This is to certify that the

dissertation entitled

A MIXED LINEAR MODEL WITH TWO-WAY CROSSED RANDOM EFFECTS AND ESTIMATION VIA THE EM ALGORITHM

presented by

SANG-JIN KANG

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in MEASUREMENT AND QUANTITATIVE RESEARCH METHODS

Stephen W. Raudenbush Ed.D.
Major professor

Date Feb. 26, 1992

A MIXED LINEAR MODEL WITH TWO-WAY CROSSED RANDOM EFFECTS AND ESTIMATION VIA THE EM ALGORITHM

By

Sang-Jin Kang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology
and Special Education

1992

ABSTRACT

A MIXED LINEAR MODEL WITH TWO-WAY CROSSED RANDOM EFFECTS AND ESTIMATION VIA THE EM ALGORITHM

By

Sang-Jin Kang

In the past two decades, there has been a prominent methodological effort in educational statistics to develop analytical methods that account for the multilevel characteristics of the educational data. As a result, methodologists have developed estimation procedures appropriate for nested multilevel data assumed normally distributed. One limitation shared by all of the new multilevel analytic approaches is that they apply only to those multilevel data which are purely hierarchical. Although educational systems typically have hierarchical organizations-students are grouped together for learning within classrooms, classrooms within schools - very often the structure of a system is not a 'pure' hierarchy. Students may belong to more than one group simultaneously. If the data are nested within the cell of two cross-classified grouping factors, we call it crossed multilevel data.

This study presented an appropriate statistical model for the analysis of crossed multilevel data for general applicability. The crossed multilevel model presented in this thesis allows us to consider simultaneously the multilevel and crossed features of higher level units. Such a model promises to increase the descriptive power and inferential

advantages for both micro-and macro-parameters in the crossed multilevel context. Research interests in this thesis were five fold: a) to conceptualize the crossed multilevel model; b) to present estimation theory for the model using the empirical Bayes viewpoint; c) to provide a computational algorithm for parameter estimation using the EM algorithm; d) to provide empirical evidences for the accuracy of the computing algorithm; e) to present the application of the model; f) to provide the substantive applicability of the model in educational research.

The conceptualization of the model was described via a crossed random effects ANOVA model and by linking it to the crossed multilevel model. Estimation theory was reviewed from an empirical Bayes viewpoint and the computational procedure implemented via the EM algorithm. The accuracy of the algorithm was tested using randomly generated data against the standard statistical packages (SAS and BMDP) and via a simulation study. An actual experimental data set was analyzed for illustration and the substantive applicability of the model in educational research was assessed.

This dissertation is dedicated to my mother, for her love and special life for the family.

To my brother, Sang-Yul Kang, and his wife, In-Sook Kim, for their support and sincere life.

To my wife, Eun-Mi, for her love, patience, and ongoing support. I love you.

To my daughters, Bona and Boyoung, who enriched my family life in East Lansing and my nephews, Bio, Bundo, and Yohan, for their growth and future lives.

To all of my other relatives and friends who have understood and have been generous while I have worked in United States.

ACKNOWLEDGEMENTS

I wish to send my heartfelt thanks to Dr. Stephen W. Raudenbush, my advisor and committee chair, who has been my teacher, role model, and provided rock of support throughout my doctoral program. His encouragement, mentoring and his faith in me greatly aided in the completion of this dissertation. Dr. Raudenbush's insight, courage, and unusual motivation of hard work will continue to be an inspiration for me.

I also wish to send my special thanks to Dr. Brian P. Rowan, who has enlightened, and mentored my professional work. Dr. Rowan has provided a constant financial support for my advanced study. Dr. Rowan's generosity and open-minded advices for my work have inspired me.

My special appreciation goes to Dr. John Schwille who provided me financial support for more than one year and has willingness to give his valuable time for my needs. It has been my luck that he was near me.

Dr. William H. Schmidt, my former advisor and a committee member, gave generously one-year research assistantship when I was a first year doctoral student. Dr. Schmidt also gave helpful comments and suggestions on my dissertation. Thanks.

I would like to send my recognition with gratitude to my other committee members. To Dr. James Stapleton, who taught me statistics from the basic to the advance, for his kind suggestions and comments. To Dr. Robert Floden for his tough questions. To Dr. Richard Houang for his willingness to use his valuable time for me.

I wish to acknowledge the professors in the Department of Education at

Yonsei University. I realize that all my success in the advanced study is rooted in their teaching.

Finally, I wish to recognize my wife, Eun-Mi, who has shared the hard life for my entire study at Michigan State University.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

CONCEPTUALIZATION OF CROSSED MULTILEVEL MODEL

Introduction

In statistical analysis, a researcher needs to specify an appropriate model that guides inquiry as well as describes the structure of data as precisely as possible. Educational systems typically have hierarchical organizations in which "units" at one level (e.g., students) are "nested" within units at the next higher level (e.g., classrooms or schools). These educational systems often produce inherently hierarchical data. The problems of analyzing hierarchical data arise when key variables of interest are measured at different levels of an organizational hierarchy. The prime question for an appropriate statistical model in this case is whether the model takes into account the effects of variables measured at both the individual and the group levels. Failure to account for hierarchies may lead to trouble in terms of research validity and has been the core in methodological criticism of educational research (see Cronbach, 1976; Burstein, 1980; Cooley, Bond, & Mao, 1981; Raudenbush, and Bryk, 1988).

Despite such fundamental warnings, many analysts used single-level models even if the key variables were measured at two different levels. This mismatch between single-level models and the multi-level data often leads the researchers to the analytic dilemmas in the choice of unit of analysis. In most educational research, students are not randomly assigned to groups such as classrooms or schools. The lack of

independence of responses is further exacerbated by the fact that students receive treatment as a group. When students share common group histories, teachers, peer experiences, their responses will be correlated.

In addition to the violation of statistical independence assumption, hierarchical data usually produce variables of interest at both student and group levels. Traditional linear models can account for only single level variables and fail to accommodate the variables measured at both levels which restricts model specification and appropriate inferences of interest as a result. Moreover researchers may raise the questions about how group processes (i.e., some policy implementation) are interrelated with the processes within the groups (i.e., student behavior) when they have hierarchical data. These questions are hardly answered by the classical linear models. Educational researchers have long been concerned with multilevel issues. But traditional research methods have not provided adequate tools with which to analyze data arising in naturally occurring hierarchies.

New advances on model specification for multilevel data were made by the methodologists who have advised researchers to formulate explicit, multi-level models which enable testing of hypotheses about the effects occurring within each level and interrelations among them (Burstein, 1980; Cooley, Bond,& Mao, 1981; Rogosa, 1978). A number of methodologists, working independently, have developed estimation procedures appropriate for hierarchical, or multilevel data assumed normally distributed (Aitkin & Longford, 1986; Goldstein, 1986; DeLeeuw & Kreft, 1986; Mason, Wong,& Entwisle, 1984; Raudenbush & Bryk, 1986). They made a substantial methodological advances in two key reasons

(Raudenbush,1988):

"First, such methods enable researchers to formulate and test explicit statistical models for process occurring within and between educational units. These models solve, in principle, the problem of aggregation bias .... Second, these models enable researchers to specify an appropriate error structure, including random intercepts and random coefficients .... In most settings, appropriate specification of error components solves the problems of misestimated precision which have plagued hypothesis testing in nested unbalanced data sets."(p.86).

One limitation shared by all of the new multilevel analytic approaches is that they apply only to those multilevel data structures which are purely hierarchical. Although educational systems typically have a hierarchical organization - students are grouped together for learning within classrooms, classrooms within schools - very often the structure of a system is not a "pure" hierarchy. Students may belong to more than one group simultaneously. For example, all students are members of neighborhoods as well as of schools and not all students in the same neighborhood are in the same school and vice versa. Therefore schools and neighborhoods in this case are crossed with different number of students or no students within each cell classified by the two factors. If such a classification of students is possible, again a statistical model must reflect the crossed structure so that a researcher may investigate the variables at both higher level units, neighborhoods and schools. Because students are "nested" within cells of a neighborhoods by schools cross-classification, the structure of data is cross-classified as well as hierarchical. Also both neighborhood and school effects are considered random because we can assume the two higher

level

Goldst

hierar

Fo

pose

crosse

issue

Commi

1971;

appli

varia

which

and

contr

corp

the

mixe

mode

appl

blo

des

by

foo

deg

exa

ar

level units are randomly selected from larger populations of the two. Goldstein (1987) called this "crossed-multilevel" structure while the hierarchical data is "nested-multilevel".

For the analysis of crossed-multilevel data, a researcher ought to pose a model that considers simultaneously both the multilevel and crossed features of data. Many researchers implicitly worked on this issue in the research on variance component analysis (Henderson, 1953; Cunningham & Henderson, 1968; Hartley & Rao, 1967; Patterson & Thompson, 1971; LaMotte, 1973). The results from these researchers, however, are applicable under restrictive conditions, as when no continuous group variables are available, and there was no multilevel model formulation which enables estimation of covariance components of micro-parameters and testing the effects of crosslevel interactions in multilevel contexts. Their results are limited to the estimation of variance components.

Researchers working on multilevel models have already anticipated the analysis of crossed-multilevel data in the view point of general mixed model. Only a few computational attempts were made with simple models. Lindley and Smith (1972) showed how Bayes estimation can be applied to the analysis of crossed-multilevel data with a randomized block design. But their attempt was limited to the case of balanced design with one observation per cell. A more complete attempt was made by Dempster, Rubin, and Tsutakawa (1981) in the example of "Professional football scores" where the game scores were nested within offensive and defensive effects of the teams involved in each game. These two examples, however, do not lead directly to a general crossed-multilevel model. Rather researchers need to understand these exemplary results as

pointing to the possible emergence of a general model. Unfortunately we do not have such a general crossed-multilevel model where any number of fixed and random effects, in principle, can be estimated at both within and between unit level along with the interaction effects for both balanced and unbalanced designs.

Crossed-multilevel data are both common and poorly understood in educational research. Consider a school testing program. Students in each grade level receive multiple tests. Here students are nested within the cell of grade levels by test forms; grade level effects and test form effects are viewed as random; grade levels and test forms are crossed. Again suppose schools have an annual testing program for graduates. Students, in this case, are nested within schools as well as years; school effects and year effects may be viewed as random; and both schools and years are crossed.

Objectives

This research attempts to achieve six objectives corresponding to the six chapters of the thesis:

1. Conceptualization of crossed-multilevel model is described in chapter 1 through two parts. The first part provides background information for crossed multilevel model. The second part describes the mathematical model for crossed multilevel analysis and considers the research inquiries that it makes possible.

2. Estimation theory for the general crossed-multilevel model is presented via the empirical Bayes viewpoint for both general and mixed linear models in chapter 2.

3. A computational algorithm is provided for the estimation of

parameters in the posed model using the EM Algorithm (Dempster, Laird, and Rubin, 1977) in chapter 3.

4. A computer program is developed in the Gauss language (version 2.0), and accuracy of the algorithm is examined in chapter 4.

5. Application of the crossed multilevel model to real data set is presented through illustration in chapter 5.

6. The properties of the estimators and value of the study is discussed with summary in chapter 6.

Problems with Multilevel Contexts

As mentioned, most educational research has tended to ignore the nesting of individuals within groups and, instead, used single-level models to analyze the multilevel data. This mismatch between single-level models and multilevel data causes the following major problems.

1. Unit of Analysis

Analyzing multilevel data with single-level models often leads the researchers to the analytic dilemmas in the choice of unit of analysis. Researchers have stated that significance tests based on individual-level analysis are unacceptable, due to the violation of the independence assumption when subjects receive treatments as a group, even if individuals are randomly assigned to groups (Cronbach, 1975; Glass & Stanley, 1970; Page, 1975; Aitkin, Anderson, & Hinde, 1981; Knapp, 1977; Walsh, 1947). Inferences about individual behavior, such as individual achievement, based on group level analysis can cause aggregation bias (Cronbach & Webb, 1975; Robinson, 1950; Hopkins, 1982). Some investigators may choose separate analysis at each level in the hope that

the cor

critici

provide

Cron

"choice

multil

intere

2. Con

In

need

resea

organ

imple

of th

due

achie

stud

ever

stud

scho

be

con

sch

dis

sch

the convergent results of the two analyses would avoid the methodological criticism. Raudenbush and Bryk (1988) have stated that this strategy provides no guidelines on how to interpret the results when they diverge.

Cronbach (1976), Burstein (1980) and others have recognized that "choice of unit of analysis" is the wrong question for analyzing multilevel data because the variation at each level is potentially of interest and ought not be ignored.

2. Confounding Factors

In the multilevel contexts, the variables measured at different levels need to be taken into account in the model specification. Educational researchers are often interested in the association between school organization and student achievement. Consider a school district that implements a certain policy to improve student achievement. The effect of the policy on student achievement may be different across students due to the differences in student demographic backgrounds such as prior achievement and motivation, variables defined at student-level. If such student differences are ignored, one must assume the policy effects are evenly distributed across the students regardless the differences of student characteristics, which is h rdly acceptable. The effect of school organizational variables, i.e. policy variables, may also tend to be different across schools due to the differences of the contextual conditions across the schools. For example, schools may differ in their school climates such as principal leadership, staff cooperation, student disorder, and in socioeconomic levels of the communities in which schools are located.

Whenever we study the effect of organizational variables on students

or teachers nested in the organization, we must adjust for confounding effects that could occur at both the individual level and the organizational level. Otherwise the estimates of the effects of the organizational variables will be biased. The restriction of the standard single-level linear model for the analysis of multilevel data is that it does not simultaneously control the variables defined at different levels.

3. Cross-Level Interactions

The effects of organizational variables may depend on variables defined at different levels of the processes of the organization. For example, the fast pacing of the curriculum is beneficial for high-aptitude students but not for low-aptitude student (Gamoran, 1991), an example of an aptitude-by-treatment interaction (Cronbach and Snow, 1977). Such interaction is a "cross-level interaction" (Bryk & Raudenbush, 1989) in terms of the multilevel viewpoint because a variable measured at higher level (i.e. pacing of the curriculum) interacts with a variable measured at a lower level (i.e. student aptitude).

Cross-level interaction effects may have important implications for the formulation of scientific theories because the effects constrain the generalizability of findings. Cronbach and Webb (1975) realized that the characteristics of the classroom interacted with individual backgrounds. Such interactions rendered the treatment effect unique in each study and constrained the generalizability of a finding across the settings with the treatment.

The lack of generalizability of organizational effects may often be

found in the studies of school effects because the effect of a certain policy or reform may depends on the characteristics of students or on the contextual conditions of the schools. Traditional single-level models do not allow the investigation of the cross-level interactions because they must assume that effects are homogeneous across the groups.

Educational Research and Crossed Multilevel Data

Crossed Multilevel Contexts

Consider the three educational research cases below which require the analysis of crossed multilevel data.

Case 1. Assessment of School Effects A number of district administrators want to diagnose the secondary high school education of their districts and therefore implement an annual testing program for the graduates. The administrators want to know whether the schools provide stable education with good quality, (i.e., excellence), in terms of graduates test scores across the years. If the test scores of the graduates are not stable across the years then they want to describe the changes. The administrators also want to know how much the schools differ in the test scores of the graduates. If the schools differ substantially in test scores, then they hope to see what characteristics of the schools are determinants of such differences. Another important question is whether the effects of student background characteristics are homogeneous across schools and years. Are outcomes more or less equitably distributed in different schools? Is the equity of distribution stable over time? Available data to the administrators in this case are graduates test scores in each year, students demographic

backgr

teache

Case 2

adminis

systems

primary

adminis

element

at seco

the se

achieve

charact

achieve

serve

data to

schools

charact

Case 3.

evaluat

test (R

supposed

is expe

after

exhauste

whether

background variables, and school characteristics (e.g., school size, teacher/student ratio).

Case 2. Relationaship between the schooling levels Again the administrators want to get fundamental information about the schooling systems of the districts. In schooling systems, children attending a primary school may go to one of several secondary schools. The administrators need information about what characteristics of the elementary schools support the long-term progress of student achievement at secondary high schools. They also need to know the characteristics of the secondary high schools that have positive or negative effects on achievement. The more interesting question may be what combination of characteristics of the elementary and secondary schools supports student achievement. Another question is how well schools of different types serve children of different ethnic and social background. Available data to the administrator are achievement test scores in elementary schools and secondary schools, student background variables, and characteristics of the schools at both levels.

Case 3. Evaluation of standardized tests A psychometrician wants to evaluate a standardized achievement test that was developed as a power test (Rudman & Raudenbush, 1987). Unlike a speed test, a power test is supposed to be insensitive to the duration of test administration. It is expected that examinees can not answer further questions correctly after the prescribed test time because their knowledge would be exhausted after that elapsed time. Research hypotheses in this case are whether there are the effects of excess test time on examinees' test

scores.

The psychometrician first blocks classrooms on background characteristics and then randomly assigns classrooms to excess time conditions. Hence children are nested within cells of a blocks-by-treatment cross-classification. Both treatment and blocks are conceived as random. Available data to the psychometrician are student demographic variables, teacher characteristics, students' prior achievement, and the standard test scores.

Analytical Commonalities

The design characteristics of the above three research cases share four points. First, data of the all studies are basically multilevel. The studies use the variables measured at both individual and at higher levels. Second, the higher-level units of the studies are all considered as random. In the first case, we consider the sample schools as selected from the larger population and the years of the test administration are also random. In the second case, the elementary schools and secondary high schools are all considered as random. The third case uses a randomized block design where the block effects are typically viewed as random and the categories of different excess test times are viewed as random as long as the researcher increases the number of conditions.

Third, each study uses two kinds of higher level units and the two factors are crossed. The first example shows that graduates are nested within cells classified by the schools and years. In the second example, students have dual membership of their elementary schools and secondary high schools. The current organization of schools implies that primary

schools will not be nested within secondary schools. Rather the two factors will be crossed. The last example utilizes the block design where the blocks are crossed with the categories of different time allocations for testing.

Finally, the data in every case are unbalanced. In the first example, the number of students are not the same across the years as well as across the schools. This unbalanced character of the sample size may be more serious in the second example. Students attending a secondary high school do not come from all elementary schools. Again, the number of students from each elementary school is different in each high school. Therefore, in a natural situation, the numbers of students classified into each elementary-by-secondary cell are unequal and there will be many missing cells. A similar situation occurs in the third example. In education, experimental studies usually use volunteers as the source of data. The class sizes will vary, meaning that cell sizes will vary. Also not every block may contain the same number of students.

In sum, each of the three studies has two crossed random factors within which individuals are nested. Because the higher level units are crossed and the lower level units are nested within cells in the two-way classification of the higher level units, we call the data "crossed-multilevel" (Goldstein, 1987[a]). The appropriate model should then consider simultaneously both the multilevel and crossed features of data. The multilevel and crossed features require the specification of four random effects in a model: random individual effects, random effects of each of the two higher level units, and random interaction effects. In addition, individual scores within each cell can be described as a function of multiple individual characteristics; and some of the within-

cell effects vary significantly over rows and columns while the others may vary only over rows.

Methodological Advantages of Two-Way Crossed Multilevel Models

The limitation of the current available multilevel models for the analysis of crossed multilevel data is that those models have been developed for the analysis of one-way nested data. Those models are appropriate to the analysis of nested multilevel data, but only when one level is purely nested within the other level. The nested multilevel model allows estimation of covariance-components at each level.

For the analysis of crossed multilevel data, a model should estimate the covariance components at each level but it should also decompose the covariance-components at higher levels into three components corresponding to specified random effects for two higher-level units and for their interactions. The model should handle with unbalanced data and covariates having either fixed or random effects. We now consider the additional information obtainable from crossed-multilevel models in connection with the previous three research examples.

1. Covariance-components decomposition

The crossed multilevel model allows specification of the appropriate error terms in a model where schools and years are crossed. For the first example, the crossed multilevel model can estimate a within-cell variance, σ^2, and between-cell covariance components. The between-cell covariance-components are decomposed into the three parts; one is the covariance matrix for school residual effect, say τ_a, the

other is for year residual effect, say τ_b, and for the residual interaction effect between schools and years, τ_c.

With these covariance components we could get various intra-unit correlations which inform us of the proportion of the total observed variance that lies within cells, between rows, between columns and between cells. We could also do a statistical test of the significance of random variation for each effect. These two pieces of information, the intra-unit correlations and the χ^2-test, provide us both practical and statistical information about further model specification. For example, the presence of a large intra-school correlation informs us that a large percentage of variation of the student scores lies among schools. The significant results of a χ^2-test for school residual dispersion effect, τ_a, informs us that there still remains a significant random residual effect not explained by the school characteristics in the model. Similar information is obtainable for the analyses of the two other examples.

2. Random interaction effects

We return now to the case in which primary and secondary schools are crossed. A major difference between the nested multilevel model and crossed multilevel model is that crossed multilevel model captures the crossed features of the two higher level units while the nested multilevel models do not.

Because the crossed multilevel model allows the specification of random interaction effects of the two higher level units, primary and secondary schools in the second example, it produces estimates of the covariance components of random interaction effects of the two higher

level units. The presence of substantial random interaction effects tells us that certain primary school effects depend upon the secondary school attended; or, equivalently that the effects of attending a particular secondary school depend on the primary school attended. If we find a significant interaction effect of a particular primary school characteristic and a particular secondary school characteristic on student achievement, then the effect of the involved school characteristic can't be generalized to the entire schooling system.

3. Statistical precision

Imagine that a researcher has applied a nested-multilevel model to analyze multilevel data with a two-way classification. He has performed the analysis as a compromise because a crossed-multilevel model was not available. The nested-multilevel model specifies a single error term for higher level units. There are, however, three error sources in the higher level units: two higher level units (e.g., schools, years) and their interactions.

The estimates of higher level covariance components in his analysis are then the sum of the three covariance components from each error source. In the multilevel contexts, the estimates of fixed and random effects are the functions of covariance components at both individual and group levels. Nested multilevel models use the sum of covariance components of the higher level units for parameter estimation without knowing the size of covariance components from each error source. Although the exact functional relationship between the nested and crossed multilevel models regarding the parameter estimation is not clear to the author at this point, the author suspects the precision of

the parameter estimation may be sub-optimal when the nested multilevel model is applied to the analysis of crossed multilevel data. Consider the case in which primary and secondary schools are crossed. Again suppose we are estimating the fixed effects of primary and secondary school characteristics. Then nested multilevel model uniformly applies the sum of covariance components for estimating the fixed effects of both primar and secondary schools while the size of covariance components pertinent to each higher level units are different.

Crossed multilevel models allow specification of the error terms precisely for the analysis of nested data with two-way classification and properly decomposes the covariance components at higher level. Hence it uses the decomposed covariance components at higher level for parameter estimation of the model.

Past Methodological Research on Crossed Multilevel Models

Standard texts on experimental design provide methods for simple crossed random effects models in which a single random component is associated with each cell in a fully balanced two-way cross-classification (see, for example, Kirk, 1982). However, as in the above examples, the interesting designs in education will typically be substantially unbalanced. For example, many cells in a neighborhood-by-schools cross-classification will be empty or small; and enrollments will vary in a schools-by-time cross-classification.

Researchers working on variance component analysis have tried to elaborate models in order to meet various conditions (Henderson, 1953; Cunningham & Henderson, 1968; Hartley & Rao, 1967; Patterson & Thompson, 1971; LaMotte, 1973). The results from these researchers, however, are

applicable under restrictive conditions, as when no continuous group variables are available and their results are limited to the estimation of variance components.

Harville (1977) reviews methods of variance components estimation based on maximum likelihood for unbalanced designs with crossed random factors. However, although these methods are appropriate for variance components, they do not allow covariance components, which will often be of interest in ducation. For example, regression coefficients describing the relationship between social background and achievement may vary across schools or neighborhoods, or across time.

Lindley and Smith (1972) showed how Bayes estimation can be applied to the analysis of crossed-multilevel data with a randomized block design. But their attempt was limited to the case of balanced designs with one observation per cell. Dempster, Rubin, and Tsutakawa (1981) provided numerical results in the example of "Professional football scores" where the game scores were nested within offensive and defensive effects of the teams involved in each game. Dempster et al. provided maximum likelihood estimates using the EM algorithm and explained the empirical Bayes estimation theory. These two examples are useful; however, they do not lead directly to the general crossed multilevel model.

Multilevel methodologists anticipated the analysis of crossed-multilevel data in the viewpoint of the general mixed model. In this stream, Goldstein (1987[b]) outlined a procedure based on iterative generalized least squares for statistical estimation in crossed-random covariance components. He, however, supplied no computational strategy.

The Two-Way Crossed Multilevel Model

In this section, the two-way crossed-multilevel model is conceptualized and formulated. In order to promote the conce tual understanding of the crossed-multilevel model, we first consider crossed-random effects ANOVA model and formulate the crossed multilevel model from the ANOVA model. Then the general crossed multilevel model will be presented in two forms: matrix form, and no-subscript form.

The Two-Way Crossed Random Effect ANOVA Model

Suppose a researcher wants to investigate the effects of schools and neighborhoods on students' achievement. He may first select J schools and obtain the information about students' K residential areas. The framework of this design shows; a) students are nested within the cells of a schools-by-neighborhoods cross-classification, b) school effects on students achievement are random, c) neighborhood effects are random.

From this description, we may estimate some or all of the following effects; a) random main effects of schools, b) random main effects of neighborhoods, c) fixed effects of school characteristics, d) fixed effects of neighborhood characteristics, e) random interaction effects (school by neighborhood), and f) fixed interaction effects (fixed school-by-fixed neighborhood predictors).

If we consider just the random effects, we can specify a statistical model as below

$$Y_{ijk} = \mu + a_j + b_k + c_{jk} + e_{i:jk} \qquad (1.1)$$

for $i = 1,2,\ldots,n_{jk}$, $j = 1,2,\ldots,J_k$, $k = 1,2,\ldots,K_j$; and Y_{ijk} is ith student score in school j in neighborhood k; μ is grand mean of all students in the population; a_j is random effect of school j; b_k is random effect of neighborhood k; c_{jk} is random interaction effect between school j and neighborhood k; and finally $e_{i:jk}$ is a random error. The notation "i:jk" means student i is nested in school j and neighborhood k. The subnotation of the cell size, n_{jk}, the number of schools, J_k, and the number of neighborhood, K_j, imply the design characteristics which allow unbalanced and incomplete data. Since all effects are random, we have a series of assumptions: $a_j \sim N(0, \sigma_a^2)$; $b_k \sim N(0, \sigma_b^2)$; $c_{jk} \sim N(0, \sigma_c^2)$; $e_{i:jk} \sim N(0, \sigma_e^2)$ are all required assumptions of Equation (1.1). The effects of a_j, b_k, c_{jk}, and $e_{i:jk}$ are mutually independent.

Equation (1.1) considers only the random effects. Random main effects may or may not be of central interest for the inquiry of a researcher. A researcher interested how much of the variance among students' scores is attributable to the differences among schools and neighborhoods, may use this model for analyzing data from the above design.

Very often, researchers inquire about the effects of school and neighborhood characteristics on student achievement and how students scores in a school differ across neighborhoods. Equation (1.1), assuming all predictors are centered around their means, can be elaborated into the model below in this case:

$$Y_{ijk} = \mu + \beta_1 X_{1j} + u_j + \beta_2 X_{2k} + v_k + \beta_3 X_{3jk} + \pi_{jk} + e_{i:jk} \qquad (1.2),$$

where β_1 is fixed effect of school characteristic, X_1 (e.g., school size); β_2 is fixed effect of a neighborhood characteristic, X_2 (e.g.,community SES level); and β_3 is the fixed effect of the interaction between X_1 and X_2. Assuming X_1, X_2, and X_3 are orthogonal, the random school effect, a_j, in Equation (1.1) is decomposed into two parts in Equation (1.2); $a_j = \beta_1 X_{1j} + u_j$. Thus u_j is a residual random school effect after accounting for the fixed school effect, $\beta_1 X_{1j}$. Similarly, the random neighborhood effect, b_k, and interaction effect c_{jk} of Equation (1) are also decomposed into two parts respectively; $b_k = \beta_2 X_{2k} + v_k$, and $c_{jk} = \beta_3 X_{3jk} + \pi_{jk}$. This model needs assumptions for the random terms: $u_j \sim N(0, \sigma_u^2)$, $v_k \sim N(0, \sigma_v^2)$, $\pi_{jk} \sim N(0, \sigma_\pi^2)$, and $e_{ijk} \sim N(0, \sigma_e^2)$. Other alternative model specifications are possible if a researcher has different aims of inquiry. For example, a researcher may utilize covariates with students level variables, say X_{4ijk} (e.g., prior achievement) centered around the mean of group jk. Then the student level random error can be decomposed into two parts; $e_{i:jk} = \beta_4 X_{4ijk} + f_{i:jk}$, where β_4 is the fixed effect of student prior achievement level, X_4. Whatever model a researcher poses, the model tells us that all specified effects in the model need to be estimated.

We call Equation (1.1) a crossed-random effect ANOVA model which can be understood as multilevel model because it takes into account the contributions from group level units (the a_j, b_k, c_{jk}) and individual effect, $e_{i:jk}$. Equation (1.2) is also considered as a multilevel model because there are two kinds random terms; u_j, v_k, and π_{jk} are group level residuals and $e_{i:jk}$ is individual level residual. These two models show the basic features of crossed-random effect models and how

we specify a model of interest. Because these two models incorporate both individual and group level random effects, we can get considerable information from the random effect dispersion matrix through variance decomposition.

Random Effect Dispersion Matrix and Variance Decomposition

Equation (1.2) contains four fixed parameters; μ, β_1, β_2, β_3 and the random variables; u_j, v_k, π_{jk}, $e_{i:jk}$. Here we assume the model is additive and that the four random variables are mutually independent each other.

The variance of Y_{ijk}, for given fixed effects, is

$$\text{Var}\,(Y_{ijk} \mid \mu, \beta_1 X_{1j}, \beta_2 X_{2k}, \beta_3 X_{3jk}) = \sigma_u^2 + \sigma_v^2 + \sigma_\pi^2 + \sigma_e^2,$$

and the conditional covariance between students' score within the cell of school by neighborhood is

$$\text{Cov}\,(u_j+v_k+\pi_{jk}+e_{i:jk}, u_j+v_k+\pi_{jk}+e'_{i:jk}) = \sigma_u^2 + \sigma_v^2 + \sigma_\pi^2.$$

Hence the correlation is given by definition,

$$\rho = (\sigma_u^2 + \sigma_v^2 + \sigma_\pi^2)/(\sigma_u^2 + \sigma_v^2 + \sigma_\pi^2 + \sigma_e^2)$$

which we can consider "intra-cell" correlation. This correlation itself shows the proportion of between-cell variances over the total variance. Thus it tells us the proportion of variance attributable to variation between cells of school-by-neighborhood.

Again the conditional covariances for students in different schools but the same neighborhood is

$$\mathrm{Cov}(u_j+v_k+\pi_{jk}+e_{i:jk},\ u'_j+v_k+\pi'_{jk}+e'_{i:jk}) = \sigma_v^2$$

and for students in different neighborhood but the same school is

$$\mathrm{Cov}(u_j+v_k+\pi_{jk}+e_{i:jk},\ u_j+v'_k+\pi'_{jk}+e'_{i:jk}) = \sigma_u^2$$

From these covariances we can get two more "intra-unit" correlations; the "intra-neighborhood" correlation,

$$\rho = \sigma_v^2 \,/\, (\sigma_u^2+\sigma_v^2+\sigma_\pi^2+\sigma_e^2)$$

and the intra-school correlation

$$\rho = \sigma_u^2 \,/\, (\sigma_u^2+\sigma_v^2+\sigma_\pi^2+\sigma_e^2).$$

Again each intra-correlation shows the proportion of total variance between students which is due to the differences between neighborhood or between schools respectively.

Now we have all necessary information to form the full dispersion matrix, V, for total sample. The full variance-covariance matrix, V, contains cell matrices, V_{jk}, of school j and year k where there are n_{jk} students. We can form these cell matrices which can be classified into four types. The first type of cell matrices represent the variance-covariance matrices for those students in the same school and

neighborhood. These matrices take the diagonal positions of the full matrix, V. This matrix contains the variances of students within a cell, $\sigma_u^2+\sigma_v^2+\sigma_\pi^2+\sigma_e^2$, as the diagonal elements, and the covariances, $\sigma_u^2+\sigma_v^2+\sigma_\pi^2$, as the off-diagonal elements. For off-diagonal block matrices of the full matrix of V, there are three kinds of block matrices. One represents the blocks which consider the students in the same school but in the different neighborhood. In this case all elements are σ_u^2. The other one represents the students in the same neighborhood but different schools. This block contains σ_v^2 in all elements. The last type of block matrices are for those students who have different neighborhood and school membership. This last type of matrices are null. Thus the full matrix, V, has somewhat complex structure, but is composed of only four kinds of block matrices; one kind of diagonal block matrices and three kinds off-diagonal block matrices.

Crossed Multilevel Model Formulation

The crossed random effects ANOVA model presented at the previous section is now reformulated into the crossed-multilevel model. Such reformulation will show the logic of crossed-multilevel modelling.

Base Model

Recall the Equation (1.1) which reflects the research design where students are nested within the cells of a schools-by-neighborhoods cross-classification and both schools' and neighborhoods' effects are random. Since Equation (1.1) has a hierarchical structure, it can be represented into two stage formulae in terms of within-cell and between-

cell models. The within-cell model is

$$Y_{ijk} = \beta_{jk} + e_{ijk} \qquad (1.3)$$

where the individual score, Y_{ijk}, is composed of the average score of the cell classified by jth school and kth neighborhood plus an individual effect e_{ijk} - $N(0, \sigma^2)$. Equation (1.3) is the traditional regression model with no predictors in the model except that β_{jk} is allowed to vary randomly. Thus we pose between-cell model to explain the variation of the regression coefficients in the within-cell model. The mean score of a cell, β_{jk}, can be decomposed into the grand mean, μ, across all cells plus the effects of school j, neighborhood k, and their interactions. Thus the between-cell model is specified as

$$\beta_{jk} = \mu + a_j + b_k + c_{jk} \qquad (1.4)$$

where μ is the grand mean of all cell means, a_j is the effect of school j, b_k is the effect of neighborhood k, c_{jk} is the effect of cell jk. Since the effects of school, neighborhood, and their interactions are all random, the distributional assumptions are; a_j - $N(0, \tau_a)$, b_k - $N(0, \tau_b)$, and c_{jk} - $N(0, \tau_c)$. If we substitute the Equation (1.4) into the Equation (1.3) we get the Equation (1.1), crossed random effect ANOVA model. So the two models are identical.

Here the random factor a_j reflects the variation of cell means that is attributed to the effect of a school. Thus τ_a is understood as a true parameter variance due to differences between schools. A similar meaning is applied to the τ_b and τ_c. Using this crossed-multilevel

model we can get various information. First we can test the hypothesis that H_0: $\mu = 0$ (i.e. the grand mean is zero), but it is not usually of interest in this model. We also pose hypotheses about the parameter variances as well. The null hypotheses are: $var(a_j) = 0$; $var(b_k) = 0$; and $var(c_{jk}) = 0$.

Study Model

The basic rationale of model specification of crossed-multilevel model for each study is the same as in the other available multilevel models (i.e. H M). First, we examine the variability among students in the hierarchical structure via the base model. Second, we specify the within-cell model only to reduce the within-cell variation and no predictor variables are used at between-cell model. In this stage we can determine whether the effects of the within-cell variables (i.e., student variables) are random or fixed across the higher level units by examining the intra-unit correlations of each within-cell slope. If the intra-unit correlations of a certain within-cell slope is close to zero then we can fix the effect of that predictor and the within-cell model become a mixed linear model. Third, after completion of the within-cell model specification we start the specification of between-cell model in order to identify variability among higher level units as a function of between-unit variables. The criterion of the model specification is a coefficient of determination, R^2, which has similar meaning in the regression analysis. While R^2 means the proportion of the explained variation by a model given the total observed variation in regression analysis, the R^2 in crossed-multilevel model is obtained based on the true parameter variances. Therefore the meaning of R^2 in this case is

that the proportion of explained true variance by a model given the total parameter variances from the model based on no predictors.

Suppose we complete the model specification via the above three steps and have a final model where two within cell variables are employed but one of them has fixed effect, i.e. pretest score, while the intercept is random across the groups and the remaining student variable (e.g., SES level) has random effects across the schools only. Again suppose one school variable (e.g., school size) and one neighborhood variable (e.g., crime rate) are identified as an significant predictors in the final model. By using matrix form, the within-cell model is

$$Y_{jk} = X_{jk} \beta_{jk} + e_{jk} \qquad (1.5)$$

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}_{jk} = \begin{bmatrix} 1 & X_{11} & X_{21} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & X_{1n} & X_{2n} \end{bmatrix}_{jk} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{jk} + \begin{bmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix}_{jk}$$

where Y_{jk} is a vector of students' posttest scores, X_{jk} is the matrix of student predictors where X_1 is students SES level and X_2 is the pretest scores that have fixed effects. Since the within-cell model has both fixed and random effects it is a mixed linear model itself. In the following between-cell model we need to model the within-cell parameters, the β's, as a function of the group level variables. The between-cell model is

$$\beta_{jk} = W_{jk} \Gamma$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{jk} = \begin{bmatrix} 1 & w_1 & w_2 & w_3 & 0 & 0 & 0 \\ 0 & 0^1 & 0^2 & 0^3 & 1 & w_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0^1 & 1 \end{bmatrix}_{jk} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{02} \\ \gamma_{03} \\ \gamma_{10} \\ \gamma_{11} \\ \gamma_{20} \end{bmatrix}$$

$$+ R_{jk} a_j + C_{jk} b_k + T_{jk} c_{jk}$$

$$+ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}_{jk} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}_j \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{jk} [\, b_0 \,]_k + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{jk} [\, c_0 \,]_{jk} \qquad (1.6)$$

Here β_{jk} is a vector of within cell parameters; W_{jk} is the matrix of group level variables; Γ is the matrix of fixed effects of group level predictors; R_{jk}, C_{jk}, and T_{jk} are the selection matrices for random residual effects; a_j, b_k, and c_{jk} are the random residual effect vectors with distributions $a_j \sim N(0, \tau_a)$, $b_k \sim N(0, \tau_b)$, $c_{jk} \sim N(0, \tau_c)$.

An interesting feature of the above Equation is the presence of R_{jk}, C_{jk}, and T_{jk}, which serve for identifying the residual random effects. If the all three within-cell parameters, β_0, β_1, β_2, are random across the schools and neighborhoods including the interactions between the two factors, then the three selection matrices, R_{jk}, C_{jk}, and T_{jk}, are all identity matrices with the dimension of the number of within-cell parameters. If one represents the between-cell model in a scalar form, one will see how the within cell parameter, β_2, has a fixed effect. These three selection matrices are useful for a practical purpose. Analysts may need to set some particular residual effects to zero to fix the relevant variable effects on outcome. If the three selection matrices are null, then analysts are assuming that the within-cell slopes are all fixed and the crossed-multilevel model become the

standard ordinary least squares regression equation with single variance component, σ^2.

Another interesting feature of Equation (1.6) is the format of W_{jk} which shows a different set of predictors for each within-cell slope. Here, W_1, W_2, and W_3 are the school size, crime rate, and their interactions of the jkth cell respectively. Because the within-cell slope, β_2, has no group level variation, it has been fixed and no group level predictors are included in the matrix of group level variables. Thus the model has great flexibility in model specification by allowing a different set of predictors for each within-cell slope.

The above between-cell model is a multivariate model because there are multiple outcome variables (β's) for each cell, jk. If we combine the two equations by substituting the Equation (1.6) into the Equation (1.5), the model becomes

$$Y_{jk} = X_{jk} W_{jk} \Gamma + X_{ajk} a_j + X_{bjk} b_k + X_{cjk} c_{jk} + e_{jk} \qquad (1.7)$$

where $X_{ajk} = X_{jk}R_{jk}$; $X_{bjk} = X_{jk}C_{jk}$; $X_{cjk} = X_{jk}T_{jk}$. Equation (1.7) is considered as a mixed linear model where the first term of the right hand side is a fixed portion and and remaining terms are all random.

General Crossed Multilevel Model

The crossed random effect ANOVA model and reformulation of it into the crossed multilevel model presented above explicitly describes the design characteristics of the data and allow specification of appropriate error structures that solve the problem of misestimated precision. We shall now see a general form of crossed-multilevel

model. To make the presentation concrete, suppose again that we wish to estimate a regression equation for each of many cells classified by schools and neighborhoods. Our aim is to discover whether these regression equations differ across the cells and , if they do, to explore the reasons why they vary. To the extent that these regression parameters do vary, we want to ask: what school and neighborhood characteristics are associated with variation in these regression coefficients? To investigate this kind of problem we formulate a multilevel model which is composed of two submodels: a within-cell model and a between-cell model. The parameters of the within cell model are conceived as outcome variables in a between cell model. After formulation of multilevel model we will see how the model can be viewed as a special case of general mixed linear model.

The presentation of crossed multilevel model is ordered in two forms according to its generalization: 1. matrix form with subscript 2. no subscript form.

Matrix Form with subscripts

The regression formulas provided in the previous section are useful for seeing the exact structure of the equations. The formulation of a general model expands notation and facilitates derivation of the estimation formula.

We first write the within-cell model which corresponds to the equation (1.5) as

$$Y_{jk} = X_{jk}\beta_{jk} + e_{jk}, \qquad (1.8)$$

where $Y_{jk} = [Y_{1jk}, \ldots, Y_{njk}]^T$ is an n_{jk} by 1 vector of achievement scores;

$\beta_{jk} = [\beta_{0jk}, \ldots, \beta_{pjk}, \ldots, \beta_{r-1jk}]^T$ is an r by 1 vector of micro parameters;

$e_{jk} = [e_{1jk}, \ldots, e_{njk}]^T$ is the n_{jk} by 1 vector of random errors assumed normally distributed with a mean vector of zero and dispersion matrix Ψ_{jk};

and

$$X_{jk} = \begin{bmatrix} 1 & X_{11jk} & \cdots & X_{r-1,1jk} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1njk} & \cdots & X_{r-1,njk} \end{bmatrix}.$$

The between-cell model in matrix notation corresponding to the equation (1.6) is

$$\beta_{jk} = W_{jk}\gamma + R_{jk}a_j + C_{jk}b_k + T_{jk}c_{jk} \qquad (1.9)$$

where $\gamma = [\gamma_0^T, \ldots, \gamma_p^T, \ldots, \gamma_{r-1}^T]^T$ and $\gamma_p^T = [\gamma_{p0}, \gamma_{pa1}, \cdots \gamma_{pas}, \gamma_{pb1}, \ldots, \gamma_{pbt}, \gamma_{pc1}, \ldots, \gamma_{pcu}]^T$. The elements of γ_p^T are the parameters capturing the structural relationship between the p th within-cell slope and the predictor variables of schools, neighborhoods and their interactions.

$a_j = [a_0, \ldots, a_{r1-1}]_j^T$ and $a_j \sim N(0, \tau_a)$;

$b_k = [b_0, \ldots, b_{r2-1}]_k^T$ and $b_k \sim N(0, \tau_b)$;

$c_{jk} = [c_0,\ldots,c_{r3-1}]_{jk}^T$ and $c_{jk} \sim N(0,\tau_c)$.

The dispersion matrices τ_a, τ_b, and τ_c are all full symmetric matrices with the dimension of r1, r2, and r3 which reflect the number of micro-parameters that vary randomly across the schools, neighborhoods and school-by-neighborhood interactions respectively. The matrix of group level variables is

$$W_{jk} = \begin{bmatrix} 1\ W_{0alj}\cdots W_{0cujk} & & & \\ & \ddots & & 0 \\ & & 1\ W_{palj}\cdots W_{pcujk} & \\ 0 & & & \ddots \\ & & & 1\ W_{r-1alj}\cdots W_{r-1cujk} \end{bmatrix} .$$

The elements of W_{jk} are somewhat complex, because the predictors are chosen form a variety of resources; schools, neighborhoods and the interactions of the two macro units. For example, the first element, W_{0alj}, means that it is the value of first school characteristic, a_1, at school j, predicting the intercept of within-cell model, β_0. The design characteristics of W_{jk} allows a different set of predictors for each within-cell slope.

R_{jk} is a random school effect indicator matrix with the dimension of r x r_1, where r is the number of all micro parameters and r_1 is the number of micro parameters that vary randomly across the schools. C_{jk} is a random neighborhood effect indicator matrix with the dimension of r x r_2, where r_2 is the number of micro parameters of within-cell model that vary randomly across the neighborhoods. T_{jk} is a interaction random effects indicator matrix with the dimension of r x r_3, and r_3 is

the number of randomly varying micro parameters across the cells.

W_{jk} becomes a partitioned matrix with block diagonal structure where the row vectors are stacked along the main diagonal. γ is constructed of subvectors, one for each of the r outcomes. The subvectors are "stacked" on top of each other. The total number of elements of γ is F, where $F = \Sigma F_p$ and $F_p = 1 + s_p + t_p + u_p$ for $p = 0,\ldots,r-1$, and s_p is the number of fixed row effects predicting p th within cell slope, β_{pjk}; t_p is the number of fixed column effects predicting β_{pjk}; u_p is the number of fixed interaction effects predicting β_{pjk}.

The between cell model is a multivariate model because there are "r" outcome variables for each cell, jk. Because we allow a different set of predictors for each within cell slope, β_{pjk}, we need to note $F_p \neq F_{p'}$ for β_p and $\beta_{p'}$.

Matrix Form without Subscripts

To further simplify notation and subsequent presentation of estimation formulas, we now rewrite the general crossed multilevel models without subscripts. This presentation is useful in two ways: First, it provides the matrix structure of the crossed multilevel model. Second, using the model without subscripts, it is easy to present estimation method. The within cell model becomes

$$Y = X\beta + e, \qquad e \sim N(0, \Psi) \qquad (1.10)$$

where $Y = [Y_{11}^T, \ldots, Y_{JK}^T]^T$,

$\beta = [\beta_{11}^T, \ldots, \beta_{JK}^T]^T$,

$e = [e_{11}^T, \ldots, e_{JK}^T]^T$, and

$$X = \text{Diag}\,[X_{jk}] = \begin{bmatrix} X_{11} & & & 0 \\ & \ddots & & \\ & & X_{jk} & \\ & & & \ddots \\ 0 & & & X_{JK} \end{bmatrix} .$$

The between cell model is

$$\beta = W\gamma + Ra + Cb + Tc \,, \qquad (1.11)$$

where $\gamma = [\gamma_0^T, \ldots, \gamma_p^T, \ldots, \gamma_{r-1}^T]^T$,

$W = [W_{11}^T, \ldots, W_{JK}^T]^T$,

$a = [a_1^T, \ldots, a_J^T]^T$, and $a \sim N(0, \Omega_a)$,

$b = [b_1^T, \ldots, b_K^T]^T$, and $b \sim N(0, \Omega_b)$,

$c = [c_{11}^T, \ldots, c_{JK}^T]^T$, and $c \sim N(0, \Omega_c)$,

where $\Omega_a = \text{subdiag}(\tau_a)$, $\Omega_b = \text{subdiag}(\tau_b)$, and $\Omega_c = \text{subdiag}(\tau_c)$.

$R = R_{jk} \otimes M_R$ where R_{jk} are always $r \times r_1$ matrices where each column has elements of a single '1' and all other elements are zero. As noted earlier the elements of '1s' in R_{jk} indicate the presence of random effects of within cell slopes. M_R is G x J matrix which determines the row membership of a cell; "G" is the total number of cells with data and "J" is the total number of row groups, that is schools. The operator " $\otimes$ " means Kronecker product. $C = C_{jk} \otimes M_C$ where M_C is G x K matrix which determines the column membership of a cell. $T = T_{jk} \otimes M_T$, where M_T is G x G identity matrix.

General Mixed Linear Model

If we substitute the between-cell model (Equation 1.11) into the within-cell model (Equation 1.10), we obtain the single model

$$Y = XW\gamma + XRa + XCb + XTc + e$$

This model can be restated in simpler notation as

$$Y = XW\gamma + X_1a + X_2b + X_3c + e \qquad (1.12),$$

The new matrices of X_1 to X_3 represent the XR, XC, and XT. This combined model can be reformulated into general mixed linear model as

$$Y = A_1\theta_1 + A_2\theta_2 + e, \qquad (1.13)$$

where $A_1 = [X\ W]$, θ_1 is γ, $A_2 = [X_1|X_2|X_3]$, $\theta_2 = [a^T, b^T, c^T]^T$, and e is the same as defined previously. The author will use these two models, Equation (1,12) and (1.13), for the presentation of estimation theory in chapter two.

CHAPTER II

EMPRICAL BAYES ESTIMATION OF RANDOM EFFECTS IN THE CROSSED MULTILEVEL MODEL

Introduction

At present, the estimation theories contributing to crossed multilevel data may be classified in two research streams. The first stream may be viewed as classical variance component analysis which started as early as the 1930s. In this stream the researchers attempted to estimate various random effects ANOVA or ANCOVA models. Graybill (1961) extensively treated the estimation problems for the constants and variances in the linear model with balanced designs and demonstrated the optimality properties of the classical analysis of variance procedures. In the case of 'unbalanced factorial and nested data', there may be four subgroups of estimation methods. They may be labeled, 1) Method of Moments (Henderson, 1953; Searle and Henderson, 1961; Cunningham and Henderson, 1968), 2) MIVQUE or MINQUE (Harville, 1969; LaMotte, 1970; Rao, 1972), 3) Maximum Likelihood (Hartley and Rao, 1967; Harville, 1977), 4) Restricted Maximum Likelihood (REML) (Patterson and Thompson, 1971; 1974; Harville, 1977).

These four methods are currently implemented in standard statistical packages, such as SAS and BMDP but all are limited to the estimation of variance components. Numerical applications are available under mixed ANOVA or ANCOVA models which do not allow continuous group-level variables (see BMDP,1985,pp413-435).

The more general approach for hierarchically nested data was made in the second research stream. Researchers in this stream use the term of

'multilevel analysis' because they attempt to estimate both micro- and macro-parameters. There are several researchers who developed estimation methods independently. Lindley and Smith (1972) derived Bayesian estimates for the hierarchical linear model; Smith (1973) compared Bayesian and least squares estimates for hierarchical linear model; Dempster, Laird, and Rubin (1977) established the EM algorithm for ML estimation for covariance components models; Longford (1985) developed the Fisher scoring algorithm for ML estimation of covariance components in multilevel mixed linear models; Goldstein (1986) developed an iterative generalized least squares estimation; Laird and Ware (1982), Strenio, Weisberg, and Bryk (1983), Mason, Wong, and Entwisle (1984), and Raudenbush and Bryk (1986) developed approaches to estimation using empirical Bayes estimation method via EM algorithm.

These methods have been widely applied for the analysis of hierarchically nested data and can be considered as candidate methods for the analysis of crossed multilevel model. This thesis follows the flow of empirical Bayes estimation using the EM algorithm for the numerical analysis of crossed multilevel model. While the Bayesian theory formulates a prior density for the variance and covariances, the empirical Bayes theory uses maximum likelihood point estimates of the variance and covariances of the prior that maximize their marginal posterior distribution (Efron and Morris, 1975). The term empirical Bayes came from the fact that it uses empirical data to estimate the variance and covariance parameters.

For the application of empirical Bayes estimation, the EM algorithm produces the ML estimates of parameters by substituting the expected complete data sufficient statistics into the formula for complete-data ML

estimation of the parameters. The present chapter discusses how empirical Bayes estimation theory can be applied to the estimation of crossed multilevel model. The next chapter presents computational EM formulas for obtaining ML estimates of the parameters. Here, I will describe Bayesian estimation theory for the model given the known fixed effects, variances, and covariances. Then in chapter 3 I will show how EM algorithm can be used for estimation of crossed multilevel model.

The Bayesian Model with Known Covariances

Using the notation and results of Raudenbush (1988), the Bayesian linear model takes the form of general linear model,

$$Y = A\theta + e, \qquad e \sim N(0, \Psi) \quad (2.1),$$

where Y is N by 1 vector of outcomes;

A is N by P matrix of predictors;

θ is P by 1 vector of parameters;

e is N by 1 vector of random errors.

The critical difference between the classical linear model and the Bayesian model lies on the conception of the parameters, θ. The conception of the parameters in classical linear model can be summarized as follows:

1) θ is a fixed unknown; 2) we compute $\hat{\theta}$ as our estimate based on the sample of data; 3) if we replicate the same study many times, the $\hat{\theta}$ values computed will vary with a mean of θ and a dispersion var($\hat{\theta}$) and the distribution of $\hat{\theta}$ is called the sampling distribution; 4) we use the

sampling distribution of $\hat{\theta}$ for the inferences about θ.

In the Bayesian point of view, the parameters, θ, of Equation (2.1) themselves have a general linear structure in terms of other quantities which they call hyperparameters (Lindley, & Smith, 1972). Therefore all parameters of θ are viewed as random and hyperparameters are viewed as fixed. Since θ is random, we may propose its distribution,

$$\theta \sim N(\bar{\theta}, \Omega) \qquad (2.2).$$

The Bayesian views this as our prior distribution of θ and the prior distribution represents our state of knowledge on θ (Smith, 1973). Therefore our prior knowledge about the location of the parameter vector is θ, and the precision of this knowledge is measured by Ω^{-1}.

In the context of Bayesian estimation, the prior distribution of parameters is assumed to be known, and the posterior distribution of parameters given the data and prior parameters needs to be found. The procedure for finding the posterior distribution involves Bayes theorem, which states that the posterior density function $f(\theta|Y)$ is proportional to the product of two independent density functions: the likelihood $L(Y|\theta)$ and the prior distribution $P(\theta)$. Hence the posterior density function is,

$$f(\theta|Y) \propto L(Y|\theta)P(\theta) \qquad (2.3).$$

Bayes theorem incorporates the prior information of the parameters and the observed information from the sample. The right hand side of

Equation (2.3) can be easily seen as joint probability density function of observed data Y and prior information of θ. Under the assumption of known dispersion matrices, Ψ and Ω, the joint normal distribution of the data, Y, and the parameter θ is

$$f(Y,\theta) = L(Y|\theta)P(\theta) \qquad (2.4)$$
$$= C_1 \exp[-1/2(Y-A\theta)^T\Psi^{-1}(Y-A\theta)]C_2\exp[-1/2(\theta-\bar{\theta})^T\Omega^{-1}(\theta-\bar{\theta})]$$

where C_1 and C_2 are normalizing constants for the two normal densities. Thus the posterior distribution of θ is proportional to the joint normal density function.

$$f(\theta|Y) \propto \exp[-1/2(Y-A\theta)^T\Psi^{-1}(Y-A\theta)]\ \exp[-1/2(\theta-\bar{\theta})^T\Omega^{-1}(\theta-\bar{\theta}] \qquad (2.5).$$

Equation (2.5) determines θ^*, the posterior mean of θ, and the dispersion D_θ^*. That is

$$f(\theta|Y) \propto \exp[-1/2\ (\theta-\theta^*)D_\theta^{*-1}(\theta-\theta^*)], \qquad (2.6)$$
$$\text{where } \theta^* = D_\theta^*(\Omega^{-1}\bar{\theta} + A^T\Psi^{-1}Y) \qquad (2.7)$$
$$\text{and} \quad D_\theta^* = (\Omega^{-1} + A^T\Psi^{-1}A)^{-1}. \qquad (2.8)$$

Estimation theory for the Bayes linear model is presented in Lindley and Smith (1972) and Dempster, Rubin, and Tsutakawa (1981). Bayesian inferences are based on these posterior estimators from which we can get point estimates and intervals.

Theory of EM Algorithm

The Bayesian theory presented above requires known variances and covariances in order to get posterior means and dispersion matrices for parameters under consideration. In the empirical Bayes method, the parameters are estimated from the joint normal posterior distribution given the ML estimates or other consistent point estimates of the hyper-parameters for their unknown values. This empirical Bayes estimation method is applied in the EM algorithm, where the ML estimates of hyperparameters defined through empirical Bayesian methods are used for finding the expected complete data sufficient statistics which in turn are used for ML estimation of the parameters (Dempster, Laird, and Rubin, 1977).

The name EM algorithm comes from the characteristics of an iterative routine which cycles through an Expectation step and a Maximization step at each iteration. The expectation step finds the posterior expectation of the sufficient statistics based on the 'complete data' given the 'observed data' and current estimates of parameters. The Maximization step then uses the expected sufficient statistics to produce an estimated values of unknown parameters under estimation. These two steps cooperate to increase the likelihood function of the estimated parameters.

To make the procedure of EM algorithm concrete, consider a linear model with a sample size of n and we want to estimate the variance components of the model as,

$$Y = X\beta + e,$$

where Y is a n by 1 vector of observations;

X is a n by p matrix of predictor variables;

β is a p by 1 vector of regression coefficients;

e is a n by 1 vector of random error, and the distributional assumption is $e \sim N(0, \sigma^2 I)$ and $\beta \sim N(\theta, \Gamma)$ as a prior.

The EM algorithm uses the expected value of sufficient statistics of the complete data conditioned on observed data and current estimates of parameters as a proxy for the summary statistics of "complete data". The "complete data" consist of Y and the true values of parameters, here β and e. By employing the assumption of having complete data, we estimate the sufficient statistics of complete data by the conditional expectation as $E(e'e|Y,\sigma^{2i})$ in E-step, where $e'e$ is the complete data sufficient statistics, Y is the observed data, σ^{2i} is the current estimate of σ^2 at the ith iteration. The term '|' can be read as 'given the data of'. Suppose we have a vague prior on β. Then the variance components of Γ become infinitely large and $\Gamma^{-1} \to 0$ and the Equation (2.9) become the functionally same model as a standard linear model. Hence we pose the conditional distribution of β given the data as

$$\beta|Y,\sigma^{2i} \sim N(\beta^i, V) \qquad (2.10)$$

where $\beta^i = (X'X)^{-1}X'Y$

$V = \sigma^{2i}(X'X)^{-1}$.

The sufficient statistic is then

$$E(e'e|Y,\sigma^{2i}) = E[(Y-X\beta)'(Y-X\beta)|Y,\sigma^{2i}] \qquad (2.11)$$
$$= E[(Y-X\beta^{i}+X\beta^{i}-X\beta)'(Y-X\beta^{i}+X\beta^{i}-X\beta)|Y,\sigma^{2i}]$$
$$= (Y-X\beta^{i})'(Y-X\beta^{i}) + E[(\beta-\beta^{i})'X'X(\beta-\beta^{i})|Y,\sigma^{2i}]$$
$$= SSres + tr[X'XVar(\beta|Y,\sigma^{2i})]$$
$$= SSres + tr(X'X)\sigma^{2i}(X'X)^{-1} = SSres + p\sigma^{2i}.$$

The result of Equation (2.11) is the proxy of the sufficient statistics for the complete data. In the M-step we simply get maximum likelihood estimates of σ^2 as

$$\sigma^{2(i+1)} = [SSres + p\sigma^{2i}]/n \qquad (2.12).$$

The resulting value of $\sigma^{2(i+1)}$ is then used as input for the next E-step. The reader will notice that if the current estimate $\hat{\sigma}^{2i} = SSres/(n-p)$, then $\hat{\sigma}^{2(i+1)} = SSres/(n-p)$ also. Indeed, if $\hat{\sigma}^{2i}$ differ from $SSres/(n-p)$, the iteration will nevertheless converge to that estimate.

For the use of EM algorithm, there remains a choice between two likelihood functions for ML estimation of variances and covariances, which is labeled as MLF and MLR (Dempster, Rubin, and Tsutakawa, 1981). Consider the general mixed Bayesian model as

$$Y = A_1\theta_1 + A_2\theta_2 + e,$$
$$\text{where} \quad \begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\Gamma & 0\\ 0 & T\end{bmatrix} \right)$$
$$\text{and} \quad e \sim N(0, \Psi)$$

The MLF method treats θ_1, Ψ, and T as fixed parameters to be estimated.

To achieve such estimation requires the conditional distribution $f(\theta_2|Y,\theta_1,\Psi,T)$. In contrast, the MLR method treats Ψ and T as fixed parameters to be estimated. To achieve such estimation requires $f(\theta_1,\theta_2|Y,\Psi,T)$. In this thesis, I have chosen the MLF approach.
Consider the general mixed linear model as Equation (1.13), which was

$$Y = A_1\theta_1 + A_2\theta_2 + e,$$

where the $A_1\theta_1$ is the fixed part and remaining right hand side of the equation is random part. To obtain the conditional density of θ_2 given Y, θ_1, Ψ, and T, the model is modified as,

$$d = A_2\theta_2 + e,$$

where $d = Y - A_1\theta_1$ (Dempster, Rubin, & Tsutakawa, 1981). Then the equation become a general Bayesian model which enables us to find posterior means and dispersion matrix of θ_2 using Bayesian method presented in the previous section in this chapter.

Bayesian Formulation of Crossed Multilevel Model

In order to use EM algorithm, we need to formulate the crossed multilevel model into the general Bayesian form of Equation (2.1). Recall the combined model Equation (1.12) was

$$Y = XW\gamma + X_1a + X_2b + X_3c + e$$

and the mixed linear model, Equation (1.13), was

$$Y = A_1\theta_1 + A_2\theta_2 + e.$$

Since these two models are equivalent, we can posit the following identities.

$$A_1 = [XW] \qquad (2.13)$$

$$A_2 = [X_1 \mid X_2 \mid X_3]$$

and

$$\theta_1 = \gamma$$

$$\theta_2^T = [a^T \mid b^T \mid c^T].$$

The related assumptions are that the fixed and random parameters are independent and that the random parameters, a, b, and c, are also independent. We change the general mixed linear model into the model with only random parameters such as

$$d = A_2\theta_2 + e \qquad (2.14),$$

where $d = Y - A_1\theta_1$ and the distributional assumptions on the new outcome variable is

$$d \sim N(0 , A_2TA_2^T + \Psi) \qquad (2.15)$$

Other terms of the model (2.14) keep the same distributional assumptions as the mixed model.

In order to make the equivalent crossed multilevel model, we also change the model (2.14) into;

$$d = X_1 a + X_2 b + X_3 c + e \qquad (2.16),$$

where $d = Y - XW\gamma$ and the distributional assumption is

$$d \sim N(0 , X_1 \Omega_a X_1^T + X_2 \Omega_b X_2^T + X_3 \Omega_c X_3^T + \sigma^2 I) \qquad (2.17)$$

Now, the models (2.16) and (2.14) are equivalent.

Joint Normal Distribution

The necessary step for posterior estimation is the specification of joint normal distribution of the modified model.

The expected mean vector of the dependent variable is

$E(d) = E(A_2\theta_2 + e) = 0$

and the expected mean vector of the parameter is

$E(\theta_2) = 0.$

The variance of outcome vector is

$Var(d) = A_2 T A_2^T + \Psi$

and the parameter variance is

$Var(\theta_2) = T.$

The covariance between 'd' and 'θ_2' is

$Cov(d, \theta_2) = Cov(A_2\theta_2 + e, \theta_2) = A_2 T.$

Hence, the joint normal distribution is

$$\begin{bmatrix} d \\ \theta_2 \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A_2 T A_2^T + \Psi & A_2 T \\ T A_2^T & T \end{pmatrix} \right] \qquad (2.18)$$

The equivalent distribution of the crossed multilevel model can be

obtained by substituting A_2 and T with their corresponding terms. Hence

$$\begin{bmatrix} d \\ a \\ b \\ c \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} X_1\Omega_a X_1^T + X_2\Omega_b X_2^T + X_3\Omega_c X_3^T + \sigma^2 I & X_1\Omega_a & X_2\Omega_b & X_3\Omega_c \\ \Omega_a X_1^T & \Omega_a & 0 & 0 \\ \Omega_b X_2^T & 0 & \Omega_b & 0 \\ \Omega_c X_3^T & 0 & 0 & \Omega_c \end{bmatrix} \right] \quad (2.19).$$

Posterior Distribution of Parameters

Having specified the joint normal distribution of modified crossed multilevel model, we now need to get the posterior distribution of the parameters, a, b, and c. By using the definitions of (2.13) and by applying the posterior estimation method, we now can pose the posterior distribution of parameter vector of the modified version of the Bayesian model. Then we translate the results into the terms of crossed multilevel model using the identities defined before. The posterior distribution of the parameter given the data is

$$\theta_2 | d, T, \Psi \sim N(\theta_2^*, D_{\theta 2}^*) \quad (2.20),$$

where

$$\theta_2^* = D_{\theta 2}^* A_2^T \Psi^{-1} d \quad (2.21)$$

and

$$D_{\theta 2}^* = (A_2^T \Psi^{-1} A_2 + T^{-1})^{-1} \quad (2.22)$$

Equations of (2.21) and of (2.22) show that the posterior distribution of parameters requires the prior information of the parameters; Ψ, T, and θ_1. Here we assume that the information of the Ψ, T, and θ_1 are known.

I will first show the posterior dispersion matrix and then move on to the posterior mean of parameter vector in the crossed multilevel model.

Posterior Dispersion Matrix of Parameters

Equation (2.22) is now rewritten in terms of crossed multilevel model,

$$D_{\theta 2}^{*} = \sigma^2 \begin{bmatrix} X_1^T X_1 + \sigma^2 \Omega_a^{-1} & X_1^T X_2 & X_1^T X_3 \\ X_2^T X_1 & X_2^T X_2 + \sigma^2 \Omega_b^{-1} & X_2^T X_3 \\ X_3^T X_1 & X_3^T X_2 & X_3^T X_3 + \sigma^2 \Omega_c^{-1} \end{bmatrix}^{-1} \quad (2.23).$$

To derive the inversion of the above matrix, the following definitions are useful.

$$B^{+} = \begin{bmatrix} X_1^T X_1 + \sigma^2 \Omega_a^{-1} & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 + \sigma^2 \Omega_b^{-1} \end{bmatrix} = \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix} \quad (2.24)$$

$$B = \begin{bmatrix} X_1^T X_3 \\ X_2^T X_3 \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

$$U = X_3^T X_3 + \sigma^2 \Omega_c^{-1}$$

With these definitions we say

$$D_{\theta 2}^{*} = \begin{bmatrix} B^{+} & B \\ B^T & U \end{bmatrix}^{-1}. \quad (2.25)$$

The advantage of this representation is that we can now apply the partitioned matrix inversion method since (2.25) is partitioned in two-by- two form. The results of the inversion (Searle, 1982) is

$$D_{\theta 2}^{*} = \begin{bmatrix} H^{-1} & -(GH^{-1})^{T} \\ -(GH^{-1}) & U^{-1} + GH^{-1}G^{T} \end{bmatrix}, \qquad (2.26)$$

where $H = B^{+} - BU^{-1}B^{T}$

$G = U^{-1}B^{T}$.

Equation (2.26) requires the two inverted matrices, U^{-1} and H^{-1}. The dimension of U^{-1} is determined by the number of columns of the matrix X_3. But the 'H' matrix has two-by-two partitioned form again. We apply the same inversion method as applied for (2.25). For the inversion of the 'H' we use the definitions at Equation (2.24) and rewrite it into simpler notation as

$$H = \begin{bmatrix} B^{11} - B_1U^{-1}B_1^{T} & B^{12} - B_1U^{-1}B_2^{T} \\ B^{21} - B_2U^{-1}B_1^{T} & B^{22} - B_2U^{-1}B_2^{T} \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \qquad (2.27)$$

Then $$H^{-1} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}^{-1} = \begin{bmatrix} H^{11} & H^{12} \\ H^{21} & H^{22} \end{bmatrix} \qquad (2.28)$$

where $H^{11} = (H_{11} - H_{12}H_{22}^{-1}H_{21})^{-1}$

$H^{12} = -H^{11T}H_{12}H_{22}^{-1}$

$H^{22} = H_{22}^{-1} + (H_{22}^{-1}H_{21})H^{11T}(H_{12}H_{22}^{-1})$ or $(H_{22} - H_{21}H_{11}^{-1}H_{12})^{-1}$.

Once we get the inverted matrix 'H', we can apply the results to the inversion of the matrix (2.26). From (2.26) we have

$$D_{\theta 2}^{*} = \sigma^2 \begin{bmatrix} H^{11} & H^{12} & -H^{11}B_1U^{-1} - H^{12}B_2U^{-1} \\ H^{21} & H^{22} & -H^{21}B_1U^{-1} - H^{22}B_2U^{-1} \\ & (\text{symmetric}) & U^{-1} + U^{-1}[B_1{}^{T}H^{11}B_1 + B_2{}^{T}H^{21}B_1 \\ & & + B_1{}^{T}H^{12}B_2 + B_2{}^{T}H^{22}B_2]U^{-1} \end{bmatrix} \qquad (2.29)$$

Now the task is to reexpress the each block of $D_{\theta 2}^{*}$ in terms of the crossed multilevel model. The Equation (2.29) shows that the solutions of each block are linked each other. Thus more compact notation for the matrix (2.29) is useful.

$$\text{Let } D_{\theta 2}^{*} = \sigma^2 \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \qquad (2.30).$$

Using the definitions at Equations of (2.24) and (2.27), and by collecting the results of Equations (2.26) through (2.29), we can obtain the following results:

$$v_{11} = H^{11} = (H_{11} - H_{12}H_{22}^{-1}H_{21})^{-1} = (H_{11} - G_a)^{-1} \qquad (2.31)$$

$$v_{12} = H^{12} = -H^{11T}H_{12}H_{22}^{-1} = -v_{11}{}^{T}X_1{}^{T}{}_{M}X_2H_{22}^{-1}$$

$$v_{22} = H^{22} = H_{22}^{-1} + (H_{22}^{-1}H_{21})H^{11T}(H_{12}H_{22}^{-1})$$
$$= H_{22}^{-1} + H_{22}^{-1}X_2{}^{T}MX_1v_{11}{}^{X}{}_{1}{}^{T}MX_2H_{22}^{-1}$$
$$\text{or} \quad (H_{22} - H_{21}H_{11}^{-1}H_{12})^{-1} = (H_{22} - G_b)^{-1}$$

$$v_{13} = -(v_{11}X_1{}^{T}X_3 + v_{12}X_2{}^{T}X_3)U^{-1}$$

$$v_{23} = -(v_{21}X_1^TX_3 + v_{22}X_2^TX_3)U^{-1}$$

$$v_{33} = U^{-1} - (v_{13}^TX_1^TX_3 + v_{23}^TX_2^TX_3)U^{-1}$$

$$v_{21} = v_{12}^T, \quad v_{31} = v_{13}^T$$

where $H_{11} = X_1^TMX_1 + \sigma^2\Omega_a^{-1}$,

$H_{22} = X_2^TMX_2 + \sigma^2\Omega_b^{-1}$,

$M = I - X_3^TU^{-1}X_3$,

$G_a = X_1^TMX_2H_{22}^{-1}X_2^TMX_1$,

$G_b = X_2^TMX_1H_{11}^{-1}X_1^TMX_2$.

We need to note that the two terms H_{11} and H_{22} are subdiagonal matrices and their inversions are easily obtainable. The complexity of these computation can be reduced if we note the fact that each component is the function of other components. Hence once we have information of U^{-1}, and H_{22}^{-1}, we can get v_{11} which serves for the estimation of v_{12} which, in turn, serves for the computation of v_{13}, and so on.

Posterior Mean of Parameter Vector

Having obtained the posterior dispersion matrix of the crossed multilevel model, we now consider the posterior expectation. Recall the Equation (2.21) which was

$$\theta_2^* = D_{\theta 2}^* A_2^T\Psi^{-1}d.$$

Assuming we have the estimates of θ_1 and we have $D_{\theta 2}^*$, the matrix operation become straitforward.

Concerning the dispersion matrix, recall the assumptions for the crossed multilevel model in which the column, row, and interaction effects are mutually independent. In other words, the proposed prior dispersion matrix has block diagonal form where Ω_a, Ω_b, and Ω_c are at its diagonal position. Ω_a, Ω_b, and Ω_c are also block diagonal matrices with submatrices of τ_a, τ_b, and τ_c at their diagonal position. These three submatrices are full matrices with dimension of r_1, r_2, and r_3 and they take the diagonal position of Ω_a, Ω_b, and Ω_c respectively. However the posterior dispersion matrix which can be obtained after data observation is not block diagonal. The off diagonal submatrices, v_{12}, v_{13}, and v_{23}, are not null. We need to use the obtained full matrix of $D_{\theta 2}^{*}$ for posterior estimation of the parameter vector as the equation (2.21).

The equation (2.21) become as follow.

$$\theta_2^{*} = \begin{bmatrix} a^{*} \\ b^{*} \\ c^{*} \end{bmatrix} = \begin{bmatrix} v_{11}X_1^{T} + v_{12}X_2^{T} + v_{13}X_3^{T} \\ v_{21}X_1^{T} + v_{22}X_2^{T} + v_{23}X_3^{T} \\ v_{31}X_1^{T} + v_{32}X_2^{T} + v_{33}X_3^{T} \end{bmatrix} d \qquad (2.32)$$

where X_1, X_2, and X_3 terms reflect the data of the predictors classified as column, row, and their interaction variables respectively. To simplify the results of Equation (2.32), we need to first expand the results, using the operational results of Equation (2.31), and simplify them by noting the interrelationship of the submatrices of Equation (2.31). After a somewhat complicated operation of the relevant submatrices, we can arrive the following results.

$$a^* = v_{11}Q_a, \qquad (2.33)$$

$$b^* = v_{22}Q_b,$$

$$c^* = U^{-1}X_3^T[d - (X_1a^* + X_2b^*)],$$

where $Q_a = X_1^TMd - X_1^TMX_2H_{22}^{-1}X_2^TMd,$

$$Q_b = X_2^TMd - X_2^TMX_1H_{11}^{-1}X_1^TMd.$$

All terms included in Equation (2.33) are previously defined at Equation (2.31).

CHAPTER III.

COMPUTING ESTIMATES OF CROSSED MULTILEVEL MODEL PARAMETERS

This chapter will present the technical aspects of implementing crossed multilevel analysis. The author will show how the EM algorithm provides the estimates of crossed multilevel model.

Model

The model defined for MLF estimation was

$$d = A_2\theta_2 + e,$$

where $d = Y - A_1\theta_1$;

$\theta_2 \sim N(0, T)$;

$e \sim N(0, \Psi)$

with appropriate dimension. θ_1 is fixed parameter while θ_2 is considered as random. The corresponding crossed multilevel model is

$$d = X_1a + X_2b + X_3c + e,$$

where $d = Y - XW\gamma$

$a \sim N(0, \Omega_a)$

$b \sim N(0, \Omega_b)$

$c \sim N(0, \Omega_c)$

and $e \sim N(0, \sigma^2 I)$.

As before, the following definitions are useful to link the above two models;

$$A_1 = XW, \quad A_2 = [X_1 \mid X_2 \mid x_3], \quad \theta_1 = \gamma, \quad \theta_2^T = [a^T \mid b^T \mid c^T].$$

The error, e, is an N by 1 vector which is stacked from the top with all within cell error vectors, e_{jk}, which are uncorrelated across the cells. The within cell errors are n_{jk} by 1 vectors where the n_{jk} elements are also independent. Therefore the dispersion matrix, Ψ, is a N by N diagonal matrix which composed of the dispersion matrices of all non-null cells. Hence the dispersion matrix of each cell is $\sigma^2 I$ and the overall dispersion matrix is

$$\Psi = \begin{bmatrix} \sigma^2 In_{11} & & \\ & \ddots & 0 \\ & 0 & \\ & & \sigma^2 In_{JK} \end{bmatrix} = \sigma^2 I,$$

where $N = \Sigma \Sigma n_{jk} = \Sigma n_g$.

The random parameter variance matrices, Ω_a, Ω_b, and Ω_c are both diagonal block matrices with submatrices of τ_a, τ_b, and τ_c at their diagonal position across the row, column units and their interactions. The three submatrices, τ_a, τ_b, and τ_c, are of full matrices with the dimensions of the number of within-cell random slopes across the row, column units and their interactions as stated in chapter I.

Sufficient Statistics of Complete Data

In the empirical Bayes method, random effects are estimated given the ML estimates of the fixed unknown parameters. In ML estimation, the logic of EM algorithm is to use the expected value of sufficient statistics of the 'complete data' given the observed data and the previous estimates of these fixed parameters as a substitute for the

summary statistics of 'complete data', and perform ML estimation based on the assumption that we had observed the complete data. In our case, the complete data consist of the outcome variable Y and the true values of random effects, θ_2 and e. The fixed unknown parameters include σ^2, τ_a, τ_b, τ_c, and γ. By employing the assumption of having the complete data, ML estimation can be simple to derive.

To find the complete data sufficient statistics, I will first use the joint likelihood function of the complete data and parameters for MLF estimation and then translate the results in terms of the crossed multilevel model. Given the model defined at Equation (2.14) and (2.15) and by referencing the Equation (2.4), the joint density function of the parameters and the complete data is

$$f(d,\theta_2|\theta_1,T,\Psi) = f(d|\theta_2,\theta_1,T,\Psi)\, f(\theta_2|\theta_1,T,\Psi) \qquad (3.1),$$

$$\text{where } f(d|\theta_2,\theta_1,T,\Psi) = [(2\pi)^N|\Psi|]^{-1/2}\exp[(-1/2)(d-A_2\theta_2)^T\Psi^{-1}(d-A_2\theta_2)],$$

$$f(\theta_2|\theta_1,T,\Psi) = [(2\pi)^F|T|]^{-1/2}\exp[(-1/2)(\theta_2{}^T T^{-1}\theta_2)].$$

The term, F, in the second equation on the right hand side of Equation (3.1) is the total number of elements of θ_2. Hence the log-likelihood function will be

$$L \propto (-1/2)(\hat{d}-A_2\theta_2)^T\Psi^{-1}(\hat{d}-A_2\theta_2) - (1/2)\theta_2{}^T T^{-1}\theta_2$$

where $\hat{d} = y - A_1\hat{\theta}_1$ and $\hat{\theta}_1 = (A_1{}^T A_1)^{-1}A_1{}^T(Y - A_2\theta_2)$. Let $\hat{e} = \hat{d}-A_2\theta_2$. From the above complete data log-likelihood, we see that $\hat{e}^T\hat{e}$ and $\theta_2\theta_2{}^T$ are the sufficient statistics for σ^2 and the variance components T. Also $A_1{}^T Y$ and $A_1{}^T A_2\theta_2$ are the sufficient statistics for θ_1.

Having obtained the set of complete data sufficient statistics of the

modified bayesian model; ($\hat{e}^T\hat{e}$, $\theta_2\theta_2^T$, and $A_1^T A_2\theta_2$), we now need to translate them in terms of the crossed multilevel model. Concerning the term, $\theta_2\theta_2^T$, we know that θ_2 is a vector of random effects as

$$\theta_2^T = [a_1^T, \ldots, a_J^T | b_1^T, \ldots, b_K^T | c_{11}, \ldots, c_{JK}^T],$$

$$\text{and } aj^T = [a_{0j}, \ldots, a_{r1-1,j}];$$

$$b_k^T = [b_{0k}, \ldots, b_{r2-1,k}];$$

$$c_{jk}^T = c_g^T = [c_{0g}, \ldots, c_{r3-1,g}],$$

for j = 1,...,J rows; k = 1,...,K columns; and (jk) = g = 1,...,G cells classified by the jth row and kth column; where r_1 is the number of within-cell slopes that are random across the row units; r_2 is the number of within-cell slopes that are random across the column units; r_3 is the number of random within-cell slopes regarding the cells classified by row and column units. Under the assumptions that the rows, the columns and their interaction effects are mutually independent and that each row unit is independent from other row units, each column unit is independent from other column units of complete data and, given the row and column effects, the cell is independent from other cells, the sufficient statistics we need for random residuals are;

$$\Sigma\, a_j a_j^T, \quad \Sigma\, b_k b_k^T, \quad \Sigma\, c_g c_g^T \qquad (3.5)$$

Now it become trivial to get the ML estimator of T as

$$\hat{\tau}_a = J^{-1} \Sigma\, a_j a_j^T \qquad (3.6)$$

$$\hat{\tau}_b = K^{-1} \Sigma\, b_k b_k^T$$

$$\hat{\tau}_c = G^{-1} \Sigma c_g c_g^T$$

Concerning the terms of $\hat{e}^T \hat{e}$, where $\hat{e} = \hat{d} - A_2\theta_2$ and $\hat{d} = Y - A_1\hat{\theta}_1$, we use the cell level notations of the vector $\hat{e}$ and use the identities between the Bayesian and crossed multilevel model. Then the sufficient statistic for σ^2 of complete data is

$$\Sigma \Sigma \hat{e}_{jk}^T \hat{e}_{jk} = \Sigma \Sigma [\hat{d}_{jk} - (X_{1jk}a_j + X_{2jk}b_k + X_{3jk}c_{jk})]^T \quad (3.7)$$
$$x \ [\hat{d}_{jk} - (X_{1jk}a_j + X_{2jk}b_k + X_{3jk}c_{jk})].$$

Then the ML estimator of σ^2 is

$$\hat{\sigma}^2 = (1/N) \Sigma \Sigma \ \hat{e}_{jk}^T \hat{e}_{jk} \qquad (3.8).$$

Finally the sufficient statistics of the estimator of θ_1 are $A_1^T Y$ and $A_1^T A_2 \theta_2$. the corresponding cell-level expression for $A_1^T A_2 \theta_2$ in crossed-multilevel model is

$$A_1^T{}_{jk} A_{2jk} \theta_{2jk} = W_{jk}^T X_{jk}^T (X_{1jk}a_j + X_{2jk}b_k + X_{3jk}c_{jk}).$$

Then the equation $\hat{\theta}_1 = (A_1^T A_1)^{-1} A_1^T (Y - A_2\theta_2)$ can be rewritten in computational form in terms of crossed multilevel model as

$$\hat{\gamma} = (\Sigma \Sigma W_{jk}^T X_{jk}^T X_{jk} W_{jk})^{-1} (\Sigma \Sigma W_{jk}^T X_{jk}^T Y_{jk} - \Sigma \Sigma W_{jk}^T X_{jk}^T (X_{1jk}a_j$$
$$+ X_{2jk}b_k + X_{3jk}c_{jk})) \qquad (3.9).$$

Equation (3.9) shows that the complete data sufficient statistics for estimating the fixed effects of the crossed multilevel model are;

$$\Sigma\Sigma\, W_{jk}{}^{T}X_{jk}{}^{T}Y_{jk}, \quad \Sigma\Sigma\, W_{jk}{}^{T}X_{jk}{}^{T}X_{1jk}a_j, \quad \Sigma\Sigma\, W_{jk}{}^{T}X_{jk}{}^{T}X_{2jk}b_k,$$
$$\Sigma\Sigma\, W_{jk}{}^{T}X_{jk}{}^{T}X_{3jk}c_{jk}. \qquad (3.10)$$

Having found the complete data sufficient statistics for ML estimation of τ_a, τ_b, τ_c, and σ^2 and γ, we need to obtain the expected values of the complete data sufficient statistics conditioned on the observed incomplete data Y and the estimates of the parameters. The EM algorithm then uses the values of the conditional expectation of the complete data sufficient statistics as the proxy of the complete data sufficient statistics for ML estimation of parameters. The following sections will show the EM formulas for parameter estimation.

EM Formula for Estimating the Fixed Parameter ($\theta_1{}^*$)

As noted at prior section, the complete data sufficient statistic for estimating the posterior value of θ_1 is the term $A_1{}^{T}A_2\theta_2$. If we have complete data we can directly apply the value of $A_1{}^{T}A_2\theta_2$ for ML estimation. Since the actual data is not complete, the EM algorithm uses the following conditional expectation to get the substitute for the complete data sufficient statistic for ML estimation as

$$A_1{}^{T}A_2E[\theta_2|T,\ \sigma^2,\ \theta_1,\ d] = A_1{}^{T}A_2\theta_2{}^* \qquad (3.11),$$

where the right side of '|' includes incomplete data 'd' and the current estimates of T, σ^2,and θ_1. Since A_1, A_2 are given, the posterior expectation applies only to θ_2. Therefore the result of (3.9) is $A_1{}^{T}A_2\theta_2{}^*$, where $\theta_2{}^*$ is the posterior mean of θ_2 as presented at chapter two. Once we have the estimates of σ^2, T, and γ, we can obtain

θ_2^* from Equation (2.33) and then the estimate of the fixed parameter γ^* is

$$\gamma^* = (\Sigma \Sigma W_{jk}^T X_{jk}^T X_{jk} W_{jk})^{-1} [\Sigma \Sigma W_{jk}^T X_{jk}^T Y_{jk} - \Sigma \Sigma W_{jk}^T X_{jk}^T (X_{1jk} a_j^* + X_{2jk} b_k^* + X_{3jk} c_{jk}^*)] \qquad (3.12).$$

EM Formula for Estimating the Parameter Variances

If the data is complete then the sufficient statistic is $\Sigma a_j a_j^T$ to get ML estimates of τ_a. With EM algorithm, the data are incomplete. Thus we substitute $E[\Sigma\ a_j a_j^T | d,\ \Psi,\ \tau,\ \gamma]$ for $\Sigma\ a_j a_j^T$. Finding this conditional expectation directly follow from the standard theory. The dispersion of a_j, $Var(a_j)$, is defined as,

$$Var(a_j) = E(a_j a_j^T) - E(a_j)\ E(a_j)^T \qquad \text{(Searle, 1982)}.$$

This may be restated into a more useful form,

$$E(a_j a_j^T) = E(a_j)\ E(a_j)^T + Var\ (a_j).$$

This equation is for unconditional expectation. It is the analogue for conditional expectation given the 'incomplete data' Y and current estimates of (τ_a, τ_b, τ_c, σ^2, and γ), so the expectation is

$$E(\Sigma\ a_j a_j^T | d, \tau_a, \tau_b, \tau_c, \sigma^2, \gamma) = \Sigma\ a_j^* a_j^{*T} + \sigma^2 \Sigma\ v_{11jj} \qquad (3.13).$$

v_{11jj} are the matrices taken from the diagonal position of v_{11} at equation (2.31).

The maximization step for the estimation of τ_a^* is accomplished by

the simple operation. The complete data ML estimator of τ_a^* is,

$$\tau_a^* = J^{-1}(\Sigma\, a_j^* a_j^{*T} + \sigma^2 \Sigma\, v_{11jj}) \qquad (3.14).$$

This completes one iteration for estimating τ_a^*. The posterior estimates of other two estimates, τ_b^* and τ_c^*, are presented at equation (3.15) and (3.16) below;

$$\tau_b^* = K^{-1}\,(\,\Sigma\, b_k^*\, b_k^{*T} + \sigma^2 \Sigma\, v_{22kk}) \qquad (3.15)$$

$$\tau_c^* = G^{-1}\,(\Sigma\Sigma\, c_{jk}^* c_{jk}^{*T} + \sigma^2 \Sigma\Sigma\, v_{33jkjk}). \qquad (3.16)$$

'v_{22kk}' and 'v_{33jkjk}' are the matrices taken from the diagonal positions of v_{22} and v_{33} at Equation (2.31) respectively.

The procedure used to derive the equations (3.15) and (3.16) is the same as the one for τ_a^*. The iterative routine for estimating the $\mathrm{Var}(\theta_{2jk})$ is now explicit. τ_a^*, τ_b^* and τ_c^* consist the $\mathrm{Var}(\theta_{2jk})$ as known variances for next iteration of the EM computation, where

$$\mathrm{Var}(\theta_{2jk}) = T^* = \begin{bmatrix} \tau_a^* & & \\ & \tau_b^* & \\ & & \tau_c^* \end{bmatrix}$$

The posterior estimation given the known variances was provided in chapter two. We plug the estimated variances from the previous iteration (Equation 3.14, 3.15, 3.16) into the given variances to obtain the sufficient statistics at next iteration (E-step). The new ML estimates follow easily (M-step).

EM Formula for Estimating the Within-Cell Variance (Ψ)

The within cell errors, e_{jk}, are uncorrelated across and within the cells. Therefore the variance matrix, Ψ, is a N by N diagonal matrix with off- diagonals of zero. Because the errors are uncorrelated, the variance estimates of each cell can be computed separately and the estimation procedure is the same across the G cells. The G parallel estimates are then arranged along the diagonal of Ψ to complete the matrix estimate.

In order to compute sufficient statistics, we first write down a model for e_{jk}. The model is

$$e_{jk} = d_{jk} - A_{2jk}\theta_{2jk} \qquad (3.17),$$

where $d_{jk} = Y_{jk} - A_{1jk}\theta_1$ with the dimension of n_{jk} by 1. A_{2jk}, θ_{2jk} are the same as before except the notation of cell indicators of the subscripts. Then the complete data sufficient statistic for σ^2 is

$$\Sigma\,\Sigma\,\hat{e}_{jk}^{T}\hat{e}_{jk} = \Sigma\,\Sigma\,(\hat{d}_{jk} - A_{2jk}\,\theta_{2jk})^{T}\,(\hat{d}_{jk} - A_{2jk}\,\theta_{2jk}), \qquad (3.18)$$

where $\hat{d}_{jk} = Y_{jk} - A_1\hat{\theta}_1$.

The resulting quantity of Equation (3.18) is a scalar. The expected value for a scalar quadratic is

$$E\,(e_{jk}^{T}e_{jk}) = E(e_{jk})^{T}\,E(e_{jk}) + tr[Var(e_{jk})] \qquad (3.19)$$

(Searle,1982,p.355).

As we discussed earlier, the EM algorithm uses the conditional expectation at E-step. Hence the posterior estimation of the sufficient

statistic has the same form of (3.18) except the notations of posterior estimates as

$$E(e_{jk}^T e_{jk} | d, T, \sigma^2, \theta_1) = E(e_{jk}^*)^T E(e_{jk}^*) + tr[Var(e_{jk} | d, T, \sigma^2, \theta_1)] \quad (3.20).$$

We now need each term of the right hand side of equation (3.20). By using the analogues between the unconditional and conditional expectation, and the identity (3.17), we have

$$E(e_{jk} | d, T, \sigma^2, \theta_1) = d_{jk} - A_{2jk}\theta_{2jk}^* \quad (3.21),$$

where θ_{2jk}^* is the posterior mean of θ_{2jk}. The second term in the right hand side of equation (3.20) is rewritten as

$$Var(d_{jk} - A_{2jk}\theta_{2jk} | d, T, \sigma^2, \theta_1).$$

This is the same as $Var(-A_{2jk}\theta_{2jk} | d, T, \sigma^2, \theta_1)$ because d_{jk} is given data that means constant. Hence we have

$$Var(e_{jk} | d, T, \sigma^2, \theta_1) = - A_{2jk} Var(\theta_{2jk} | d, T, \sigma^2, \theta_1) A_{2jk}^T \quad (3.22)$$

where $Var(\theta_{2jk} | d, T, \sigma^2, \theta_1) = \sigma^2 \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}$,

whose component matrices had been defined from Equation (2.30) through (2.31). The expression of (3.22) can be more computationaly convenient form if we consider the property of trace to yield,

$$tr(A_{2jk}Var(\theta_{2jk}|d,\ T,\ \sigma^2,\ \theta_1)A_{2jk}^T) = tr(A_{2jk}^T A_{2k}Var(\theta_{2jk}|d,\ T,\ \sigma^2,\ \theta_1)) \quad (3.23).$$

Equation (3.20) can then be shown to be

$$E(e_{jk}^T e_{jk}|d_{jk},\ T,\ \sigma^2,\ \theta_1) = (d_{jk} - A_{2jk}\theta_{2jk}^*)^T(d_{jk} - A_{2jk}\theta_{2jk}^*) + tr(A_{2jk}^T A_{2jk}Var(\theta_{2jk}|d,\ T,\ \sigma^2,\ \theta_1)).$$

This equation can be translated back to the crossed multilevel form as

$$E(e_{jk}^T e_{jk}|d_{jk},\ T,\ \sigma^2,\ \gamma) = [d_{jk} - (X_{1jk}a_j^* + X_{2jk}b_k^* + X_{3jk}c_{jk}^*)]^T [d_{jk} - (X_{1jk}a_j^* + X_{2jk}b_k^* + X_{3jk}c_{jk}^*)] + tr\ \sigma^2 \left(\begin{bmatrix} X_1^TX_1 & X_1^TX_2 & X_1^TX_3 \\ X_2^TX_1 & X_2^TX_2 & X2^TX_3 \\ X_3^TX_1 & X_3^TX_2 & X_3^TX_3 \end{bmatrix}_{jk} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v3_2 & v_{33} \end{bmatrix} \right) \quad (3.24).$$

Having obtained the conditional sufficient statistic, the ML estimator of σ^2 is easily followed at the M-step as,

$$\sigma^{2*} = (1/N)\ \Sigma\ \Sigma\ E(e_{jk}^T e_{jk}|d_{jk},\ \tau,\ \sigma^2,\ \gamma).$$

The resulting estimates of σ^{2*} will be used as a known variance for the next EM iteration. At this point, we have completed one iteration of EM algorithm.

The EM algorithm described so far does not require an evaluation of the likelihood function. But as the iterative routine of EM algorithm

produces successive values of estimates, it will be useful to monitor the progress of the algorithm by evaluating the likelihood at each iteration. If there is no change between the likelihood with the estimates at one iteration and the likelihood at the next iteration, then we can say the results converge and decide to stop the iteration. I will show the likelihood function for the MLF estimation and then present the log-likelihood function for the crossed multilevel model.

Observed Data Likelihood for the MLF Estimation

The derivation of the observed data likelihood is logically simple if we note the relationship among the several probability density functions in equation (2.3) which can be rewritten for the marginal density function of 'Y',

$$f(Y) = f(Y|\theta)f(\theta)/f(\theta|Y) \qquad (3.25)$$

where $f(Y)$ is the marginal probability density function of Y;

$f(\theta|Y)$ is the posterior density function of θ given Y;

$f(Y|\theta)$ is the likelihood of Y given θ;

$f(\theta)$ is the prior density function of θ.

The above equation is based on general Bayesian model stated at equation (2.1).

In the modified Bayesian mixed model, we consider (θ_1, T, Ψ) to be the parameters and $d = Y - A_1\theta_1$ as the data, hence all density functions are conditioned on (θ_1, T, Ψ) and the equation (3.25) is now restated as

$$f(d|\theta_1, T, \Psi) = f(d|\theta_2, T, \Psi, \theta_1)\ f(\theta_2|T, \Psi, \theta_1) \ /\ f(\theta_2|d, T, \Psi, \theta_1) \qquad (3.26)$$

where $f(d|\theta_1, T, \Psi)$ is the marginal probability density function of the data 'd' given (θ_1,T,Ψ); $f(d|\theta_2,T,\Psi,\theta_1)$ is the likelihood of the data given $(\theta_2,T,\Psi,\theta_1)$; $f(\theta_2|T,\Psi,\theta_1)$ is the prior density function of θ_2 given (T,Ψ,θ_1); $f(\theta_2|d,T,\theta_1)$ is the posterior density function of θ_2 given the data 'd' and the parameters (T,Ψ,θ_1).

By noting the general form of the density functions from Equation (2.4) and from Equation (2.6) and by substituting the relevant terms for the modified Bayesian model in appropriate manner we can specify the each of the right hand side functions of Equation (3.26). The two numerator functions are;

$$f(d|\theta_2,T,\Psi,\theta_1) = [(2\pi)^N|\Psi|]^{-1/2}\exp[(-1/2)(d - A_2\theta_2)^T\Psi^{-1}(d - A_2\theta_2)] \tag{3.27}$$

$$f(\theta_2|T,\Psi,\theta_1) = [(2\pi)^{-1/2}]^F|T|^{-1/2}\exp[(-1/2)(\theta_2{}^T T^{-1}\theta_2)] \tag{3.28}$$

and the denominator function is;

$$f(\theta_2|d,T,\Psi,\theta_1) = [(2\pi)^{-1/2}]^F|D_{\theta 2}^*|^{-1/2}\exp[(-1/2)(\theta_2 - \theta_2{}^*)^T D_{\theta 2}{}^{*-1}(\theta_2 - \theta_2{}^*)] \tag{3.29}$$

where $\theta_2{}^* = (A_2{}^T\Psi^{-1}A_2 + T^{-1})^{-1}A_2{}^T\Psi^{-1}d$, and $D_{\theta 2}{}^* = (A_2{}^T\Psi^{-1}A_2 + T^{-1})^{-1}$. These three equations from (3.27) to (3.29) hold for all θ_2, therefore if we evaluate Equation (3.27) at $\theta_2=\theta_2{}^*$, we can obtain more simplified equations since the only remaining term from the equation (3.29) will be

$|D_{\theta 2}^{*}|^{-1/2}$ after eliminating the constant terms. The resulting equation is then

$$f(d|\theta_1,T,\Psi) = [(2\pi)^N|\Psi|]^{-1/2}|D_{\theta 2}^{*}|^{1/2}|T|^{-1/2}\exp[(-1/2)S(\theta_2^{*})] \tag{3.30}$$

where $S(\theta_2^{*}) = (d - A_2\theta_2^{*})\Psi^{-1}(d - A_2\theta_2^{*}) + \theta_2^{*T}T^{-1}\theta_2^{*}$

$= d^{T}\Psi^{-1}(d - A_2\theta_2^{*})$ (by using $\theta_2^{*}=(A_2^{T}\Psi^{-1}A_2+T^{-1})A_2^{T}\Psi^{-1}d$)

Equation (3.30) is the likelihood we seek.

Log-likelihood for the Crossed Multilevel Model

Given the Equation (3.30), the log-likelihood function is

$$LLF(\theta_1,T,\sigma^2|d) = (-1/2)\log|\Psi| + 1/2\log|D_{\theta 2}*| - 1/2\log|T| - 1/2S(\theta_2^{*}) \tag{3.31}$$

In order to translate this log-likelihood in terms of the crossed-multilevel model, let us consider the four terms; $|\Psi|$, $|T|$, $|D_{\theta 2}^{*}|$, and $S(\theta_2^{*})$. First we consider the equivalence, $\Psi=\sigma^2 I$, in crossed multilevel model as noted at chapter 2. Thus

$$\log|\Psi| = N\log\sigma^2. \tag{3.32}$$

The prior parameter variance T is the block diagonal with the submatrices of cell-level dispersion matrices, r_a, r_b, and r_c that have dimensions of r_1, r_2, and r_3 respectively. Hence the dimension of T is $(Jr_1 + Kr_2 + Gr_3)$. Therefore the log of the determinant of T is the sum of all diagonal terms as

$$\log|T| = J\log|r_a| + K\log|r_b| + G\log|r_c|. \qquad (3.33)$$

For the determinant of posterior dispersion matrix $D_{\theta 2}^{*}$ which is a three-by-three partitioned full matrix, we represent the matrix into two by two partitioned matrix and obtain the determinant of the whole matrix as

$$|D\theta_2^{*}| = \left|\sigma^2 \left|\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}\right|\right| = \sigma^{(Jr_1+Kr_2+Gr_3)} \left|\begin{matrix} d_{22} & d_{23} \\ d_{32} & d_{33} \end{matrix}\right| \qquad (3.34),$$

where $d_{22} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$,

$d23 = \begin{bmatrix} v_{13} \\ v_{23} \end{bmatrix}$, $\quad d_{32} = d_{23}^{T}$,

$d_{33} = v_{33}$.

The second term of the right hand side of the Equation (3.34) is then

$$\left|\begin{matrix} d_{22} & d_{23} \\ d_{32} & d_{33} \end{matrix}\right| = |d_{22}||d_{33}-d_{32}d_{22}^{-1}d_{23}| \qquad (3.35)$$

$$-|v_{11}||v_{22}-v_{21}v_{11}^{-1}v_{12}||d_{33}-d_{32}d_{22}^{-1}d_{23}|.$$

Equation (3.35) can be simplified more by using the previous results of Equations of (2.26) and (2.31) in chapter 2. Using the Equation (2.31), we can obtain

$$|v_{22}-v_{21}v_{11}^{-1}v_{12}| = |H_{22}^{-1}|,$$

where H_{22} is the block diagonal matrix with dimension of Kr_2. Again using the Equation (2.26) we get

$$|d_{33}-d_{32}d_{22}^{-1}d_{23}| = |U^{-1}|,$$

where U is the block diagonal matrix with dimension of Gr_3. Finally $S(\theta_2^*) = d^T\Psi^{-1}(d - A_2\theta_2^*)$ can be replaced with

$$S(\theta_2^*) = \Sigma\Sigma[d_{jk}^T d_{jk} - d_{jk}^T(X_{1jk}a_j^* + X_{2jk}b_k^* + X_{3jk}c_{jk}^*)]/\sigma^2 \qquad (3.36).$$

By substituting $\log|\Psi|$, $\log|D_{\theta 2}^*|$, $\log|T|$, and $S(\theta_2^*)$ of Equation (3.31) with Equation (3.32) through (3.36), we arrive the final form of log-likelihood function for the crossed multilevel model.

$$\begin{aligned} LLF(\sigma^2, T, \theta_1|d) = {} & (Jr_1+Kr_2+Gr_3-N)\log\sigma^2 - (J\log|r_a|+K\log|r_b|+G\log|r_c|) \\ & + \log|v_{11}| + \Sigma\log|H_{22k}^{-1}| + \Sigma\Sigma\log|U_{jk}^{-1}| - [\Sigma\Sigma d_{jk}^T d_{jk} \\ & - \Sigma\Sigma d_{jk}^T(X_{1jk}a_j^* + X_{2jk}b_k^* + X_{3jk}c_{jk}^*)]/\sigma^2 \qquad (3.37). \end{aligned}$$

The EM algorithm evaluates the log-likelihood at each iteration to monitor the progress of the algorithm. The deviance, say δ, between the log-likelihood at i<u>th</u> iteration and its value at (i+1)<u>th</u> iteration will be computed at each iteration. A value of δ close to zero indicates that the log-likelihood at the i<u>th</u> iteration to be very small in comparison with its value at (i+1)<u>th</u> iteration. The actual EM iteration will stop at $\delta \leq k$, where k is predetermined level.

(CW

mul

des

inp

ite

und

bas

be

pro

cro

of

rea

est

est

all

whi

spe

fix

pre

the

the

CHAPTER IV

CHECKING THE ACCURACY OF THE COMPUTING ALGORITHM

The author developed the computer program, "Crossed Multi-Level (CML) algorithm," that provides estimates derived from the crossed multilevel model using Gauss (version 2.0) language. The program is designed to use sufficient statistics of the cross-product matrix as input data and to perform thousands of calculations over numerous iterations of the EM algorithm. It is also designed to analyze data under general crossed multilevel modelling. Therefore an accuracy check based on hand calculation of all equations in numerous situations would be unreasonably demanding and quite unreliable.

An alternative reliable way is to utilize already available computer programs, such as SAS and BMDP, that support some special cases of the crossed multilevel analyses and to use simulation methods for other cases of the model. This approach for checking accuracy of the algorithm is reasonable in two points. First, the computer program performs MLF estimation of the EM algorithm which produces maximum likelihood estimates as noted earlier. Some standard packages, such as SAS and BMDP allow maximum likelihood estimation for variance components models in which only an intercept of regression model at the cell level can be specified as random and all other regression coefficients have to be fixed. The author decided to compare the estimates between the programs presented in this thesis and the standard packages under the assumption these standard packages produce true maximum likelihood estimates when the crossed multilevel model has a random intercept. For the analysis

of crossed multilevel model with multiple random slopes in the within-cell model, a covariance components model, I used a simulation method to see if the program recovers the known parameter values.

The procedure for checking the accuracy of the program is organized in Table 1. The first column of the table lists the crossed multilevel models from the simplest case to the most complicated. The second column of table 1 tells us how the computational results were confirmed to be accurate.

Table 1. Procedure for checking the accuracy of the algorithm

Models	Empirical Evidence
A. Crossed Random Effects Anova Within-cell model: $Y_{ijk} = \beta_{jk} + e_{ijk}$ Between-cell model: $\beta_{jk} = \gamma_0 + a_j + b_k + c_{jk}$	compared to SAS
B. Variance Components Model I Within-cell model: $Y_{ijk} = \beta_{jk} + e_{ijk}$ Between-cell model: $\beta_{jk} = \gamma_0 + \gamma_1 W_{1jk} + \gamma_2 W_{2jk} + a_j + b_k + c_{jk}$	compared to BMDP
C. Variance Components Model II Within-cell model: $Y_{ijk} = \beta_{0jk} + \beta_1 X_{1ijk} + e_{ijk}$ Between-cell model: $\beta_{0jk} = \gamma_0 + \gamma_1 W_{1jk} + \gamma_2 W_{2j} + a_j + b_k + c_{jk}$	compared to BMDP
D. Covariance Components Model Within-cell model: $Y_{ijk} = \beta_{0jk} + \beta_{1jk} X_{ijk} + e_{ijk}$ Between-cell model: $\beta_{0jk} = \gamma_{00} + \gamma_{01} W_{1jk} + \gamma_{02} W_{2jk} + a_{0j} + b_{0k} + c_{0jk}$ $\beta_{1jk} = \gamma_{10} + a_{1j} + b_{1k} + c_{1jk}$	simulation study

Model A to model C represent the variance components models because there are only variances to be estimated. Thus the resulting estimates of variance components are all scalars. Model D is the covariance components model in which the two random within-cell slopes are correlated and the resulting parameter variance-covariance matrices are all two-by-two full matrices. The SAS (see SAS User's Guide, 5th Ed., chapter 41) program was used to obtain the maximum likelihood estimates of the posed models of A. BMDP (see BMDP Manual, Vol. 2., general mixed model analysis, pp.1144-1153) was used for estimating the model B where two group level predictors are involved and the model C where a fixed effect covariate is involved at individual level, an analysis which SAS cannot perform.

The common features of variance-components models in BMDP and SAS lie in the point that they are experimental in nature. The fixed effects sum to zero and the random effects are assumed to be sampled from normal populations with zero means. Although BMDP supports the model with fixed effects covariates at the individual level, SAS requires that all predictors in the model are the group level categorical variables and provides only the estimates of variance components even if fixed effects predictors are in the model. The algorithm for the crossed multilevel model presented in this thesis does not require the predictors in the model to be constrained as class or discrete variables. Because no available computer program allows the estimation of a crossed multilevel model with continuous variables in the model, the author selected the model with discrete variables only for computational comparison. Later at chapter 4 of this thesis, a crossed multilevel model with a continuous covariate in the model will be estimated.

In the case of model D, I conducted simulation analyses for balanced data, since no computer programs are available for estimation of crossed multilevel model when it has multiple random slopes. The simulation method is somewhat judgemental because one should decide the number of replications of the simulation. The current computer program for the crossed multilevel model involves complicated computation and takes a long time for computation. So the author decided to use relatively large size data, N=8400, but limited the replications to twenty times. For the unbalanced data of model D, another simulation analysis would be better for checking the accuracy of the program. However, I analyzed one unbalanced data set that originated from an already used balanced data and compared the results to the balanced case in the hope that if the two results were close enough, then the algorithm is believed to analyze the unbalanced data properly.

Computational Results

The author produced the all data sets to estimate the posed models through random generation using the Gauss programs. To explain the models and data used for accuracy check, following definitions are useful:

1. Design characteristics:

   J = number of macro units for rows (e.g. schools);

   K = number of macro units for columns (e.g. neighborhoods);

   G = number of cells classified by the two sets of macro units;

   n = number of observations of each cell;

   N = total sample size.

2. Model characteristics

Y_{ijk} = outcome score of person i in the cell classified by jth row and kth column units;

β_{qjk} = qth random regression slope of a cell classified by jth row and kth column units;

e_{ijk} = random effect of person i of cell jk. $e_{jk} \sim N(0, \sigma^2)$;

γ = fixed effect of a parameter;

a_j = random effect of jth row units, $a_j \sim N(0, \tau_a)$;

b_k = random effect of kth column units, $b_k \sim N(0, \tau_b)$;

c_{jk} = random interaction effect of jth row and kth column units, $c_{jk} \sim N(0, \tau_c)$.

Crossed Random Effects Anova

Model A is a crossed random effects ANOVA model. The design characteristics are: J=10, K=14, G=140, n=10, N=1400. The distributions from which the values of random effects selected are: $a_j \sim N(0, 16)$, $b_k \sim N(0, 25)$, $c_{jk} \sim N(0, 36)$, and $e_{ijk} \sim N(0, 100)$. The fixed parameter value was assigned as $\gamma = 10$. The fixed effect, γ, and the random parameters at the higher level, a_j, b_k, and c_{jk} constituted the data for each cell mean, β_{jk}. β_{jk} and the individual random effect, e_{ijk} together produced the outcome values, Y_{ijk}. Using Y_{ijk}, two analyses were performed using ML estimation of SAS and the algorithm developed by the author.

For the analysis of unbalanced data, fifteen cells of data were arbitrarily selected out. Hence the design characteristics are changed as: J=10, K=14, g=125, n=10, N=1250. The model is the same as in the balanced case. Table 2 shows the computational results of the CML

algorithm in comparison with the results of the ML estimation of SAS program.

Table 2. Computational comparison between SAS program and the CML algorithm for crossed random effect ANOVA

Covariance components	SAS	CML algorithm
Balanced Data		
τ_a	37.287992	37.287925
τ_b	29.999248	29.999221
τ_c	31.867203	31.867350
σ^2	94.260318	94.260317
γ	12.671820	12.671820
Unbalanced Data		
τ_a	40.036423	40.036586
τ_b	36.3315730	36.315495
τ_c	30.811034	30.811064
σ^2	94.260322	94.260317
γ	12.799353	12.804980

The two sets of results clearly show that two programs produced identical estimates. One thing to note in Table 2 is that the variance components analysis of the SAS program does not produce the estimates of the fixed effects, γ, which is the grand mean of the sample in Model A, so I obtained the observed grand mean to check the accuracy of the estimates of fixed effects, γ, of the CML algorithm.

cla
Hen
des
acc
mac
dum
0 f
Alt
the
cor

when
mean
uni
uni
mul
α_1,
way
is
two
the
BMDP
the

Variance Components Model I

Model B is the typical variance components model for which most classical research on variance components have centered (see Searle and Henderson, 1961; Rao,1972; Hartley and Rao, 1967; Harville 1977). The design characteristics and model characteristics of the data used for the accuracy check are the same as the case of Model A except for the two macro variables in Model B. The two macro variables, W_{1jk} and W_{2jk}, are dummy variables defined on rows and columns respectively. They are coded 0 for the first half units and 1 for the remaining units respectively. Although BMDP provides the estimates of the fixed effects in the model, the parameters are estimated under a general mixed ANOVA model. The corresponding variance components model for model B that BMDP uses is

$$Y_{ijk} = \mu + \alpha_l + \beta_m + a_j + b_k + c_{jk} + e_{ijk} \qquad (4.1),$$

where Y_{ijk} is the outcome value of student i in jkth cell; μ is the grand mean; α_l is the fixed effect of W_1, the macro variable defined on row units; β_m is the fixed effect of W_2, the macro variable defined on column units; a_j, b_k, c_{jk}, and e_{ijk} are the random effects as in the crossed multilevel model. In Equation (4.1), the fixed effects parameters, μ, α_l, and β_m are not the same as γ_0, γ_1, and γ_2 in Model B. An alternative way to check the accuracy of the estimated fixed effects of CML algorithm is to compute the predicted mean values of the groups classified by the two group-level variables using the estimates of the fixed effects from the two programs. The design matrix of the variance components model in BMDP, whose columns are the orthogonal contrasts, and the dummy coding of the fixed effects variables in the crossed multilevel model enable us to

compute the predicted group means from the two analyses. I will first show the computational comparison of the variance components and then move to the case of the fixed effects.

The computational results for the estimates of the variance components from BMDP and crossed multilevel algorithm for balanced and unbalanced data are appeared at Table 3.

Table 3. Computational comparison for the estimates of variance components between BMDP program and the CML algorithm for Model B.

Covariance components	BMDP	CML algorithm
Balanced Data		
τ_a	12.665	12.665328
τ_b	28.111	28.111434
τ_c	31.879	31.879463
σ^2	94.260	94.260318
Unbalanced Data		
τ_a	12.266	12.239884
τ_b	34.666	34.696003
τ_c	31.070	31.080777
σ^2	94.260	94.260289

The first panel of Table 3 shows that the results from each program are identical. For the results from the analysis of unbalanced data, the results are virtually identical but the differences among the estimates, .03 for parameter variances and .0001 for within-cell variance, are a little larger than those of balanced case.

The computational comparisons for the estimates of the fixed effects

Table 4. Computational comparison for the estimates of the fixed effects between BMDP and CML algorithm for Model B (Balanced case)

BMDP			CML		
μ	α_1	β_m	γ_0	γ_1	γ_2
12.672	-4.909	-1.109	6.653551	9.818753	2.217782

Group Codes		Predicted Group Means	
W_1	W_2	BMDP	CML
0	0	$\mu+\alpha_1+\beta_m$= 6.654	γ_0= 6.654
1	0	$\mu+\alpha_1-\beta_m$=16.472	$\gamma_0+\gamma_1$=16.472
0	1	$\mu-\alpha_1+\beta_m$= 8.871	$\gamma_0+\gamma_2$= 8.871
1	1	$\mu-\alpha_1-\beta_m$=18.690	$\gamma_0+\gamma_1+\gamma_2$=18.690

Table 5. Computational comparison for the estimates of the fixed effects between BMDP and CML algorithm for Model B (Unbalanced case)

BMDP			CML		
μ	α_1	β_m	γ_0	γ_1	γ_2
12.717	-5.098	- .911	6.480460	10.322841	1.667013

Group Codes		Predicted Group Means	
W_1	W_2	BMDP	CML
0	0	$\mu+\alpha_1+\beta_m$=6.709	γ_0=6.480
1	0	$\mu+\alpha_1-\beta_m$=16.904	$\gamma_0+\gamma_1$=16.803
0	1	$\mu-\alpha_1+\beta_m$=8.530	$\gamma_0+\gamma_2$=8.148
1	1	$\mu-\alpha_1-\beta_m$=18.725	$\gamma_0+\gamma_1+\gamma_2$=18.470

are appeared in Table 4 for the balanced case and in Table 5 for the unbalanced case.

The first panels of Table 4 and Table5 shows the estimated values of the fixed parameters and the second panels of the tables show the computational comparisons of the predicted group means obtained from the estimated fixed effects from the two programs. The equations of the second panel shows that the predicted mean values of the groups classified by the two group level variables are obtained using the design matrix in BMDP and the dummy coding of the variables in CML algorithm. The results show that the predicted mean values of the groups classified by the group level variables, W_1 and W_2, are virtually the same in both balanced and unbalanced case.

Variance Components Model II

Model C is a more complicated type of variance components model that has fixed effect variables at both levels of units. SAS cannot analyze the data using this model because SAS allows only group level predictors in model specification and the predictors should not be continuous (see SAS User's Guide, chapter 41). BMDP can perform the analysis of this model but it also constrains the group level predictors not to be continuous. In order to compare the computational results, the author coded the individual level covariates, X_{1ijk}, as zero, one, and two (a linear contrasts) and kept the coding systems for group level predictors as before. The sample size of each cell was increased as n=20 because the model has a within-cell covariate.

The design characteristics for the balanced case are then: J=10, K=14, G=140, n=20, N=2800. The chosen random effect distributions are

the same as before. For the analysis of unbalanced data, fourteen cells were arbitrarily selected out and the resulting design characteristics are: J=10, K=14, G=126, n=20, N=2520. The computational results for variance components from BMDP and CML algorithm are presented at Table 6.

Table 6. Computational comparison for variance components between BMDP and CML algorithm for Model C.

Covariance components	BMDP	CML algorithm
Balanced Data		
τ_a	13.565	13.565323
τ_b	17.402	17.402192
τ_c	25.609	25.609418
σ^2	140.532	140.53248
Unbalanced Data		
τ_a	12.569	12.565704
τ_b	14.497	14.501222
τ_c	26.852	26.853643
σ^2	140.532	140.53247

Table 6 shows that the two sets of results from each program are equivalent. The results of covariance components of balanced case are closer each other than those of the unbalanced case.

Concerning the estimates of the fixed effects, the variance components model that BMDP uses is

$$Y_{ijk} = \mu + \beta_1 X_{ijk} + \alpha_l + \beta_m + a_j + b_k + c_{jk} + e_{ijk} \qquad (3.38),$$

where β_1 is the fixed effect of the characteristic of student i in jk<u>th</u>

cell, and all other terms are the same as in model B. As in the case of model B, BMDP uses the design matrix for estimating the fixed effects, μ, α_1, and β_m, thus those fixed effects parameters are different from CML algorithm that estimates the regression coefficients in terms of general linear model.

Table 7 and Table 8 show the computational comparisons for fixed effects between the BMDP and the CML algorithm.

Table 7. Computational comparison for the estimates of the fixed effects between BMDP and CML algorithm for Model C (Balanced case)

BMDP				CML			
μ	α_1	β_m	β_1	γ_0	γ_1	γ_2	β_1
12.185	.113	.207	7.363	12.505	-.414	-.225	7.363

Group Codes		Adjusted Group Means (after covariate)	
W_1	W_2	BMDP	CML
0	0	$\mu+\alpha_1+\beta_m$=12.505	γ_0=12.505
1	0	$\mu+\alpha_1-\beta_m$=12.091	$\gamma_0+\gamma_1$=12.091
0	1	$\mu-\alpha_1+\beta_m$=12.279	$\gamma_0+\gamma_2$=12.279
1	1	$\mu-\alpha_1-\beta_m$=18.865	$\gamma_0+\gamma_1+\gamma_2$=18.865

Table 8. Computational comparison for the estimates of the fixed effects between BMDP and CML algorithm for Model C (Unbalanced case)

BMDP				CML			
μ	α_1	β_m	β_1	γ_0	γ_1	γ_2	β_1
12.215	-.029	.286	7.363	12.583	-.686	.084	7.363

Group Codes		Adjusted Group Means (after covariate)	
W_1	W_2	BMDP	CML
0	0	$\mu+\alpha_1+\beta_m$=12.472	γ_0=12.583
1	0	$\mu+\alpha_1-\beta_m$=11.900	$\gamma_0+\gamma_1$=11.897
0	1	$\mu-\alpha_1+\beta_m$=12.530	$\gamma_0+\gamma_2$=12.668
1	1	$\mu-\alpha_1-\beta_m$=11.958	$\gamma_0+\gamma_1+\gamma_2$=11.982

The first panels of Table 7 and Table 8 show the estimated values of the fixed parameters and the second panels of the tables show the computational comparisons of the predicted group means after accounting for the effect of the fixed effect of individual covariate. The equations of the second panel show that the adjusted mean values of the groups classified by the two group level variables are obtained using the design matrix in BMDP and the dummy coding of the variables in CML algorithm. The results show that the two sets of adjusted group means are virtually the same in both balanced and unbalanced case.

Covariance Components Model

Model D has two random within-cell parameters and the variation of the two parameters across the macro units was explained with different sets of predictors. The two within-cell parameters are not necessarily independent. Hence the resulting dispersion matrices, τ_a, τ_b, and τ_c,

are no longer scalars, rather they are all two by two full matrices. Standard statistical packages cannot analyze data under this kind of covariance components model. Since there is no available computer program for the Model D, the author conducted a simulation study to check the accuracy of the computational results for balanced data. For the complete accuracy check of the program performance, the number of replications of analyses should be large enough, for example ten thousands times, which is unreasonably demanding for the present thesis work. As a compromise, the author decided to conclude that the algorithm works properly if the results show sensible evidence of accurate results. The number of replications of analysis was limited to twenty times but the sample size was taken large enough, N=8400, in order to compensate for the small number of replications. The design characteristics for the simulation data are: J=14, K=20, G=280, n=30, N=8400. The distribution of the individual random effects are designated as e_{ijk} - N(0, 49). For the group level random effects regarding β_{0jk}, the distributions are: a_{0j} - N(0, 16), b_{0k} - N(0, 25), C_{0jk} - N(0, 36). The other set of distributions of the group level random effects regarding β_{1jk} are: a_{1j} - N(0, 9), b_{1k} - N(0, 4), c_{1jk} - N(0,16). The fixed parameter values were assigned as: $\gamma_{00} = \gamma_{10} = 10$, $\gamma_{01} = 5$, and $\gamma_{02} = 3$. The group level predictors, W_{1jk} and W_{2jk}, are all dummy variables coded zeros and ones. The within-cell random effect covariate, X_{1ijk}, has the values of zero, one, and two.

The preassigned parameter values are now compared with the mean values of the estimates obtained through twenty replications of analysis. Table 9 shows the results.

Table 9. Computational comparison between the preassigned parameter values and the average values of the estimates from twenty simulations.

Covariance components	Parameter values	Estimates			
		Mean	SD	Min	Max
σ^2	49	49.09	.72	47.54	50.14
τ_{0a}	16	16.97	7.80	5.75	38.39
τ_{0b}	25	23.82	9.31	8.12	36.57
τ_{0c}	36	37.84	3.63	32.00	44.22
τ_{1a}	9	7.70	3.82	3.19	15.32
τ_{1b}	4	4.52	2.21	1.70	8.52
τ_{1c}	16	15.96	1.64	13.26	20.35
Fixed effects					
γ_{00}	10	9.96	3.52	3.10	15.23
γ_{01}	5	3.47	3.09	-2.47	11.14
γ_{02}	3	3.50	2.92	-4.05	7.50
γ_{10}	10	9.99	1.21	7.33	12.32

Table 9 shows first that the mean values of the estimates are all reasonably close to the preassigned parameter values. Second, all the preassigned parameter values fall within the limit of one standard deviation from the mean estimates. Third, the σ^2, τ_{0c}, and τ_{1c} are closest among the covariance components because these parameters are estimated with the large sample sizes, they are N=8400, G=280. Other estimates of the covariance components are not as close to the parameter values as are σ^2 τ_{0c} or τ_{1c}, due to the insufficient sample sizes such as J=14, K=20. Fourth, for the fixed effects, the mean value of γ_{10} which is the mean of within-cell slopes, β_{1jk}, is closest to the parameter value and has high precision, SD=1.21. Generally speaking, the

slopes have less variation than does the intercept. The results of the fixed effect estimation support this fact. These four characters of the results establish that the CML algorithm produces sensible results for Model D even if the replications of the analysis were limited to twenty times.

For the analysis of unbalanced data, I first built the data by selecting out, using a random digit table, the data of 30 cells from an already used balanced data set. Hence the design characteristics are; J=14, K=20, G=250, n=30, N=7500, but the model characteristics and the preassigned parameter values are the same as in the balanced case. Table 10 show the results of the analysis that compared to the balanced case and the preassigned parameter values.

Table 10. Computational comparison between the preassigned parameter values and the results of the analyses of the balanced and unbalanced data.

Parameters	Parameter values	Estimates	
		Balanced Data	Unbalanced Data
Random effects			
σ^2	49	49.504685	49.152115
τ_{0a}	16	38.388547	23.765979
τ_{0c}	25	17.432644	18.404436
τ_{1a}	9	11.717677	5.432532
τ_{1b}	4	6.440947	6.959900
τ_{1c}	16	18.057778	13.437822
Fixed effects			
γ_{00}	10	14.881167	10.308522
γ_{01}	5	2.130050	3.770741
γ_{02}	3	-1.010730	1.931768
γ_{10}	10	12.318740	10.085683

Table 10 shows that two sets of estimated values have some differences from the preassigned parameter values. These differences occurred because the estimates were obtained from the single samples of the population with the preassigned parameter values. Similarly the differences between the two sets of estimates of the balanced and the unbalanced data are attributable to the fact that the observation was taken from one sample of data. If we replicate the attempt to make an unbalanced data set from the same balanced data and compute the mean values of the estimates from the analyses of the unbalanced data, then the two sets of estimates would be closer. Nevertheless, we can see some patterns of the results in Table 10 as we found in Table 9. The

estimates of σ^2 show the closest values to the parameter values and, both balanced and unbalanced cases show little differences, because the estimates were taken from the large sample size of 8400 and 7500 respectively. The estimates of τ_{0a} and τ_{1a} show somewhat large differences from the preassigned parameter values as well as between the two data cases, because they were taken from the sample size of J=14. The overall information of Table 10 is that the distribution of the estimated values are centered on their parameter values in either cases.

CHAPTER V

ILLUSTRATION

In this chapter, crossed multilevel analysis is illustrated by reanalyzing the data collected by Rudman and Raudenbush (1987). The experience with the crossed multilevel analysis will enhance our understanding on the logic of model specification and answer the following questions:

1. What parameters can be estimated ?
2. How are the hypothesis tested ?
3. How are the results interpreted meaningfully ?

Data and Hypotheses

Rudman and Raudenbush's (1987) study of the effects of excess testing time on test scores provides the data for illustration. The purpose of the study was to assess the influence of providing excess testing time on standardized reading comprehension test scores. The test was not intended to be a speed test, but rather a power test. The time limit of the standardized test has been determined by the "90 % criterion." Using this criterion, testing time is the time elapsed until 90 % of the examinees complete the item analysis edition of the instrument. The assumption underlying the procedure is that the 90 % completing the test had arrived at correct answers to many of the items and had used informed guessing on the remainder. It is also assumed that the remaining 10 % who had not completed the test would merely employ random guessing of given more time. Hence more time would not translate

into mean test score gains. If the test results are sensitive to the provided excess time it becomes important to discover the optimal testing time by estimating the functional form of the relationship between excess time and test scores for the susceptible test. Moreover, such tests may be sensitive to variations in test adminstration procedures and hence would have impact on decisions on student placement in advanced classes, promotion to a higher grade, teacher awards, or other recognitions because of their class' higher test scores.

Sample and Design

In the original study, 29 5th grade teachers from 16 of the 33 elementary schools in the Lansing school district in Michigan volunteered to serve as participants in the study. However data could be collected for only 23 of these classrooms. These 23 fifth grade classrooms supplied usable data for 471 pupils. The data set contains both demographic characteristics (including ethnicity, sex, eligibility for free lunch, etc.), and test scores, pretest scores (including reading subtest scores and total reading scores).

The design of the study involved first, creation of seven blocks based on mean pretest scores each containing four classrooms from the original 29 classes. Four treatment groups were established, each represent testing-time allotments defined by having 0, 5, 10, or 15 minutes excess time to complete the test. Within each of the seven blocks, classrooms were assigned at random to one of the four treatments. One remaining class was added to one of the classes of which the class size was only five. Therefore there are 22 cells of data in two-way classification. The data were well suited to a

polynomial trend analysis. The sample sizes in the design are presented at Table 11.

Table 11. Design and Sample Sizes

	Excess Time				
	None	5 Min	10 Min	15 Min	Overall
Block 1	-	-	-	24	24
Block 2	19	22	27	18	86
Block 3	28	23	20	-	71
Block 4	24	-	24	24	72
Block 5	22	22	23	21	88
Block 6	25	20	15	-	60
Block 7	21	13	17	19	70
Overall	139	100	126	106	471

If the sample had included 28 classrooms, the design would have been a balanced randomized block design (Kirk, 1982, chapter 6) with seven blocks and four treatments. In the study, however, six cells were missing and the sample sizes of the cells classified by the seven blocks and four treatments are not the same. The original study (Rudman and Raudenbush, 1987) used "the least squares solution" (see Searle, 1971) by applying a series of regression models with increasing complexity, and computed the reduction in residual variation on each step. This approach solves the inter-correlation of the main and interaction effects caused by the unbalanced character of the data. However, the amount of variation assigned to each effect will depend on the order of its inclusion. The interpretation of hypothesis testing is also conditional to the order of

its inclusion (Searle, 1971). In addition, the two crossed-factors are considered to have fixed effects in that study. However, the blocks are typically considered as having random effects and the excess time effects are viewed as random also because they are considered as a sample from the population of the excess testing times. Again classrooms are nested within the cells classified by the blocks and treatment groups. Hence the design reflects crossed multilevel data.

Analysis

The general strategy for crossed multilevel analysis can be summarized in three steps: 1. examining variability among students in the hierarchical structure (base model), 2. examining variability among students within the cell, here classroom, classified by blocks and treatment groups (within-cell model specification), 3. identifying variability as a function of group level variables (between-cell model specification).

Base Model

In the first stage of model specification we examine variability of the data where students are nested within the cells, classrooms, classified by the blocks and treatment groups. Addleman (1970) stated two analytical principles related to the data structure. The first one may be considered as a general principle in experimental design, that is

" The design and analysis of experiments should take into account all of the major sources of variability that are expected to influence the responses (p.1,095)."

Second, since students are nested within classrooms, both experimental unit (classrooms) error and the observational unit (individual) error should be included in the model (p. 1,097). With these classical conceptions of model specification, we can see that the present design reflects four sources of variability of students scores; individual differences, block differences, treatment group differences, and the interactions from different combinations between the blocks and the treatment groups. The errors from individual differences are the observational unit errors because the data were taken from the individual scores. The errors from the classroom differences are the experimental unit errors because the classrooms are the units that are randomly assigned to the cells of two-way classification.

In the multilevel conceptions, the observational unit errors are specified under the within-cell model and the experimental unit errors are specified under the between-cell model where the errors are decomposed into three parts. We first pose the within cell model as

$$Y_{ijk} = \beta_{jk} + e_{i:jk}$$

for $i = 1, \ldots, n_{jk}$ student in the classroom that belongs to block j, $j = 1, \ldots, 7$, and treatment group k, $k = 1, \ldots, 4$. Y_{ijk} is the test score of student i in the classroom jk. β_{jk} is a mean test score of students in classroom jk, and $e_{i:jk}$ is an individual effect of student i nested within ":" the classroom jk. It is assumed that $e_{i:jk}$ is normally distributed within each treatment by block with mean 0 and constant variance σ^2. The within-cell model is a traditional

regression model with no predictors in the model except β_{jk}, class means.

Thus we pose between-cell model to examine variability of the classroom means, β_{jk}. The between-cell model is

$$\beta_{jk} = \gamma_0 + a_j + b_k + c_{jk}.$$

Each class mean score, β_{jk}, becomes the outcome variable which is a function of the grand mean, γ_0, plus the jth block effect, a_j, the kth treatment effect, b_k, and the effect of interaction between jth block and kth treatment group. The distributional assumptions of the three error terms are; $a_j \sim N(0, \tau_a)$; $b_k \sim N(0, \tau_b)$; and $c \sim N(0, \tau_c)$. The three error terms are mutually independent. Hence the variance of class mean scores, parameter variance, is $\mathrm{Var}(\beta_{jk}) = \tau_a + \tau_b + \tau_c$, which is the total variance attributable to the differences of the classrooms. The full results for this model are reported in Table 12.

Table 12. Results of crossed multilevel analysis of base model

Fixed Effects	Estimates	SE	t
γ_0	41.535	.459	90.569
Covariance Components			
Within-cell variance			
σ^2	99.059		
Between-cell variance			
Overall	τ_a	τ_b	τ_c
5.221	0.901	1.667	2.647
-2ln(maximum likelihood)	3515.8118		

Table 12 shows the estimation results of both fixed effect, γ_0 (intercept), and the covariance components. These estimates are maximum likelihood estimates since the CML model uses MLF estimation method of EM algorithm. For the fixed effect, γ_0 (intercept), we can test the hypothesis, H_0: $\gamma_0 = 0$, but it is not of interest at this point. The intercept is the grand mean of all classroom mean scores.

The critical point in Table 12 is decomposition of the between-cell variance into three parts; τ_a , τ_b, τ_c. The total observed variance is $Var(Y_{ijk}) = \sigma^2 + \tau_a + \tau_b + \tau_c = 104.28$. The between-cell variance is 5.221 and the within-cell variance is 99.059. We can compute various intra-unit correlations using the decomposed covariance components. First, we can estimate the proportion of variance that lies within- and between- classrooms. About 95 % of the observed variance is at individual level. Only 5 % of the observed variance is attributable to the differences among classroom memberships. Classical-single level models, i.e., regression or ANOVA, do not provide the parameter variance.

Nested-multilevel models such as ML2 or HLM provide the parameter variance but the analysis of variance components would stop here. The crossed multilevel model allows further decomposition of the variance. We would like to know what proportion of the variance lies between blocks or treatments or their interactions. Seventeen percents of the parameter variance and one percent of the observed variance reflects variation among blocks. About 32 % of the parameter variance is attributable to the treatment group differences. Again 51 % of the parameter variance and 2.5 % of the observed variance reflect random interaction effects, which means the variation among students with different classroom membership after taking out the effect of blocks and treatment.

. Kang and Raudenbush (1988) performed a comparison between a classical analysis and HLM using the same data. In that study the observed variance was decomposed as σ^2=98.73 and τ=5.5397. The two estimates are virtually the same as the estimates in Table 12. An interesting point is that the parameter variance from HLM analysis is the sum of the three variance components at macro level. With the HLM analysis we don't know the proportion of the parameter variance that is solely attributable to the differences among treatment group memberships of students. The tiny differences among the estimates from HLM and CML analysis may be caused by two things. One thing is that HLM uses MLR estimation which is equivalent to REML (Patterson and Thompson, 1971; 1974; Harville, 1977) but CML uses MLF estimation which produce ML estimates. The other thing is rounding errors when the two programs compute the estimates.

In sum, the crossed multilevel analysis provides all necessary information regarding variance decomposition of the base model and guides further model specification. Having decomposed the covariance

components, we now move our interest to account for variability of the student scores at each level. We first attempt to account for the individual level variance.

Within-Cell Model Specification

As shown in Table 12, about 95 % of the observed variance lies at the individual level within each classroom. To account for the variation we need to identify the covariates from individual level variables. In the original study (Rudman & Raudenbush, 1987), the best single covariate proved to be the total reading pretest, r=.75. Only one covariate was needed because other likely covariates, (e.g. ethnicity, sex, parent education) were not significantly related to the outcome after adjusting for the effects of the best covariate. The traditional single-level analysis, as in the original study, treats the effects of individual-level variables as fixed, which implies that the effects of the individual characteristics on the outcome variable are constant across the all classrooms, blocks, and the treatment conditions. However we don't know whether the effects of the chosen covariate on students' test scores are constant across the all classrooms. The effects of the covariate may work differently across the classrooms. Suppose certain teachers provide individual teaching for the students who need remedial study after the pretest, but some other teachers do not. Then the correlations between the posttest scores and pretest scores may vary across the classes. The correlations of the two test scores of the classes with remedial teaching would be smaller than those of the other classes. Therefore we first examine the variation of the covariate effects across the classes, and specify it as having a random effect. We

pose the within-cell model as

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}(\text{Tread})_{ijk} + e_{i:jk}.$$

The variable code "Tread" indicates total reading pretest scores. Note the effect of the covariate, β_{1jk}, has subscripts in the model, which allows the parameter to have different values across the classes. The between-cell model does not need any predictors at this moment, but we need to present a pair of equations because the two within-cell parameters are specified as random. The between-cell models are

$$\beta_{0jk} = \gamma_{00} + a_{0j} + b_{0k} + c_{0jk}$$
$$\beta_{1jk} = \gamma_{10} + a_{1j} + b_{1k} + c_{1jk}$$

where $a_{0j} \sim N(0, \tau_{0a})$, $b_{0k} \sim N(0, \tau_{0b})$, $c_{0jk} \sim N(0, \tau_{0c})$. Similarly $a_{1j} \sim N(0, \tau_{1a})$, $b_{1k} \sim N(0, \tau_{1b})$, $c_{1jk} \sim N(0, \tau_{1c})$. The random effects within each between-cell equation are mutually independent. However the two outcome variables, β_{0jk} and β_{1jk}, are not necessarily independent of each other. Thus the random effects within each macro unit, for example a_{0j} and a_{1j}, are correlated across the two equations, and resulting covariance components at macro level will be two by two full matrices. The results of estimating this crossed multilevel model are shown at Table 13.

Table 13. Results of crossed multilevel analysis with random slopes at within-cell model

Fixed Effects	Estimates	SE	t
γ_{00}	12.093	1.19576	10.114
γ_{10}	.3848	.01516	25.3826
Covariance Components			
Within-cell variance			
σ^2	42.474		
Between-cell variances			

Overall		τ_a		τ_b		τ_c	
9.2207	-.1025	2.4059	-.0123	1.8698	-.0405	4.9500	-.0497
	.0016		8.346E-5		.0010		.0006

-2ln(likelihood) 3122.8877

The major concern of this analysis was to see whether the covariate, Tread, has a fixed or random effect. The results of Table 13 clearly shows the slope of the covariate does not vary so much across the blocks, τ_{1a}=8.346E-5; the treatment groups, τ_{1b} =.0010; and the interactions, τ_{1c}=.0006. We can test the hypotheses, H_0: $Var(a_1j) = Var(b_{1k}) = Var(c_{1jk}) = 0$ with the likelihood ratio test by using the values of -2ln(likelihood). It is known that the statistic, $-2\ln(L_1/L_2)$, has an asymptotic χ^2 distribution, where L_1 is the maximum likelihood value of a less complex model and L_2 is the value of a more complex model and the degrees of freedom of χ^2 statistic is the difference of the number of parameters to be estimated in each model. Hence the deviance between the two values of -2ln(likelihood) will have χ^2 distribution with the difference of the number of parameters between the two models as the degrees of freedom. The number of parameters of the model in Table 13 is twelve, while only five parameters were estimated in the model of

Table 12. The deviance between the two values of -2ln(likelihood) is 392.9241 which is large enough to reject the χ^2 statistic with seven degrees of freedom. Therefore we reject the composite null hypotheses, H_0: $\gamma_{10} = 0$, $Var(a_{1j}) = Var(b_{1k}) = Var(c_{jk}) = 0$. We, however, do not know which particular hypotheses are rejected given the all posed hypotheses. To identify the significant effects by using the loglikelihood ratio test, we need to specify another model that has fewer parameters.

Table 13 shows that the within-cell variance has been reduced from 99.059 to 42.474 by virtue of the covariate. Concerning the fixed effects, the average slope of the covariate across the all cells is .3848. There are other results regarding the within-cell intercept, β_{0jk}, but the intercept is meaningless because the covariate was coded with its raw scores.

Results shown at Table 13 reveal some characteristics of the crossed multilevel model. Traditional variance components models and single-level models that are available through SAS or BMDP can't perform this analysis,, where multiple random within-cell parameters are modeled with appropriate error terms. Under classical variance components models, all covariates must have fixed effects and estimation of random effects is limited to variance components.

The difference between nested and crossed multilevel analysis is again shown at Table 13. While crossed multilevel analysis partitions the covariance components into three matrices, τ_a, τ_b, and τ_c, nested multilevel analysis, however would provide the overall matrix only. In Table 13, both overall and the three separate variance-covariance matrices show that the sizes of the variance of the covariate effect are small. Hence both analyses may agree to fix the effect of covariate in

this case. However when the overall parameter variance of the covariate is substantial, a nested multilevel analysis would provide no information about which macro units cause the substantial size of parameter variance while crossed-multilevel model identify what particular random effect among macro units are significant. Thus the crossed multilevel model provides clear guide for further model specification.

Although the sizes of the variances of the covariate is small, we still need to test the significance of the variance components by fixing the effects of the covariate to get a deviance of -2ln(likelihood) between the model of Table 13 and the new model with fixed covariate effects. Thus the second between-cell model is cancelled and the within-cell model become

$$Y_{ijk} = \beta_{0jk} + \beta_1 \, (\text{Tread})_{ijk} + e_{ijk}$$

and the between-cell model has only one equation as

$$\beta_{0jk} = \gamma_{00} + a_{0j} + b_{0k} + c_{0jk}.$$

The new results of the analysis are shown at Table 14.

Table 14. Results of crossed multilevel analysis with a fixed within cell slope

Fixed Effects	Estimates	SE	t
γ_{00}	12.2879	1.2044	10.2029
β_1	.3826	.0153	25.055
Covariance Components			
Within-cell variance			
σ^2	43.0875		
Between-cell variances			
Overall	τ_a	τ_b	τ_c
2.984	1.1102	1.2682	.6058
-2ln(likelihood)	3125.2379		

Table 9 shows a highly significant (t=25.055) covariate effect. The within-cell variance at Table 12 was 99.059 but it was reduced to 43.0875 here. About 57 % of the within-cell variance was accounted by the covariate. The overall between-cell variance was also reduced from 5.2205 to 2.984. The last line of Table 14 shows that -2ln(likelihood) is 3125.2379 and the deviance between the values in Table 13 and Table 14 is 2.3502. The χ^2 statistic with six degrees of freedom is 12.59 at 5 % significance level. Hence we do not reject the null hypotheses, H_0: $Var(a_{1j}) = Var(b_{1k}) = Var(c_{1jk}) = 0$ and decide to fix the effects of the covariate across the blocks, treatment groups and the block-by-treatment interactions.

The parameter variance estimates, $Var(\beta_{0jk}|Tread)$, become conditional variances. They measure the amount of variability remaining among the class means. Raudenbush and Bryk (1986) used a measure of model performance, R^2 as in regression, by comparing it to the

unconditional parameter variance estimates from the first stage. Equivalent measures can be used for crossed multilevel model. Hence the proportion of explained parameter variance by the model is

$$R^{2*} = \frac{\text{Var}(\beta_{jk}) - \text{Var}(\beta_{jk} \mid \text{Tread})}{\text{Var}(\beta_{jk})} = \frac{5.2205-2.984}{5.2205} = .428.$$

About 43 % of the parameter variance was explained by the model or the covariate in this case. The R^{2*} presented above is never less than the one obtained through OLS estimation because the OLS estimates, $\hat{\beta}_{jk}$, include true parameter value plus sampling error as $\hat{\beta}_{jk} = \beta_{jk} + e_{jk}$ but the EM estimates, β_{jk}, is true parameter value. In the previous study (Kang and Raudenbush, 1988) the within-cell variance was 43.22, the overall parameter variance was 2.915, and the total OLS between-cell variance, $\text{Var}(\hat{\beta}_{jk})$, was 10.987. Based on these statistics, the proportion of explained parameter variance by HLM is 42.4 % and by regression approach is 23.9 %. It clearly shows that both multilevel models are more reliable than the classical regression model in terms of the coefficient of determination (R^2). It also shows that the parameter variance from nested multilevel analysis is the sum of the three covariance components at macro-level variances in crossed multilevel model. The nested multilevel model cannot identify the separate error sources in the model in crossed multilevel contexts.

Having completed the specification of the within-cell model, we now need to identify the variability among classrooms as a function of

between-classroom variables.

Between-Cell Model Specification

The general principle for modelling strategy for the statistical model specification is to build a parsimonious model. In the crossed multilevel model, the χ^2 statistic which is the deviance of the-2ln(maximum likelihood) and which allows us to test a composite hypothesis, serves as a criterion for model specification. Available group level predictors in the original study (Rudman and Raudenbush, 1987) were linear trend of the blocks, and the polynomial trends variables (linear, quadratic, and cubic) of the treatment groups. The linear trends of the blocks and treatment groups were significant in that study after examining all possible sets of predictors. In the present study, the author decided to use only the polynomial trends variables of the treatment groups for illustration purpose. By using the predictors taken from one dimension of two-way classification, we are able to see changes of each covariance component at the group level.

We first pose the model with all polynomial trends variables and then take out the nonsignificant predictors from the model. The within-cell model is the same as before.

$$Y_{ijk} = \beta_{0jk} + \beta_1(\text{Tread})_{ijk} + e_{i:jk}$$

but the between-cell model includes the predictors as

$$\beta_{0jk} = \gamma_{00} + \gamma_{01}(\text{Lin})_{jk} + \gamma_{02}(\text{Quad})_{jk} + \gamma_{03}(\text{Cub})_{jk} + a_{0j} + b_{0k} + c_{0jk'}$$

where Lin = linear trend, Quad = quadratic trend, and Cub = cubic trend of the treatment effects variables. All other notations are the same as before.

The results of the analysis of the above crossed multilevel model are shown at Table 15.

Table 15. Results of crossed multilevel analysis with both within- and between- cell variables.

Fixed Effects	Estimates	SE	t
γ_{00}	12.3233	1.2079	10.20
γ_{01}	.5728	.1342	4.27
γ_{02}	-.0670	.3049	-.22
γ_{03}	.0158	.1383	.11
β_1	.3824	.0153	25.02
Covariance Components			
Within-cell Variance			
σ^2	42.9445		
Between-cell Variances			
Overall	τ_a	τ_b	τ_c
1.39728	1.08924	.01114	.2970
-2ln(likelihood)	3116.4663		

The results shows that the χ^2 value of the log-likelihood ratio test is 8.77 with three degrees of freedom, but χ^2 value at 5 % significance level is 7.82. So the composite hypotheses, H_0: $\gamma_{01} = \gamma_{02} = \gamma_{03} = 0$ is rejected. The first panel of Table 10 shows the results of t-test for the fixed effects and the fixed effects of quadratic and of cubic trends

effects are not significant in predicting the variation of within-cell intercept. So these two predictors will be dropped in next model specification. The residual parameter variances are reduced by virtue of the predictors at group-level. But the residual parameter variance of the random block effect, τ_a, shows the least change because the employed predictors are mainly supposed to account for the random treatment effects. One notable result in Table 15 is that the within-cell variance was not changed by the effects of the group level predictors. This is because the group level variables predict the variation of the responses only at group level.

Since we know only the linear effect of the treatment is significant in predicting the variation of within-cell intercept, we pose the final crossed multilevel model. The within- and the between-cell models are

$$Y_{ijk} = \beta_{0jk} + \beta_1(\text{Tread})_{ijk} + e_{i:jk}$$
$$\beta_{0jk} = \gamma_{00} + \gamma_{01}(\text{Lin})_{jk} + a_{0j} + b_{0k} + c_{0jk}.$$

The results are shown at Table 16.

Table 16. Results of crossed multilevel analysis of the final model

Fixed Effects	Estimates	SE	t
γ_{00}	12.30087	1.20408	10.21
γ_{01}	.57793	.13336	4.33
β_1	.38256	.01526	25.07
Covariance Components			
Within-Cell Variance			
σ^2	42.9431		
Between-Cell Variances			
Overall	τ_a	τ_b	τ_c
1.4134	1.1045	.0113	.2976
-2ln(likelihood)	3116.5100		

Table 16 shows that the estimated fixed effect of the linear trend is .578 and the change of the -2ln(likelihood) is almost zero. The classical interpretation of the estimated effects in regression analysis can be used for the fixed effects in the crossed multilevel model, that is the average change of outcome scores for one unit change of linear trend variable is .578. The linear trend variable was coded as; group 1 = -3; group 2 = -1; group 3 = 1; group 4 = 3; in the study. Hence the expected mean change of reading test scores for 5 minutes excess testing time is the twice of the coefficient which becomes 1.156.

Concerning the covariance-components at macro level, we found the size of the variances at both within- and between-cell levels was not changed a lot after deleting the two non-significant trends effects. Comparison of the results with the results of Table 14 tells us that the within-cell variance was not affected by the effect of the linear trend

variable which is from treatment groups, but the magnitude of residual parameter variance of random treatment effects were reduced from 1.268 to .0113. Hence about 99 % of the random treatment effects were accounted by the linear trend variable.

We can compute the coefficient of determination for parameter variance here. The proportion of the explained parameter variance by the model is

$$R^{2*} = \frac{Var(\beta_{jk}) - Var(\beta_{jk} \mid Tread, Lin)}{Var(\beta_{jk})} = \frac{5.2205 - 1.4134}{5.2205} = .729$$

About 73 % of the parameter variance was explained by the model.

Based on this results regarding fixed effects, we could conclude as in the original study: 1. Test designers should consider more precise methods of setting testing time limits, 2. tests which are sensitive to variations in the procedures of administration should not be used for high-stakes decisions unless the procedures can be carefully monitored (Rudman and Raudenbush, 1987, p.14).

CHAPTER VI

CONCLUSION

Summary

Educational systems typically have hierarchical organizations in which "units" at one level are "nested" within units at the next higher level. These educational systems often produce hierarchical data. There have been many controversies over results found through traditional statistical analyses. As Cronbach (1976) and others emphasized, many educational studies have used inappropriate analyses, including many important evaluation studies. A number of methodologists have developed multilevel models with estimation procedures appropriate for multilevel data. Although these multilevel models have made substantial methodological advances in analyzing multilevel data, they apply only to those multilevel data structures in which each lower-level unit belongs to only one unit at the next higher level. In many cases, however, the structure of a system is not so simple. Students may belong to more than one group simultaneously. For example, we could cross-classify students both by the school they attend and by the neighborhood they live in. We call this kind data as "crossed multilevel data."

Despite the recognition of a need for an appropriate multilevel model analyzing crossed multilevel data, a major difficulty has been constrained by existing technologies such as one-way nested multilevel models, variance components models, and OLS regression models. This thesis has now expanded the multilevel techniques to include the two-way crossed multilevel model. The major products of this thesis are five things:

1. A crossed multilevel statistical model has been presented in general form.

2. The Empirical Bayes estimation procedure has been adapted to the estimation of the crossed multilevel model.

3. A computing algorithm (the CML algorithm) for numerical analysis of crossed multilevel data has been developed using the Gauss language.

4. The accuracy of the computing algorithm has been tested in reasonably comprehensive situations of crossed multilevel modelling both for balanced and unbalanced data sets.

5. The application of the crossed multilevel model to real data set was learned through illustration.

In chapter one, the issues of crossed multilevel analysis in educational research were identified and the statistical model for crossed multilevel analysis was presented. The problems of existing statistical models for analyzing crossed multilevel data resulted because those models can't specify the error terms from all sources in crossed multilevel contexts. Traditional single level models, such as regression models, fail to incorporate the error sources from multilevel structure. Such a limitation of the models leads researchers to have an enforced choice of either individual level or group level analysis. Because the observations within a group are not independent, significance tests based on individual level analysis are not acceptable, due to the violation of independence assumption. Dependencies among observations cause over-estimation of the precision which is a function of sample size and intra-class correlations. The group level analysis does not violate the independence assumption but cannot use individual variables in the model. Inferences about individual behavior (e.g., individual student achievement) based on group level analysis can cause aggregation bias. The limitation of classical statistical analysis is further exacerbated

when researchers raise the inquiries about how the process in higher level units (e.g., organizational treatment) influence the process at lower level units (e.g., students achievement). The effect of treatment at the organization level on the individual students may be different due to the differences of students backgrounds. Traditional analysis must assume that the organizational treatment effects on individual behavior are homogeneous, which is hardly acceptable.

The currently available nested multilevel models virtually solve these problems when they are applied for the analysis of one-way nested multilevel data. Those nested multilevel models, however, are not appropriate for the analysis of crossed multilevel data because the models specify only one group-level error term, while crossed multilevel data have three error sources, two sets of macro units plus their interactions, at the group level. Application of nested multilevel model for the analysis of crossed multilevel data would produce large variance, T, at group level which will be the sum of three group level variances, $T = \tau_a + \tau_b + \tau_c$, obtainable from a crossed multilevel model. The exact functional relationship between nested multilevel and crossed multilevel models regarding the group level precision is not clear to the author at this point. It will be a interesting study to investigate the significance tests when a nested multilevel model is used for the analysis of a crossed multilevel data. One thing we know is that the nested multilevel model does not identify the all error sources of the crossed multilevel data at group level.

Inquiries into the use of the crossed multilevel model have begun with several concrete examples. In the case of school effects studies, many schools can be examined for multiple time points in a single crossed

multilevel analysis if the observed students are different across the time periods. Furthermore, the crossed multilevel model allows the estimation of random interaction effects which can't be achieved through the existing nested multilevel models. Significant random interaction effects constrain the generalizability of the research findings and lead us to identify what particular fixed effects of group characteristics are interactively at work.

In chapter two, the empirical Bayes estimation method using EM algorithm was adapted for the crossed multilevel model. The essential idea of the estimation theory is twofold: First, the crossed multilevel model is viewed as a case of general mixed linear model, which enables us to build the identities between the two models. The general mixed linear model can be modified into either a Bayesian model or a linear model. Then Bayesian estimation method and generalized least squares method were utilized for the estimation of random, θ_2^* and D_θ^*, and fixed parameters, θ_1^*, respectively.

Second, the MLF method of EM algorithm was used for finding ML estimates of variance and covariances. In EM algorithm, the most complex computation is to obtain the posterior dispersion matrix which requires matrix inversion, $D_\theta^* = (A'\Psi^{-1}A + T^{-1})^{-1}$. The advantage of MLF over MLR method for computing the posterior dispersion matrix of crossed multilevel model is that the dimension of the matrix to be inverted is smaller in MLF (3 x 3 partitioned matrix) than in MLR (4 x 4 partitioned matrix).

Based on the estimation method presented in chapter two, chapter three provides actual computational EM formula which have been used for the CML computer program. In chapter four, the accuracy of the

performance of the CML algorithm was examined through reasonably comprehensive evaluations and all the resulting estimates were virtually identical to ML estimates computed from either standard packages (SAS, BMDP) where those programs apply or simulations. The study provided the evidence confirming the accuracy of the algorithm.

In chapter five, the use of crossed multilevel model was illustrated by applying the model to the analysis of actual experimental data. The experience with the crossed multilevel model brought up insights about the logic of model specification, hypothesis testing, parameters to be estimated, and the interpretation of the findings. It also showed that the use of the crossed-multilevel model in crossed-multilevel contexts has advantages over single-level models in the estimation of covariance components, inference on the hypothesis testing, and the coefficient of determination (R^2). The basic procedure of model specification consists of three steps. The first step is to examine variability of the observed outcome variable with respect to the crossed multilevel structure. We call the model for this purpose the base model since no predictors are included in the model. Using the base model, we obtained variance-covariance components estimates at both the individual and group levels. This decomposition of the observed total variance enables us to obtain three intra-unit correlations: intra-cell, intra-block, and intra-treatment. The classical single-level models do not allow the estimation of parameter variance and the OLS estimate of the group-level variance is contaminated with sampling-errors. Nested-multilevel models provide the parameter variance but cannot decompose the parameter variance into the three components.

Having examined the decomposition of the variances and covariances,

we tried to explain the process within each classroom through within-cell model specification. The power of crossed multilevel model in this step is that the model clearly shows what effects of individual characteristics are random across what particular macro factors. Classical-single level models always assume that the effects of individual characteristics are constant across all macro units, which is hardly acceptable. Nested-multilevel models allow the effects of individual characteristics to vary across the macro units but they cannot identify what particular macro factors, among the three in crossed-multilevel contexts, are the source of variation of the slopes at individual level. In our example, the effect of total pretest score (Tread) showed little variation across the all three macro factors; treatment groups, blocks, and the classrooms classified by the blocks and treatment groups. Thus we decided to fix the effect of the variable.

Finally, group level variables were examined in order to identify the variability of the observed outcome score as a function of group level variables. In this step, I examined the treatment variables to show the changes of parameter variances between the blocks, τ_a, and the treatment groups, τ_b. It clearly showed that the parameter variance pertinent to blocks, τ_a, did not change for blocks when we used the variables of treatment characteristics in the model. Thus we could know that the treatment variables mainly explain the variability of the outcome values across the treatment groups.

As many multilevel methodologists anticipated the emergence of a general crossed multilevel model, we now have the model. The crossed multilevel model presented in this thesis, however, is not yet completed for those audiences who want to use it. The limitations and the expected necessary future work for this model are now considered.

Limitations

1. Computational Complexity

Complexity of computing estimates of crossed multilevel model come from the three facts: First, the crossed multilevel model uses the variables at both micro and two sets of macro units. Because the CML algorithm is designed to use the model for general applicability, the algorithm allows estimation for a wide range of models with various combinations of predictors at both levels as in chapter three. These complex model characteristics and general applicability of the model make the algorithm nearly 1000 lines long in Gauss language. If we consider that Gauss, unlike the fortran, is a matrix operation oriented program language, we could imagine the complexity of computation.

Second, the crossed multilevel model can be applied for experimental data as well as for large scale survey data such as national data sets. For the analysis of large scale data, computation using the whole data matrices is not always possible due to the limitation of virtual memory, especially in micro computers. An alternative way is to compute the sum of squares and cross-products of each cell separately and recapture the necessary statistics based on computed cell statistics. This unit by unit computation approach cause the program complex but useful for computation of large scale data sets. Although computation is complex, this computational strategy allows us to perform the crossed multilevel analysis using micro computers of 386, AT, and even XT.

Although more work will enable us to improve the efficiency of performance of the CML algorithm, the final version would still definitely be complex. The current CML algorithm was written with Gauss

language (version 2.0) so it can be used only for micro computers with appropriate math-coprocessors. The Gauss software (version 2.0) allows matrix inversion up to a 90 X 90 matrix, which enables us to use a fairly large data set. Increasing popularity of micro computer use for computation would increase the use of the crossed multilevel analysis, but we still need to have the algorithm for mainframe computer.

2. Sensitivity to Assumptions

Empirical Bayes estimation requires the assumption that data and the prior distribution of the random effects are normal. Although the distribution of outcome measures can be examined, the prior distribution of parameter effects is actually unknown. The sensitivity of results to violations of this assumption are not known.

Future Work

1. Elaboration of the Algorithm

The current version of CML algorithm is not an efficient one at this points. It reads the sufficient statistics multiple times at each computational step because it involves the computational components for each set of macro units along with their interactions. Simplification of the program is possible with more experienced computer programming. Additional work needed for the program is to include all necessary test statistics. The current program allows us to estimate the fixed effects parameters with significance tests and the variance-covariance components at each level. Various other test statistics, for example confidence intervals need to be added in the program and these are expected to be done very soon.

One strong advantage of empirical Bayes estimation is that it can decompose the within-cell parameter variance into true and sampling variances. For example, $\hat{\beta}_{jk} = \beta_{jk} + e_{jk}$, where the outcome is the OLS estimate of a cell coefficient, β_{jk} is the true coefficient value and e_{jk} is a sampling error. Then $Var(\hat{\beta}_{jk}) = Var(\beta_{jk}) + Var(e_{jk}) = (\tau_a+\tau_b+\tau_c) + v$, where the left hand side of the equation is the observed parameter variance and $Var(\beta_{jk})$ is the true parameter variance and $Var(e_{jk})$ is the sampling variance. At this point, the CML algorithm produces only the true parameter variances, τ_a, τ_b, τ_c and within-cell variance, σ^2. The importance of this decomposition of true and sampling variance of the observed within-cell parameters lies in the points that we could compute the reliability, $\rho = Var(\beta_{jk})/Var(\hat{\beta}_{jk})$, of the estimation of within-cell parameters and could compare the performance of EB estimation and traditional OLS estimations and obtain the technical properties of EB estimates. EB estimators were known as "shrunken estimators",i.e., estimators shrunken toward their unconditional grand mean (unconditional) or their conditional grand mean (conditional) (Rubin,1981; Strenio, 1981; Raudenbush, 1984). This property of EB estimator can be proved by using the decomposition of the observed parameter variance into true and sampling variance.

Finally the CML algorithm is not user friendly. More technical programming work is necessary for the popular use of the algorithm.

2. Integration of Crossed Multilevel and Nested Multilevel Models

The crossed multilevel model in this thesis is the expansion of a nested multilevel model that shares the same modelling strategies, hypothesis testing , and the interpretation of the findings. These

similarities encourage us to combine the two models. To denote the possibilities, let me consider the simple case in which the crossed multilevel model can work like the nested multilevel model. In nested multilevel model, the base model is

$$Y_j = \beta_j + e_j \qquad e_j \sim N(0, \sigma^2 I) \quad \text{(within-group model)}$$
$$\beta_j = \gamma_0 + u_j \qquad u_j \sim N(0, \tau) \quad \text{(between-group model)}$$

The within-group model shows a vector of outcome scores within j<u>th</u> group is a group mean, β_j, plus random errors, e_j. Each group mean, β_j, is the function of a grand mean, γ_0, and the random effect of the group, u_j.

Consider now the base crossed multilevel model,

$$Y_{jk} = \beta_{jk} + e_{jk} \qquad e_{jk} \sim N(0, \sigma^2 I)$$
$$\beta_{jk} = \gamma_0 + a_j + b_k + c_{jk} \qquad a_j \sim N(0, \tau_a),\ b_k \sim N(0, \tau_b),\ c_{jk} \sim N(0, \tau_c)$$

The evident difference between the two models is that crossed multilevel model has two subscripts, macro row units j and macro column units k, and the cells, jk, classified by the two sets of macro units. The relationship between the two models regarding the parameter variances is $\tau = \tau_a + \tau_b + \tau_c$. Suppose the within-cell parameter, β_{jk}, does vary only across the macro row units, so $\tau_b = \tau_c = 0$, then we would fix the effects of column and the cell units on the within-cell parameters as zeros. In this case the model become

$$Y_{jk} = \beta_{jk} + e_{jk} \qquad e_{jk} \sim N(0, \sigma^2 I)$$
$$\beta_{jk} = \gamma_0 + a_j \qquad a_j \sim N(0, \tau_a).$$

So the within-cell parameter has only one random effect defined on rows. In this new model, $\beta_{jk} = \beta_j$ and $e_{jk} = e_j$, also $\tau_a = \tau$. Thus the base nested model and the new crossed multilevel model become identical.

Since the CML algorithm is developed for general applicability, it supports the model above. A distinction between the nested multilevel and crossed multilevel model aries when we have longitudinal data sets or time-series data. Consider a number of students are observed across the multiple time points. The design may be viewed two-way crossed (student j and time k) design with one observation in each cell. But the observations across the times are not independent. While the nested multilevel model views the design as an oneway nested design, multiple observations are nested within students, and so allow the dependencies of the observations, the crossed multilevel model treats the multiple observations, time k, as macro units and considers them as independent. So the crossed multilevel model can't analyze the data appropriately.

Similar problem arise when we have longitudinal data. Suppose many schools are observed through multiple time points using the same samples of students from each school. The design of this data may be viewed as two way cross-sectional, the schools (row macro units), the time points (column macro units). But the column units are not independent because observations are taken from the same sample at different time points. The crossed multilevel model assumes the two sets of macro units are all independent. The 3-level nested multilevel model handles this longitudinal data because it considers the multiple observations nested within students which are also nested within schools at a time point (see, Raudenbush, 1989). The difficulty of the integration of the two

models stems from the fact that the crossed multilevel model uses the time points as macro units but the 3-level nested multilevel model considers the time points as the micro units nested within each student. If an algorithm allows the transition of micro- and macro-units in its configuration of data, there may be a way to integrate the two models. But we need to have a comprehensive comparison of the characteristics of the two models to improve our understanding about the combination of the two models.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aitkin,M.,Anderson,K.,& Hinde,J.(1981). Statistical modeling of data on teaching styles.Journal of the Royal Statistical Society,Series A, 144(4), 419-461.

Aitkin,M.,& Longford,N.(1986). Statistical modeling issues in school effectiveness studies. Journal of the Royal Statistical Society, Series A, 144(4),419-461.

Burstein,L.(1980). The analysis of multi-level data in educational research and evaluation. Review of Research in Education,8,158-233.

Cooley,W.W.,Bond,L.,& Mao,B.(1981). Analyzing multi-level data. In R.A.Berk(Ed.),Educational evaluation methodology(pp.64-83). Baltimore, ML:Johns Hopkins University Press.

Cronbach,L.J.(1976). Research on classrooms and schools:Formulation of questions, design and analysis. Occasional paper of the Stanford Evaluation Consortium, School of Education, Stanford University.

Cronbach,L.J., & Snow,R.E.(1977). Aptitudes and instructional methods. New York: Irvington.

Cronbach,L.J.,& Webb,N.(1975).Between and within-class effects in a reported aptitude-by-treatment interaction: Reanalysis of a study by G.L. Anderson. Journal of Educational Psychology,6, 717-724.

Cunningham,E.P., & Henderson,C.R.(1968). An iterative procedure for estimating fixed effects and variance components in mixed model situations. Bioketrics,24,13-25.

DeLeeuw,J., & Kreft,I.(1986). Random coefficient models for multilevel analysis. Journal of Educational Statistics, 11(1), 57-85.

Dempster,A.P., Laird,N.M., & Rubin,D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Soceity (Series B),39, 1-18.

Dempster,A.P.,Rubin,D.B.,& Tsutakawa,R.D.(1981). Estimation in covariance components models,Journal of the American Statistical Association,76,341-53.

Efron,B., & Morris,C.(1975). Data Analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association, 74, 311-319.

Gamoran, A.(1991). Schooling and Achievement: Additive versus Interactive Models. In Raudenbush,W.S., & Willms,J.D.(Ed.) School, Classrooms, and Pupils: International Studies of Schooling from a Multilevel Perspective.

San Diego, CA: Academic Press, Inc.

Gauss version 2.0 system and Graphics Manual (1984-1988),Aptech systems, Inc..

Glass,G.V.,& Stanley,J.C.(1970).Statistical methods in education and psychology, Englewood Cliffs, N.J.:Prentice-Hall

Goldstein,H.(1986). Multilevel mi ed linear model analysis using iterative generalized least squares. Biometrika,73(1), 43-56.

Goldstein,H.(1987[a]). Multilevel Models in Educational and Social Research. London:Charles Griffin & Company Limited.

Goldstein,H.(1987[b]). Multilevel Covariance Component Models. Biometrika. 74, 2, 430-431.

Graybill,F.A.(1961). An Introduction to Linear Statistical Models, V.1. N.Y.:McGraw-Hill Book Company,Inc.

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

Harville,D.A.(1969). Quadratic unbiased estimation of variance compinents for the one-way classification, 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. 358, 320-340.

Henderson,C.R.(1953). Estimation of Variance and Covariance Components, Biometrics, 9, 226-252.

Hopkins,K.D. (1982). The unit of analysis:Group means versus individual observations, American Educational Research Journal, 19(1), 5-18.

Kang, S.J., and Raudenbush,S.W.(1988). An analytical Comparison between Classical ANOVA and Hierarchical Linear Models in the Analysis of Nested Unbalanced Experimental Data. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, April.

Kirk,R.E.(1982). Experimental Design: Procedures for the behavioral Sciences (2nd ed.) Belmont, CA: Brooks/Cole.

Knapp,T.R.(1977). The unit of analysis problemin applications of simple correlational research.Journal of Educational Statistics,2(3),171-186.

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

LaMotte,L.R.(1970). A class of estimators of variance components, Technical report no.10, Department of Statistics, University of Kentucky, Lexington, Kentucky.

LaMotte,L.R.(1973). Quadratic estimation of variance components,Biometrics,29,311-330.

Lindley,D.V.,& Smith,A.F.M. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society.(series B).34, 1-41.

Longford,N.T.(1985). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. Unpublished manuscript, Institute for applied statistics. Lancaster University, Lancaster, England.

Mason,W.M.,Wong,G.Y.,& Entwisle,B.(1984). Contextual analysis through the multilevel linear model. In S.Leinhardt (Ed.),Sociological methodology 1983-1984(pp.72-103). San Franc sco:Jossey-Bass.

Page,E.B.(1975). Recapturing the richness within the classroom. Paper presented at the annual meeting of AERA, Chicago.

Patterson,H.D.,& Thompson,R.(1971). Recovery of inter block information when block sizes are unequal,Biometrika,58,545-554.

Patterson,H.D.,& Thompson,R.(1974). Maximum likelihood estimation of components of variance, Proceedings of the 8th International Biometric Conference, 197-207.

Rao,C.R.(1972). Estimation of variance and covariance components in linear models. Journal of the American Statistical Association.67. 112-15.

Raudenbush, S.W. (1984). Application of a Hierarchical Linear Model in Educational Research. Unpublished dissertation, Cambridge,MA: Harvard University.

Raudenbush, S.W. (1988). Educational Applications of Hierarchical Linear Models: A Review. Journal of Educational Statistics,13(2),85-116.

Raudenbush,S.W. (1989). The analysis of longitudinal, multilevel data, International Journal of Educational Research, 13, 721-740.

Raudenbush,S.W., & Bryk,A.S.(1986). A hierarchical model for studying school effects, Sociology of Education,59, 1-17.

Raudenbush,S.W.,& Bryk,A.S.(1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. Review of Research in Education,15,423-475.

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

Rogosa,D.(1978). Politics, process, and pyramids. Journal of Educational Statistics, 3(1), 79-86.

Rudman,H.,& Raudenbush,S.W. (1987). The effect of exceeding prescribed time limits in the administration of a Stanford Test of Reading

Comprehension and Mathematics Application. paper presented at NCME., Washington,D.C., April.

SAS Institute Inc. SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc., 1985.

Searle,S.R.(1982).Matrix algebra useful for statistics.New York:John Wiley & Sons,Inc.

Searle,S.R., & Henderson,C.R.(1961). Computing procedures for estimating compineents of variance in the two-way classification mixed model. Biometrics,17. 607-16.

Smith,A.F.M. (1973). A general Bayesian linear model. Journal of the Royal Statistical Society, (series B),35, 61-75.

Strenio,J.L.F.(1981). Empirical estimation for a hierarchical linear model. Unpublished dissertation, Cambridge, MA: Harvard University.

Strenio,J.L.F., Weisberg,H.I.,& Bryk,A.S.(1983). Empirical Bayes estimation of individual growth curve parameters ande their relationship to covariates. Biometrics,39. 71-86.

Walsh,J.D.(1947). Concerning the effect of the interclass correlation on certain significance tests. Annals of Mathematical Statistics. 18. 88-96.