This is to certify that the

dissertation entitled

Two Level Nested Hierarchical Linear Model with Random Intercepts via the Bootstrap

presented by

Joshua Gisemba Bagaka's

has been accepted towards fulfillment
of the requirements for

Ph. D degree in Counseling, Educational Psychology & Special Education (Statistics & Research Design)

Major professor

Date March 13, 1992

DATE DUE	DATE DUE	DATE DUE
JUN 12		
MAR 11 1		

TWO LEVEL NESTED HIERARCHICAL LINEAR MODEL WITH RANDOM INTERCEPTS VIA THE BOOTSTRAP

by

Joshua Gisemba Bagaka's

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

TWO LEVEL NESTED HIERARCHICAL LINEAR MODEL WITH RANDOM INTERCEPTS VIA THE BOOTSTRAP

by

Joshua Gisemba Bagaka's

In statistical linear models, most procedures available for estimating the variance components of the mixed model are usually based on the assumption that the error terms and each set of random effects in the model are normally distributed with zero means and some variance–covariance structure. However, in certain research situations, there is little doubt that the error terms and each set of random effects in the mixed model can be characterized as moderately or even distinctly non–normal with heavy tails or badly skewed distributions.

Efron (1979) discussed the use of a technique called the bootstrap to generate sampling distributions of statistics and thereby to draw inferences about parameters without requiring any distributional properties. Besides the fact that the bootstrap liberates statisticians from over–reliance on distributional assumptions, the method makes it possible to attack more complicated problems which may not have closed–form expressions.

This study utilized the bootstrap procedure to estimate the sampling distribution of estimators, their standard errors and thereby setting confidence intervals about the parameters of a mixed HLM under a variety of conditions.

Applicability of the bootstrap on data originating from real research situations was demonstrated through the estimation of the effects of school, classroom, and teacher variables on the teachers' self–efficacy.

Based on the usual MINQUE and bootstrap estimators, the study showed that the success of estimation is typically affected by the nature and size of the tails of the distribution of the errors and sets of random effects parameters of the model. The bootstrap generally followed MINQUE quite closely in estimating the fixed and random effects of the model under both the normal and double exponential distributions. Particularly in estimating the population inter–class variance, τ^2 at the 0.01 level of the intra–class correlation, the bootstrap was surprisingly closer to the parameter value than the MINQUE.

Due to the fact that the bootstrap procedure is highly dependent on the computer, the study recommended that software to implement the bootstrap algorithm be developed to make the method available to research practitioners. Availability of the method to research practitioners will provide an important and flexible tool, typically unavailable through classical methods, of estimating the sampling distributions of statistics, their standard errors, and thereby setting confidence intervals about parameters.

This dissertation is dedicated

TO

my parents

Ludiah and Andaraniko Bagaka

and to

my sister Milcah Bosibori who has

been fighting hard for her

life during the period

of this dissertation.

ACKNOWLEDGEMENT

This dissertation would not have been realized without the direct and indirect assistance and encouragement of many people. Very deep appreciation goes to Dr. Stephen Raudenbush, my advisor, friend and chair of my committee for his support, direction, and encouragement throughout my doctoral studies. Very sincere gratitude goes to Dr. James Stapleton, my masters program advisor and member of my doctoral committee, whose special friendship, encouragement, endless patience, and advise fostered the congenial atmosphere in which my entire graduate studies were undertaken. Very grateful acknowledgement also goes to the other members of my committee, Dr. Betsy Becker and Dr. James Gaveleck, for their support, prudent guidance, and encouragement. The writer also wishes to extend a special acknowledgement to Cathy Sparks for her tireless service of typing this dissertation.

I wish to express my deep gratitude to the members of my family: To my children Cliff Onserio, Kathrine Kerubo, Edward Makoyo, and my wife Hellen Moraa, who sacrificed their time and enjoyment; to my parents, Ludiah and Andaraniko Bagaka for the manner they raised me, their belief in me, and their endless patience, love and support; to my parent in–laws, Bathseba and Sospeter Mariera for their continuous love; to my brothers and sisters, Musa Nyakeri, Jeliah Mongina, Yunuke Nyamunsi, Milcah Bosibori, Sarah Makori, Anna Kwamboka, Dorika Tabitha, Isabela Kemunto, and Henry Nyabuto for their special love and encouragement; to all my brother– and sister–in–laws; and to the entire extended Bagaka clan and family.

Very special acknowledgment goes to my family friends, Ogega and Phoebe Mokogi Omete and their children; my nephew and friend, Thomas Ogega Orenge; Linda and Richard Solomon and their children; my sincere friends, David Wafula Makanda, the Mwakikotis, the Saginis, Zablon Oonge, Roselyne and Zipporah Chisnell, Samuel Muga, the Mapatis, Annie Woo, Zora Ziazi, the Navarros, Vicky Lutherford, Sohed, Kamuyu–wa–Kangethe, John Lelei, and the entire Kenyan family in the greater Lansing Area for their family–like support and friendship.

I also wish to extend a special acknowledge to my cousin Charles Nyakeri; my nephew and role model, Dr. Benson Mochoge; and all members of the Greater Gionseri community for their encouragement and support. To Chris Obure, Nyaundi, Oigara, Paul Sikini, my brother in–law Peter Omwoyo, the 1983 NHS and NASS staff and all friends and relatives who rendered their encouragement as well as moral and financial support. Their support is deeply appreciated by the writer, his family and the people of Kenya. May God bless everyone. ASHANTE SANA.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Estimation is frequently based on subpopulations which can be combined collectively into one underlying population. For instance, educational performance can be examined through a sample from several schools in a nation or state. It is quite natural to estimate the mean achievement and the spread of students' achievement in each school. Yet groups such as the school district personnel may be interested in knowing the achievement of students in their school district relative to the national average, while the school principal may be interested in the performance of the school relative to the statewide or national average performance. Several other interest groups (parents, teachers, education ministers) may have interest in different "levels" of data, making it necessary to examine the data in stages.

Recently methodologists (Aitking and Longford, 1986; Burstein, Linn, & Campell, 1978; Burstein & Miller, 1980; Goldstein, 1986; Raudenbush & Bryk, 1986) have developed techniques to address studies involving data which have a hierarchical character. Most of these studies have been conducted under the assumption that the observations are independent and normally distributed. Mason, Wong, and Entwistle (1983) and Raudenbush (1984) formulated similar mathematical models for hierarchical data within–macro units with, say, students in a school as "micro units" of analysis and schools as the "macro units" of analysis. The resulting within– and between–macro units models were based on the usual independence and normality assumptions.

Since the manner of obtaining data typically affects inferences that can be drawn from such models, we consider a sampling process in which the "macro" units are randomly drawn from a population before a random sample of "micro" units are drawn from each "macro" unit. The resulting data are thus associated with two random components (the between and within macro variance components) and the model is correspondingly called the random effects model. Many situations arise where "macro" units are nested within some fixed factors (not drawn randomly from a population), together with other micro fixed effects and covariate(s), resulting in a mixed model with both fixed and random effects. Analysis of variance is traditionally employed in situations involving fixed effects models, to estimate the fixed effects parameters.

Although statistical procedures are available for estimating the variance components of the random parts of the mixed model, these procedures have several limitations, which take one or more of the following forms. First is the problem of unbalanced designs (unequal subclass numbers). Estimating variance components from unbalanced data is not as straightforward as from balanced data (Searle, 1971). Secondly, estimating variance components often involves relatively cumbersome algebra which makes it difficult for most methods to estimate model parameters when covariate(s) are involved as part of the fixed factors. Third is the problem of negative estimates of the variance components and last but not least, the problem that most variance component estimation procedures are based on the assumption that the random error terms and sets of random effects are normally distributed.

The limitation of unbalancedness is certainly clear since balanced designs are rare in research situations. Thus, procedures of estimating variance components limited to balanced designs may not be at all useful. On the situation of cumbersome algebra and negative estimates, no one method has yet been clearly

established as superior either in minimizing the amount of computation required to estimate the variance components or in obtaining non–negative estimates of the variance components. These are the situations in which attempts can be made to minimize but not necessarily to overcome the problem.

Most procedures available for estimating the variance components of the mixed model are based on the assumption that the error terms and each set of random effects in the model are normally distributed with zero means and some variance—covariance structure. Then for the balanced random component model, it can be shown that the sum of squares in the analysis of variance are distributed independently of each other; and each sums of squares divided by the expected values of its mean square has a central chi–square distribution with the corresponding degrees of freedom (Searle, 1971). This holds true only for the random component model. For the mixed model, this distributional property only holds for those sums of squares whose expected values are not functions of fixed effects; otherwise the same ratio of sums of squares that do involve fixed effects, will have a non–central chi–square distribution. Thus, the normality assumption for the error term and each set of random effects in the model is the basis of the distributional properties of variance component estimators, on which most variance component estimation procedures are based.

However experience has shown that in certain research situations, there is little doubt that the error terms and each set of random effects in the mixed model can be characterized as moderately or even distinctly non–normal. For example, educationally oriented variables such as number of days absent from school, number of times a student answers a question (or talks in class) and many other variables are likely to produce non–normal data that are heavy tailed or badly skewed. Thus

the results of statistical methods based on the Gaussian assumptions may not always be reliable.

Approaches are available for dealing with non–normality. Most involve transforming the original data to a form more closely resembling a normal distribution such that normal theory methods can be applied (Box and Cox, 1964). Efron (1982) examined a family of six transformations and cautioned against uncritical use of normality as a criterion for successful transformation. Perhaps variance stabilizing transformations may be preferable. Efron (1982) discusses situations in which it is better to transform to homoscedasity and ignore non–normality than vice–versa. Otherwise the alternative could be to do a complete analysis to recover the lost information during transformation. However, the practical motivation of transformation theory is to avoid complicated analysis, especially in already complicated situations (Efron, 1982).

What complicates the issue of using transformations even more is the fact that the underlying distribution of the original variable must be known before one decides on the appropriate transformation. In many situations, the underlying distribution of the original data may not be known and thus appropriate transformation of the data becomes difficult.

Rao (1971) proposed the Minimum Norm Quadratic Unbiased Estimator (MINQUE) for variance components, which does not require the normal distribution properties of the error term and each of the sets of random effects. The method is quite general, applicable most experimental situations, and the computations are relatively simple (Rao, 1971). The approach of the MINQUE involves estimating a linear function of the variance components using a quadratic function of the observations, using pre–assigned weights in the norm. The MINQUE estimates therefore may vary with the choice of weights.

In addressing the problem of dependency on the weights when using MINQUE, Brown (1976) suggested a procedure in which, after calculating a MINQUE estimate as usual, the values therein are used as weights and the MINQUE is calculated again. The process is repeated iteratively until two successive estimates are equal, to some degree of approximation. The method has been named iterative MINQUE or I–MINQUE (Brown, 1976) and it has been shown that MINQUE and I–MINQUE estimators are asymptotically normal. However, because the I–MINQUE estimators are obtained iteratively, they do not have the properties used in deriving MINQUE (unbiasedness, translation invariant and minimum norm), and thus they are not necessarily unbiased or "best" in any sense (Searle, 1979).

This study adopted the MINQUE procedure as a useful method of estimating the variance component since the procedure does not require the normal distributional properties. In addition, perhaps one of the most useful features of the study was in the specific manner in which MINQUE was implemented. The study used a crude ANOVA–type estimate of the variance components of the mixed model as in Hanushek (1974). The values from this prior estimator are used to determine the weights which are then used in the computation of the MINQUE estimators. However, this does not in any way constitute the focus of the present dissertation. The primary focus stands to be an attempt to liberate statisticians from over–reliance on the normal assumptions in estimating the variance components of a mixed model.

Efron (1979) has discussed the use of a technique called the bootstrap to generate sampling distributions of statistics and thereby to draw inferences about parameters without requiring any distributional properties. Although Efron avoids making any general claim for the origin of the name "bootstrap," Efron's examples may suggest to some that it is indeed a technique of "pulling ourselves up by our

bootstraps" in a data analysis, that is, for obtaining inferences insensitive to model assumptions (Rubin, 1981). Indeed, the name reflects the fact that one available sample gives rise to many others (Diaconis & Efron, 1983).

The development of the bootstrap starts with a sample $X = \{X_1,...,X_n\}$ of n observations. From this sample, a random sample of size n is drawn with replacement from which a first bootstrap estimate is calculated. The replicated sample is denoted by $X^* = \{X_1^*,...,X_n^*\}$ and the bootstrap replicated estimate computed from X^* is denoted by $\hat{\theta}^*$. The process is repeated a large number B times resulting in a sequence $\hat{\theta}_b^*$ of estimates for $b = 1,...,B$. If F designates the unknown distribution of X, then Efron (1979,1982) argues that the empirical bootstrap distribution F^* of X^* can provide a very good approximation of F for a wide variety of interesting statistics. The bootstrap, therefore, which is an elaboration of the jackknife invented by Quenouille (1949), provides a general method which can be applied to complicated situations where theoretical analysis is not possible. Under quite general conditions, the bootstrap gives asymptotically consistent results and for some simple problems which can be analyzed completely, for example, ordinary linear regression, it automatically produces results which are comparable to standard solutions (Efron, 1981b). Through a series of examples, Efron has shown that the bootstrap method works reasonably well under a variety of situations. A more detailed discussion of the bootstrap method is offered in Chapter III.

Dependence on the Normal Assumptions

The distributional assumptions imputed to the random error terms and each set of random effects in the mixed model are that they are independent and normally distributed with mean $\underset{\sim}{0}$ and some variance—covariance structure. But

in order to realize the increased flexibility of hierarchical linear models, careful attention needs to be paid to these statistical assumptions (Bryk and Raudenbush, 1987). Though methods are available to assess the degree to which these assumptions are realized in research situations, many researchers proceed with computations under the normal assumptions regardless of whether or not the normality condition is met. However, there are several situations in educational research where hierarchical models may be applicable, but the normal assumptions may not be guaranteed. For example, in the model involving student achievement scores, or number of days absent there is doubt that the error terms are normally distributed in certain situations.

The most common macro unit of analysis for the between group hierarchical model in education is the school. Often, a random sample of schools is drawn from which a sample of students is also drawn at random. Schools with different characteristics may be in the sample resulting in an hierarchical data set with certain response variables with different distributions for each school. Certain educationally oriented variables, either at the student or school or classroom level may be observed. Yet as mentioned earlier, for some of the variables, under the assumption of random sampling from these subpopulations, there may be doubt about the normality of the population distribution. Some schools may have data sets whose underlying distribution is negatively skewed, others positively skewed, others heavy–tailed and others may even be normally distributed. Under these conditions using the standard methods to calculate parameter estimates may not provide better estimates. Attempts to transform data to a form more closely resembling a normal distribution will involve identifying the underlying original distributions for variables in each context (e.g., school) first before deciding on the most appropriate transformation strategies for each subpopulation. Even if the

underlying distributions of the subpopulations were known, transforming variables differently for each "macro" unit may deteriorate into a welter of calculations. In such a situation, therefore, the bootstrap algorithm becomes handy and appropriate not only to determine the expected values of the estimates without worrying about the distributional properties but also to estimate the standard errors of the estimates and the empirical distributions of the estimators, thereby setting confidence intervals about the parameters.

Negative Variance Estimates

The usefulness of variance component techniques is frequently limited by the occurrence of negative estimates of essentially positive parameters (Thompson, 1962). Though methods like the Restricted Maximum Likelihood (REML) were primarily designed to remove this objectionable characteristic for certain experimental designs, the problem still remains unsolved. Thompson (1962) described an algorithm for solving the problem of negative estimates of variance components for all random effects models by considering that their expected mean square column forms a mathematical tree in a certain sense. The algorithm was described as follows:

> "Consider the maximum mean square in the entire array; if this mean square is the root of the tree then equate it to its expectation. If the minimum mean square is not the root then pool it with its predecessor." Thompson, 1962, p. 273.

In either case the problem is reduced to an identical one having estimates of the variance components. The estimates are non–negative and have a maximum likelihood property.

Other methods like the method of moments have ways of controlling for the occurrence of negative estimates by simply equating any negative estimate to zero. It is anticipated that the bootstrap method used in this study will provide another useful way of controlling for non–negativity of estimates of the variance components particularly when the population interclass variance is small but positive.

For the bootstrap method, the estimate $\hat{\theta}^*_b$ of the variance component is computed at each replication b for b = 1,2,...,B where B is a large number. The expected value of the estimate is then the average over all B replicated values of the estimator. It is anticipated that if the parameter value of the variance component is non–negative, then the sum and average over all the replicated values will be non–negative. In this case therefore, we view the bootstrap as a means of providing the MINQUE method with B opportunities to prove the positiveness of the estimate of the parameter value which is essentially positive.

Purpose of the Study

The interest of the study lies in a two–level mixed and nested hierarchical linear model (HLM) with random intercepts. The general problem is one of estimating the fixed effects and the variance components (group and individual level variances) under several situations including conditions under which the normality assumption may not be guaranteed. Besides the non–normality problem, the problem of negative estimates of the between "macro" unit variance component (especially in the case of boundary situations when the true variance component is small) is not new to statisticians.

In the present study two different estimators of the model parameters were obtained and compared against each other. The first estimator was the MINQUE based on the original sample. The other was the bootstrap estimate computed though resampling from a sample with replacement.

Objectives of the Study

The study demonstrates the use of the bootstrap in providing estimates of the parameters (fixed and random) of a general two–level mixed and nested hierarchical linear model, determining the standard errors of the estimators and their empirical bootstrap distributions. In addition to demonstrating that the bootstrap algorithm liberates statisticians from over–reliance on the Gaussian assumptions (Diaconis and Efron, 1983), through Monte Carlo simulations, this study also

(1) Determines the bootstrap standard errors of the variance components and thereby allow construction of bootstrap confidence intervals about the parameters.

(2) Assesses the performance of the bootstrap method in determining the estimates of the sampling distributions, the standard errors, and the interval estimates of the variance component estimates of the model when the response variable,

   a) is normally distributed.

   b) has a distribution with fairly heavy tails (e.g., the double exponential distribution)

(3) Evaluate the bootstrap estimates and usual MINQUE estimates of the variance components.

(4) Examines the relative accuracy of the bootstrap method in estimating the variance components of the model in the case of boundary situations, particularly when the population interclass variance is small but positive.

(5) Determines the bootstrap estimate of the fixed effects parameters and their standard errors.

Summary

The present dissertation concentrates on the problem of estimating the parameters of a mixed and nested hierarchical linear model. Chapter II will describe the hierarchical model and the estimation of variance components in balanced and unbalanced designs when the models are with and without covariate(s). Chapter III will concentrate on the discussion of the bootstrap method. The design of the study will be provided in Chapter IV and an application of the bootstrap method in estimating model parameters in higher order teaching (HOT) research will be presented in Chapter V. MINQUE and bootstrap Monte Carlo simulations results will be presented in Chapter VI and conclusions and recommendations set out in Chapter VII.

CHAPTER II

THE MULTILEVEL MODEL AND ESTIMATION

Introduction

It is common in educational research to study the effect of character of the educational group (e.g., school, school district, classroom, province). These group–oriented variables (e.g., school policies, teacher/student ratio, per–pupil spending) may form part of a set of independent variables hypothesized to have an effect on some individual–level outcome variable(s). For example, student learning activities occur within organizations in which the individual students belong (Burstein, 1980). It is therefore necessary for educational researchers to understand and be able to explain the complex influence of not only individual level variables but also group–oriented variables on some individual student outcome variables.

Data in this class of research is typically available at two levels of observation, individual (or micro) and group (or macro) levels, giving rise to a hierarchical structure of data. Similar arguments can apply as well to more complicated nesting situations (students within classrooms, classrooms within schools, schools within school districts, and school districts within states or provinces) without loss of generality (Burstein, 1980). The problem then, is that of analyzing such multilevel data when certain key independent variables are measured at different levels of an organizational hierarchy.

Traditional statistical methods of data analysis like multiple regression and analysis of variance have been found to be ineffective in studies involving such data of hierarchical structure. Methodologists (Burstein, 1980; Burstein & Linn, 1978;

Mason, Wong, & Entwisle, 1984; Raudenbush & Bryk, 1986; Raudenbush, 1984) have not only warned against the use of such classical linear models but have also provided general statistical models of investigation when data exists in hierarchical structures. These models are commonly referred to as hierarchical linear models (HLM).

Studies of school effectiveness (e.g., Brookover, et.al., 1982) have been interested not only in student achievement levels as measured by the average achievement scores but also in overall group achievement or "equity" as measured by the variability (or spread) of achievement scores. From this viewpoint, more effective schools for example, not only produce high average achievement scores but also help students of varied backgrounds to achieve mastery (Raudenbush, 1984). The notion of evaluating the effectiveness of schools in achieving "equity" by observing the within–school variability of scores can also be extended to examining effectiveness of the state, province or country by observing between–school variability in student achievement scores. Coupled with the fact that the "macro" units (e.g., schools or classrooms) in the study may constitute a random sample of such units from a population, then the mixed model conceptualization is certainly appealing. Thus, one important class of investigation in such situations would involve estimation of the variance components (or equity) in addition to examining the effects of other fixed factors in the mixed hierarchical linear models.

The Multilevel Model

The structure of data considered in this multilevel framework is assumed to involve two levels of observations; the individual (micro) level and some higher (macro) level. The structure can be characterized by contexts such as schools or countries as "macro" units of analysis and individual subjects as the "micro" units of analysis. The fundamental assumption underlying this multi–level hierarchical

framework is that the micro values of the response variable depend in some way on context and that the effect of the micro determinants may vary as a function of context (Mason, et. al., 1983). At the lowest level, some measure of outcome for individual subjects and other individual characteristics may be appropriate.

Suppose we begin by posing a within–context model that defines a micro equation with one micro response variable Y and one micro regressor X, which is identical for each macro unit j as,

$$(2.1) \qquad Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \epsilon_{ij}$$

where $j = 1, 2,...,J$ macro units and $i = 1, 2,...,n_j$ micro units within the macro units. In this case Y_{ij} is the response and X_{1ij} the regressor value for subject i in macro unit j and ϵ_{ij} is the random error term. The usual assumption is that ϵ_{ij} is distributed normally with mean zero and variance σ_e^2. The micro parameters β_{0j} and β_{1j} are assumed to vary across context as a function of some macro regressor variable W.

Since β_{0j} and β_{1j} are defined for each context, we pose the between–context models using β_{0j} and β_{1j} as response variables as

$$(2.2) \qquad \beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + e_{0j}$$

$$(2.3) \qquad \beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + e_{1j}$$

where β_{0j} and β_{1j} are the intercept and regression slope respectively for context j, as defined in Equation 2.1. Both the intercept and the slope are assumed to be random with e_{0j} and e_{1j} as their respective random error terms. It is most common to assume that the error terms e_{0j} and e_{1j} are normally distributed with mean zero and variances η_{00} and η_{11}, respectively, with the covariance of e_{0j} and e_{1j} denoted by η_{01}.

A single equation for this simple case of a multilevel model can be obtained by substituting Equations 2.2 and 2.3 into Equation 2.1 as

$$(2.4) \qquad Y_{ij} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{10}X_{1ij} + \gamma_{11}W_{1j}X_{1ij} + (e_{0j} + X_{1ij}e_{1j} + \epsilon_{ij}).$$

Although Equation 1.4 involves one micro regressor and one macro regressor, it is quite general in the sense that other models of potential interest can evolve from it (Mason, et. al., 1983). For example, a random effects one way analysis of variance (ANOVA) model can be derived from it by setting to zero all of the coefficients of X_1 and W_1, i.e., $\gamma_{01} = \gamma_{10} = \gamma_{11} = e_{1j} = 0$ resulting in the model equation,

$$(2.5) \qquad Y_{ij} = \gamma_{00} + e_{0j} + \epsilon_{ij}.$$

Similarly, a fixed effects regression model is obtained by setting e_{0j} and e_{1j} to zero to obtain the model equation

$$(2.6) \qquad Y_{ij}=\gamma_{00} + \gamma_{01}W_{ij} + \gamma_{10}X_{1ij} + \gamma_{11}W_{1j}X_{1ij} + \epsilon_{ij}.$$

For this study, a hierarchical linear model with random intercepts is considered. Consequently, the random error e_{1j} associated with the random slope model given in Equation 2.3 is set to zero. Model Equation 2.4 then reduces to the variance component model given by

$$(2.7) \qquad Y_{ij} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{10}X_{1ij} + \gamma_{11}W_{1j}X_{1ij} + (e_{0j} + \epsilon_{ij})$$

which is a mixed model with the term $(e_{0j} + \epsilon_{ij})$ as the random part and $(\gamma_{00} + \gamma_{01}W_1 + \gamma_{10}X_{1ij} + \gamma_{11}W_{1j}X_{1ij})$ as the fixed part of the model. The fixed part of the model in Equation 2.7 may take a more general form with multiple X's (i.e. $X_{2ij}, X_{3ij}, \ldots$) and W's ($W_2, W_3, \ldots$) which may also include interactions.

One of the fixed factors of the model, for example, may be the sector in which the random factor of context is nested. The fixed effects factor (sector) is taken to consist of K levels, bringing the total number of fixed effects parameters (including covariates) to P.

In terms of the general linear model matrix notation, and if we allow for any number L of regressor variables, Equation 2.7 can be expressed for the jth context as

$$(2.8) \qquad \underset{\sim}{Y}_j = \underset{\sim}{X}_j\underset{\sim}{\alpha}_j + \underset{\sim}{Z}_j\underset{\sim}{b}_j + \underset{\sim}{\epsilon}_j$$

$j = 1, 2,...J.$ where $\underset{\sim}{Y}_j$ is a $(n_j \times 1)$ vector of response values; $\underset{\sim}{X}_j$ is a $(n_j \times p)$ matrix of known constants; $\underset{\sim}{\alpha}_j$ is a $(p \times 1)$ vector of unknown fixed effects parameters; $\underset{\sim}{Z}_j$ is $(n_j \times 1)$ vector of 1's; $\underset{\sim}{b}_j$ is $(q \times 1)$ vector of unknown random effects parameters and $\underset{\sim}{\epsilon}_j$ is an $(n_j \times 1)$ vector of random error terms.

The Two–level Hierarchical Linear Model with Random Intercepts

The model illustrated thus far reduces to simpler models under specific conditions. The present study involves two factor levels, a fixed and a random factor where the random effects are nested within fixed effects. Application of this model can be seen in education research with the school background or sector (public, private or religious) as the fixed factor and individual schools as the random factor. At the lowest level, some measure of outcome, for example academic achievement may be of interest. Other school characteristics (teacher–student ratio, school financial means, school policy, inner city or suburban location) together with student characteristics (social economic status, IQ) may be included in the model as covariates. An example of the use of "Micro" and "Macro" variables can be found in Mason et. al. (1983) who used a model for a multilevel analysis of the determinants of children born in fifteen less–developed countries. In this study which was part of the Michigan Comparative Fertility project, Mason et. al. (1983) used countries as the macro units of analysis, while married respondents served as the micro units of analysis. Some of the macro variables used to define the context within which individual childbearing took place included socioeconomic

development, family planning program effort, and per capita gross national product. The micro specifications used included contraceptive use patterns, and the wife's education level.

Model Assumptions

Two levels of distributional assumptions can be specified for the multilevel hierarchical linear model described above. First is the assumption related to the micro specification model shown in Equation 2.1. For this model, the error terms, ϵ_{ij} are assumed to be independently and normally distributed with mean vector $\underset{\sim}{0}$ and variance $\sigma_j^2 I_{n_j}$ for $j = 1,...,J$. With Equations 2.2 which describe the random intercept part of the model, the following assumptions are made:

(i) the error terms e_{0j} associated with the intercept β_{0j} are assumed to be distributed normally with mean 0 and some variance η_{00}.

(ii) the micro errors, ϵ_{ij} are independent of the macro errors e_{0j}.

While attainability of the distributional assumptions related to the micro error terms ϵ_{ij} can be easily accessed by observing the distribution of the response values Y_{ij}, accessing the distributional assumptions of the macro error terms e_{0j} is more difficult since β_{0j} is not directly observable. This worsens the situation in dealing with methods which are overly dependent on distributional assumptions.

The assumption of independence in multilevel models also takes two forms, within– and between–group independence. Robustness to within–group dependence (dependence among observations) is of primary interest in the statistical literature (Burstein, 1980). The statistical consequences of ignoring the intraclass correlation structure that results by ignoring group membership can be quite serious (Burstein, 1980). In educational research involving student achievement, we realize that instruction is primarily group–based. Instruction of students within the same class

is likely to be more similar than that for students from different classes. Under these and similar circumstances, the between– and within–group error terms are likely to be correlated.

In general, standard statistical estimation techniques like ordinary least squares are ineffective in the presence of within–group dependencies. Yet in several educational research situations, depending on the nature of the outcome and effect variables under study, dependence among observations may be an inevitable phenomena. Thus it may be reasonable for researchers to assume that dependencies among observations exist, such that more effort may be spent on ways to identify and adjust for these dependencies rather than assuming independence when dependence may exist.

Variance Components Estimation

The problem of estimating the variance components in mixed linear models, containing both fixed and random effects is not new to research methodologists. Several methods of estimation have been suggested (Henderson, 1953; Hartley, 1967; Searle, 1970; Henderson, et.al., 1959; Rao, 1970; 1971a, 1971b, 1972; Thompson, 1962). The deficiencies and/or difficulties in the application of these methods are also well known (Searle, 1978). Estimates could be negative. Compuatational problems could arise, particularly when covariates are involved as part of fixed factors of the mixed model. There is no general method to cover all situations and problems. The problem of variance component estimation also varies with the design, whether the data is balanced (equal subclass numbers) or unbalanced.

Balanced Designs

Balanced designs are those in which there are equal numbers of observations in all the macro units. The analysis of variance method (or the method of moments) is traditionally employed in estimating the variance components of mixed

(or random) balanced designs. The method involves equating statistics to their expected values and solving the resulting equations for the parameters (Hocking, 1985). But due to the infrequency of balanced designs in real world research, methods which are limited to balanced designs are not at all useful.

Unbalanced Designs

Unbalanced designs are to those in which the number of observations in the sub–classes or macro groups are not all the same. Besides the problem of cumbersome algebra and a confusion of symbols in variance components estimation in unbalanced designs, other problems arise. Whereas with balanced data there is only one set of quadratic forms to use (the analysis of variance mean squares), there are many sets of quadratic forms that can be used for unbalanced data. And unlike in balanced data, most quadratic forms in unbalanced designs lead to estimates that have few optimal properties. As Searle (1971) indicated, none of the earlier methods were clearly established as superior in variance component estimation.

Efforts to adapt variance component estimation methods to unbalanced data were led by Henderson (1953) who described an analogous to the analysis of variance method used with balanced data, but designed to correct that deficiency. Other methods evolved thereafter (see Searle, 1968), but Searle (1971) indicated that most of these methods reduced in some way to the method of moments for balanced data. The methods involve relatively cumbersome algebra such that a discussion of unbalanced data easily deteriorate into a welter of symbols.

Other more recent methods of variance components estimation evolved which are not necessarily allergic to unbalancedness. The maximum likelihood (ML) estimator of the variance components is one such method. The ML estimators of the variance components are those values of the components which maximize the likelihood over the positive space of the variance components parameters (Corbeil and Searle, 1976). Application of the ML method therefore requires assuming a

probability density function for the random variables, and then writing down the likelihood function of the sample data. Though the general ML procedure can be used for almost any probability density function, for variance component estimation, it is customary to assume normality (Searle, 1979). Then maximizing the logarithm of the likelihood function is fairly straightforward. However, as indicated earlier, requiring the normality assumptions in certain research situations may be expecting too much.

An alternative to the ML estimator of the variance components is the restricted maximum likelihood (REML) which was first suggested by Thompson (1962) and later formally described by Patterson and Thompson (1971) and Corbeil and Searle (1976). The method is based on a transformation that partitions the likelihood under normality into two parts, one being free of the fixed effects and the other involving fixed effects. Maximizing the part that is free of fixed effects yield the REML estimators (Corbeil and Searle, 1976). The REML estimators are translation invariant, but because maximum likelihood restricts the estimator to the allowable parameter space (positive), then REML estimators are biased. Thus, in terms of assumption requirements, neither ML nor REML offers no solutions to the estimation of variance components without assuming certain distributional properties.

From the late 1960's till the early 1970's, statisticians were involved in seeking methods of variance components estimation that posess more desirable properties than just being unbiased and translation invariant. LaMette (1973) and Rao (1970) though working independently, derived the minimum variance quadratic unbiased estimators (MIVQUE) from the theoretical viewpoints without offering ways of applying the method to actual data analysis. Henderson (1973) derived computational formulae for MIVQUE based on the mixed model equations (MME) and indeed showed that LaMotte's and Rao's methods were identical.

MIVQUE assumes normality and that V, the variance–covariance matrix of the observations is known. Then the variance of quadratic forms is minimized for this V. Since V is not known in reality, the procedure requires utilizing some prior information about V, and as a result, the variance of quadratic forms is only minimized if this prior V is the true population value. However, MIVQUE is unbiased and translation invariant. In addition, if the prior V is the same as the true V, then MIVQUE is also minimum variance. Thus, MIVQUE in general, is not minimum variance in practice, but only as good as the prior V. For the present study, the MIVQUE too does not offer solutions to variance component estimation since the procedure requires the normality assumption.

Rao (1970, 1971a,b, 1972) proposed a minimum norm, quadratic, unbiased estimator (MINQUE) of the variance component to estimate a linear function $\underset{\sim}{P}'\underset{\sim}{\sigma}$ of the variance component (for known $\underset{\sim}{P}'$). The method utilizes a quadratic function $\underset{\sim}{Y}'\underset{\sim}{A}\underset{\sim}{Y}$ of the observations for $\underset{\sim}{Y}$, a vector of observations and $\underset{\sim}{A}$, a symmetric matrix. The quadratic function $\underset{\sim}{Y}'\underset{\sim}{A}\underset{\sim}{Y}$ used to estimate $\underset{\sim}{P}'\underset{\sim}{\sigma}$ is taken to possess the properties of translation invariance, unbiasedness and minimum norm (Searle, 1979). More importantly, the MINQUE theory is developed without reference to normality or the variance of the estimator and the method is highly flexible in the choice of norm while at the same time preserving the desirable properties of the estimator (Rao, 1971). Since their invention these estimators have gained much recognition. See in particular, Seely (1971), Hartley et. al. (1978), and Searle (1979). Also the naive form of the MINQUE which corresponds to the rather uninformative prior value $\underset{\sim}{V} = w_0\underset{\sim}{I}_n$ (MINQUE0) is provided by Statistical Analysis Systems (SAS). Due to its desirable properties, particularly the fact that the procedure does not require the normal assumptions, the present study adopted

the MINQUE technique in estimating the parameters of the mixed model via the bootstrap method.

MINQUE for Two–level HLM with Random Intercepts

The minimum norm quadratic unbiased estimator (MINQUE) for the variance components of a mixed model is based on the statistical linear model whose general form is represented by

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}\underset{\sim}{b} + \underset{\sim}{\epsilon} \tag{2.9}$$

with the following definitions:

$\underset{\sim}{Y}$ is a $(n\times 1)$ vector of n observations

$\underset{\sim}{X}$ is an $(n\times P)$ known matrix of rank $r(\underset{\sim}{X}) < n$

$\underset{\sim}{\alpha}$ is a $(P\times 1)$ vector of P fixed effects parameters

$\underset{\sim}{Z}$ is an $(n\times J)$ known matrix, often consisting of 1's and 0's

$\underset{\sim}{b}$ is a $(J\times 1)$ vector of J unobservable random effects parameters, and

$\underset{\sim}{\epsilon}$ is a $(n\times 1)$ vector of random error terms.

In order to identify the variance components corresponding to the random effects in $\underset{\sim}{b}$, this vector $\underset{\sim}{b}$ is partitioned as

$$\underset{\sim}{b}' = [\underset{\sim}{b}_1'\ldots\underset{\sim}{b}_k\ldots\underset{\sim}{b}_c'] \text{ for } k = 1,\ldots,c, \tag{2.10}$$

where the vector $\underset{\sim}{b}_k$ contains j_k effects for the levels of the k^{th} random factor. Corresponding to $\underset{\sim}{b}_k$ of 2.10 the incidence matrix $\underset{\sim}{Z}$ is accordingly partitioned as

$$\underset{\sim}{Z} = [\underset{\sim}{Z}_1\ldots\underset{\sim}{Z}_k\ldots\underset{\sim}{Z}_c] \text{ for } k = 1,\ldots,c, \tag{2.11}$$

such that 2.9 can be written as,

$$(2.12) \qquad \underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \sum_{k=1}^{c} \underset{\sim}{Z}_k \underset{\sim}{b}_k + \underset{\sim}{\epsilon}$$

with the model elements defined as before. Equation 2.10 and 2.11 are similar to 5.3 in Rao (1971b).

A compact way of writing (2.12) is to define $\underset{\sim}{\epsilon}$ as another $\underset{\sim}{b}_k$ namely, $\underset{\sim}{b}_0$ and the corresponding $\underset{\sim}{Z}_0$ as I_n. The model Equation 2.12 becomes

$$(2.13) \qquad \underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \sum_{k=0}^{c} \underset{\sim}{Z}_k \underset{\sim}{b}_k$$

with the following distribution properties:

$$(2.14) \qquad E(\underset{\sim}{b}_k) = \underset{\sim}{0},\ var(\underset{\sim}{b}_k) = \sigma_k^2 I_{jk},\ cov(\underset{\sim}{b}_k, \underset{\sim}{b}_{k'}) = \underset{\sim}{0}$$

for $k,k' = 0,1,\ldots,c$

where $cov(\underset{\sim}{b}_k, \underset{\sim}{b}_{k'})$ is the matrix of covariance of the elements of $\underset{\sim}{b}_k$ with those of $\underset{\sim}{b}_{k'}$ for $k \neq k'$. The variance of $\underset{\sim}{b}$ is given by

$$(2.15) \qquad Var(\underset{\sim}{b}) = D = diag\{\sigma_k^2 I_{jk}\} \text{ for } k = 0,\ldots,c.$$

With this formulation, we notice that Equation 2.7 is a special case of 2.13 with $c = 1$ whose compact form may be given by

$$(2.16) \qquad \underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}_0\underset{\sim}{b}_0 + \underset{\sim}{Z}_1\underset{\sim}{b}_1$$

where

$\underset{\sim}{Y}$ is a $(n \times 1)$ vector of n observations,

$\underset{\sim}{X}$ is an $(n \times P)$ known matrix of rank $r(x) < n$

representing fixed effects parameters,

$\underset{\sim}{\alpha}$ is a $(P \times 1)$ vector of P fixed effects parameters,

$\underset{\sim}{Z}_0$ is an $(n \times n)$ identity matrix,

$\underset{\sim}{b}_0$ is a $(n \times 1)$ vector of residual error terms,

$\underset{\sim}{Z}_1$ is a $(n \times J)$ known matrix, often consisting of 1's and 0's and

$\underset{\sim}{b}$ is a $(J \times 1)$ vector of J unobservable random effects

parameters.

The distributional properties imputed to 2.16 are according to 2.14 given by

(2.17) $\quad E(\underset{\sim}{b}_0\underset{\sim}{b}_1) = \underset{\sim}{0},\ \mathrm{cov}(\underset{\sim}{b}_0,\underset{\sim}{b}_1) = \underset{\sim}{0}$

(2.18) $\quad \underset{\sim}{D} = \mathrm{var}(\underset{\sim}{b}_k) = \mathrm{diag}\{\sigma_e^2 I_n,\ \tau^2 I_J\}$ for $k = 0,1$

where σ_e^2 is the variance component of the residual errors and τ^2 is the variance component of the random effects of the model.

For the two–level mixed model of the form given in Equation 2.16, the MINQUE estimate $\hat{\underset{\sim}{\sigma}}$ of the variance components of $\underset{\sim}{b}_0$ and $\underset{\sim}{b}_1$ using weights w_0 and w_1 in the norm is given by

(2.19) $\quad \hat{\underset{\sim}{\sigma}} = \{\mathrm{tr}(\underset{\sim}{P}_w\underset{\sim}{Z}_k\underset{\sim}{Z}_k'\underset{\sim}{P}_w\underset{\sim}{Z}_{k'}\underset{\sim}{Z}_{k'}')^{-1}\{\underset{\sim}{Y}'\underset{\sim}{P}_w\underset{\sim}{Z}_k\underset{\sim}{Z}_k'\underset{\sim}{P}_w\underset{\sim}{Y}\},$

for $k,k' = 0,1$

where $\underset{\sim}{P}_w$ which is given by

(2.20) $\quad \underset{\sim}{P}_w = \underset{\sim}{V}_w^{-1} - \underset{\sim}{V}_w^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}$

is the projection operator on the space generated by the columns of $\underset{\sim}{X}$ similar to 1.2 in Rao (1971b). $\underset{\sim}{V}_w = \underset{\sim}{Z}\underset{\sim}{D}_w\underset{\sim}{Z}'$ for $\underset{\sim}{D}_w = \mathrm{diag}\{w_0\underset{\sim}{I}_n,\ w_1\underset{\sim}{I}_J\}$ is a dispersion matrix of b where $w_0 = 1 - \rho$ and $w_1 = \rho$ for ρ the intraclass correlation coefficient. In practice, the weights w_0 and w_1 are pre–assigned numbers hence $\underset{\sim}{V}_w$ and $\underset{\sim}{P}_w$ are matrices which can be calculated easily.

To advance the MINQUE estimates associated with the weights w_0 and w_1 for this special case, define $\underset{\sim}{F}_w$ and $\underset{\sim}{U}_w$ as follows:

(2.21) $\quad F_w = \{\mathrm{tr}(P_wZ_kZ_k'P_wZ_{k'}Z_{k'}')\}$

(2.22) $\quad U_w = \{Y'P_wZ_kZ_k'P_wY\}$

for $k, k' = 0,1$. $\underset{\sim}{F}_w$ is a (2×2) matrix and $\underset{\sim}{U}_w$ is a 2–dimensional vector both originating from 2.19 such that the MINQUE estimator $\hat{\underset{\sim}{\sigma}}$ is given by

$$\hat{\underset{\sim}{\sigma}} = \underset{\sim}{F}_w^{-1}\underset{\sim}{U}_w \tag{2.23}$$

Define the matrices $\underset{\sim}{K}$ and $\underset{\sim}{A}_w$ as part of the projection operator 2.20 as

$$\underset{\sim}{K} = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^- \tag{2.24}$$

$$\underset{\sim}{A}_w = \underset{\sim}{V}_w^{-1}\underset{\sim}{X}\underset{\sim}{K}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}. \tag{2.25}$$

If we let $f_{kk'}$ to be elements of a (2×2) matrix $\underset{\sim}{F}_w$ for $k,k' = 0,1$, then the following definitions can be given:

$$f_{00} = tr(\underset{\sim}{P}_w\underset{\sim}{P}_w) = tr(\underset{\sim}{V}_w^{-2}) - tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w) \tag{2.26}$$

$$\begin{aligned} f_{01} = f_{10} &= tr(\underset{\sim}{P}_w\underset{\sim}{P}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1') \\ &= tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1') - tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1') \end{aligned} \tag{2.27}$$

$$\begin{aligned} f_{11} &= (tr(\underset{\sim}{P}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{P}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1') \\ &= tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') - tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1'). \end{aligned} \tag{2.28}$$

In order to simplify the vector $\underset{\sim}{U}_w$ of quadratic forms, we notice that

$$\begin{aligned} \underset{\sim}{P}_w\underset{\sim}{Y} &= (\underset{\sim}{V}_w^{-1} - \underset{\sim}{V}_w^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{V}_w^{-1})\underset{\sim}{Y} \\ &= \underset{\sim}{V}_w^{-1}(\underset{\sim}{Y} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}) \end{aligned}$$

such that

$$\underset{\sim}{P}_w\underset{\sim}{Y} = \underset{\sim}{V}_w^{-1}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}}) \tag{2.29}$$

where $\hat{\underset{\sim}{\alpha}}$ is the estimate of the fixed effects parameters of the model given by

$$\hat{\underset{\sim}{\alpha}} = \underset{\sim}{K}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}. \tag{2.30}$$

With this simplification, if we let u_0 and u_1 be the elements of the two

dimensional vector $\underset{\sim}{U}_w$ of quadratic forms, then the following definitions can be given:

$$(2.31) \qquad u_0 = \underset{\sim}{Y}'\underset{\sim}{P}_w\underset{\sim}{P}_w\underset{\sim}{Y} = (\underset{\sim}{Y} - \underset{\sim}{X}\hat{\alpha})'\underset{\sim}{V}_w^{-2}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\alpha})$$

$$(2.32) \qquad u_1 = \underset{\sim}{Y}'\underset{\sim}{P}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{P}_w\underset{\sim}{Y} = (\underset{\sim}{Y} - \underset{\sim}{X}\hat{\alpha})'\underset{\sim}{V}_w^{-1}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{V}_w^{-1}(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\alpha}).$$

We notice that $\underset{\sim}{Z}_1\underset{\sim}{Z}_1'$ which occurs extensively in 2.27, 2.28, and 2.32 is block diagonal with submatrices m_j of size $(n_j \times n_j)$ whose elements are all 1's and $\underset{\sim}{V}_w^{-1}$ is block diagonal with submatrices $\underset{\sim}{V}_j^{-1}$ also of size $(n_j \times n_j)$ given by

$$(2.33) \qquad \underset{\sim}{V}_j^{-1} = w(\underset{\sim}{I}_{n_j} - c_j\underset{\sim}{m}_j)$$

for $w = \frac{1}{1-w_1}$ and $c_j = \frac{w_1}{(1+(n_j-1)w_1}$ for $j = 1,2,\ldots,J$. Let $\underset{\sim}{X}_j$ be the n_j rows of the matrix $\underset{\sim}{X}$ associated with fixed effects in context j. Let $\underset{\sim}{m}_j = Z_{1j}Z_{1j}'$ for Z_{1j} a $(n_j \times 1)$ vector of 1's. Then $\underset{\sim}{K}$, $\hat{\alpha}$ and the elements of $\underset{\sim}{F}_w$ and $\underset{\sim}{U}_w$ will simplify significantly as follows:

$$(2.34) \qquad \underset{\sim}{K} = \{w\,\Sigma\,(\underset{\sim}{X}_j'\underset{\sim}{X}_j - c_j\underset{\sim}{S}_j\underset{\sim}{S}_j')\}$$

for $\underset{\sim}{S}_j = \underset{\sim}{X}_j'Z_{1j}$ a $(P \times 1)$ vector of column sums of $\underset{\sim}{X}_j$;

$$(2.35) \qquad tr(\underset{\sim}{V}_w^{-2}) = w^2\,\Sigma\,n_j\{(1-c_j)^2 + c_j^2(n_j-1)\}$$

$$(2.36) \qquad tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w) = w^3\,\Sigma\,tr(\underset{\sim}{t}_j) - a_jc_j\{(1-c_jn_j)^2 + (2-c_jn_j)\}$$

for $\underset{\sim}{t}_j = \underset{\sim}{X}_j'\underset{\sim}{X}_j\underset{\sim}{K}$ and $a_j = tr(\underset{\sim}{X}_{1j}'\underset{\sim}{K}\underset{\sim}{X}_j\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}') = tr(\underset{\sim}{S}_j'\underset{\sim}{K}\underset{\sim}{S}_j) = \underset{\sim}{S}_j'\underset{\sim}{K}\underset{\sim}{S}_j$;

$$(2.37) \qquad tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^2\,\Sigma\,n_j(1-c_jn_j)^2$$

$$(2.38) \qquad tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^3\,\Sigma\,a_j(1-c_jn_j)^3$$

$$(2.39) \qquad tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^2\,\Sigma\,n_j^2(1-c_jn_j)^2$$

$$(2.40) \qquad tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^3\,\Sigma\,n_ja_j(1-c_jn_j)^3$$

$$(2.41) \qquad \tilde{\alpha} = w\,\Sigma\,(\underset{\sim}{K}\underset{\sim}{X}_j'\underset{\sim}{Y}_j - c_jr_j\underset{\sim}{K}\underset{\sim}{S}_j)$$

for $r_j = \underset{\sim}{Z}_{1j}'\underset{\sim}{Y}_j$ is the sum of $\underset{\sim}{Y}$ elements in context j. If we define $\underset{\sim}{d}_j = \underset{\sim}{Y}_j - \underset{\sim}{X}_j\hat{\alpha}$, $h_j = \underset{\sim}{Z}_{1j}'\underset{\sim}{d}_j = \underset{\sim}{d}_j'\underset{\sim}{Z}_{1j} = \Sigma(\underset{\sim}{Y}_j - \underset{\sim}{X}_j\hat{\alpha})$; and $g_j = \underset{\sim}{d}_j'\underset{\sim}{d}_j$, then the elements of the vector U_w can be computed as

$$(2.42) \qquad u_0 = w^2 \Sigma (g_j - c_j h_j^2(2 - c_j n_j))$$

$$(2.43) \qquad u_1 = w^2 \Sigma h_j^2(1 - c_j n_j)^2.$$

As a result of adopting the formulae to the specific model of this dissertation shown in Equation 2.16, the MINQUE estimators $\hat{\underset{\sim}{\sigma}}$ of the variance components associated with the weights w_0 and w_1 are given by

$$(2.44) \qquad \hat{\underset{\sim}{\sigma}} = \begin{bmatrix} \hat{\sigma}_e^2 \\ \hat{\tau}^2 \end{bmatrix}$$

where the components $\hat{\sigma}_e^2$ and $\hat{\tau}^2$ are defined by

$$(2.45) \qquad \hat{\sigma}_e^2 = (f_{11}u_0 - f_{01}u_1)/D \text{ and}$$

$$(2.46) \qquad \hat{\tau}^2 = (f_{00}u_1 - f_{10}u_0)/D$$

where $D = f_{11}f_{00} - f_{01}f_{10}$ is the determinant of the (2×2) matrix $\underset{\sim}{F}_w$ given in 2.21.

A special case of MINQUE, namely MINQUEO is obtained by using zero for all w_k's except w_0. With such weights, $\underset{\sim}{V}_w$ reduces to $w_0\underset{\sim}{I}_n$ and $\underset{\sim}{P}_w$ to $\frac{1}{w_0}\underset{\sim}{M}$ for M given by

$$(2.47) \qquad \underset{\sim}{M} = \underset{\sim}{I}_n - \underset{\sim}{L}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{L}$$

where $\underset{\sim}{L}$ is the Cholesky decomposition of $\underset{\sim}{V}_w$. This special MINQUE0 estimator denoted by $\hat{\sigma}_0$ is given by

$$(2.48) \qquad \hat{\underset{\sim}{\sigma}}_0 = \{tr(\underset{\sim}{M}\underset{\sim}{Z}_k\underset{\sim}{Z}_k'\underset{\sim}{M}\underset{\sim}{Z}_{k'}\underset{\sim}{Z}_{k'}')\}^{-1}\{\underset{\sim}{Y}'\underset{\sim}{M}\underset{\sim}{Z}_k\underset{\sim}{Z}_k\underset{\sim}{M}\underset{\sim}{Y}\}$$

for $k,k' = 0,1$ in the place of 2.19. The estimators given in 2.48 were the first estimators suggested by Rao (1970) and they are the estimators provided by Statistical Analysis System (SAS). They also appeared in Hartley et.al. (1978) and in Seely (1971). However, Searle (1979) has indicated that, other than the fact that

these MINQUE0 estimators correspond to the rather uninformative prior value $\underset{\sim}{V}_w = w_0\underset{\sim}{I}_n$, they have no particular merit, besides being relatively easy to compute.

Choice of Initial Weights

One feature which is perhaps one of the most important features of the MINQUE estimator is in the choice of the weights w_k, $k = 0,1,...,c$. Searle (1979) indicated that regardless of the choice of the weights the MINQUE estimators will possess the properties of unbiasedness, translation invariance and minimum norm, provided the w_k's are chosen such that $\underset{\sim}{F}_w^{-1}$ exists. A version of MINQUE (Brown, 1976) known as Iterative MINQUE (I–MINQUE) is obtained as follows: After calculating a MINQUE estimate $\hat{\underset{\sim}{\sigma}}$ using arbitrary weights as in 2.19, the values therein are used as weights w_k and $\hat{\underset{\sim}{\sigma}}$ is calculated again. The process in repeated iteratively until two successive values of $\hat{\underset{\sim}{\sigma}}$ are equal (to some degree of approximation). However, because I–MINQUE estimates are obtained iteratively, they do not have the properties used in deriving 2.19; and as such they are not necessarily unbiased or "best" in any sense (Searle, 1979).

Instead of using arbitrary weights, the present study uses the method of estimating $\underset{\sim}{\sigma}$ as in Hanushek (1974). These prior estimates are then utilized to determine the weights w_k. It is then these prior weights which are used in the computation of the MINQUE estimator $\hat{\underset{\sim}{\sigma}}$ in 2.19.

It is expected that, by determining MINQUE estimates through weights established from some prior estimate of $\underset{\sim}{\sigma}$ we may obtain more efficient MINQUE estimates of the variance components than through estimation based on arbitrary weights.

Summary

Chapter II discussed the mathematical model for the hierarchical data together with the model assumption. Methods of variance component estimation were reviewed and their limitations, weaknesses and strengths discussed, for both balanced and unbalanced designs. Since the minimum norm quadratic unbiased estimation (MINQUE) procedure was adopted for the present study to be used via the bootstrap a more detailed discussion of the MINQUE was offered, particularly for the two–level hierarchical linear model. An improved method of choosing the initial weights for MINQUE was presented. A detailed review and discussion of the bootstrap method is presented in Chapter III.

CHAPTER III

THE BOOTSTRAP METHOD

Introduction

A typical problem in statistics involves estimation of an unknown population parameter θ. Two main questions arise in connection with the problem:

Question 1: Perhaps among several possible estimators of the parameter θ, what estimator $\hat{\theta}$ should be used to estimate θ?

Question 2: Having chosen some estimator $\hat{\theta}$, how accurate is it as an estimator of θ?

This situation is easily adopted to most research problems in the real world.

For the problem of estimating the parameters of a mixed hierarchical linear model, the issue of Question 1 was discussed in the greater part of Chapter II. The minimum norm quadratic unbiased estimator (MINQUE) of the variance component was adopted as the method of estimation.

The bootstrap method is generally concerned with the issues of Question 2. As mentioned earlier, the bootstrap is a computer–intensive method, which substitutes considerable amounts of computation in place of theoretical analysis (Efron and Tibshirani, 1986). The method can routinely answer questions which are far too complicated for traditional statistical analysis. Even for relatively simple problems, the computer–based bootstrap is an increasingly good data analytic bargain in an era of exponentially declining computational costs on extremely fast computers (Efron and Tibshirani, 1986).

Somewhat unfortunately, the name "bootstrap" conveys the wrong impression of "something for nothing"—of statisticians idly resampling from their samples, presumably having about as much success as they would if they tried to pull themselves up by their bootstraps (Hall, 1990). This is by no means the case. The bootstrap is a technique with a sound and promising theoretical basis (Hall, 1990).

As indicated in Chapter I (and later in this chapter), the bootstrap approach involves computing an estimate of the parameter of interest and repeating the process a large number of times by resampling with replacement. Thus, in utilizing the technique in estimating the variance component of the mixed hierarchical model described in this study, MINQUE will be used because of its desirable properties and the fact that it does not require the normality assumptions.

The primary objective of the dissertation is to demonstrate the estimation of parameters of a mixed hierarchical linear model in the total absence of distributional assumptions of the model. It was due to this fact that MINQUE was adopted since it does not require to normality (Rao, 1971). MINQUE then provides a comparable partnership with the bootstrap which is a method designed to liberate statisticians from over–reliance on the normal assumptions in data analysis.

In the literature, a distinction is made among three types of bootstrap method namely, nonparametric, smoothed and parametric bootstrap. Since it is the backbone of the study, a detailed discussion of the nonparametric bootstrap is deemed necessary. It is this discussion to which this Chapter is committed.

Nonparametric Bootstrap

In the nonparametric problem, we define a resample X^* as an unordered collection of n items drawn from X with replacement, so that each X_i^* has

probability $1/n$ of being equal to any given one of the X_j's. In other words,

$$(3.1) \qquad P(X_i^* = X_j | X) = 1/n, \ 1 \leq i, j \leq n.$$

Of course, (3.1) means that X^* is likely to contain repeats, all of which must be listed in the collection X^*. We will here formally present the nonparametric bootstrap with its associated terminologies and notation.

Let $X_1, X_2, \ldots, X_n$ be independent random variables with an unknown common distribution function F. Suppose $\hat{\theta}$ is chosen as a statistic to estimate a parameter θ of the distribution F. The bootstrap distribution of $\hat{\theta}$ is generated by taking repeated samples from $X_1, X_2, \ldots, X_n$. One such bootstrap sample is a simple random sample $X_1^*, X_2^*, \ldots, X_n^*$ of size n drawn from $X_1, X_2, \ldots, X_n$ with replacement. One bootstrap replication of the statistic $\hat{\theta}$ is then the value $\hat{\theta}^*$ of $\hat{\theta}$ computed on the bootstrap replicated sample $(X_1^*, X_2^*, \ldots, X_n^*)$. Thus, the bootstrap distribution of $\hat{\theta}$ is generated by considering a large number B of bootstrap replications $\hat{\theta}^*$ of $\hat{\theta}$.

In this process of "resampling," the n data points $X_1, X_2, \ldots, X_n$ are treated as a population with distribution function F. Let $\hat{F}$ be the empirical distribution of $X_1, X_2, \ldots, X_n$ which puts mass $1/n$ on each X_i, for $i = 1, 2, \ldots, n$. When we resample the data with replacement $X_1^*, X_2^*, \ldots, X_n^*$ are independent, with common distribution function F. The idea, then, is that the behavior of the bootstrap quantity $\hat{\theta}^*$ mimics the behavior of $\hat{\theta}$. Thus, the distribution of $\hat{\theta}^*$ could be generated from the data and used to approximate the unknown sampling distribution of $\hat{\theta}$.

Another way of looking at the bootstrap estimation is by supposing that $X_1, X_2, \ldots, X_n$ are independently and identically distributed random variables from a population with unknown cumulative density function, F, and suppose the objective is to draw inferences about some parameter θ of the population. If $\hat{\theta}$ is an estimator of θ and $\hat{F}$ is the sample cdf that assigns mass $1/n$ to each X_i,

$i = 1, 2,..., n$, then Schenker (1985) has clarified that the bootstrap approximates the sampling distribution of $\hat{\theta}$ under F by the sampling distribution of $\hat{\theta}^*$ under $\hat{F}$.

Several studies (Efron & Gong, 1983; Bickel & Freedman, 1981; Efron, 1981a; Schenker, 1985) have presented more straightforward procedure for performing Monte Carlo simulation through steps. As a simple example, suppose our goal is to estimate the standard error of the sample mean $\bar{X}$. Let $\hat{\theta} = \bar{X} = \Sigma X_i/n$ (the sample mean). The bootstrap algorithm will approach this problem in the following steps:

Step 1 Given the sample $X_1, X_2, ..., X_n$, construct $\hat{F}$ by assigning mass $1/n$ to each of the X_i, $i = 1, 2,....,n$.

(3.2) $\hat{F}$: mass $1/n$ at X_i, $i = 1,2,...,n$

Step 2 Draw a bootstrap sample $X_1^*, X_2^*, ..., X_n^*$ from F with replacement and calculate

$$\hat{\theta}^* = \frac{1}{n} \sum_{1}^{n} X_i^*$$

Step 3 Independently do Step 2 some number B times, obtaining bootstrap replications $\hat{\theta}_1^*, \hat{\theta}_2^*, ..., \hat{\theta}_B^*$.

Step 4 Calculate the standard error $\sigma(\hat{\theta}^*)$ of $\hat{\theta}^*$ by

$$\sigma(\hat{\theta}^*) = \left[\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \hat{\theta}_\cdot^*)^2 \right]^{1/2}$$

where

$$\hat{\theta}_\cdot^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^*.$$

However, the question of how well the empirical distribution of $\hat{\theta}^*$ under $\hat{F}$ approximate the sampling distribution of $\hat{\theta}$ under F is certainly crucial. Freedman and Bickel (1981) applied the bootstrap algorithm to estimate the standard error of the sample mean $\bar{X}$ assuming they did not know the formula

$\sigma/\sqrt{n}$. With a population of 6,672 Americans aged 18–79 in Cycle I of the Health Examination Survey, Freedman and Bickel were interested in the estimate of the mean systolic blood pressure in millimeters of mercury. Using B = 100, the bootstrap algorithm yielded a mean systolic blood pressure of 129.6 with a standard deviation of 21.4 compared to the sample mean of 130.3 and a standard deviation of 23.2 millimeters of mercury. On plotting the bootstrap and the theoretical distribution of the mean, the bootstrap distribution followed the theoretical distribution rather closely.

In estimating the distribution of the 5%, 10%, and 25% trimmed means, Efron (1986) used B = 200 bootstrap replications for several trials along with the jackknife and the theoretical optimization. The simulation results showed that the bootstrap clearly out performed the jackknife and its results were surprisingly close to the theoretical optimum for a scale–invariant standard deviation estimate assuming full knowledge of the parametric family.

In spite of a great deal of encouragement from several studies (Beran, 1984; Lunneborg, 1985; Rasmussen, 1987; Singh, 1981) in assessing the accuracy of Efron's bootstrap, others like Dolker, Halperin, and Divgi (1982) have expressed doubt. The key question about the bootstrap technique is its accuracy in a situation where the sample size is small. What can one say in the case of small samples, especially if the true distribution has long, thick tails like in the case of the Cauchy distribution (Nash, 1981)? On the other hand, the advocates of the bootstrap may question whether the problem of small sample sizes and long, fat tails is unique to the bootstrap or whether it is a problem affecting most statistical methods.

The Bootstrap Estimate of Bias

The bootstrap technique of estimating bias is based on the idea that for a statistic $\hat{\theta}$, which estimates a parameter θ, $\hat{\theta}^*$ is the bootstrap estimate of $\hat{\theta}$.

Thus, the bootstrap estimate of bias, denoted by BIAS is given by

$$(3.3) \qquad \text{BIAS} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}^*_b - \hat{\theta}) = \hat{\theta}^*_\cdot - \hat{\theta},$$

where $\hat{\theta}$ is the usual estimate of θ and $\hat{\theta}^*_\cdot$ is the average of $\hat{\theta}^*_b$ over all bootstrap replicated values of $\hat{\theta}^*$. Efron (1982) used this principle to compare the relative accuracy of the bootstrap with other methods of estimation. By using 1000 bootstrap replications for the law school data which yielded $\hat{\rho}^* = 0.779$ compared to $\hat{\rho} = 0.776$, Efron found that the bootstrap BIAS = 0.003, compared to –0.007 for the jackknife and –0.011 for normal theory. In this study, a similar principle will be utilized to demonstrate the relative accuracy of the bootstrap, and the usual estimates based on this principle of the estimate of bias.

The Bootstrap Confidence Intervals

Approximate confidence limits for the parameter can be found via the bootstrap using either the t–method or the bootstrap percentile method.

(i) The t–method: Suppose, for example, that $\hat{\theta}$ is the usual estimator of the parameter θ. The usual confidence interval are based on the assumption that, the statistic T, given by

$$(3.4) \qquad T = \frac{\hat{\theta} - \theta}{\hat{\sigma}_{\hat{\theta}}},$$

is approximately distributed N(0, 1). Then the $(1-\alpha)100\%$ confidence interval is given by

$$(3.5) \qquad \hat{\theta} \pm Z_{1-\alpha/2}\hat{\sigma}_{\hat{\theta}}.$$

Using the same principle, we can set the bootstrap confidence intervals by utilizing the fact that the bootstrap estimator, $\hat{\theta}^*$ is the estimator of $\hat{\theta}$. Let the statistic T^*_b at replication b be defined by

$$(3.6) \qquad T_b^* = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}_{\hat{\theta}^*}}$$

where $\hat{\theta}^*$ is based on the bootstrap sample and $\hat{\sigma}^2_{\hat{\theta}^*}$ is the estimator of $\mathrm{Var}(\hat{\theta})$ based on the bootstrap sample. If we observe T_b^* a large number B times, for b = 1,2,....B, we will estimate the distribution of T by the empirical bootstrap distribution of T*'s. Then the bootstrap confidence interval about the parameter θ are determined by the following process: find a number T_c such that the proportion of T*'s between $-T_c$ and T_c is $(1-\alpha)$. The bootstrap $(1-\alpha)100\%$ confidence interval is then given by

$$(3.7) \qquad \hat{\theta} \pm T_c \hat{\sigma}_{\hat{\theta}}.$$

(ii) Percentile method: The bootstrap percentile method for constructing confidence intervals about the parameter θ may be presented as follows.

Let $\hat{\theta}$ be the usual statistic estimating a parameter θ based on the original sample. Define the statistic $d = \hat{\theta} - \theta$. We can approximate the distribution of d by the bootstrap distribution of $d^* = \hat{\theta}_b^* - \hat{\theta}$, for b = 1,2,...,B, where $\hat{\theta}_b^*$ are the estimates of θ based on repeated samples from the original sample. If we find d_c such that $(1-\alpha)100\%$ of all d_b^*'s are between $-d_c$ and d_c, then the $(1-\alpha)100\%$ confidence interval about θ is given by

$$(3.8) \qquad \theta \pm d_c.$$

Correction for Bias in Bootstrap Estimation

Perhaps the most important application of the bootstrap is in estimating the distributions and standard errors of statistics based on independent observations. But in many problems of practical interests, the bootstrap is employed to estimate an expected value (Hall, 1990). For example, if $\hat{\theta}$ is an estimate of an unknown

parameter θ, then we might want to estimate bias, $E(\hat{\theta}-\theta)$, or the distribution function of $d = \hat{\theta} - \theta$.

In the case of bias, using Efron's notation, we have,

$$\text{BIAS} = E(\hat{\theta}) - \theta = E(\hat{\theta}-\theta)$$

Using $\theta = (\theta-\hat{\theta}) + \hat{\theta}$

$$\theta = \hat{\theta} - (\hat{\theta}-\theta) = \hat{\theta} - d \quad \text{for} \quad d = \hat{\theta} - \theta, \tag{3.9}$$

we can approximate the distribution of d by the distribution of $d^* = \hat{\theta}^* - \hat{\theta}$ where $\hat{\theta}^*$ is the version of $\hat{\theta}$ computed from bootstrap sample $\{X_1^*, X_2^*, \ldots, X_n^*\}$, rather than from the original sample $\{X_1, X_2, \ldots, X_n\}$. Thus we have an estimate of bias, $\hat{\text{BIAS}}$ given by

$$\hat{\text{BIAS}} = \hat{\theta}_\cdot^* - \hat{\theta}$$

and by 3.9 θ can be estimated by

$$\begin{aligned} \hat{\theta}_{\text{BOOT}} &= \hat{\theta} - (\hat{\theta}_\cdot^* - \hat{\theta}) \\ &= \hat{\theta} - \hat{\theta}_\cdot^* + \hat{\theta} \end{aligned}$$

$$\hat{\theta}_{\text{BOOT}} = 2\hat{\theta} - \hat{\theta}_\cdot^* \tag{3.10}$$

Equation 3.10 is thus the bootstrap point estimate of the parameter θ which is corrected for bias. This method which is referred to as "pivoting" has been stressed in many studies of the bootstrap method for confidence intervals (Abramovitch and Singh, 1985; DiCiccio and Tibshirani, 1987; Hinkley and Wei, 1984; Hall, 1990; and Schenker, 1985). In most instances, pivoting amounts to "studentizing" or correcting for scale. However, in certain situations, pivoting is difficult to sustain in problems where scale cannot be estimated in a stable way (Hall, 1990). But if we perceive that the major advantage of "standard" bootstrap methods is the ability to construct confidence intervals of particularly difficult problems, and if we see respect of transformation as a major property, then pivoting is advocated (Hall, 1990).

CHAPTER IV

DESIGN OF THE STUDY

Introduction

The primary purpose of the present study was to estimate the parameters of a two level nested hierarchical linear mixed model under a variety of conditions. One such condition which forms the focus of the study is in estimating the model parameters (fixed and random) in situations where the random error terms and sets of random effects are not normally distributed. The also demonstrated the ability of the bootstrap algorithm in providing the estimates of the fixed and random effects of the model, generating bootstrap empirical distributions and standard errors of the statistics and thereby setting confidence intervals about the parameters. This chapter presents the methodology and design employed in the study.

Implementation of the study design required a method of generating samples drawn from populations of known parameters. Data generation and analysis were performed on an IBM 3090 VF mainframe computer at Michigan State University. Programs used in generating data and Monte Carlo simulations were coded in Statistical Analysis Systems (SAS), mostly using Interactive Matrix Language (SAS/IML). Data were generated in such a way so that sets of data were sampled from populations of known parameter values in order to provide a check for the performance of estimation procedures and their properties. Data were sampled were the normal and double exponential (or Laplace) population distributions.

The normal distribution represented the situation where the classical estimation methods are usable. The double exponential or Laplace distribution (an

example of a distribution with long and fat tails) represented a departure from normality. In order to assess the relative usefulness of the bootstrap method, estimation of the mixed model parameters were studied under the two distributional models (normal and double exponential) and three specifications of the population intraclass correlation condition. Specification of the intraclass correlation, the study design, and parameter values are presented in the later part of this chapter.

Generation of Data

The uniform random number generator (UNIFORM) function in SAS/IML was used to generate random uniform deviates in the interval (0,1). In certain instances, this uniform random number generating function was used as a basis for generating other random numbers by applying some standard linear transformations to these uniform deviates. The SAS/IML software provides a procedure which generates independent values from a standard univariate normal distribution. The package also provides a procedure (REPEAT) which creates a matrix or vector by repeating the values of the argument. These three features of SAS/IML were used in generating data sets for the study.

In generating samples drawn from a population of known parameters, the data created had to fit certain assumptions. Each subject's observed score Y_{ij} was assumed to have been influenced by a combination of factors and effects. The pool of factors and effects included the fixed effects factor α; a covariate X whose coefficient is denoted by β; the random effects factor b_j which are assumed to be distributed with mean 0 and variance τ^2; and the random error term ϵ also distributed with mean 0 and variance σ_e^2. A typical observed value Y_{ij} containing all these features is generated through the equation.

$$Y_{ij} = \mu + \alpha_k + \beta X_{ij} + b_j + \epsilon_{ij} \qquad (4.1)$$

where

Y_{ij} is the observed value of subject i in context j;

α_k is the effect of level k of the fixed factor;

X_{ij} is the covariate value of subject i in context j whose coefficient is denoted by β;

b_j is the effect of level j (context j) of the random factor;

ϵ_{ij} is the random error term associated with subject i in context j.

With $k = 1,2,...,P-2$; $j = 1,2,...,J$; and $i = 1,2,...,n_j$, and by using the matrix notation, it can be shown that Equation 4.1 is similar to Equation 2.9 in Chapter II where the term $\underset{\sim}{X}\underset{\sim}{\alpha}$ in 2.9 represents the first three terms of 4.1.

While the SAS/IML procedure NORMAL was utilized to generate independent values from a standard univariate normal distribution, SAS/IML program segments were coded to generate double exponential variates. The most direct way to generate double exponential (or Laplace) variates involves first generating two uniform random variates, U_1 and U_2 in the interval (0,1). Set $X_1 = -\ln(U_1)$ and $X_2 = -1$ if $U_2 < 0.5$. If $U_2 \geq 0.5$, then set $X_2 = 1$. Then the variates Y defined by the equation

$$(4.2) \qquad Y = \frac{1}{\sqrt{2}} X_1 X_2$$

are distributed as double exponential with mean zero and variance 2.

SAS/IML code segments used to generate the normal and double exponential distributions are given in Appendix B.

Study Design and Parameter Values

The structure of data in the present study is assumed to involve a random factor consisting of J levels, nested within some fixed factor levels. The random

factor may be characterized by contexts such as schools or countries and the fixed factor characterized by sector (e.g. public, private, or religious) in the case of schools as context. In the case of countries as context, the fixed factor levels may be taken to be levels of economic or industrial development (e.g. developed, less developed, developing or underdeveloped) or may be world regions.

As noted earlier, two design factors in this study are expected to influence the success of the estimation of model parameters. These are the population distribution of the random components and the population intraclass correlation. The intraclass correlation denoted by ρ is given by

$$\rho = \frac{\tau^2}{\tau^2 + \sigma_e^2} \tag{4.3}$$

where σ_e^2 and τ^2 are the intra– and inter–class variances of the model respectively. As Raudenbush and Bryk (1988) indicated, the intraclass correlation has two useful and mathematically equivalent interpretations. First, it is the correlation between pairs of values within the J contexts such that it measures the degree of dependence among observations sharing a context. Secondly, as a ratio, it represents the proportion of the total variation in the response values which is between contexts. Estimation of variance components is often difficult when ρ is quite small, sometimes resulting in negative estimates of the variance components. Due to this feature, three levels of the intraclass correlation for each of the two distributional models were introduced in the study as part of the design factors. In order to vary the intraclass correlation, σ_e^2 was fixed at 100 while τ^2 was allowed to take values 1, 5.26, and 25 resulting in ρ taking values of 0.01, 0.05, and 0.20 respectively. Table 4.1 presents design factor combination trials.

Table 4.1

Design Factor Combination Trials*

Intraclass Correlations (ρ)	Distribution Model: Normal	Distribution Model: Double Exponential	All
0.01	a	b	i
0.05	c	d	j
0.20	e	f	k
All	g	h	l

* 400 trials (different sets of data) were specified for each cell (a through f).

The design factor specification shown in Table 4.1 provided for a total of 2400 Monte Carlo simulation trials, each consisting of a different data set. As a result, 1200 trials were performed for each of the two distributional models (normal and double exponential) and 800 trials for each of the three levels of the intraclass correlation, such that, $g = h = 1200$; $i = j = k = 800$ and $g + h = i + j + k = 2400$.

The specific mixed model used in the study has two factors, a random factor with J levels, nested within a fixed factor with three levels and a micro level covariate variable. However, it should be noted that it is possible to extend this model by including additional covariates (at micro or macro level). For the purpose of the present study, all data sets used in the study were unbalanced (unequal number of subjects in each context) consisting of 50 macro units.

Fixing the parameter value of τ^2 at unit (near boundary value of zero) provided an additional advantage to the study. This is due to the interest of the study in estimating the random effect variance component τ^2 near the boundary conditions. It is in these situations where most variance component estimation

procedures experience problems of giving negative estimates of τ^2 when the parameter value they are estimating is essentially positive. Thus, in an attempt to understand the performance of the bootstrap procedure in estimating τ^2 near boundary conditions, out of the total 2400 trials 800 (or 33.3%) were performed for $\tau^2 = 1$ ($\rho = 0.01$), 800 (or 33.3%) for $\tau^2 = 5.26$ (or $\rho = 0.05$), and 800 (or 33.3%) for $\tau^2 = 25$ (or $\rho = 0.20$).

It should be emphasized that, in using the bootstrap algorithm to estimate the distribution of the parameters of the mixed model described in the study, estimation is done at each of b bootstrap replication, for b = 1,2,...,B, where B is a large number. For the present study, B was set at 200 bootstrap replications for each trial shown in Table 4.1.

Implementation of the Bootstrap using MINQUE

The MINQUE method of estimating the variance components requires using weights w_k associated with b_k for $k = 0,1$. Ordinarily, arbitrary weights are chosen provided one ensures that F_w^{-1} exists. According to Rao (1972), regardless of the choice of weight, w_k's, the MINQUE estimators will still possess the properties of unbiasedness, translation invariance and minimum norm. However, though the MINQUE estimators may generally possess the properties used in deriving the estimators (unbiasedness, translation invariant, and minimum norm), one would expect that, in practice, these estimators may be as good as the prior weights that were utilized. In other words, the MINQUE estimators depend to a certain extent on the prior weights used in the norm. Indeed, this condition was the motivation behind Brown (1976) who suggested iterative MINQUE (I–MINQUE). But since I–MINQUE estimators are obtained iteratively, they do not possess the properties used in deriving MINQUE. Thus, I–MINQUE estimators are not necessarily unbiased or "best" in any sense (Searle, 1979).

Instead of using arbitrary weights in implementing MINQUE, the present study employed an ANOVA–type method of independently estimating the variance components of the mixed model as in Hanushek (1974). The values of this prior estimates are used to derive the weights used in MINQUE. Using Hanushek's method, we fit ordinary regression model with all independent variables as predictors. The prior estimator, $\hat{\sigma}^2_M$ of the random error variance σ^2_e is taken as usual MSE in the multiple regression model. However, the Hanushek estimator $\hat{\tau}^2_M$ for the variance, τ^2 of the random effects of the model is given by

$$\hat{\tau}^2_M = \frac{w-(N-P)\hat{\sigma}^2_M}{N-T} \qquad (4.4)$$

where w is the sums of squares of residual in the regression model

$T = \Sigma\, tr\{S_j'(X_j'X_j)^{-1}S_j\}$ for $S_j = (P\times 1)$ vector of column sums of X_j for context j

N = sample size, and

P = number of fixed effects parameters in the model.

In order to use the Hanushek estimators $\hat{\tau}^2_M$ and $\hat{\sigma}^2_M$ to derive the weight w_0 and w_1, define the ratio $R_M = \hat{\sigma}^2_M/\hat{\tau}^2_M$. The weights w_0 and w_1 can then be obtained by

$$w_0 = \frac{R_M}{1+R_M} \text{ and } w_1 = \frac{1}{1+R_M}. \qquad (4.5)$$

We notice that $w_0 = 1-\rho_M$ and $w_1 = \rho_M$ where $\rho_M = \frac{\hat{\tau}^2_M}{\hat{\tau}^2_M+\hat{\sigma}^2_M}$. The value ρ_M is the intraclass correlation based on the Hanushek estimates, $\hat{\tau}^2_M$ and $\hat{\sigma}^2_M$ of τ^2 and σ^2_e respectively. The weights w_0 and w_1 obtained through 4.5 are the values used in the MINQUE procedure. It is reasonable to expect that the MINQUE based on weights established from some prior estimates of σ^2_e and τ^2 could be an improvement over the conventional MINQUE based on arbitrary weights.

Implementation of the bootstrap algorithm to estimate the parameters of the mixed, hierarchical linear model in the present study requires a random sampling procedure with replacement. First a random sample of J macro units (e.g. countries, schools) with replacement from the available sample of J macro units is drawn. From each of the selected macro units, a random sample of size n_j micro units are selected with replacement for $j = 1,2,...,J$. The resulting data set is termed the bootstrap replication sample (Efron, 1981). Based on the bootstrap replicated sample, the MINQUE procedure is used to determine the estimate of the parameters of the model. The process is repeated a large number B times yielding B MINQUE estimates. This technique may be presented in a sequence of steps as follows:

Step 1. Construct the distributions F_J by assigning mass $1/_J$ to each of the macro units.

Step 2. From the J macro units, select a random sample of size J with replacement

Step 3. For each of the J selected macro units containing n_j micro units, construct distributions F_{n_j} by assigning mass $1/n_j$ to the jth macro unit, for $j = 1,2,...,J$.

Step 4. From each of the J macro units whose distributions were constructed at Step 3 above, draw a random sample size n_j with replacement for $j = 1,2,...,J$.

At the end of Step 4, the resulting data is termed the bootstrap replicated sample. The vector of observations at this stage is denoted by $\underset{\sim}{Y}^*$.

Step 5. From the bootstrap replicated data set generated at Step 4, determine the MINQUE estimate of the parameters of the model given by $\underset{\sim}{\hat{\sigma}}^*$ and $\hat{\alpha}^*$.

Step 6. Independently, repeat 2, 4 and 5 a large number B times to obtain a sequence of MINQUE estimates of the parameters of the model

$$\hat{\underset{\sim}{\sigma}}^*_b \text{ and } \hat{\underset{\sim}{\alpha}}^*_b \text{ for } b = 1,2,....B$$

Step 7: Observe the distribution of the values $\hat{\underset{\sim}{\sigma}}^*_b$ and $\hat{\underset{\sim}{\alpha}}^*_b$ as the empirical bootstrap distribution of the estimates of the variance component and the fixed effects of the model.

The bootstrap standard error of each of the component of the estimates is given by

$$(4.6) \qquad s.e.(\hat{\theta}^*) = [(B-1)^{-1} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta}^*.)^2]^{1/2}$$

where $\hat{\theta}.^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}^*_b$ for $\hat{\theta}^*$ being any one of the components of $\hat{\underset{\sim}{\sigma}}$ or $\hat{\underset{\sim}{\alpha}}$.

The Computer Programs

Three main tasks in this study required the use of a computer program. These were: Generating data sets from population with known parameter and distribution; Monte Carlo simulations; and bootstrapping. Independent computer programs were coded for each task using SAS/IML package.

SAS/IML available in the MSU IBM 3090 VF mainframe computer system is a double precision and multilevel, interactive programming language. SAS/IML software is both flexible and powerful since it combines the advantages of high–level and low–level languages (SAS/IML User's Guide, 1985, p. xi).

Though SAS provides a procedure which computes the MINQUE that corresponds to the rather uninformative prior by using zero weights as an option to PROC VARCOMP, this procedure does not handle models that involve covariates. The independent variables handled by the procedure PROC VARCOMP are limited to main effects, interaction and nested effects; but no covariate effects are allowed in

the PROC VARCOMP Statement (SAS User's Guide: Statistics, 1985, p. 819). However, the present study is not limited to models which do not involve covariates. Consequently, the more flexible software SAS/IML was utilized in the study, not only to estimate parameters of the model but also to generate data.

As indicated earlier in this chapter, the first computer program generates sample observations, $\underset{\sim}{Y}$ and covariate $\underset{\sim}{X}$, and passes them over to the program that implements the bootstrap algorithm. The bootstrap estimates at each replication are written to a standard SAS file for further analysis. The Monte Carlo simulation computer program is implemented like the bootstrap program except that while the bootstrap samples data from a sample generated from the population, the Monte Carlo simulation program samples data directly from the population. The bootstrap SAS/IML code used in this study is thus flexible and available to be used to estimate parameters of a model using data obtained from real world research.

Applicability of the bootstrap method using the present SAS/IML code is demonstrated in the present study. The computer code and method are applied on actual field research data to estimate the parmeters of the model, the sampling distribution of the statistics and to set bootstrap confidence intervals about the parameters of teachers' self–efficacy prediction model. Estimation results for the fixed and random effects of the teachers' self–efficacy model are presented in Chapter V of this dissertation.

CHAPTER V

APPLICATION OF BOOTSTRAP AND MINQUE: HIGHER ORDER TEACHING

Introduction

The bootstrap is a new method whose time has come with the advent of modern computers. Though its applicability in generating sampling distributions of statistics and in construction of confidence intervals about parameters is highly promising, the method has not been widely used in educational and social science research. Strengths of the method are often demonstrated in situations where parametric modeling is difficult and/or normal assumptions are not possible. These situations are not uncommon in educational and social science research.

The interest of the present study was to demonstrate the operation of the bootstrap in a two–level hierarchical linear model. The focus of the study was upon the estimation of the group and individual level variances and fixed effects parameters of the mixed model. A highly promising approach offered by the method in this study was that of estimating the sampling distribution of the statistics and thereby setting confidence intervals about the parameters. The study used computer–simulated data to extensively assess the distributional behavior of parameter estimates under varying distributional assumptions of the errors and sets of random effects parameters.

In this chapter, applicability of the bootstrap algorithm on data originating from a real research situation is demonstrated. The method is applied on the actual field research data to estimate the parameters of the model, the sampling distribution of the estimators and to set the bootstrap confidence intervals about

the parameters. Data used in this demonstration of the applicability of the bootstrap method was part of the data gathered earlier to investigate the contextual effects on the self–efficacy of high school teachers.

Description of data and variables

The data was obtained through a survey of teachers in sixteen schools who taught Mathematics, Science, English, or Social Science. Each teacher was assigned to teach one or more classes in the school. Though the individual teacher was viewed as the basic unit of analysis, each teacher provided information on several classes. As a result, we view the teachers as the "macro" units of analysis with the classes they taught as the "micro" units of analysis. The teacher effects therefore, constitute the random factor of the model.

The dependent variable in the study was teachers' perception of self–efficacy which was measured at the class level. A measure of teachers' self–efficacy represents a person's perceived expectancy of enacting a desired level or type of performance through personal effort (Bandura, 1986). For instance, a teacher who possess a high level of self–efficacy will be of the view that, no matter the nature of students or facilities he or she is provided with, he or she will produce an excellent level of performance. On the other hand, a teacher with low self–efficacy will feel paralyzed if he or she is given "poor" children. The phenomenon has been identified to have an effect on both students' and teachers' performance (Fuller, et.al., 1982).

In the present study, the extent to which teachers' self–efficacy is influenced by institutional, classroom and individual teacher characteristics is examined. Academic subject taught (Mathematics, Science, English, or Social Science) represented the primary fixed factor of the model used to predict teachers' self–efficacy. Other independent variables of the model which were viewed as covariates fell into two categories, namely, between– and within–teacher variables.

The between teacher variables included: STAFCOOP, cooperation of staff; TCONTROL, Teacher control; and PLEADER, Principal leadership. The within teacher (or classroom) level independent variables included: STUDACH, class average student achievement level; LVLPREP, class level of preparation; and SIZE, class size.

Selection of valid data for the variables of interest resulted in a sample of 244 teachers who provided information on 1634 classes taught. Breakdown of the number of teachers and number of classes by academic subject areas were as follows: Mathematics had 63 teachers with 370 classrooms; Science had 59 teachers with 391 classrooms; English had 69 teachers with 509 classrooms; and Social Science had 53 teachers with 364 classrooms. The average number of classes for which each teacher provided information was about 6.6.

The Model Statements

We begin by posing a within–teacher model that defines a "micro" equation with EFFICACY as the response variable and LVLPREP, SIZE and STUDACH as "micro" regressors which are identical for each teacher j as,

$$(5.1)\quad (\text{EFFICACY})_{ij} = \beta_{0j} + \sum_{h=1}^{4} (\text{SUBJECT})\, h_j + \beta_{1j}(\text{LULPREP}_i)_j + \beta_{2j}(\text{SIZE}_i)_j + \beta_{3j}(\text{STUDACH}_i)_j + \epsilon_{ij}$$

where $j = 1,\ldots,J$ teachers and $i = 1,\ldots,n_j$ classes for each teacher j. Since β_{0j}, β_{1j}, β_{2j}, and β_{3j} are defined for each teacher, we can pose the between–teacher model using these coefficients as responses similar to Equations 2.2 and 2.3 in Chapter II of this dissertation. Specifically, we consider the intercept β_{0j} to be random and dependent on the between–teacher independent variables such that the associated "macro" model is given by

(5.2) $$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{STAFCOOP})_j + \gamma_{02}(\text{TCONTROL})_j + \gamma_{03}(\text{PLEADER})_j + e_{oj}$$

where $j = 1,2,...,J$ teachers. Combining Equation 5.1 and 5.2 yields,

(5.3) $$(\text{EFFICACY})_{ij} = [\gamma_{00} + \gamma_{01}(\text{STAFCOOP})_j + \gamma_{02}(\text{TCONTROL})_j + \gamma_{03}(\text{PLEADER})_j + \sum_{h=1}^{4}(\text{SUBJECT})\,h_j + \beta_{1j}(\text{LVLPREP}_i)_j + \beta_{2j}(\text{SIZE}_i)_j + \beta_{3j}(\text{STUDACH}_i)_j] + [e_{0j} + \epsilon_{ij}]$$

similar to Equation 2.7 in Chapter II. Model Equation 5.3 can be written in the general linear matrix notation as in Equation 2.8 in Chapter II for teacher j with,

(5.4) $$\underset{\sim}{Y}_j = (\text{EFFICACY})_{ij}$$

(5.5) $$\underset{\sim}{X}_j\underset{\sim}{\alpha}_j = \gamma_{00} + \gamma_{01}(\text{STAFCOOP})_j + \gamma_{02}(\text{TCONTROL})_j + \gamma_{03}(\text{PLEADER})_j + \sum_{h=1}^{4}(\text{SUBJECT})\,h_j + \beta_{1j}(\text{LVLPREP}_i)_j + \beta_{2j}(\text{SIZE}_i)_j + \beta_{3j}(\text{STUDACH}_i)_j$$

(5.6) $$\underset{\sim}{Z}_j\underset{\sim}{b}_j = e_{0j} \text{ for } b_j = e_{0j} \text{ and } Z_j = (1,...,1)'$$

(5.7) $$\underset{\sim}{\epsilon}_j = \epsilon_{ij}.$$ Equations 5.5 and 5.6 represent the fixed and random effects of the model respectively, while Equation 5.7 is the expression for the random errors of the model. The intent then is to estimate both the fixed and random effects of the model on the measure of teachers' self–efficacy.

Estimation Procedure

The ability of the bootstrap to estimate parameters of the model given in Equation 5.3 was demonstrated through the used of MINQUE. For each parameter, the usual MINQUE estimates were provided based on the original sample. The bootstrap estimates based on $B = 1000$ repeated resampling with replacement were also obtained. Due to the bootstrap's ability to generate sampling distributions

through resampling, 95% bootstrap confidence intervals about each of the parameters were also provided.

Estimates were provided for a total of 14 parameters of the model. There were four effect levels of the factor, SUBJECT, denoted by α_1, α_2, α_3, and α_4 corresponding to Mathematics, Science, English and Social Science respectively. Parameters for other fixed factors (or covariates) were denoted by $(\beta_1,\beta_2,\beta_3)$ corresponding to the within teacher (or class–room) effects (LVLPREP, SIZE, STUDACH) and $(\gamma_1,\gamma_2,\gamma_3)$ corresponding to the effects between teacher (LVLPREP, SIZE, STUDACH, STAFCOOP, TCONTROL, and PLEADER). Besides SUBJECT, all other fixed factor were viewed as covariates in the model. The inter–teacher variance of the model was denoted by τ^2 while σ_e^2 denoted the variance of the random errors (or intra–teacher variance). In addition, the intra–teacher correlation denoted by ρ, and computed as in Equation 4.3 in Chapter IV and the constant common to all observations denoted by γ_{00} were estimated through both MINQUE and bootstrap.

Results of Estimation

Table 5.1 presents the MINQUE and bootstrap results for the estimation of the fourteen parameters of the teacher self–efficacy prediction model. The results provides the usual MINQUE estimate, the bootstrap estimate which is the average over all B = 1000 bootstrap replications, the bootstrap standard error, and 95% bootstrap confidence intervals about each parameter. The bootstrap estimate of bias given by $\hat{\theta}^*_\cdot - \hat{\theta}$, where $\hat{\theta}^*_\cdot$ is the average of the estimator over the B bootstrap replications and $\hat{\theta}$ is the usual estimator based on the original sample is also represented. For purposes of consistency with notation given by Efron (1979), the bootstrap estimate of bias is denoted by BIAS.

Table 5.1

Bootstrap and MINQUE estimates of the effects of type of subject, school climate, and classroom variables on the teachers' perceived self–efficacy.

			Bootstrap (B=1000)		95% Bootstrap C.I.			
Parameter		MINQUE	Average	S.D.	L.C.L.	U.C.L.	Width	Bias
Intra–Teacher Variance:	τ^2	0.1320	0.1397	0.0115	0.1000	0.146	0.0460	0.0077
Inter–Teacher Variance:	σ^2	0.2800	0.2817	0.0123	0.2546	0.3026	0.0480	0.0017
Intra–Class Correlation:	ρ	0.3204	0.3314	0.0219	0.2645	0.3509	0.0864	0.0110
Constant:	β_0	1.6880	1.8004	0.0624	1.4593	1.6948	0.2355	0.1124
Academic Subjects								
Mathematics:	α_1	0.4830	0.5066	0.0289	0.4049	0.5164	0.1115	0.0236
Science:	α_2	0.4560	0.4779	0.0267	0.3827	0.4832	0.1005	0.0219
English:	α_3	0.4290	0.4610	0.0271	0.3457	0.4467	0.1010	0.0320
Social Science:	α_4	0.3210	0.3557	0.0283	0.2383	0.3428	0.1045	0.0347
Class Level of Preparation:	β_1	0.0720	0.0582	0.0101	0.0642	0.1052	0.041	-0.0138
Class Size:	β_2	0.0390	0.0384	0.0154	0.0080	0.0700	0.0620	-0.0006
Average Student Achievement Level:	β_3	0.2510	0.2283	0.0138	0.2505	0.3000	0.0495	-0.0227
Staff–Cooperation:	γ_1	0.0570	0.0484	0.0240	0.0160	0.1105	0.0945	-0.0086
Teacher Control:	γ_2	0.1230	0.1594	0.0303	0.0252	0.1437	0.1185	0.0364
Principal Leadership:	γ_3	-0.0160	-0.0108	0.0275	-0.0745	0.0355	0.1100	0.0052

From Table 5.1, it is shown that the average of the bootstrap estimates of the parameters over $B = 1000$ replications did not differ much from the usual MINQUE estimates. In addition, the bootstrap feature which was not available through the MINQUE procedure was the estimation of the standard error of the estimate. The statistic showed low bootstrap standard errors of the estimate for all fourteen parameters of the model. The lowest value of the bootstrap estimate of the standard error was observed for the estimates of the effect of class SIZE and the estimate of the intra–teacher variance, $\hat{\sigma}_e^2$. For these two parameter estimates, the bootstrap estimate of bias was less than 0.002. Except for the estimator of the constant, γ_{00} whose bootstrap estimate of bias was 0.1124, the bootstrap estimate of bias for all other statistic was no more than 0.04.

An accomplishment of the bootstrap method which is not readily available through the usual MINQUE was the construction of confidence intervals about each of the parameters of the model. The 95% bootstrap confidence intervals were used as a means of testing for the significance of both the fixed and random effects on the teachers' self–efficacy. Based on the 95% confidence intervals, the results showed that all factors with the exception of Principal Leadership have statistically significant effect on teachers' Self–Efficacy.

The intra–class correlation denoted by ρ was significantly different from zero, with 0.3204 and 0.3314 being the MINQUE and bootstrap estimates of ρ respectively. Estimates of ρ through both methods indicated that, approximately 30% of the total variance in teachers' Self–Efficacy is between teachers. The MINQUE and bootstrap estimate of the inter–teacher variance denoted by τ^2 was 0.1320 and 0.1397 respectively.

In some problems of practical interest, we may wish to observe the behavior of the statistic used to estimate a parameter. This requires knowledge of the sampling distribution of the statistic, often based on the Gaussian theory. In

situations where this theory is not available, it is often difficult to draw conclusions about the sampling distribution of the statistic. In such situations, the bootstrap offers perhaps one of the most significant contributions to statistics. The method's applicability to even complicated problems involving statistics which may not have closed form expressions may be a major promise of the bootstrap. But the bootstrap can be applied to simple problems as well.

In the present study, the bootstrap method was used to generate the sampling distributions of the statistics used to estimate the parameters of the teachers self–efficacy prediction model. The distributions were based on 1000 bootstrap replications.

Figure 5.1 presents three percentage polygons of the estimators of the inter–teacher variance, τ^2, the intra–teacher variance, σ_e^2, and the intra–teacher correlation, ρ, based on B = 1000 bootstrap replications. Though the distribution of $\hat{\tau}^{2*}$ appeared to be slightly positively skewed, the distributions of all the three estimators are fairly symmetric with very low dispersion.

Figure 5.2 presents four percentage polygons of the estimators of the effects of Mathematics, α_1, Science, α_2, English, α_3, and Social Science, α_4 on teachers' self–efficacy based on B = 1000 bootstrap replications. All four charts represent the empirical bootstrap distribution of the estimator of SUBJECT effects on the teachers' perception of self–efficacy appears to be fairly symmetric with moderate variability. However, the estimates of the effect of Mathematics, Social Science, and Science seem to be slightly negatively skewed. But most importantly, all the charts show that the replicated estimates are centered extremely close to the MINQUE estimates of the effects of Mathematics, Science, English, and Social Science.

Figure 5.1

Percentage polygons for the bootstrap estimate of the inter and intra–teacher variances and the intra–teacher correlation for the teachers' self–efficacy prediction (B = 1000 replications).

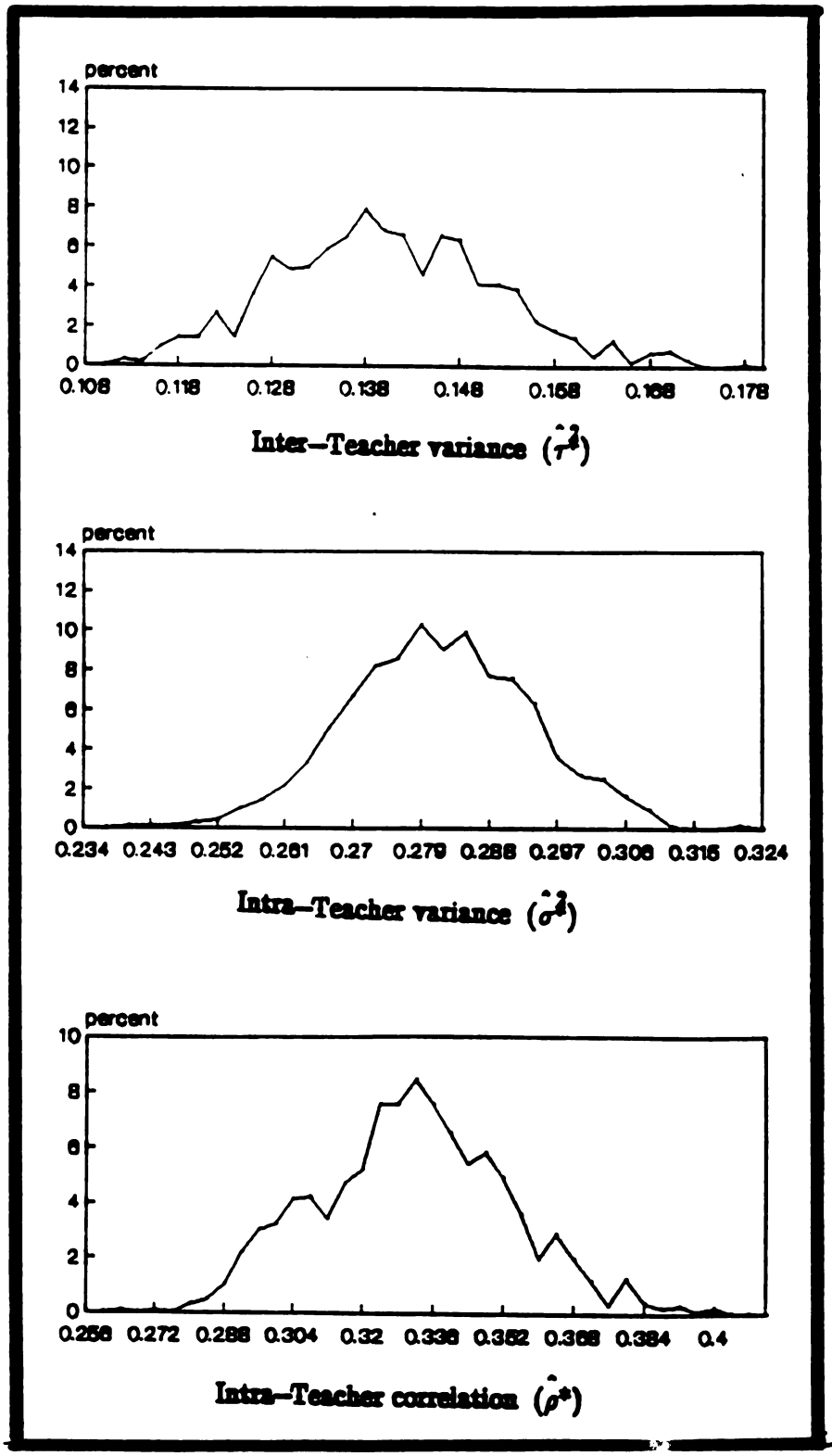
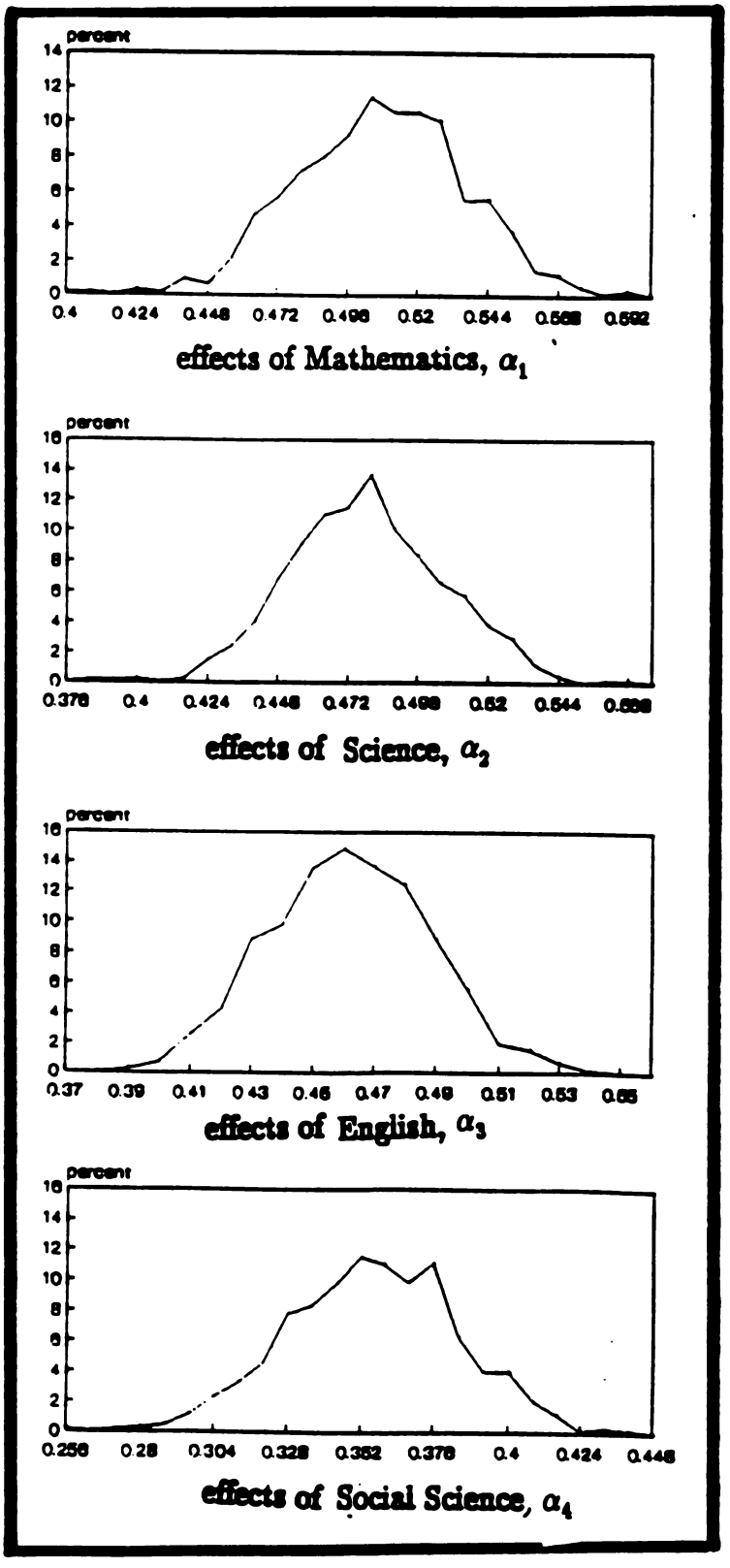


Figure 5.2

Percentage polygons for the bootstrap estimate of the effects of Mathematics, Science, English, and Social Science on the teachers' self–efficacy (B = 1000 replications).



CHAPTER VI

SIMULATIONS AND BOOTSTRAP RESULTS

Overview

The purpose of the study was to demonstrate the use of the bootstrap in providing estimates of the parameters of a general two level mixed hierarchical linear model, determining the standard error of the estimates and their empirical bootstrap distributions. The objective was to observe the behavior of the bootstrap and MINQUE estimates of the fixed effects and the variance components of the model under several conditions including situations where the normal distributional assumptions may be violated. The study examined the influence of the magnitude of the population intraclass correlation and the tail size of the distribution on the estimation of the parameters of the mixed model. Double exponential (or LaPlace) represented a distribution with fairly long and thick tails.

MINQUE and bootstrap abilities to estimate parameters of the mixed HLM model were demonstrated by estimating the parameters from a large number of independent samples generated from populations of known distributions and parameter values. Applying the estimation procedures on sets of data generated from a population of known parameters provided a means of evaluating the relative effectiveness of the methods of estimation. The independent samples consisted of 50 groups with each group containing 25 to 45 observations.

Estimation of parameters was studied for two underlying population distributions, namely the normal and double exponential (or Laplace), and three levels of the intraclass correlation. These two design factors provided for a total of six design factor combinations (or cells). A total of 400 trials (based on independent

samples) were performed for each design factor combination. As a result, 2400 Monte Carlo simulation trials, each based on a different data set, were performed for the study. MINQUE and bootstrap point estimates, the 95% and 90% bootstrap confidence intervals, empirical bootstrap distribution and standard errors were provided for each trial. The MINQUE and bootstrap summary results are presented in the remaining part of this chapter.

Results of Estimation Procedures

Simulated data represented observations from two population distributions of random errors and sets of random effects characterized by three levels of the intraclass correlation. The mixed model contained seven parameters of which three were random effects parameters and four fixed. The random effects parameters were the within and between group variances denoted by σ_e^2 and τ^2 respectively and the intraclass correlation denoted by ρ. The fixed effects parameters included α_1, α_2 and α_3 for levels of the fixed factor and β, the coefficient of the covariate. The MINQUE and bootstrap estimates were obtained for each of the seven parameters of the model. While one MINQUE estimate was obtained at each trial, (based on the original sample) the bootstrap estimate at each trial was the average over 200 bootstrap replicated values. Thus, the average of the bootstrap estimate over 400 trials is the average of the 400 averages each computed from 200 bootstrap replications. Ten functions of the MINQUE and/or bootstrap estimates were computed for the six non–redundant estimates for both models under normal and double exponential error terms and sets of random effects at each trial. The average and standard deviation of the estimates of these functions over 400 trials for each design factor combination (cells denoted by a through f of Table 4.1 in Chapter IV) are presented in Tables 6.1 through 6.6.

Ten functions of the estimates consisted of: MINQUE and bootstrap estimates, the bootstrap estimate of bias denoted by BIAS, the MINQUE bias and bootstrap estimate of bias denoted by D_1 and D_2 respectively, the MINQUE ratio, R_2 and its corresponding bootstrap ratio denoted by R_1, the MINQUE and bootstrap mean square error denoted by MSE1 and MSE2 respectively, and the bootstrap/MINQUE measure of relative efficiency.

Table 6.1 presents the average and standard deviation of ten functions of $\hat{\tau}^2$ and/or $\hat{\tau}^2_*$ over the 400 trials under the normal and double exponential error terms and sets of random effects for three levels of the intraclass correlation. From Table 6.1 it is shown that the bootstrap overestimated τ^2 with a bias of 0.3432 and 0.3765 under normal and double exponential respectively for $\rho = 0.01$. For this low value of the intraclass correlation, the bootstrap estimate of bias, denoted by BIAS and given by $\hat{\tau}^2_* - \hat{\tau}^2$ was also high and positive indicating that the bootstrap method on average overestimated the value of τ^2 under both normal and double exponential error terms and sets of random effects of the model. At this level of the intraclass correlation ($\rho = 0.01$), though MINQUE on average also overestimated τ^2, its bias was relatively low, at 0.0292 under normal and 0.0597 under the double exponential distribution. The ratio, R_2 expected to be 1.00 was observed at 1.0292 while its bootstrap estimate, R_1 was 0.7811 under the normal distribution. The same ratios were 1.0597 and 2.3032 respectively under the double exponential. At this level of the intraclass correlation condition, the bootstrap estimate seemed to be more efficient both under the normal and double exponential. From these results, it is apparent that MINQUE and bootstrap estimates were fairly close both under the normal and double exponential distributions.

Table 6.1

Average and standard deviation of the functions of the estimates $\hat{\tau}^2$ and/or $\hat{\tau}^2_*$ under the normal and double exponential error and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.01	Bootstrap, $\hat{\tau}^2_*$	1.00	1.3432	0.7495	1.3765	0.7298
	MINQUE, $\hat{\tau}^2$	1.00	1.0292	0.9010	1.0597	0.8879
	BIAS=$\hat{\tau}^2_*-\hat{\tau}^2$		0.3140	0.2113	0.3168	0.2207
	$D_1=\hat{\tau}^2-\tau^2$		0.0292	0.9010	0.0597	0.8879
	$D_2=\hat{\tau}^2_*-\tau^2$		0.3432	0.7495	0.3765	0.7298
	$R_1=\hat{\tau}^2_*/\hat{\tau}^2$		0.7811	11.0436	2.3032	14.4144
	$R_2=\hat{\tau}^2/\tau^2$		1.0292	0.9010	1.0597	0.8879
	MSE1=$(\hat{\tau}^2-\tau^2)^2$		0.8105	1.2806	0.7899	1.2086
	MSE2=$(\hat{\tau}^2_*-\tau^2)^2$		0.6781	1.3690	0.6729	1.2438
	Rel. Efficiency@		0.8366		0.8519	

Table 6.1 (continued)

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.05	Bootstrap, $\hat{\tau}^2_*$	5.26	5.2900	1.6634	5.5309	2.1743
	MINQUE, $\hat{\tau}^2$	5.26	5.1755	1.6621	5.3996	2.1786
	BIAS=$\hat{\tau}^2_*-\hat{\tau}^2$		0.1144	0.1201	0.1313	0.1316
	$D_1=\hat{\tau}^2-\tau^2$		−0.0845	1.6621	0.1396	2.1786
	$D_2=\hat{\tau}^2_*-\tau^2$		0.0299	1.6634	0.2709	2.1743
	$R_1=\hat{\tau}^2_*/\hat{\tau}^2$		1.0261	0.0314	1.0354	0.0972
	$R_2=\hat{\tau}^2/\tau^2$		0.9839	0.3160	1.0265	0.4142
	MSE1=$(\hat{\tau}^2-\tau^2)^2$		2.7627	3.5337	4.7539	8.4574
	MSE2=$(\hat{\tau}^2_*-\tau^2)^2$		2.7610	3.6216	4.7891	8.7037
	Rel Efficiency@		0.9994		1.0074	
0.20	Bootstrap,$\hat{\tau}^2_*$	25.00	24.9553	5.8896	25.5167	8.8370
	MINQUE, $\hat{\tau}^2$	25.00	24.8520	5.8766	25.4018	8.8398
	BIAS=$\hat{\tau}^2_*-\hat{\tau}^2$		0.1032	0.2050	0.1149	0.2077
	$D_1=\hat{\tau}^2-\tau^2$		−0.1480	5.8766	0.4018	8.8398
	$D_2=\hat{\tau}^2_*-\tau^2$		−0.0447	5.8896	0.5167	8.8370
	$R_1=\hat{\tau}^2_*/\hat{\tau}^2$		1.0043	0.0085	1.0052	0.0089
	$R_2=\hat{\tau}^2/\tau^2$		0.9941	0.2351	1.0161	0.3536
	MSE1=$(\hat{\tau}^2-\tau^2)^2$		34.4700	49.1902	78.1073	153.6137
	MSE2=$(\hat{\tau}^2_*-\tau^2)^2$		34.6030	49.6024	78.1651	154.7723
	Rel. Efficiency@		1.0039		1.0007	

@ Rel. Efficiency = MSE2/MSE1

At the second level of the intraclass correlation ($\rho = 0.05$), the average values of $\hat{\tau}^2$ and $\hat{\tau}^2_*$ were 5.1755 and 5.2900 in the normal case compared to the true parameter value set at 5.26. At this level of the intraclass correlation, both the bootstrap and MINQUE estimates were very close to the parameter τ^2 with 0.0299 and –0.0845 as their respective biases under normality. The estimates were slightly off under double exponential with $\hat{\tau}^2 = 5.3996$ with a bias of 0.1396 and $\hat{\tau}^2_* = 5.5309$ with a bias of 0.2709.

Under this condition of the population intraclass correlation however, the strength of the bootstrap was demonstrated in the estimates R_1 and R_2. The average value of R_1 was 1.0261 under the normal and 1.0354 under the double exponential compared to the average values of R_2 which was observed at 0.9839 under the normal and 1.0265 under the double exponential.

Perhaps the most successful estimation of τ^2 was attained in the situation where the population intraclass correlation was 0.20, particularly under the normal distribution. Compared to the true parameter value of $\tau^2 = 25$, $\hat{\tau}^2_*$ was observed at 24.9553 and $\hat{\tau}^2$ at 24.8520 under the normal. The average values of $\hat{\tau}^2$ and $\hat{\tau}^2_*$ were 25.4018 and 25.5167 respectively under the double exponential distribution. Based on the bias of these estimators, the results shows that the bootstrap with a bias of –0.0447 and the MINQUE with a bias of –0.1480 were very close under the normal distribution. The bootstrap estimate of bias was also observed at 0.1032 under the normal. The bias for the MINQUE and the bootstrap estimates were observed at 0.4018 and 0.5167 respectively under the double exponential, with the bootstrap estimate of bias equal to 0.1149.

It may be important to note that, in both the normal and double exponential distributions, the average values of R_1 and R_2 were quite close to 1.00. In particular, the ratio R_1 was surprisingly close to 1.00 indicating a very successful bootstrap estimation process. The bootstrap replicated values of R_1 were not only centered near 1.00 but also were less variable under both the normal and double exponential. The measure of relative efficiency for the two estimators was also extremely close to 1.00 both under the normal and double exponential.

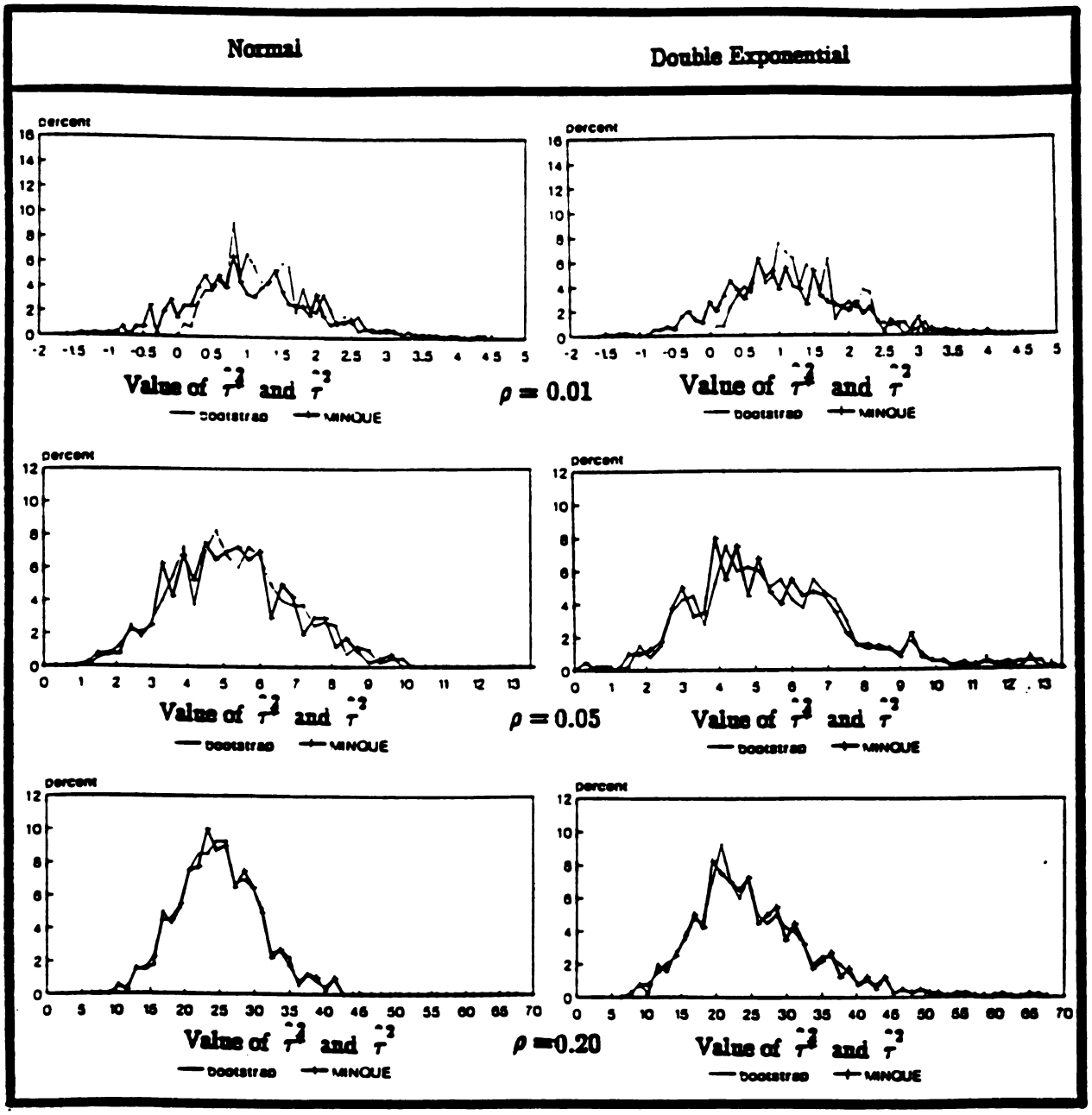
Figure 6.1 displays the percentage polygons of the 400 bootstrap and MINQUE estimates of τ^2 under the normal and double exponential errors and sets of random effects at each of the three levels of the population intraclass correlation. At $\rho = 0.01$, both MINQUE and the bootstrap estimates at each trial were centered near the true parameter value set at 1.00 under both the normal and double exponential. However, the percentage polygon for the bootstrap was positively skewed while that of the MINQUE was nearly symmetrical under both the normal and double exponential distributions. This is mainly due to the fact that the bootstrap was protected from giving negative estimates of τ^2 while MINQUE was not.

From Figure 6.1 it is apparent that a greater mass of observations were around 1.00 for the bootstrap frequent polygon than for the MINQUE polygon. It can therefore be argued that, at this level of the population intraclass correlation, the bootstrap seemed to be a good complement to the MINQUE estimator of τ^2.

Percentage polygons for the 400 MINQUE and bootstrap estimates under the normal and double exponential distributions at $\rho = 0.05$ shows that, both MINQUE and the bootstrap were free of giving negative estimates and both were

Figure 6.1

Percentage polygons for the MINQUE and bootstrap estimate of τ^2 over 400 trials under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.



centered near the true parameter value of τ^2 which was set at 5.26. However, for both MINQUE and the bootstrap, the estimates were more variable under the double exponential than under the normal distribution.

The difference in the variability of both the MINQUE and bootstrap estimators between the normal and double exponential were more apparent at $\rho = 0.20$ (see Figure 6.1). The percentage polygons for both estimators showed more variability under the double exponential than under the normal. Estimation results at the three levels of the population intraclass correlation show that, though the bootstrap seems to be a more stable estimator of τ^2, particularly at the low level of the intraclass correlation, the characteristic of the tails of the distribution seem to be equally affecting the bootstrap and MINQUE in estimating τ^2.

Table 6.2 presents the average and standard deviation of the ten estimable functions of $\hat{\sigma}_e^2$ and/or $\hat{\sigma}_*^2$ over 400 trials under the normal and double exponential error terms and sets of random effects for three levels of the population intraclass correlation. From Table 6.2 it is shown that the bootstrap slightly underestimated σ_e^2 under both the normal and double exponential distributions for $\rho = 0.01$. MINQUE slightly underestimated σ_e^2 under normality but slightly overestimated σ_e^2 under the double exponential for $\rho = 0.01$.

At this level of the population intraclass correlation, the bias for MINQUE was –0.0680 under the normal and 0.0698 under the double exponential. The bias for the bootstrap estimate was –0.2025 under the normal and –0.0730 under the double exponential. The bootstrap estimate of bias was observed at –0.1344 under the normal and –0.1428 under the double exponential. Average results for R_1 and R_2 demonstrated a very successful estimation process at this level of the population intraclass correlation. R_1 was observed at 0.9987 for the normal and at 0.9986 for

Table 6.2

Average and standard deviation of the functions of the estimates $\hat{\sigma}^2$ and/or $\hat{\sigma}_*^2$ under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.01	Bootstrap $\hat{\sigma}_*^2$	100.00	99.7975	3.6352	99.9270	5.9080
	MINQUE, $\hat{\sigma}^2$	100.00	99.9320	3.6223	100.0698	5.9100
	BIAS=$\hat{\sigma}_*^2-\hat{\sigma}^2$		−0.1344	0.2331	−0.1428	0.3939
	$D_1=\hat{\sigma}^2-\sigma^2$		−0.0680	3.6223	0.0698	5.9100
	$D_2=\hat{\sigma}_*^2-\sigma^2$		−0.2025	3.6352	−0.0730	5.9080
	$R_1=\hat{\sigma}_*^2/\hat{\sigma}^2$		0.9987	0.0023	0.9986	0.0039
	$R_2=\hat{\sigma}^2/\sigma^2$		0.9993	0.0362	1.0007	0.0591
	MSE1=$(\hat{\sigma}^2-\sigma^2)^2$		13.0927	17.0420	34.8462	50.6894
	MSE2=$(\hat{\sigma}_*^2-\sigma^2)^2$		13.2225	17.1911	34.8222	51.0716
	Rel. Efficiency@		1.0099		0.9993	

Table 6.2 (continued)

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.05	Bootstrap, $\hat{\sigma}^2_*$	100.00	99.6360	3.4799	99.9349	5.8466
	MINQUE, $\hat{\sigma}^2$	100.00	99.7693	3.4733	100.0755	5.8629
	BIAS=$\hat{\sigma}^2_*-\hat{\sigma}^2$		−0.1333	0.2398	−0.1406	0.3942
	$D_1=\hat{\sigma}^2-\sigma^2$		−0.2307	3.4733	0.0755	5.8629
	$D_2=\hat{\sigma}^2_*-\sigma^2$		−0.3640	3.4799	−0.0651	5.8466
	$R_1=\hat{\sigma}^2_*/\hat{\sigma}^2$		0.9987	0.0024	0.9987	0.0039
	$R_2=\hat{\sigma}^2/\sigma^2$		0.9977	0.0347	1.0008	0.0586
	MSE1=$(\hat{\sigma}^2-\sigma^2)^2$		12.0867	17.0071	34.3938	49.9322
	MSE2=$(\hat{\sigma}^2_*-\sigma^2)^2$		12.2122	17.2576	34.1013	50.2230
	Rel. Efficiency@		1.0104		0.9915	
0.20	Bootstrap, $\hat{\sigma}^2_*$	100.00	100.0437	3.7669	99.7328	5.7180
	MINQUE, $\hat{\sigma}^2$	100.00	100.1660	3.7717	99.8545	5.7151
	BIAS=$\hat{\sigma}^2_*-\hat{\sigma}^2$		−0.1223	0.2361	−0.1216	0.3878
	$D_1=\hat{\sigma}^2-\sigma^2$		0.1660	3.7718	−0.1455	5.7151
	$D_2=\hat{\sigma}^2_*-\sigma^2$		0.0437	3.7669	−0.2672	5.7180
	$R_1=\hat{\sigma}^2_*/\hat{\sigma}^2$		0.9988	0.0024	0.9988	0.0039
	$R_2=\hat{\sigma}^2/\sigma^2$		1.0017	0.0377	0.9985	0.0572
	MSE1=$(\hat{\sigma}^2-\sigma^2)^2$		14.2177	19.9419	32.6019	44.6796
	MSE2=$(\hat{\sigma}^2_*-\sigma^2)^2$		14.1557	19.9124	32.6852	45.4206
	Rel. Efficiency@		0.9956		1.0026	

@ Rel. Efficiency = MSE2/MSE1

the double exponential while R_2 was observed at 0.9993 under the normal and at 1.0007 under double exponential.

Similarly, surprisingly accurate results were observed at the 0.05 level of the intraclass correlation. At this level, the average of bootstrap estimate of σ_e^2 was observed at 99.6360 with a bias of –0.3640 under the normal and at 99.9349 with a bias of –0.00651 under double exponential. The average of the MINQUE estimate of σ_e^2 was observed at 99.7693 with a bias of –0.2307 under the normal and at 100.0755 with a bias of 0.0755 under double exponential.

Compared to the expected ratio of the estimates at 1.00, both MINQUE and the bootstrap very closely estimated the ratio with $R_1 = 0.9987$ both under the normal and double exponential. R_2 was observed at 0.9977 and 1.0008 under the normal and double exponential respectively. At the two levels of the intraclass correlation condition ($\rho = 0.01$ and $\rho = 0.05$), the bootstrap and MINQUE were very close both under the normal and double exponential distribution.

At the 0.20 level of the intraclass correlation, both MINQUE and the bootstrap slightly overestimated σ_e^2 under the normal and slightly underestimated σ_e^2 under double exponential. The bootstrap average was closer to the true value of the parameter than the MINQUE with a bias of 0.0437 under the normal distribution. On the other hand, the MINQUE was closer to the parameter than the bootstrap with a bias of –0.1455 under the double exponential distribution. Average values of R_1 and R_2 were very close to their expected value (1.00) at this level of the intraclass correlation under both the normal and double exponential distributions.

On average therefore, results in Table 6.2 shows that both MINQUE and the bootstrap very closely estimated the parameter σ_e^2 under both the normal and

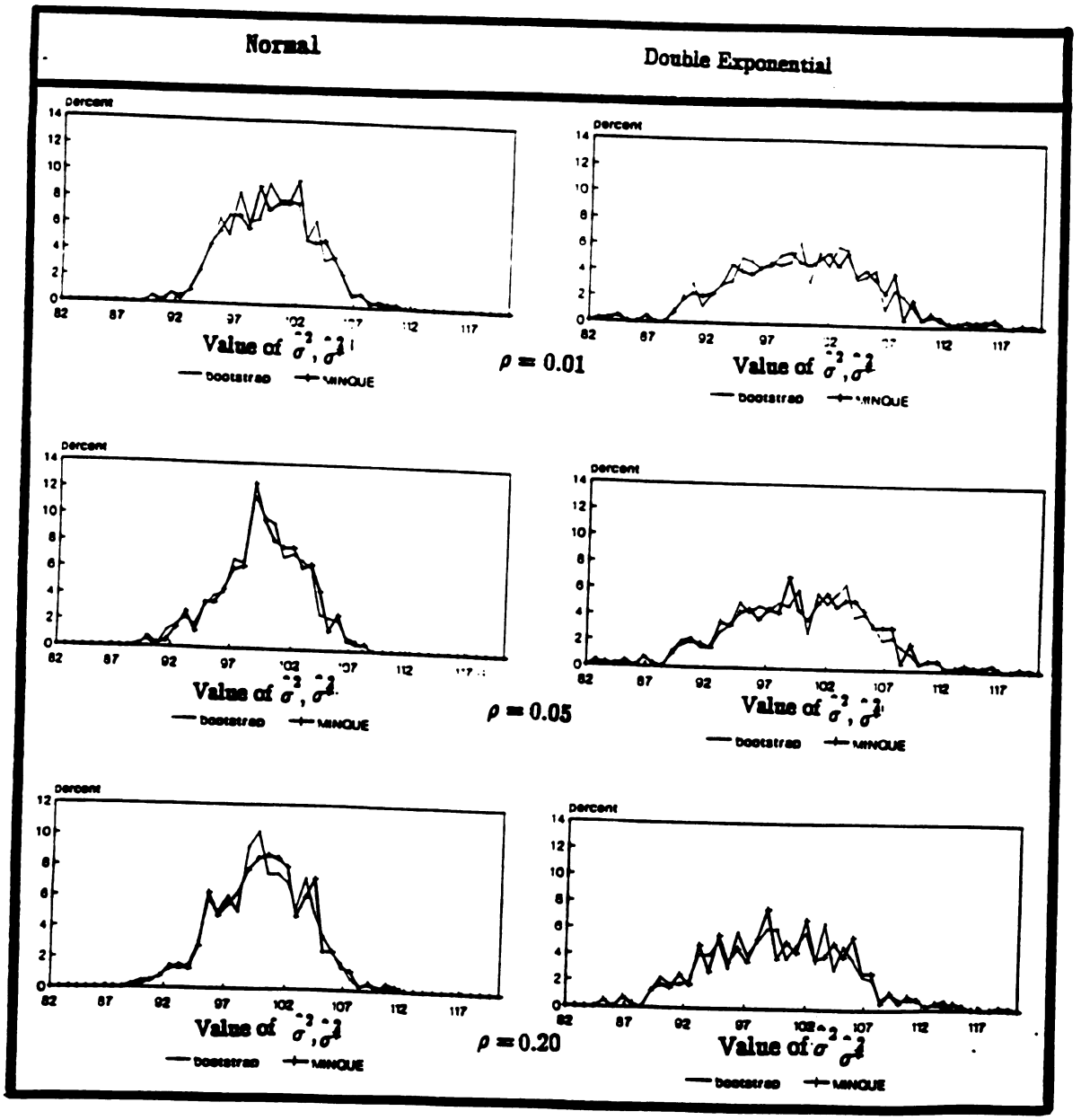
double exponential errors and sets of random effects at all three levels of the intraclass correlation. However, at all three levels of the intraclass correlation, the standard deviation of the functions of the estimates was relatively higher under double exponential that under the normal distribution. Regardless of the underlying distribution of the errors and sets of random effects of the model, the estimation of the ratio of the estimators, R_1 and R_2 was quite close to 1.00 at all levels of the intraclass correlation.

Figure 6.2 displays the percentage polygons of the 400 bootstrap and MINQUE estimates of σ_e^2 under the normal and double exponential errors and sets of random effects at each of the three levels of the population intraclass correlation. From Figure 6.2 we see that, at all levels of the population intraclass correlation, the bootstrap estimator followed the MINQUE quite closely. Percentage polygons for both estimators were centered near the true parameter value set at 100. However, differences in variation of the estimates by distribution was quite obvious. The spread of both MINQUE and bootstrap frequent polygons was clearly higher under the double exponential than under the normal distribution. For the estimation of σ_e^2, therefore, it can be argued that while both MINQUE and the bootstrap fairly closely estimated σ_e^2, their efficiency was severely affected by the nature and size of the tails of the distribution of the errors and sets of random effects. Both estimators were less efficient under a distribution with long and thick tails (like that of the double exponential) than under a distribution with short and lighter tails.

The intraclass correlation is given as a function of τ^2 and σ_e^2 whose formula is shown in Equation 4.3. It is the index which measures the degree of

Figure 6.2

Percentage polygons for the MINQUE and bootstrap estimate of σ_e^2 over 40 trials under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.



dependence among observations sharing a context as well as providing the proportion of the total variation in the response values that is between contexts (Randenbush and Bryk, 1988). Success of estimation of model parameters often depends on this measure, with less success when ρ is quite small. Due to this feature, the population intraclass correlation was used as an important design factor in the present study.

The MINQUE estimator $\hat{\rho}$ of ρ is obtained by substituting $\hat{\tau}^2$ for τ^2 and $\hat{\sigma}_e^2$ for σ_e^2 in Equation 4.3. Likewise, the bootstrap estimator $\hat{\rho}^*$ is obtained by substituting $\hat{\tau}_*^2$ and $\hat{\sigma}_{e*}^2$ for τ^2 and σ_e^2 respectively in Equation 4.3.

Bootstrap and Monte Carlo results for the estimation of ten estimable functions of $\hat{\rho}$ and/or $\hat{\rho}_*$ under the normal and double exponential errors and sets of random effects of the model for the three levels of the population intraclass correlation are presented in Table 6.3. Summary statistics in Table 6.3 show that both MINQUE and the bootstrap slightly overestimated ρ under both the normal and double exponential distributions at the 0.01 level of the intraclass correlation. The bias for the MINQUE and bootstrap were correspondingly 0.00013 and 0.0032 under the normal and 0.0005 and 0.0035 under double exponential. The bootstrap estimate of bias was 0.0031 under the normal and 0.003 under double exponential. The results show that though R_1 was poorly estimated at $\rho = 0.01$, estimation of the other nine functions of $\hat{\rho}$ and/or $\hat{\rho}_*$ near their expected value at this level of the population intraclass correlation. Mean square errors for both MINQUE and the bootstrap were surprisingly close to zero, both under the normal and double exponential distributions for $\rho = 0.01$.

At the 0.05 level of the intraclass correlation, all the ten estimable functions of $\hat{\rho}$ and/or $\hat{\rho}_*$ were very close to their expected values estimated. The average

Table 6.3

Average and standard deviation of the functions of the estimates $\hat{\rho}$ and/or $\hat{\rho}_*$ under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.01	Bootstrap, $\hat{\rho}_*$	0.01	0.0132	0.0072	0.0134	0.0070
	MINQUE, $\hat{\rho}$	0.01	0.0101	0.0088	0.0105	0.0087
	$BIAS=\hat{\rho}_*-\hat{\rho}$		0.0031	0.0021	0.0030	0.0022
	$D_1=\hat{\rho}-\rho$		0.00013	0.0088	0.0005	0.0088
	$D_2=\hat{\rho}_*-\rho$		0.0032	0.0072	0.0035	0.0070
	$R_1=\hat{\tau}_*/\hat{\rho}$		1.1407	3.4598	1.3958	3.3220
	$R_2=\hat{\rho}/\rho$		1.0132	0.8798	1.0460	0.8672
	$MSE1=(\hat{\rho}-\rho)^2$		0.0001	0.0001	0.0001	0.0001
	$MSE2=(\hat{\rho}-\rho)^2$		0.0001	0.0001	0.0001	0.0001
	Rel. Efficiency@		1.0000		1.0000	

Table 6.3 (continued)

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.05	Bootstrap, $\hat{\rho}_*$	0.05	0.0501	0.0151	0.0521	0.0193
	MINQUE, $\hat{\rho}$	0.05	0.0491	0.0151	0.0509	0.0193
	$\hat{BIAS}=\hat{\rho}_*-\hat{\rho}$		0.0010	0.0011	0.0012	0.0012
	$D_1=\hat{\rho}-\rho$		-0.0009	0.0151	0.0009	0.0193
	$D_2=\hat{\rho}_*-\rho$		0.0001	0.0151	0.0021	0.0193
	$R_1=\hat{\rho}_*/\hat{\rho}$		1.0237	0.0306	1.0306	0.0962
	$R_2=\hat{\rho}/\rho$		0.9829	0.3023	1.0182	0.3866
	$MSE1=(\hat{\rho}-\rho)^2$		0.0002	0.0003	0.0004	0.0006
	$MSE2=(\hat{\rho}_*-\rho)^2$		0.0002	0.0003	0.0004	0.0006
	Rel. Efficiency①		1.0000		1.0000	
0.20	Bootstrap, $\hat{\rho}_*$	0.20	0.1978	0.0381	0.2002	0.0534
	MINQUE, $\hat{\rho}$	0.20	0.1972	0.0381	0.1992	0.0534
	$\hat{BIAS}=\hat{\rho}_*-\hat{\rho}$		0.0006	0.0014	0.0009	0.0014
	$D_1=\hat{\rho}-\rho$		-0.0028	0.0381	-0.0008	0.0534
	$D_2=\hat{\rho}_*-\rho$		-0.0022	0.0381	0.0002	0.0534
	$R_1=\hat{\rho}_*/\hat{\rho}$		1.0033	0.0074	1.0051	0.0080
	$R_2=\hat{\rho}/\rho$		0.9860	0.1906	0.9962	0.2671
	$MSE1=(\hat{\rho}-\rho)^2$		0.0015	0.0021	0.0028	0.0043
	$MSE2=(\hat{\rho}_*-\rho)^2$		0.0015	0.0021	0.0028	0.0044
	Rel. Efficiency①		1.0000		1.0000	

① Rel. Efficiency = MSE2/MSE1

values of $\hat{\rho}$ and $\hat{\rho}_*$ were 0.0491 and 0.0501 respectively under the normal and 0.0509 and 0.0521 respectively under double exponential. Accordingly, the biases for the bootstrap and MINQUE were respectively 0.0001 and –0.0009 under the normal and 0.0021 and 0.0009 respectively under double exponential. The bootstrap estimate of bias was 0.0010 under the normal and 0.0012 under double exponential. Under this condition of the intraclass correlation, R_1 and R_2 were fairly close to 1.00 with $R_1 = 1.0237$ and 1.0336 under the normal and double exponential respectively and $R_2 = 0.9829$ and 1.0182 under normal and double exponential respectively. The mean square error for both MINQUE and bootstrap was quite low under both normal and double exponential at this level of the intraclass correlation.

The bootstrap slightly underestimated ρ under the normal but very accurately estimated ρ under double exponential at the 0.20 condition of the intraclass correlation. On the other hand, on average MINQUE slightly underestimated ρ both under the normal and double exponential at the 0.20 condition of the intraclass correlation. The MINQUE and bootstrap biases were observed at –0.0028 and –0.0022 respectively under the normal and –0.0008 and 0.0002 respectively under double exponential. The bootstrap estimate of bias was 0.0006 under the normal and 0.0009 under double exponential. R_1 was surprisingly close to 1.00 under both the normal and double exponential but R_2 was slightly less than 1.00 under both the normal and double exponential.

At this level of the intraclass correlation condition, both MINQUE and bootstrap mean square errors were quite low, both observed at 0.0015 under the normal and 0.0028 under double exponential. Based on these results therefore, it is apparent that, while estimation of functions of $(\hat{\tau}_*^2, \hat{\tau}^2)$ and $(\hat{\sigma}_*^2, \hat{\sigma}^2)$ may not have been very successful, estimation of functions of $(\hat{\rho}_*, \hat{\rho})$ which in turn depends on $\hat{\tau}_*^2$, $\hat{\tau}^2$, $\hat{\sigma}_*^2$, and $\hat{\sigma}^2$ seemed to have been fairly successful for both MINQUE and the

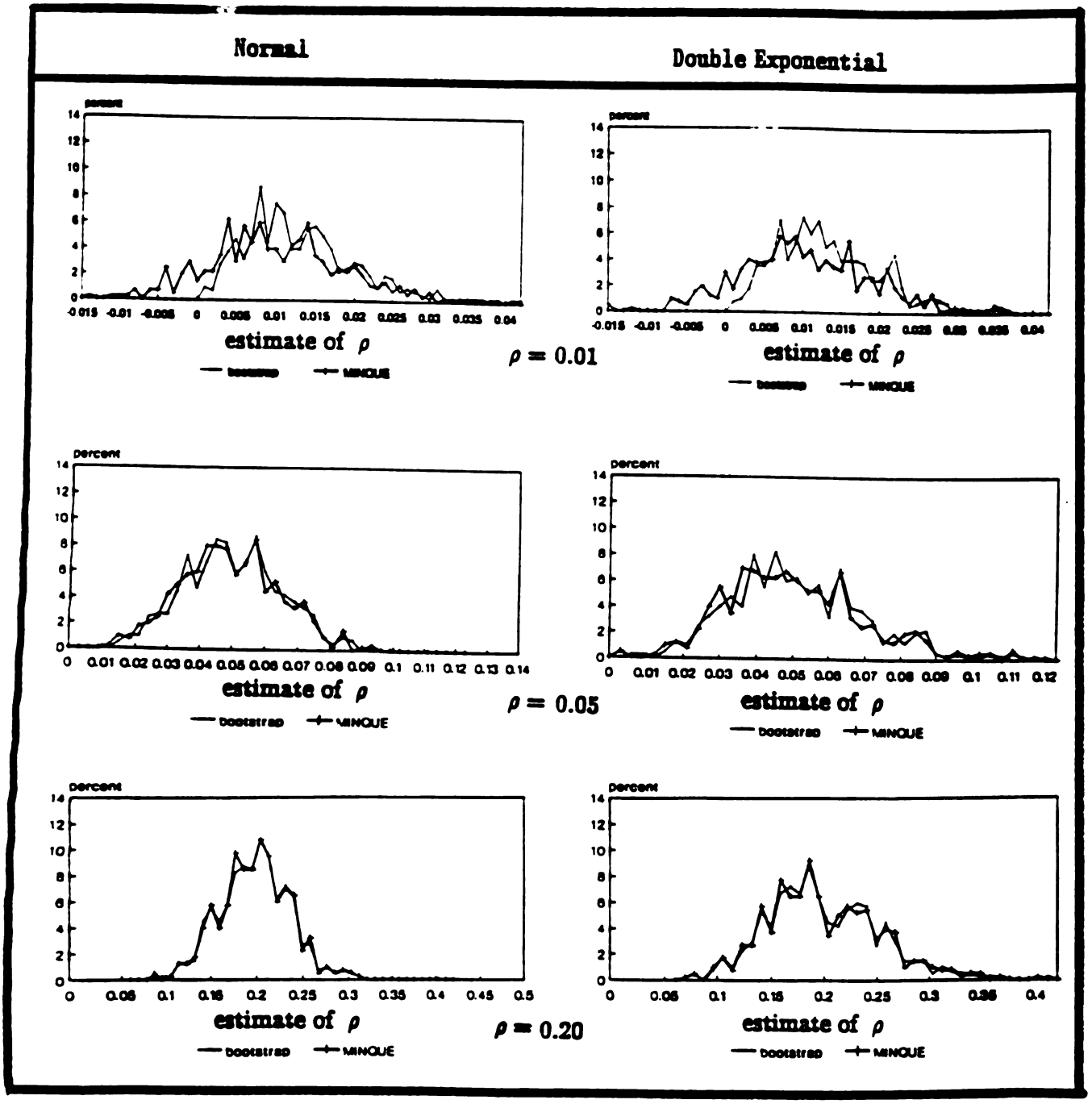
bootstrap. However, at this conditions of the intraclass correlation, the bootstrap ration R_1 was closer to 1.00 than R_2 under both normal and double exponential errors and sets of random effects of the model. It may also be important to note that, for the estimation of the parameter ρ, the bootstrap/MINQUE measure of relative efficiency at all three levels of the intraclass correlation was extremely close to 1.00 under both the normal and double exponential distribution.

Figure 6.3 shows the percentage polygons of the 400 bootstrap and MINQUE estimates of ρ under the normal and double exponential distributions at each of the three levels of the population intraclass correlation. At the 0.01 level of the intraclass correlation condition, though both MINQUE and the bootstrap percentage polygons were centered near the population parameter value of ρ, it is evident that a greater mass of observations were around the parameters value under the bootstrap frequent polygon than under the MINQUE polygon, for both normal and double exponential distributions. Thus, once again, the bootstrap method has been shown to be a more efficient estimator of ρ than MINQUE at the 0.01 level of the intraclass correlation condition.

Percentage polygons for the 400 MINQUE and bootstrap estimates under the normal and double exponential distributions at the 0.05 level of the intraclass correlation condition show that MINQUE and the bootstrap followed each other very closely. However, the percentage polygons under this condition of the intraclass correlation indicated that the values of both estimates were more variable under the double exponential than under the normal. The percentage polygons under the 0.20 level of the intraclass condition shows that the bootstrap and MINQUE followed each other even more closely than at the 0.05 intraclass

Figure 6.3

Percentage polygons for the MINQUE and bootstrap estimate of ρ over 400 trials under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.



correlation condition. Likewise, values of both estimates were more variable under double exponential than under the normal distributions.

The estimation results at the three levels of the intraclass correlation conditions indicate that the bootstrap is a more stable estimator of ρ, particularly at the 0.01 level of the intraclass correlation condition. However, nature and size of the tail of the distribution of the errors and sets of random effects equally influence the bootstrap and MINQUE in estimating ρ. Estimation tends to be less successful under a distribution with long and thick tails (like that of the double exponential) than under a less thick and short tailed distribution.

Fixed effects parameters of the model which included α_1, α_2, and α_3 for the three levels of the fixed factor and β, the coefficient of the covariate was estimated at the three levels of the intraclass correlation conditions under the normal and double exponential errors and sets of random effects of the model. Due to the fact that α_j for $j = 1,2,3$ are linearly dependent, estimation is only required for any two of α_j. For the purpose of the present dissertation, estimation results for α_1, α_3, and β are presented for each of the six design factor combinations.

Table 6.4 presents the summary statistics over the 400 trials for the ten estimable functions of $\hat{\alpha}_1$ and/or $\hat{\alpha}_1^*$ under the three levels of the intraclass correlation condition for the normal and double exponential distributions. Means and standard deviations presented in Table 6.4 shows that both MINQUE and the bootstrap very closely estimated the fixed effect parameter α_1 at all the three levels of the intraclass correlation condition under both normal and double exponential distributions. At the 0.01 level of the intraclass correlation, the bias for MINQUE and bootstrap were 0.0261 and 0.0264 respectively under the normal and

Table 6.4

Average and standard deviation of the functions of the estimates $\hat{\alpha}_1$ and/or $\hat{\alpha}_1^*$ under the normal and double exponential error and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.01	Bootstrap $\hat{\alpha}_1^*$	–5.00	–4.9736	0.9302	–5.0251	0.9073
	MINQUE, $\hat{\alpha}_1$	–5.00	–4.9739	0.9264	–5.0243	0.9032
	$\hat{BIAS}=\hat{\alpha}_1^*-\hat{\alpha}_1$		0.0003	0.0587	–0.0008	0.0619
	$D_1=\hat{\alpha}_1-\alpha_1$		0.0261	0.9264	–0.0243	0.9032
	$D_2=\hat{\alpha}_1^*-\alpha_1$		0.0264	0.9302	–0.0251	0.9073
	$R_1=\hat{\alpha}_1^*/\hat{\alpha}_1$		0.9999	0.0128	1.0000	0.0133
	$R_2=\hat{\alpha}_1/\alpha_1$		0.9948	0.1853	1.0049	0.1806
	$MSE1=(\hat{\alpha}_1-\alpha_1)^2$		0.8568	1.3327	0.8143	1.1918
	$MSE2=(\hat{\alpha}_1^*-\alpha_1)^2$		0.8638	1.3334	0.8217	1.2026
	Rel. Efficiency of		1.0082		1.0091	

Table 6.4 (continued)

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.05	Bootstrap $\hat{\alpha}_1^*$	–5.00	–5.0139	1.0867	–5.0211	1.0484
	MINQUE, $\hat{\alpha}_1$	–5.00	–5.0143	1.0889	–5.0221	1.0438
	BIAS=$\hat{\alpha}_1^*-\hat{\alpha}_1$		0.0004	0.0606	0.0010	0.0632
	$D_1=\hat{\alpha}_1-\alpha_1$		–0.0143	1.0889	–0.0221	1.0438
	$D_2=\hat{\alpha}_1^*-\alpha_1$		–0.0139	1.0867	–0.0211	1.0484
	$R_1=\hat{\alpha}_1^*/\hat{\alpha}_1$		1.0001	0.0128	0.9996	0.0138
	$R_2=\hat{\alpha}_1/\alpha_1$		1.0029	0.2178	1.0044	0.2088
	MSE=$(\hat{\alpha}_1-\alpha_1)^2$		1.1830	1.9324	1.0873	1.5678
	MSE=$(\hat{\alpha}_1^*-\alpha_1)^2$		1.1781	1.9236	1.0968	1.5833
	Rel. Efficiency@		0.9959		1.0087	
0.20	Bootstrap $\hat{\alpha}_1^*$	–5.00	–4.9529	1.6260	–5.0502	1.5745
	MINQUE, $\hat{\alpha}_1$	–5.00	–4.9522	1.6186	–5.0488	1.5736
	BIAS=$\hat{\alpha}_1-\hat{\alpha}_1$		–0.0007	0.0613	–0.0014	0.0660
	$D_1=\hat{\alpha}_1-\alpha_1$		0.0478	1.6186	–0.0488	1.5736
	$D_2=\hat{\alpha}_1^*-\alpha_1$		0.0471	1.6260	–0.0502	1.5745
	$R_1=\hat{\alpha}_1^*/\hat{\alpha}_1$		1.0000	0.0183	1.0001	0.0172
	$R_2=\hat{\alpha}_1/\alpha_1$		0.9904	0.3237	1.0098	0.3147
	MSE1=$(\hat{\alpha}_1-\alpha_1)^2$		2.6156	3.9830	2.4724	3.2204
	MSE2=$(\hat{\alpha}_1^*-\alpha_1)^2$		2.6394	4.0097	2.4754	3.2304
	Rel. Efficiency@		1.0091		1.0012	

–0.0243 and –0.0251 under double exponential. The bootstrap estimate of bias at this level was 0.0003 under the normal and –0.0008 under double exponential. At this level of the intraclass correlation, R_1 was observed at 0.9999 under the normal and 1.000 under the double exponential compared to R_2 which was 0.9948 under the normal and 1.0049 under double exponential. Relative efficiency for the two estimators was extremely close to 1.00 both under normal and double exponential.

For $\rho = 0.05$, the average of bootstrap and MINQUE estimates over the 400 trials were –5.0139 and –5.0221 under double exponential. Their respective biases were –0.0139 and –0.0143 under the normal and –0.0211 and –0.0221 under double exponential. Compared to the expected ratio of the estimates at 1.00, both MINQUE and the bootstrap were very close in estimated the ratio with $R_1 = 1.0001$ under the normal and $R_1 = 0.9996$ under double exponential while $R_2 = 1.0029$ under normal and $R_2 = 1.0044$ under double exponential.

Estimation of α_1 at the 0.20 level of the intraclass correlation was equally successful with R_1 being closer to 1.00 than R_2 under both normal and double exponential distributions. The bootstrap estimate of bias was lower under the normal than under double exponential. However, the bootstrap and MINQUE biases differed by no more than 0.007 under normal or double exponential distributions, and the measure of efficiency was extremely close to 1.00 under the normal and double exponential.

The effect of the intraclass correlation condition on the estimation of the functions of $\hat{\alpha}_1$ and/or $\hat{\alpha}_1^*$ was apparent in the MINQUE and bootstrap mean square errors. Both mean square errors tended to increase with increasing ρ under both normal and double exponential distributions. For instance, the bootstrap mean square errors were 0.8638, 1.1781, and 2.6394 for $\rho = 0.01$, 0.05 and 0.20

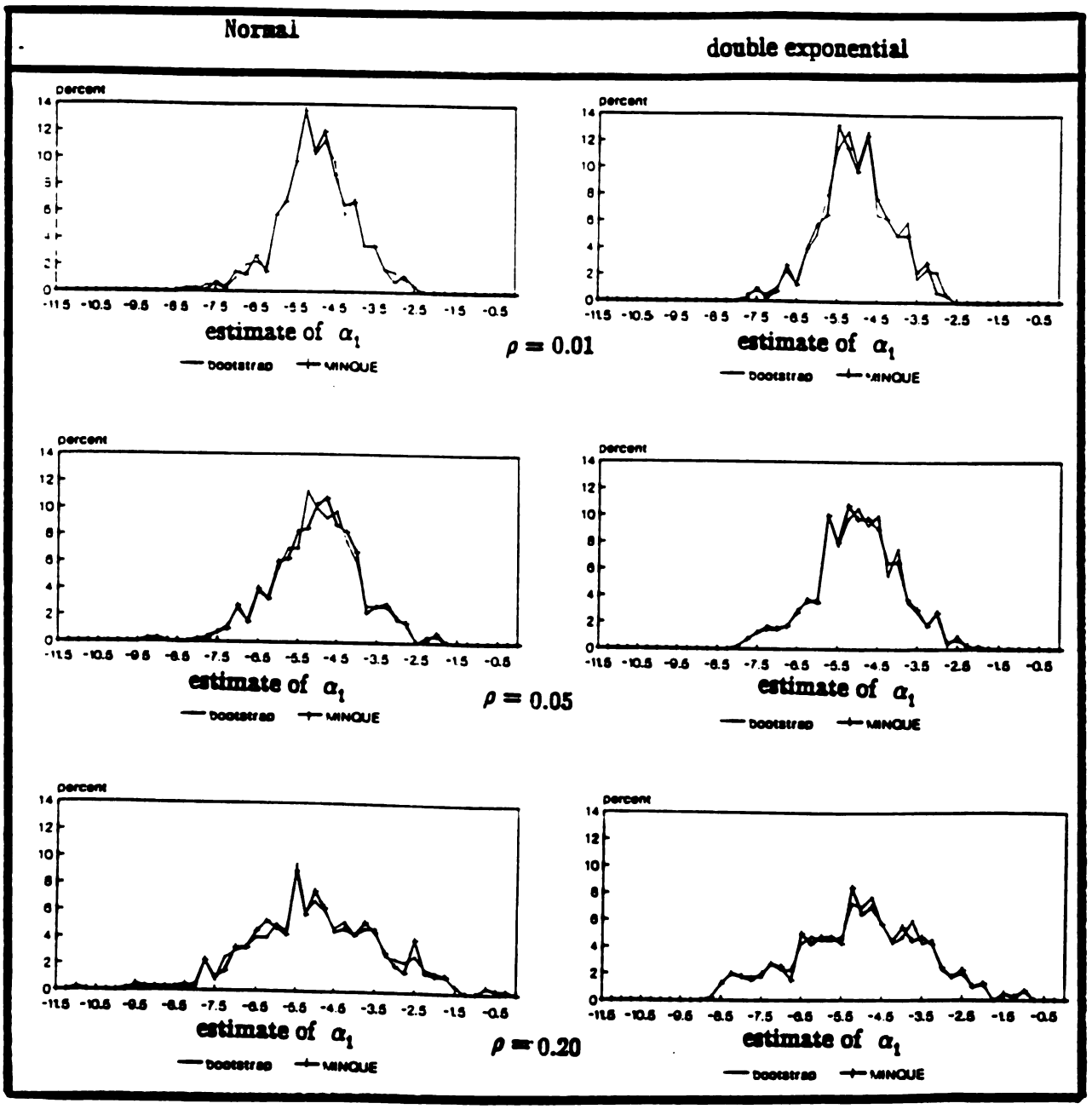
respectively under the normal and 0.8217, 1.0968, and 2.4754 for $\rho = 0.01, 0.05,$ and 0.20 respectively under double exponential. The mean square errors for the MINQUE estimates were quite close to those of the bootstrap at all levels of the intraclass correlation under both normal and double exponential. Under both the normal and double exponential, the bootstrap/MINQUE measure of relative efficiency was extremely close to 1.00 indicating that the bootstrap and MINQUE estimators very closely estimated the parameter α_1.

In general, therefore, results in Table 6.4 show that, both MINQUE and bootstrap very closely estimated the parameter α_1 at all levels of the intraclass correlation conditions under both normal and double exponential distributions. The ratio R_1 was consistently closer to 1.00 compared to R_2 at all the six design factor combinations; indicating a great deal of promise through the bootstrap method.

Figure 6.4 presents six percentage polygons of the 400 bootstrap and MINQUE estimates of α_1 under the normal and double exponential errors and sets of random effects at each of the three levels of the population intraclass correlation condition. From these charts it is clear that the bootstrap generally followed the MINQUE very closely at all levels of the intraclass correlation. The figures also showed no obvious difference in estimation between the normal and double exponential distributions. However, the highest spread of the estimates for both bootstrap and MINQUE were observed at the 0.20 level of the intraclass correlation followed by 0.05 level. The spread of both estimates was lowest at 0.01. Percentage polygons shown in Figure 6.4 therefore indicate that, though MINQUE and bootstrap do not differ in estimating α_1, both their ability to produce efficient (less variable) estimates depends on the level of the population intraclass correlation.

Figure 6.4

Percentage polygons for the MINQUE and bootstrap estimate of α_1 over 40 trials under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.



Both methods yield less efficient estimates when the population intraclass correlation is high. Their mean square errors increased at the same rate with increasing intraclass correlation.

Despite differences in variability of the estimates at different levels of the intraclass correlation, the percentage polygons indicate that the estimates were centered nearly at the same point. Estimates were expected to be centered at the true population parameters value which was set at –5.00. The results showed that all the six percentage polygons were centered no more than 0.05 away from the true parameter value.

Summary results for the bootstrap and MINQUE estimates of the parameter α_3 based on the ten estimable functions of $\hat{\alpha}_3$ and/or $\hat{\alpha}_3^*$ over the 400 trials are presented in Table 6.5. Summary results are presented for each of the three levels of the intraclass correlation conditions under both the normal and double exponential distribution of the errors and sets of random effects of the model. The true population parameter, α_3 was set at 3.00. MINQUE and bootstrap estimates are compared against the true population parameter value.

Averages and standard deviations over 400 trials presented in Table 6.5 shows that, both MINQUE and the bootstrap very closely estimated the parameter α_3. At the 0.01 level of the intraclass correlation, both MINQUE and bootstrap slightly overestimated the parameter α_3 under the normal and slightly underestimated α_3 under the double exponential errors and sets of random effects of the model. The biases for MINQUE and bootstrap were 0.0366 and 0.0362 respectively under the normal and –0.0123 and –0.0101 under double exponential. The bootstrap estimate of bias was observed at –0.0004 under the normal and 0.0021 under double exponential. The ratio R_1 was surprisingly close to 1.00 under

Table 6.5

Average and standard deviation of the functions of the estimates $\hat{\alpha}_3$ and/or $\hat{\alpha}_3^*$ under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.01	Bootstrap, $\hat{\alpha}_3^*$	3.00	3.0362	0.9030	2.9878	0.9053
	MINQUE, $\hat{\alpha}_3$	3.00	3.0366	0.9029	2.9899	0.9109
	$\hat{BIAS}=\hat{\alpha}_3^*-\hat{\alpha}_3$		−0.0004	0.05566	0.0021	0.0625
	$D_1=\hat{\alpha}_3-\alpha_3$		0.0366	0.9029	−0.0123	0.9053
	$D_2=\hat{\alpha}_3^*-\alpha_3$		0.0362	0.9030	−0.0101	0.9109
	$R_1=\hat{\alpha}_3^*/\hat{\alpha}_3$		1.0005	0.0228	1.0016	0.0357
	$R_2=\hat{\alpha}_3/\alpha_3$		1.0122	0.3010	0.9959	0.3018
	$MSE1=(\hat{\alpha}_3-\alpha_3)^2$		0.8145	1.1035	0.8176	1.2312
	$MSE2=(\hat{\alpha}_3^*-\alpha_3)^2$		0.8146	1.0961	0.8277	1.2530
	Rel. Efficiency@		1.0001		1.0124	

Table 6.5 (continued)

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.05	Bootstrap, $\hat{\alpha}_3^*$	3.00	3.0380	1.0368	3.0045	1.0326
	MINQUE, $\hat{\alpha}_3$	3.00	3.0349	1.0367	3.0002	1.0278
	$\hat{BIAS}=\hat{\alpha}_3^*-\hat{\alpha}_3$		0.0031	0.0582	0.0043	0.0630
	$D_1=\hat{\alpha}_3^*-\alpha_3$		0.0349	1.0366	0.0002	1.0278
	$D_2=\hat{\alpha}_3^*-\alpha_3$		0.0380	1.0368	0.0045	1.0326
	$R_1=\hat{\alpha}_3^*/\hat{\alpha}_3$		1.0015	0.0264	1.0003	0.0415
	$R_2=\hat{\alpha}_3^*/\alpha_3$		1.0116	0.3455	1.0001	0.3426
	$MSE1=(\hat{\alpha}_3-\alpha_3)^2$		1.0732	1.4720	1.0537	1.5478
	$MSE2=(\hat{\alpha}_3^*-\alpha_3)^2$		1.0738	1.4677	1.0635	1.5753
	Rel. Efficiency@		1.0006		1.0093	
0.20	Bootstrap, $\hat{\alpha}_3^*$	3.00	3.0379	1.4153	3.0711	1.4837
	MINQUE, $\hat{\alpha}_3$	3.00	3.0380	1.4134	3.0688	1.4777
	$\hat{BIAS}=\hat{\alpha}_3^*-\hat{\alpha}_3$		−0.0002	0.00610	0.0023	0.0646
	$D_1=\hat{\alpha}_3-\alpha_3$		0.0380	1.4134	0.0688	1.4777
	$D_2=\hat{\alpha}_3^*-\alpha_3$		0.0379	1.4153	0.0711	1.4837
	$R_1=\hat{\alpha}_3^*/\hat{\alpha}_3$		0.9979	0.0529	1.0005	0.1067
	$R_2=\hat{\alpha}_3/\alpha_3$		1.0127	0.4711	1.0229	0.4926
	$MSE1=(\hat{\alpha}_3-\alpha_3)^2$		1.9940	2.9561	2.1828	3.0902
	$MSE2=(\hat{\alpha}_3^*-\alpha_3)^2$		1.9995	2.9688	2.2009	3.1362
	Rel. Efficiency@			1.0028		1.0083

@ Rel. Efficiency = MSE2/MSE1

both the normal and double exponential errors and sets of random effects of the model. The usual MINQUE ratio, R_2 was not as close to 1.00 as R_1, indicating that the bootstrap more successfully estimated the ratio than MINQUE.

A more successful estimation of the parameter α_3 was achieved by both the MINQUE and bootstrap under the double exponential errors at the 0.05 level of the intraclass correlation condition. At this level, the average values of MINQUE and bootstrap were 3.0349 and 3.0380 respectively under the normal and 3.0002 and 3.0045 respectively under double exponential distributions. The bootstrap and MINQUE bias were 0.0380 and 0.0349 under the normal and 0.0045 and 0.0002 respectively under double exponential distributions. The bootstrap estimate of bias was 0.0031 under the normal and 0.0043 under the double exponential. At this level of the intraclass correlation also, the ratio R_1 was closer to 1.00 than R_2 both under the normal and double exponential errors and sets of random effects of the model. However, at the 0.05 level of the intraclass correlation condition, both MINQUE and bootstrap mean square errors were greater than at the 0.01 level of the intraclass correlation.

At the 0.20 level of the intraclass correlation, both MINQUE and bootstrap slightly overestimated the parameter α_3 under the normal and double exponential distribution of errors and sets of random effects of the model. However, both biases were greater under double exponential distribution than under the normal. The bootstrap estimate of bias was no more than 0.0025 under both distributions. On average, R_1 was closer to 1.00 than R_2 under both the normal and double exponential. The bootstrap and MINQUE mean square errors were near 2.00 at this level of the intraclass correlation compared to about 1.05 at the 0.05 level and about 0.80 at the 0.01 level of the intraclass correlation. Thus, both mean square errors

seemed to increase with increasing level of the intraclass correlation condition. The bootstrap/MINQUE measure of relative efficiency at all three levels of the intraclass correlation were very close to 1.00 under both the normal and double exponential distribution.

Figure 6.5 shows the percentage polygons of the 400 bootstrap and MINQUE estimates of α_3 for each of the six design factor combinations. Separate frequent polygons are presented for the normal and double exponential distributions at each level of the intraclass correlation condition. From these charts, it is shown that the bootstrap followed the MINQUE quite closely such that it was difficulty to distinguish the two at certain points. All the six charts were centered near the true parameter value of α_3 which was set at 3.00.

Though the frequency polygons showed no variation by the distribution of errors and sets of random effects, the spread of the estimates seems to vary by level of the intraclass correlation condition. The highest spread was observed at the 0.20 level of the intraclass correlation while the lowest spread was seen at the 0.01 level.

The other fixed effects parameter of the model which was examined in the study was β, the coefficient of the covariate. The true parameter value was set at 1.00. As in the other parameters of the model, the bootstrap and MINQUE estimates of β were calculated at each of the three levels of the intraclass correlation under both normal and double exponential errors and sets of random effects of the model. Summary results for the bootstrap and MINQUE estimates over the 400 trials for the six design factor combinations are presented in Table 6.6. Here, averages and standard deviations of ten function of $\hat{\beta}$ and $\hat{\beta}^*$ are presented

Figure 6.5

Percentage polygons for the MINQUE and bootstrap estimate of α_3 over 40 trials under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

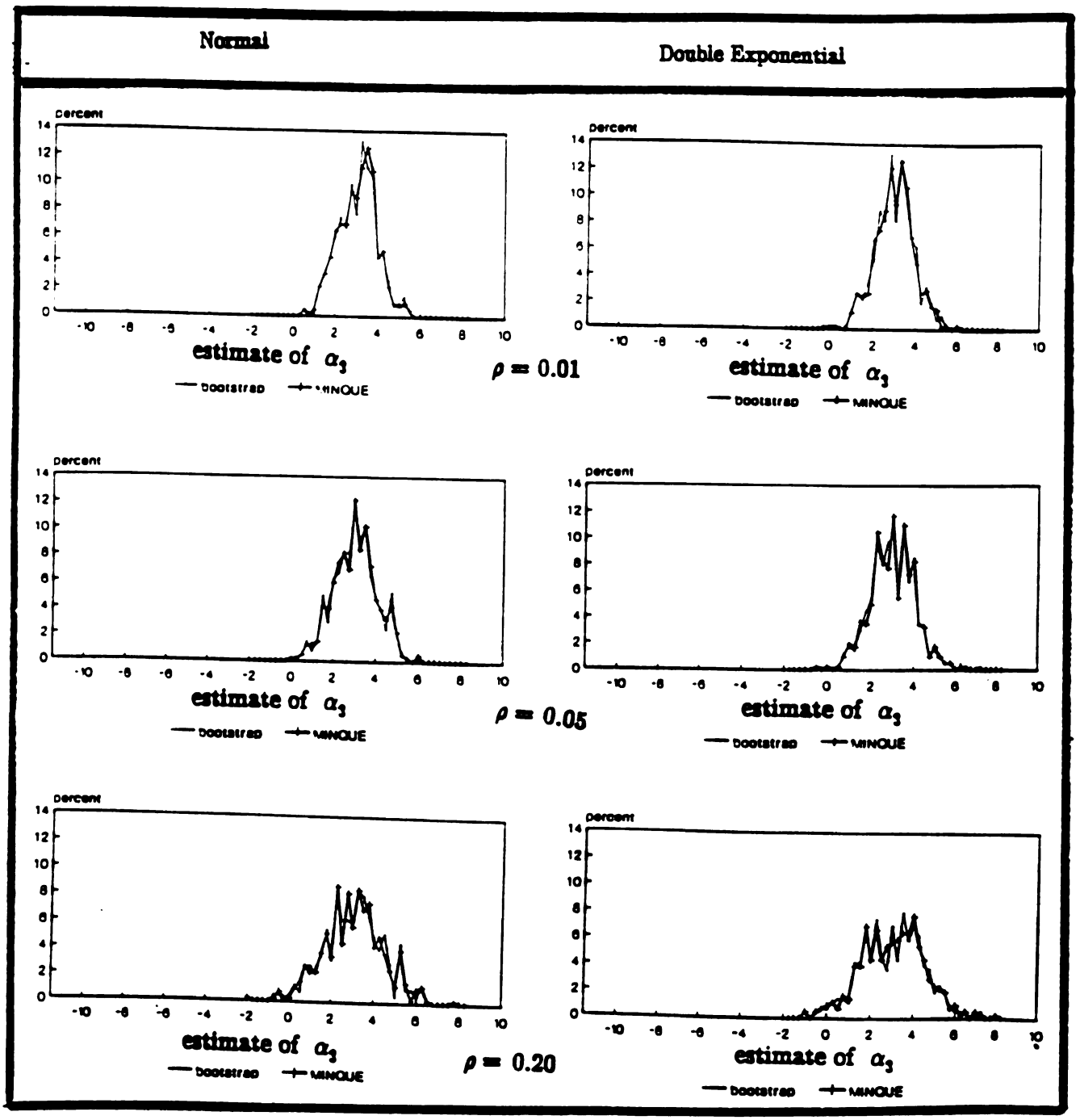


Table 6.6

Average and standard deviation of the functions of the estimates $\hat{\beta}$ and/or $\hat{\beta}^*$ under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.01	Bootstrap, $\hat{\beta}^*$	1.00	1.0000	0.0117	1.0005	0.0123
	MINQUE, $\hat{\beta}$	1.00	1.000	0.0117	1.0005	0.0122
	BIAS=$\hat{\beta}^*-\hat{\beta}$		–0.0000	0.0008	–0.0000	–.0009
	$D_1=\hat{\beta}-\beta$		–0.0000	0.0117	0.0005	0.0122
	$D_2=\hat{\beta}^*-\beta$		–0.0000	0.0117	0.0005	0.0122
	$R_1=\hat{\beta}^*/\hat{\beta}$		1.0000	0.0008	1.0000	0.0009
	$R_2=\hat{\beta}/\beta$		1.0000	0.0117	1.0005	0.0122
	MSE1=$(\hat{\beta}-\beta)^2$		0.0001	0.0002	0.0001	0.0002
	MSE2=$(\hat{\beta}^*-\beta)^2$		0.0001	0.0002	0.0002	0.0002
	Rel. Efficiency@		1.0000		2.0000	

Table 6.6 (continued)

			Normal		Double Exponential	
Value of ρ	Estimate	Par.Value	Average	S.D.	Average	S.D.
0.05	Bootstrap, $\hat{\beta}^*$	1.00	0.9999	0.0125	1.0005	0.0123
	MINQUE, $\hat{\beta}$	1.00	0.9999	0.0124	1.0005	0.0123
	$\hat{BIAS}=\hat{\beta}-\beta$		0.0000	0.0009	–0.0001	0.0009
	$D_1=\hat{\beta}-\beta$		–0.0001	0.0124	0.0005	0.0123
	$D_2=\hat{\beta}^*-\beta$		–0.0001	0.0125	0.0005	0.0123
	$R_1=\hat{\beta}^*/\hat{\beta}$		1.0000	0.0009	0.9999	0.0009
	$R_2=\hat{\beta}/\beta$		0.9999	0.0124	1.0005	0.0123
	$MSE1=(\hat{\beta}-\beta)^2$		0.0002	0.0002	0.0002	0.0002
	$MSE2=(\hat{\beta}^*-\beta)^2$		0.0002	0.0002	0.0002	0.0002
	Rel. Efficiency@		1.0000		1.0000	
0.20	Bootstrap, $\hat{\beta}^*$	1.00	0.9997	0.0121	1.0005	0.0128
	MINQUE, $\hat{\beta}$	1.00	0.9997	0.0120	1.0005	0.0127
	$\hat{BIAS}=\hat{\beta}^*-\hat{\beta}$		0.0000	0.0009	0.0000	0.0009
	$D_1=\hat{\beta}-\beta$		–0.0003	0.0120	0.0005	0.0127
	$D_2=\hat{\beta}^*-\beta$		–0.0003	0.0121	0.0005	0.0128
	$R_1=\hat{\beta}^*/\hat{\beta}$		1.0000	0.0009	1.0000	0.0009
	$R_2=\hat{\beta}/\beta$		0.9997	0.0120	1.0005	0.0127
	$MSE1=(\hat{\beta}-\beta)^2$		0.0001	0.0002	0.0002	0.0002
	$MSE2=(\hat{\beta}^*-\beta)^2$		0.0001	0.0002	0.0002	0.0003
	Rel. Efficiency@		1.0000		1.0000	

@ Rel. Efficiency = MSE2/MSE1

at each of the three levels of the intraclass correlation by each of the distribution of the errors and sets of random effects of the model.

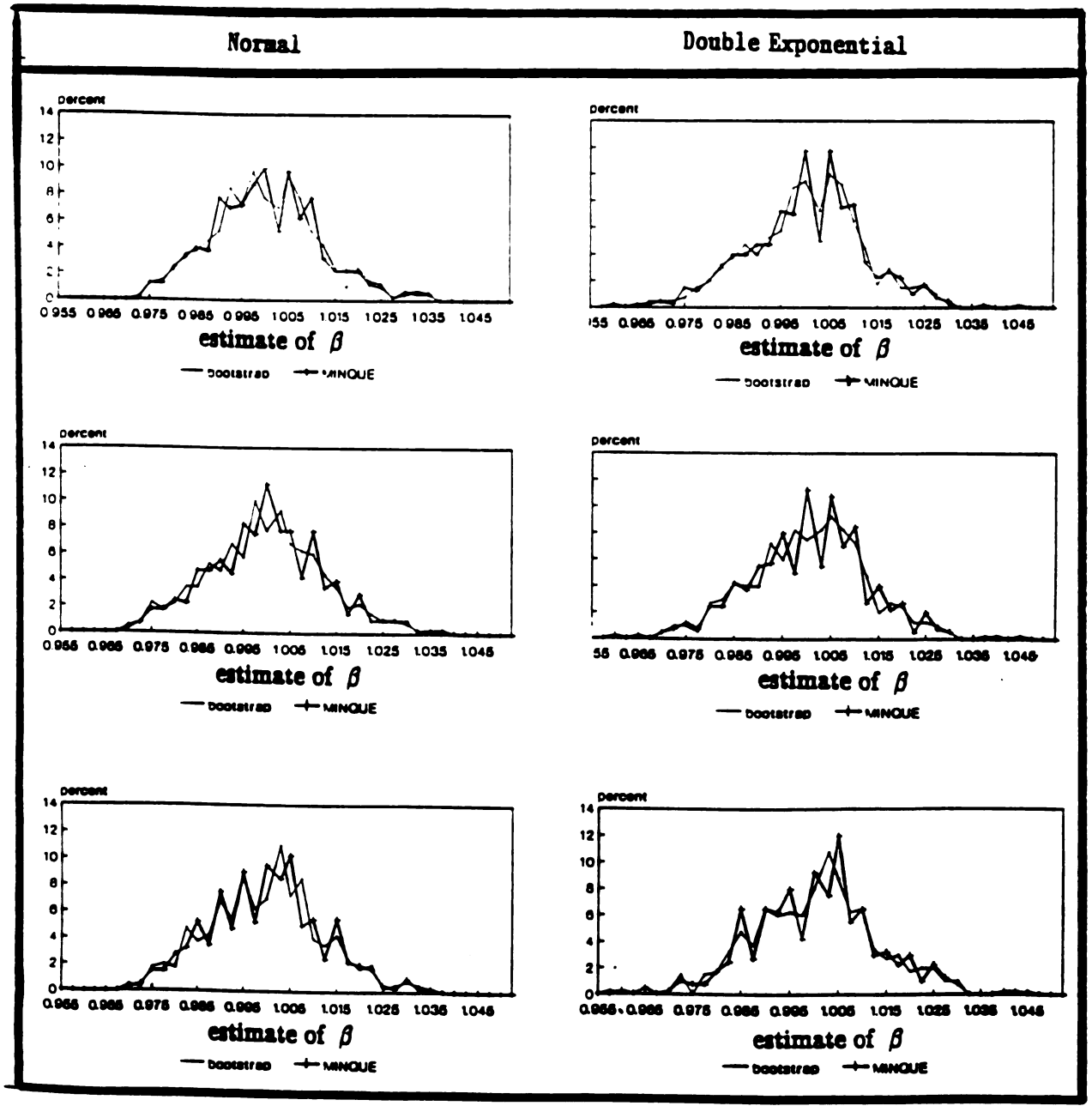
From Table 6.6 it is shown that the average values of the bootstrap and MINQUE were extremely close to the true parameter value regardless of the level of the intraclass correlation or distribution of the errors and sets of random effects of the model. At all levels of the intraclass correlation condition, the bootstrap and MINQUE biases were never greater than 0.0005 and the bootstrap estimate of bias was perfectly nil under both normal and double exponential distributions.

The average of the ratios R_1 was almost always equal to 1.00 at all levels of the intraclass correlation for both normal and double exponential errors and sets of random effects of the model. However, the average of the ratios, R_2 slightly differed from 1.00 for some design factor combinations. The bootstrap and MINQUE mean square errors were surprisingly small at all levels of the intraclass correlation for both normal and double exponential distributions. Thus, based on these summary statistics, it is clear that the parameter β was very successfully estimated by both MINQUE and the bootstrap regardless of the level of the intraclass correlation condition and the distribution of the errors and sets of random effects of the model. In terms of their relative accuracy, neither method (MINQUE or Bootstrap) was superior to the other. Their measure of relative efficiency was extremely close to 1.00 at all levels of the intraclass correlation particularly under the normal distribution.

Figure 6.6 is a display of the percentage polygons of the 400 bootstrap and MINQUE estimates of β under the normal and double exponential errors and sets of random effects at each of the three levels of the population intraclass correlation condition. From these charts, it is apparent that the bootstrap on average followed

Figure 6.6

Percentage polygons for the MINQUE and bootstrap estimate of β over 400 trials under the normal and double exponential errors and sets of random effects for $\rho = 0.01$, 0.05, and 0.20.



the MINQUE quite closely at all levels of the intraclass correlation condition. The percentage polygons also showed no obvious difference in estimation between the normal and double exponential distributions. All the six figures were centered extremely close to 1.00 as expected.

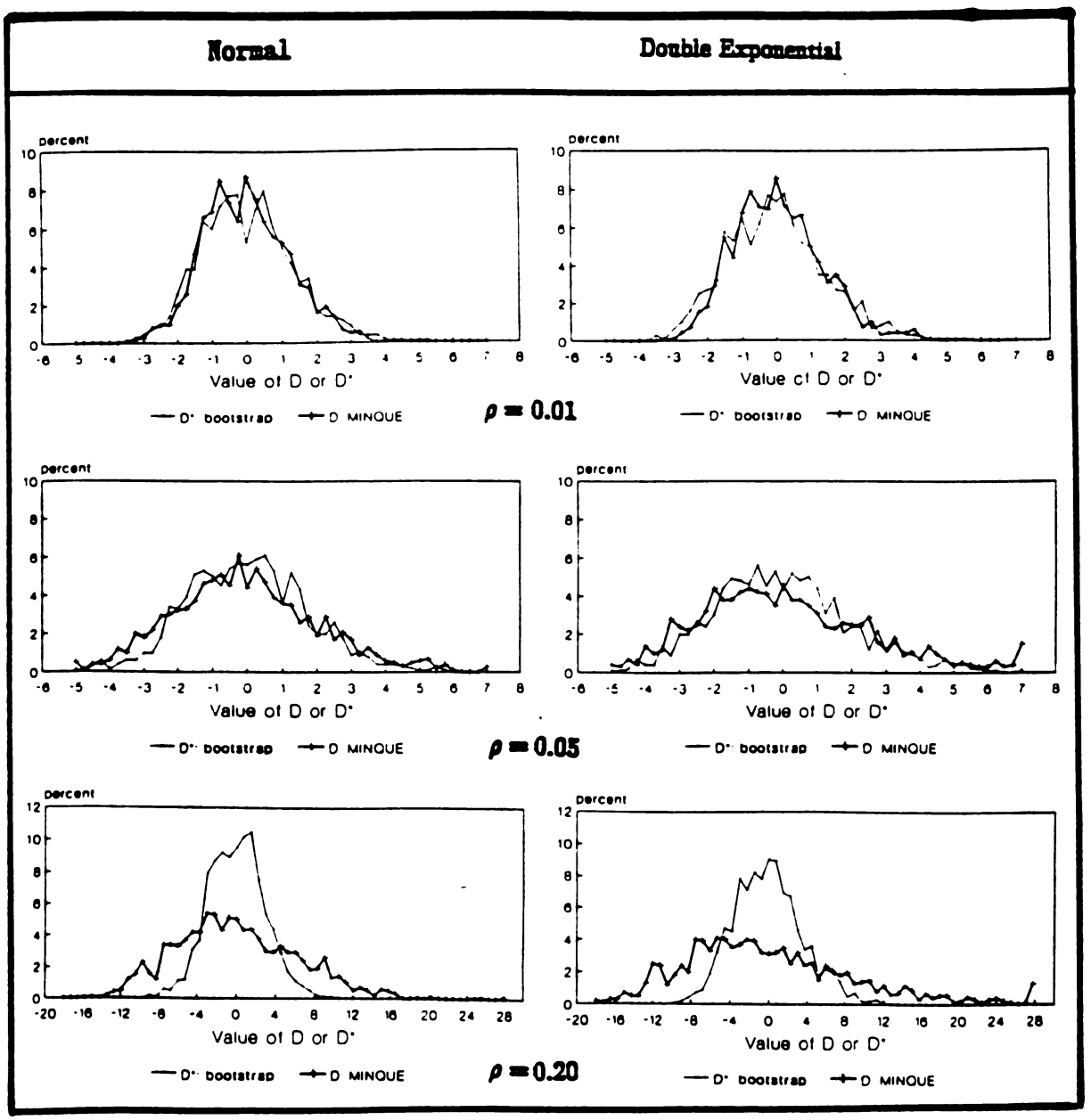
Results of Bootstrap Confidence Intervals

The bootstrap procedure for constructing confidence intervals is perhaps one of the most significant accomplishments of the bootstrap method. The procedure can be applied to even more complicated problems involving statistics whose sampling distributions cannot be determine analytically. Derivation of the bootstrap method for the confidence interval is based on the following assumption. For an estimator, $\hat{\theta}$ of the parameter θ, let $D = \hat{\theta}-\theta$. Define $D^* = \hat{\theta}^*-\hat{\theta}$ to be the bootstrap quantity observed at each bootstrap replication. The bootstrap distribution of D^* estimates the unknown distribution of D. As an illustration, percentage polygons based on 1000 repetitions to demonstrate the relationship between the distribution of functions D and D^* of τ^2, $\hat{\tau}^2$ and/or $\hat{\tau}^{2*}$ at 0.01, 0.05, and 0.20 intraclass correlation conditions are presented in Figure 6.7. But it is important for readers to be reminded that, while D was derived from 1000 independent samples drawn from a population with predetermined parameters, D^* was derived from 1000 repeated resampling drawn from one such sample with replacement.

If the distribution of D were known, then the $(1-\alpha)100\%$ confidence interval can be defined using real values D_L and D_U by the probability statement,

Figure 6.7

Percentage polygons for the relationship between the distribtion of the function $D = \hat{\tau}^2 - \tau^2$ and $D^* = \hat{\tau}^{2*} - \hat{\tau}^2$ for $\rho = 0.01$, 0.05, and 0.20.



$$P(D_L \leq D \leq D_U) \simeq 1 - \alpha$$

or

$$P(D_L \leq \hat{\theta} - \theta \leq D_U) \simeq 1 - \alpha$$

which can be written as,

$$(6.1) \qquad P(\hat{\theta} - D_U \leq \theta \leq \hat{\theta} - D_L) \simeq 1 - \alpha.$$

Since D_U and D_L are not observable the probability statement in Equation 6.1 is estimated by the bootstrap probability statement,

$$(6.2) \qquad P(\hat{\theta} - D_U^* \leq \theta \leq \hat{\theta} - D_L^*) \simeq 1 - \alpha$$

where D_U^* and D_L^* are bootstrap versions of D_U and D_L respectively computed from bootstrap samples. Equation 6.2 gives the bootstrap $(1-\alpha)100\%$ confidence interval via the percentile method. The procedure is highly flexible and can be applied to complicated problems in a wide range of situations, where classical methods may fail to be useful. Indeed, this was one of the aspects of the present study where the bootstrap delivers while the MINQUE does not.

In the present study, 90 and 95 percent bootstrap confidence intervals were constructed for each of the six parameters of the mixed model. The confidence intervals were constructed for all six design factor combinations (cells a through f in Table 4.1).

Table 6.7 presents the averages and standard deviations for the 90 and 95 percent confidence limits based on the bootstrap over the 400 trials at the 0.01 level of the intraclass correlation condition. The summary statistics for the lower confidence limit (L.C.L.), upper confidence limit (U.C.L.) and the width of the confidence interval are presented both under the normal and double exponential distributions.

Table 6.7

Averages and standard deviations of the bootstrap confidence limits and the width of the confidence intervals about the six parameters of the model under the normal and double exponential for $\rho = 0.01$.

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
$\tau^2(a,b)$	1.00	95% C.I:	L.C.L.	−1.7015	0.7953	−1.8480	0.8505
			U.C.L.	3.4778	1.1100	3.5388	1.1171
			Width	5.1793	0.5601	5.3868	0.6503
		90% C.I:	L.C.L.	−1.1889	0.7853	−1.2793	0.8324
			U.C.L.	3.1248	1.0733	3.1941	1.0889
			Width	4.3137	0.4677	4.4735	0.5417
$\sigma^2(a,b)$	100.00	95% C.I:	L.C.L.	92.7051	3.5458	88.8798	5.3956
			U.C.L.	106.8318	4.0241	110.4088	7.1031
			Width	14.1267	1.1860	21.5290	2.8913
		90% C.I:	L.C.L.	93.9623	3.5644	90.8313	5.4802
			U.C.	105.7354	3.9994	108.7660	6.9171
			Width	11.7730	0.9369	17.9348	2.2998

Table 6.7 (continued)

Parameter	Par.Value	Interval Estimate		Normal Average	Normal S.D.	Double Exponential Average	Double Exponential S.D.
ρ(a,b)	0.01	95% C.I:	L.C.L.	–0.0162	0.0081	–0.0167	0.0084
			U.C.L.	0.0347	0.0108	0.0355	0.0110
			Width	0.0508	0.0048	0.0523	0.0050
		90% C.I:	L.C.L.	–0.0113	0.0079	–0.0116	0.0082
			U.C.L.	0.0311	0.0103	0.0320	0.0107
			Width	0.0424	0.0039	0.0436	0.0041
α_1(a,b)	–5.00	95% C.I:	L.C.L.	–6.6875	0.9412	–6.7399	0.9009
			U.C.L.	–3.2586	0.9366	–3.3184	0.9316
			Width	3.4289	0.2391	3.4216	0.2669
		90% C.I:	L.C.L.	–6.4109	0.9348	–6.4495	0.8991
			U.C.L.	–3.5447	0.9318	–3.6082	0.9179
			Width	2.8662	0.1860	2.8413	0.2103

Table 6.7 (continued)

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
α_3(a,b)	3.00	95% C.I:	L.C.L.	1.3659	0.9143	1.3039	0.9057
			U.C.L.	4.7217	0.9216	4.6672	0.9371
			Width	3.3558	0.2547	3.3633	0.2684
		90% C.I:	L.C.L.	1.6407	0.9100	1.5883	0.9003
			U.C.L.	4.4418	0.9130	4.3855	0.9271
			Width	2.8011	0.1893	2.7972	0.2060
β(a,b)	1.00	95% C.I:	L.C.L.	0.9767	0.0119	0.9771	0.0126
			U.C.L.	1.0240	0.0120	1.0239	0.0122
			Width	0.0467	0.0034	0.0467	0.0037
		90% C.I:	L.C.L.	0.9806	0.0118	0.9811	0.0126
			U.C.L.	1.0195	0.0118	1.0199	0.0122
			Width	0.0389	0.0027	0.0389	0.0029

From Table 6.7 it is shown that bootstrap 95% confidence intervals about the parameter τ^2 at the 0.01 level of the intraclass correlation had a width of 5.1793 and 5.3868 under the normal and double exponential respectively. The average width of the 90% confidence interval was 4.3137 and 4.4735 under the normal and double exponential respectively. Corresponding standard deviation to these averages show clearly how precise these intervals were.

Averages and standard deviations of the confidence limits and widths of confidence intervals about the parameter σ_e^2 also showed a fairly precise bootstrap interval estimation process. The average width of the confidence intervals were fairly low, particularly under the normal distribution. Standard deviations corresponding to these averages were 0.9369 under the normal and 2.2998 under double exponential.

Since the bootstrap interval estimation process about the parameters τ^2 and σ_e^2 both of which are component of ρ (see Equation 4.3) was rather unsuccessful, the results showed an equally successful bootstrap interval estimation process about the parameter ρ. The average width of the 95% confidence interval was 0.0508 under the normal and 0.0523 under the double exponential distribution.

At this level of the intraclass correlation condition, the highest success of the bootstrap confidence interval estimation procedure was achieved about the parameter β, the coefficient of the covariate of the model. For this parameter, very precise bootstrap confidence intervals were obtained both under the normal and double exponential distributions. For instance, the average width of the 95% bootstrap confidence interval was 0.0467 under the normal and 0.0467 under the double exponential. The standard deviations corresponding to these average widths

were 0.0034 and 0.0037 under the normal and double exponential respectively. With these results, it is apparent that, regardless of the distribution of the errors and sets of random effects of the model, the bootstrap confidence intervals about the parameter β were extremely precise.

Table 6.8 shows the summary statistics for 90% and 95% bootstrap confidence limits over the 400 trials at the 0.05 level of the intraclass correlation condition. Averages and standard deviations for the lower (L.C.L.) and upper (U.C.L.) confidence limits and the width of the confidence interval are presented under the normal and double exponential distributions.

Summary statistics in Table 6.8 shows that, the bootstrap confidence interval about the parameter τ^2 at the 0.05 level of the intraclass correlation were by far more successful than the same intervals at the 0.01 level of the intraclass correlation condition. The average widths were much smaller and less variable. Summary statistics for bootstrap confidence intervals about the parameters σ^2, α_1, α_3, and β at the 0.05 and 0.01 levels of the intraclass correlation showed a more precise interval estimation under both the normal and double exponential. For the parameters τ^2 and ρ, the bootstrap confidence interval estimation procedure at the 0.05 level of the intraclass correlation was equally successful as at the 0.01 level of the intraclass correlation condition. For instance, compared to the average width of the 95% confidence interval about ρ of 0.0508 when $\rho = 0.01$, the same average was 0.0615 when $\rho = 0.05$ under the normal distribution. Similar low differences between the two levels of the intraclass correlation in the width of the bootstrap confidence intervals about the parameter ρ were observed under the double exponential distribution.

Table 6.8

Averages and standard deviations of the bootstrap confidence limits and the width of the confidence intervals about the six parameters of the model under the normal and double exponential for $\rho = 0.05$.

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
$\tau^2(c,d)$	5.26	95% C.I:	L.C.L.	1.6878	1.4390	1.7302	1.8989
			U.C.L.	8.4066	1.9646	8.6379	2.5444
			Width	6.7188	0.7833	6.9077	0.9608
		90% C.I:	L.C.L.	2.3419	1.4577	2.4302	1.9352
			U.C.L.	7.9180	1.9092	8.1630	2.4868
			Width	5.5761	0.6189	5.7328	0.7734
$\sigma^2(c,d)$	100.00	95% C.I:	L.C.L.	92.5692	3.4670	88.9170	5.3877
			U.C.L.	106.6525	3.8824	110.4178	7.0169
			Width	14.0832	1.2034	2.7504	7.5647
		90% C.I:	L.C.L.	93.7975	3.4149	90.8487	5.4584
			U.C.L.	105.5263	3.8263	108.7716	6.8586
			Width	11.7289	0.9328	17.9229	2.2340

Table 6.8 (continued)

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
ρ(c,d)	0.05	95% C.I:	L.C.L.	0.0179	0.0139	0.0191	0.0181
			U.C.L.	0.0794	0.0173	0.0811	0.0220
			Width	0.0615	0.0057	0.0620	0.0065
		90% C.I:	L.C.L.	0.0235	0.0141	0.0249	0.0183
			U.C.L.	0.0748	0.0169	0.0766	0.0217
			Width	0.0513	0.0045	0.0517	0.0052
α_1(c,d)	–5.00	95% C.I:	L.C.L.	–6.7327	1.1207	–6.7432	1.0410
			U.C.L.	–3.2951	1.0962	–3.3127	1.0705
			Width	3.4376	0.2559	3.4305	0.2616
		90% C.I:	L.C.L.	–6.4476	1.1031	–6.4488	1.0369
			U.C.L.	–3.5808	1.0909	–3.6006	1.0542
			Width	2.8669	0.1906	2.8482	0.2109

Table 6.8 (continued)

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
α_3(c,d)	3.00	95% C.I:	L.C.L.	1.3547	1.0546	1.3165	1.0330
			U.C.L.	4.7268	1.0481	4.6813	1.0602
			Width	3.3721	0.2653	3.3648	0.2764
		90% C.I:	L.C.L.	1.6298	1.0526	1.6000	1.0197
			U.C.L.	4.4483	1.0425	4.3985	1.0469
			Width	2.8185	0.1987	2.7985	0.2080
β(c,d)	1.00	95% C.I:	L.C.L.	0.9763	0.0124	0.9772	0.0127
			U.C.L.	1.0233	0.0126	1.0239	0.0124
			Width	0.0470	0.0033	0.0467	0.0037
		90% C.I:	L.C.L.	0.9803	0.0124	0.9811	0.0125
			U.C.L.	1.0193	0.0125	1.0199	0.0124
			Width	0.0390	0.0025	0.0388	0.0028

Summary results for the 90% and 95% confidence intervals for the 400 trials at the 0.20 level of the intraclass correlation are presented in Table 6.9. Averages for the lower (L.C.L.), upper (U.C.L.) confidence limits and the width of the confidence intervals are presented under both the normal and double exponential errors and sets of random effects of the model.

At this level of the intraclass correlation, these summary statistics showed slightly wider bootstrap confidence procedure about the parameters τ^2 and σ_e^2 than at the other levels of the intraclass correlation. No differences in the level of success of the method were noticed in estimating the confidence intervals about the other parameters of the model among different levels of the intraclass correlation condition. For the fixed effects parameters, the bootstrap procedure for confidence interval was also always successful at all levels of the intraclass correlation condition.

These results revealed an important feature of the bootstrap procedure for confidence intervals about the parameters τ^2 and σ_e^2. The success of the procedure depends on the level of the intraclass correlation. When the population intraclass correlation is small, the bootstrap procedure for confidence intervals using the percentile method about the fixed and random effects parameters of the model is quite precise. At high values of the intraclass correlation condition, the bootstrap interval estimation procedure about the random effects parameters is slightly less precise. However, regardless of the level of the intraclass correlation, the bootstrap method for the confidence interval about the fixed effects parameters of the model $(\alpha_1, \alpha_3, \beta)$ seemed to be a remarkable success.

Table 6.9

Averages and standard deviations of the bootstrap confidence limits and the width of the confidence intervals about the six parameters of the model under the normal and double exponential for $\rho = 0.20$.

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
$\tau^2(e,f)$	25.00	95% C.I:	L.C.L.	18.9969	5.3161	19.4793	8.0518
			U.C.L.	30.4033	6.4824	30.9582	9.6290
			Width	11.4065	1.4426	11.4789	1.9123
		90% C.I:	L.C.L.	20.0441	5.4238	20.5034	8.1698
			U.C.L.	29.5548	6.3752	30.1033	9.5247
			Width	9.5107	1.1626	9.5999	1.6083
$\sigma^2(e,f)$	100.00	95% C.I:	L.C.L.	92.9721	3.7505	88.7349	5.2749
			U.C.L.	107.0575	4.2087	110.1482	6.8213
			Width	14.0853	1.1690	21.4133	2.7346
		90% C.I:	L.C.L.	94.2014	3.7305	90.6651	5.3502
			U.C.	105.9621	4.1663	108.4825	6.6614
			Width	11.7607	0.9424	17.8175	2.2111

Table 6.9 (continued)

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
ρ(e,f)	0.20	95% C.I:	L.C.L.	0.1589	0.0380	0.1598	0.0526
			U.C.L.	0.2360	0.0395	0.2386	0.0552
			Width	0.0771	0.0055	0.0788	0.0073
		90% C.I:	L.C.L.	0.1656	0.0378	0.1666	0.0526
			U.C.L.	0.2299	0.0393	0.2325	0.0550
			Width	0.0643	0.0045	0.0659	0.0058
α_1(e,f)	–5.00	95% C.I:	L.C.L.	–6.7019	1.6240	–6.7908	1.5732
			U.C.L.	–3.1895	1.6174	–3.3161	1.5938
			Width	3.5124	0.2596	3.4746	0.2583
		90% C.I:	L.C.L.	–6.4198	1.6122	–6.4918	1.5694
			U.C.L.	–3.4902	1.6144	–3.6062	1.5845
			Width	2.9296	0.1946	2.8856	0.2091

Table 6.9 (continued)

				Normal		Double Exponential	
Parameter	Par.Value	Interval Estimate		Average	S.D.	Average	S.D.
α_3(e,f)	3.00	95% C.I:	L.C.L.	1.3369	1.4220	1.3603	1.4780
			U.C.L.	4.7563	1.4244	4.7698	1.4999
			Width	3.4194	0.2616	3.4095	0.2724
		90% C.I:	L.C.L.	1.6144	1.4178	1.6568	1.4707
			U.C.L.	4.4715	1.4182	4.4869	1.4944
			Width	2.8570	0.2037	2.8300	0.2100
β(e,f)	1.00	95% C.I:	L.C.L.	0.9755	0.0122	0.9767	0.0130
			U.C.L.	1.0236	0.0122	1.0244	0.0128
			Width	0.0481	0.0034	0.0476	0.0037
		90% C.I:	L.C.L.	0.9796	0.0121	0.9808	0.0130
			U.C.L.	1.0197	0.0120	1.0203	0.0127
			Width	0.0401	0.0027	0.0396	0.0029

The reader should be reminded that, the bootstrap procedure for the confidence interval demonstrated above, represents perhaps the greatest charm of the bootstrap technique. Using the technique, confidence intervals which may be difficult to obtain through the usual MINQUE are possible.

Accuracy of Bootstrap Confidence Intervals

In this simulation study, 90% and 95% bootstrap confidence interval were constructed at each of the 400 trials. Table 6.10 shows the percentage of the number of times each of the six population parameter fell within the 90% and 95% bootstrap confidence interval for all the six design factor combinations (cells a through f). Ideal percentages are expected to be near $(1-\alpha)100$, for $\alpha = 0.01$ or 0.05.

From these results, it is shown that, near perfect percentages were observed for all the six parameters at the 0.05 level of the intraclass correlation under the normal distribution. At this level of the population intraclass correlation condition, percentages under the double exponential, though not as good as those under the normal, were not far off from the expected quantity $(1-\alpha)100$. For most of the parameters, disappointingly low percentages were observed at the 0.20 level of the intraclass correlation, particularly under the double exponential errors and sets of random effects of the model. Even at the other two levels of the intraclass correlation condition, there were more coverage probabilities which were below the expected quantity $(1-\alpha)$ than those above the expected quantity. This finding perhaps sends a precautionally message to research practitioners to aim higher

coverage probabilities when setting confidence intervals rather than investing high hopes at the conventional 0.9 or 0.95 coverage probabilities. However, for the parameters σ^2 and β, percentages extremely close to $(1-\alpha)100\%$ were observed for all the six design factor combinations.

Table 6.10

Percentage of times that the true population parameters fell within the confidence intervals formed using the bootstrap procedure at the three levels of the intra-class correlation.

		Normal		Double Exponential	
Value ρ	Expected	90% C.I.	95% C.I.	90% C.I.	95% C.I.
0.01 τ^2	1.00	97.5	99.0	97.8	98.8
0.05	5.26	89.5	94.0	81.0	88.3
0.20	25.00	58.8	70.3	42.3	52.0
0.01 σ^2	100.00	89.8	94.8	85.0	91.8
0.05	100.00	88.3	94.0	85.5	92.0
0.20	100.00	98.5	92.8	85.5	92.3
0.01 ρ	0.01	98.0	99.0	97.5	99.0
0.05	0.05	89.3	93.8	82.0	88.0
0.20	0.20	59.8	68.8	45.0	52.5
0.01 α_1	–5.00	87.8	93.0	88.0	93.0
0.05	–5.00	81.8	88.3	81.8	88.3
0.20	–5.00	64.5	74.0	62.8	71.3
0.01 α_3	3.00	87.5	93.5	86.8	92.8
0.05	3.00	82.3	88.3	84.0	89.3
0.20	3.00	70.3	76.8	66.5	76.8
0.01 β	1.00	89.0	95.0	87.5	93.5
0.05	1.00	86.5	93.3	87.8	93.5
0.20	1.00	88.5	95.0	87.5	93.5

CHAPTER VII

SUMMARY, TECHNICAL DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Overview

The primary purpose of the study was to demonstrate the operation of the bootstrap in estimating parameters of a mixed hierarchical linear model with random intercepts. The study demonstrated the ability of the bootstrap algorithm in providing the estimates of the fixed and random effects of the model, generating bootstrap empirical distributions and standard errors of the statistics and thereby constructing confidence intervals about the parameters.

The design of the study utilized samples generated from populations of known parameters. Computer programs used to generate independent samples and perform Monte Carlo simulations were coded in Statistical Analysis Systems (SAS), mostly using Interactive Matrix Language (SAS/IML). Generation of data from populations of known parameter values provided a check for the performance of the estimation procedures.

Population distributions from which samples were drawn from represented the normal and double exponential (or Laplace) distributions. The double exponential distribution (an example of a distribution with fairly long and thick tails) represented a distribution with some departure from normality, a situation which most classical statistical methods are typically not usable.

The Minimum Norm Quadratic Unbiased Estimation (MINQUE) procedure was adopted as a useful method of estimating the parameters of the model at each bootstrap replication. The method provided a comparable partnership with the

bootstrap since they both do not require the normal distributional properties. Thus, for each parameter of the model, two estimators were provided. These estimators represented the MINQUE estimator based on the original sample and the bootstrap estimator based on the resampled data.

Though the derivation of the MINQUE is based on arbitrary weights in the norm, the present study adopted an ANOVA–type method of independently estimating the variance components of the model as in Hanushek (1974). The values therein those prior estimates were used to determine the weights to be used in MINQUE.

In order to extensively assess the behavior of the bootstrap and MINQUE estimators, a total of 2400 Monte Carlo simulation trials, each consisting of a different data set were performed for six design factor combinations. The six design factor combinations represented the three levels of the intraclass correlation by the two distributional models (normal and double exponential).

In addition to simulated data, the bootstrap method and MINQUE were also applied on actual field research data to estimate both the fixed and random effects of the model involving the effect of institutional, classroom and individual teacher variables on self–efficacy of high school teachers. For each parameter of this specific model, the two estimators, the MINQUE and the bootstrap estimates were provided side by side. However, the bootstrap's additional estimation advantage was demonstrated by providing the bootstrap standard errors, empirical bootstrap sampling distributions of estimators, and the 95% bootstrap confidence intervals about each of the parameters of the teachers' self–efficacy prediction model. The bootstrap estimate of bias for each estimator were also provided.

Summary

Teachers' Self–Efficacy Model

Three parameters of the random part of the model, representing the intra– and inter–teacher variances, denoted by σ_e^2 and τ^2 respectively and the intra–teacher correlation denoted by ρ were estimated using both MINQUE and the bootstrap. In addition, eleven fixed effects parameters of the fixed part of the model representing the effects of Mathematics (α_1), Science (α_2), English (α_3), Social Science (α_4), class level of preparation (β_1), class size (β_2), average student achievement level (β_3), staff cooperation (γ_1), teacher control (γ_2), principal leadership (γ_3) and the constant common to all classrooms denoted by γ_{00} were studied.

The bootstrap estimates of the effects of intra–teacher variance, inter–teacher variance, the effect of class size, staff cooperation, and principal leadership were close to the MINQUE estimates. For these estimators, the bootstrap estimate of bias was no more than 0.008. However, except for the estimate of the constant γ_{00}, whose bootstrap estimate of bias was 0.1124, the bootstrap estimate of bias for the remaining nine estimators was no more than 0.04.

The bootstrap provided additional estimation informations which was not available through the MINQUE. These included the bootstrap standard error of each estimator, the 95% confidence intervals about the parameters, and the empirical bootstrap distribution of the statistics. Extremely low values of the bootstrap standard errors were observed particularly for the inter– and intra–teacher variances, the effects of the class level of preparation, class size, teacher control, and principal leadership. The bootstrap standard errors for these estimators were all close to 0.01.

As a means of testing hypothesis about the parameters of the teachers' self–efficacy prediction model, the 95% bootstrap confidence intervals about the

parameters were constructed. Based on these intervals, hypotheses of whether each of these parameters was different from zero were tested. Based on this bootstrap fashion of testing hypothesis, all factors, with an exception of principal leadership were found to have a statistically significant effect on teachers' self–efficacy.

Seven percentage polygons based on B = 1000 bootstrap replications of the estimators of the inter–teacher variance (τ^2), intra–teacher variance (σ_e^2), the intra–teacher correlation (ρ), the fixed effects of Mathematics (α_1), Science (α_2), English (α_3), and Social Science (α_4) were presented. Though the estimate of the sampling distribution of the inter–teacher variance (τ^2) was slightly positively skewed, and that of the effects of Mathematics (α_1), Social Science (α_4), and Science (α_2) were slightly negatively skewed, estimates of sampling distributions for all other estimators were fairly symmetric. But perhaps more importantly, percentage polygons for all seven estimators were centered extremely close to their corresponding usual MINQUE point estimators.

By applying the bootstrap method on actual field research data, the study demonstrated three features which research practitioners may find useful. First is the bootstrap's ability to provide the standard errors, empirical bootstrap sampling distributions of estimators, and setting confidence intervals about each of the parameters. This feature is typically not available through classical methods in the absence of certain distributional assumptions. The second feature was the flexability of the design in accommondating a wide range of independent variables (both continuous and discrete) in the model. The ability of a design to accommodate all types of independent effects is important given the limitations of most available statistical packages. For instance, the procedure VARCOM in SAS allows only for independent effects limited to main effects, interactions, and nested effects but not continuous effects. The third feature and perhaps the least expected was the efficiency of the bootstrap computer code. Though the bootstrap is

typically perceived as computer intensive, the present study utilized a simple program coded in SAS/IML through the MSU IBM 3090 VF mainframe computer. With this program, one bootstrap trial on the full model (seven independent variables) of B = 1000 replications took approximately 18:31.16 CPU time, which was not very expensive.

The Simulated Models

Different simulated models corresponding to each of the six design factor combinations (see Table 4.1) were studied. The six models represented the two distributional models (normal and double exponential) by three levels of the population intraclass correlation condition (ρ = 0.01, 0.05, and 0.20). The six design factor combinations are denoted by cells a through f. Each of the six models specified according to the design factor combination contained seven parameters of which three were random and four fixed. The three random effects parameters represented the inter–class variance (τ^2), the intra–class variance (σ_e^2), and the intraclass correlation (ρ). The four fixed effects parameters, α_1, α_2, α_3, and β represented the three levels of the fixed factor and the coefficient of the covariates. Estimation of non–redundant effects were based on α_1, α_3, and β.

A total of 400 Monte Carlo simulation trials (based on independent samples) were performed for each of the six design factor combinations (cells a through f), resulting in a grand total of 2400 Monte Carlo simulation trials, each based on a different data set drawn according to the specified design factor combination parameters. Ten estimable functions expressed in terms of the usual MINQUE and/or bootstrap estimates were used to assess the estimation of both the usual MINQUE and the bootstrap estimators. The ten estimable functions were carefully chosen to provide meaningful statistics like the Mean Square Errors, MSE1 and

MSE2, for the usual MINQUE and bootstrap estimator respectively, and the bootstrap estimate of bias denoted by BIAS.

Since data were drawn from populations of known parameter values, estimable functions were checked against their expected parameters values. The following is a presentation of the summary of estimation results organized by the population parameters of the models.

Inter–class variance (τ^2)

At the 0.01 intra–class correlation conditions, both MINQUE and bootstrap overestimated the parameter τ^2 with biases equal to 0.0292 and 0.0597 under normal and double exponential respectively for MINQUE, and 0.3432 and 0.3765 under normal and double exponential respectively for the bootstrap. Bootstrap estimates clearly improved for $\rho = 0.05$ both under the normal and double exponential with biases of 0.0299 and 0.2709 under the normal and double exponential respectively. Corresponding biases for the MINQUE estimate were –0.0845 under normal and 0.1396 under the double exponential. The bootstrap estimate of bias was observed at 0.1144 for $\rho = 0.05$ compared to 0.3140 for $\rho = 0.01$ under the normal distribution.

Particularly successful estimation results for the parameter τ^2 were attained at the $\rho = 0.20$ intraclass correlation condition. The bootstrap with a bias of –0.0447 was clearly close to the usual MINQUE with a bias of –0.1480 under the normal distribution. The bootstrap was also fairly close to the usual MINQUE under the double exponential with the former registering a bias of 0.4018 and the later having a bias of 0.5167. The estimate of bias at this level was 0.1149, and the ratio R_1 was surprisingly close to 1.00.

On average therefore, the MINQUE was closer to the parameter than the bootstrap only at the 0.01 level of the intraclass correlation condition under the

normal. At $\rho = 0.05$ and 0.20, the bootstrap was close to the parameter value compared to the MINQUE under the normal distribution. Both methods failed to produce better estimates for τ^2 at all levels of the intraclass correlation under the double exponential distribution.

Percentage polygons for the 400 MINQUE and bootstrap estimates were centered near the true population parameter value of τ^2 at all levels of this intraclass correlation condition. However, though the bootstrap percentage polygon appeared to be positively skewed while the MINQUE polygon was fairly symmetric, at the 0.01 level of the intraclass correlation condition it was observed that a greater mass of observations were around 1.00 for the bootstrap percentage polygon than for the MINQUE polygon.

The bootstrap confidence intervals about the parameter τ^2 were extremely tight under the double exponential as well as under the normal distribution at the 0.01 level of the intraclass correlation condition. Bootstrap confidence intervals about τ^2 were fairly short, both under the normal and double exponential at the 0.05 level of the intraclass correlation condition, but were wider at the 0.20 level of the intraclass correlation condition. The percentage of times the true parameter value of τ^2 fell within the 90 or 95 percentage bootstrap confidence intervals were close to either 90 or 95 at the 0.05 level of the intraclass correlation condition. The percentage of times the parameter value was captured by the bootstrap confidence intervals were furthest from the expected confidence coefficient at the 0.20 level of the intraclass correlation condition (see Table 6.10).

Intra–class variance (σ_e^2)

MINQUE and bootstrap fairly accurately estimated the population inter–class variance both under the normal and double exponential distribution of errors and sets of random effects parameters, at all levels of the intraclass

correlation condition. However, at the 0.20 level of the intraclass correlation, the bootstrap was closer to the parameter value than the MINQUE with a bias of 0.0437 compared to the MINQUE bias of 0.1660 under the normal distribution.

At the 0.01 level of the intraclass correlation the statistic R_2 was extremely close to unity, as expected. At all three levels of the intraclass correlation, the standard deviation of the functions of the estimates were relatively high under double exponential than under the normal distribution.

Percentage polygons for the MINQUE and bootstrap estimates at all levels of the intraclass correlation, showed the bootstrap following the usual MINQUE quite closely. Percentage polygons for both estimators were centered extremely close to the true parameter value of σ_e^2, which was set at 100. Thus it can be argued that, while both MINQUE and the bootstrap fairly accurately estimate σ_e^2; efficiency of these estimates is severely affected by the nature and size of the tails of the distribution of the errors and sets of random effects parameters. Both estimators are less efficient under a distribution with fairly long and/or thick tails than under a distribution with short and/or thin tails. But the measure of their relative efficiency was extremely close to unity.

The bootstrap confidence intervals about the parameter σ_e^2 showed a very successful bootstrap interval estimation process. The average widths of the confidence intervals were quite low, particularly under the normal distribution. At all levels of the population intraclass correlation condition, the percentage of times the true parameter value of σ_e^2 fell within the 90 or 95 percentage bootstrap confidence intervals were extremely close to either 90 or 95 under both normal and double exponential distributions.

Intraclass Correlation (ρ)

At the 0.01 level of the population intraclass correlation condition, both MINQUE and bootstrap very slightly overestimated ρ under both the normal and double exponential. The biases were extremely close to zero under both distributions. With an exception of R_1 which was poorly estimated, all other nine functions of $\hat{\rho}$ and/or $\hat{\rho}^*$ were fairly accurately estimated. At this level of the intraclass correlation conditions the mean square errors for both MINQUE and the bootstrap were particularly close to zero, under both distributions.

At the 0.05 level of the intraclass correlation condition, all ten estimable functions of $\hat{\rho}$ and/or $\hat{\rho}^*$ were very successfully estimated, both by the MINQUE and the bootstrap. The biases for the bootstrap and MINQUE estimators were extremely close to zero under both the normal and double exponential distributions.

The bootstrap slightly underestimated ρ under the normal, but very accurately estimated ρ under the double exponential at the 0.20 level of the intraclass correlation condition. The MINQUE, on the other hand, slightly underestimated ρ both under the normal and double exponential at this level of the population intraclass correlation condition. Under both distributions, the statistics R_1 and R_2 were extremely close to unity. However, at this condition of the intraclass correlation, the bootstrap performed as well as the MINQUE in estimating the ratio of the estimate to the parameter, ρ under both the normal and double exponential distributions.

Percentage polygons for the 400 MINQUE and bootstrap estimates of ρ under the normal and double exponential distributions showed that the two methods followed each other very closely. For both methods however,the estimates of ρ were more variable under the double exponential than under the normal distribution, at the 0.05 and 0.20 levels of the intraclass correlation condition.

The bootstrap interval estimation about the parameter τ^2, as a component of ρ (see Equation 4.3) was successful, particularly at the 0.01 level of the intraclass correlation; 90 and 95 percentage confidence intervals about ρ were fairly successful under both normal and double exponential. However, the percentage of times, the parameter value of ρ fell within the bootstrap 90 and 95 percent confidence intervals were furthest from the expected confidence coefficient at the 0.20 levels of the intraclass correlation.

Fixed effects parameters ($\alpha_1, \alpha_2, \alpha_3$)

Since α_1, α_2, and α_3 are linearly dependent, estimation was only required for any two of them. Estimation results for α_1 and α_3 were presented.

At the 0.01 level of the intraclass correlation condition, both MINQUE and bootstrap fairly accurately estimated both α_1 and α_3 with biases of no more than 0.027 for α_1 and 0.037 for α_3. The statistics R_1 and R_2 at this level of the intraclass correlation were extremely close to 1.00 under both the normal and double exponential distributions. Mean square errors for both MINQUE and bootstrap estimates of α_1 and α_3 were no more than 0.87 under both normal and double exponential distributions. The measure of their relative efficiency was quite clsoe to one.

At the 0.05 level of the intraclass correlation, all ten estimable functions of $\hat{\alpha}_1$ and/or $\hat{\alpha}_1^*$ were very accurately estimated by both MINQUE and bootstrap, under the normal. However, the bootstrap and MINQUE estimates of α_3 were surprisingly accurate under the double exponential than under the normal. The average values of the functions R_1 and R_2 for both MINQUE and bootstrap estimates of α_1 and α_3 were extremely close to 1.00 under both normal and double exponential for $\rho = 0.05$.

At the 0.20 level of the intraclass correlation, though the bootstrap was closest to the parameter α_1 under the normal than under double exponential, the biases for both MINQUE and bootstrap were no more than 0.05. Both biases for bootstrap and MINQUE estimators of α_3 were close to 0.04 under the normal but near 0.07 under double exponential.

In general therefore, both MINQUE and bootstrap very successfully estimated the parameter α_1 and α_3 at all levels of the intraclass correlation under both normal and double exponential distributions. However, the mean square error for both MINQUE and bootstrap estimates of α_1 and α_3 tended to increase with intraclass correlation under both normal and double exponential distributions.

The bootstrap confidence intervals about the parameters α_1 and α_3 at the 0.05 and 0.01 levels of the intraclass correlation showed a more precise bootstrap interval estimation process under both normal and double exponential distributions. Except for $\rho = 0.20$, the percentage of times the 90 or 95 percent bootstrap confidence intervals captured the parameters α_1 and α_3 were extremely close to the expected confidence coefficient, $(1 - \alpha)\ 100\%$.

Coefficient of the covariates (β)

Perhaps the most accurate bootstrap and MINQUE estimation results were obtained for the parameter β, the coefficient of the covariates. For this parameter, the bootstrap and MINQUE average estimates over 400 trials were extremely close to the true parameter value regardless of the level of the intraclass correlation or distribution of the errors and sets of random effects parameters of the model. At all levels of the intraclass correlation condition, the bootstrap and MINQUE biases were never greater than 0.0005 and the bootstrap estimate of bias was perfectly nil under both normal and double exponential distribution.

Average values for the functions R_1 and R_2 of $\hat{\beta}$ and/or $\hat{\beta}^*$ were either extremely close to unit or exactly equal to unit at all levels of the intraclass correlation condition. The mean squares errors for the bootstrap and MINQUE estimators of β were no more than 0.0002 under the normal and 0.0003 under the double exponential at all three levels of the intraclass correlation condition. Thus, based on these results it was evident that the parameter β was extremely accurately estimated by both MINQUE and the bootstrap, regardless of the level of the intraclass correlation condition and the nature and size of the tail of the distribution.

Percentage polygons for the MINQUE and bootstrap estimates of β showed no obvious differences between MINQUE and the bootstrap nor between their estimation ability under the normal or the double exponential. At all levels of the intraclass correlation, the percentage polygons showed that a large mass of the bootstrap and MINQUE estimates were within 0.015 points from the true parameter value of β which was set at 1.00. Also, regardless of the level of the intraclass correlation, the bootstrap method for the confidence intervals about the parameter β was a remarkable success. Even the percentage of times the 90 or 95 percent confidence intervals captured the true parameter value of β were extremely close to the expected confidence coefficient at all levels of the intraclass correlation for both normal and double exponential distributions.

Technical Discussion

Much of statistical inference amounts to describing the relationship between a sample and the population from which the sample was drawn. Consider for instance, the statistic $\hat{\theta}$ used to estimate an unknown parameter θ. Suppose we define a function R given by $R = \hat{\theta}/\theta$. Since the behavior of R is unobservable, we may wish to approximate its distribution. The main principle of the bootstrap is

to estimate the unknown distribution of a function, such as R by the distribution of $R^* = \hat{\theta}^*/\hat{\theta}$, where $\hat{\theta}^*$ is the bootstrap version of $\hat{\theta}$, computed from repeated resampling.

The key feature of this argument is the hypothesis that the relationship between $\hat{\theta}$ and $\hat{\theta}^*$ should closely resemble that between $\hat{\theta}$ and θ. Under the assumption that the relationships are identical, we equate the two ratios, R and R^* and obtain the estimate of θ which is a function of data. Similar arguments can be made for other functions like say, $D^* = \hat{\theta}^* - \hat{\theta}$ whose distribution will resemble that of $D = \hat{\theta} - \theta$. Bootstrap confidence intervals are then constructed based on this approximation as demonstrated in Equation 6.1 through 6.3 in Chapter VI of this dissertation.

In the present study, through Monte Carlo simulations, the distributions of R and R^* were observed through two types of resampling. The distribution of R was examined by drawing a random sample from a population having known parameters, computing the statistic $\hat{\theta}$ and repeating the process a large number of times. On the other hand, the distribution of R^* was observed by drawing one sample from the population similar to the one used in resampling for R. From this sample, a random sample of the same size is drawn with replacement, the statistic $\hat{\theta}^*$ computed, and the process repeated a large number of times. The statistic $\hat{\theta}$ based on the original sample was also computed. The distributions of R and R^* were then derived from this systems. The purpose was then to empirically examine the resemblence of the distribution of R^* and that of R.

Figures 7.1 and 7.2 presents the percentage polygons for the distributions of R and R^* representing the ratios of the estimators of the random and fixed parameters of a mixed hierarchical linear model discussed in Chapter II of this

Figure 7.1

Percentage polygons for the distributions of R and R* representing the ratios of the estimates of the random parameters τ^2, σ_e^2, and ρ.

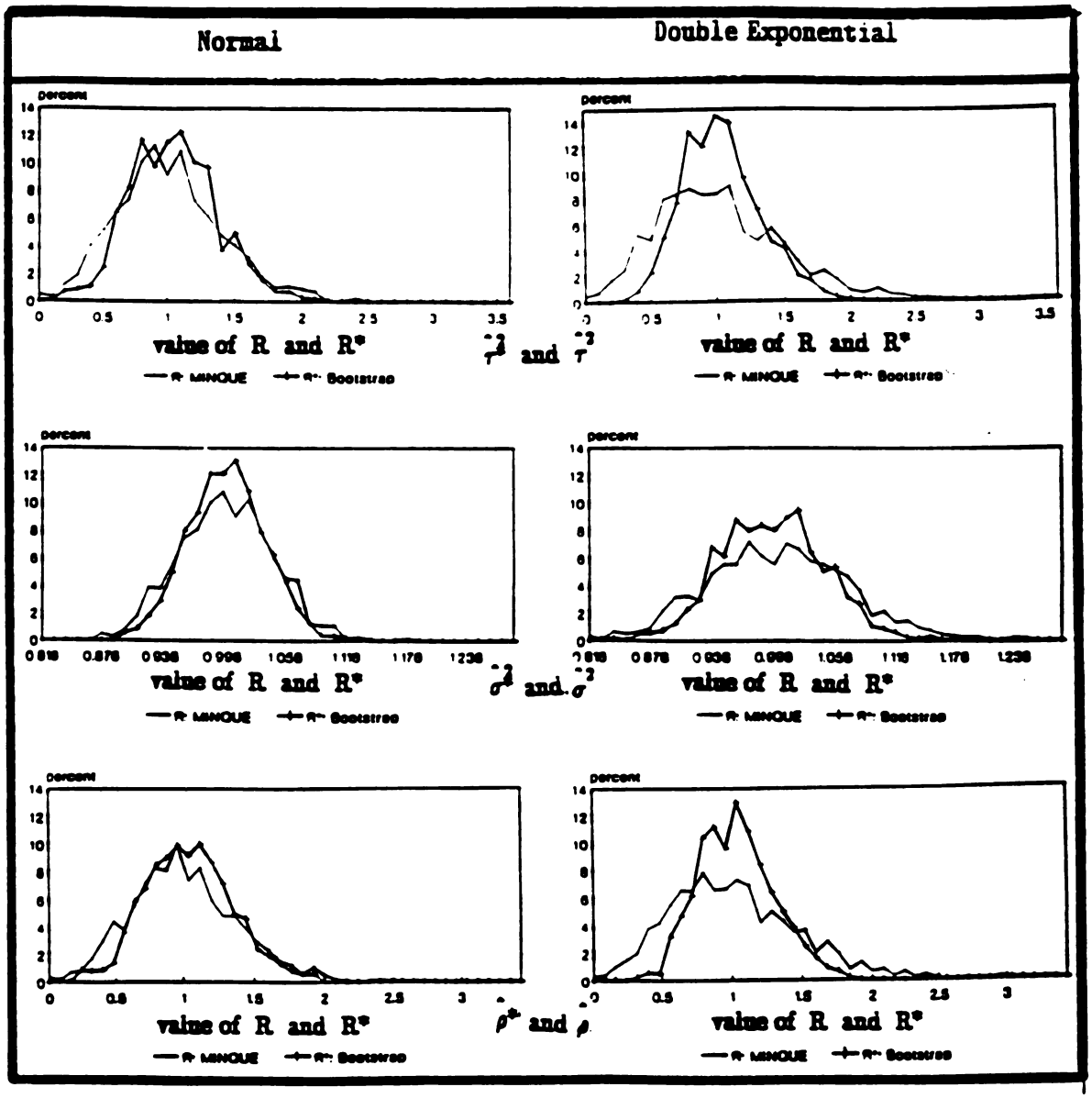
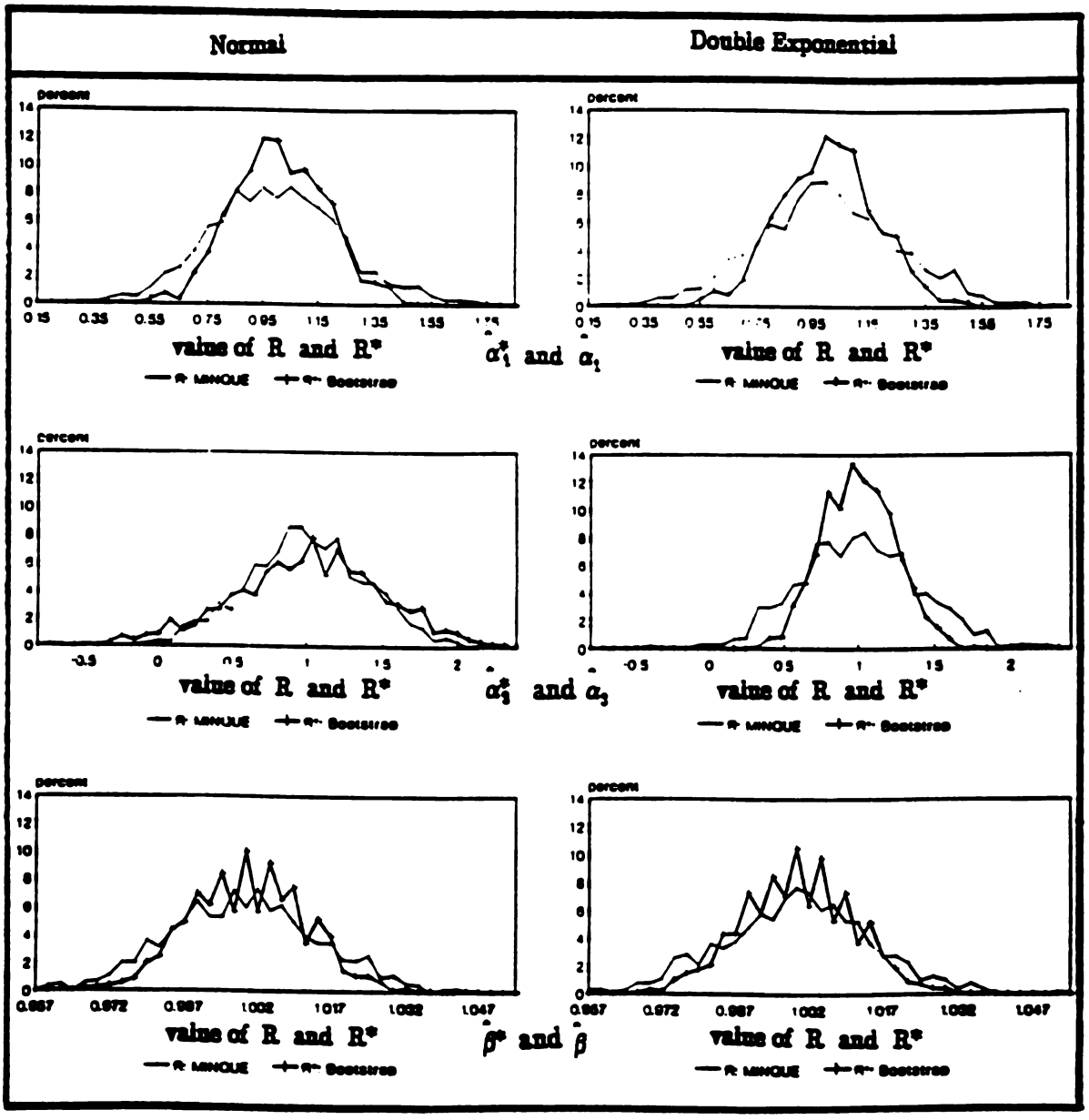


Figure 7.2

Percentage polygons for the distribution of R and R* representing the ratios of the estimates of the fixed parameters α_1, α_3, and β.



dissertation. The estimators represented in Figure 7.1 correspond to the random parameters τ^2, σ_e^2, and ρ while those represented in Figure 7.2 correspond to the fixed parameters α_1, α_3, and β. Estimation of these parameters was done at the 0.05 level of the intraclass correlation condition.

The expected value of both R and R* is 1.00. Consequently, both percentage polygons derived from the resampled data should be centered near 1.00. Indeed, Figure 7.1 and 7.2 shows that both percentage polygons were centered extremely close to 1.00.

It is important to emphasize that the distribution of R represents a sampling distribution of a statistic which is unobservable in actual research situations. Properties of this distribution can only be viewed theoretically for certain statistics, typically via the normal theory. On the other hand, the distribution of R* represents an approximation of the distribution of R. More importantly, the distribution of R* is almost always observable via the bootstrap algorithm. If the distribution of R* fairly accurately approximates the distribution of R, then the bootstrap proves itself as a highly promising method in statistics.

From Figures 7.1 and 7.2, it is apparent that the distributions of R and R* are fairly similar, particularly in terms of their location (or central tendency). They differ slightly in variability. However, the distribution of R* surprisingly appear to be even "better" than that of R in the sense that, a greater mass of observations are near 1.00 under the R* curve than under the R curve. This variations were clearly marked under the double exponential than under the normal distribution. Such variations in the distributions of R and R*, though slight, by underlying distribution of errors and the random effects of the model were demonstrated to be consistent for all estimators of the six parameters of the mixed hierarchical linear model considered in the study (see Figures 7.1 and 7.2).

Conclusions

The following conclusions were drawn from the results of the Monte Carlo simulation study and the results of the application the bootstrap and MINQUE on the estimation of the teachers' self–efficacy prediction model.

1. Though the main mission of the bootstrap is not point estimation, the average of the bootstrap estimates over B bootstrap replications can sometimes be closer to the parameter value that the estimator based on the original sample. Thus, the bootstrap may be viewed as both a point and interval estimation technique.

2. Efficiency of the usual MINQUE and the bootstrap estimators of the parameters of a model are typically affected by the nature and size of the tails of the distribution of the errors and sets of random effects of the model. Both estimators are less efficient under a distribution with fairly long tick tails than under a distribution with short think tails. In addition, the effect of the nature and size of the tails of distribution tends to be more severe in estimating random effects than fixed effects of the model.

3. The bootstrap percentile method for the confidence intervals about the parameters τ^2, ρ, α_1, and α_3 were successful at low intraclass correlation conditions. At the 0.20 level of the intraclass correlation, the coverage probabilities of the confidence intervals about these parameters was quite low. However, at and below the 0.05 level of the intraclass correlation condition, the bootstrap percentile method of the confidence intervals was shown to be highly promising.

4. The bootstrap's ability to estimate the standard error of the statistics, generating empirical sampling distribution of the estimators and thereby setting confidence intervals about parameters, without reference to any distributional properties is the single most promising feature of the bootstrap. This ability was very successfully demonstrated in the present study. Most importantly, the success of the bootstrap point and interval estimation abilities were proved by comparing the bootstrap estimates against the pre–determined true values of the model parameters.

5. The MINQUE and bootstrap estimate of the coefficient of the covariates of the model was surprisingly accurate. The bootstrap standard errors were extremely low and bias was minimal. Even the bootstrap confidence intervals about the parameter β were extremely precise.

6. In applying the bootstrap and MINQUE methods on the teachers' self–efficacy prediction model, which contained several predictors, showed the promising ability of the bootstrap and MINQUE. The MINQUE which was once considered computationally prohibitive can be used on such a large model with easy; even via the bootstrap which involves repeated computation. The bootstrap algorithm can be implemented on a large model of seven independent variables at a cost of no more than 20 CPU time for one trial of 1000 replications.

7. For a statistic $\hat{\theta}$ used to estimate a parameter θ, the function R, defined by $R = \hat{\theta}/\theta$ was used to represent the relationship between $\hat{\theta}$ and θ. Given $\hat{\theta}^*$ as the bootstrap version of $\hat{\theta}$ computed from repeated resampling, we define the

function $R^* = \hat{\theta}^*/\hat{\theta}$ as an approximation to R. Through Monte Carlo simulations, the distributions of R and R^* were found to be fairly similar, particularly in terms of central tendency. The distributions differed slightly in variability. The distribution of R^* was slightly less variable than the distribution of R.

Recommendations

Through Monte Carlo simulations, the bootstrap was demonstrated as a promising approach to estimating the standard error of the statistic, generating its sampling distribution and thereby setting confidence intervals about a parameter. This approach was empirically shown to work very well in estimating the parameters of a mixed hierarchical model whose errors and random effects parameters are either normally or double exponentially distributed. Applicability of the bootstrap approach was further demonstrated in estimating the parameters of the teachers' self–efficacy prediction model.

Implementation of the bootstrap method requires a great deal of computer usage. Though modern fast and relatively inexpensive computers are readily available, software to implement the bootstrap algorithm are currently unavailable. Development of such software is highly recommended to make the bootstrap available to research practitioners.

Recommendations for further research

Results of a simulation study are typically limited in their generalization to the conditions examined in the study. The present study examined the operation of the bootstrap via MINQUE in estimating parameters of a mixed hierarchical model when the errors and random effects are either normally or double exponentially distributed. The study was done under three levels of the intraclass correlation conditions.

The effectiveness of the bootstrap approach under severely skewed or heavily tailed distributions remain to be seen. Studies to implement the bootstrap method in examining the sampling distribution of estimators of parameters whose underlying distributions are badly skewed like the gamma or heavily tailed like the Cauchy are deemed necessary to fully understand the abilities and limitations of the bootstrap approach.

The present study considered an hierarchical model consisting of "micro" and "Macro" models with the assumption that only the intercepts were random. By fixing other coefficients of the "micro" models simplified the study to one of examining the variance components without covariates. A study to examine the operation of the bootstrap in models involving not only variance components but also covariance component will shed more light on the understanding of the bootstrap in the hierarchical context.

The use of the bootstrap percentile method for the confidence interval at $\rho = 0.20$ was not very successful in estimating certain parameters. A more promising bootstrap t–method for the confidence interval was not used due to the fact that the standard error of the MINQUE estimator was not known. Further research geared to determining the standard error of MINQUE is deemed necessary in order for the bootstrap users to utilize the t–method for the confidence intervals.

APPENDICES

APPENDIX A

SUMMARY OF COMPUTATIONAL FORMULAE

The object of MINQUE the study was to find the estimate $\hat{\underset{\sim}{\sigma}}$ of the variance compont $\underset{\sim}{\sigma}$ of the two–level mixed model $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}\underset{\sim}{b}$. The estimate $\underset{\sim}{\sigma}$ using weights w_0 and w_1 in the norm is given by

$$\hat{\underset{\sim}{\sigma}} = \begin{bmatrix} \hat{\sigma}_e^2 \\ \hat{\tau}^2 \end{bmatrix} = \underset{\sim}{F}_w^{-1}\,\underset{\sim}{U}_w$$

where

$$\underset{\sim}{F}_w = \{tr(\underset{\sim}{P}_w\underset{\sim}{Z}_k\underset{\sim}{Z}_k'\underset{\sim}{P}_w\underset{\sim}{Z}_{k'}\underset{\sim}{Z}_{k'}')\} = \begin{bmatrix} f_{00} & f_{01} \\ f_{10} & f_{11} \end{bmatrix} \text{ for } k, k' = 0,1$$

and

$$\underset{\sim}{U}_w = \{\underset{\sim}{Y}'\underset{\sim}{P}_w\underset{\sim}{Z}_k\underset{\sim}{Z}_k'\underset{\sim}{P}_w\underset{\sim}{Y}\} = \begin{bmatrix} u_0 \\ u_1 \end{bmatrix}$$

for $\underset{\sim}{P}_w = \underset{\sim}{V}_w^{-1} - \underset{\sim}{V}_w^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}$.

Let $\underset{\sim}{K} = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}$ and $\underset{\sim}{A}_w = \underset{\sim}{V}_w^{-1}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}$

such that $\underset{\sim}{P}_w = \underset{\sim}{V}_w^{-1} - \underset{\sim}{A}_w$.

If we define $w = \frac{1}{1-w_1}$ and $c_j = \frac{w_1}{1+(n_j-1)w_1}$ for n_j = number of micro units in macro group j, then the following is the summary of the computational formulae:

1. $\underset{\sim}{K}$ and $\underset{\sim}{A}_w$:

(a) Let $\underset{\sim}{K} = \underset{\sim}{X}' \underset{\sim}{V}_w^{-1} \underset{\sim}{X} = \sum_j \underset{\sim}{X}_j' \underset{\sim}{V}_j^{-1} \underset{\sim}{X}_j$

$$= \{w \Sigma (\underset{\sim}{X}_j' \underset{\sim}{X}_j - c_j \underset{\sim}{X}_j' \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}' \underset{\sim}{X}_j)\}^-$$

$$= \{w \Sigma (\underset{\sim}{X}_j' \underset{\sim}{X}_j - c_j \underset{\sim}{S}_j \underset{\sim}{S}_j')\}^-$$

(b) $\underset{\sim}{A}_w = \underset{\sim}{V}^{-1} \underset{\sim}{X} \underset{\sim}{K} \underset{\sim}{X}' \underset{\sim}{V}_w^{-1}$

$$= w^2 \Sigma \{\underset{\sim}{I}_{n_j} - c_j \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}') \underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j' (\underset{\sim}{I}_{n_j} - c_j \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}')\}$$

$$= w^2 \Sigma (\underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j' - c_j \underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j' \underset{\sim}{Z}_j \underset{\sim}{Z}_j' - c_j \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}' \underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j'$$

$$+ c_j^2 \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}' \underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j' \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}').$$

2. $\underset{\sim}{F}_w$ matrix:

(a) $f_{00} = tr(\underset{\sim}{V}_w^{-2}) - tr(\underset{\sim}{V}_w^{-1} \underset{\sim}{A}_w)$

where

$tr(\underset{\sim}{V}_w^{-2}) = w^2 \Sigma n_j \{(1-c_j)^2 + c_j^2 (n_j - 1)\}$

$tr(\underset{\sim}{V}_w^{-1} \underset{\sim}{A}_w) = w^3 \Sigma \{tr(\underset{\sim}{t}_j) - c_j a_j [(1-c_j n_j)^2 + (2-c_j n_j)]\}$

where $\underset{\sim}{t}_j = \underset{\sim}{X}_j' \underset{\sim}{X}_j \underset{\sim}{K}$ and $a_j = tr(\underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j' \underset{\sim}{Z}_{1j} \underset{\sim}{Z}_{1j}')$

which is a scalar simplified by $a_j = tr(\underset{\sim}{S}_j' \underset{\sim}{K} \underset{\sim}{S}_j) = \underset{\sim}{S}_j' \underset{\sim}{K} \underset{\sim}{S}_j$.

(b) $f_{01} = f_{10} = tr(\underset{\sim}{V}_w^{-2} \underset{\sim}{Z}_1 \underset{\sim}{Z}_1) - tr(\underset{\sim}{V}_w^{-1} \underset{\sim}{A} \underset{\sim}{Z}_1 \underset{\sim}{Z}_1')$

where

$tr(\underset{\sim}{V}_w^{-2} \underset{\sim}{Z}_1 \underset{\sim}{Z}_1') = w^2 \Sigma n_j (1-c_j n_j)^2$

$tr(\underset{\sim}{V}_w^{-1} \underset{\sim}{A}_w \underset{\sim}{Z}_1 \underset{\sim}{Z}_1' \underset{\sim}{Z}_1 \underset{\sim}{Z}_1') = w^1 \Sigma a_j (1-c_j n_j)^3$

(c) $f_{11} = tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') - tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1')$

where

$tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^2 \Sigma\, n_j^2(1-c_jn_j)^2$

$tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^3 \Sigma\, n_ja_j(1-c_jn_j)^3$

3. $\underset{\sim}{\alpha}$ and $\underset{\sim}{U}_w$:

(a) $\hat{\alpha} = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}$

$= \underset{\sim}{K}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}$

$= w \Sigma\, \underset{\sim}{K}\underset{\sim}{X}_j' (I_{n_j} - c_j\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}')\underset{\sim}{Y}_j$

$= w \Sigma\, \underset{\sim}{K}\underset{\sim}{X}_j' - c_j\underset{\sim}{K}\underset{\sim}{X}_j'\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}'\underset{\sim}{Y}_j)$

$= w \Sigma\, (\underset{\sim}{K}\underset{\sim}{X}_j' - c_jr_j\underset{\sim}{K}\underset{\sim}{S}_j)$ where $r_j = \underset{\sim}{Z}_{1j}'\underset{\sim}{Y}_j$ is the sum of $\underset{\sim}{Y}$ elements in context j

(b) $u_0 = w^2 \Sigma\, \underset{\sim}{d}_j'(I_{n_j} - 2c_j\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}' + c_j^2n_j\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}')\underset{\sim}{d}_j$ for $\underset{\sim}{d}_j = \underset{\sim}{Y}_j - \underset{\sim}{X}_j\hat{\alpha}$

$w^2 \Sigma\, (\underset{\sim}{d}_j'\underset{\sim}{d}_j - 2c_j\underset{\sim}{d}_j'\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}'\underset{\sim}{d}_j + c_j^2n_j\underset{\sim}{d}_j'\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}'\underset{\sim}{d}_j)$

$w^2 \Sigma\, (\underset{\sim}{d}_j\underset{\sim}{d}_j - 2c_jh_j^2 + c_jn_jh_j^2)$

for $h_j = \underset{\sim}{Z}_{1j}'\underset{\sim}{d}_j = \underset{\sim}{d}_j'\underset{\sim}{Z}_{1j} = \Sigma\, (\underset{\sim}{Y}_j - \underset{\sim}{X}_j\hat{\alpha})$

(c) $u_1 = w^2 \Sigma\, (1-c_jn_j)^2\, \underset{\sim}{d}_j'\underset{\sim}{Z}_{1j}\underset{\sim}{Z}_{1j}'\underset{\sim}{d}_j$

$= w^2 \Sigma\, h_j^2(1-c_jn_j)^2.$

4. MINQUE for the Fixed Effects:

Model: $\underset{\sim}{Y} = \underset{\sim}{X}\alpha + \underset{\sim}{Z}\underset{\sim}{b}$

where $\underset{\sim}{Y}$ is $(n\times 1)$ vector of n observations

$\underset{\sim}{X}$ is an $(n\times\rho)$ matrix of known constants

α is a $(\rho\times 1)$ vector of fixed effects parameters

$\underset{\sim}{Z}$ is an (n×n) identify matrix (usually denoted by Z_0 in the random or mixed models).

$\underset{\sim}{b}$ is (n×1) vector of residual error terms.

$$\underset{\sim}{D}_w = w_0 \underset{\sim}{I}_n$$

$$\underset{\sim}{V}_w = \underset{\sim}{Z}\underset{\sim}{D}_w\underset{\sim}{Z}' = \underset{\sim}{D}_w = w_0\underset{\sim}{I}_n$$

which implies that

$$V_w^{-1} = \frac{1}{w_0}\underset{\sim}{I}_n \quad \text{and}$$

$$V_w^{-2} = \frac{1}{w_0^2}\underset{\sim}{I}_n$$

$$K = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^- = w_0(\underset{\sim}{X}'\underset{\sim}{X})^-$$

$$\underset{\sim}{A}_w = \underset{\sim}{V}_w^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{V}_w^{-1} = \underset{\sim}{V}_w^{-1}\underset{\sim}{X}K\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}$$

$$= \frac{1}{w_0^2}\cdot w_0\,\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'$$

$$= \frac{1}{w_0}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'$$

$$\underset{\sim}{F}_w = \mathrm{tr}(\underset{\sim}{P}_w\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{P}_w\underset{\sim}{Z}\underset{\sim}{Z}') = \mathrm{tr}(\underset{\sim}{P}_w\underset{\sim}{P}_w)$$

where $\underset{\sim}{P}_w = \underset{\sim}{V}_w^{-1} - \underset{\sim}{V}_w^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}$

$$= \mathrm{tr}(\underset{\sim}{V}_w^{-2} - \underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w)$$

$$= \mathrm{tr}(\underset{\sim}{V}_w^{-2}) - \mathrm{tr}(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w)$$

$$= \frac{n}{w_0^2} - \frac{p}{w_0^2}$$

$$\frac{1}{w_0^2}(n-p)$$

$$\underset{\sim}{U}_w = \underset{\sim}{Y}'\underset{\sim}{P}_w\underset{\sim}{Z}\underset{\sim}{Z}'\underset{\sim}{P}_w\underset{\sim}{Y} = \underset{\sim}{Y}\,\underset{\sim}{P}_w\underset{\sim}{P}_w\underset{\sim}{Y}\,.$$

$$\text{But } \underset{\sim}{P}_w \underset{\sim}{Y} = (\underset{\sim}{V}_w^{-1} - \underset{\sim}{V}_w^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1})\underset{\sim}{Y}$$

$$= \underset{\sim}{V}_w^{-1}(\underset{\sim}{Y} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y})$$

$$= \underset{\sim}{V}_w^{-1}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$\text{for } \hat{\underset{\sim}{\alpha}} = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}$$

thus,

$$\underset{\sim}{U}_w = (\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\alpha})'\underset{\sim}{V}_w^{-2}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$= \frac{1}{w_0^2}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$\text{where } \hat{\underset{\sim}{\alpha}} = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}$$

$$= w_0(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\frac{1}{w_0}\underset{\sim}{Y}$$

$$= (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y}.$$

Thus, for the general fixed effects model, the MINQUE estimator $\hat{\sigma}_0^2 = \hat{\sigma}_e^2$ is given by

$$\hat{\sigma}_e^2 = \underset{\sim}{F}_w^{-1}\underset{\sim}{U}_w$$

$$= \frac{w_0^2}{n-p} \cdot \frac{1}{w_0^2}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$= \frac{1}{n-p}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

which is independent of the weights w_0 and w_1. Consider for example the simplist and naive model given by,

$$Y_{ij} = \mu + \epsilon_{ij}.$$

In the notation of the form

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}\underset{\sim}{b},$$

$\underset{\sim}{X}$ is a $(n \times 1)$ vector of 1's

$\alpha = \mu$ is a scalar

$\underset{\sim}{Z} = \underset{\sim}{I}_n$ is an (n×n) indentity matrix

$\underset{\sim}{b}$ is a (n×1) vector of residual error terms.

In this specific case, P = 1 and

$$\hat{\alpha} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Y} = \bar{Y}.. \text{ (Grand mean).}$$

$$\underset{\sim}{U}_w = \frac{1}{w_0^2}(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\alpha})'(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\alpha})$$

$$= \frac{1}{w_0^2}\Sigma(Y_{ij} - \bar{Y}..)^2$$

such that, with $F_w^{-1} = \frac{w_0^2}{n-p}$,

the MINQUE estimator $\hat{\sigma}_0^2$ is given by

$$\hat{\sigma}_0^2 = \hat{\sigma}_e^2 = F_w^{-1}U_w$$

$$= \frac{w_0^2}{n-p} \cdot \frac{1}{w_0^2}\Sigma(Y_{ij} - \bar{Y}..)^2$$

$$= \frac{1}{n-1}\Sigma(Y_{ij} - \bar{Y}..)^2$$

$$= S^2$$

which is the moment estimator.

5. MINQUE for the one way random effects balanced model:

Model: $\underset{\sim}{Y} = \underset{\sim}{X}\alpha + \underset{\sim}{Z}\underset{\sim}{b}$ $\quad\quad$ $\{Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$

where $\underset{\sim}{Y}$ is (N×1) vector of N observations for N = nJ,

J = # of levels.

$\underset{\sim}{X}$ is (N×1) vector of 1's

$\alpha = \mu$

$\underset{\sim}{Z} = [\underset{\sim}{Z}_0\underset{\sim}{Z}_1]$, $\underset{\sim}{Z}_0 = \underset{\sim}{I}_N$

and $\underset{\sim}{Z}_1$ is (N×J) block diagonal, each block

being a column of 1's.

$\underset{\sim}{b} = [\underset{\sim}{b}_0\ \underset{\sim}{b}_1]$,

$\underset{\sim}{b}_0 = (N \times 1)$ vector of residual error terms

$\underset{\sim}{b}_1 = (J \times 1)$ vector of J unobservable random effects parameters.

(a) $\underset{\sim}{K} = (\underset{\sim}{X}' \underset{w}{\overset{-1}{\underset{\sim}{V}}} \underset{\sim}{X})^{-} = \{w \Sigma (\underset{\sim}{X}_j' (\underset{nj}{\underset{\sim}{I}} - c_j \underset{ij}{\underset{\sim}{Z}} \underset{ij}{\underset{\sim}{Z}}') \underset{\sim}{X}_j)\}^{-}$

$= \{w \Sigma (n - cn^2)\}^{-}$ since $n_j = n$; $c_j = c$

$= \{wnJ(1-\lambda)\}^{-1}$.

Thus

$$K = \frac{1}{wnJ(1-\lambda)} \quad \text{for } \lambda = \frac{\tau^2}{\tau^2 + \sigma^2/n}$$

(b) $\underset{w}{\underset{\sim}{A}} = w^2 \Sigma (1-c_j n_j)^2 \underset{\sim}{X}_j \underset{\sim}{K} \underset{\sim}{X}_j'$

$= w^2 K (1-\lambda)^2 \Sigma \underset{\sim}{X}_j \underset{\sim}{X}_j'$

(c) $f_{00} = tr(\underset{w}{\overset{-2}{\underset{\sim}{V}}}) - tr(\underset{w}{\overset{-1}{\underset{\sim}{V}}} \underset{w}{\underset{\sim}{A}})$

(i) $tr(\underset{w}{\overset{-2}{\underset{\sim}{V}}}) = w^2 \Sigma n_j \{(1-c_j)^2 + c_j^2 (n_j - 1)\}$

$= w^2 nJ((1-c)^2 + c^2(n-1)\}$

$= w^2 nJ(1-2c + nc^2)$

$= w^2 J(1-\lambda)^2 + (n-1)w^2 J$

(ii) $tr(\underset{w}{\overset{-1}{\underset{\sim}{V}}} \underset{w}{\underset{\sim}{A}}) = w^3 K \Sigma n_j (1-c_j n_j)^3$

$= w^3 KnJ(1-\lambda)^3$

$= w^3 \frac{1}{wnJ(1-\lambda)} nJ(1-\lambda)^3$

$= w^2 (1-\lambda)^2$.

Thus $f_{00} = w^2 J(1-\lambda)^2 + (n-1)w^2 J - w^2(1-\lambda)^2$

$= w^2 \{(J-1)(1-\lambda)^2 + (n-1)J\}$

$= w^2 \{(J-1)(1-\lambda)^2 + J(n-1)\}$

(d) $f_{01} = f_{10} = tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1') - tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1')$

(i) $tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^2 \Sigma n_j(1-c_jn_j)^2$

$= w^2nJ(1-\lambda)$

(ii) $tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^3K \Sigma n_j^2(1-c_jn_j)^3$

$= w^3Kn^2J(1-\lambda)^3$

$= w^3 \frac{1}{wnJ(1-\lambda)} n^2J(1-\lambda)^3$

$= w^2n(1-\lambda)^2$

Thus, $f_{01} = f_{10} = w^2nJ(1-\lambda)^2 - w^2n(1-\lambda)^2$

$= w^2n(1-\lambda)^2 (J-1)$

$= w^2n(J-1) (1-\lambda)^2$

(e) $f_{11} = tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') - t(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1')$

(i) $tr(\underset{\sim}{V}_w^{-2}\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^2 \Sigma n_j(1-c_jn_j)^2$

$= w^2n^2J(1-\lambda)^2$

(ii) $tr(\underset{\sim}{V}_w^{-1}\underset{\sim}{A}_w\underset{\sim}{Z}_1\underset{\sim}{Z}_1'\underset{\sim}{Z}_1\underset{\sim}{Z}_1') = w^3 \Sigma Kn_j(1-c_jn_j)^3$

$= w^3Kn^3J(1-\lambda)^3$

$= w^3 \frac{1}{wnJ(1-\lambda)} n^3J(1-\lambda)^3$

$= w^2n^2(1-\lambda)^2$

$\Rightarrow f_{11} = w^2n^2J(1-\lambda)^2 - w^2n^2(1-\lambda)^2$

$= w^2n^2(1-\lambda)^2 (J-1)$

$= w^2n^2(J-1) (1-\lambda)^2$

(f) $\hat{\underset{\sim}{\alpha}} = (\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{X})^{-} \underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}$

$= K\underset{\sim}{X}'\underset{\sim}{V}_w^{-1}\underset{\sim}{Y}$

$= wK \Sigma (1-\lambda)r_j$ (for r_j = sum of Y_{ij} in context j)

$= wK(1-\lambda) \Sigma rj$ $(\Sigma r_j = \Sigma\Sigma Y_{ij})$

$= wKnJ(1-\lambda) \bar{Y}..$

$$= w \frac{1}{wnJ(1-\lambda)} nJ(1-\lambda)\, \bar{Y}..$$

$$= \bar{Y}..$$

(g) (i) $\underset{\sim}{d}_j = (\underset{\sim}{Y}_j - \underset{\sim}{X}_j \hat{\underset{\sim}{\alpha}})$

(ii) $h_j = \underset{\sim}{Z}'_{ij}\underset{\sim}{d}_j = \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}..)$

$$\Rightarrow h_j^2 = \sum_{j=1}^{J} [\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}..]^2 = \sum_{j=1}^{J} (n_j \bar{Y}._j - n_j \bar{Y}..)^2$$

$$= \sum_{j=1}^{J} n^2(\bar{Y}._j - \bar{Y}..)^2 = n^2 \Sigma(\bar{Y}._j - \bar{Y}..)^2$$

$$= n(SSB)$$

(iii) $g_j = \underset{\sim}{d}_j \underset{\sim}{d}_j = \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}..)^2$

$$\Rightarrow \Sigma\, g_i = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}..)^2 = SST$$

(h) (i) $u_0 = w^2 \Sigma\, (g_i - 2c_j h_j^2 + c_j^2 n_j h_j^2)$

$$= w^2 \{\Sigma\, g_i - 2c\Sigma\, h_j^2 + c^2 n\, \Sigma h_j^2\}$$

$$= w^2 \{SST - (2cn)SSB + (c^2 n^2)SSB\}$$

$$= w^2 \{SSB(1-2\lambda+\lambda^2) + SSW\}$$

$$= w^2 \{SSB(1-\lambda)^2 + SSW\}$$

(ii) $u_1 = w^2 \Sigma(1-c_j n_j)^2 h_j^2$

$$= w^2(1-\lambda)^2\, n.SSB$$

$$= w^2 n(1-\lambda)^2\, SSB$$

(i) $D=\det(F_w) = f_{00}f_{11} - f_{01}f_{10} = f_{00}f_{11} - f_{01}^2$

(i)$f_{00}f_{11} = w^2\{(J-1)(1-\lambda)^2 + J(n-1)\}\{w^2 n^2 (J-1)(1-\lambda)^2\}$

$$= w^4 n^2 (J-1)^2 (1-\lambda)^4 + w^4 n^2 J(J-1)(n-1)(1-\lambda)^2$$

(ii) $f_{01}^2 = [w^2 n(J-1)(1-\lambda)^2]^2 = w^4 n^2 (J-1)^2 (1-\lambda)^4$

Thus, $D = f_{00}{}_{f}{}_{11} - f_{01}^2$

$$= w^4 n^2 J(J-1)(n-1)(1-\lambda)^2$$

(j) $\hat{\tau}^2 = (f_{00}u_1 - f_{01}u_0)/D$

(i) $f_{00}u_1 = W^2\{(J-1)(1-\lambda)^2 + J(N-1)\}\{w^2n(1-\lambda)^2SSB\}$

$= w^4n(J-1)(1-\lambda)^4\ SSB + w^4nJ(n-1)(1-\lambda)^2SSB$

(ii) $f_{01}u_0 = \{w^2n(J-1)(1-\lambda)^2\}\{w(SSB(1-\lambda)^2 + SSW)\}$

$= w^4n(J-1)(1-\lambda)^4\ SSB + w^4n(J-1)(1-\lambda)^2\ SSW$

Thus $f_{00}u_1 - f_{01}u_0 = wn(1-\lambda)^2\ [J(n-1)SSB - (J-1)SSW]$.

Thus

$$\hat{\tau}^2 = \frac{w^4n(1-\lambda)^2[J(n-1)SSB-(J-1)SSW]}{w^4n^2J(J-1)(n-1)(1-\lambda)^2}$$

$$= \frac{[J(n-1)SSB] - [(J-1)SSW]}{nJ(J-1)(n-1)}$$

$$= \frac{J(n-1)SSB}{nJ(J-1(n-1)} - \frac{(J-1)SSW}{nJ(J-1)(n-1)}$$

$$= \frac{SBB}{n(J-1)} - \frac{SSW}{nJ(n-1)} = \frac{MSB}{n} - \frac{MSW}{n}$$

As a result,

$$\hat{\tau}^2 = \frac{MSB-MSW}{n}$$

(which is the same as the method of moments).

(k) $\hat{\sigma}_e^2 = (f_{11}u_0 - f_{01}u_1)/D$

(i) $f_{11}u_0 = \{w^2n^2(J-1)(1-\lambda)^2\}\{w^2SB(1-\lambda)^2 + SSW\}$

$= w^4n^2(J-1)(1-\lambda)^4SSB + w^4n^2(J-1)(1-\lambda)^2SSW$

(ii) $f_{01}u_1 = \{w^2n(J-1)(1-\lambda)^2\}\{w^2n(1-\lambda)^2SSB\}$

$= w^4n^2(J-1)(1-\lambda)^4SSB$

which implies that,

$$f_{11}u_0 - f_{01}u_1 = w^4n^2(J-1)(1-\lambda)^2\ SSW$$

such that

$$\hat{\sigma}_e^2 = (f_{11}u_0 - f_{01}u_1)/D$$

$$= \frac{w^4n^2(J-1)(1-\lambda)^2SSW}{w^4n^2J(J-1)(n-1)(1-\lambda)^2}$$

$$= \frac{SSW}{J(n-1)}$$

$$= MSW$$

(which is the same as in the method of moments).

APPENDIX B

SAS/IML COMPUTER PROGRAMS

PART 1

```
*    COMPUTER PROGRAM TO IMPLEMENT THE BOOTSTRAP  ALGORITHM      ;
*      ON A SAMPLE HIERARCHICAL DATA DRAWN FROM A NORMAL         ;
*              POPULATION OF KNOWN PARAMETERS.                   ;
*_______________________________________________________________;
*                                                                ;
* THE PROGRAM FIRST SETS UP THE Z AND THE FIRST PART OF THE      ;
* X MATRIX EXCLUDING THE COVARIATES.  THE CONSTRUCTION OF        ;
* THESE MATRICES ARE BASED ON THE NUMBER OF OBSERVATIONS IN      ;
* CELL TO SATISFY THE REQUIREMENTS AS IN EQUATION 2.9 IN         ;
* CHAPTER II.  THE PROGRAM IDENTIFIES THE COMPONENTS ON EACH     ;
* AS DEMONSTRATED BY EQUATION 2.11 IN CHAPTER II.                ;
* THE WEIGHT w1 WAS DETERMINED SEPARATELY USING THE              ;
* HANUSHEK (1974) METHOD.                                        ;
PROC IML;
  START;
  W=1/(1-W1);
    GROUPS=50;
    NV11=REPEAT(20,2,1);
    NV21=REPEAT(25,5,1);
    NV31=REPEAT(30,10,1);
    NV1=NV11//NV21//NV31;
    NV12=REPEAT(35,5,1);
    NV22=REPEAT(40,3,1);
    NV32=REPEAT(20,3,1);
    NV42=REPEAT(25,5,1);
    NV2=NV12//NV22//NV32//NV42;
    NV13=REPEAT(30,10,1);
    NV23=REPEAT(35,5,1);
    NV33=REPEAT(40,2,1);
    NV3=NV13//NV23//NV33;
    NV=NV1//NV2//NV3;
  CV=W1/(1+(NV-1)*W1);
  X11=REPEAT(1,465,1);
  X12=REPEAT(1,480,1);
  X13=REPEAT(1,555,1);
  X01=REPEAT(0,465,1);
  X02=REPEAT(0,480,1);
  X03=REPEAT(0,555,1);
  X1=X11//X02//X03;
  X2=X01//X12//X03;
  X3=X01//X02//X13;
  X4=X1||X2||X3;


*----------------------------------------------------------------;
```

```
*  PROGRAM SEGMENT TO GENERATE DATA FROM A NORMAL POPULATION     ;
*  OF SPECIFIED PARAMETER VALUES.  THE PROGRAM FIRST DETERMINES  ;
*  THE FIXED EFFECTS PARAMETERS AND THE COVARIATE WHICH IN       ;
*  ARE USED IN TURN TO GENERATE THE OBSERVATIONS Y THROUGH       ;
*  THE EQUATION GIVEN BY:                                       ;
*                                                               ;
*                       Y = (X*ALPHA) + B + E                   ;
*                                                               ;
* SEED: IS ANY NUMBER USED TO CREATE A RANDOM NUMBER OF         ;
*  OBSERVATIONS FROM SOME POPULATION.                           ;

    SEED = 10199;
*                                                               ;
*TIS THE INDEX COUNTER FOR THE NUMBER OF SIMULATION TRIALS      ;
*---------------------------------------------------------------;
DO T = 1 TO 400;
    REFFECTS = 2.2935 * NORMAL(REPEAT(SEED,GROUPS,1));
    B1 = REFFECTS[1,1];
    N1 = NV[1,1];
    BJ1 = REPEAT(B1,N1,1);
DO I = 2 TO GROUPS;
    BJ = REFFECTS[I,1];
    N =  NV[I,1];
    BJ1=BJ1//REPEAT(BJ,N,1);
END;
    B=BJ1;
    E = 10 * NORMAL(REPEAT(SEED,1500,1));
    X41 = REPEAT(25,1500,1);
    X5 = INT(75 * UNIFORM(REPEAT(SEED,1500,1))) + X41;
    X=X4||X5;
    ALPHA = {-5,2,3,1.0};
    Y = (X*ALPHA) + B + E;
*---------------------------------------------------------------;
* AT THIS POINT A SPECIFIC DATA SET HAS BEEN GENERATED WITH     ;
* THE FIXED EFFECTS ALPHA AND B AND E AS THE RANDOM       ;
* PARTS OF THE MODEL.  WHILE THE FIXED EFFECTS REMAINED AT      ;
* THESE VALUES, THE RANDOM EFFECTS PARAMETERS TOOK THE VALUES   ;
* AS SHOWN BELLOW:                                              ;
*                                                               ;
*          ------------------------------------------------     ;
*                        INTRA-CLASS                            ;
*          DATA SET      CORRELATION   TAU SQUARE   SIGMA SQUARE;
*          ------------------------------------------------     ;
*                                                               ;
*              1           0.01          1.00          100     ;
*              2           0.05          5.26          100     ;
*              3           0.20         25.00          100     ;
*          ------------------------------------------------     ;
*                                                               ;
```

```
*-------------------------------------------------------------------*;
* K IS A MATRIX WHICH IS PART OF THE PROJECTION MATRIX PW           *;
* GIVEN IN EQUATION 2.20 IN CHAPTER II.  THE MATRIX K IS            *;
* GIVEN BY:                                                         *;
*                                                                   *;
*                    K=INV(X'VWIX)                                  *;
*                                                                   *;
* THE FOLLOWING PROGRAM SEGMENT COMPUTES THE ELEMENTS OF THE        *;
* MATRIX K.                                                         *;
*-------------------------------------------------------------------*;
  K=0;
  K1=0;
  K2=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  Z1J=REPEAT(1,NJ,1);
  CJ=CV[J,1];
  SJ=XJ`*Z1J;
  K1=K1+(XJ`*XJ);
  K2=K2+(CJ*SJ*SJ`);
  M=M+NJ;
END;
  K=W*(K1-K2);
  K=INV(K);
*-------------------------------------------------------------------*;
*                 DETERMINATION OF THE MATRIX FW:                   *;
* Fw IS A (2x2) MATRIX ASSOCIATED WITH WEIGHTS Wk SHOWN             *;
* IN EQUATION 2.21, AND WHOSE ELEMENTS ARE DETERMINED THROUGH       *;
* EQUATION 2.26 THROUGH 2.28.                                       *;
* ALPHAH IS A VECTOR OF THE ESTIMATES OF THE FIXED EFFECTS          *;
* PARAMETERS OF THE MODEL BASED ON THE ORIGINAL DATA SET.           *;
* THUS, THE FOLLOWING PROGRAM SEGMENT DETERMINES THE MATRICES       *;
* USED TO COMPUTE THE USUAL MINQUE ESTIMATES THAT ARE BASED         *;
* ON THE ORIGINAL DATA SET.                                         *;
*-------------------------------------------------------------------*;
  F001=0;
  F002=0;
  F011=0;
  F012=0;
  F111=0;
  F112=0;
  ALPHA1={0,0,0,0};
  ALPHA2={0,0,0,0};
  M=1;
  N1=0;
```

```
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  CJ2=CJ*CJ;
  NJ2=NJ*NJ;
  C2=(1-CJ)*(1-CJ);
  TJ=TRACE(XJ`*XJ*K);
  SJ=XJ`*Z1J;
  AJ=SJ`*K*SJ;
  AC=AJ*CJ;
  CN1=1-CN;
  CN12=CN1*CN1;
  CN13=CN1*CN12;
  AN=AJ*NJ;
  RJ=Z1J`*YJ;
  F001=F001+(NJ*(C2+(CJ2*(NJ-1))));
  F002=F002+(TJ-AC*(CN12+(2-CN)));
  F011=F011+(NJ*CN12);
  F012=F012+(AJ*CN13);
  F111=F111+(NJ2*CN12);
  F112=F112+(AN*CN13);
  ALPHA1=ALPHA1+(K*XJ`*YJ);
  ALPHA2=ALPHA2+(CJ*RJ*K*SJ);
  ALPHAH=ALPHA1-ALPHA2;
  M=M+NJ;
END;
  W2=W*W;
  W3=W2*W;
  F001=W2*F001;
  F002=W3*F002;
  F011=W2*F011;
  F012=W3*F012;
  F111=W2*F111;
  F112=W3*F112;
  F00=F001-F002;
  F01=F011-F012;
  F11=F111-F112;
  ALPHAH=W*ALPHAH;
  ALPHAHT=ALPHAH`;
```

```
*-------------------------------------------------------------------;
*                  DETERMINATION OF THE MATRIX UW:                  ;
* Uw IS A 2 DIMENSIONAL VECTOR OF QUADRATIC FORMS WHOSE             ;
* ELEMENTS ARE DENOTED BY u0 AND u1 (SEE EQUATION 2.22, 2.31        ;
* AND 2.32).                                                       ;
* DETF IS THE DETERMINANT OF THE MATRIX FW USED TO OBTAIN          ;
* THE INVERSE OF THE (2x2) MATRIX Fw.                              ;
* SIGMAH IS THE INTRA-CLASS VARIANCE COMPONENT BASED ON THE        ;
* ORIGINAL HIERARCHICAL DATA SET.                                  ;
* TAUH IS THE INTER-CLASS VARIANCE COMPONENT ESTIMATE BASED        ;
* ON THE ORIGINAL DATA SET.                                        ;
* LAMDA IS THE INTRA-CLASS CORRELATION BASED ON THE ORIGINAL       ;
* SAMPLE AND COMPUTED BY THE FORMULA,                              ;
*                                                                  ;
*                    LAMDA = TAUH/(TAUH+SIGMAH)                    ;
*-------------------------------------------------------------------;
  U01=0;
  U11=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  N2=NJ*NJ;
  DJ=YJ-(XJ*ALPHAH);
  HJ=Z1J`*DJ;
  GJ=DJ`*DJ;
  HJ2=HJ*HJ;
  CH2=CJ*HJ2;
  CN12=(1-CN)*(1-CN);
  DJ=YJ-(XJ*ALPHAH);
  U01=U01+(GJ-(CH2*(2-CN)));
  U11=U11+(HJ2*CN12);
  M=M+NJ;
END;
*-------------------------------------------------------------------;
* THIS MARKS THE END OF THE COMPUTATION OF THE USUAL MINQUE         ;
* ESTIMATES BASED ON THE ORIGINAL SAMPLE.  THE USUAL MINQUE        ;
* ARE PRINTED AT THE FIRST LINE.  ESTIMATES PRINTED AT THE         ;
* PROCEEDING LINES ARE THE BOOTSTRAP REPLICATED ESTIMATES          ;
* BASED ON THE RESAMPLED DATA FROM THE ORIGINAL SAMPLE             ;
*-------------------------------------------------------------------;
  U0=W2*U01;
  U1=W2*U11;
  DETF=(F00*F11)-(F01*F01);
  SIGMAH=((F11*U0)-(F01*U1))/DETF;
  TAUH=((F00*U1)-(F01*U0))/DETF;
```

```
*-------------------------------------------------------------------;
*        A PROCEDURE TO BOOTSTRAP THE PARAMETER ESTIMATES           ;
*     BY COMPUTING THE ESTIMATE B TIMES THROUGH RESAMPLING.         ;
* B IS THE INDEX COUNTER WHICH COUNTS THE BOOTSTRAP REPLICATED      ;
* SAMPLES.  SEED1 IS THE RANDOM GENERATOR FOR THE BOOTSTRAP.        ;
*-------------------------------------------------------------------;
SEED1 = 10199;
DO B=1 TO 200;
*-------------------------------------------------------------------;
* A PROCEDURE USED TO RESAMPLE DATA FROM THE ORIGINAL DATA SET      ;
* BY FIRST CREATING AN INDEX FOR EACH OBSERVATION.  THIS            ;
* PROCESS IS REPEATED B TIMES FOR SOME LARGE B REPRESENTING         ;
* THE NUMBER OF BOOTSTRAP REPLICATIONS.                             ;
* ------------------------------------------------------------------;
NT=1;
CONSTANT=REPEAT(NT,NV[1,],1);
INDEX=NV[1,]*(UNIFORM(REPEAT(SEED1,NV[1,],1)))+CONSTANT;
DO S=2 TO GROUPS;
  NT=NT+NV[S-1,];
  CONSTANT=REPEAT(NT,NV[S,],1);
  INDEX1=NV[S,]*(UNIFORM(REPEAT(SEED1,NV[S,],1)))+CONSTANT;
  INDEX=INDEX//INDEX1;
END;
    INDEX=INT(INDEX);
 YSTAR=Y[INDEX];
 X5STAR=X5[INDEX];
 XSTAR=X4||X5STAR;
 YSTART=YSTAR`;
*-------------------------------------------------------------------;
*   DETERMINE K=INV(X'VWIX) BASED ON THE REPLICATED COVARIATE       ;
*   VALUES                                                          ;
*-------------------------------------------------------------------;
  K=0;
  K1=0;
  K2=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=XSTAR[M:N1,];
  YJ=YSTAR[M:N1,];
  Z1J=REPEAT(1,NJ,1);
  CJ=CV[J,1];
  SJ=XJ`*Z1J;
  K1=K1+(XJ`*XJ);
  K2=K2+(CJ*SJ*SJ`);
  M=M+NJ;
END;
  K=W*(K1-K2);
  K=INV(K);
```

```
*-------------------------------------------------------------------;
*  DETERMINATION OF THE MATRIX FW BASED ON THE REPLICATED K         ;
*  MATRIX.                                                          ;
* ALPHAH1 IS THE ESTIMATE OF THE FIXED EFFECTS PARAMETERS BASED     ;
* ON THE REPLICATED DATA SET.                                       ;
*-------------------------------------------------------------------;
  F001=0;
  F002=0;
  F011=0;
  F012=0;
  F111=0;
  F112=0;
  ALPHA1={0,0,0,0};
  ALPHA2={0,0,0,0};
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=XSTAR[M:N1,];
  YJ=YSTAR[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  CJ2=CJ*CJ;
  NJ2=NJ*NJ;
  C2=(1-CJ)*(1-CJ);
  TJ=TRACE(XJ`*XJ*K);
  SJ=XJ`*Z1J;
  AJ=SJ`*K*SJ;
  AC=AJ*CJ;
  CN1=1-CN;
  CN12=CN1*CN1;
  CN13=CN1*CN12;
  AN=AJ*NJ;
  RJ=Z1J`*YJ;
  F001=F001+(NJ*(C2+(CJ2*(NJ-1))));
  F002=F002+(TJ-AC*(CN12+(2-CN)));
  F011=F011+(NJ*CN12);
  F012=F012+(AJ*CN13);
  F111=F111+(NJ2*CN12);
  F112=F112+(AN*CN13);
  ALPHA1=ALPHA1+(K*XJ`*YJ);
  ALPHA2=ALPHA2+(CJ*RJ*K*SJ);
  ALPHAH1=ALPHA1-ALPHA2;
  M=M+NJ;
END;
  W2=W*W;
  W3=W2*W;
  F001=W2*F001;
  F002=W3*F002;
  F011=W2*F011;
  F012=W3*F012;
```

```
  F111=W2*F111;
  F112=W3*F112;
  F00=F001-F002;
  F01=F011-F012;
  F11=F111-F112;
  ALPHAH1=W*ALPHAH1;
  ALPHAH1T=ALPHAH1`;
*-------------------------------------------------------------------*;
*  DETERMINATION OF THE VECTOR UW BASED ON THE RESAMPLED Y          *;
*  AND THE REPLICATED K MATRIX                                      *;
* SIGMAH1 IS THE INTRA-CLASS VARIANCE COMPONENT ESTIMATE            *;
* BASED ON THE REPLICATED SAMPLE                                    *;
* TAUH1 IS THE INTER-CLASS VARIANCE COMPONENT ESTIMATE BASED        *;
* ON THE REPLICATED SAMPLE.                                         *;
* LAMDA1 IS THE INTRA-CLASS CORRELATION ESTIMATE BASED ON           *;
* THE REPLICATED SAMPLE.                                            *;
*-------------------------------------------------------------------*;
  U01=0;
  U11=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=XSTAR[M:N1,];
  YJ=YSTAR[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  N2=NJ*NJ;
  DJ=YJ-(XJ*ALPHAH1);
  HJ=Z1J`*DJ;
  GJ=DJ`*DJ;
  HJ2=HJ*HJ;
  CH2=CJ*HJ2;
  CN12=(1-CN)*(1-CN);
  DJ=YJ-(XJ*ALPHAH1);
  U01=U01+(GJ-(CH2*(2-CN)));
  U11=U11+(HJ2*CN12);
  M=M+NJ;
END;
  U0=W2*U01;
  U1=W2*U11;
  DETF=(F00*F11)-(F01*F01);
  TAUH1=((F00*U1)-(F01*U0))/DETF;
  SIGMAH1=((F11*U0)-(F01*U1))/DETF;
*-------------------------------------------------------------------*;
* THE NEXT PROGRAM SEGMENT PRINTS THE VALUE OF THE BOOTSTRAP        *;
* AT EACH OF THE B BOOTSTRAP REPLICATION.                           *;
*-------------------------------------------------------------------*;
```

```
  PRINT T (|FORMAT=4.0|) B (|FORMAT=4.0|) TAUH (|FORMAT=8.3|)
      TAUH1 (|FORMAT=8.3|) SIGMAH (|FORMAT=8.3|)
      SIGMAH1 (|FORMAT=8.3|);
  PRINT ALPHAHT (|FORMAT=8.3|) ALPHAH1T (|FORMAT=8.3|);
SEED1 = SEED1 + 100;
END;
*-------------------------------------------------------------------;
*  THE END OF THE BOOTSTRAP TRIAL BASED ON THE RESAMPLED DATA.     ;
*  ANOTHER TRIAL WILL BE PERFORMED AFTER CHANGING THE SEED FOR     ;
*  THE RANDOM SAMPLING ALGORITH.                                   ;
*-------------------------------------------------------------------;
SEED = SEED + 100;
END;
*-------------------------------------------------------------------;
*  THIS MARKS THE END OF THE SIMULATION TRIAL.  EACH SUCH TRIAL    ;
*  RESULTS IN ONE SET OF THE USUAL MINQUE ESTIMATES AND B SETS     ;
*  OF THE BOOTSTRAP REPLICATED ESTIMATES.  THE SUMMARY             ;
*  STATISTICS FOR THE BOOTSTRAP REPLICATED ESTIMATES ARE ALSO      ;
*  COMPUTED.                                                       ;
*-------------------------------------------------------------------;
FINISH;
RUN;
```

PART 2

```
*     COMPUTER PROGRAM TO IMPLEMENT THE BOOTSTRAP ALGORITHM       ;
*      ON A SAMPLE HIERARCHICAL DATA DRAWN FROM A DOUBLE          ;
*         EXPONENTIAL POPULATION OF KNOWN PARAMETERS.             ;
*-------------------------------------------------------------------;
* THE PROGRAM FIRST SETS UP THE Z AND THE FIRST PART OF THE       ;
* X MATRIX EXCLUDING THE COVARIATES.  THE CONSTRUCTION OF         ;
* THESE MATRICES ARE BASED ON THE NUMBER OF OBSERVATIONS IN       ;
* CELL TO SATISFY THE REQUIREMENTS AS IN EQUATION 2.9 IN          ;
* CHAPTER II.  THE PROGRAM IDENTIFIES THE COMPONENTS ON EACH      ;
* AS DEMONSTRATED BY EQUATION 2.11 IN CHAPTER II.                 ;
* THE WEIGHT w1 WAS DETERMINED SEPARATELY USING THE               ;
* HANUSHEK (1974) METHOD.                                         ;
PROC IML;
  START;
  W=1/(1-W1);
    GROUPS=50;
    NV11=REPEAT(20,2,1);
    NV21=REPEAT(25,5,1);
    NV31=REPEAT(30,10,1);
    NV1=NV11//NV21//NV31;
    NV12=REPEAT(35,5,1);
    NV22=REPEAT(40,3,1);
    NV32=REPEAT(20,3,1);
    NV42=REPEAT(25,5,1);
    NV2=NV12//NV22//NV32//NV42;
    NV13=REPEAT(30,10,1);
```

```
    NV23=REPEAT(35,5,1);
    NV33=REPEAT(40,2,1);
    NV3=NV13//NV23//NV33;
    NV=NV1//NV2//NV3;
  CV=W1/(1+(NV-1)*W1);
  X11=REPEAT(1,465,1);
  X12=REPEAT(1,480,1);
  X13=REPEAT(1,555,1);
  X01=REPEAT(0,465,1);
  X02=REPEAT(0,480,1);
  X03=REPEAT(0,555,1);
  X1=X11//X02//X03;
  X2=X01//X12//X03;
  X3=X01//X02//X13;
  X4=X1||X2||X3;
*-------------------------------------------------------------------;
*  PROGRAM SEGMENT TO GENERATE DATA FROM A DOUBLE EXPONENTIAL       ;
*  POPULATION OF SPECIFIED PARAMETER VALUES.                        ;
*  THE PROGRAM FIRST DETERMINES THE FIXED EFFECTS PARAMETERS        ;
*  TOGETHER WITH THE RANDOM EFFECTS PARAMETERS                      ;
*  AND THE COVARIATE WHICH ARE USED IN TURN TO GENERATE THE         ;
*  OBSERVATIONS Y THROUGH THE EQUATION GIVEN BY:                    ;
*                                                                   ;
*                        Y = (X*ALPHA) + B + E                      ;
*                                                                   ;
* SEED1 AND SEED2: ARE ANY NUMBERS USED TO CREATE A RANDOM          ;
*  NUMBER OF OBSERVATIONS FROM A DOUBLE EXPONENTIAL POPULATION.     ;
*                                                                   ;
* T IS THE INDEX COUNTER FOR THE NUMBER OF SIMULATION TRIALS        ;
*-------------------------------------------------------------------;
DO T = 1 TO 400;
  SEED1 = 100999;
  SEED2 = 12399;
*-------------------------------------------------------------------;
* T IS THE INDEX COUNTER FOR THE NUMBER OF SIMULATION TRIALS        ;
*-------------------------------------------------------------------;
DO T = 1 TO 1;
  UR1=UNIFORM(REPEAT(SEED1,GROUPS,1));
  UR2=UNIFORM(REPEAT(SEED2,GROUPS,1));
  LR= -1 * LOG(UR1);
  TR1 = REPEAT(1,GROUPS,1);
  TR2 = TR1 # (UR2 >= 0.5);
  TR3 = -1*(TR1 # (UR2 < 0.5));
  TR4 = TR2 + TR3;
  REFFECTS = 2.2935 * ((LR#TR4)/SQRT(2));
*-------------------------------------------------------------------;
    B1 = REFFECTS[1,1];
    N1 = NV[1,1];
    BJ1 = REPEAT(B1,N1,1);
DO I = 2 TO GROUPS;
    BJ = REFFECTS[I,1];
    N =  NV[I,1];
    BJ1=BJ1//REPEAT(BJ,N,1);
END;
```

```
    B=BJ1;
    E = 10 * NORMAL(REPEAT(SEED,1500,1));
    X41 = REPEAT(25,1500,1);
    X5 = INT(75 * UNIFORM(REPEAT(SEED,1500,1))) + X41;
    X=X4||X5;
    ALPHA = {-5,2,3,1.0};
    Y = (X*ALPHA) + B + E;
*-------------------------------------------------------------------;
* AT THIS POINT A SPECIFIC DATA SET HAS BEEN GENERATED WITH        ;
* THE FIXED EFFECTS ALPHA AND B AND E AS THE RANDOM                ;
* PARTS OF THE MODEL.  WHILE THE FIXED EFFECTS REMAINED AT         ;
* THESE VALUES, THE RANDOM EFFECTS PARAMETERS TOOK THE VALUES      ;
* AS SHOWN BELLOW:                                                 ;
*                                                                  ;
*          ------------------------------------------------------  ;
*                        INTRA-CLASS                               ;
*          DATA SET      CORRELATION    TAU SQUARE    SIGMA SQUARE ;
*          ------------------------------------------------------  ;
*                                                                  ;
*              1            0.01           1.00          100      ;
*              2            0.05           5.26          100      ;
*              3            0.20          25.00          100      ;
*          ------------------------------------------------------  ;
*                                                                  ;
* K IS A MATRIX WHICH IS PART OF THE PROJECTION MATRIX PW          ;
* GIVEN IN EQUATION 2.20 IN CHAPTER II.  THE MATRIX K IS           ;
* GIVEN BY:                                                        ;
*                                                                  ;
*                    K=INV(X'VWIX)                                 ;
*                                                                  ;
* THE FOLLOWING PROGRAM SEGMENT COMPUTES THE ELEMENTS OF THE       ;
* MATRIX K.                                                        ;
*-------------------------------------------------------------------;
  K=0;
  K1=0;
  K2=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  Z1J=REPEAT(1,NJ,1);
  CJ=CV[J,1];
  SJ=XJ`*Z1J;
  K1=K1+(XJ`*XJ);
  K2=K2+(CJ*SJ*SJ`);
  M=M+NJ;
END;
  K=W*(K1-K2);
  K=INV(K);
```

```
*-------------------------------------------------------------*;
*                DETERMINATION OF THE MATRIX FW:              *;
* Fw IS A (2x2) MATRIX ASSOCIATED WITH WEIGHTS Wk SHOWN       *;
* IN EQUATION 2.21, AND WHOSE ELEMENTS ARE DETERMINED THROUGH *;
* EQUATION 2.26 THROUGH 2.28.                                 *;
* ALPHAH IS A VECTOR OF THE ESTIMATES OF THE FIXED EFFECTS    *;
* PARAMETERS OF THE MODEL BASED ON THE ORIGINAL DATA SET.     *;
* THUS, THE FOLLOWING PROGRAM SEGMENT DETERMINES THE MATRICES *;
* USED TO COMPUTE THE USUAL MINQUE ESTIMATES THAT ARE BASED   *;
* ON THE ORIGINAL DATA SET.                                   *;
*-------------------------------------------------------------*;
  F001=0;
  F002=0;
  F011=0;
  F012=0;
  F111=0;
  F112=0;
  ALPHA1={0,0,0,0};
  ALPHA2={0,0,0,0};
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  CJ2=CJ*CJ;
  NJ2=NJ*NJ;
  C2=(1-CJ)*(1-CJ);
  TJ=TRACE(XJ`*XJ*K);
  SJ=XJ`*Z1J;
  AJ=SJ`*K*SJ;
  AC=AJ*CJ;
  CN1=1-CN;
  CN12=CN1*CN1;
  CN13=CN1*CN12;
  AN=AJ*NJ;
  RJ=Z1J`*YJ;
  F001=F001+(NJ*(C2+(CJ2*(NJ-1))));
  F002=F002+(TJ-AC*(CN12+(2-CN)));
  F011=F011+(NJ*CN12);
  F012=F012+(AJ*CN13);
  F111=F111+(NJ2*CN12);
  F112=F112+(AN*CN13);
  ALPHA1=ALPHA1+(K*XJ`*YJ);
  ALPHA2=ALPHA2+(CJ*RJ*K*SJ);
  ALPHAH=ALPHA1-ALPHA2;
  M=M+NJ;
END;
```

```
  W2=W*W;
  W3=W2*W;
  F001=W2*F001;
  F002=W3*F002;
  F011=W2*F011;
  F012=W3*F012;
  F111=W2*F111;
  F112=W3*F112;
  F00=F001-F002;
  F01=F011-F012;
  F11=F111-F112;
  ALPHAH=W*ALPHAH;
  ALPHAHT=ALPHAH`;
*-------------------------------------------------------------------;
*                DETERMINATION OF THE MATRIX UW:                    ;
* Uw IS A 2 DIMENSIONAL VECTOR OF QUADRATIC FORMS WHOSE             ;
* ELEMENTS ARE DENOTED BY u0 AND u1 (SEE EQUATION 2.22, 2.31        ;
* AND 2.32).                                                        ;
* DETF IS THE DETERMINANT OF THE MATRIX FW USED TO OBTAIN           ;
* THE INVERSE OF THE (2x2) MATRIX Fw.                               ;
* SIGMAH IS THE INTRA-CLASS VARIANCE COMPONENT BASED ON THE         ;
* ORIGINAL HIERARCHICAL DATA SET.                                   ;
* TAUH IS THE INTER-CLASS VARIANCE COMPONENT ESTIMATE BASED         ;
* ON THE ORIGINAL DATA SET.                                         ;
* LAMDA IS THE INTRA-CLASS CORRELATION BASED ON THE ORIGINAL        ;
* SAMPLE AND COMPUTED BY THE FORMULA,                               ;
*                                                                   ;
*                    LAMDA = TAUH/(TAUH+SIGMAH)                     ;
*-------------------------------------------------------------------;
  U01=0;
  U11=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  N2=NJ*NJ;
  DJ=YJ-(XJ*ALPHAH);
  HJ=Z1J`*DJ;
  GJ=DJ`*DJ;
  HJ2=HJ*HJ;
  CH2=CJ*HJ2;
  CN12=(1-CN)*(1-CN);
  DJ=YJ-(XJ*ALPHAH);
  U01=U01+(GJ-(CH2*(2-CN)));
  U11=U11+(HJ2*CN12);
  M=M+NJ;
END;
```

```
*-------------------------------------------------------------------;
* THIS MARKS THE END OF THE COMPUTATION OF THE USUAL MINQUE        ;
* ESTIMATES BASED ON THE ORIGINAL SAMPLE.  THE USUAL MINQUE        ;
* ARE PRINTED AT THE FIRST LINE.  ESTIMATES PRINTED AT THE         ;
* PROCEEDING LINES ARE THE BOOTSTRAP REPLICATED ESTIMATES          ;
* BASED ON THE RESAMPLED DATA FROM THE ORIGINAL SAMPLE             ;
*-------------------------------------------------------------------;
  U0=W2*U01;
  U1=W2*U11;
  DETF=(F00*F11)-(F01*F01);
  SIGMAH=((F11*U0)-(F01*U1))/DETF;
  TAUH=((F00*U1)-(F01*U0))/DETF;
*-------------------------------------------------------------------;
*        A PROCEDURE TO BOOTSTRAP THE PARAMETER ESTIMATES           ;
*     BY COMPUTING THE ESTIMATE B TIMES THROUGH RESAMPLING.         ;
* B IS THE INDEX COUNTER WHICH COUNTS THE BOOTSTRAP REPLICATED      ;
* SAMPLES.  SEED1 IS THE RANDOM GENERATOR FOR THE BOOTSTRAP.       ;
*-------------------------------------------------------------------;
SEED1 = 10199;
DO B=1 TO 200;
*-------------------------------------------------------------------;
* A PROCEDURE USED TO RESAMPLE DATA FROM THE ORIGINAL DATA SET      ;
* BY FIRST CREATING AN INDEX FOR EACH OBSERVATION.  THIS            ;
* PROCESS IS REPEATED B TIMES FOR SOME LARGE B REPRESENTING         ;
* THE NUMBER OF BOOTSTRAP REPLICATIONS.                             ;
* ------------------------------------------------------------------;
NT=1;
CONSTANT=REPEAT(NT,NV[1,],1);
INDEX=NV[1,]*(UNIFORM(REPEAT(SEED1,NV[1,],1)))+CONSTANT;
DO S=2 TO GROUPS;
  NT=NT+NV[S-1,];
  CONSTANT=REPEAT(NT,NV[S,],1);
  INDEX1=NV[S,]*(UNIFORM(REPEAT(SEED1,NV[S,],1)))+CONSTANT;
  INDEX=INDEX//INDEX1;
END;
    INDEX=INT(INDEX);
 YSTAR=Y[INDEX];
 X5STAR=X5[INDEX];
 XSTAR=X4||X5STAR;
 YSTART=YSTAR`;
*-------------------------------------------------------------------;
*   DETERMINE K=INV(X'VWIX) BASED ON THE REPLICATED COVARIATE       ;
*   VALUES                                                          ;
*-------------------------------------------------------------------;
  K=0;
  K1=0;
  K2=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=XSTAR[M:N1,];
  YJ=YSTAR[M:N1,];
```

```
  Z1J=REPEAT(1,NJ,1);
  CJ=CV[J,1];
  SJ=XJ`*Z1J;
  K1=K1+(XJ`*XJ);
  K2=K2+(CJ*SJ*SJ`);
  M=M+NJ;
END;
  K=W*(K1-K2);
  K=INV(K);
*-------------------------------------------------------------------;
*  DETERMINATION OF THE MATRIX FW BASED ON THE REPLICATED K        ;
*  MATRIX.                                                         ;
* ALPHAH1 IS THE ESTIMATE OF THE FIXED EFFECTS PARAMETERS BASED    ;
* ON THE REPLICATED DATA SET.                                      ;
*-------------------------------------------------------------------;
  F001=0;
  F002=0;
  F011=0;
  F012=0;
  F111=0;
  F112=0;
  ALPHA1={0,0,0,0};
  ALPHA2={0,0,0,0};
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=XSTAR[M:N1,];
  YJ=YSTAR[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  CJ2=CJ*CJ;
  NJ2=NJ*NJ;
  C2=(1-CJ)*(1-CJ);
  TJ=TRACE(XJ`*XJ*K);
  SJ=XJ`*Z1J;
  AJ=SJ`*K*SJ;
  AC=AJ*CJ;
  CN1=1-CN;
  CN12=CN1*CN1;
  CN13=CN1*CN12;
  AN=AJ*NJ;
  RJ=Z1J`*YJ;
  F001=F001+(NJ*(C2+(CJ2*(NJ-1))));
  F002=F002+(TJ-AC*(CN12+(2-CN)));
  F011=F011+(NJ*CN12);
  F012=F012+(AJ*CN13);
  F111=F111+(NJ2*CN12);
  F112=F112+(AN*CN13);
  ALPHA1=ALPHA1+(K*XJ`*YJ);
  ALPHA2=ALPHA2+(CJ*RJ*K*SJ);
```

```
  ALPHAH1=ALPHA1-ALPHA2;
  M=M+NJ;
END;
  W2=W*W;
  W3=W2*W;
  F001=W2*F001;
  F002=W3*F002;
  F011=W2*F011;
  F012=W3*F012;
  F111=W2*F111;
  F112=W3*F112;
  F00=F001-F002;
  F01=F011-F012;
  F11=F111-F112;
  ALPHAH1=W*ALPHAH1;
  ALPHAH1T=ALPHAH1`;
*-------------------------------------------------------------------;
*  DETERMINATION OF THE VECTOR UW BASED ON THE RESAMPLED Y         ;
*  AND THE REPLICATED K MATRIX                                     ;
* SIGMAH1 IS THE INTRA-CLASS VARIANCE COMPONENT ESTIMATE           ;
* BASED ON THE REPLICATED SAMPLE                                   ;
* TAUH1 IS THE INTER-CLASS VARIANCE COMPONENT ESTIMATE BASED       ;
* ON THE REPLICATED SAMPLE.                                        ;
* LAMDA1 IS THE INTRA-CLASS CORRELATION ESTIMATE BASED ON          ;
* THE REPLICATED SAMPLE.                                           ;
*-------------------------------------------------------------------;
  U01=0;
  U11=0;
  M=1;
  N1=0;
DO J=1 TO GROUPS;
  NJ=NV[J,1];
  N1=N1+NJ;
  XJ=XSTAR[M:N1,];
  YJ=YSTAR[M:N1,];
  CJ=CV[J,1];
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  N2=NJ*NJ;
  DJ=YJ-(XJ*ALPHAH1);
  HJ=Z1J`*DJ;
  GJ=DJ`*DJ;
  HJ2=HJ*HJ;
  CH2=CJ*HJ2;
  CN12=(1-CN)*(1-CN);
  DJ=YJ-(XJ*ALPHAH1);
  U01=U01+(GJ-(CH2*(2-CN)));
  U11=U11+(HJ2*CN12);
  M=M+NJ;
END;
```

```
  U0=W2*U01;
  U1=W2*U11;
  DETF=(F00*F11)-(F01*F01);
  TAUH1=((F00*U1)-(F01*U0))/DETF;
  SIGMAH1=((F11*U0)-(F01*U1))/DETF;
*-------------------------------------------------------------;
* THIS PROGRAM SEGMENT PRINTS THE VALUE OF THE BOOTSTRAP      ;
* AT EACH OF THE B BOOTSTRAP REPLICATION.                     ;
*-------------------------------------------------------------;
  PRINT T (|FORMAT=4.0|) B (|FORMAT=4.0|) TAUH (|FORMAT=8.3|)
      TAUH1 (|FORMAT=8.3|) SIGMAH (|FORMAT=8.3|)
      SIGMAH1 (|FORMAT=8.3|);
  PRINT ALPHAHT (|FORMAT=8.3|) ALPHAH1T (|FORMAT=8.3|);
SEED1 = SEED1 + 100;
SEED2 = SEED2 + 100;
END;
*-------------------------------------------------------------;
*  THE END OF THE BOOTSTRAP TRIAL BASED ON THE RESAMPLED DATA.;
*  ANOTHER TRIAL WILL BE PERFORMED AFTER CHANGING THE SEED FOR;
*  THE RANDOM SAMPLING ALGORITH.                              ;
*-------------------------------------------------------------;
END;
*-------------------------------------------------------------;
*  THIS MARKS THE END OF THE SIMULATION TRIAL.  EACH SUCH TRIAL;
*  RESULTS IN ONE SET OF THE USUAL MINQUE ESTIMATES AND B SETS;
*  OF THE BOOTSTRAP REPLICATED ESTIMATES.  THE SUMMARY        ;
*  STATISTICS FOR THE BOOTSTRAP REPLICATED ESTIMATES ARE ALSO ;
*  COMPUTED.                                                  ;
*-------------------------------------------------------------;
FINISH;
RUN;
```

PART 3

```
*    COMPUTER PROGRAM TO SIMULATE THE SAMPLING DISTRIBUTION   ;
*    OF THE MINQUE ESTIMATE FOR A SAMPLE DRAWN FROM A         ;
*    NORMAL POPULATION OF KNOWN PARAMETERS.                   ;
*-------------------------------------------------------------;
* THE PROGRAM FIRST SETS UP THE Z AND THE FIRST PART OF THE   ;
* X MATRIX EXCLUDING THE COVARIATES.  THE CONSTRUCTION OF     ;
* THESE MATRICES ARE BASED ON THE NUMBER OF OBSERVATIONS IN   ;
* CELL TO SATISFY THE REQUIREMENTS AS IN EQUATION 2.9 IN      ;
* CHAPTER II.  THE PROGRAM IDENTIFIES THE COMPONENTS ON EACH  ;
* AS DEMONSTRATED BY EQUATION 2.11 IN CHAPTER II.             ;
* THE WEIGHT w1 WAS DETERMINED SEPARATELY USING THE           ;
* HANUSHEK (1974) METHOD.                                     ;
*-------------------------------------------------------------;
*                                                             ;
```

```
PROC IML;
  START;
  W=1/(1-W1);
    GROUPS=50;
    NV11=REPEAT(20,2,1);
    NV21=REPEAT(25,5,1);
    NV31=REPEAT(30,10,1);
    NV1=NV11//NV21//NV31;
    NV12=REPEAT(35,5,1);
    NV22=REPEAT(40,3,1);
    NV32=REPEAT(20,3,1);
    NV42=REPEAT(25,5,1);
    NV2=NV12//NV22//NV32//NV42;
    NV13=REPEAT(30,10,1);
    NV23=REPEAT(35,5,1);
    NV33=REPEAT(40,2,1);
    NV3=NV13//NV23//NV33;
    NV=NV1//NV2//NV3;
  CV=W1/(1+(NV-1)*W1);
  X11=REPEAT(1,465,1);
  X12=REPEAT(1,480,1);
  X13=REPEAT(1,555,1);
  X01=REPEAT(0,465,1);
  X02=REPEAT(0,480,1);
  X03=REPEAT(0,555,1);
  X1=X11//X02//X03;
  X2=X01//X12//X03;
  X3=X01//X02//X13;
*-------------------------------------------------------------------;
*  PROGRAM SEGMENT TO GENERATE DATA FROM A NORMAL POPULATION       ;
*  OF SPECIFIED PARAMETER VALUES.  THE PROGRAM FIRST DETERMINES    ;
*  THE FIXED EFFECTS PARAMETERS AND THE COVARIATE WHICH IN         ;
*  ARE USED IN TURN TO GENERATE THE OBSERVATIONS Y THROUGH         ;
*  THE EQUATION GIVEN BY:                                          ;
*                                                                  ;
*                        Y = (X*ALPHA) + B + E                     ;
*                                                                  ;
* SEED: IS ANY NUMBER USED TO CREATE A RANDOM NUMBER OF            ;
*  OBSERVATIONS FROM SOME POPULATION.                              ;

    SEED = 10199;
*                                                                  ;
*TIS THE INDEX COUNTER FOR THE NUMBER OF SIMULATION TRIALS         ;
*-------------------------------------------------------------------;
    SEED = 199;
DO T = 1 TO 1000;
    REFFECTS = 2.2935 * NORMAL(REPEAT(SEED,GROUPS,1));
    B1 = REFFECTS[1,1];
    N1 = NV[1,1];
    BJ1 = REPEAT(B1,N1,1);
```

```
DO I = 2 TO GROUPS;
    BJ = REFFECTS[I,1];
    N =  NV[I,1];
    BJ1=BJ1//REPEAT(BJ,N,1);
END;
    B=BJ1;
    E = 10 * NORMAL(REPEAT(SEED,1500,1));
    X41 = REPEAT(25,1500,1);
    X4 = INT(75 * UNIFORM(REPEAT(SEED,1500,1))) + X41;
    X=X1||X2||X3||X4;
    ALPHA = {-5,2,3,1.0};
    Y = (X*ALPHA) + B + E;
*-------------------------------------------------------------------;
*                  DETERMINE K=INV(X'VWIX)                          ;
*-------------------------------------------------------------------;
  K=0;
  K1=0;
  K2=0;
  M=1;
  N1=NV[1,1];
DO J=1 TO GROUPS;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  NJ=NROW(YJ);
  Z1J=REPEAT(1,NJ,1);
  CJ=CV[J,1];
  SJ=XJ`*Z1J;
  K1=K1+(XJ`*XJ);
  K2=K2+(CJ*SJ*SJ`);
  M=M+NV[J,1];
  N1=N1+NV[J,1];
END;
  K=W*(K1-K2);
  K=INV(K);
*-------------------------------------------------------------------;
*                DETERMINATION OF THE MATRIX FW                     ;
*-------------------------------------------------------------------;
  F001=0;
  F002=0;
  F011=0;
  F012=0;
  F111=0;
  F112=0;
  ALPHA1={0,0,0,0};
  ALPHA2={0,0,0,0};
  M=1;
  N1=NV[1,1];
DO J=1 TO GROUPS;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  NJ=NROW(YJ);
```

```
   Z1J=REPEAT(1,NJ,1);
   CN=CJ*NJ;
   CJ2=CJ*CJ;
   NJ2=NJ*NJ;
   C2=(1-CJ)*(1-CJ);
   TJ=TRACE(XJ`*XJ*K);
   SJ=XJ`*Z1J;
   AJ=SJ`*K*SJ;
   AC=AJ*CJ;
   CN1=1-CN;
   CN12=CN1*CN1;
   CN13=CN1*CN12;
   AN=AJ*NJ;
   RJ=Z1J`*YJ;
   F001=F001+(NJ*(C2+(CJ2*(NJ-1))));
   F002=F002+(TJ-AC*(CN12+(2-CN)));
   F011=F011+(NJ*CN12);
   F012=F012+(AJ*CN13);
   F111=F111+(NJ2*CN12);
   F112=F112+(AN*CN13);
   ALPHA1=ALPHA1+(K*XJ`*YJ);
   ALPHA2=ALPHA2+(CJ*RJ*K*SJ);
   ALPHAH=ALPHA1-ALPHA2;
   M=M+NV[J,1];
   N1=N1+NV[J,1];
END;
   W2=W*W;
   W3=W2*W;
   F001=W2*F001;
   F002=W3*F002;
   F011=W2*F011;
   F012=W3*F012;
   F111=W2*F111;
   F112=W3*F112;
   F00=F001-F002;
   F01=F011-F012;
   F11=F111-F112;
   ALPHAH=W*ALPHAH;
   ALPHAHT=ALPHAH`;
*-------------------------------------------------------------------;
*                DETERMINATION OF THE MATRIX UW                     ;
*-------------------------------------------------------------------;
   U01=0;
   U11=0;
   M=1;
   N1=NV[1,1];
DO J=1 TO GROUPS;
   XJ=X[M:N1,];
   YJ=Y[M:N1,];
   CJ=CV[J,1];
   NJ=NROW(YJ);
   Z1J=REPEAT(1,NJ,1);
   CN=CJ*NJ;
```

```
  N2=NJ*NJ;
  DJ=YJ-(XJ*ALPHAH);
  HJ=Z1J`*DJ;
  GJ=DJ`*DJ;
  HJ2=HJ*HJ;
  CH2=CJ*HJ2;
  CN12=(1-CN)*(1-CN);
  DJ=YJ-(XJ*ALPHAH);
  U01=U01+(GJ-(CH2*(2-CN)));
  U11=U11+(HJ2*CN12);
  M=M+NV[J,1];
  N1=N1+NV[J,1];
  SEED = SEED + 10;
END;
  U0=W2*U01;
  U1=W2*U11;
  DETF=(F00*F11)-(F01*F01);
  SIGMAH=((F11*U0)-(F01*U1))/DETF;
  TAUH=((F00*U1)-(F01*U0))/DETF;
  LAMDA=TAUH/(TAUH+SIGMAH);
  PRINT S TAUH SIGMAH LAMDA ALPHAHT;
END;
PRINT SEED;
FINISH;
RUN;
```

PART 4

```
*     COMPUTER PROGRAM TO SIMULATE THE SAMPLING DISTRIBUTION     ;
*     OF THE MINQUE ESTIMATE FOR A SAMPLE DRAWN FROM A           ;
*     DOUBLE EXPONENTIAL POPULATION OF KNOWN PARAMETERS.         ;
*-------------------------------------------------------------------;
* THE PROGRAM FIRST SETS UP THE Z AND THE FIRST PART OF THE      ;
* X MATRIX EXCLUDING THE COVARIATES.  THE CONSTRUCTION OF        ;
* THESE MATRICES ARE BASED ON THE NUMBER OF OBSERVATIONS IN      ;
* CELL TO SATISFY THE REQUIREMENTS AS IN EQUATION 2.9 IN         ;
* CHAPTER II.  THE PROGRAM IDENTIFIES THE COMPONENTS ON EACH     ;
* AS DEMONSTRATED BY EQUATION 2.11 IN CHAPTER II.                ;
* THE WEIGHT w1 WAS DETERMINED SEPARATELY USING THE              ;
* HANUSHEK (1974) METHOD.                                        ;
*-------------------------------------------------------------------;
*                                                                  ;
PROC IML;
  START;
  W=1/(1-W1);
    GROUPS=50;
    NV11=REPEAT(20,2,1);
    NV21=REPEAT(25,5,1);
    NV31=REPEAT(30,10,1);
    NV1=NV11//NV21//NV31;
    NV12=REPEAT(35,5,1);
    NV22=REPEAT(40,3,1);
    NV32=REPEAT(20,3,1);
```

```
    NV42=REPEAT(25,5,1);
    NV2=NV12//NV22//NV32//NV42;
    NV13=REPEAT(30,10,1);
    NV23=REPEAT(35,5,1);
    NV33=REPEAT(40,2,1);
    NV3=NV13//NV23//NV33;
    NV=NV1//NV2//NV3;
  CV=W1/(1+(NV-1)*W1);
  X11=REPEAT(1,465,1);
  X12=REPEAT(1,480,1);
  X13=REPEAT(1,555,1);
  X01=REPEAT(0,465,1);
  X02=REPEAT(0,480,1);
  X03=REPEAT(0,555,1);
  X1=X11//X02//X03;
  X2=X01//X12//X03;
  X3=X01//X02//X13;
*-------------------------------------------------------------------;
*  PROGRAM SEGMENT TO GENERATE DATA FROM A DOUBLE EXPONENTIAL      ;
*  POPULATION OF SPECIFIED PARAMETER VALUES.                       ;
*-------------------------------------------------------------------;
  SEED1 = 10199;
  SEED2 = 11099;
DO S = 1 TO 1000;
  UR1=UNIFORM(REPEAT(SEED1,GROUPS,1));
  UR2=UNIFORM(REPEAT(SEED2,GROUPS,1));
  LR= -1 * LOG(UR1);
  TR1 = REPEAT(1,GROUPS,1);
  TR2 = TR1 # (UR2 >= 0.5);
  TR3 = -1*(TR1 # (UR2 < 0.5));
  TR4 = TR2 + TR3;
  REFFECTS = 2.2935 * ((LR#TR4)/SQRT(2));
    B1 = REFFECTS[1,1];
    N1 = NV[1,1];
    BJ1 = REPEAT(B1,N1,1);
DO I = 2 TO GROUPS;
    BJ = REFFECTS[I,1];
    N =  NV[I,1];
    BJ1=BJ1//REPEAT(BJ,N,1);
END;
    B=BJ1;
  UE1=UNIFORM(REPEAT(SEED1,1500,1));
  UE2=UNIFORM(REPEAT(SEED2,1500,1));
  LE= -1 * LOG(UE1);
  TE1 = REPEAT(1,1500,1);
  TE2 = TE1 # (UE2 >= 0.5);
  TE3 = -1*(TE1 # (UE2 < 0.5));
  TE4 = TE2 + TE3;
  E = 10 * ((LE#TE4)/SQRT(2));
    X41 = REPEAT(25,1500,1);
    X4 = INT(75 * UNIFORM(REPEAT(SEED1,1500,1))) + X41;
    X=X1||X2||X3||X4;
```

```
    ALPHA = {-5,2,3,1.0};
    Y = (X*ALPHA) + B + E;
*-------------------------------------------------------------;
*                  DETERMINE K=INV(X'VWIX)                    ;
*-------------------------------------------------------------;
  K=0;
  K1=0;
  K2=0;
  M=1;
  N1=NV[1,1];
DO J=1 TO GROUPS;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  NJ=NROW(YJ);
  Z1J=REPEAT(1,NJ,1);
  CJ=CV[J,1];
  SJ=XJ`*Z1J;
  K1=K1+(XJ`*XJ);
  K2=K2+(CJ*SJ*SJ`);
  M=M+NV[J,1];
  N1=N1+NV[J,1];
END;
  K=W*(K1-K2);
  K=INV(K);


*-------------------------------------------------------------;
*                DETERMINATION OF THE MATRIX FW               ;
*-------------------------------------------------------------;
  F001=0;
  F002=0;
  F011=0;
  F012=0;
  F111=0;
  F112=0;
  ALPHA1={0,0,0,0};
  ALPHA2={0,0,0,0};
  M=1;
  N1=NV[1,1];
DO J=1 TO GROUPS;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  NJ=NROW(YJ);
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  CJ2=CJ*CJ;
  NJ2=NJ*NJ;
  C2=(1-CJ)*(1-CJ);
  TJ=TRACE(XJ`*XJ*K);
  SJ=XJ`*Z1J;
  AJ=SJ`*K*SJ;
  AC=AJ*CJ;
```

```
  CN1=1-CN;
  CN12=CN1*CN1;
  CN13=CN1*CN12;
  AN=AJ*NJ;
  RJ=Z1J`*YJ;
  F001=F001+(NJ*(C2+(CJ2*(NJ-1))));
  F002=F002+(TJ-AC*(CN12+(2-CN)));
  F011=F011+(NJ*CN12);
  F012=F012+(AJ*CN13);
  F111=F111+(NJ2*CN12);
  F112=F112+(AN*CN13);
  ALPHA1=ALPHA1+(K*XJ`*YJ);
  ALPHA2=ALPHA2+(CJ*RJ*K*SJ);
  ALPHAH=ALPHA1-ALPHA2;
  M=M+NV[J,1];
  N1=N1+NV[J,1];
END;
  W2=W*W;
  W3=W2*W;
  F001=W2*F001;
  F002=W3*F002;
  F011=W2*F011;
  F012=W3*F012;
  F111=W2*F111;
  F112=W3*F112;
  F00=F001-F002;
  F01=F011-F012;
  F11=F111-F112;
  ALPHAH=W*ALPHAH;
  ALPHAHT=ALPHAH`;
*-------------------------------------------------------------------;
*                 DETERMINATION OF THE MATRIX UW                    ;
*-------------------------------------------------------------------;
  U01=0;
  U11=0;
  M=1;
  N1=NV[1,1];
DO J=1 TO GROUPS;
  XJ=X[M:N1,];
  YJ=Y[M:N1,];
  CJ=CV[J,1];
  NJ=NROW(YJ);
  Z1J=REPEAT(1,NJ,1);
  CN=CJ*NJ;
  N2=NJ*NJ;
  DJ=YJ-(XJ*ALPHAH);
  HJ=Z1J`*DJ;
  GJ=DJ`*DJ;
  HJ2=HJ*HJ;
  CH2=CJ*HJ2;
  CN12=(1-CN)*(1-CN);
  DJ=YJ-(XJ*ALPHAH);
```

```
   U01=U01+(GJ-(CH2*(2-CN)));
   U11=U11+(HJ2*CN12);
   M=M+NV[J,1];
   N1=N1+NV[J,1];
   SEED1 = SEED1 + 100;
   SEED2 = SEED2 + 100;
END;
   U0=W2*U01;
   U1=W2*U11;
   DETF=(F00*F11)-(F01*F01);
   SIGMAH=((F11*U0)-(F01*U1))/DETF;
   TAUH=((F00*U1)-(F01*U0))/DETF;
   LAMDA=TAUH/(TAUH+SIGMAH);
   PRINT T TAUH SIGMAH LAMDA ALPHAHT;
END;
FINISH;
RUN;
```

BIBLIOGRAPHY

Abramovitch, L., and Singh, K. (1985). Edgeworth corrected pivital statistics and the bootstrap. The Annals of Statistics, 13, 1, 116–132.

Aitkin, M., and Longford, N. (1986). Statistical modeling issues in school effectiveness studies. Journal of the Royal Statistical Society, (Series A), 149, 1–43.

Arlin, M. (1984b). Time, equality, and mastery learning. Review of Educational Research, 541, 65–86.

Arlin, M., & Webster, J. (1983). Time cost of mastery learning. Journal of Educational Psychology, 75, 187–195.

Ashton, P., & Webb, R. (1986). Making a Difference: Teachers' Sense of Efficacy and Student Achievement. New York: Longman.

Bagaka's J. G. (1989). Empirical Bayes Bootstrap: Estimation of parameter distribution through the bootstrap in hierarchical data. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, New Jersey: Prentice–Hall.

Bangert–Drowns, R. L. (1986). A review of developments in meta–analytic method. Psychological Bulletin, 99, 388–399.

Beran, R. (1984). Jackknife approximations to bootstrap estimates. The Annals of Statistics, 12(1), 101–118.

Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. The Annals of Statistics, 9(6), 1196–1217.

Block, J. H. (Ed.) (1971). Mastery Learning: Theory and Practice. New York: Holt, Rinehart and Winston.

Bloom, B. S. (1982). Human Characteristics and School Learning. New York: McGraw–Hill Book Company.

Bloom, B. S. (1984a). The 2 Sigma Problem: The search for methods of instruction as effective as one–to–one tutoring. Educational Researcher, 13, 6, 4–16.

Bloom, B. S. (1984b). The search for methods of group instruction as effective as one–to–one tutoring. Educational Leadership, 41(8), 4–17.

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations. Journal of Royal Statistical Society, B26, 211–246.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., Wisenbaker, J. (1979). School Social Systems and Student Achievement: Schools Can Make A Difference. New York: Praeger.

Brookover, W., Beamer, L., Efthim, H., Hathaway, P., Lezotte, L., Miller, S., and Passalacqua, J. (1982). Creating Effective Schools. Learning Publication.

Brown, K.G. (1976). Asymptotic behavior of MINQUE–type estimators of variance components. Annals of Statistics, 4, 746–754.

Burstein, L., Linn, R. L., & Campell, F. (1978). Analyzing multi–level data in the presence of heterogeneous within–class regressions. Journal of Educational Statistics, 3 (4), 347–389.

Burstein, L., & Miller, M.D. (1980). Regression–based analysis of multi–level educational data. New Directions for Methodology of Social and Behavioral Sciences, 6, 194–211.

Cochran, W. G. (1977). Sampling Technique. New York: John Wiley & Sons.

Corbeil, R. R., and Searle, R. S. (1976). A comparison of of variance component estimators. Biometrics, 32, 779–791.

Corno, L., & Snow, R. E. (1985). Adapting teaching to individual differences among learners. In M. Wittlock (ed.), Handbook of Research on Teaching. 3rd ed., Chicago, IL: Rand NcNally.

Carrol, J. B. (1963). A model for school learning. Teachers College Record, 64, 723–733.

Carroll, R. (1979). On estimating variances of robust estimates when the errors are asymmetric. Journal of American Statistics, 74, 674–679.

Dembo, M.H., & Gibson, S. (1985). Teachers' Sense of Efficacy: An important factor in school improvement. The Elementary School Journal, 86, 2, 173–184.

Diaconis, P., & Efron, B. (1983). Computer–intensive methods in statistics. Scientific American, 248, 116–130.

Dicoccio, T., and Tibshirani, R. (1987). Bootstrap confidence intervals and bootstrap approximations. Journal of American Statistical Association, 82, 397, 163–170.

Dolker, M., Halperin, S., & Divgi, D. R. (1982). Problems with bootstrapping Pearson correlations in very small bivariate samples. Psychometrics, 47 (4), 529–530.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7, 1–26.

Efron, B. (1981a). Nonparametric standard errors and confidence intervals. The Canadian Journal of Statistics, 9 (2), 139–172.

Efron, B. (1981b). Censored data and the bootstrap. Journal of American Statistical Association, 76, 374, 312–319.

Efron, B. (1982). The jackknife, the bootstrap, and other sampling plans. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1982). Transformation theory: How normal is a family of distributions? The Annals of Statistics, 10, 2, 323–339.

Efron, B., & Gong, C. (1983). A leisurely look at the bootstrap, the jackknife, and cross–validation. The American Statistician, 37 (1), 36–48.

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science, 1, 1, 54–77.

Freedman, D. A. (1981). Bootstrapping regression models. Annals of Statistics, 9(6), 1218–1228.

Fuller, B., Wood, K., Rapoport, T., & Dombusch, S.M. (1982). The organizational context of individual efficacy. Review of Educational Research, 52, 1, 7–30.

Goldstein, H.I. (1986). Efficient statistical modeling of logitudinal data. Annals of Human Biology, 13, 129–142.

Gover, J. C. (1962). Variance component estimation for unbalanced hierarchical classifications. Biometrics, 18, 537–542.

Hall, P. (1986a). On the bootstrap and confidence intervals. The Annals of Statistics, 14, 4, 1431–1452.

Hall, P. (1986b). On the number of bootstrap simulations required to construct a confidence interval. The Annals of Statistics, 14, 4, 1453–1462.

Hall, P. (1988). Theoretical coomparison of bootstrap confidence intervals. The Annals of Statistics, 16, 3, 927–953.

Hall, P. (1989). Bootstrap confidence regions for directional data. Journal of American Statistical Association, 84, 408, 996–1002.

Hall, P. (1990). The Bootstrap and Edgeworth Expansion. (in Print).

Hanushek, E. A. (1974). Efficient estimators for regressing regression coefficients. The American Statistican, 28, 2, 66–67.

Hartley, H.O. (1967). Expectation, variances and covariances of ANOVA mean squares by 'synthesis'. Biometrics, 23, 105–114.

Hartley, H. O., Rao, J. N. K., and LaMotte, L. R. (1978). A simple 'synthesis' based method of variance component estimation. Biometrics, 34, 233–242.

Harville, D. A. (1977). Maximum Likelihood approaches to variance component estimation and to related problems. Journal of American Statistical Association, 72, 358, 320–340.

Henderson, C. R. (1953). Estimation of variance and covariance components. Biometrics, 9, 226–252.

Henderson, C. R. (1959). Design and analysis of animal husbandry experiments. Techniques and Procedures in Animal Production Research. American Society of Animal Production Monograph.

Hinkley, D., and Wei, B. (1984). Improvements of jackknife confidence limit method. Biometrika, 71, 2, 331–339.

Hocking, R. R. (1985). The analysis of linear models. Monterey: Brooks/Cole Publishing Company.

Kennedy, W. J., & Gentle, J. E. (1980). Statistical Computing. New York: Marcell Dekker, Inc.

Laird, N. M., & Louis, T. A. (1987). Empirical Bayes confidence interval based on Bootstrap samples. Journal of American Statistical Association, 82, 399, 739–750.

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

Leestma, S., & Nyhoff, L. (1987). PASCAL programming and problem solving. New York: Macmillan Publishing Company.

Lunneborg, C. E. (1985). Estimating the correlation coefficient: The bootstrap approach. Psychological Bulletin, 98 (1), 209–215.

Mason, W. M., Wong, G. Y., & Entwistle, B. (1983). Contextual analysis through the multi–level linear model. Sociological Methodology. San Francisco, CA: Jossey–Bass.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). Introduction to the Theory of Statistics. New York: McGraw–Hill Book Company.

Nash, S. N. (1981). Comments on nonparametric standard errors and confidence intervals. The Canadian Journal of Statistics, 9(2), 163–164.

Newman, F.M., Rutter, R.A., & Smith, M.S. (1989). Organizational factors that affect school sense of efficacy, community, and expectations. Sociology of Education, 62, 221–238.

Njunji, A. (1974). Transformation of education in Kenya since independence. Education Eastern Africa, 4, 107–125.

Patterson, H.D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. Biometrika, 58, 545–554.

Peterson, P. (1972). A Review of the Research on Mastery Learning Strategies. Unpublished Manuscript, International Association for the Evaluation of Educational Achievement.

Quenouille, M. (1949). Approximate tests of correlation in time series. Journal of the Royal Statistical Society, Series B, 11, 18–84.

Rao, J. N. K. (1968). "On expectations, variances and covariances of ANOVA Mean squares by 'synthesis'." Biometrics, 24, 963–978.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. Journal of American Statistical Association, 67, 337, 112–115.

Rao, C. R. (1970). "Estimation of Heteroscedastic Variances in Linear models. Journal American Statistical Association, 65, 161–172.

Rao, C. R. (1971 a). Estimation of variance and covariance components – MINQUE theory. Journal of Multivariate Analysis, 1, 257–275.

Rao, C. R. (1971 b). Minimum variance quadratic unbiased estimation of variance components. Journal of Multivariate Analysis, 1, 445–456.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models Journal of American Statistical Association, 67, 112–115.

Rasmussen, J. L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches. Psychological Bulletin, 101 (1) 136–139.

Raudenbush, S. W. (1984). Application of a hierarchical linear model in educational research. Unpublished doctoral dissertation, Harvard University.

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. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101(1) 147–158.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In Ernst Z. Rothkopf (ed.), Review of Research in Education 1988–89. Washington, D.C.: American Educational Research Association.

Rubin, D. B. (1981). The Bayesian bootstrap. The Annals of Statistics, 9(1), 130–134.

SAS Institute, Inc. (1985). SAS User's Guide: Statistics. Cary, NC: SAS Institute, Inc.

SAS Institute, Inc. (1985). SAS/IML User's Guide. Cary, NC: SAS Institute, Inc.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. Journal of American Statistical Association, 80, 390, 360–361.

Searle, S. R. (1971). Topics in variance component estimation. Biometrics, 27, 1–76.

Searle, S. R. (1979). Notes on Variance Component Estimation: A detailed account of maximum likelihood and kindred methodology. Biometrics Unit, Cornell University, New York.

Seely, J. (1971). Quadratic subspaces and completeness. Annals of Mathematical Statistics, 42, 710–721.

Singh, J. (1981). On the asymptotic accuracy of Efron's bootstrap. The Annals of Statistics, 9(6), 1187–1195.

Slavin, R. E. (1987). Mastery learning reconsidered. Review of Educational Research, 57, 2, 175–213.

Thompson, W. A. (1962). The problem of negative estimates of variance components. Annals of Mathematical Statistics, 33, 273–289.