THE USE OF AUXILIARY INFORMATION IN THE LINEAR

LEAST-SQUARES PREDICTION APPROACH TO CLUSTER

SAMPLING IN A FINITE POPULATION

by

Ragan Burt Madden

Dissertation submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Dr. Richard L. Greenstreet, Chairman
Advisory Committee

Dr. Jesse Arnold, Chairman
Department of Statistics

Dr. Raymond H. Myers
Department of Statistics

Dr. Walter R. Pirie
Department of Statistics

Dr. H. F. Van Landingham
Department of Electrical
Engineering

October, 1976

Blacksburg, Virginia

ACKNOWLEDGMENTS

The author would like to thank his advisor, Dr. Richard L. Greenstreet, for his suggestion of the problem and his many helpful comments. Dr. Walter Pirie deserves special praise for his careful reading of the manuscript and his offering of constructive criticisms. The Statistics Department of Virginia Polytechnic Institute and State University should also be acknowledged for providing the computer funds necessary for the many hours of computer use. I am also grateful to Miss Judy Galliher for her patience and willingness in typing the manuscript.

Finally, I would like to thank my parents, Mr. and Mrs. Ragan D. Madden, for instilling in me and my two sisters the importance of education and the ambition to pursue it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

1.1 General Remarks and the Concept of a Super Population

When one wishes to estimate the total or average of all the elements in a finite population by taking a sample survey, it is often practical to consider the elements as being grouped into clusters. A two-stage sampling process can then be adopted. The first stage consists of a selection of sample clusters, and the second consists of a selection of elements from the sample clusters. The clusters are called first-stage or primary units, and the elements within the clusters are called second-stage or secondary units. One could, of course, have situations which call for a three-or-more-stage sampling process, but here we are concerned only with a two-stage scheme.

Cochran [2] points out that cluster sampling may be a feasible method when (i) a complete list of the secondary elements to be sampled is not available or (ii) if a complete list is available, it may be more economical to adopt a cluster sampling approach than to take just a random sample of the elements.

For example, suppose one wishes to estimate the total amount of a prescribed drug which has been sold in Virginia during the current year. A complete list of all drugstores in Virginia may not be available; however, one can consider all of the drugstores in each county as constituting a cluster. The sampling method would then involve first taking a sample of counties, and then selecting certain drugstores within each sampled county. If, on the other hand, a complete list of all drugstores were available, then a random selection of them may result in the selected stores being located great distances apart. The cost incurred in traveling to these stores may exceed that incurred by first selecting a number of counties and then traveling to a number of stores within each selected county.

Often one may have knowledge of some auxiliary information about the clusters. Frequently, this information is either (i) the cluster totals at a previous census, (ii) values of the cluster totals and some of the secondary units at a previous census, or (iii) values of all of the secondary units at a previous census. If it is felt that the present cluster totals are related to the previous totals, then this auxiliary information should be utilized in the analysis.

Some authors--Cochran [2] and Sukhatme and Sukhatme [19]--give a slightly different definition of cluster sampling. They define it to be the process whereby one first takes a sample of clusters, and then each element in the sampled clusters is selected. The sampling scheme we are using in which only a portion of the elements in the sample clusters may be chosen is termed just two-stage sampling or sub-sampling by these authors.

It should be mentioned that all of the subsequent analysis assumes that the clusters have already been formed. It is not our intention to investigate the problem of actually forming or constructing the clusters. Usually we consider the clusters as being geographic sub-units as, for example, the counties of a state or the precincts of a county.

Also, it is assumed that all units in the first and second-stages are distinguishable and sampling at both stages is done without replacement.

In the last few years the concept of using a super-population model in sampling from finite populations has received some attention. This approach considers the values of the population units to be realized values of random variables with unknown means and known or unknown variances and covariances. In contrast, conventional sampling theory considers the values of the units as being just fixed unknown constants, and the only random nature of the problem is in the selection of the sample units.

In the drugstore example the quantities of the prescribed drug which have been sold by the various stores at the time the sample is taken can be thought of as being generated by an underlying stochastic model. If there are K total stores in the population, then the super-population would be the set of all K-tuples where the, say, i^{th} component of each K-tuple would represent a possible value for the i^{th} store. The finite population is then a realization of one of these K-tuples. Another example which illustrates the concept is a dice tossing problem. Suppose that K dice have been tossed independently, and it is desired to estimate the sum of all K faces that appeared based on a sample of size k. Furthermore, suppose it is not known if the dice are fair, but it can be assumed the outcomes of each are identically distributed with mean β. Then for each $i = 1, \ldots, K$; the outcome y_i of the i^{th} die is a realized value of a random variable Y_i with mean β and variance σ^2 where β and σ^2 are unknown. The super population is the set of all 6^K K-tuples $(y_1, \ldots, y_K)$. The finite population is a realization of one of these K-tuples. In both of these examples the fact that the values of the units in the finite population are fixed at the time the sample is taken does not preclude their having been generated by an underlying model.

In all the ensuing analyses it is not necessary to know the actual values comprising the super population, as in the dice tossing example. Only the form of the basic model is assumed. When necessary, any unknown parameters present in the model are estimated by the techniques described.

The assumption of an underlying model allows one to obtain many optimization criteria which are not obtainable in the conventional theory. For example, optimum estimators of the total of the units and their mean square errors can be derived using linear least-square (l.l.s.) prediction methods, other estimators can be analyzed and compared to the optimum ones, and optimum selection of the units to be sampled can often be determined.

Also, any inferences, such as biasedness or mean square error, concerning an estimator in the super-population approach are with respect to the particular underlying model which describes the random variables. Inferences in the conventional approach are with respect to the particular sampling process--e.g., simple random sampling or probability proportional to size random sampling.

1.2 Literature Review

One of the first to look at properties of estimators in a super-population model was Des Raj [3]. He considers the elements y_i in the population to be generated according to the linear model $Y_i = \beta X_i + \varepsilon_i$

where β is an unknown regression coefficient, X_i is a known auxiliary variable associated with the i^{th} unit, $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = aX_i^g$, $g \geq 0$, and the ε_i are independent. Comparisons are made between ratio, regression, and probability proportional to size estimators in the super-population model. Brewer [1] investigates a similar model and makes comparisons between sampling with replacement and sampling without replacement for various ratio type estimators.

One of the first attempts at utilizing l.l.s. prediction methods in finite population sampling was Royall [13] and [14], who considers the same model with the exception of $Var(\varepsilon_i)$. He assumes $Var(\varepsilon_i) = \sigma^2 v(X_i)$ where $v(X_i)$ is a known function of X_i. Comparisons are made between his derived optimum estimators of the total and the conventional estimators. Royall also presents empirical evidence showing the conventional estimators of the mean square error of the ratio estimator to be inappropriate as a measure of precision. Royall and Herson [15] generalize the preceding model by allowing $Y_i = h(X_i) + \varepsilon_i [v(X_i)]^{1/2}$ where $h(X_i) = \sum_{j=0}^{J} \delta_j \beta_j X_i^j$ with $\delta_j = 0$ or 1 and the β_j unknown. The random errors are assumed to be independent each with mean zero and variance σ^2. They introduce the concept of balanced sampling as a means of protecting against bias when the regression part of the model fails. Royall and Herson [16] discuss the technique of stratifying on the auxiliary variable X in the preceding model as another means of protecting against bias introduced by a failure of the model. In both [15] and [16] l.l.s. prediction methods are employed.

Inferences made on the estimators derived under the previous models are based on the assumption of repeated sampling from different finite populations each with the same values of the auxiliary variables. Hartley and Sielken [9] extend the concept somewhat by allowing inferences to be made on estimators under the assumption of repeated sampling from different finite populations each with possibly different values of the auxiliary variables. They obtain confidence intervals, under normality assumptions, on the regression parameters.

Greenstreet [5], [6], and Greenstreet and Madden [8] extend Royall's work to the case of multivariate characteristics and multivariate auxiliary variables on each unit. Greenstreet and Madden [7] compare the performances of two different models (additive and interactive) for the case of stratified sampling on qualitative auxiliary variables in the univariate case.

As for cluster-sampling problems, Scott and Smith [18] utilize a Bayesian approach by assuming prior distribution on the parameters of the population. Royall [17] considers the cluster-sampling problem by l.l.s. prediction methods, but does not allow for the presence of any auxiliary information.

Cochran [2] and Sukhatme and Sukhatme [19] give a thorough treatment of cluster-sampling processes in the conventional theory.

1.3 Preliminaries

Another aspect of the super-population model is that the quantity we are trying to estimate--the total T of all the units--is a realized value of a random variable. Whenever we say that an estimator $\hat{T}$ is unbiased for T, we mean that $\hat{T}$ is an unbiased estimator of the expected value of T. That is, $E(\hat{T}) = E(T)$ where the expected values are with respect to the super-population model. Theorem 1.1 gives the optimum estimator for the total of the unsampled units.

Theorem 1.1. Whittle [21] In the linear model $\underline{Y} = Z\,\underline{\beta} + \underline{\varepsilon}$, let $\underline{Y}$ be a vector of K random variables, Z a K×p matrix (full rank) of known auxiliary variables, $\underline{\beta}$ a p×1 vector of unknown regression coefficients, and let $E[\underline{Y}] = Z\,\underline{\beta}$ and $Var[\underline{Y}] = V$. Let a sample of size k be selected and let $\underline{Y}_I$ be the k×1 vector of sampled $\underline{Y}$'s, $\underline{Y}_{II}$ the (K-k)×1 vector of unsampled $\underline{Y}$'s, Z_i the i^{th} submatrix (i = I, II) of Z associated with $\underline{Y}_i$. Let V_I be the k×k covariance matrix of $\underline{Y}_I$, V_{II} the (K-k)×(K-k) covariance matrix of $\underline{Y}_{II}$, and $V_{II,I}$ the (K-k)×k matrix representing the covariances between $\underline{Y}_{II}$ and $\underline{Y}_I$. We can therefore write the model as

$$\begin{pmatrix} \underline{Y}_I \\ \underline{Y}_{II} \end{pmatrix} = \begin{pmatrix} Z_I \\ Z_{II} \end{pmatrix} \underline{\beta} + \underline{\varepsilon} \text{ , and}$$

$$V = \begin{pmatrix} V_I & V'_{II,I} \\ V_{II,I} & V_{II} \end{pmatrix} .$$

Now, let $\underline{c}$ be a vector of (K-k) known constants.

Then, the best, linear, unbiased, estimator (b.l.u.e.) $\widehat{\underline{c}'\,\underline{Y}_{II}}$ of $\underline{c}'\,\underline{Y}_{II}$ in the sense of minimizing the mean square error (m.s.e.),

$$E(\widehat{\underline{c}'\,\underline{Y}_{II}} - \underline{c}'\,\underline{Y}_{II})^2$$

is given by

$$\widehat{\underline{c}'\,\underline{Y}_{II}} = \underline{c}'[Z_{II}\,\hat{\underline{\beta}} + V_{II,I}\,V_I^{-1}(\underline{Y}_I - Z_I\,\hat{\underline{\beta}})] \tag{1.3.1}$$

where

$$\hat{\underline{\beta}} = (Z_I'\,V_I^{-1}\,Z_I)^{-1}\,Z_I'\,V_I^{-1}\,\underline{Y}_I .$$

We observe that $\hat{\underline{\beta}}$ is the b.l.u.e. of $\underline{\beta}$.

The m.s.e. is given by

$$E(\widehat{\underline{c}'\underline{Y}_{II}} - \underline{c}'\underline{Y}_{II})^2 = \underline{c}'[V_{II} - V_{II,I}\, V_I^{-1}\, V'_{II,I} + (Z_{II} - V_{II,I}\, V_I^{-1}\, Z_I) \times (Z'_I\, V_I^{-1}\, Z_I)^{-1}\, (Z_{II} - V_{II,I}\, V_I^{-1}\, Z_I)']\, \underline{c}. \quad (1.3.2)$$

Utilizing Theorem 1.1 enables one to obtain the b.l.u.e. of the total T of all units in the finite population.

Theorem 1.2: Given the linear model $\underline{Y} = Z\,\underline{\beta} + \underline{\varepsilon}$ as defined in Theorem 1.1. Let $T=\sum_1^K y_i$ be the quantity to be estimated (See Section 1.4). Then, the b.l.u.e. $\hat{T}_B$ of T in the sense of minimizing $E(\hat{T}_B - T)^2$ is given by

$$\hat{T}_B = \sum_S y_i + \widehat{\underline{c}'\, \underline{Y}_{II}}\ ,$$

where $\sum_S y_i$ is the sum of the y's in the sample, $\underline{c}$ is a vector of (K-k) ones, and $\widehat{\underline{c}'\, \underline{Y}_{II}}$ is given by Theorem 1.1.

Proof: The total T can be written as the sum of the units in the sample plus the sum outside the sample. That is,

$$T = \sum_S y_i + \sum_{\bar{S}} y_i\ .$$

Now, any linear estimator for T can be written as

$$\hat{T} = \sum_S y_i + f(\underline{Y}_I)$$

where $f(\underline{Y}_I)$ is some linear function of the y's in the sample. Then

$$E(\hat{T} - T)^2 = E(f(\underline{Y}_I) - \sum_{\bar{S}} y_i)^2\ .$$

But from Theorem 1.1, this is minimized among all linear, unbiased, estimators $\hat{T}$ by that particular one for which

$$f(\underline{Y}_I) = \widehat{\underline{c}' \, \underline{Y}_{II}}.$$

Hence, the optimum estimator $\hat{T}_B$ is given by

$$\hat{T}_B = \sum_S y_i + \widehat{\underline{c}' \, \underline{Y}_{II}}.$$

1.4 Further Remarks

Strictly speaking, we are trying to estimate the random variable $\sum_1^K Y_i$. However, the realized values y are unknown prior to sampling and thus have the same *a priori* probabilistic structure as the corresponding random variables Y. Operations such as expected values, variances, and covariances of expressions containing y are to be interpreted in an *a priori* sense, and are equivalent to the same operations where y is replaced by its corresponding random variable Y.

II. A CLUSTER MODEL WHEN AUXILIARY INFORMATION IS AVAILABLE ONLY ON CLUSTER TOTALS

2.1 Notation

The following notation will be used throughout:

N = Total number of clusters in finite population

n = Number of clusters sampled at first stage

M_i = Total number of elements in population for i^{th} cluster (known for all i)

m_i = Number of elements sampled at second stage from i^{th} sampled cluster

K = Total number of elements in finite population = $\sum_1^N M_i$

k = Number of elements in sample = $\sum_{i \in S} m_i$

X_i = Auxiliary variable associated with i^{th} cluster (known for all i)

y_{ij} = Value of j^{th} unit in i^{th} cluster = realized value of random variable Y_{ij}

$\bar{y}_i$ = Mean of the m_i second-stage units from i^{th} sampled cluster

T_i = Total of all elements in i^{th} cluster = $\sum_{j=1}^{M_i} y_{ij}$

T = Grand total of all elements in population = $\sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij}$

ρ_i = Correlation of any two elements in i^{th} cluster

σ_i^2 = Variance of each element in i^{th} cluster

β = Unknown regression coefficient

S = Set of numbers identifying the n sample clusters

$\bar{S}$ = Set of numbers identifying the (N-n) unsampled clusters

S_i = Set of numbers identifying the m_i sampled units in i^{th} sampled cluster

2.2 The Model

If X_i represents the total of the units† in the i^{th} cluster at a previous census, then we shall consider the model

$$Y_{ij} = \beta X_i/M_i + \varepsilon_{ij} \;;\; i = 1,\dots, N \qquad (2.2.1)$$
$$j = 1,\dots, M_i$$

where

$$E(Y_{ij}) = \beta X_i/M_i$$

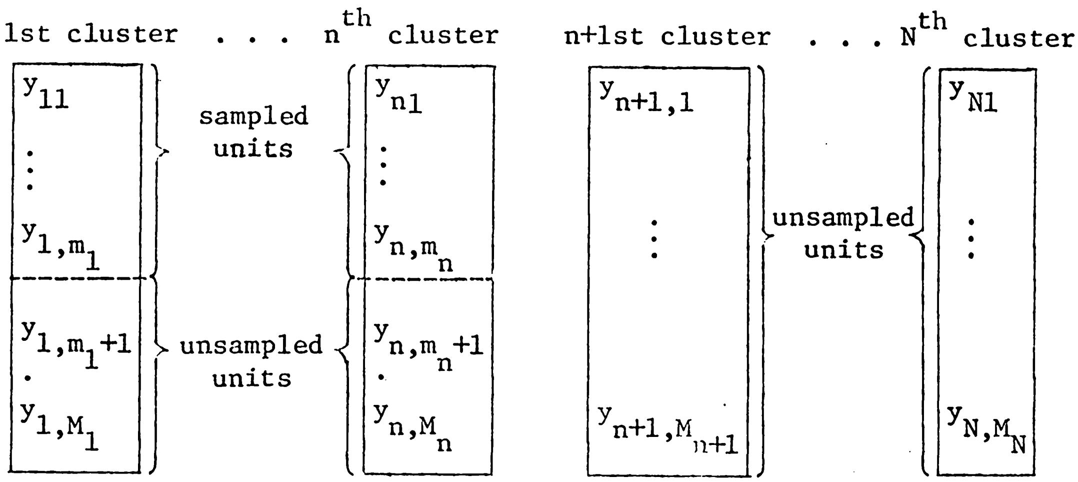
$$\text{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma_i^2 \;;\; i = i', j = j' \\ \rho_i\sigma_i^2 \;;\; i = i', j \neq j' \\ 0 \;;\; i \neq i' \end{cases} .$$

We are assuming that units in a given cluster are correlated; whereas units in different clusters are uncorrelated. Also we are assuming that all elements in a given cluster have the same mean, variance, and covariances. This is the same covariance structure as used by Royall [17]. However, he does not consider the case where the regression part of the model depends on any auxiliary information.

Note that under our model we have $T_i = \beta X_i + \varepsilon_i$. Or, the present cluster total is a linear function of the previous total. In Chapter IV we shall consider a model which uses auxiliary information on each secondary unit.

Our objective is to estimate the grand total T of all the units based on a sample of k units selected from n sample clusters. The clusters, sample units, and unsampled units are illustrated as follows:

†Henceforth, the term "units" or "elements" will, unless otherwise modified, refer to the second-stage units.



We have assumed for convenience that the clusters are numbered from 1 to N such that the first n of them represent the sample clusters, and the M_i units in the i^{th} sample cluster have been numbered from 1 to M_i such that the first m_i of them represent the sample units. We are not, at this point, interested in how the sample clusters should be selected or how the units within the sample clusters should be chosen. More will be said about this later.

2.3 The b.l.u.e. of the Total

In the notation of Theorem 1.1 we have that

$$\underline{Y}_I = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1m_1} \\ \hline \vdots \\ \hline y_{n1} \\ \vdots \\ y_{nm_n} \end{pmatrix} \qquad \underline{Y}_{II} = \begin{pmatrix} y_{1,m_1+1} \\ \vdots \\ y_{1M_1} \\ \hline \vdots \\ \hline y_{n,m_n+1} \\ \vdots \\ y_{nM_n} \\ \hline y_{n+1,1} \\ \vdots \\ y_{n+1,M_{n+1}} \\ \hline \vdots \\ \hline y_{N1} \\ y_{NM_N} \end{pmatrix}$$

$$Z_I = \left.\begin{pmatrix} X_1/M_1 \\ \vdots \\ X_1/M_1 \\ \hline \vdots \\ \hline X_n/M_n \\ \vdots \\ X_n/M_n \end{pmatrix}\right. \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} m_1 \text{ units} \\ \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} m_n \text{ units} \end{matrix} \quad , \qquad Z_{II} = \begin{pmatrix} X_1/M_1 \\ \vdots \\ X_1/M_1 \\ \hline \vdots \\ \hline X_n/M_n \\ \vdots \\ X_n/M_n \\ \hline \vdots \\ \hline X_{n+1}/M_{n+1} \\ \vdots \\ X_{n+1}/M_{n+1} \\ \hline \vdots \\ \hline X_N/M_N \\ \vdots \\ X_N/M_N \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} (M_1-m_1) \text{ units} \\ \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} (M_n-m_n) \text{ units} \\ \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} M_{n+1} \text{ units} \\ \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} M_N \text{ units} \end{matrix}$$

The covariance matrix V is given on the following page. Note that V has been partitioned into

$$V = \begin{pmatrix} V_I & V'_{II,I} \\ V_{II,I} & V_{II} \end{pmatrix} .$$

$$V = \begin{pmatrix} \begin{pmatrix} A_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & A_n \end{pmatrix} & \begin{pmatrix} & & \\ & \text{symm.} & \\ & & \end{pmatrix} \\ \begin{pmatrix} C_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & C_n \\ 0 & \cdots & 0 \\ \vdots & & \\ 0 & \cdots & 0 \end{pmatrix} & \begin{pmatrix} B_1 & & & & & \bigcirc \\ & \ddots & & & & \\ & & B_n & & & \\ & & & B_{n+1} & & \\ & & & & \ddots & \\ \bigcirc & & & & & B_N \end{pmatrix} \end{pmatrix} \tag{2.3.1}$$

The submatrices A_i are each $(m_i \times m_i)$ for $i = 1, \ldots, n$ and are defined to be

$$A_i(j, k) = \begin{cases} \sigma_i^2 & ; j = k \\ \rho_i \sigma_i^2 & ; j \neq k \end{cases} .$$

The submatrices B_i are each $(M_i - m_i) \times (M_i - m_i)$ for $i = 1, \ldots, n$ and are $(M_i \times M_i)$ for $i = n+1, \ldots, N$. The B_i are defined to be

$$B_i(j, k) = \begin{cases} \sigma_i^2 & ; j = k \\ \rho_i \sigma_i^2 & ; j \neq k \end{cases} .$$

The submatrices C_i are each $(M_i - m_i) \times m_i$ for $i = 1, \ldots, n$ and are defined to be

$$C_i(j, k) = \rho_i \sigma_i^2 \text{ for all } j, k \quad .$$

we have that

$$V_I^{-1} = \begin{pmatrix} A_1^{-1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & A_n^{-1} \end{pmatrix}$$

where

$$A_i^{-1} = \frac{1}{D_i} \begin{pmatrix} (m_i-2)\rho_i+1; & -\rho_i; & \ldots & ; -\rho_i \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & ; (m_i-2)\rho_i+1 \end{pmatrix}.$$

The dimensions of A_i^{-1} are $m_i \times m_i$ and

$$D_i = \sigma_i^2 [1 + (m_i - 2)\ \rho_i - (m_i - 1)\ \rho_i^2]\ .$$

Letting $\underline{c}$ be a vector of (K-k) ones, we have from Theorems 1.1 and 1.2 that the b.l.u.e. $\hat{T}_B$ of the total T is given by

$$\hat{T}_B = \sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij} + \sum_{i\varepsilon S} (M_i - m_i)[\omega_i \bar{y}_i + \hat{\beta} X_i (1-\omega_i)/M_i] + \hat{\beta} \sum_{i\varepsilon \bar{S}} X_i \qquad (2.3.2)$$

where

$$\hat{\beta} = \frac{\sum_{i\varepsilon S} m_i\ X_i\ \bar{y}_i/M_i\ \sigma_i^2[\rho]_i}{\sum_{i\varepsilon S} m_i\ X_i^2/M_i^2\ \sigma_i^2[\rho]_i}\ ,$$

$$\omega_i = \rho_i\ m_i/[\rho]_i\ ,$$

$$[\rho]_i = 1 - \rho_i + m_i\ \rho_i\ ,$$

and

$$\bar{y}_i = \sum_{j \in S_i} y_{ij}/m_i \quad .$$

From Theorem 1.1, we obtain the m.s.e. to be

$$E(\hat{T}_B - T)^2 = \sum_{i \in S} (M_i - m_i)\sigma_i^2[1-\rho_i+\rho_i(M_i-m_i)] + \sum_{i \in \bar{S}} M_i\sigma_i^2(1-\rho_i+\rho_iM_i)$$

$$- \sum_{i \in S} (M_i-m_i)^2\omega_i\rho_i\sigma_i^2 + \frac{\left[\sum_{i \in S} \frac{(M_i-m_i)X_i(1-\rho_i)}{M_i[\rho]_i} + \sum_{i \in \bar{S}} X_i\right]^2}{\sum_{i \in S} \frac{m_i X_i^2}{M_i^2 \sigma_i^2 [\rho]_i}} \quad . \qquad (2.3.3)$$

To see that $\hat{T}_B$ is unbiased for T, we have

$$E(\hat{T}_B) = \beta(\sum_{i \in S} \sum_{j \in S_i} X_i/M_i) + \beta \sum_{i \in S} (M_i-m_i)[\omega_iX_i/M_i + X_i(1-\omega_i)/M_i] + \beta\sum_{\bar{S}} X_i$$

$$= \beta \sum_{i \in S} m_iX_i/M_i + \beta \sum_{i \in S} (M_i-m_i)X_i/M_i + \beta \sum_{i \in \bar{S}} X_i$$

$$= \beta \sum_{1}^{N} X_i = E(\sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij})$$

$$= E(T) \; .$$

2.4 Discussion of the Parameters and the b.l.u.e.

2.4.1 Estimation of ρ_i and σ_i^2. Let $\sigma_i^2 = \sigma^2$ for all $i \in S$. Then the covariance matrix V_I of the units in the sample can be written as $V_I = \sigma^2 H$. Assuming the ρ_i are known, an unbiased estimator for σ^2 is

$$\hat{\sigma}^2 = (\underline{Y}_I - Z_I\hat{\beta})' H^{-1} (\underline{Y}_I - Z_I\hat{\beta})/(k-1) \qquad (2.4.1)$$

where we are using the notation of Theorem 1.1.

If the σ_i^2 are not all assumed equal, and the ρ_i are known, the same general expression can be used to obtain an unbiased estimator for σ_i^2 provided that on the right-hand-side we use the values corresponding to just the i^{th} cluster. That is, Y_I, Z_I, and H are appropriate for only the i^{th} cluster, $\hat{\beta}$ is calculated from observations only from the i^{th} cluster, and the divisor is (m_i-1).

The usual estimate of the correlation ρ between two random variables, W and Z say, is based on observing n ordered pairs (w_i, z_i) and using the estimator

$$\hat{\rho} = \frac{\sum_1^n (w_i - \bar{w})(z_i - \bar{z})}{[\sum_1^n (w_i - \bar{w})^2 \sum_1^n (z_i - \bar{z})^2]^{1/2}} .$$

It can be shown--Kendall and Stuart [10]--that for a bivariate normal distribution for W and Z that $\hat{\rho}$ is slightly biased for ρ when $\rho \neq 0$, but the bias is very small and decreases with increasing n.

Now, assume $\rho_i = \rho$ for all $i \in S$. The data are not presented as ordered pairs, and so the usual estimator for ρ can not be directly used. For the type data that we have, Kendall and Stuart [10] recommend estimating ρ by means of the intra-class correlation coefficient described below.

For each cluster, form all possible ordered pairs of the observations. So, for the i^{th} cluster we will have $m_i(m_i - 1)$ ordered pairs. Hence, we will have a total of $\sum_{i\varepsilon S} m_i(m_i - 1) = L$ ordered pairs for the k observations. Then, using these L ordered pairs estimate ρ with the usual estimate $\hat{\rho}$. Kendall and Stuart [10] show that the estimator $\hat{\rho}$ becomes, for this case

$$\hat{\rho} = \frac{\sum_{i\varepsilon S} m_i^2(\bar{y}_i - \bar{\bar{y}})^2 - \sum_{i\varepsilon S}\sum_{j\varepsilon S_i}(y_{ij} - \bar{\bar{y}})^2}{\sum_{i\varepsilon S}(m_i - 1)\sum_{j\varepsilon S_i}(y_{ij} - \bar{\bar{y}})^2} \qquad (2.4.2)$$

where

$$\bar{\bar{y}} = \frac{1}{L}\sum_{i\varepsilon S}(m_i - 1)\sum_{j\varepsilon S_i} y_{ij} ,$$

and represents the sample mean of each of the two variates. If $m_i = m$ for all $i\varepsilon S$, then (2.4.2) simplifies to

$$\hat{\rho} = \frac{1}{m-1}\left(\frac{m\,s^2}{t^2} - 1\right) ,$$

where $$s^2 = \frac{1}{n}\sum_{i\varepsilon S}(\bar{y}_i - \bar{\bar{y}})^2$$

and $$t^2 = \frac{1}{L}\sum_{i\varepsilon S}(m_i - 1)\sum_{j\varepsilon S_i}(y_{ij} - \bar{\bar{y}})^2 .$$

It should be noted that for this case if there is no variation among the observations within each cluster, then $t^2 = s^2$ and $\hat{\rho} = 1$ which is certainly the estimate one would like. For unequal correlations, ρ_i can be estimated by applying (2.4.2) to the m_i observations from the i^{th} cluster.

<u>2.4.2 Hypothesis Testing on ρ_i and σ_i^2</u>. Several standard methods are available for testing the hypothesis that $\sigma_i^2 = \sigma^2$ for all $i\varepsilon S$,

e.g., Bartlett's Test. Many of these tests are very dependent on the assumption that the units in each cluster are independent and normally distributed, however.

The testing of the hypothesis that $\rho_i = \rho$ for all $i \varepsilon S$ can be based on Fisher's z-transformation of the sample correlation coefficients $\hat{\rho}_i$. See Rao [12] and Kendall and Stuart [10]. For small values of ρ the test is robust to departures from normality; whereas for large values the test is more sensitive to departures from normality but becomes less so as the sample sizes increase.

2.4.3 Restrictions on the ρ_i. We have a sample of k observations from n clusters with covariance matrix V_I given in (2.3.1). Let V_I = diag.$\{A_1, A_2, \ldots, A_n\}$ where A_i is the $m_i \times m_i$ covariance matrix of elements from the i^{th} sample cluster. In order to examine values of the ρ_i which result in V_I being positive definite, let $\underline{Z}$ be any vector of k real variables partitioned as $\underline{Z}' = (\underline{Z}_1', \ldots, \underline{Z}_n')$ where $\underline{Z}_i'$ is a vector (row) of m_i variables. Then, we have

$$\underline{Z}' V_I \underline{Z} = \sum_1^n \underline{Z}_i' A_i \underline{Z}_i ,$$

and V_I is thus positive definite if each A_i is. But from a theorem in Graybill [4], A_i is positive definite if and only if

$$-1/(m_i - 1) < \rho_i < 1 . \qquad (2.4.3)$$

The case where some of the $\rho_i = 1$ can be examined as follows. Recall that if two random variables, Y_{ij}, and $Y_{ij'}$, have a correlation of unity, they must be related by

$$Y_{ij} = a\ Y_{ij'} + b$$

where a and b are constants. But under our model (2.2.1) we have

$$E[Y_{ij}] = \beta\ X_i/M_i = a\ \beta\ X_i/M_i + b$$

and

$$Var[Y_{ij}] = \sigma_i^2 = a^2\ \sigma_i^2\ .$$

Hence, a = 1 and b = 0, which implies that $Y_{ij} = Y_{ij'}$. That is, all elements in each cluster for which $\rho_i = 1$ are equal. Thus, it suffices to sample only one element from these clusters. But if $m_i = 1$ for all sample clusters for which $\rho_i = 1$, then the corresponding $A_i = \sigma_i^2$ which is certainly positive definite. Hence, inequality (2.4.3) can be relaxed to include values of $\rho_i = 1$ provided we adopt the criterion of sampling only one element from these sample clusters.

Further discussion of cases for which $\rho_i = 1$ is given at the end of Section 2.5.

Also, in most of the ensuing analyses we shall assume that $\rho_i \geq 0$.

2.4.4 Properties of the b.l.u.e. (i) If $\sigma_i^2 = \sigma^2$ for all $i\varepsilon S$ we note from (2.3.2) that $\hat{T}_B$ is independent of σ^2. (ii) It is easy to see that $\hat{T}_B$ is unbiased for T even though one chooses the σ_i^2 and ρ_i incorrectly. That is, a change in the covariance structure does not affect the unbiasedness of $\hat{T}_B$. (iii) The estimator $\hat{T}_B$ and its mse are independent of the dimensions used for the X's so long as any other set of values, say Z, are related to X by $Z_i = b\, X_i$ where b is a constant. (iv) When some of the sampled clusters are such that $M_i = 1$, the corresponding ρ_i in (2.3.2) and (2.3.3) have no meaning. It is easy to see that these two expressions are completely independent of ρ_i whenever $M_i = m_i = 1$. (v) If all $M_i = 1$, $i = 1,\ldots, N$; model (2.2.1) becomes

$$Y_i = \beta\, X_i + \varepsilon_i \;,$$

which is just a non-cluster model. Letting $\mathrm{Var}(Y_i) = \sigma^2$, $\sigma^2\, X_i$, and $\sigma^2\, X_i^2$; we obtain the estimators

$$\hat{T}_0 = \sum_S y_i + (\sum_S X_i\, y_i) \sum_{\bar{S}} X_i / \sum_S X_i^2 \;,$$

$$\hat{T}_1 = \Sigma\, y_i + \sum_S y_i \sum_{\bar{S}} X_i / \sum_S X_i \;,$$

and

$$\hat{T}_2 = \sum_S y_i + (\sum_S y_i/X_i) \sum_{\bar{S}} X_i/n$$

respectively. Royall [13] obtains these estimators by beginning with the above model. So, our model is an extension of his non-cluster model. (vi) We observe from (2.3.2) that $\hat{T}_B$ is written as the sum of three terms. The first obviously is just the sum of the sample units. To see what the other two represent, it is helpful to take $E(\hat{T}_B)$. We have

$$E(\hat{T}_B) = E[\sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij}] + \beta \sum_{i\varepsilon S} (M_i - m_i) X_i / M_i + \beta \sum_{i\varepsilon \bar{S}} X_i$$

$$= E[\sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij}] + E[\sum_{i\varepsilon S} \sum_{j\notin S_i} y_{ij}] + E[\sum_{i\varepsilon \bar{S}} \sum_{j=1}^{M_i} y_{ij}] \;.$$

Thus, $\hat{T}_B$ is written as the sum of (a) the sample elements, (b) an unbiased estimator of the sum in the sample clusters but not in the sample itself, and (c) an unbiased estimator of the sum in the unsampled clusters. (vii) In order to examine the behavior of $\hat{T}_B$, in terms of its expected value, when the regression part of model (2.2.1) is incorrect, suppose that the true model is

$$Y_{ij} = (\beta_0 + \beta_1 X_i)/M_i + \varepsilon_{ij} \quad . \tag{2.4.4}$$

It can then be shown that

$$\begin{aligned} E(\hat{T}_B - T) = \beta_0 \{ &\sum_{i\varepsilon S} (M_i - m_1)[\omega_i/M_i + b\, X_i(1 - \omega_i)/M_i] \\ &+ b \sum_{i\varepsilon \bar{S}} X_i - [\sum_{i\varepsilon S} (M_i - m_i)/M_i + (N - n)]\} \end{aligned} \tag{2.4.5}$$

where

$$b = \frac{\sum_{i\varepsilon S} m_i X_i/M_i^2 \sigma_i^2[\rho]_i}{\sum_{i\varepsilon S} m_i X_i^2/M_i^2 \sigma_i^2[\rho]_i} \quad .$$

In general (2.4.5) does not equal zero. That is, $\hat{T}_B$ is a biased estimator of T when the true model is (2.4.4). It can also be easily shown that $\hat{T}_B$ is biased under higher-order regression models.

It should be noted that for non-cluster models; that is, $M_i = 1$; Royall and Herson [15] give conditions for which optimum estimators are unbiased under alternative regression models.

In Section 2.6 we consider a conventional estimator and show that by a proper selection of sample clusters it can be made unbiased under a wide variety of alternative models.

<u>2.4.5 Optimum Selection of the Clusters</u>. The determination of which clusters should be sampled so as to minimize (2.3.3) appears to be intractable in the general case. However, in a special case, given by the following theorem, optimum first-stage selection of the clusters can be determined.

<u>Theorem 2.1</u>: Let the number of clusters to be sampled, n, be fixed, and let the m_i also be fixed. Assume that all elements in all clusters have the same variance and correlation (non-negative), i.e., $\sigma_i^2 = \sigma^2$ and $\rho_i = \rho \geq 0$. Then, if $M_k > M_j$ implies that $X_k/M_k > X_j/M_j$, the largest n clusters in terms of M (equivalently, in terms of X) will minimize the m.s.e. (2.3.3).

<u>Proof</u>: Assuming equal variances and correlations, we can write (2.3.3) as

$$E(\hat{T}_B - T)^2/\sigma^2 = (K - k)(1 - \rho) + \rho\{\sum_{i\varepsilon \bar{S}} M_i^2 + \sum_{i\varepsilon S} (M_i - m_i)^2(1 - \rho)/[\rho]_i\}$$

$$+ [\sum_1^N X_i - \sum_{i\varepsilon S} m_i X_i(1 - \rho + M_i \rho)/M_i[\rho]_i]^2 / \sum_{i\varepsilon S} \frac{m_i X_i^2}{M_i^2[\rho]_i} . \tag{2.4.6}$$

Now, suppose the largest n clusters have not all been sampled. Consider a new sample where the, say j^{th} cluster, is replaced with the k^{th} cluster where $M_k > M_j$, $m_k = m_j$, and all other sample clusters and m's remain the same. The m.s.e. of the estimate, say $\hat{T}_{B'}$, obtained from this new sample can be written as

$$E(\hat{T}_{B'} - T)^2/\sigma^2 = (K - k)(1 - \rho) + \rho\{ \sum_{i\varepsilon\bar{S}} M_i^2 - (M_k^2 - M_j^2)$$

$$+ \sum_{i\varepsilon S} (M_i - m_i)^2 (1 - \rho)/[\rho]_i + (M_k - m_j)^2(1 - \rho)/[\rho]_j$$

$$- (M_j - m_j)^2 (1 - \rho)/[\rho]_j\}$$

$$+ \{\left(\sum_{i\varepsilon S} m_i X_i^2/M_i^2 [\rho]_i - m_j X_j^2/M_j^2[\rho]_j + m_j X_k^2/M_k^2[\rho]_j\right)^{-1}$$

$$\times \left(\sum_1^N X_i - \sum_{i\varepsilon S} m_i X_i(1-\rho+M_i\rho)/M_i[\rho]_i - m_j X_k(1-\rho+M_k\rho)/M_k[\rho]_j\right.$$

$$\left. + m_j X_j(1 - \rho + M_j \rho)/M_j [\rho]_j\right)^2\} \qquad (2.4.7)$$

where the summations are over exactly the same clusters as in (2.4.6). The expression inside the first set of braces can be written as

$$\sum_{i\varepsilon\bar{S}} M_i^2 + \sum_{i\varepsilon S} (M_i - m_i)^2 (1 - \rho)/[\rho]_i - \left((M_k^2 - M_j^2) \rho m_j/[\rho]_j\right.$$

$$\left. + 2(1 - \rho)m_j (M_k - M_j)/[\rho]_j\right)$$

which, under the conditions of the theorem, is smaller than the corresponding expression in (2.4.6). Now, the expression inside the second set of braces will be smaller than the corresponding expression in (2.4.6) if

$$\left(\sum_{i\varepsilon S} m_i X_i^2/[\rho]_i M_i^2\right)\left(\sum_1^N X_i - \sum_{i\varepsilon S} m_i X_i(1 - \rho + M_i\rho)/M_i [\rho]_i\right.$$

$$\left. - m_j X_k(1 - \rho + M_k\rho)/M_k [\rho]_j + m_j X_j(1 - \rho + M_j\rho)/M_j [\rho]_j\right)^2$$

$$< \left(\sum_{i\varepsilon S} m_i X_i^2/[\rho]_i M_i^2 - m_j X_j^2/M_j^2[\rho]_j + m_j X_k^2/M_k^2[\rho]_j\right)$$

$$\times \left(\sum_1^N X_i - \sum_{i\varepsilon S} m_i X_i(1 - \rho + M_i\rho)/M_i [\rho]_i\right)^2 .$$

Again, under the conditions of the theorem the terms inside the first and second set of brackets on the left side of the inequality are smaller than the terms inside the first and second set of brackets on the right side respectively. Hence, the inequality is valid, and we have that the new sample decreases the m.s.e. It therefore follows that the largest clusters will minimize (2.4.6).

We note that under the assumption of equal variances and correlations (non-negative) and n and m_i fixed the conditions of the theorem will be satisfied whenever the auxiliary variables satisfy the relation $X_i = cM_i^q$ where $c > 0$ and $q \geq 1$. Also, if $M_i = M$, i.e., all clusters are of equal size, then it is easy to see that clusters with largest X's will minimize (2.4.6).

2.5 A Suboptimal Estimator of the Total

Equations (2.3.2) and (2.3.3) give respectively the b.l.u.e. and its m.s.e. under model (2.2.1). We have seen that with equal variances and correlations the b.l.u.e. is independent of the common variance σ^2 but does depend on the common correlation ρ. We now wish to look at the case for which one correctly assumes that the variances and correlations are equal but an incorrect choice of the correlation is made.

Let $\rho_i = \rho$ and $\sigma_i^2 = \sigma^2$ for all $i = 1,\ldots, N$. Let $\hat{T}_B(\rho_a)$ be the estimator (2.3.2) when one assumes ρ to be ρ_a. Let

$$E[\hat{T}_B(\rho_a) - T]^2_{\rho_t}$$

be the m.s.e. of $\hat{T}_B(\rho_a)$ when the true value of ρ is ρ_t. Now, $\hat{T}_B(\rho_a)$ is still unbiased for T-Property (ii) of Section 2.4.4. So it is a suboptimal estimator in the sense that it is the b.l.u.e. only when $\rho_a = \rho_t$. We now wish to determine its m.s.e. when the true correlations are ρ_t. We have first that

$$E[\hat{T}_B(\rho_a) - T]^2_{\rho_t} = \mathrm{Var}[\hat{T}_B(\rho_a) - T]_{\rho_t} .$$

Dropping the subscript ρ_t, letting $\underline{c}$ be a vector of (K-k) ones, and $\underline{Y}_{II}$ be the vector of all (K-k) unsampled units, we have

$$\begin{aligned}
\mathrm{Var}[\hat{T}_B(\rho_a) - T] &= \mathrm{Var}\left\{\sum_{i\varepsilon S} (M_i - m_i)[\omega_i \bar{y}_i + \hat{\beta} X_i (1 - \omega_i)/M_i] \right. \\
&\quad \left. + \hat{\beta} \sum_{i\varepsilon \bar{S}} X_i - \underline{c}' \underline{Y}_{II}\right\} \\
&= \sum_{i\varepsilon S} (M_i - m_i)^2 \omega_i^2 \mathrm{Var}\, \bar{y}_i + \left\{\sum_{i\varepsilon S} \frac{(M_i - m_i) X_i (1-\omega_i)}{M_i}\right\}^2 \mathrm{Var}\, \hat{\beta} \\
&\quad + \left\{\sum_{i\varepsilon \bar{S}} X_i\right\}^2 \mathrm{Var}\, \hat{\beta} + \mathrm{Var}(\underline{c}' \underline{Y}_{II}) \\
&\quad + 2\left\{\sum_{i\varepsilon S} (M_i - m_i)\omega_i \operatorname{cov}(\bar{y}_i, \hat{\beta})\right\}\left\{\sum_{i\varepsilon S} \frac{(M_i - m_i) X_i (1-\omega_i)}{M_i}\right\} \\
&\quad + 2 \sum_{i\varepsilon S} (M_i - m_i)\omega_i \operatorname{cov}(\bar{y}_i, \hat{\beta}) \sum_{i\varepsilon \bar{S}} X_i \\
&\quad - 2 \sum_{i\varepsilon S} (M_i - m_i)\omega_i \sum_{j\notin S_i} \operatorname{cov}(\bar{y}_i, y_{ij}) \\
&\quad + 2 \sum_{i\varepsilon S} (M_i - m_i) X_i (1-\omega_i)/M_i \sum_{i\varepsilon \bar{S}} X_i \mathrm{Var}\, \hat{\beta} \\
&\quad - 2 \sum_{i\varepsilon S} (M_i - m_i) X_i (1-\omega_i)/M_i \operatorname{cov}(\hat{\beta}, \underline{c}' \underline{Y}_{II}) \\
&\quad - 2 \sum_{i\varepsilon \bar{S}} X_i \operatorname{cov}(\hat{\beta}, \underline{c}' \underline{Y}_{II}) . \qquad (2.5.1)
\end{aligned}$$

The following expressions are needed in the evaluation of (2.5.1):

$$\mathrm{Var}\, \bar{y}_i = \frac{\sigma^2}{m_i} [1 + \rho_t (m_i - 1)]$$

$$\mathrm{Var}\, \hat{\beta} = \sum_{i\varepsilon S} m_i^2 X_i^2 \mathrm{Var}(\bar{y}_i)/M_i^2 [\rho_a]_i^2 \Big/ \left\{\sum_{i\varepsilon S} m_i X_i^2 / M_i^2 [\rho_a]_i\right\}^2$$

$$\mathrm{cov}(\bar{y}_i,\hat{\beta}) = m_i\ X_i\ \mathrm{Var}(\bar{y}_i)/M_i[\rho_a]_i/\left(\sum_{i\varepsilon S} m_i\ X_i^2/M_i^2\ [\rho_a]_i\right)$$

$$\mathrm{Var}(\underline{c}'\ \underline{Y}_{II}) = \sigma^2\left((K-k)(1-\rho_t) + \sum_{i\varepsilon S}(M_i - m_i)^2\ \rho_t + \sum_{i\varepsilon \bar{S}} M_i^2\ \rho_t\right)$$

$$\mathrm{cov}(\bar{y}_i, y_{ij}) = \rho_t\ \sigma^2 \text{ for } j \not\in S_i$$

$$\mathrm{cov}(\hat{\beta},\underline{c}'\underline{Y}_{II}) = \frac{\rho_t\sigma^2}{\sum\limits_{i\varepsilon S}\frac{m_i X_i^2}{M_i^2[\rho_a]_i}}\left(\sum_{i\varepsilon S}(M_i - m_i)m_i\ X_i/M_i\ [\rho_a]_i\right)$$

In these relations,

$$\omega_i = \rho_a\ m_i/[\rho_a]_i,\ [\rho_a]_i = 1 - \rho_a + m_i\ \rho_a\ ,\ [\rho_t]_i = 1 - \rho_t + m_i\ \rho_t\ .$$

Substituting these expressions into (2.5.1), we obtain, after simplification, that

$$\frac{E[\hat{T}_B(\rho_a) - T]^2_{\rho_t}}{\sigma^2} = (K-k)(1-\rho_t)+\rho_t \sum_{i\varepsilon\bar{S}} M_i^2 + \sum_{i\varepsilon S}\frac{(M_i-m_i)^2}{[\rho_a]_i^2}\Big(\rho_a m_i(\rho_a-\rho_t)$$

$$- \rho_a\rho_t m_i\ [\rho_a]_i + \rho_t[\rho_a]_i^2\Big)$$

$$+ \frac{2}{\sum\limits_{i\varepsilon S}\frac{m_i X_i^2}{M_i^2[\rho_a]_i}}\left(\sum_{i\varepsilon S}\frac{m_i X_i}{M_i[\rho_a]_i^2}\{(M_i - m_i)(\rho_a - \rho_t)\right.$$

$$\left.\times\ [(1-\rho_a)(\sum_{i\varepsilon S} X_i/[\rho_a]_i - \sum_{i\varepsilon S} m_i X_i/M_i[\rho_a]_i) + \sum_{i\varepsilon\bar{S}} X_i]\}\right)$$

$$+ \frac{1}{\left(\sum\limits_{i\varepsilon S}\frac{m_i\ X_i^2}{M_i^2[\rho_a]_i}\right)^2}\left(\sum_{i\varepsilon S}\frac{m_i\ X_i^2}{M_i^2[\rho_a]_i^2}[\rho_t]_i\left((1-\rho_a)(\sum_{i\varepsilon S} X_i/[\rho_a]_i\right.\right.$$

$$\left.\left. - \sum_{i\varepsilon S} m_i\ X_i/M_i\ [\rho_a]_i) + \sum_{i\varepsilon\bar{S}} X_i\right)^2\right) . \qquad (2.5.2)$$

It is easy to show that (2.5.2) and (2.4.6) are equivalent whenever $\rho_a = \rho_t = \rho$. In particular the two m.s.e.'s are equivalent for $\rho_a = \rho_t = 1$ and any $m_i > 1$. Thus the derivation of (2.4.6), which utilized Theorem 1.1, which in turn depended on V_I^{-1} is still valid even though V_I is singular for $\rho_i = \rho = 1$ and $m_i > 1$. Of course, for this case the covariance matrix of the sampled elements is not positive definite.

Now, from Section 2.4.3 we have agreed to let $m_i = 1$ whenever $\rho_i = 1$. As we saw, this criterion results in V_I being positive definite. We can now show that for a fixed sample size k, $\rho_i = \rho = 1$, and $\sigma_i^2 = \sigma^2$, there is another advantage to allowing $m_i = 1$. From (2.4.6) we have, for this case, and any choice of the m_i that

$$E(\hat{T}_B - T)^2/\sigma^2 = \sum_{i\varepsilon\bar{S}} M_i^2 + (\sum_1^N X_i - \sum_{i\varepsilon S} X_i)^2 / \sum_{i\varepsilon S} X_i^2/M_i^2 \ . \quad (2.5.3)$$

Thus, the m.s.e. is independent of m_i. Hence, if one has a budget which allows for a total of k units to be sampled, and if $k < N$ (number of clusters in population), then (2.5.3) is minimized <u>with respect to the number of clusters sampled</u> by sampling k clusters, each with $m_i = 1$. Of course, the particular k clusters selected should be those which minimize (2.5.3)--see Theorem 2.1. If $k \geq N$ then (2.5.3) is minimized by choosing one unit from each of the N clusters, in which case the m.s.e. is zero.

In Chapter III we shall examine the behavior of (2.5.2) with respect to various choice of ρ_a and ρ_t.

2.6 A Conventional Estimator of the Total

The conventional estimator recommended with simple random sampling at both stages is

$$\hat{T}_N = \sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij} + \sum_{i\varepsilon S} (M_i - m_i)\bar{y}_i + \sum_{i\varepsilon S} M_i \bar{y}_i \sum_{i\varepsilon \bar{S}} X_i / \sum_{i\varepsilon S} X_i \ . \quad (2.6.1)$$

Actually, the estimator given by Cochran [2] is a function of auxiliary variables X_{ij} being available for each secondary unit. However, by letting $X_{ij} = X_i/M_i$ and rewriting the estimator one obtains (2.6.1).

To examine $\hat{T}_N$ for unbiasedness under model (2.2.1), we have

$$\begin{aligned} E(\hat{T}_N) &= \beta \sum_{i\varepsilon S} m_i(X_i/M_i) + \beta \sum_{i\varepsilon S} (M_i - m_i)X_i/M_i + \frac{\beta \sum_{i\varepsilon S} M_i(X_i/M_i) \sum_{i\varepsilon \bar{S}} X_i}{\sum_{i\varepsilon S} X_i} \\ &= \beta\left(\sum_{i\varepsilon S} X_i + \sum_{i\varepsilon \bar{S}} X_i\right) \\ &= E\left(\sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij}\right) \\ &= E(T). \end{aligned}$$

Thus, $\hat{T}_N$ is unbiased for the total T. As in (2.3.2) the estimator is written as the sum of (i) the elements in the sample, (ii) an unbiased estimator of the sum in the sample clusters but not in the sample itself, and (iii) an unbiased estimator of the sum in the unsampled clusters.

We have seen in property (vii) of Section 2.4.4 that the b.l.u.e. $\hat{T}_B$ derived under model (2.2.1) is biased for T when the regression part of the model is incorrect. Furthermore, except in non-cluster models, it is difficult to find conditions for the sample clusters which would result in $\hat{T}_B$ being unbiased. For the conventional estimator $\hat{T}_N$ conditions can be found, as the following theorem illustrates, such that the estimator is unbiased for a wide choice of models.

Theorem 2.2: Consider the model

$$Y_{ij} = (\delta_0\beta_0 + \delta_1\beta_1X_i + \delta_2\beta_2X_i^2 + \ldots + \beta_pX_i^p)/M_i + \varepsilon_{ij}$$

where

$$\delta_i = 0 \text{ or } 1 \; ; \; i = 0,\ldots, p-1 \; ,$$

$E(\varepsilon_{ij}) = 0$, and the y's have the same covariance structure as in the previous model (2.2.1). Let $t = 1,\ldots, p$ if $p \geq 1$ and let $t = 1$ if $p = 0$. Then $\hat{T}_N$ is unbiased for T for any selection of the δ_i and any values of the β_i if and only if for each t the X's satisfy the relation

$$\overline{X_S^t} = \overline{X^t} \tag{2.6.2}$$

where

$$\overline{X_S^t} = \sum_{i\varepsilon S} X_i^t/n$$

and

$$\overline{X^t} = \sum_1^N X_i^t/N \; .$$

Proof: We have

$$E[T-\hat{T}_N] = \delta_0\beta_0\left((N-n)\sum_{i\varepsilon S} X_i - n\sum_{i\varepsilon \bar{S}} X_i\right) / \sum_{i\varepsilon S} X_i + \delta_2\beta_2\left(\sum_{i\varepsilon S} X_i \sum_{i\varepsilon \bar{S}} X_i^2\right.$$

$$\left. - \sum_{i\varepsilon \bar{S}} X_i \sum_{i\varepsilon S} X_i^2\right) / \sum_{i\varepsilon S} X_i$$

$$+ \dots + \beta_p\left(\sum_{i\varepsilon S} X_i \sum_{i\varepsilon \bar{S}} X_i^p - \sum_{i\varepsilon \bar{S}} X_i \sum_{i\varepsilon S} X_i^p\right) / \sum_{i\varepsilon S} X_i$$

Now, this expression is zero for $\delta_i = 0$ or 1 and any values of the β_i if and only if

$$\frac{\sum_{i\varepsilon S} X_i}{n} = \frac{\sum_{i\varepsilon \bar{S}} X_i}{N-n}$$

$$\frac{\sum_{i\varepsilon S} X_i}{\sum_{i\varepsilon \bar{S}} X_i} = \frac{\sum_{i\varepsilon S} X_i^2}{\sum_{i\varepsilon \bar{S}} X_i^2}$$

$$\vdots$$

$$\frac{\sum_{i\varepsilon S} X_i}{\sum_{i\varepsilon \bar{S}} X_i} = \frac{\sum_{i\varepsilon S} X_i^p}{\sum_{i\varepsilon \bar{S}} X_i^p}$$

But these conditions can be written as

$$\bar{X}_S = \bar{X}_{\bar{S}} \; ; \quad \overline{X_S^2} = \overline{X_{\bar{S}}^2} \; ; \; \dots \; ; \; \overline{X_S^p} = \overline{X_{\bar{S}}^p} \; .$$

But $\bar{X}_S = \bar{X}_{\bar{S}} \Rightarrow \bar{X}_S = \bar{X}$ and similarly for the other conditions, Q.E.D.

Sample clusters which satisfy relation (2.6.2) are said to be <u>balanced</u> on the X's up to and including the p^{th} moment.

<u>2.6.1 The m.s.e. and Optimum Selection of the Clusters.</u> Since $\hat{T}_N$ is unbiased for T, we have

$$E(\hat{T}_N - T)^2 = \mathrm{Var}(\hat{T}_N - T) \quad .$$

Letting $\underline{c}$ be a vector of (K-k) ones and $\underline{Y}_{II}$ the vector of all unsampled units, it follows that

$$\mathrm{Var}(\hat{T}_N - T) = \mathrm{Var}\left[\sum_{i\varepsilon S} (M_i - m_i)\bar{y}_i + \sum_{i\varepsilon S} M_i \bar{y}_i \sum_{i\varepsilon \bar{S}} X_i / \sum_{i\varepsilon S} X_i - \underline{c}' \underline{Y}_{II}\right]$$

$$= \sum_{i\varepsilon S} (M_i - m_i)^2 \mathrm{Var}\ \bar{y}_i + \left(\sum_{i\varepsilon\bar{S}} X_i / \sum_{i\varepsilon S} X_i\right)^2 \sum_{i\varepsilon S} M_i^2 \mathrm{Var}\ \bar{y}_i$$

$$+ \mathrm{Var}(\underline{c}' \underline{Y}_{II}) + 2\left(\sum_{i\varepsilon\bar{S}} X_i\right) \sum_{i\varepsilon S} (M_i - m_i) M_i (\mathrm{Var}\ \bar{y}_i) / \sum_{i\varepsilon S} X_i$$

$$- 2 \sum_{i\varepsilon S} (M_i - m_i) \sum_{j\notin S_i} \mathrm{cov}(\bar{y}_i, y_{ij})$$

$$- 2 \sum_{i\varepsilon\bar{S}} X_i \left(\sum_{i\varepsilon S} M_i \sum_{j\notin S_i} \mathrm{cov}(\bar{y}_i, y_{ij})\right) / \sum_{i\varepsilon S} X_i \ . \qquad (2.6.3)$$

We have

$$\mathrm{Var}(\bar{y}_i) = \sigma_i^2 [1 + \rho_i (m_i - 1)]/m_i$$

$$\mathrm{Var}(\underline{c}'\underline{Y}_{II}) = \sum_{i\varepsilon S} (M_i - m_i) \sigma_i^2 [1 - \rho_i + \rho_i (M_i - m_i)]$$

$$+ \sum_{i\varepsilon\bar{S}} M_i \sigma_i^2 (1 - \rho_i + \rho_i M_i)$$

$$\mathrm{cov}(\bar{y}_i, y_{ij}) = \rho_i \sigma_i^2 \quad , j\notin S_i \ .$$

Substituting these expressions into (2.6.3), we obtain, after simplification, that

$$E(\hat{T}_N - T)^2 = \sum_{i\varepsilon S} (M_i - m_i)\, \sigma_i^2 (1 - \rho_i) + \sum_{i\varepsilon \bar{S}} M_i\, \sigma_i^2 (1 - \rho_i + \rho_i M_i)$$

$$+ \sum_{i\varepsilon S} M_i^2 \rho_i \sigma_i^2 \left[\sum_{i\varepsilon \bar{S}} X_i / \sum_{i\varepsilon S} X_i\right]^2$$

$$+ \sum_{i\varepsilon S} \frac{\sigma_i^2 (1 - \rho_i)}{m_i} \left[(M_i - m_i) + M_i \sum_{i\varepsilon \bar{S}} X_i / \sum_{i\varepsilon S} X_i\right]^2 .$$

Now, letting $\sigma_i^2 = \sigma^2$ and $\rho_i = \rho$ for all $i = 1,\ldots, N$; the m.s.e. becomes

$$E(\hat{T}_N - T)^2/\sigma^2 = (K - k)(1 - \rho) + \rho\left[\sum_{i\varepsilon \bar{S}} M_i^2 + \sum_{i\varepsilon S} M_i^2 \left(\sum_{i\varepsilon \bar{S}} X_i / \sum_{i\varepsilon S} X_i\right)^2\right]$$

$$+ \frac{(1 - \rho)}{\left[\sum_{i\varepsilon S} X_i\right]^2} \sum_{i\varepsilon S} \left(M_i \sum_{1}^{N} X_i - m_i \sum_{i\varepsilon S} X_i\right)^2 / m_i . \qquad (2.6.4)$$

Explicit conditions, similar to those for Theorem 2.2 which minimize (2.6.4) are difficult to find. However, one notes that in the case of equal cluster sizes; i.e., $M_i = M$; and $\rho \geq 0$, then selection of clusters with the largest X's will minimize (2.6.4).

<u>2.6.2 Optimum Number of Sample Clusters for Fixed Sample Size and Equal Cluster Sizes</u>. We have seen in Section 2.6.1 that when $M_i = M$, the largest clusters in terms of X should be sampled so as to minimize the m.s.e. (2.6.4). This result does not answer the question, however, of how many clusters should be sampled. That is, if the total number of units, k, to be sampled is fixed, then should all k units be selected from the largest cluster if possible, or perhaps the units should be allocated among the largest two clusters, or perhaps only one unit should be chosen from each of the k largest clusters if possible.

Due to the large number of parameters present in (2.6.4), a complete answer to this question is difficult. However, for special cases the behavior of the m.s.e. for increasing number of sample clusters can be determined by the following three theorems.

<u>Theorem 2.3</u> Let $\rho_i = \rho = 0$, $\sigma_i^2 = \sigma^2$, $M_i = M$, $m_i = m$, and the number of units, k, to be sampled be fixed. Then the minimum number of largest (in terms of X) clusters will minimize $E(\hat{T}_N - T)^2$.

<u>Proof</u>: We have seen in Section 2.6.1 that the largest clusters should be sampled for a given number, n, of sample clusters. To show that n should be the smallest possible value compatible with the fixed sample size k and equal m's, consider two samples, S and S', of clusters containing n_1 and n_2 clusters respectively. Let

$$n_1 = p_1 N ,$$

$$n_2 = p_2 N ,$$

and

$$0 < p_1 < p_2 \leq 1 .$$

Let the largest clusters be sampled for S and S'. So, the clusters in S are a subset of clusters in S'. Let the number of units sampled per cluster for S and S' be m_1 and m_2 respectively. Then for fixed k, we have from (2.6.4) that

$$[(\text{m.s.e.})_S - (\text{m.s.e.})_{S'}]/\sigma^2 = \frac{1}{\left(\sum\limits_{i\varepsilon S} X_i\right)^2} n_1 (M\sum_1^N X_i - m_1 \sum_{i\varepsilon S} X_i)^2/m_1$$

$$- \frac{1}{\left(\sum\limits_{i\varepsilon S'} X_i\right)^2} n_2 (M\sum_1^N X_i - m_2 \sum_{i\varepsilon S'} X_i)^2/m_2$$

$$= \frac{1}{k(\sum_{i\varepsilon S} X_i)^2}(Mn_1\sum_1^N X_i - k\sum_{i\varepsilon S} X_i)^2 - \frac{1}{k(\sum_{i\varepsilon S'} X_i)^2}(Mn_2\sum_1^N X_i - k\sum_{S'} X_i)^2$$

$$= \frac{1}{k}\left[M^2(\sum_1^N X_i)^2\left[\frac{n_1^2}{(\sum_S X_i)^2} - \frac{n_2^2}{(\sum_{S'} X_i)^2}\right] - 2Mk\sum_1^N X_i\left[\frac{n_1}{\sum_{i\varepsilon S} X_i} - \frac{n_2}{\sum_{i\varepsilon S'} X_i}\right]\right]$$

$$= \frac{M\sum_1^N X_i}{k}\left[M\sum_1^N X_i\left(\frac{n_1}{\sum_{i\varepsilon S} X_i} + \frac{n_2}{\sum_{i\varepsilon S'} X_i}\right) - 2k\right]\left(\frac{n_1}{\sum_{i\varepsilon S} X_i} - \frac{n_2}{\sum_{i\varepsilon S'} X_i}\right)$$

$$= \frac{M\sum_1^N X_i}{k\sum_{i\varepsilon S} X_i \sum_{i\varepsilon S'} X_i}\left[(n_1\sum_{i\varepsilon S'} X_i - n_2\sum_{i\varepsilon S} X_i)\{M\sum_1^N X_i\left(\frac{n_1\sum_{i\varepsilon S'} X_i + n_2\sum_{i\varepsilon S} X_i}{\sum_{i\varepsilon S} X_i \sum_{i\varepsilon S'} X_i}\right)\right.$$

$$\left. - n_1 m_1 - n_2 m_2\}\right]$$

$$= \frac{M\sum_1^N X_i}{k\sum_{i\varepsilon S} X_i \sum_{i\varepsilon S'} X_i}(n_1\sum_{i\varepsilon S'} X_i - n_2\sum_{i\varepsilon S} X_i)\left[\left(\frac{M\sum_1^N X_i}{\sum_{i\varepsilon S} X_i} - m_1\right)n_1\right.$$

$$\left. + \left(\frac{M\sum_1^N X_i}{\sum_{i\varepsilon S'} X_i} - m_2\right)n_2\right] .$$

Now S' contains n_2/n_1 times as many clusters as S. Furthermore, S' consists of those clusters in S together with clusters with smaller X's than those in S. Hence, it must follow that

$$\frac{n_2}{n_1}\sum_{i\varepsilon S} X_i \geq \sum_{i\varepsilon S'} X_i \quad .$$

Therefore,

$$[(\text{m.s.e.})_S - (\text{m.s.e.})_{S'}]/\sigma^2 \leq 0 \ .$$

So, if the number of sampled clusters is increased, the m.s.e. cannot decrease. It follows then that the minimum number of sample clusters will minimize the m.s.e..

Intuitively, one would hope that whenever $\rho_i = \rho = 1$, that sampling the maximum number of clusters would minimize the m.s.e.. This would be desirable since as seen in Section 2.4.3 complete knowledge of all secondary units in the i^{th} sample cluster is obtained whenever $\rho_i = 1$. Now, we have agreed to sample only one element from each cluster for which $\rho_i = 1$--see Section 2.4.3. So, the sample size k is just equal to the number of clusters sampled for this case. An analogy to the previous theorem, which requires a fixed sample size, then is not possible for the case $\rho_i = 1$ since a change in the number of clusters sampled implies a change in the total number of units sampled.

A major theoretical drawback of the conventional estimator $\hat{T}_N$ is the fact that in certain cases the m.s.e. (2.6.4) may actually increase as the number of clusters sampled (equivalently, as the total sample size) increase whenever $\rho_i = \rho = 1$. For example, let

$$M_i = M, \ \sigma_i^2 = \sigma^2, \text{ and } \sum_1^N X_i = 10^6 \ .$$

Let the two largest X's be 1000 and 1, and let the sample S consist of the largest X and the sample S' consist of the two largest X's. We have then that

$$[(\text{m.s.e.})_S - (\text{m.s.e.})_{S'}]/\sigma^2 = M^2\ [1 + \left(\frac{990000}{1000}\right)^2 - 2\left(\frac{998999}{1001}\right)^2$$

$$= (-9.94\times 10^5)\ M^2\ .$$

The following theorem gives sufficient conditions on the behavior of the X's such that the m.s.e. will decrease as the number of sampled clusters increase whenever $\rho_i = \rho = 1$.

Theorem 2.4 Let $\rho_i = \rho = 1$, $\sigma_i^2 = \sigma^2$, and $M_i = M$. Also, let S, S', n_1, n_2, p_1, and p_2 be defined as in Theorem 2.3. Let

$$\left(\sum_{i\epsilon S'} X_i\right)^2 > \frac{n_2}{n_1}\left(\sum_{i\epsilon S} X_i\right)^2\ .$$

Then

$$E(\hat{T}_N - T)_S^2 > E(\hat{T}_N - T)_{S'}^2\ .$$

Proof: We have from (2.6.4) that

$$\begin{aligned}
[(\text{m.s.e.})_S - (\text{m.s.e.})_{S'}]/\sigma^2 &= M^2\Bigg(n_2 - n_1 + n_1[(\sum_1^N X_i)^2 - 2\sum_1^N X_i \sum_{i\epsilon S} X_i \\
&\quad + (\sum_{i\epsilon S} X_i)^2]/(\sum_{i\epsilon S} X_i)^2 \\
&\quad - n_2[(\sum_1^N X_i)^2 - 2\sum_1^N X_i \sum_{i\epsilon S'} X_i + (\sum_{i\epsilon S'} X_i)^2]/(\sum_{i\epsilon S'} X_i)^2\Bigg) \\
&= M^2\Bigg(n_1[(\sum_1^N X_i)^2 - 2\sum_1^N X_i \sum_{i\epsilon S} X_i]/(\sum_{i\epsilon S} X_i)^2 \\
&\quad - n_2[(\sum_1^N X_i)^2 - 2\sum_1^N X_i \sum_{i\epsilon S'} X_i]/(\sum_{i\epsilon S'} X_i)^2\Bigg)
\end{aligned}$$

$$= \frac{M^2 \sum_1^N X_i}{(\sum_{i\varepsilon S} X_i)^2 (\sum_{i\varepsilon S'} X_i)^2} \left\{ \sum_1^N X_i [n_1(\sum_{i\varepsilon S'} X_i)^2 - n_2(\sum_{i\varepsilon S} X_i)^2] \right.$$

$$\left. + 2 \sum_{i\varepsilon S} X_i \sum_{i\varepsilon S'} X_i [n_2 \sum_{i\varepsilon S} X_i - n_1 \sum_{i\varepsilon S'} X_i] \right\} .$$

But

$$\sum_{i\varepsilon S'} X_i \leq \frac{n_2}{n_1} \sum_{i\varepsilon S} X_i ,$$

and by the conditions of the theorem

$$n_1(\sum_{S'} X_i)^2 - n_2(\sum_{i\varepsilon S} X_i)^2 > 0 .$$

Q.E.D.

The next theorem gives the behavior of the m.s.e. in the case of balanced samples.

Theorem 2.5 Let $\rho_i = \rho \geq 0$, $\sigma_i^2 = \sigma^2$, $M_i = M$, $m_i = m$, and the number of units, k, to be sampled be fixed. Then, if for any selection of the clusters we have $\overline{X}_S = \overline{X}$, it follows that the maximum possible number of sample clusters will minimize the m.s.e. $E(\hat{T}_N - T)^2$.

Proof: Consider two samples of first-stage units, S and S', containing n_1 and n_2 clusters respectively. Let $n_1 < n_2$ and $k = pK$ where $0 < p \leq 1$. Then, from (2.6.4) and using the restriction that the sample clusters are balanced, we have

$$(\text{m.s.e.})_S = (K - pK)(1 - \rho) + \rho M^2[N - n_1 + n_1(\frac{N-n_1}{n_1})^2]$$

$$+ \frac{(1-\rho)N^2}{n_1^2} \frac{n_1}{m_1} (M - m_1 n_1/N)^2$$

$$= K(1 - p)(1 - \rho) + \rho M^2 N(N/n_1 - 1)$$

$$+ \frac{(1-\rho)}{n_1 m_1} (MN - m_1 n_1)^2$$

$$= K(1 - p(1 - \rho) + \rho KM(N/n_1 - 1) + \frac{(1-\rho)}{pK} (K - pK)^2$$

$$= K(1 - p)(1 - \rho) + \rho KM(N/n_1 - 1) + \frac{(1-\rho)}{p} K(1 - p)^2. \tag{2.6.5}$$

So,

$$(\text{m.s.e.})_S - (\text{m.s.e.})_{S'} = \rho KM(N/n_1 - N/n_2) \geq 0 .$$

Hence, the addition of more clusters cannot increase the m.s.e.

Note that if the conditions of Theorem 2.5 are met, and if the number of clusters to be sampled is fixed; then the m.s.e. is independent of the particular choice of clusters one makes.

2.6.3 Optimum Allocation of Second-Stage Sample Units

Theorem 2.6 Let $\rho_i = \rho$ and $\sigma_i^2 = \sigma^2$. Then for a fixed total sample size k and a given selection of clusters, the values of m_i which minimize (2.6.4) are obtained by proportional allocation. That is

$$m_i = kM_i / \sum_{i\varepsilon S} M_i \ ; \ i\varepsilon S .$$

Proof: We wish to minimize (2.6.4) subject to the constraint that $\sum_{i\in S} m_i = k$. Letting τ be a Lagrangian multiplier, we have

$$\frac{\partial}{\partial m_i}\left[E(\hat{T}_N - T)^2 + \tau\left(\sum_{i\in S} m_i - k\right)\right]$$

$$= -\sigma^2(1-\rho)\left\{\left(\sum_1^N X_i\right)^2 M_i^2 / \left(\sum_{i\in S} X_i\right)^2 m_i^2 - 1\right\} + \tau .$$

Setting the derivative equal to zero, we obtain

$$m_i = M_i \sum_1^N X_i / \sum_{i\in S} X_i \ \{\tau/\sigma^2(1-\rho) + 1\}^{\frac{1}{2}}$$

and

$$k = \sum_{i\in S} m_i = \sum_{i\in S} M_i \sum_1^N X_i / \sum_{i\in S} X_i \ \{\tau/\sigma^2(1-\rho) + 1\}^{\frac{1}{2}} .$$

Or,

$$\frac{m_i}{k} = \frac{M_i}{\sum_{i\in S} M_i} .$$

2.6.4 Comparison of Conventional Estimator and b.l.u.e. Based on the results in Chapters I and II, the following comparisons can be made between $\hat{T}_B$ and $\hat{T}_N$: (i) Since $\hat{T}_N$ is unbiased for T and since $\hat{T}_B$ is the b.l.u.e., it follows that for any choice of clusters the m.s.e. of $\hat{T}_N$ is no smaller than the m.s.e. of $\hat{T}_B$. (ii) In using $\hat{T}_N$, one can protect against bias caused by lower or higher order terms possibly being present in the model. The protection is achieved, conceptually, by selecting clusters which are balanced on the X's. The b.l.u.e. offers no convenient method of protecting against bias. (iii) In certain situations, i.e., if the conditions of Theorems 2.3, 2.4, and 2.5 are met; the behavior of the m.s.e. of $\hat{T}_N$ as the number of sample clusters increases can be determined. A similar type of analysis for $\hat{T}_B$ appears to be intractable. (iv) Optimum allocation of second-stage units can be determined when using $\hat{T}_N$--Theorem 2.6. Optimum allocation for $\hat{T}_B$ is very difficult.

It should be pointed out that in the conventional theory, see Sukhatme and Sukhatme [19], the estimator $\hat{T}_N$ is biased with respect to simple random sampling at both stages. Also, in the conventional theory the m.s.e. of $\hat{T}_N$ can only be approximated, and optimum properties regarding the choice of clusters and the type of allocation of the second-stage units are hard to find.

2.7 Selection of Second-Stage Units

The relations for the m.s.e.'s of $\hat{T}_B$, $\hat{T}_B$ (ρ_a), and $\hat{T}_N$ imply that they are completely independent of the second-stage units that are selected. Now, the derivation of these m.s.e.'s was made under the assumption that one has no prior knowledge concerning the actual values of the m_i second-stage units selected. However, the following theorem shows that when using $\hat{T}_N$ if the second-stage units satisfy a certain condition, the m.s.e. (2.6.4) can be reduced for a given sample of clusters.

Theorem 2.7. Let $\rho_i = \rho$ and $\sigma_i^2 = \sigma^2$. Let the m_i values and the choice of clusters to be sampled, S, be given. Let $\bar{y}_i$ be the mean of the m_i sampled second-stage units and $\bar{Y}_i$ the mean of all M_i second-stage units in the i^{th} sampled cluster. Then, if

$\bar{y}_i = \bar{Y}_i$ for each sampled cluster, the m.s.e. of $\hat{T}_N$ is less than (2.6.4).

<u>Proof</u>: If $\bar{y}_i = \bar{Y}_i$ for all $i \epsilon S$, then the first two terms in (2.6.1) give the <u>exact</u> sum of <u>all</u> secondary units in the sampled clusters. We thus have that

$$\begin{aligned} \mathrm{Var}(\hat{T}_N - T) &= \mathrm{Var}[\sum_{i\epsilon S} M_i \bar{y}_i \sum_{i \epsilon \bar{S}} X_i / \sum_{i\epsilon S} X_i - \sum_{i\epsilon \bar{S}} \sum_{j=1}^{M_i} y_{ij}] \\ &= \sigma^2 [\sum_{i\epsilon \bar{S}} X_i / \sum_{i\epsilon S} X_i]^2 \sum_{i\epsilon S} M_i^2 [1 + \rho(m_i - 1)]/m_i \\ &+ \sigma^2 \sum_{i\epsilon \bar{S}} M_i (1 - \rho + \rho M_i) \ . \end{aligned}$$

or,

$$\begin{aligned} \mathrm{Var}(\hat{T}_N - T)/\sigma^2 &= (K - \sum_{i\epsilon S} M_i)(1-\rho) + \rho\{ \sum_{i\epsilon \bar{S}} M_i^2 + \sum_{i\epsilon S} M_i^2 (\sum_{i\epsilon \bar{S}} X_i / \sum_{i\epsilon S} X_i)^2\} \\ &+ \frac{(1-\rho)}{(\sum_{i\epsilon S} X_i)^2} \sum_{i\epsilon S} (M_i \sum_1^N X_i - M_i \sum_{i\epsilon S} X_i)^2 / m_i \ . \end{aligned}$$

It is easy to see that this expression is smaller than (2.6.4).

Thus, when using $\hat{T}_N$, one would like to select the second-stage units such that $\bar{y}_i = \bar{Y}_i$ for each $i\epsilon S$. In attempting to achieve these equalities, one may consider the following criteria:

1. Suppose previous values are available on some of the secondary units for each sample cluster.
2. Suppose that for each $i\epsilon S$, a particular selection of m_i second-stage units would result in the previous sample mean of these units

being equal to the previous population mean.

3. Then if it is felt likely that the same m_i units would have a present mean close to the present population mean for each $i \varepsilon S$, these particular units would be the ones desirable to be sampled.

Now, obviously if no previous values of any secondary units were available, the criteria could not be used. Cochran [2] show that for any finite population of M_i elements all simple random samples† of m_i elements will result in the sample means being unbiased for the population mean with respect to simple random sampling. Hence, in this case a simple random sample of m_i secondary units for each $i \varepsilon S$ is recommended. Of course, random sampling would also be recommended if it is felt that 3. is not met.

---

†Simple random sampling (without replacement) is that sampling scheme such that for each $i \varepsilon S$ each of the $\binom{M_i}{m_i}$ possible samples has an equal chance of being selected.

III. A COMPUTER IMPLEMENTED STUDY

3.1 Remarks and Objectives

Chapters I and II were mainly concerned with the theoretical analysis of the b.l.u.e. $\hat{T}_B$, the suboptimal estimator $\hat{T}_B\ (\rho_a)$, and a conventional estimator $\hat{T}_N$. In order to further analyze the performance of these estimators for a wide range of parameter values, it is necessary to utilize a computer-implemented study. The estimators are to be compared in terms of their m.s.e.'s under the assumption of equal variances and correlations for all clusters.

The objectives of the study are as follows: Let K, k, N, n, M_i, m_i, and the X_i be fixed. Let a method of selecting the n clusters be given, and let $\rho_i = \rho$ and $\sigma_i^2 = \sigma^2$ for i = 1, ..., N.

1. Suppose one uses the sub-optimum estimator $\hat{T}_B\ (\rho_a)$ where the true (unknown) value of the correlation is, say, ρ_t. Then, it is of interest to investigate the relative m.s.e.'s of $\hat{T}_B\ (\rho_a)$ and the optimum estimator $\hat{T}_B$ as ρ_a departs from ρ_t. This investigation will then indicate the sensitivity of $\hat{T}_B\ (\rho_a)$ to choices of ρ_a.

2. Since $\hat{T}_N$ is never the optimum estiator for any choice of ρ; that is, as seen from (2.6.1) and (2.3.2) $\hat{T}_N \neq \hat{T}_B$ for any value of ρ; it may turn out that one of the sub-optimum estimators will out-perform $\hat{T}_N$ for various values of ρ. The two simplest sub-optimum estimators are those obtained by assuming ρ to be zero or one. Thus, the relative

m.s.e.'s of $\hat{T}_B$ (0), $\hat{T}_B$ (1), and $\hat{T}_N$ are investigated.

3. If one of $\hat{T}_B$ (0), $\hat{T}_B$ (1), and $\hat{T}_N$ consistently out perform the other two; then it is of interest to compare this estimator with $\hat{T}_B$ (ρ_a) for various choices of ρ_a. This investigation will indicate whether in certain cases one of the three simpler estimators may be safely used instead of the more complex $\hat{T}_B$ (ρ_a).

4. By Theorem 2.5 we know that the maximum number of sample clusters will minimize the m.s.e. of $\hat{T}_N$ for the case of balanced clusters and equal M_i. However, it is of interest to know the percent decrease in m.s.e. as the number of sample clusters increases. This investigation could then be used to indicate whether or not it would be worthwhile, in terms of added costs, to actually sample more clusters.

3.2 Parameter Values

For the first three objectives the following parameter values are used:

$$N = 20$$

$$M_i = \begin{cases} i \\ 300 + 10i \end{cases} ; \; i = 1, \ldots, 20$$

$$K = \begin{cases} 210; & M_i = i \\ 8100; & M_i = 300 + 10i \end{cases}$$

$$k = 0.05 \; K = \begin{cases} 10; & M_i = i \\ 405; & M_i = 300 + 10i \end{cases}$$

$$n = \begin{cases} 5, 10 & ; M_i = i \\ 5, 10, 15, 20; & M_i = 300 + 10i \end{cases}$$

$\rho_i = \rho,\ 0 \le \rho \le 1$ (Actual values used are shown in Tables)

$\sigma_i^2 = \sigma^2$ (values are not necessary since ratios of m.s.e's are always calculated)

$m_i = kM_i / \sum_{i \in S} M_i$ (proportional allocation, non-integer values of m_i are allowed)

The values of the X's are allowed to range from 1.0 to 2.0 and from 1.0 to 201.0. Furthermore, four types of spacings of the X's within each of the two ranges are considered. They are referred to as TYPE I, TYPE II, TYPE III, and TYPE IV spacings. The histograms of the first three types are as follows:

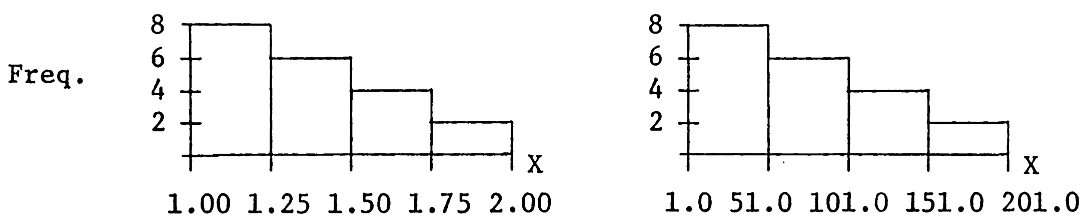


FIGURE 3.1 Histogram of TYPE I X's

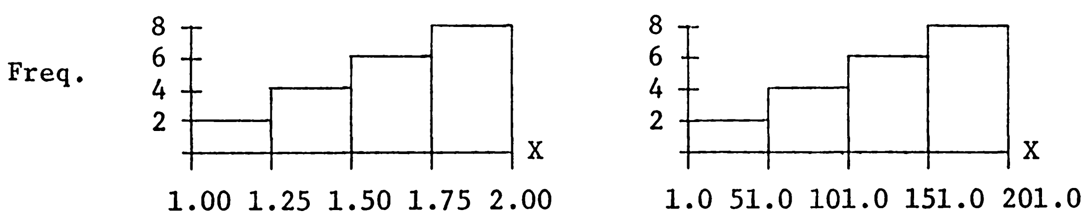


FIGURE 3.2 Histogram of TYPE II X's

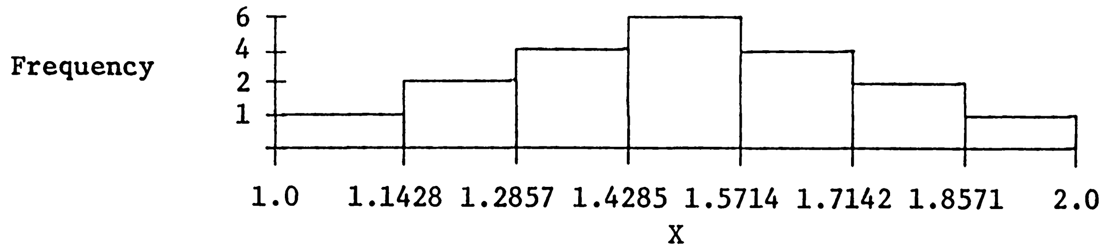


FIGURE 3.3 Histogram of TYPE III X's

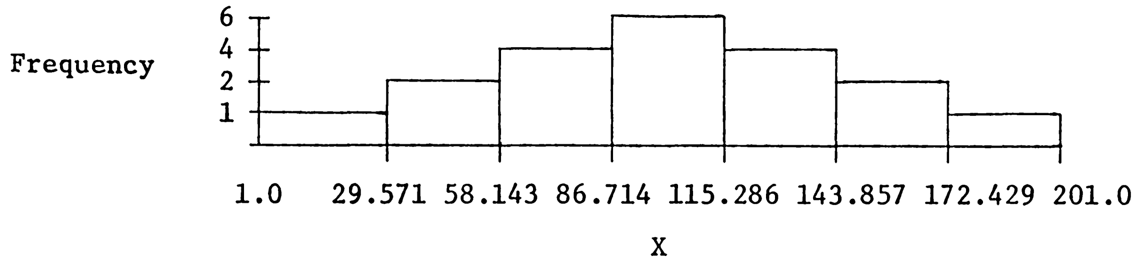


Figure 3.3a

For these three types of spacings, the X's are assumed to be equally spaced within any given subinterval. For example, for TYPE II X's distributed between 1.0 and 201.0 there are four values in the subinterval 51 to 101. These four values are equally spaced within this interval; i.e., they are 61, 71, 81, and 91.

Note that for a given range on the X's, a TYPE I spacing implies more small X's than large, a TYPE II spacing implies more large X's than small, and a TYPE III spacing implies that X's are spaced symmetrically about some point.

The TYPE IV spacing on the X's is a <u>uniform</u> spacing within the given range. The X values are given by

$$X_i = \begin{cases} 1.0 + (i-1)(0.0526) \text{ for } 1.0 \leq X \leq 2.0 \\ 1.0 + (i-1)(10.5263) \text{ for } 1.0 \leq X \leq 201.0 \text{ .} \end{cases}$$

The X's are always assumed to increase with the M's. The values of the X's and corresponding M's are given in Table 3.32 at the end of Section 3.6. The X/M values increase with M for the cases $M_i = i$, $X\sim 1.0\text{-}2.0$, and TYPE IV spacing; $M_i = 300+10i$, $X\sim 1.0\text{-}2.0$, and TYPE IV spacing; and $M_i = 300+10i$, $X\sim 1.0\text{-}201.0$, and all four spacings. For these cases Theorem 2.1 says that the largest clusters should be selected to use with $\hat{T}_B$. For the other cases and estimators one has no criterion to use in the choice of sample clusters. For convenience the <u>largest</u> clusters are <u>always</u> the ones sampled.

Regarding objective 4 and the parameter values used, consider two samples S and S' containing n_1 and n_2 clusters, respectively. Let $n_1 = p_1N$, $n_2 = p_2N$, and $k = pK$ where $0 < p_1 < p_2 \leq 1$ and $0 < p \leq 1$. Then, using (2.6.5), the percent decrease in m.s.e. by going from S to S' is given by

$$\frac{\rho M(1/p_1 - 1/p_2)}{(1-p)(1-\rho)/p + \rho M(1/p_1 - 1)} \times 100 \text{ .} \tag{3.2.1}$$

The values of the parameters used in the computer study are

$$M = 10 \text{ and } 100;\ p = 0.05\ ;\ \rho = 0.0, 0.2, \ldots, 1.0;$$

$$p_1 = 0.05, 0.10, \ldots, 0.95; \text{ and}$$

$$p_2 = p_1 + .05$$

3.3 Description of Tables

Tables 3.1--3.4 give the sensitivity of the m.s.e. of $\hat{T}_B\ (\rho_a)$ as ρ_a departs from the true value of the correlation, ρ_t. In particular, the tables give the ratio

$$E[\hat{T}_B\ (\rho_a) - T]^2_{\rho_t} / E[\hat{T}_B - T]^2_{\rho_t} = (2.5.2)/(2.4.6).$$

For example, from Table 3.1 (b) if the true value of ρ is 0.8, but one assumes the value of ρ to be 0.0, then the ratio of the two m.s.e.'s is 1.003. This implies that the m.s.e. of the optimum estimator $\hat{T}_B$ is increased by only 0.3%. So, the values in Tables 3.1--3.4 can be used to determine the percent increase in the m.s.e. of the optimum estimator $\hat{T}_B$ by using the sub-optimum estimator $\hat{T}_B\ (\rho_a)$.

Tables 3.1--3.4 include cases when $\rho_a = 1$ and also when $\rho_t = 1$. As pointed out in Section 2.5 the two m.s.e.'s (2.5.2) and (2.4.6) are valid for these cases. However, the optimum estimator for $\rho_t = 1$ would be obtained by using $\hat{T}_B$ with $\rho = 1$ and $m_i = 1$, since as was mentioned in Section 2.5 this estimator has smaller m.s.e. than the same estimator with $m_i > 1$. Hence, the values in the first four tables for these cases are not, strictly speaking, the ratio of the m.s.e. of $\hat{T}_B\ (\rho_a)$ to the optimum estimator. They are simply the ratio of (2.5.2) to (2.4.6) with the m_i possibly being larger than one.

Tables 3.5--3.8 are the column averages for Tables 3.1--3.4 for each ρ_a, and represent the average percent increase in the optimum m.s.e. obtained by using the sub-optimum estimator $\hat{T}_B\ (\rho_a)$.

For example, from Table 3.5 (b) and TYPE I X's, if one always assumes a value of ρ to be 0.0, then, on the average, use of $\hat{T}_B$ (0.0) will increase the optimum m.s.e. $\hat{T}_B$ by 0.18%.

Tables 3.9--3.12 give the ratio of the m.s.e. of $\hat{T}_B$ (0) to $\hat{T}_N$ for various values of ρ_t. That is, the entries give the ratio of (2.5.2), where $\rho_a = 0$, to (2.6.4).

Tables 3.13--3.16 give the ratio of the m.s.e. of $\hat{T}_B$ (1) to $\hat{T}_N$ for various values of ρ_t. That is, the entries give the ratio of (2.5.2) where $\rho_a = 1$ to (2.6.4).

Tables 3.17--3.20 give the ratio of the m.s.e. of $\hat{T}_B$ (1) to $\hat{T}_B$ (0) for various values of ρ_t. That is, the entries give the ratio of (2.5.2) where $\rho_a = 1$ to (2.5.2) where $\rho_a = 0$.

Tables 3.21--3.26 give the averages, with respect to ρ_t, of the ratios in Tables 3.9--3.20. These tables can be used to determine the average relative performances of $\hat{T}_B$ (0), $\hat{T}_B$ (1), and $\hat{T}_N$.

For example, from Table 3.25 (b), n = 10, and TYPE I X's; the entry 0.982 implies that on the average (average taken with respect to ρ_t) the m.s.e. of $\hat{T}_B$ (1) is 1.8% smaller than that of $\hat{T}_B$ (0).

At this point, a result of the previous tables should be given before describing the next set. From Tables 3.21--3.26 one notes that the estimator $\hat{T}_B$ (1) appears to consistently outperform either $\hat{T}_B$ (0) or $\hat{T}_N$ on the average. So, regarding objective 3 of Section 3.1, $\hat{T}_B$ (1) is the estimator chosen to compare with $\hat{T}_B$ (ρ_a).

Tables 3.27 and 3.28 then show the ratio of the m.s.e. of $\hat{T}_B$ (ρ_a) to that of $\hat{T}_B$ (1) for various choices of ρ_a and ρ_t, and can be used to determine the relative performance of the sub-optimum estimator to the estimator $\hat{T}_B$ (1).

For example, from Table 3.27 (b) if one <u>assumes</u> ρ to be zero in $\hat{T}_B(\rho_a)$ but ρ is <u>actually</u> 0.4 then the m.s.e. of $\hat{T}_B(\rho_a)$ is 3.5% greater than that of $\hat{T}_B(1)$.

Tables 3.29 and 3.30 give the average, with respect to ρ_t, of the entries in Tables 3.27 and 3.28 for each choice of ρ_a. These tables given an indication of the average relative performance of $\hat{T}_B(\rho_a)$ to $\hat{T}_B(1)$.

For example, from Table 3.30 (b) and X TYPE I, the entry 1.090 corresponding to $\rho_a = 0.2$ implies that if one always <u>assumes</u> ρ to be 0.2, then on the average the m.s.e. of $\hat{T}_B(0.2)$ will be 9% greater than that of $\hat{T}_B(1)$.

Table 3.31 gives the percent decrease in m.s.e. of $\hat{T}_N$ for the case of balanced clusters and equal M_i. The entries in the table are values of (3.2.1).

Table 3.32 gives the M and X values used in the study.

<u>3.4 Results of the Study</u>

Regarding the sensitivity of $\hat{T}_B(\rho_a)$ to choices of ρ_a, Tables 3.1--3.8 show the following:

1. In all cases, sampling only 25% (n = 5) of the total clusters results in $\hat{T}_B(\rho_a)$ being very insensitive to ρ_a. From Tables 3.5--3.8 one sees that for any ρ_a and X TYPE the average percent increase in the m.s.e. of the optimum estimator is less than one percent when n = 5.
2. These results also apply in most cases even when sampling 50% (n = 10) of the total clusters. Tables 3.5--3.8 show that the average percent increase in the optimum m.s.e. is no greater than 5.55 when n = 10.

3. If more than 50% of the clusters are sampled, $\hat{T}_B(\rho_a)$ becomes more sensitive to choices of ρ_a for many of the cases--especially when all clusters are sampled and the X's range over a wide interval as shown in Tables 3.4 (c), (d), (g), (h), (k), (l), (o), and (p). However, even when more than 50% of the clusters are selected, Tables 3.3 and 3.4 show that for nearly all cases the m.s.e. of $\hat{T}_B(\rho_a)$ is nearly equal to that of the optimum estimator provided that ρ_a is within 0.2 of ρ_t.
4. In nearly all cases $\hat{T}_B(\rho_a)$ is <u>least</u> sensitive to values of ρ_a if the X's are of TYPE II and the <u>most</u> sensitive for TYPE I X's.
5. The sensitivity appears to increase as the range of the X's increases.

Regarding the relative performances of the estimators $\hat{T}_B$ (0), $\hat{T}_B$ (1), and $\hat{T}_N$, Tables 3.9--3.26 show the following:
1. From Tables 3.13--3.16 it is seen that $\hat{T}_B$ (1) outperforms $\hat{T}_N$ in nearly all cases. Even in cases in which it does not, the ratio of the m.s.e. of the two estimators is no greater than 1.003.
2. Tables 3.17--3.20 show that T_B (1) outperforms $\hat{T}_B$ (0) in most cases. Even when $\rho_t = 0$, in which case $\hat{T}_B$ (0) is the optimum estimator, $\hat{T}_B$ (1) performs very well except when $M_i = 300 + 10i$ and $n = 15$ or 20 (Tables 3.20 (c) and (d)).
3. Tables 3.23 and 3.24 show that on the average $\hat{T}_B$ (1) is never worse than $\hat{T}_N$ and is usually better.
4. Tables 3.25 and 3.26 show that, on the average, $\hat{T}_B$ (1) is never worse and is usually better than $\hat{T}_B$ (0).

It should be pointed out that these results do not imply that $\hat{T}_B$ (1) should always be preferred to $\hat{T}_N$. Especially, if one wishes to protect against bias caused by higher or lower order terms being present in the model, then a strong case can be made for the estimator $\hat{T}_N$ by appeal to Theorem 2.2. Also, Theorems 2.3--2.6 give certain optimum properties of $\hat{T}_N$ which are not easily obtained for $\hat{T}_B$ (1). In addition, for nearly all cases there was very little difference in the two estimators.

Regarding the selection of one of $\hat{T}_B$ (0), $\hat{T}_B$ (1), or $\hat{T}_N$ to compare to $\hat{T}_B$ (ρ_a), it was decided to use $\hat{T}_B$ (1) since it outperforms on the average either $\hat{T}_B$ (0) or $\hat{T}_N$. Now, the results of Tables 3.1--3.8 showed that if no more than 50% of the clusters are sampled, then $\hat{T}_B$ (ρ_a) is very insensitive to choices of ρ_a. Hence, comparisons between $\hat{T}_B$ (ρ_a) and $\hat{T}_B$ (1) are made only for n = 15 and 20.

Tables 3.27--3.30 show the following:

1. Table 3.27 shows that if the X's range over a very small interval and if $\rho_a \neq 0$, then there is very little difference in $\hat{T}_B$ (ρ_a) and $\hat{T}_B$ (1). The exception to this is when all clusters are sampled (n = 20) and ρ_t = 1.0, in which case the m.s.e. of $\hat{T}_B$ (1) is zero.
2. From the same table, if ρ_a = 0, then $\hat{T}_B$ (1) beats $\hat{T}_B$ (ρ_a) for every value of the true correlation except zero.
3. From Table 3.28 one sees that if the X's range over a large interval, then $\hat{T}_B$ (1) tends to outperform $\hat{T}_B$ (ρ_a) whenever $\rho_t > \rho_a$; whereas $\hat{T}_B$ (ρ_a) is usually the better estimator whenever $\rho_t < \rho_a$. Of course, if $\rho_a = \rho_t$ then $\hat{T}_B$ (ρ_a) is the optimum estimator and hence can never be worse than $\hat{T}_B$ (1).
4. Tables 3.29 and 3.30 imply that for nearly all cases if $\rho_a \leq 0.2$, then on the average $\hat{T}_B$ (1) is at least as good and in many instances better than $\hat{T}_B$ (ρ_a). From the same tables one notes that if $\rho_a > 0.2$,

then on the average $\hat{T}_B$ (ρ_a) tends to be the better estimator in all cases. However, even for these cases there is not much difference in the two.

Regarding the examination of the percent decrease in m.s.e. of $\hat{T}_N$ for the case of balanced clusters and equal M_i; Table 3.31 shows the following:

1. For small M and small ρ $(\rho \leq 0.2)$ the m.s.e. is reduced very little by a 5% increase in total clusters sampled provided at least 50% of the total clusters are sampled.
2. In nearly all other cases the m.s.e. is reduced significantly by a 5% increase in total clusters sampled.
3. The percentage decrease in m.s.e. increases with ρ for any given 5% increase in total clusters sampled.
4. For small M and $\rho \leq 0.8$ the percent decrease in m.s.e. decreases as the percent of total clusters sampled increases.
5. For large M and $\rho \leq 0.2$ the same result holds as in (4).
6. As M increases the percent decreases in m.s.e. increases for a given value of ρ and a given increase in percent of total clusters sampled.

3.5 Conclusions Regarding Selection of Estimator

Many of the results in Section 3.4 depend on the absolute and relative values of ρ_a and ρ_t. Now, in a practical problem one may let ρ_a be the intra-class correlation coefficient $\hat{\rho}$ given by (2.4.2). However, one may have no knowledge at all about ρ_t other than the fact, perhaps, that $\rho_t \neq 1$. That is, other than possibly knowing that the secondary units in each cluster are not all equal; the sampler may not be willing to place ρ_t in any narrow interval. How, then, should one proceed to select an estimator under this lack of knowledge of ρ_t?

It should be mentioned that most of the subsequent conclusions are based on the particular computer study described in this chapter. As pointed out in Section 3.6 much additional work is needed before more general results can be obtained.

Also, the conclusions are concerned only with the selection of an estimator. Other necessary considerations such as optimum selection of the clusters, optimum number of sample clusters, optimum allocation of the second-stage units, and the selection of the second-stage units are important factors in the design of any two-stage process. Chapter II was concerned mostly with these parts of the design.

Before stating any conclusions, we first assume (i) the model (2.2.1) adequately describes the data (ii) $\rho_i = \rho_t$ and $\sigma_i^2 = \sigma^2$ for all $i = 1,\ldots, N$. (iii) $0 \leq \rho_t < 1$ and (iv) when applicable, the intra-class correlation coefficient $\hat{\rho}$ given by (2.4.2) is non-negative.

With these assumptions the following conclusions can be used as a guide in the selection of an estimator.

1. If no more than 50% of the total clusters in the population are sampled, use the estimator $\hat{T}_B\ (\rho_a)$ where $\rho_a = \hat{\rho}$.
2. If around 75% of the total clusters are sampled, and the X's range over a small interval, i.e., $\max(X)/\min(X) \sim 2.0$, then use $\hat{T}_B\ (\rho_a)$ where $\rho_a = \hat{\rho}$.
3. If about 75% of the total clusters are sampled, and the X's range over a wide interval, i.e., $\max(X)/\min(X) \sim 201.0$, then use $\hat{T}_B\ (1)$.

4. If all clusters in the population are sampled, use $\hat{T}_B$ (1).

5. If it is felt that assumption (i) may not be valid, i.e., the model should perhaps contain higher or lower order terms; then use $\hat{T}_N$ with balanced clusters. See Theorem 2.2.

3.6 Limitations of the Study

The computer study undertaken was limited in its scope, and certainly one should be aware of the difficulty in drawing any general conclusions from it. In order to obtain more conclusive results, one should look at cases in which (i) more clusters are present in the population, (ii) a higher percentage of second-stage units are sampled, (iii) the M_i and X_i vary over different ranges than those considered, (iv) some of the M_i and/or X_i are equal, (v) the M_i do not increase with the X_i, (vi) allocations, other than proportional, of the m_i are used, (vii) selection of the clusters is not restricted to the largest ones, (viii) elements in different clusters have unequal correlations and variances, and (ix) negative values of ρ_a, which may arise when letting $\rho_a = \hat{\rho}$ given by (2.4.2), are used in $\hat{T}_B$ (ρ_a).

TABLE 3.1

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 2.0$

(a) TYPE I X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.002	1.002	1.003
0.2	1.000	1.000	1.000	1.001	1.001	1.001
0.4	1.001	1.000	1.000	1.000	1.000	1.001
0.6	1.002	1.001	1.000	1.000	1.000	1.000
0.8	1.003	1.001	1.000	1.000	1.000	1.000
1.0	1.005	1.002	1.001	1.000	1.000	1.000

(b) TYPE I X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.001	1.001	1.002
0.2	1.000	1.000	1.000	1.000	1.001	1.001
0.4	1.000	1.000	1.000	1.000	1.000	1.001
0.6	1.001	1.000	1.000	1.000	1.000	1.001
0.8	1.003	1.002	1.001	1.000	1.000	1.000
1.0	1.007	1.004	1.003	1.001	1.000	1.000

TABLE 3.1. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 2.0$

(c) TYPE II X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.001	1.002
0.2	1.000	1.000	1.000	1.000	1.001	1.001
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.002	1.001	1.000	1.000	1.000	1.000
1.0	1.003	1.001	1.001	1.000	1.000	1.000

(d) TYPE II X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000
0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.001	1.000	1.000	1.000	1.000

TABLE 3.1. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 2.0$

(e) TYPE III X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.002	1.003
0.2	1.000	1.000	1.000	1.001	1.001	1.001
0.4	1.001	1.000	1.000	1.000	1.000	1.001
0.6	1.002	1.001	1.000	1.000	1.000	1.000
0.8	1.003	1.001	1.000	1.000	1.000	1.000
1.0	1.004	1.002	1.001	1.000	1.000	1.000

(f) TYPE III X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.001
0.2	1.000	1.000	1.000	1.000	1.000	1.001
0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.001	1.000	1.000	1.000	1.000
1.0	1.003	1.002	1.001	1.000	1.000	1.000

TABLE 3.1. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 2.0$

(g) TYPE IV X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.002	1.002
0.2	1.000	1.000	1.000	1.000	1.001	1.001
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.002	1.000	1.000	1.000	1.000	1.000
0.8	1.002	1.001	1.000	1.000	1.000	1.000
1.0	1.003	1.002	1.001	1.000	1.000	1.000

(h) TYPE IV X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000
0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.001	1.000	1.000	1.000	1.000
1.0	1.002	1.002	1.001	1.000	1.000	1.000

TABLE 3.2.

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 201.0$

(a) TYPE I X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.002	1.003	1.005	1.006
0.2	1.001	1.000	1.000	1.001	1.002	1.003
0.4	1.003	1.000	1.000	1.000	1.001	1.002
0.6	1.005	1.002	1.000	1.000	1.000	1.001
0.8	1.009	1.003	1.001	1.000	1.000	1.000
1.0	1.012	1.006	1.002	1.001	1.000	1.000

(b) TYPE I X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.003	1.006	1.011	1.017
0.2	1.001	1.000	1.001	1.003	1.007	1.013
0.4	1.004	1.001	1.000	1.001	1.004	1.009
0.6	1.012	1.005	1.001	1.000	1.001	1.005
0.8	1.032	1.018	1.008	1.002	1.000	1.002
1.0	1.106	1.067	1.038	1.007	1.004	1.000

TABLE 3.2. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e.of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 201.0$

(c) TYPE II X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.001	1.001	1.002
0.2	1.000	1.000	1.000	1.000	1.001	1.001
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.002	1.001	1.000	1.000	1.000	1.000
1.0	1.003	1.001	1.001	1.000	1.000	1.000

(d) TYPE II X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000
0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.002	1.001	1.001	1.000	1.000	1.000

TABLE 3.2. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 201.0$

(e) TYPE III X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.002	1.003	1.004
0.2	1.001	1.000	1.000	1.001	1.002	1.002
0.4	1.002	1.000	1.000	1.000	1.001	1.001
0.6	1.004	1.001	1.000	1.000	1.000	1.000
0.8	1.006	1.002	1.001	1.000	1.000	1.000
1.0	1.008	1.004	1.002	1.001	1.000	1.000

(f) TYPE III X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.002	1.003	1.005
0.2	1.000	1.000	1.000	1.001	1.002	1.004
0.4	1.001	1.000	1.000	1.000	1.001	1.002
0.6	1.004	1.002	1.000	1.000	1.000	1.001
0.8	1.009	1.005	1.002	1.001	1.000	1.001
1.0	1.027	1.016	1.009	1.004	1.001	1.000

TABLE 3.2. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = i$

$X \sim 1.0 - 201.0$

(g) TYPE IV X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.002	1.002
0.2	1.000	1.000	1.000	1.000	1.001	1.001
0.4	1.001	1.000	1.000	1.000	1.000	1.001
0.6	1.002	1.001	1.000	1.000	1.000	1.000
0.8	1.003	1.001	1.000	1.000	1.000	1.000
1.0	1.004	1.002	1.001	1.000	1.000	1.000

(h) TYPE IV X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.001	1.002	1.003
0.2	1.000	1.000	1.000	1.001	1.001	1.002
0.4	1.001	1.000	1.000	1.000	1.001	1.002
0.6	1.002	1.001	1.000	1.000	1.000	1.001
0.8	1.006	1.003	1.001	1.000	1.000	1.000
1.0	1.018	1.012	1.007	1.003	1.001	1.000

TABLE 3.3.

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

M_i = 300 + 10i

X ~ 1.0 - 2.0

(a) TYPE I X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.001	1.001	1.001	1.001
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(b) TYPE I X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.002	1.003	1.003	1.003	1.003
0.2	1.005	1.000	1.000	1.000	1.000	1.000
0.4	1.006	1.000	1.000	1.000	1.000	1.000
0.6	1.007	1.000	1.000	1.000	1.000	1.000
0.8	1.007	1.000	1.000	1.000	1.000	1.000
1.0	1.007	1.000	1.000	1.000	1.000	1.000

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(c) TYPE I X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.003	1.003	1.004	1.004	1.004
0.2	1.010	1.000	1.000	1.000	1.000	1.000
0.4	1.016	1.000	1.000	1.000	1.000	1.000
0.6	1.019	1.000	1.000	1.000	1.000	1.000
0.8	1.020	1.000	1.000	1.000	1.000	1.000
1.0	1.022	1.000	1.000	1.000	1.000	1.000

(d) TYPE I X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.002	1.002	1.002	1.002	1.002
0.2	1.012	1.000	1.000	1.000	1.000	1.000
0.4	1.035	1.000	1.000	1.000	1.000	1.000
0.6	1.082	1.001	1.000	1.000	1.000	1.000
0.8	1.223	1.004	1.001	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

M_i = 300 + 10i

X ~ 1.0 - 2.0

(e) TYPE II X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000
0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(f) TYPE II X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(g) TYPE II X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(h) TYPE II X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.001	1.001	1.001	1.001
0.2	1.004	1.000	1.000	1.000	1.000	1.000
0.4	1.013	1.000	1.000	1.000	1.000	1.000
0.6	1.030	1.001	1.000	1.000	1.000	1.000
0.8	1.082	1.003	1.000	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0.0

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(i) TYPE III X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.001	1.001	1.001	1.001
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(j) TYPE III X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.001	1.001	1.001	1.001
0.2	1.002	1.000	1.000	1.000	1.000	1.000
0.4	1.003	1.000	1.000	1.000	1.000	1.000
0.6	1.003	1.000	1.000	1.000	1.000	1.000
0.8	1.003	1.000	1.000	1.000	1.000	1.000
1.0	1.003	1.000	1.000	1.000	1.000	1.000

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(k) TYPE III X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.002	1.000	1.000	1.000	1.000	1.000
0.6	1.003	1.000	1.000	1.000	1.000	1.000
0.8	1.003	1.000	1.000	1.000	1.000	1.000
1.0	1.003	1.000	1.000	1.000	1.000	1.000

(l) TYPE III X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.001	1.001
0.2	1.003	1.000	1.000	1.000	1.000	1.000
0.4	1.009	1.000	1.000	1.000	1.000	1.000
0.6	1.020	1.000	1.000	1.000	1.000	1.000
0.8	1.054	1.001	1.000	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(m) TYPE IV X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.001	1.001
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(n) TYPE IV X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.001	1.001	1.002	1.002
0.2	1.003	1.000	1.000	1.000	1.000	1.000
0.4	1.003	1.000	1.000	1.000	1.000	1.000
0.6	1.004	1.000	1.000	1.000	1.000	1.000
0.8	1.004	1.000	1.000	1.000	1.000	1.000
1.0	1.004	1.000	1.000	1.000	1.000	1.000

TABLE 3.3. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(o) TYPE IV X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.002	1.002	1.002	1.002	1.002
0.2	1.007	1.000	1.000	1.000	1.000	1.000
0.4	1.010	1.000	1.000	1.000	1.000	1.000
0.6	1.012	1.000	1.000	1.000	1.000	1.000
0.8	1.013	1.000	1.000	1.000	1.000	1.000
1.0	1.014	1.000	1.000	1.000	1.000	1.000

(p) TYPE IV X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.002	1.003	1.003	1.003	1.004
0.2	1.014	1.000	1.000	1.000	1.001	1.001
0.4	1.042	1.000	1.000	1.000	1.000	1.000
0.6	1.098	1.002	1.000	1.000	1.000	1.000
0.8	1.266	1.007	1.001	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.4.

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(a) TYPE I X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.004	1.004	1.005	1.005	1.005
0.2	1.006	1.000	1.000	1.000	1.000	1.000
0.4	1.006	1.000	1.000	1.000	1.000	1.000
0.6	1.006	1.000	1.000	1.000	1.000	1.000
0.8	1.006	1.000	1.000	1.000	1.000	1.000
1.0	1.006	1.000	1.000	1.000	1.000	1.000

(b) TYPE I X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.028	1.032	1.033	1.034	1.034
0.2	1.054	1.000	1.000	1.000	1.000	1.001
0.4	1.065	1.000	1.000	1.000	1.000	1.000
0.6	1.069	1.000	1.000	1.000	1.000	1.000
0.8	1.072	1.001	1.000	1.000	1.000	1.000
1.0	1.073	1.001	1.000	1.000	1.000	1.000

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(c) TYPE I X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.003	1.099	1.106	1.109	1.111
0.2	1.025	1.000	1.002	1.004	1.005	1.006
0.4	1.371	1.003	1.000	1.000	1.001	1.001
0.6	1.438	1.006	1.000	1.000	1.000	1.000
0.8	1.479	1.008	1.001	1.000	1.000	1.000
1.0	1.508	1.009	1.002	1.000	1.000	1.000

(d) TYPE I X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.177	1.225	1.296	1.257	1.264
0.2	1.819	1.000	1.013	1.025	1.034	1.039
0.4	3.404	1.030	1.000	1.005	1.011	1.016
0.6	6.597	1.125	1.010	1.000	1.003	1.007
0.8	16.190	1.430	1.061	1.007	1.000	1.003
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(e) TYPE II X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.001	1.001	1.001	1.001
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(f) TYPE II X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.002	1.002	1.002	1.002	1.003
0.2	1.004	1.000	1.000	1.000	1.000	1.000
0.4	1.005	1.000	1.000	1.000	1.000	1.000
0.6	1.006	1.000	1.000	1.000	1.000	1.000
0.8	1.006	1.000	1.000	1.000	1.000	1.000
1.0	1.006	1.000	1.000	1.000	1.000	1.000

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(g) TYPE II X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.008	1.010	1.010	1.011	1.011
0.2	1.027	1.000	1.000	1.000	1.001	1.001
0.4	1.041	1.000	1.000	1.000	1.000	1.000
0.6	1.049	1.001	1.000	1.000	1.000	1.000
0.8	1.054	1.001	1.000	1.000	1.000	1.000
1.0	1.057	1.001	1.000	1.000	1.000	1.000

(h) TYPE II X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.050	1.065	1.071	1.075	1.079
0.2	1.248	1.000	1.005	1.009	1.012	1.014
0.4	1.744	1.011	1.000	1.002	1.004	1.006
0.6	2.799	1.044	1.004	1.000	1.001	1.003
0.8	5.768	1.153	1.023	1.003	1.000	1.001
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(i) TYPE III X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.002	1.002	1.002	1.002	1.002
0.2	1.003	1.000	1.000	1.000	1.000	1.000
0.4	1.003	1.000	1.000	1.000	1.000	1.000
0.6	1.003	1.000	1.000	1.000	1.000	1.000
0.8	1.004	1.000	1.000	1.000	1.000	1.000
1.0	1.004	1.000	1.000	1.000	1.000	1.000

(j) TYPE III X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.008	1.009	1.010	1.010	1.010
0.2	1.018	1.000	1.000	1.000	1.000	1.000
0.4	1.022	1.000	1.000	1.000	1.000	1.000
0.6	1.023	1.000	1.000	1.000	1.000	1.000
0.8	1.024	1.000	1.000	1.000	1.000	1.000
1.0	1.025	1.000	1.000	1.000	1.000	1.000

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(k) TYPE III X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.016	1.019	1.020	1.020	1.021
0.2	1.056	1.000	1.000	1.001	1.001	1.001
0.4	1.084	1.001	1.000	1.000	1.000	1.000
0.6	1.100	1.001	1.000	1.000	1.000	1.000
0.8	1.110	1.002	1.000	1.000	1.000	1.000
1.0	1.117	1.002	1.000	1.000	1.000	1.000

(l) TYPE III X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.050	1.064	1.070	1.073	1.075
0.2	1.262	1.000	1.004	1.008	1.011	1.013
0.4	1.781	1.010	1.000	1.002	1.004	1.006
0.6	2.830	1.042	1.003	1.000	1.001	1.003
0.8	5.982	1.146	1.021	1.003	1.000	1.001
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(m) TYPE IV X's, n = 5

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.001	1.001	1.001	1.001	1.001
0.2	1.002	1.000	1.000	1.000	1.000	1.000
0.4	1.002	1.000	1.000	1.000	1.000	1.000
0.6	1.002	1.000	1.000	1.000	1.000	1.000
0.8	1.002	1.000	1.000	1.000	1.000	1.000
1.0	1.002	1.000	1.000	1.000	1.000	1.000

(n) TYPE IV X's, n = 10

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.010	1.011	1.012	1.012	1.012
0.2	1.020	1.000	1.000	1.000	1.000	1.000
0.4	1.024	1.000	1.000	1.000	1.000	1.000
0.6	1.026	1.000	1.000	1.000	1.000	1.000
0.8	1.026	1.000	1.000	1.000	1.000	1.000
1.0	1.027	1.000	1.000	1.000	1.000	1.000

TABLE 3.4. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ When

M_i = 300 + 10i

X ~ 1.0 - 201.0

(o) TYPE IV X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.043	1.051	1.054	1.056	1.057
0.2	1.131	1.000	1.001	1.002	1.003	1.003
0.4	1.196	1.001	1.000	1.000	1.000	1.001
0.6	1.232	1.003	1.000	1.000	1.000	1.000
0.8	1.254	1.004	1.001	1.000	1.000	1.000
1.0	1.269	1.005	1.001	1.000	1.000	1.000

(p) TYPE IV X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.136	1.174	1.190	1.199	1.205
0.2	1.628	1.000	1.011	1.021	1.028	1.033
0.4	2.860	1.025	1.000	1.004	1.009	1.014
0.6	5.343	1.104	1.008	1.000	1.002	1.006
0.8	12.810	1.359	1.051	1.006	1.000	1.002
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.5.

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$,

With Respect to ρ_t, When

$M_i = i$

$X \sim 1.0 - 2.0$

(a) n = 5

ρ_a	X TYPE I	II	III	IV
0.0	1.0018	1.0012	1.0017	1.0013
0.2	1.0007	1.0003	1.007	1.0005
0.4	1.0003	1.0003	1.0003	1.0003
0.6	1.0005	1.0002	1.0003	1.0002
0.8	1.0005	1.0003	1.0005	1.0005
1.0	1.0008	1.0005	1.0008	1.0005

(b) n = 10

ρ_a	X TYPE I	II	III	IV
0.0	1.0018	1.0002	1.0007	1.0005
0.2	1.0010	1.0002	1.0005	1.0005
0.4	1.0007	1.0000	1.0002	1.0002
0.6	1.0003	1.0000	1.0000	1.0000
0.8	1.0003	1.0000	1.0000	1.0000
1.0	1.0008	1.0000	1.0003	1.0002

TABLE 3.6.

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$,

With Respect to ρ_t, When

$M_i = i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_a	X TYPE I	II	III	IV
0.0	1.0050	1.0012	1.0035	1.0017
0.2	1.0020	1.0003	1.0012	1.0007
0.4	1.0008	1.0002	1.0007	1.0003
0.6	1.0008	1.0002	1.0007	1.0002
0.8	1.0013	1.0003	1.0010	1.0005
1.0	1.0020	1.0005	1.0012	1.0007

(b) n = 10

ρ_a	X TYPE I	II	III	IV
0.0	1.0258	1.0005	1.0068	1.0045
0.2	1.0153	1.0002	1.0038	1.0027
0.4	1.0085	1.0002	1.0020	1.0013
0.6	1.0032	1.0000	1.0013	1.0008
0.8	1.0095	1.0000	1.0012	1.0008
1.0	1.0077	1.0000	1.0022	1.0013

TABLE 3.7.

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$,

With Respect to ρ_t, When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(a) n = 5

ρ_a	X TYPE I	II	III	IV
0.0	1.0008	1.0005	1.0008	1.0008
0.2	1.0002	1.0000	1.0002	1.0000
0.4	1.0002	1.0000	1.0002	1.0000
0.6	1.0002	1.0000	1.0002	1.0002
0.8	1.0002	1.0000	1.0002	1.0002
1.0	1.0002	1.0000	1.0002	1.0002

(b) n = 10

ρ_a	X TYPE I	II	III	IV
0.0	1.0053	1.0008	1.0023	1.0030
0.2	1.0003	1.0000	1.0002	1.0000
0.4	1.0005	1.0000	1.0002	1.0000
0.6	1.0005	1.0000	1.0002	1.0002
0.8	1.0005	1.0000	1.0002	1.0003
1.0	1.0005	1.0000	1.0002	1.0003

TABLE 3.7. (Continued)

Average Ratio m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$,

With Respect to ρ_t, When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(c) n = 15

ρ_a	X TYPE I	II	III	IV
0.0	1.0145	1.0007	1.0020	1.0093
0.2	1.0005	1.0000	1.0000	1.0003
0.4	1.0005	1.0000	1.0000	1.0003
0.6	1.0007	1.0000	1.0000	1.0003
0.8	1.0007	1.0000	1.0000	1.0003
1.0	1.0007	1.0000	1.0000	1.0003

(d) n = 20

ρ_a	X TYPE* I	II	III	IV
0.0	1.0704	1.0258	1.0172	1.0840
0.2	1.0014	1.0010	1.0000	1.0022
0.4	1.0010	1.0000	1.0000	1.0008
0.6	1.0000	1.0000	1.0000	1.0006
0.8	1.0000	1.0000	1.0000	1.0008
1.0	1.0000	1.0000	1.0000	1.0010

*Values when $\rho_t = 1$ are not included in average.

TABLE 3.8.

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$,

With Respect to ρ_t, When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_a	X TYPE I	II	III	IV
0.0	1.0050	1.0008	1.0028	1.0017
0.2	1.0007	1.0000	1.0003	1.0002
0.4	1.0007	1.0002	1.0003	1.0002
0.6	1.0008	1.0002	1.0003	1.0002
0.8	1.0008	1.0002	1.0003	1.0002
1.0	1.0008	1.0002	1.0003	1.0002

(b) n = 10

ρ_a	X TYPE I	II	III	IV
0.0	1.0555	1.0045	1.0187	1.0205
0.2	1.0050	1.0003	1.0013	1.0017
0.4	1.0053	1.0003	1.0015	1.0018
0.6	1.0055	1.0003	1.0017	1.0020
0.8	1.0057	1.0003	1.0017	1.0020
1.0	1.0058	1.0005	1.0017	1.0020

TABLE 3.8. (Continued)

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$,

With Respect to ρ_t, When

M_i = 300 + 10i

X ~ 1.0 - 201.0

(c) n = 15

ρ_a	X TYPE I	II	III	IV
0.0	1.3407	1.0380	1.0778	1.1803
0.2	1.0182	1.0018	1.0037	1.0093
0.4	1.0173	1.0017	1.0032	1.0090
0.6	1.0183	1.0017	1.0035	1.0093
0.8	1.0192	1.0020	1.0035	1.0098
1.0	1.0197	1.0020	1.0037	1.0102

(d) n = 20

ρ_a	X TYPE* I	II	III	IV
0.0	5.8020	2.5018	2.5710	4.7282
0.2	1.1524	1.0516	1.0496	1.1248
0.4	1.0618	1.0192	1.0184	1.0488
0.6	1.0566	1.0170	1.0166	1.0442
0.8	1.0610	1.0184	1.0178	1.0476
1.0	1.0658	1.0202	1.0196	1.0520

*Values when ρ_t = 1 are not included in average.

TABLE 3.9.

Ratio of m.s.e. of $\hat{T}_B$ (0) to m.s.e. of $\hat{T}_N$ When

$M_i = i$

$X \sim 1.0 - 2.0$

(a) n = 5

X TYPE

ρ_t	I	II	III	IV
0.00	1.000	0.997	0.999	0.998
0.10	0.999	0.995	0.999	0.998
0.25	0.999	0.994	0.998	0.997
0.50	0.998	0.991	0.997	0.995
0.75	0.998	0.989	0.996	0.993
0.90	0.998	0.987	0.995	0.992
1.00	0.997	0.986	0.995	0.991

(b) n = 10

X TYPE

ρ_t	I	II	III	IV
0.00	0.993	0.981	0.986	0.990
0.10	0.992	0.980	0.985	0.989
0.25	0.990	0.977	0.982	0.986
0.50	0.985	0.969	0.975	0.980
0.75	0.976	0.955	0.962	0.969
0.90	0.965	0.939	0.947	0.956
1.00	0.953	0.921	0.931	0.941

TABLE 3.10.

Ratio of m.s.e. of $\hat{T}_B$ (0) to m.s.e. of $\hat{T}_N$ When

$M_i = i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	0.994	0.999	0.998	1.000
0.10	0.996	0.998	0.999	1.000
0.25	0.998	0.997	1.000	1.000
0.50	1.003	0.995	1.002	1.000
0.75	1.007	0.994	1.005	1.000
0.90	1.010	0.993	1.006	1.000
1.00	1.012	0.992	1.007	1.000

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	0.987	0.994	0.998	1.000
0.10	0.989	0.993	0.998	1.000
0.25	0.993	0.992	0.999	1.000
0.50	0.993	0.992	0.999	1.000
0.75	1.024	0.982	0.999	1.002
0.90	1.055	0.973	1.000	1.004
1.00	1.104	0.960	1.002	1.007

TABLE 3.11.

Ratio of m.s.e. of $\hat{T}_B$ (0) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	0.999	1.000	0.999	1.000
0.10	1.000	0.999	1.000	1.000
0.25	1.001	0.999	1.000	1.000
0.50	1.001	0.999	1.000	1.000
0.75	1.001	0.999	1.001	1.000
0.90	1.001	0.999	1.001	1.000
1.00	1.001	0.999	1.001	1.000

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	0.998	1.000	1.000	1.000
0.10	1.004	0.999	1.000	1.002
0.25	1.006	0.998	1.001	1.003
0.50	1.007	0.998	1.002	1.003
0.75	1.007	0.998	1.002	1.003
0.90	1.007	0.998	1.002	1.003
1.00	1.007	0.998	1.002	1.003

TABLE 3.11. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (0) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(c) n = 15

ρ_t	X TYPE I	II	III	IV
0.00	0.997	1.000	1.000	0.999
0.10	1.005	0.999	1.000	1.004
0.25	1.012	0.999	1.000	1.007
0.50	1.017	0.998	1.000	1.011
0.75	1.019	0.998	1.000	1.012
0.90	1.020	0.998	1.000	1.013
1.00	1.021	0.998	1.000	1.013

(d) n = 20

ρ_t	X TYPE I	II	III	IV
0.00	0.998	0.999	0.993	0.996
0.10	1.004	1.001	1.000	1.004
0.25	1.016	1.006	1.004	1.019
0.50	1.054	1.020	1.013	1.064
0.75	1.167	1.061	1.041	1.199
0.90	1.506	1.186	1.123	1.604
1.00	∞	∞	∞	∞

TABLE 3.12.

Ratio of m.s.e. of $\hat{T}_B$ (0) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	0.986	1.000	0.992	0.997
0.10	1.000	1.001	1.000	1.001
0.25	1.002	1.001	1.001	1.001
0.50	1.002	1.001	1.001	1.002
0.75	1.003	1.001	1.001	1.002
0.90	1.003	1.001	1.001	1.002
1.00	1.003	1.001	1.001	1.002

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	0.952	0.998	0.985	0.984
0.10	1.030	1.003	1.010	1.012
0.25	1.053	1.005	1.017	1.020
0.50	1.063	1.005	1.021	1.024
0.75	1.067	1.006	1.023	1.025
0.90	1.069	1.006	1.023	1.026
1.00	1.069	1.006	1.023	1.026

TABLE 3.12. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (0) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(c) n = 15

ρ_t	X TYPE I	II	III	IV
0.00	0.891	0.989	0.968	0.942
0.10	1.112	1.013	1.026	1.059
0.25	1.276	1.031	1.063	1.147
0.50	1.404	1.046	1.093	1.215
0.75	1.468	1.053	1.108	1.298
0.90	1.492	1.056	1.114	1.262
1.00	1.506	1.057	1.117	1.269

(d) n = 20

ρ_t	X TYPE I	II	III	IV
0.00	0.791	0.929	0.930	0.830
0.10	1.217	1.063	1.070	1.162
0.25	2.070	1.332	1.351	1.825
0.50	4.629	2.137	2.192	3.816
0.75	12.305	4.554	4.715	9.789
0.90	35.332	11.804	12.285	27.708
1.00	∞	∞	∞	∞

TABLE 3.13.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_N$ When

$M_i = i$

$X \sim 1.0 - 2.0$

(a) $n = 5$

ρ_t	X TYPE I	II	III	IV
0.00	1.002	0.998	1.002	1.000
0.10	1.001	0.997	1.000	0.999
0.25	1.000	0.994	0.999	0.997
0.50	0.997	0.990	0.996	0.994
0.75	0.995	0.987	0.993	0.991
0.90	0.994	0.985	0.991	0.989
1.00	0.993	0.984	0.990	0.988

(b) $n = 10$

ρ_t	X TYPE I	II	III	IV
0.00	0.995	0.982	0.987	0.990
0.10	0.994	0.980	0.985	0.989
0.25	0.991	0.977	0.982	0.987
0.50	0.985	0.969	0.975	0.980
0.75	0.974	0.955	0.961	0.969
0.90	0.961	0.939	0.945	0.955
1.00	0.946	0.920	0.928	0.939

TABLE 3.14.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_N$ When

$M_i = i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	1.000	1.000	1.002	1.002
0.10	1.000	0.999	1.002	1.002
0.25	1.000	0.997	1.001	1.000
0.50	1.000	0.995	1.000	0.999
0.75	1.000	0.992	0.999	0.998
0.90	1.000	0.991	0.999	0.997
1.00	1.000	0.990	0.999	0.996

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	1.003	0.994	1.003	1.003
0.10	1.003	0.994	1.002	1.003
0.25	1.003	0.992	1.001	1.002
0.50	1.003	0.989	0.998	1.001
0.75	1.002	0.981	0.993	0.998
0.90	1.000	0.972	0.986	0.995
1.00	0.998	0.959	0.976	0.989

TABLE 3.15.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(a) $n = 5$

ρ_t	X TYPE I	II	III	IV
0.00	1.000	1.000	1.000	1.000
0.10	1.000	0.999	1.000	1.000
0.25	1.000	0.999	1.000	1.000
0.50	1.000	0.999	1.000	1.000
0.75	1.000	0.999	1.000	1.000
0.90	1.000	0.999	1.000	1.000
1.00	1.000	0.999	1.000	1.000

(b) $n = 10$

ρ_t	X TYPE I	II	III	IV
0.00	1.000	1.000	1.000	1.001
0.10	1.000	0.998	0.999	1.000
0.25	1.000	0.998	0.999	1.000
0.50	1.000	0.997	0.999	0.999
0.75	1.000	0.997	0.998	0.999
0.90	1.000	0.997	0.998	0.999
1.00	1.000	0.997	0.998	0.999

TABLE 3.15. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(c) n = 15

ρ_t	X TYPE I	II	III	IV
0.00	1.000	1.000	1.000	1.001
0.10	1.000	0.999	0.999	1.000
0.25	1.000	0.998	0.998	1.000
0.50	1.000	0.998	0.997	1.000
0.75	0.999	0.997	0.997	0.999
0.90	0.999	0.997	0.997	0.999
1.00	0.999	0.997	0.997	0.999

(d) n = 20

ρ_t	X TYPE I	II	III	IV
0.00	1.000	1.000	1.000	1.000
0.10	1.000	1.000	1.000	1.000
0.25	1.000	1.000	1.000	1.000
0.50	1.000	1.000	1.000	1.000
0.75	1.000	1.000	1.000	1.000
0.90	1.000	1.000	1.000	1.000
1.00	0/0	0/0	0/0	0/0

TABLE 3.16.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	0.990	1.000	0.994	0.999
0.10	0.996	1.000	0.997	1.000
0.25	0.996	1.000	0.997	1.000
0.50	0.996	1.000	0.997	1.000
0.75	0.996	1.000	0.997	1.000
0.90	0.996	1.000	0.997	1.000
1.00	0.996	1.000	0.997	1.000

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	0.985	1.000	0.995	0.996
0.10	0.993	1.000	0.997	0.998
0.25	0.995	1.000	0.998	0.999
0.50	0.996	1.000	0.998	0.999
0.75	0.996	1.000	0.999	0.999
0.90	0.997	1.000	0.999	0.999
1.00	0.997	1.000	0.999	0.999

TABLE 3.16. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_N$ When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(c) n = 15

ρ_t	X TYPE I	II	III	IV
0.00	0.990	1.000	0.998	0.995
0.10	0.993	1.000	0.999	0.997
0.25	0.995	1.000	0.999	0.998
0.50	0.997	1.000	0.999	0.999
0.75	0.998	1.000	1.000	0.999
0.90	0.999	1.000	1.000	0.999
1.00	0.999	1.000	1.000	1.000

(d) n = 20

ρ_t	X TYPE I	II	III	IV
0.00	1.000	1.000	1.000	1.000
0.10	1.000	1.000	1.000	1.000
0.25	1.000	1.000	1.000	1.000
0.50	1.000	1.000	1.000	1.000
0.75	1.000	1.000	1.000	1.000
0.90	1.000	1.000	1.000	1.000
1.00	0/0	0/0	0/0	0/0

TABLE 3.17.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_B$ (0) When

$M_i = i$

$X \sim 1.0 - 2.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	1.003	1.002	1.003	1.002
0.10	1.002	1.001	1.002	1.001
0.25	1.000	1.000	1.000	1.000
0.50	0.999	0.999	0.999	0.999
0.75	0.997	0.998	0.997	0.998
0.90	0.996	0.998	0.996	0.997
1.00	0.995	0.997	0.996	0.997

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	1.002	1.000	1.000	1.000
0.10	1.002	1.000	1.000	1.000
0.25	1.001	1.000	1.000	1.000
0.50	1.000	1.000	1.000	1.000
0.75	0.998	1.000	0.999	0.999
0.90	0.996	0.999	0.998	0.999
1.00	0.993	0.999	0.997	0.998

TABLE 3.18.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_B$ (0) When

$M_i = i$

$X \sim 1.0 - 201.0$

(a) $n = 5$

ρ_t	X TYPE I	II	III	IV
0.00	1.006	1.002	1.004	1.002
0.10	1.004	1.001	1.003	1.002
0.25	1.002	1.000	1.001	1.000
0.50	0.997	1.000	0.998	0.999
0.75	0.992	0.998	0.995	0.997
0.90	0.990	0.998	0.993	0.996
1.00	0.988	0.997	0.992	0.996

(b) $n = 10$

ρ_t	X TYPE I	II	III	IV
0.00	1.017	1.000	1.005	1.003
0.10	1.014	1.000	1.004	1.003
0.25	1.010	1.000	1.003	1.002
0.50	1.000	1.000	1.000	1.000
0.75	0.978	1.000	0.993	0.996
0.90	0.948	1.000	0.985	0.991
1.00	0.904	0.998	0.974	0.982

TABLE 3.19.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_B$ (0) When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	1.001	1.000	1.000	1.000
0.10	0.999	1.000	0.999	0.999
0.25	0.999	1.000	0.999	0.999
0.50	0.999	0.999	0.999	0.999
0.75	0.999	0.999	0.999	0.999
0.90	0.999	0.999	0.999	0.999
1.00	0.999	0.999	0.999	0.999

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	1.003	1.000	1.001	1.002
0.10	0.996	0.999	0.998	0.998
0.25	0.994	0.999	0.997	0.997
0.50	0.993	0.999	0.997	0.997
0.75	0.993	0.999	0.997	0.996
0.90	0.993	0.999	0.997	0.996
1.00	0.993	0.999	0.997	0.996

TABLE 3.19. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_B$ (0) When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(c) n = 15

ρ_t	X TYPE I	II	III	IV
0.00	1.004	1.000	1.000	1.003
0.10	0.995	1.000	0.999	0.997
0.25	0.988	0.999	0.998	0.993
0.50	0.983	0.999	0.998	0.989
0.75	0.980	0.999	0.997	0.987
0.90	0.979	0.999	0.997	0.987
1.00	0.979	0.999	0.977	0.986

(d) n = 20

ρ_t	X TYPE I	II	III	IV
0.00	1.003	1.001	1.000	1.004
0.10	0.996	0.999	0.999	0.996
0.25	0.984	0.994	0.996	0.981
0.50	0.949	0.981	0.987	0.940
0.75	0.857	0.942	0.961	0.834
0.90	0.664	0.843	0.890	0.623
1.00	0.000	0.000	0.000	0.000

TABLE 3.20.

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_B$ (0) When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(a) n = 5

ρ_t	X TYPE I	II	III	IV
0.00	1.005	1.001	1.003	1.001
0.10	0.995	0.999	0.997	0.999
0.25	0.994	0.999	0.997	0.998
0.50	0.994	0.999	0.997	0.998
0.75	0.994	0.999	0.997	0.998
0.90	0.994	0.999	0.996	0.998
1.00	0.994	0.999	0.996	0.998

(b) n = 10

ρ_t	X TYPE I	II	III	IV
0.00	1.034	1.003	1.010	1.012
0.10	0.964	0.997	0.988	0.986
0.25	0.945	0.995	0.981	0.979
0.50	0.937	0.995	0.978	9.976
0.75	0.934	0.994	0.977	0.974
0.90	0.933	0.994	0.976	0.974
1.00	0.932	0.994	0.976	0.974

TABLE 3.20. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (1) to m.s.e. of $\hat{T}_B$ (0) When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(c) n = 15

ρ_t	X TYPE I	II	III	IV
0.00	1.111	1.011	1.021	1.057
0.10	0.893	0.987	0.974	0.941
0.25	0.780	0.970	0.940	0.870
0.50	0.710	0.956	0.915	0.822
0.75	0.680	0.950	0.902	0.800
0.90	0.669	0.947	0.898	0.792
1.00	0.663	0.946	0.895	0.788

(d) n = 20

ρ_t	X TYPE I	II	III	IV
0.00	1.264	1.077	1.075	1.205
0.10	0.821	0.941	0.934	0.881
0.25	0.482	0.751	0.790	0.548
0.50	0.216	0.468	0.456	0.262
0.75	0.081	0.220	0.212	0.102
0.90	0.028	0.085	0.081	0.036
1.00	0.000	0.000	0.000	0.000

TABLE 3.21.

Average of m.s.e. Ratios of $\hat{T}_B$ (0) to $\hat{T}_N$,

With Respect to ρ_t, When

$M_i = i$

(a) X ~ 1.0 - 2.0

n	X TYPE I	II	III	IV
5	0.998	0.991	0.997	0.995
10	0.979	0.960	0.967	0.973

(b) X ~ 1.0 - 201.0

n	X TYPE I	II	III	IV
5	1.003	0.995	1.002	1.000
10	1.022	0.983	0.999	1.002

TABLE 3.22.

Average of m.s.e. Ratios of $\hat{T}_B$ (0) to $\hat{T}_N$,

With Respect to ρ_t, When

$M_i = 300 + 10i$

(a) X ~ 1.0 - 2.0

n	X TYPE I	II	III	IV
5	1.001	0.999	1.000	1.000
10	1.005	0.998	1.001	1.002
15	1.013	0.999	1.000	1.008
20*	1.124	1.046	1.029	1.148

(b) X ~ 1.0 - 201.0

n	X TYPE I	II	III	IV
5	1.000	1.001	1.000	1.001
10	1.043	1.004	1.015	1.017
15	1.307	1.035	1.071	1.183
20*	9.391	3.637	3.757	7.522

*m.s.e. ratios for $\rho_t = 1$ are not included in average.

TABLE 3.23.

Average of m.s.e. Ratios of $\hat{T}_B$ (1) to $\hat{T}_N$,

With Respect to ρ_t, When

$M_i = i$

(a) X ~ 1.0 - 2.0

n	X TYPE I	II	III	IV
5	0.997	0.991	0.996	0.994
10	0.978	0.960	0.966	0.973

(b) X ~ 1.0 - 201.0

n	X TYPE I	II	III	IV
5	1.000	0.995	1.000	0.999
10	1.002	0.983	0.994	0.999

TABLE 3.24.

Average of m.s.e. Ratios of $\hat{T}_B$ (1) to $\hat{T}_N$,

With Respect to ρ_t, When

$M_i = 300 + 10i$

(a) X ~ 1.0 - 2.0

n	X TYPE I	II	III	IV
5	1.000	0.999	1.000	1.000
10	1.000	0.998	0.999	1.000
15	1.000	0.998	0.998	1.000
20*	1.000	1.000	1.000	1.000

(b) X ~ 1.0 - 201.0

n	X TYPE I	II	III	IV
5	0.995	1.000	0.997	1.000
10	0.994	1.000	0.998	0.999
15	0.996	1.000	0.999	0.998
20*	1.000	1.000	1.000	1.000

*m.s.e. ratios for $\rho_t = 1$ are not included in average.

TABLE 3.25.

Average of m.s.e. Ratios of $\hat{T}_B$ (1) to $\hat{T}_B$ (0),

With Respect to ρ_t, When

$M_i = i$

(a) X ~ 1.0 - 2.0

n	X TYPE I	II	III	IV
5	0.999	0.999	0.999	0.999
10	0.999	1.000	0.999	1.000

(b) X ~ 1.0 - 201.0

n	X TYPE I	II	III	IV
5	0.997	0.999	0.998	0.999
10	0.982	1.000	0.995	0.997

TABLE 3.26.

Average of m.s.e. Ratios of $\hat{T}_B$ (1) to $\hat{T}_B$ (0),

With Respect to ρ_t, When

$M_i = 300 + 10i$

(a) X ~ 1.0 - 2.0

n	X TYPE I	II	III	IV
5	0.999	0.999	0.999	0.999
10	0.995	0.999	0.998	0.997
15	0.987	0.999	0.998	0.992
20	0.779	0.823	0.833	0.768

(b) X ~ 1.0 - 201.0

n	X TYPE I	II	III	IV
5	0.996	0.999	0.998	0.999
10	0.954	0.996	0.984	0.982
15	0.787	0.967	0.935	0.867
20	0.413	0.506	0.500	0.431

TABLE 3.27.

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(a) TYPE I X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.996	0.991	1.000	1.000	1.000	1.000
0.2	1.010	1.000	1.000	1.000	1.000	1.000
0.4	1.016	1.000	1.000	1.000	1.000	1.000
0.6	1.019	1.000	1.000	1.000	1.000	1.000
0.8	1.020	1.000	1.000	1.000	1.000	1.000
1.0	1.022	1.000	1.000	1.000	1.000	1.000

(b) TYPE I X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.998	0.993	1.000	1.000	1.000	1.000
0.2	1.012	1.000	1.000	1.000	1.000	1.000
0.4	1.035	1.000	1.000	1.000	1.000	1.000
0.6	1.082	1.001	1.000	1.000	1.000	1.000
0.8	1.223	1.004	1.001	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.27. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

M_i = 300 + 10i

X ~ 1.0 - 2.0

(c) TYPE II X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.000	1.000	1.000	1.000	1.000	1.000
0.4	1.001	1.000	1.000	1.000	1.000	1.000
0.6	1.001	1.000	1.000	1.000	1.000	1.000
0.8	1.001	1.000	1.000	1.000	1.000	1.000
1.0	1.001	1.000	1.000	1.000	1.000	1.000

(d) TYPE II X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.999	1.000	1.000	1.000	1.000	1.000
0.2	1.004	1.000	1.000	1.000	1.000	1.000
0.4	1.013	1.000	1.000	1.000	1.000	1.000
0.6	1.030	1.001	1.000	1.000	1.000	1.000
0.8	1.082	1.003	1.000	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.27. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(e) TYPE III X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000
0.2	1.001	1.000	1.000	1.000	1.000	1.000
0.4	1.002	1.000	1.000	1.000	1.000	1.000
0.6	1.003	1.000	1.000	1.000	1.000	1.000
0.8	1.003	1.000	1.000	1.000	1.000	1.000
1.0	1.003	1.000	1.000	1.000	1.000	1.000

(f) TYPE III X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.999	1.000	1.000	1.000	1.000	1.000
0.2	1.003	1.000	1.000	1.000	1.000	1.000
0.4	1.009	1.000	1.000	1.000	1.000	1.000
0.6	1.020	1.000	1.000	1.000	1.000	1.000
0.8	1.055	1.001	1.000	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.27. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(g) TYPE IV X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.998	0.999	1.000	1.000	1.000	1.000
0.2	1.006	1.000	1.000	1.000	1.000	1.000
0.4	1.010	1.000	1.000	1.000	1.000	1.000
0.6	1.012	1.000	1.000	1.000	1.000	1.000
0.8	1.013	1.000	1.000	1.000	1.000	1.000
1.0	1.014	1.000	1.000	1.000	1.000	1.000

(h) TYPE IV X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.997	0.999	1.000	1.000	1.000	1.000
0.2	1.013	0.999	1.000	1.000	1.000	1.000
0.4	1.041	1.000	1.000	1.000	1.000	1.000
0.6	1.098	1.002	1.000	1.000	1.000	1.000
0.8	1.267	1.007	1.001	1.000	1.000	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.28.

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(a) TYPE I X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.900	0.975	0.989	0.995	0.998	1.000
0.2	1.240	0.994	0.996	0.998	0.999	1.000
0.4	1.369	1.001	0.999	0.999	1.000	1.000
0.6	1.437	1.005	1.000	1.000	1.000	1.000
0.8	1.479	1.008	1.001	1.000	1.000	1.000
1.0	1.508	1.009	1.002	1.000	1.000	1.000

(b) TYPE I X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.791	0.931	0.969	0.985	0.994	1.000
0.2	1.750	0.962	0.975	0.987	0.995	1.000
0.4	3.350	1.014	0.984	0.989	0.995	1.000
0.6	6.548	1.117	1.003	0.993	0.995	1.000
0.8	16.140	1.426	1.058	1.005	0.997	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.28. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(c) TYPE II X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.989	0.997	0.989	0.995	0.998	1.000
0.2	1.026	0.999	1.000	1.000	1.000	1.000
0.4	1.041	1.000	1.000	1.000	1.000	1.000
0.6	1.049	1.001	1.000	1.000	1.000	1.000
0.8	1.054	1.001	1.000	1.000	1.000	1.000
1.0	1.057	1.001	1.000	1.000	1.000	1.000

(d) TYPE II X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.929	0.975	0.989	0.995	0.998	1.000
0.2	1.231	0.986	0.991	0.995	0.998	1.000
0.4	1.734	1.005	0.994	0.996	0.998	1.000
0.6	2.741	1.042	1.001	0.997	0.998	1.000
0.8	5.762	1.152	1.021	1.002	0.999	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.28. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(e) TYPE III X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.980	0.995	0.998	0.999	1.000	1.000
0.2	1.054	0.999	0.999	1.000	1.000	1.000
0.4	1.084	1.000	1.000	1.000	1.000	1.000
0.6	1.100	1.001	1.000	1.000	1.000	1.000
0.8	1.110	1.002	1.000	1.000	1.000	1.000
1.0	1.117	1.002	1.000	1.000	1.000	1.000

(f) TYPE III X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.930	0.977	0.990	0.995	0.998	1.000
0.2	1.245	0.987	0.991	0.995	0.998	1.000
0.4	1.771	1.005	0.995	0.996	0.998	1.000
0.6	2.823	1.040	1.001	0.998	0.999	1.000
0.8	5.977	1.145	1.000	1.002	0.999	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.28. (Continued)

Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1) When

$M_i = 300 + 10i$

$X \sim 1.0 - 201.0$

(g) TYPE IV X's, n = 15

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.946	0.986	0.994	0.997	0.999	1.000
0.2	1.127	0.997	0.998	0.999	1.000	1.000
0.4	1.196	1.001	0.999	1.000	1.000	1.000
0.6	1.232	1.003	1.000	1.000	1.000	1.000
0.8	1.254	1.004	1.001	1.000	1.000	1.000
1.0	1.269	1.005	1.001	1.000	1.000	1.000

(h) TYPE IV X's, n = 20

ρ_t \ ρ_a	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.830	0.942	0.974	0.988	0.995	1.000
0.2	1.576	0.968	0.979	0.989	0.995	1.000
0.4	2.821	1.011	0.986	0.990	0.996	1.000
0.6	5.309	1.097	1.002	0.994	0.996	1.000
0.8	12.780	1.356	1.049	1.004	0.998	1.000
1.0	∞	∞	∞	∞	∞	0/0

TABLE 3.29.

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1),

With Respect to ρ_t, When

$M_i = 300 + 10i$

$X \sim 1.0 - 2.0$

(a) n = 15

ρ_a	X TYPE I	II	III	IV
0.0	1.014	1.001	1.002	1.009
0.2	0.999	1.000	1.000	1.000
0.4	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000
1.0	1.000	1.000	1.000	1.000

(b) n = 20*

ρ_a	X TYPE I	II	III	IV
0.0	1.070	1.026	1.017	1.083
0.2	1.000	1.001	1.000	1.001
0.4	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000
1.0	1.000	1.000	1.000	1.000

*Averages do not include m.s.e. ratios when $\rho_t = 1.0$.

TABLE 3.30.

Average Ratio of m.s.e. of $\hat{T}_B$ (ρ_a) to m.s.e. of $\hat{T}_B$ (1),

With Respect to ρ_t, When

M_i = 300 + 10i

X $\sim$ 1.0 - 201.0

(a) n = 15

ρ_a	X TYPE I	II	III	IV
0.0	1.322	1.036	1.074	1.171
0.2	0.999	1.000	1.000	0.999
0.4	0.998	1.000	1.000	0.999
0.6	0.999	1.000	1.000	0.999
0.8	1.000	1.000	1.000	1.000
1.0	1.000	1.000	1.000	1.000

(b) n = 20*

ρ_a	X TYPE I	II	III	IV
0.0	4.763	2.479	2.549	4.663
0.2	1.090	1.032	1.031	1.075
0.4	0.998	0.999	0.999	0.998
0.6	0.992	0.997	0.997	0.994
0.8	0.995	0.998	0.998	0.996
1.0	1.000	1.000	1.000	1.000

*Averages do not include m.s.e. ratios when ρ_t = 1.0.

TABLE 3.31.

Percentage Decrease in m.s.e. of $\hat{T}_N$ as Percent of Clusters Sampled Increases and Clusters are Balanced

(a) $M_i = M = 10$

% of Clusters Sampled Goes From	ρ 0.0	0.2	0.4	0.6	0.8	1.0
5-10	0	38	46	49	51	53
10-15	0	20	28	32	35	37
15-20	0	13	20	24	27	29
20-25	0	9	15	19	22	25
25-30	0	6	11	16	19	22
30-35	0	5	9	13	17	20
35-40	0	4	8	11	15	19
40-45	0	3	6	10	14	19
45-50	0	3	5	9	13	18
50-55	0	2	5	8	12	18
55-60	0	2	4	7	12	19
60-65	0	2	4	7	11	19
65-70	0	1	3	6	11	20
70-75	0	1	3	6	11	22
75-80	0	1	3	5	10	25
80-85	0	1	2	5	10	29
85-90	0	1	2	5	10	37
90-95	0	1	2	4	10	53
95-100	0	1	2	4	10	100

TABLE 3.31. (Continued)

Percentage Decrease in m.s.e. of $\hat{T}_N$ as Percent of Clusters Sampled Increases and Clusters are Balanced

(b) $M_i = M = 100$

% of Clusters Sampled Goes From	ρ 0.0	0.2	0.4	0.6	0.8	1.0
5–10	0	51	52	52	53	53
10–15	0	34	36	37	37	37
15–20	0	26	28	29	29	29
20–25	0	21	23	24	25	25
25–30	0	18	30	21	22	22
30–35	0	15	18	19	20	20
35–40	0	14	17	18	19	19
40–45	0	12	16	17	18	19
45–50	0	11	15	16	18	18
50–55	0	10	14	16	17	18
55–60	0	10	14	16	18	19
60–65	0	9	13	16	18	19
65–70	0	8	13	17	19	20
70–75	0	8	13	17	20	22
75–80	0	8	13	18	22	25
80–85	0	7	14	20	25	29
85–90	0	7	14	22	29	37
90–95	0	7	15	25	37	53
95–100	0	6	16	29	53	100

TABLE 3.32

M and X Values

(a) X ~ 1.0 - 2.0

Cluster	M	X TYPE I	X TYPE II	X TYPE III	X TYPE IV
1	1 or 310	1.0278	1.0833	1.0714	1.0000
2	2 or 320	1.0556	1.1667	1.1904	1.0526
3	3 or 330	1.0833	1.3000	1.2381	1.1052
4	4 or 340	1.1111	1.3500	1.3143	1.1578
5	5 or 350	1.1389	1.4000	1.3428	1.2104
6	6 or 360	1.1667	1.4500	1.3714	1.2630
7	7 or 370	1.1944	1.5357	1.3999	1.3156
8	8 or 380	1.2222	1.5714	1.4489	1.3682
9	9 or 390	1.2857	1.6071	1.4693	1.4208
10	10 or 400	1.3214	1.6429	1.4897	1.4734
11	11 or 410	1.3571	1.6786	1.5102	1.5260
12	12 or 420	1.3929	1.7143	1.5306	1.5786
13	13 or 430	1.4286	1.7778	1.5510	1.6312
14	14 or 440	1.4643	1.8056	1.6000	1.6838
15	15 or 450	1.5500	1.8333	1.6285	1.7364
16	16 or 460	1.6000	1.8611	1.6571	1.7890
17	17 or 470	1.6500	1.8889	1.6856	1.8416
18	18 or 480	1.7000	1.9167	1.7618	1.8942
19	19 or 490	1.8333	1.9444	1.8095	1.9468
20	20 or 500	1.9167	1.9722	1.9285	1.9994

TABLE 3.32. (Continued)

M and X Values

(b) X ~ 1.0 - 201.0

Cluster	M	X TYPE I	II	III	IV
1	1 or 310	6.5556	17.667	15.285	1.0000
2	2 or 320	12.111	34.333	39.095	11.526
3	3 or 330	17.667	61.000	48.619	22.053
4	4 or 340	23.222	71.000	63.857	32.579
5	5 or 350	28.778	81.000	69.571	43.105
6	6 or 360	34.333	91.000	75.286	53.632
7	7 or 370	39.889	108.14	81.000	64.158
8	8 or 380	45.444	115.29	90.796	74.684
9	9 or 390	58.143	122.43	94.877	85.210
10	10 or 400	65.286	129.57	98.959	95.737
11	11 or 410	72.429	136.71	103.04	106.26
12	12 or 420	79.571	143.86	107.12	116.79
13	13 or 430	86.714	156.56	111.20	127.32
14	14 or 440	93.857	162.11	121.00	137.84
15	15 or 450	111.00	167.67	126.71	148.37
16	16 or 460	121.00	173.22	132.43	158.89
17	17 or 470	131.00	178.78	138.14	169.42
18	18 or 480	141.00	184.33	153.38	179.95
19	19 or 490	167.67	189.89	162.90	190.47
20	20 or 500	184.33	195.44	186.71	201.00

IV. A CLUSTER MODEL WHEN AUXILIARY INFORMATION IS AVAILABLE ON EACH SECONDARY UNIT

4.1 The Model and a Conventional Estimator

Often there will be available information on each of the secondary units, not just on previous cluster totals as considered up to now. If X_{ij} is the most recent value available for the j^{th} unit in the i^{th} cluster, then we shall consider the model

$$Y_{ij} = \beta X_{ij} + \varepsilon_{ij} \qquad (4.1.1)$$

where $E(Y_{ij}) = \beta X_{ij}$, and the covariance structure of the Y's is the same as given in Chapter II. See Section 2.2.

Theorems 1.1 and 1.2 could be used to obtain the b.l.u.e. of the total T under model (4.1.1). However, due to the presence of the X_{ij} terms, the derivation is very tedious and the b.l.u.e. is much more complex and difficult to analyze than the b.l.u.e. for the earlier model (2.2.1).

A conventional estimator recommended with simple random sampling at both stages is given by Cochran [2] to be

$$\hat{T}_c = (\sum_{i\varepsilon S} M_i \bar{y}_i / \sum_{i\varepsilon S} M_i (\bar{X}_i)_S) \sum_1^N X_i$$

where $(\bar{X}_i)_S$ is the mean of the m_i auxiliary variables associated with the m_i second stage units selected from the i^{th} sample cluster, and $X_i = \sum_{j=1}^{M_i} X_{ij}$. The estimator $\hat{T}_c$ can be rewritten as

$$\hat{T}_c = \sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij} + \sum_{i\varepsilon S} (M_i - m_i)\bar{y}_i + \frac{\sum_{i\varepsilon S} M_i \bar{y}_i (\sum_1^N M_i \bar{X}_i - \sum_{i\varepsilon S} M_i (\bar{X}_i)_S)}{\sum_{i\varepsilon S} M_i (\bar{X}_i)_S} \qquad (4.1.2)$$

where $\bar{X}_i = \sum_{j=1}^{M_i} X_{ij}/M_i$. That is, $\hat{T}_c$ is written as the sum of (i) the elements in the sample, (ii) an estimate of the sum in the sample clusters but not in the sample itself, and (iii) an estimate of the sum in the unsampled clusters.

4.1.1 Bias of Conventional Estimator

Using model (4.1.1), we have

$$E(\hat{T}_c - T) = \beta \{\sum_{i\varepsilon S} (M_i - m_i)(\bar{X}_i)_S + \frac{\sum_{i\varepsilon S} M_i (\bar{X}_i)_S (\sum_1^N M_i \bar{X}_i - \sum_{i\varepsilon S} M_i (\bar{X}_i)_S)}{\sum_{i\varepsilon S} M_i (\bar{X}_i)_S}$$

$$- \sum_{i\varepsilon S} (M_i - m_i)(\bar{X}_i)'_S - \sum_{i\varepsilon \bar{S}} M_i \bar{X}_i\} \qquad (4.1.3)$$

where

$$(\bar{X}_i)'_S = \sum_{j \notin S_i} X_{ij}/(M_i - m_i) \ .$$

Or,

$$E(\hat{T}_c - T) = \beta \{\sum_{i\varepsilon S} (M_i - m_i)[(\bar{X}_i)_S - (\bar{X}_i)'_S] + \sum_{i\varepsilon S} M_i [\bar{X}_i - (\bar{X}_i)_S]\} \ .$$

But

$$\bar{X}_i = \frac{1}{M_i} [m_i (\bar{X}_i)_S + (M_i - m_i)(\bar{X}_i)'_S] \ .$$

So,

$$M_i [\bar{X}_i - (\bar{X}_i)_S] = (\bar{X}_i)_S (m_i - M_i) + (M_i - m_i)(\bar{X}_i)'_S$$

$$= (M_i - m_i)[(\overline{X}_i)'_S - (\overline{X}_i)_S]$$

and

$$E(\hat{T}_c - T) = 0 \ .$$

Hence, $\hat{T}_c$ is unbiased under the super-population model (4.1.1).

We have the following theorem, analogous to Theorem 2.2, which gives conditions under which $\hat{T}_c$ is unbiased for higher order regression models.

<u>Theorem 4.1</u>. Let the y_{ij} be generated according to the model

$$Y_{ij} = \delta_0\beta_0 + \delta_1\beta_1 X_{ij} + \delta_2\beta_2 X_{ij}^2 + \dots + \beta_p X_{ij}^p + \varepsilon_{ij}$$

where $\delta_i = 0$ or 1; $i = 0, \dots, p-1$, $E(\varepsilon_{ij}) = 0$, and the Y's have the same covariance structure as in Section 2.2. Then $\hat{T}_c$ is unbiased for T for any selection of the δ_i, and any values of the β_i if for each $t = 1, \dots, p$ the X's satisfy the relations

$$\left.\begin{aligned} \overline{(X^t)}_S &= \overline{X^t} \\ \text{and} \quad \overline{(X_i^t)}_S &= \overline{(X_i^t)} \quad \text{for all } i\varepsilon S \end{aligned}\right\} \tag{4.1.4}$$

where

$$\overline{(X_i^t)}_S = \sum_{j\varepsilon S_i} X_{ij}^t / m_i \ ,$$

$$\overline{(X_i^t)} = \sum_{j=1}^{M_i} X_{ij}^t / M_i \ ,$$

$$\overline{(X^t)}_S = \sum_{i\varepsilon S} \sum_{j=1}^{M_i} X_{ij}^t / \sum_{i\varepsilon S} M_i \ ,$$

and
$$\overline{X^t} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} X_{ij}^t / \sum_1^N M_i \quad .$$
If p = 0, only the first equality (4.1.4) where t = 1 must be met.

<u>Proof</u>: We have first that

$$E(\overline{y}_i) = \frac{1}{m_i} [m_i \delta_0 \beta_0 + \delta_1 \beta_1 \sum_{j\varepsilon S_i} X_{ij} + \delta_2 \beta_2 \sum_{j\varepsilon S_i} X_{ij}^2$$
$$+ \ldots + \beta_p \sum_{j\varepsilon S_i} X_{ij}^p]$$
$$= \delta_0 \beta_0 + \delta_1 \beta_1 (\overline{X_i})_S + \delta_2 \beta_2 \ (\overline{X_i^2})_S + \ldots + \beta_p \ (\overline{X_i^p})_S \ .$$

Hence,

$$E(\hat{T}_c - T) = \sum_{i\varepsilon S} (M_i - m_i) [\delta_0 \beta_0 + \delta_1 \beta_1 \ (\overline{X_i})_S + \delta_2 \beta_2 \ (\overline{X_i^2})_S$$
$$+ \ldots + \beta_p \ (\overline{X_i^p})_S]$$
$$+ \sum_{i\varepsilon S} M_i [\delta_0 \beta_0 + \delta_1 \beta_1 \ (\overline{X_i})_S + \delta_2 \beta_2 \ (\overline{X_i^2})_S$$
$$+ \ldots + \beta_p \ (\overline{X_i^p})_S] \ \frac{[\sum_1^N M_i \ \overline{X}_i - \sum_{i\varepsilon S} M_i (\overline{X}_i)_S]}{\sum_{i\varepsilon S} M_i \ (\overline{X}_i)_S}$$
$$- \sum_{i\varepsilon S} (M_i - m_i) [\delta_0 \beta_0 + \delta_1 \beta_1 \ (\overline{X_i})_S' + \delta_2 \beta_2 \ (\overline{X_i^2})_S'$$
$$+ \ldots + \beta_p \ (\overline{X_i^p})_S'] - \sum_{i \not\varepsilon S} M_i [\delta_0 \beta_0 + \delta_1 \beta_1 \ (\overline{X_i})$$

$$+ \delta_2\beta_2 \overline{(X_i^2)} + \ldots + \beta_p \overline{(X_i^p)}]$$

where

$$\overline{(X_i^t)'_S} = \sum_{j \notin S_i} X_{ij}^t/(M_i - m_i) \ .$$

So,

$$E(\hat{T}_c - T) = \delta_0\beta_0 \left(\frac{\sum_{i \in S} M_i [\sum_1^N M_i \bar{X}_i - \sum_{i \in S} M_i (\bar{X}_i)_S]}{\sum_{i \in S} M_i (\bar{X}_i)_S} - \sum_{i \in \bar{S}} M_i \right)$$

$$+ \delta_1\beta_1 \left(\sum_{i \in S} (M_i - m_i)(\bar{X}_i)_S + \sum_{i \in S} M_i (\bar{X}_i)_S \frac{[\sum_1^N M_i \bar{X}_i - \sum_{i \in S} M_i (\bar{X}_i)_S]}{\sum_{i \in S} M_i (\bar{X}_i)_S} \right.$$

$$\left. - \sum_{i \in S} (M_i - m_i) \overline{(X_i)'_S} - \sum_{i \in \bar{S}} M_i \bar{X}_i \right)$$

$$+ \delta_2\beta_2 \left(\sum_{i \in S} (M_i - m_i) \overline{(X_i^2)}_S + \sum_{i \in S} M_i \overline{(X_i^2)}_S \frac{[\sum_1^N M_i \bar{X}_i - \sum_{i \in S} M_i (\bar{X}_i)_S]}{\sum_{i \in S} M_i (\bar{X}_i)_S} \right.$$

$$\left. - \sum_{i \in S} (M_i - m_i) \overline{(X_i^2)}'_S - \sum_{i \in \bar{S}} M_i \overline{X_i^2} \right)$$

$$+ \ldots + \beta_p \left(\sum_{i \in S} (M_i - m_i) \overline{(X_i^p)}_S + \sum_{i \in S} M_i \overline{(X_i^p)}_S \frac{[\sum_1^N M_i \bar{X}_i - \sum_{i \in S} M_i \overline{(X_i)}_S]}{\sum_{i \in S} M_i (\bar{X}_i)_S} \right.$$

$$\left. - \sum_{i \in S} (M_i - m_i) \overline{(X_i^p)}'_S - \sum_{i \in \bar{S}} M_i \overline{X_i^p} \right)$$

$$= \delta_0\beta_0 \left([\sum_{i \in S} M_i \sum_1^N M_i \bar{X}_i - \sum_{i \in S} M_i \overline{(\bar{X}_i)}_S \sum_1^N M_i] / \sum_{i \in S} M_i \overline{(\bar{X}_i)}_S \right)$$

$$+ \delta_2\beta_2 \Bigg\{ \sum_{i\varepsilon S} (M_i - m_i)[\overline{(X_i^2)}_S - \overline{(X_i^2)}_S']$$

$$+ \frac{1}{\sum_{i\varepsilon S} M_i \overline{(X_i)}_S} \Bigg\{ \sum_{i\varepsilon S} M_i \overline{(X_i^2)}_S [\sum_1^N M_i \overline{X}_i - \sum_{i\varepsilon S} M_i \overline{(X_i)}_S]$$

$$- \sum_{i\varepsilon S} M_i \overline{(X_i)}_S [\sum_1^N M_i \overline{(X_i^2)} - \sum_{i\varepsilon S} M_i \overline{(X_i^2)}] \Bigg\} \Bigg\}$$

$$+ \ldots + \beta_p \Bigg\{ \sum_{i\varepsilon S} (M_i - m_i)[\overline{(X_i^p)}_S - \overline{(X_i^p)}_S']$$

$$+ \frac{1}{\sum_{i\varepsilon S} M_i (\overline{X}_i)_S} \Bigg\{ \sum_{i\varepsilon S} M_i \overline{(X_i^p)}_S [\sum_1^N M_i \overline{X}_i - \sum_{i\varepsilon S} M_i (\overline{X}_i)_S]$$

$$- \sum_{i\varepsilon S} M_i \overline{(X_i)}_S [\sum_1^N M_i \overline{(X_i^p)} - \sum_{i\varepsilon S} M_i \overline{(X_i^p)}] \Bigg\} \Bigg\} \quad . \qquad (4.1.5)$$

Using the fact that if

$$\overline{(X_i^t)_S} = \overline{(X_i^t)_S'} ,$$

then

$$\overline{(X_i^t)_S} = \overline{(X_i^t)} ,$$

one can show that (4.1.5) is zero for any choice of the δ_i and any values of the β_i if conditions (4.1.4) are satisfied.

Note that conditions (4.1.4) imply that clusters balanced up to the p^{th} moment must first be selected, and then secondary units also balanced up to the p^{th} moment must be chosen within the sample clusters.

4.1.2 The m.s.e. and Optimum Allocation

Since $\hat{T}_c$ is unbiased for T, it follows that

$$E(\hat{T}_c - T)^2 = \mathrm{Var}(\hat{T}_c - T) \ .$$

Letting $\underline{c}$ be a vector of (K-k) ones and $\underline{y}_{II}$ the vector of all unsampled units, we have

$$\mathrm{Var}(\hat{T}_c - T) = \mathrm{Var}\Big[\sum_{i\varepsilon S} (M_i - m_i)\bar{y}_i + \frac{\sum_{i\varepsilon S} M_i \bar{y}_i [\sum_1^N M_i \bar{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]}{\sum_{i\varepsilon S} M_i (\overline{X_i})_S}$$

$$- \underline{c}' \, \underline{y}_{II}\Big]$$

$$= \sum_{i\varepsilon S} (M_i - m_i)^2 \mathrm{Var}\ \bar{y}_i + \frac{[\sum_1^N M_i \bar{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2}{[\sum_{i\varepsilon S} M_i \ (\overline{X_i})_S]^2} \sum_{i\varepsilon S} M_i^2 \ \mathrm{Var}\ \bar{y}_i$$

$$+ \mathrm{Var}\ \underline{c}' \, \underline{y}_{II} + \frac{2[\sum_1^N M_i \bar{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]}{\sum_{i\varepsilon S} M_i (\overline{X_i})_S} \sum_{i\varepsilon S} (M_i - m_i) M_i \ \mathrm{Var}\ \bar{y}_i$$

$$- 2 \sum_{i\varepsilon S} (M_i - m_i) \sum_{j \not\in S_i} \mathrm{cov}(\bar{y}_i, y_{ij})$$

$$- \frac{2[\sum_1^N M_i \bar{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]}{\sum_{i\varepsilon S} M_i \ (\overline{X_i})_S} \sum_{i\varepsilon S} M_i \sum_{j \not\in S_i} \mathrm{cov}(\bar{y}_i, y_{ij}) \ . \quad (4.1.6)$$

Using the relations given in the derivation of (2.6.4), it follows that (4.1.6) becomes

$$\mathrm{Var}(\hat{T}_c - T) = \sum_{i\varepsilon S} (M_i - m_i)^2 \, \sigma_i^2 [1 + \rho_i \, (m_i - 1)]/m_i$$

$$+ \frac{[\sum_1^N M_i \overline{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2}{\sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2} \sum_{i\varepsilon S} M_i^2 \, \sigma_i^2 \, [1 + \rho_i (m_i - 1)]/m_i$$

$$+ \sum_{i\varepsilon S} (M_i - m_i) \, \sigma_i^2 [1 - \rho_i + \rho_i (M_i - m_i)]$$

$$+ \sum_{i\varepsilon \overline{S}} M_i \, \sigma_i^2 (1 - \rho_i + \rho_i \, M_i)$$

$$+ \frac{2[\sum_1^N M_i \overline{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]}{\sum_{i\varepsilon S} M_i (\overline{X_i})_S} \sum_{i\varepsilon S} (M_i - m_i) M_i \, \sigma_i^2 [1 + \rho_i (m_i - 1)]/m_i$$

$$- 2 \sum_{i\varepsilon S} (M_i - m_i)^2 \, \rho_i \, \sigma_i^2$$

$$- \frac{2[\sum_1^N M_i \overline{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]}{\sum_{i\varepsilon S} M_i (\overline{X_i})_S} \sum_{i\varepsilon S} M_i (M_i - m_i) \, \rho_i \, \sigma_i^2 \ .$$

Letting $\sigma_i^2 = \sigma^2$ and $\rho_i = \rho$, we obtain, after simplification, that

$$\mathrm{Var}(\hat{T}_c - T)/\sigma^2 = (K-k)(1-\rho) + \rho \left(\sum_{i\varepsilon \overline{S}} M_i^2 + \sum_{i\varepsilon S} M_i^2 \, \frac{[\sum_1^N M_i \overline{X}_i - \sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2}{[\sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2} \right)$$

$$+ \frac{(1-\rho)}{[\sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2} \sum_{i\varepsilon S} [M_i \sum_1^N M_i \overline{X}_i - m_i \sum_{i\varepsilon S} M_i (\overline{X_i})_S]^2 / m_i \ . \tag{4.1.7}$$

The optimum allocation of the m_i is given by the following theorem:

Theorem 4.2. Let $\rho_i = \rho$ and $\sigma_i^2 = \sigma^2$. Then for a fixed total sample size k and a given selection of clusters, the values of m_i which minimize (4.1.7) are obtained by proportional allocation. That is,

$$m_i = k\, M_i / \sum_{i\varepsilon S} M_i \; ; \quad i\varepsilon S \; .$$

The proof is similar to that for Theorem 2.6 and is not given.

4.1.3 Comparison to Conventional Estimator of Chapter II in Terms of Bias

Under model (4.1.1) it is of interest to compare the two conventional estimators $\hat{T}_c$ and $\hat{T}_N$ given by (2.6.1). Now, the values of X_i in (2.6.1) represent previous cluster totals; hence, if previous information is available on each secondary unit, X_i is simply $\sum_{j=1}^{M_i} X_{ij}$.

The bias of $\hat{T}_N$ under model (4.1.1) is as follows:

$$E(\hat{T}_N - T) = \beta \, \{\sum_{i\varepsilon S} (M_i - m_i)\overline{(X_i)}_S + \frac{\sum_{i\varepsilon S} M_i (X_i)_S \sum_{i\varepsilon \bar{S}} M_i \bar{X}_i}{\sum_{i\varepsilon S} M_i \overline{X_i}}$$

$$- \, [\sum_{i\varepsilon S} (M_i - m_i)\overline{(X_i)}_S' + \sum_{i\varepsilon \bar{S}} M_i \overline{X_i}]\}$$

$$= \beta\{\sum_{i\varepsilon S} (M_i - m_i)[\overline{(X_i)}_S - \overline{(X_i)}_S']$$

$$+ \frac{\sum_{i\varepsilon \bar{S}} M_i \bar{X}_i}{\sum_{i\varepsilon S} M_i \bar{X}_i} \, [\sum_{i\varepsilon S} M_i \overline{(X_i)}_S - \sum_{i\varepsilon S} M_i \bar{X}_i]\}$$

$$= \beta\{\sum_{1}^{N} M_i\overline{X}_i \sum_{i\varepsilon S} M_i[(\overline{X_i})_S - \overline{X}_i]/ \sum_{i\varepsilon S} M_i\overline{X}_i\} .$$

Hence, $\hat{T}_N$ is unbiased under model (4.1.1) if

$$(\overline{X_i})_S = \overline{X}_i \quad \text{for all } i\varepsilon S.$$

That is, $\hat{T}_N$ is unbiased if the units selected from each sample cluster are balanced on the X's. One notes from (2.6.1) and (4.1.2) that if the secondary units are balanced on the X's, then $\hat{T}_N = \hat{T}_c$.

V. AN EXAMPLE

5.1 Description

The Superintendent of Public Instruction for the State of Virginia,[20], annually publishes high school enrollment for each public high school in each county. It is desired to sample a certain number of high schools in various counties during the 1972-73 school session and predict the total high school enrollment for all counties for the 1972-73 session. The counties thus play the role of clusters, and the high schools are the secondary units.

Since some of the counties had new high schools for the 1972-73 session, previous auxiliary information is not available on each secondary unit. Thus, one is forced to adopt the methods of Chapter II in which the auxiliary information is taken to be previous cluster totals. For the example, the auxiliary variable X_i represents the total high-school enrollment for the i^{th} county during the 1971-72 school year.

5.2 Estimators

All estimators and m.s.e.'s are calculated under the assumption of equal variances and correlations. For each sample, the intra-class correlation coefficient, $\hat{\rho}$, given by (2.4.2), is used to estimate ρ. The variance σ^2 is estimated by (2.4.1). It is easy to show that (2.4.1) can be written as

$$\widehat{\sigma^2} = (\underline{Y}_I' H^{-1} \underline{Y}_I - \underline{Y}_I' H^{-1} Z_I \hat{\beta})/(k-1)$$

$$= \sum_{i\varepsilon S} \{ \sum_{j\varepsilon S_i} y_{ij}^2 h_{jj}^{-1}(i) + 2 \sum_{j\varepsilon S_i} \sum_{\ell\varepsilon S_i} y_{ij} y_{i\ell} h_{j\ell}^{-1}(i)$$

$$- \frac{\hat{\beta}X_i}{M_i} \sum_{j\varepsilon S_i} y_{ij} \sum_{\ell\varepsilon S_i} h_{j\ell}^{-1}(i)\}/(k-1) \quad ,$$

where $h_{j\ell}^{-1}(i)$ is the (j, ℓ) element of the i^{th} principal submatrix of dimensions $(m_i \times m_i)$ of H^{-1}. This latter form is the one that was programmed to estimate σ^2.

Six estimators are used. They are

1. The b.l.u.e. $\hat{T}_B$ given by (2.3.2) with $\rho_i = \hat{\rho}$ and $\sigma_i^2 = \sigma^2$. The estimate of the m.s.e. is given by (2.4.6) where $\sigma^2 = \widehat{\sigma^2}$ and $\rho = \hat{\rho}$.
2. The sub-optimum estimator $\hat{T}_B$ (0) given by (2.3.2) with $\rho_i = 0$ and $\sigma_i^2 = \sigma^2$. The estimate of the m.s.e. is given by (2.5.2) where $\rho_a = 0$, $\rho_t = \hat{\rho}$, and $\sigma^2 = \widehat{\sigma^2}$.
3. The sub-optimum estimator $\hat{T}_B$ (1) given by (2.3.2) with $\rho_i = 1$ and $\sigma_i^2 = \sigma^2$. The estimate of the m.s.e. is given by (2.5.2) where $\rho_a = 1$, $\rho_t = \hat{\rho}$, and $\sigma^2 = \widehat{\sigma^2}$.
4. The conventional estimator $\hat{T}_N$ given by (2.6.1). The estimate of the m.s.e. is given by (2.6.4) where $\rho = \hat{\rho}$ and $\sigma^2 = \widehat{\sigma^2}$.

It is of interest to compare these estimators to two estimators suggested by Royall [17] which do not utilize any auxiliary information.Royall gives the estimators

5. $\hat{T}_H = \sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij} + \sum_{i\varepsilon S} (M_i - m_i)\, \overline{y}_i + \sum_{i\varepsilon S} m_i\, \overline{y}_i \sum_{i\varepsilon \overline{S}} M_i/k$

and

6. $\hat{T}_P = \sum_{i\varepsilon S} \sum_{j\varepsilon S_i} y_{ij} + \sum_{i\varepsilon S} (\overline{M}_S - m_i)\, \overline{y}_i + \sum_{i\varepsilon S} \overline{y}_i \sum_{i\varepsilon \overline{S}} M_i/n$

where

$$\overline{M}_S = \sum_{i\varepsilon S} M_i/n \ .$$

It is easy to show that both estimators are biased under model (2.2.1); hence, the m.s.e.'s are not given. Royall suggests that with equal allocation and $\hat{T}_P$, one should select the largest, in terms of M , clusters; while with equal allocation and $\hat{T}_H$, one should select large clusters of nearly equal size. With proportional allocation, Royall suggests choosing large clusters of nearly equal size for both $\hat{T}_H$ and $\hat{T}_P$.

5.3 General Discussions of Samples

For the example, many assumptions of the theorems given in Chapter II regarding which and how many clusters should be sampled are not met. In particular, it is not true that $M_i = M$ or that X /M is increasing with M . Hence, Theorems 2.1, 2.3, 2.4, and 2.5 can not be used to guide one in the selection of the clusters.

Several samples are selected. Some consisting only of largest clusters, some consisting of large clusters of nearly equal size, some consisting only of smallest clusters, and one consisting of balanced clusters--all in terms of the X's. Also, one sample is chosen which includes each cluster in the population. Equal and proportional (approximate) allocations are used.

The choice of schools to be sampled is made on the basis of the previous 1971-72 enrollment. Recall from Section 2.7 that whenevern $m_i < M_i$, and the estimator $\hat{T}_N$ is to be used; then a strong case can be made for choosing those particular m_i schools such that the previous (1971-72) average of these m_i schools equals the previous (1971-72 average of <u>all</u> schools for the i^{th} county. For the example, this criteria of selecting the schools is adopted for each of the six estimators.

5.4 Description of Tables and Figure

Table 5.1 lists each county in the population, in order of increasing X, and the enrollment for each school. Schools denoted by * are used in the sample for which <u>all</u> counties are selected.

Figure 5.1 is a graph of the 1972-73 averages of all schools in various counties vs. the corresponding X /M values. This figure gives an indication of the validity of model (2.2.1).

Table 5.2 gives values of the pertinent parameters which are needed in the evaluation of the estimators and m.s.e.'s. The total T of all schools in the population is also given.

Table 5.3 describes the various samples chosen. The column labeled $\overline{Y}$ gives the averages of <u>all</u> schools in the county.

Table 5.4 gives the estimates of ρ, σ^2, and the β's for each of the samples. The regression coefficient estimates $\hat{\beta}_B$, $\hat{\beta}_0$, and

$\hat{\beta}_1$ represent the estimates of β, given below (2.3.2), obtained by letting $\rho = \hat{\rho}$, 0, and 1 respectively. Since all $m_i = 1$ for sample S_8, the estimate (2.4.2) of ρ can not be used. The estimate of ρ obtained from sample S_4 was chosen for S_8.

Table 5.5 gives the values of each of the six estimators for each sample.

Table 5.6 shows the absolute percent error of the estimators and also the average percent error of each estimator taken over all eight samples.

Table 5.7 gives the estimates of the m.s.e.'s. The numbers in parentheses represent the power of 10 that each entry should be multiplied by to obtain the final m.s.e. estimate.

5.5 Results and Discussion

1. Tables 5.6 and 5.7 show that for many cases an estimator will out-perform another in terms of absolute percent error; whereas the estimate of the m.s.e. of the former estimator will exceed that of the latter. For example, Table 5.6 shows that sample S_6 yields better results for the estimators $\hat{T}_B$, $\hat{T}_B$ (0), $\hat{T}_B$ (1), and $\hat{T}_N$ than any other sample. However, Table 5.7 indicates that sample S_6 gives higher estimates of the m.s.e. of these estimators than many of the other samples.

The reason for this situation occurring can be explained by realizing that the values in the finite population in the example are but <u>single realizations</u> of the random variables describing the super population. Consequently, (i) the estimates of the m.s.e.'s are just that--namely, estimates, and (ii) even if the estimated m.s.e.'s were, in fact, equal to the correct m.s.e.'s, their magnitudes imply that any single realization of the estimators can fluctuate widely.

2. Each of the estimators $\hat{T}_B$, $\hat{T}_B$ (0), $\hat{T}_B$ (1), and $\hat{T}_N$ perform very well for all eight samples with the exception of S_5. Even for S_5, however, the absolute percent error for each of these estimators is less than 4%.

The excellent performance of these estimators can probably be attributed to (i) the validity of the model and (ii) the validity of the method used for selecting the second-stage units. It is seen from Figure 5.1 that the population means $\overline{Y}_i$ appear to be realized values of a random variable with mean $\beta X_i/M_i$ where β is near one. Thus the model $Y_{ij} = \beta X_i/M_i + \varepsilon_{ij}$ considered in Chapter II appears valid.

Regarding the selection of the second-stage units, one notes from Table 5.3 that the sample means $\overline{y}_i$ are fairly close to the population means $\overline{Y}_i$, and in many cases are equal. Hence, the criteria described in Section 2.7 for selecting the second-stage units to use with $\hat{T}_N$ seems to be satisfactory. Furthermore, the criteria appears to give good results when using $\hat{T}_B$, $\hat{T}_B$ (0), or $\hat{T}_B(1)$.

3. The two estimators $\hat{T}_H$ and $\hat{T}_P$, which do not use any auxiliary information perform very poorly except for $\hat{T}_H$ used with sample S_8. For samples in which all clusters in the population are selected

it is easy to see that $\hat{T}_H = \hat{T}_B\ (1)$.

4. For the estimators $\hat{T}_B$, $\hat{T}_B\ (0)$, and $\hat{T}_N$, sample S_3, which uses proportional allocation, yields better results than sample S_4, which uses the same clusters but with equal allocation.

5. The estimators $\hat{T}_B$ and $\hat{T}_B\ (0)$ give better results for samples $S_1 - S_4$, which include units from the largest county, than for samples S_5 and S_7, which do not include units from the largest county. However, Table 5.7 shows that all m.s.e. estimates using S_7 are smaller than when using $S_1 - S_4$.

6. Table 5.7 shows that sample S_8 produced the smallest estimates of the m.s.e.'s. However, Table 5.6 shows that many of the other samples actually out-performed S_8 even though S_8 contains more than twice as many second-stage units.

7. The assumption of equal variances and correlations, and the estimation of ρ with the intra-class correlation coefficient (2.4.2) produced good results for the b.l.u.e. $\hat{T}_B$.

8. It is seen from Table 5.4 that all three estimates of the regression coefficient β are nearly equal for any given sample. Thus, the estimator of β, given below (2.3.2), appears to be fairly insensitive to choices of ρ.

9. Also, Table 5.4 shows each estimate of β to be close to one. The graph of the population means in Figure 5.1 would imply a slope near one.

TABLE 5.1

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(1) Highland	1	216	228	(20) Powhatan	1	*474	825
(2) King & Queen	1	395	380		2	434	
(3) Bath	1	418	391	(21) Northum.	1	*456	847
(4) Craig	1	350	407		2	372	
(5) Rappahanock	1	439	439	(22) Richmond	1	*457	857
(6) Greene	1	437	471		2	413	
(7) King William	1	505	477	(23) Clarke	1	*619	874
(8) Cumberland	1	582	501		2	333	
(9) Surry	1	519	521	(24) Floyd	1	890	883
(10) Middlesex	1	583	566	(25) Prince Edward	1	742	889
(11) Bland	1	*277	574	(26) Goochland	1	526	916
	2	262			2	*486	
(12) Essex	1	673	610	(27) Grayson	1	*110	916
(13) New Kent	1	693	698		2	112	
(14) Amelia	1	716	706		3	574	
(15) Fluvanna	1	371	729		4	128	
	2	*362		(28) Buckingham	1	*423	930
(16) Madison	1	824	747		2	493	
(17) Mathews	1	*435	785	(29) Lunenberg	1	*473	954
	2	376			2	433	
(18) Charles City	1	831	809	(30) Sussex	1	595	1077
(19) King George	1	822	819		2	*247	

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(30) Sussex	3	224		(42) Nelson	1	*828	1357
(31) Rockbridge	1	*485	1079		2	526	
	2	578		(43) Orange	1	*1106	1362
(32) Charlotte	1	515	1102		2	1051	
	2	*601		(44) Brunswick	1	*623	1391
(33) Westmoreland	1	*536	1148		2	770	
	2	660		(45) Giles	1	*921	1508
(34) Lancaster	1	620	1181		2	566	
	2	*594		(46) Nottoway	1	410	1552
(35) Allegheny	1	1237	1217		2	*498	
(36) Northampton	1	615	1242		3	664	
	2	*644		(47) Gloucester	1	949	1570
(37) Appomattox	1	*572	1295		2	*706	
	2	821		(48) Greensville	1	*758	1578
(38) Patrick	1	1304	1299		2	716	
(39) Page	1	*646	1333	(49) Spotsylvania	1	882	1646
	2	762			2	*942	
(40) Warren	1	759	1336	(50) Carolina	1	373	1647
	2	*669			2	707	
(41) Southampton	1	*659	1341		3	*675	
	2	660		(51) Culpepper	1	*1254	1648

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(51) Culpepper	2	937		(60) Lee	2	245	
(52) Dickenson	1	732	1802		3	438	
	2	469			4	179	
	3	*566			5	753	
(53) Amherst	1	*1493	1886		6	*393	
	2	519		(61) Mecklenberg	1	*633	2459
(54) Isle of Wight	1	*1272	1967		2	692	
	2	650			3	664	
(55) Shenandoah	1	827	2043		4	663	
	2	*641		(62) Wythe	1	716	2473
	3	615			2	*593	
(56) Dinwiddie	1	*1012	2105		3	596	
	2	1031			4	511	
(57) Louisa	1	1040	2129	(63) Russell	1	722	2595
	2	*1162			2	520	
(58) Botetourt	1	*863	2159		3	791	
	2	566			4	*700	
	3	865		(64) Franklin	1	1366	2596
(54) Prince George	1	*1082	2298		2	*1233	
	2	891		(65) Pulaski	1	1326	2603
(60) Lee	1	298	2319		2	*1287	

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(66) Carroll	1	1156	2636	(70) Fauquier	1	322	3046
	2	595			2	1719	
	3	273			3	486	
	4	110			4	*510	
	5	*542		(71) Stafford	1	*1167	3224
(67) Scott	1	902	2772		2	1307	
	2	*481			3	1008	
	3	464		(72) Montgomery	1	550	3244
(68) Accomac	1	389	2776		2	1449	
	2	365			3	*1083	
	3	356			4	577	
	4	713		(73) Bedford	1	349	3310
	5	468			2	973	
	6	*350			3	*1114	
	7	91			4	932	
(69) Smyth	1	*617	2932	(74) Frederick	1	1002	3335
	2	612			2	1394	
	3	1143			3	*979	
	4	542		(75) Buchanan	1	347	3458
	5	427			2	492	
	6	141			3	876	

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(75) Buchanan	4	843		(80) Hanover	3	965	
	5	*502			4	*970	
	6	376		(81) York	1	422	4016
(76) Washington	1	1193	3663		2	*1000	
	2	483			3	758	
	3	1415			4	1575	
	4	*837			5	535	
(77) Albemarle	1	1894	3676	(82) Augusta	1	*934	4091
	2	*859			2	966	
	3	816			3	548	
	4	311			4	767	
(78) Wise	1	505	3741		5	932	
	2	*689		(83) Campbell	1	756	4701
	3	925			2	1864	
	4	567			3	*971	
	5	697			4	677	
	6	361			5	724	
(79) Halifax	1	*1675	3674	(84) Henry	1	904	4983
	2	2125			2	1251	
(80) Hanover	1	1135	3807		3	891	
	2	807			4	1090	

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(84) Henry	5	*1050		(88) Pittsylvania	3	1077	
(85) Rockingham	1	*992	4995		4	834	
	2	731			5	899	
	3	562			6	1006	
	4	660		(89) Roanoke	1	1492	11313
	5	1058			2	1416	
	6	1130			3	1237	
(86) Tazewell	1	*701	5067		4	*1204	
	2	402			5	1235	
	3	451			6	1700	
	4	1246			7	1316	
	5	828			8	622	
	6	914			9	1176	
	7	465			10	1060	
(87) Loudon	1	1041	5775	(90) Arlington	1	773	11619
	2	*1216			2	1123	
	3	976			3	1045	
	4	1153			4	837	
	5	1305			5	760	
(88) Pittsylvania	1	*1056	5652		6	1796	
	2	832			7	1739	

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(90) Arlington	8	*1095		(92) Henrico	8	2057	
	9	1901			9	1866	
(91) Chesterfield	1	557	11993		10	1398	
	2	951			11	1449	
	3	1142		(93) Prince Wm.	1	717	17532
	4	1017			2	*1203	
	5	1171			3	1533	
	6	926			4	2697	
	7	1334			5	867	
	8	*1015			6	855	
	9	1052			7	1576	
	10	1064			8	1662	
	11	967			9	1195	
	12	1340			10	1435	
(92) Henrico	1	1189	17060		11	2052	
	2	1172			12	2419	
	3	1914		(94) Fairfax	1	2297	67902
	4	1432			2	1112	
	5	1681			3	1016	
	6	1638			4	1795	
	7	*2049			5	1924	

TABLE 5.1. (Continued)

1972-73 Enrollment for Each School and Total

1971-72 Enrollment for Each County

County	School	y	X	County	School	y	X
(94) Fairfax	6	*1978		(94) Fairfax	23	1748	
	7	2583			24	969	
	8	1440			25	1126	
	9	1746			26	2391	
	10	886			27	2451	
	11	1442			28	2022	
	12	3498			29	1886	
	13	2953			30	2319	
	14	1744			31	1366	
	15	1364			32	4724	
	16	1719			33	1206	
	17	1816			34	853	
	18	2063			35	2759	
	19	941			36	622	
	20	1472			37	1066	
	21	2251			38	2899	
	22	1210					

Note: Nanesomond County was omitted from 1972-73 Report.

Schools denoted by * were those selected for sample S_8 for clusters with $M_i > 1$.

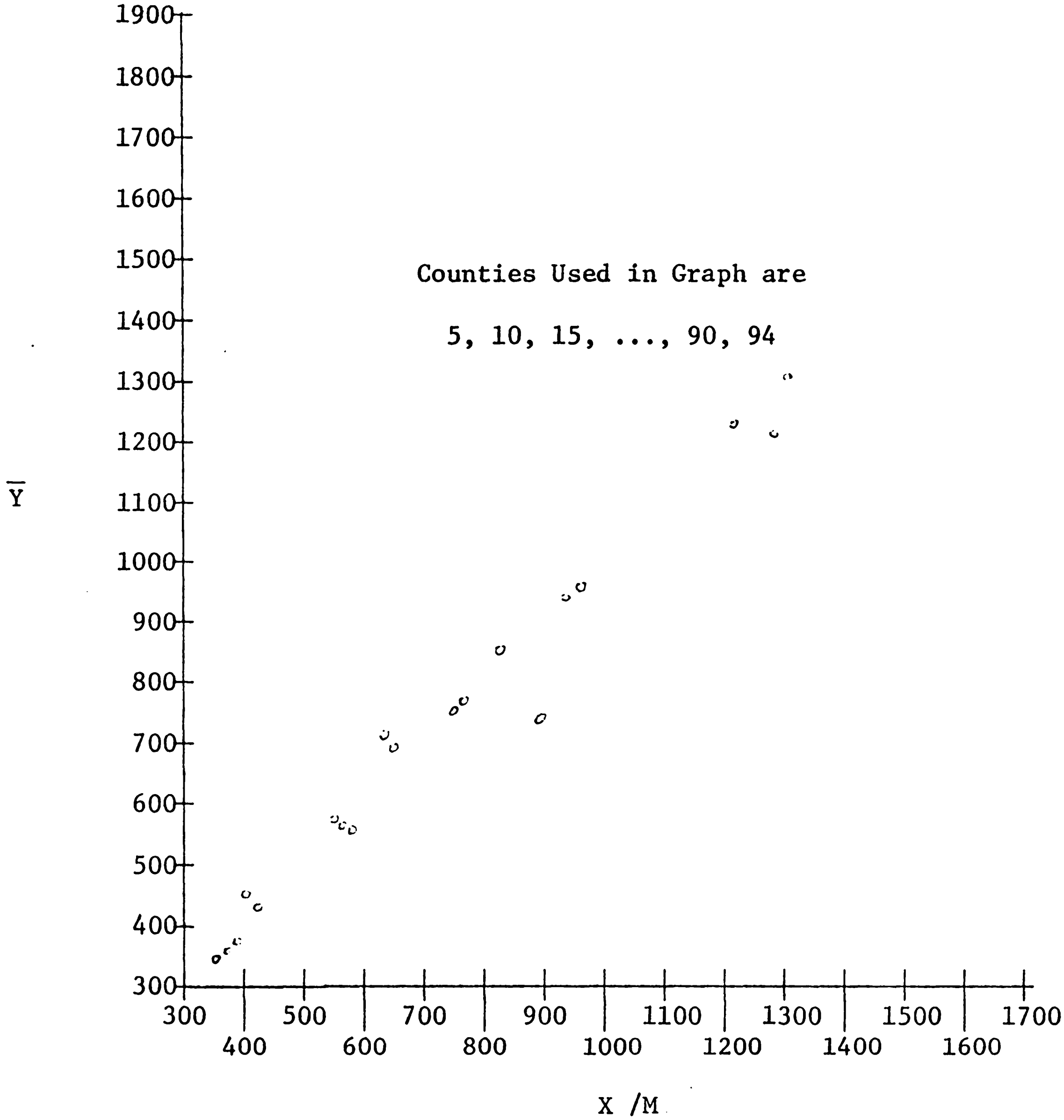


Figure 5.1

1972-73 County Averages of all Schools vs. Corresponding

X /M Values for Various Counties

TABLE 5.2

Parameter Values

$K = 329$

$k = \begin{cases} 40 \text{ for samples } S_1 - S_7 \\ 94 \text{ for sample } S_8 \end{cases}$

$N = 94$

$\sum_{1}^{N} X_i = 306,569$

$\overline{X} = 931.82$

$\sum_{1}^{N} M_i^2 = 2899$

$T = 315,768$

TABLE 5.3

Description of Samples

Sample	County	Schools	Samp. Size	Type of Allocation	Sample Mean - $\bar{y}$	Population mean - $\bar{Y}$
S_1	93	7,8	2	partial	1619	1516
	94	all	38	complete	1833	1833
S_2	93	1-6,9-12	10	prop.	1497.3	1516
	94	30,33,36,38	30		1820.8	1833
S_3	91	5-11	7	prop.	1075.6	1045
	92	2,3,5,7,9,10	6		1680	1622
	93	1,2,7-10,12	7		1458.1	1516
	94	4-6,8,11,13-18 20,21,23,24,26-28 30,36	20		1824.1	1833
S_4	91	2-7,9-12	10	equal	1096.4	1045
	92	2-11	10		1665.6	1622
	93	1-6,9-12	10		1497.3	1516
	94	4-6,11,14,17	10		1786.8	1333

TABLE 5.3 (Continued)

Description of Samples

Sample	County	Schools	Samp. Size	Type of Allocation	Sample Mean - $\bar{y}$	Population Mean - $\bar{Y}$
S_5	89	1-10	10	equal	1245.8	1245.8
	91	2-7,9-12	10		1096.4	1045
	92	2-11	10		1665.6	1622
	93	1-6,9-12	10		1497.3	1516
S_6*	27	1,3,4	3	prop.	270.7	231
Balanced on 25 Counties	78	3,4,6	3		617.7	624
	80	1,2,4	3		970.7	969
	1-15,20-23	all	M_i		$\bar{y}_i = \bar{Y}_i$	
	26,28,53	all	M_i		$\bar{y}_i = \bar{Y}_i$	
S_7	1-28	all	M_i	complete	$\bar{y}_i = \bar{Y}_i$	
S_8	1-94	Denoted by * in Table 1 for $M_i > 1$	1	equal	$\bar{y}_i=\bar{Y}_i$ for $M_i=1$. Others not shown.	

*For sample S_6, $\bar{X}_S = 931.88$; whereas for complete population $\bar{X} = 931.82$.

TABLE 5.4

Parameter Estimates

Sample	$\hat{\rho}$	$\hat{\beta}_B$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}^2$
S_1	-0.027	1.026	0.029	1.059	665710
S_2	-0.021	1.020	1.020	1.021	440642
S_3	0.120	1.035	1.030	1.039	225908
S_4	0.253	1.038	1.038	1.038	192505
S_5	0.162	1.067	1.067	1.067	182653
S_6	0.404	1.031	1.031	1.030	38497
S_7	0.149	1.013	1.014	1.010	7680
S_8	0.253	1.038	1.038	1.038	21238

TABLE 5.5

Values of the Estimators

Sample	$\hat{T}_B$	$\hat{T}_N$	$\hat{T}_B(0)$	$\hat{T}_B(1)$	$\hat{T}_H$	$\hat{T}_P$
S_1	314669	319670	315392	323227	597507	567857
S_2	312613	312751	312722	313005	562039	483887
S_3	316986	316518	316003	317735	529723	496623
S_4	317031	314220	318376	316634	504294	497285
S_5	327273	327781	327308	327255	452766	452794
S_6	315962	315390	315950	315767	172165	160906
S_7	310639	312001	310874	309820	158150	172443
S_8	319719	323243	318527	323243	323243	257191

TABLE 5.6

Absolute Percent Error in Estimates of Total

Sample \ Est.	$\hat{T}_B$	$\hat{T}_N$	$\hat{T}_B(0)$	$\hat{T}_B(1)$	$\hat{T}_H$	$\hat{T}_P$
s_1	0.348	1.236	0.119	2.362	89.223	79.834
s_2	0.999	0.955	0.965	0.875	77.991	53.241
s_3	0.386	0.238	0.074	0.623	67.757	57.275
s_4	0.400	0.490	0.826	0.274	59.704	57.484
s_5	3.643	3.804	3.655	3.638	43.386	43.458
s_6	0.061	0.120	0.058	0.003	45.477	49.043
s_7	1.624	0.119	1.550	1.884	49.916	45.389
s_8	1.251	2.367	1.015	2.367	2.367	18.551
Average	1.089	1.166	0.748	1.503	54.478	50.534

TABLE 5.7

Estimates of the m.s.e.

Sample	$\widehat{E(\hat{T}_B-T)^2}$	$\widehat{E(\hat{T}_N-T)^2}$	$\widehat{E(\hat{T}_B(0)-T)^2}$	$\widehat{E(\hat{T}_B(1)-T)^2}$
s_1	1.7328 (8)	7.3367 (8)	1.8500 (8)	1.8143 (9)
s_2	2.2914 (8)	2.3930 (8)	2.3551 (8)	3.1166 (8)
s_3	3.3346 (8)	3.4744 (8)	3.5652 (8)	3.4330 (8)
s_4	4.0480 (8)	4.8712 (8)	4.2087 (8)	4.0620 (8)
s_5	5.2912 (8)	5.4404 (8)	5.2917 (8)	5.2913 (8)
s_6	4.0334 (8)	5.0593 (8)	4.2151 (8)	4.1880 (8)
s_7	7.1389 (7)	8.5221 (7)	7.1805 (7)	7.4855 (7)
s_8	2.3632 (7)	4.0774 (7)	2.5598 (7)	4.0774 (7)

LIST OF REFERENCES

[1] Brewer, K. R. W., "Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process", The Australian Journal of Statistics, 5, 93-105, 1963.

[2] Cochran, W. G., Sampling Techniques, New York; Wiley, 1963.

[3] Des Raj, "On the Relative Accuracy of Some Sampling Techniques", Journal of the American Statistical Association, 53, 98-101, 1958.

[4] Graybill, F. A., Introduction to Matrices with Applications in Statistics, Belmont, California; Wadsworth, 1969.

[5] Greenstreet, R. L., Unpublished Ph.D. Thesis, John Hopkins School of Public Health and Hygiene, Baltimore, Md., 1973.

[6] Greenstreet, R. L., "Empirical Results in Finite Population Sampling", Proceedings of the Social Statistics Section, Journal of the American Statistical Association, 256-261, 1973.

[7] Greenstreet, R. L. and Madden, R. B., "Some Design Considerations for Stratified Sampling", Proceedings of the Social Statistics Section, Journal of the American Statistical Association, 449-454, 1975.

[8] Greenstreet, R. L. and Madden, R. B., "The Linear Least-Squares Prediction Approach to Multipurpose Sampling", submitted to Journal of the American Statistical Association, October, 1975, presently under revision.

[9] Hartley, H. O. and Sielken, R. L., Jr., "A Super-Population Viewpoint for Finite Population Sampling", Biometrics, 31, 411-422, 1975.

[10] Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, V. 2, New York; Hafner, 1967.

[11] Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, V. 3, New York; Hafner, 1968.

[12] Rao, C. R., Linear Statistical Inference and Its Applications, New York; Wiley, 1965.

[13] Royall, R. M., "On Finite Population Sampling Theory Under Certain Linear Regression Models", Biometrika, 57, 377-387, 1970.

[14] Royall, R. M., "Linear Regression Models in Finite Population Sampling Theory", in Godambe, V. P. and Spratt, P. A., eds., Foundations of Statistics Inference, Toronto; Holt, Rinehart, and Winston, 1971.

[15] Royall, R. M. and Herson, J., "Robust Estimation in Finite Populations I", Journal of the American Statistical Association, 68, 880-9, 1973.

[16] Royall, R. M. and Herson, J., "Robust Estimation in Finite Populations II: Stratification on a Size Variable", Journal of the American Statistical Association, 68, 890-3, 1973.

[17] Royall, R. M., "The Linear Least-Squares Prediction Approach to Cluster Sampling", Journal of the American Statistical Association, 71, 657-664, 1976.

[18] Scott, A. and Smith, T. M. F., "Estimation in Multi-Stage Surveys", Journal of the American Statistical Association, 64, 830-40, 1969.

[19] Sukhatme, P. V. and Sukhatme, B. V., Sampling Theory of Surveys with Applications, Ames, Iowa; Iowa State University Press, 1970.

[20] Superintendent of Public Instruction for the State of Virginia, Annual Report, Richmond, Va.; State Board of Education, 1972 and 1973.

[21] Whittle, P., Prediction and Regulation by Linear Least-Square Methods, London; The English Universities Press Ltd., 1963.

APPENDIX--A NON-CLUSTER MULTIVARIATE MODEL

A.1 Introduction

The following study represents additional work the author has done in finite population sampling utilizing a super-population viewpoint. The basic model considered is multivariate, and the study does not, in general, relate to the cluster models considered in the previous chapters.

Greenstreet, [5], was the first to investigate multivariate models in a super-population context. In [5] and also in Greenstreet and Madden, [8], several multivariate estimators are derived and analyzed. Problems which one may incur in attempting to use these estimators are discussed, and mention is made of the need for further research in the comparison of multivariate estimators to simpler univariate estimators. This study is a part of that comparison.

First, we have the following multivariate setup, as given in [5].

The population consists of N units. Associated with each unit there are p unknown characteristics and q known auxiliary variables. Let y_{ij} represent the value of the j^{th} characteristic in the i^{th} unit, $\underline{y}_j$ the $(N\times 1)$ vector of all N values of the j^{th} characteristic, $\underline{y}' = (\underline{y}_1', \ldots, \underline{y}_p')$, X_{ij} the j^{th} auxiliary variable in the i^{th} unit, and let $\underline{X}_i$ be the $(q\times 1)$ vector of all auxiliary variables for the i^{th}

unit. Consider a matrix Z defined by

$$Z = \begin{pmatrix} z_1(\underline{X}_1'), & \dots, & z_r(\underline{X}_1') \\ \vdots & & \\ z_1(\underline{X}_N'), & \dots, & z_r(\underline{X}_N') \end{pmatrix}$$

where the z's are <u>known</u> functions of the $\underline{X}$'s and Z is of full rank. For each j = 1,..., p let $\underline{y}_j$ be a realization of a random variable $\underline{Y}_j$ with $E[\underline{Y}_j] = Z\underline{\beta}_j$ where $\underline{\beta}_j' = (\beta_{j1},\dots,\beta_{jr})$ and the β_j's are <u>unknown</u>. We have

$$\underline{Y} = \begin{pmatrix} \underline{Y}_1 \\ \vdots \\ \underline{Y}_p \end{pmatrix} = \begin{pmatrix} Z & & \\ & \ddots & \\ & & Z \end{pmatrix} \begin{pmatrix} \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_p \end{pmatrix} + \underline{\varepsilon}$$

or $\underline{Y} = Z^* \, \underline{\beta} + \underline{\varepsilon}$

where $E(\underline{\varepsilon}) = \underline{0}$. Also, let

$$V = \mathrm{cov}[Y] = \begin{pmatrix} V_{11}, & V_{12}, & \dots, & V_{1p} \\ \text{symm.} & V_{22}, & \dots, & V_{2p} \\ & & \ddots & \\ & & & V_{pp} \end{pmatrix} .$$

That is, V_{jj} represents the (N×N) matrix which describes the covariation of the j^{th} characteristic among the N units, and V_{jk} represents the (N×N) matrix which describes the covariation between the j^{th} and k^{th} characteristic among the N units. Each of these submatrices being diagonal implies that the N units are uncorrelated and conversely. We shall always assume uncorrelated units.

Now, based on a sample S of n units, it is desired to estimate the total T_j of the j^{th} characteristic where $T_j = \sum_{i=1}^{N} y_{ij}$. Greenstreet [5] shows that if $\underline{\beta}_1, \ldots, \underline{\beta}_p$ are linearly independent, and the V_{ij} are diagonal matrices; i, j = 1, ..., p; then the b.l.u.e. of T_j is given by

$$\hat{T}_j = \sum_{i\varepsilon S} y_{ij} + \sum_{i\varepsilon \bar{S}} \sum_{k=1}^{r} z_k (\underline{X}_i') \hat{\beta}_{jk}$$

$$= \sum_{i\varepsilon S} y_{ij} + \underline{\omega}' \hat{\underline{\beta}}_j \qquad (A.1.1)$$

where

$$\hat{\underline{\beta}}_j' = (\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jr})$$

and is the j^{th} vector of dimension r in the (rp×1) vector

$$\hat{\underline{\beta}} = [Z^*(S)'V^{-1}(S)Z^*(S)]^{-1} Z^*(S)'V^{-1}(S)\underline{y}(S) \qquad (A.1.2)$$

where $Z^*(S)$, $V(S)$, and $\underline{y}(S)$ are analogous to Z^*, V, and $\underline{y}$ except restricted to the sample units.

In general, $\hat{T}_j$ will be a <u>multivariate</u> estimator of T_j in the sense that observations on characteristics other than the j^{th} will enter into the estimator. However, the following theorem gives sufficient conditions for which $\hat{T}_j$ is a univariate estimator in the sense that only observations on the j^{th} characteristic will enter into the estimator.

<u>Theorem--Greenstreet [5].</u> If for each $i = 1,\ldots,p$; $V_{ij} = d_{ij} V_{jj}$ where the d_{ij} are scalars and the submatrices are all diagonal,

then the b.l.u.e. of T_j is based only on observations on the j^{th} characteristic.

If the conditions of the theorem are met, then one can obtain the b.l.u.e. of T_j by observing values only for the j^{th} characteristic. Thus the covariance matrices V_{ij}; $i \neq j$, do not enter into the b.l.u.e. Also, note (i) for uncorrelated units and uncorrelated characteristics within the units, the conditions of the theorem are satisfied and (ii) if the conditions are met and if $V_{jj} = \sigma_j^2 I$; that is, the variance of y_{ij} is a constant for all units, then we see from (A.1.1) and (A.1.2) that $\hat{T}_j$ is independent of σ_j^2.

A.2 A Problem Relating to the Model

In many real problems, the elements of the matrices V_{ij}; $i, j = 1, \ldots, p$, will be functionally dependent on the auxiliary variables. Furthermore, the form of this dependence will usually be unknown. Hence, in general, the optimum estimator of T_j cannot be obtained exactly, but only approximated by assuming a particular covariance structure, say V_a, for the elements. Let this approximate estimator be denoted by $\hat{T}_j'$, obtained under the assumption that $V = V_a$.

Now, consider a univariate estimator derived under the following assumptions: (i) Suppose the auxiliary variables can be stratified into, say, H strata in such a way that within each stratum the variances and covariances of all p characteristics can be considered constant. (ii) Let T_j be considered the sum of the H strata totals T_{j_h}. That is, let $T_j = \sum_{h=1}^{H} T_{j_h}$ where T_{j_h} is the sum of the j^{th} characteristic for all units in the h^{th} stratum. (iii) For each h = 1, ..., H, obtain the optimum estimator $\hat{T}_{j_h}$ of T_{j_h} under assumption (i). By the theorem in Section A.1 and using the remarks following the theorem, we know that $\hat{T}_{j_h}$ will depend only on observations for the j^{th} characteristic in the h^{th} stratum, and will also be independent of any variances or covariances of the characteristics. (iv) Form the overall stratified estimator, say $\hat{T}_j^*$, of T_j by letting $\hat{T}_j^* = \sum_{h=1}^{H} \hat{T}_{j_h}$.

Since $\hat{T}_{j_h}$ is unbiased for T_{j_h}, it is easy to see that $\hat{T}_j^*$ is unbiased for T_j. Thus $\hat{T}_j^*$ is an unbiased, univariate estimator for T_j and does not depend on any variances or covariances.

It is of interest to compare the m.s.e. of $\hat{T}_j'$ to the m.s.e. of the stratified estimator $\hat{T}_j^*$ when the true structure of V is, say, V_t. These comparisons will then enable one to determine whether a simpler univariate estimator will outperform a more complex multivariate estimator in the face of uncertainty in the covariance structure of the characteristics.

The m.s.e. of $\hat{T}_j'$ when the true value of V(S) is $V_t(S)$ is determined as follows: Since $\hat{T}_j'$ is unbiased for T; that is, a change in the covariance structure does not affect the bias, we have

$$E(\hat{T}_j' - T_j)^2 = \mathrm{Var}(\hat{T}_j' - T_j) .$$

Using (A.1.1) and the fact that the units are assumed to be uncorrelated, it follows that

$$\mathrm{Var}(\hat{T}_j' - T_j) = \mathrm{Var}(\underline{\omega}'\ \hat{\underline{\beta}}_j - \Sigma_{i\varepsilon\bar{S}}\ y_{ij})$$

$$= \mathrm{Var}(\underline{\omega}'\ \hat{\underline{\beta}}_j) + \mathrm{Var}(\ \Sigma_{i\varepsilon\bar{S}}\ y_{ij})$$

$$= \underline{\omega}'\ \mathrm{Var}(\hat{\underline{\beta}}_j)\ \underline{\omega} + \mathrm{Var}\ \Sigma_{i\varepsilon\bar{S}}\ y_{ij} \qquad (A.2.1)$$

Now $\hat{\underline{\beta}}_j$ is obtained from (A.1.2) by assuming that $V(S) = V_a(S)$. If, in fact, the true value of $V(S)$ is given by $V_t(S)$, then we have

$$\mathrm{Var}\ \hat{\underline{\beta}} = [Z^*(S)'V_a^{-1}(S)Z^*(S)]^{-1}\ Z^*(S)'V_a^{-1}(S)V_t(S)V_a^{-1}(S)Z^*(S)$$

$$\times\ [Z^*(S)'V_a^{-1}(S)Z^*(S)]^{-1} \qquad (A.2.2)$$

The $(r\times r)$ matrix denoting $\mathrm{Var}(\hat{\beta}_j)$ is given as the j^{th} principal submatrix of (A.2.2).

The m.s.e. of the stratified estimator $\hat{T}_j^*$ when the true value of $V(S)$ is $V_t(S)$ is determined as follows: Since $\hat{T}_{j_h}$ is a univariate estimator of T_{j_h} for $h = 1, \ldots, H$, it is determined only by observations on the j^{th} characteristic within each stratum. Let S_h be that part of the overall sample S which consists of units in the h^{th} stratum., and let $\bar{S}_h$ denote the units in the population for the h^{th} stratum that are not in S_h. Then, for $h = 1, \ldots, H$, we have from (A.1.1) and (A.1.2) that

$$\hat{T}_{j_h} = \Sigma_{i\varepsilon S_h}\ y_{ij} + \Sigma_{i\varepsilon\bar{S}_h} \sum_{k=1}^{r} z_k\ (\underline{X}_i')\ \hat{\beta}_{jk_h}$$

$$= \Sigma_{i\varepsilon S_h}\ y_{ij} + \underline{\omega}_h'\ \hat{\underline{\beta}}_{j_h}$$

where

$$\hat{\underline{\beta}}_{j_h} = [Z(S_h)'\ Z(S_h)]^{-1}\ Z(S_h)'\ \underline{y}_j(S_h) \quad .$$

Since $\hat{T}_{j_h}$ is unbiased for T_{j_h}, we have for h = 1, ..., H, that if the true value of V(S) in the h^{th} stratum is $V_t(S_h)$, then

$$E(\hat{T}_{j_h} - T_{j_h})^2 = \mathrm{Var}(\hat{T}_{j_h} - T_{j_h}) = \mathrm{Var}(\underline{\omega}_h'\ \hat{\underline{\beta}}_{j_h}) + \mathrm{Var} \sum_{i\varepsilon\bar{S}_h} y_{ij}$$

$$= \underline{\omega}_h'[Z(S_h)'Z(S_h)]^{-1}\ Z(S_h)'\ V_{t_{jj}}(S_h)Z(S_h)[Z(S_h)'Z(S_h)]^{-1}\underline{\omega}_h$$

$$+ \mathrm{Var} \sum_{i\varepsilon\bar{S}_h} y_{ij} \qquad (A.2.3)$$

where $V_{t_{jj}}(S_h)$ is that particular submatrix of $V_t(S_h)$ applicable for the j^{th} characteristic. Now, using the fact that units in different strata are uncorrelated, it follows that

$$E[\hat{T}_j^* - T_j]^2 = E[\sum_{h=1}^{H} \hat{T}_{j_h} - \sum_{h=1}^{H} T_{j_h}]^2 = \sum_{h=1}^{H} E[\hat{T}_{j_h} - T_{j_h}]^2. \quad (A.2.4)$$

The objective is to evaluate the ratio of (A.2.1) to (A.2.4) for various V_a and V_t.

A.2.1 An Approach to the Solution

A population of 100 units is assumed. Associated with each unit are three characteristics (y_{i1}, y_{i2}, y_{i3}), i = 1, ..., 100 and one auxiliary variable X_i, i = 1, ..., 100. Two types of regression models are used. They are

Model I: $Y_{ij} = \beta_j + \varepsilon_{ij}$; i = 1, ..., 100; j = 1, 2, 3

and

<u>Model II</u>: $Y_{ij} = \beta_j X_i + \varepsilon_{ij}$; $i = 1, \ldots, 100$; $j = 1, 2, 3$.

For Model I we have $p = 3$, $q = 1$, $r = 1$, and

$$\underline{X}_i = X_i,\ i = 1, \ldots, 100$$
$$z_1(\underline{X}_i) = 1,\ i = 1, \ldots, 100$$
$$\underline{\beta}_j = \beta_j\ ,\ j = 1, 2, 3.$$

For Model II we have $p = 3$, $q = 1$, $r = 1$, and

$$\underline{X}_i = X_i\ ,\ i = 1, \ldots, 100$$
$$z_i(\underline{X}_i) = X_i\ ,\ i = 1, \ldots, 100$$
$$\underline{\beta}_j = \beta_j\ ,\ j = 1, 2, 3.$$

For Model II and for all covariance matrices described subsequently the X's range from 1.5 to 51.0 in increments of 0.5.

For each regression model, nine different covariance matrices are used. They are

<u>Group 1</u>

$$V_1 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i^{1/2}) \ ; \ i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = (0) \end{cases}$$

$$V_2 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i) \ ; \ i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = (0) \end{cases}$$

$$V_3 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i^2) \; ; \; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = (0) \end{cases}$$

<u>Group 2</u>

$$V_1 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i^{1/2}) ; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = \text{diag.}(1) \end{cases}$$

$$V_2 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i) \; ; \; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = \text{diag.}(1) \end{cases}$$

$$V_3 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i^2) \; ; \; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = \text{diag.}(1) \end{cases}$$

<u>Group 3</u>

$$V_1 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i^{1/2}) \; ; \; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = \text{diag.}(X_i^{1/2}-1) \; ; \; i = 1, \ldots, 100 \end{cases}$$

$$V_2 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i) \; ; \; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = \text{diag.}(X_i-1) \; ; \; i = 1, \ldots, 100 \end{cases}$$

$$V_3 \begin{cases} V_{11} = V_{22} = V_{33} = \text{diag.}(X_i^2) \; ; \; i = 1, \ldots, 100 \\ V_{12} = V_{13} = V_{23} = \text{diag.}(X_i^2 - 1) \; ; \; i = 1, \ldots, 100 \end{cases}$$

In using the stratified estimator $\hat{T}_j^*$, it is obviously impossible to satisfy assumption (i) in Section A.2 for these nine covariance structures. What is done for this case is to stratify the X's in such a way that regardless of which of the nine V's is used the maximum deviation in the variance of any characteristic within any stratum is no more than 25 percent of the total variation in the variance of that characteristic. By the nature of the V's this stratification will also give the same result for the variation in covariances of the characteristics. The criterion resulted in ten strata with the following strata boundaries:

1.5, 7.0, 13.5, 17.5, 25.5, 30.0, 32.0, 36.0, 38.5, 44.0, 51.0.

Since for each of the nine V's all three characteristics have the same variances and covariances, the m.s.e. of each characteristic will be the same. For simplicity we shall always let $j = 1$. Also note that the covariance structures for Group 1 are such that the conditions of the theorem in Section A.1 are met. Hence, $\hat{T}_j'$ is just a univariate estimator for these cases.

In the calculation of the m.s.e. (A.1.3) for $\hat{T}_1'$, the twenty units with largest X values are always the ones sampled. Regression Model I is assumed. Each of the three possible V's in Group 1 is selected as V_a and for each the m.s.e. is calculated for $V_t = V_1, V_2$, and V_3 in Group 1. This results in nine m.s.e.'s corresponding to Group 1. The same process is repeated for Group 2 and Group 3. The entire process is then repeated for regression Model II.

In the calculation of the m.s.e. (A.2.4), the two units with the largest X's in each of the ten strata are always selected. This also results in twenty units being sampled, but different units are chosen than for $\hat{T}_1'$. Regression Model I is assumed. The m.s.e. is calculated for $V_t = V_1$, V_2, and V_3 in Group 1. This results in three m.s.e.'s. The same process is repeated for Group 2 and Group 3. The entire process is then repeated for regression Model II.

A.2.2 Results. Let N_h and n_h be the number of population and sample units respectively for the h^{th} stratum, and let $v_{i1}(S_h)$ be the true variance of the first characteristic in the i^{th} sample unit for the h^{th} stratum. Then, for Model I it is easily shown that

$$\hat{T}_1^* = \sum_{h=1}^{H} \{ \sum_{i\varepsilon S_h} y_{i1}[1 + (N_h - n_h)/n_h]\} \text{ , and}$$

$$E(\hat{T}_1^* - T_1)^2 = \sum_{h=1}^{H} \{(\frac{N_h - n_h}{n_h})^2 \sum_{i\varepsilon S_h} v_{i1}(S_h) + \sum_{i\varepsilon \bar{S}_h} v_{i1}(S_h)\} \text{ .} \qquad (A.2.5)$$

For Model II, it is easily shown that

$$\hat{T}_1^* = \sum_{h=1}^{H} \{ \sum_{i\varepsilon S_h} y_{i1} + \sum_{i\varepsilon \bar{S}_h} X_i \sum_{i\varepsilon S_h} X_i y_{i1} / \sum_{i\varepsilon S_h} X_i^2\} \text{ , and}$$

$$E(\hat{T}_1^* - T_1)^2 = \sum_{h=1}^{H} \{(\sum_{i\varepsilon \bar{S}_h} X_i / \sum_{i\varepsilon S_h} X_i^2)^2 \sum_{i\varepsilon S_h} X_i^2 \, v_{i1}(S_h) + \sum_{i\varepsilon \bar{S}_h} v_{i1}(S_h)\}$$

(A.2.6)

Tables A.1 - A.3 give the ratios of the m.s.e. of $\hat{T}_1'$ to the m.s.e. of $\hat{T}_1^*$ for Model I and Groups 1, 2, and 3. The entries represent the ratio of (A.2.1) to (A.2.5)

Tables A.4 - A.6 give the same ratios for Model II and Groups 1, 2, and 3. The entries represent the ratio of (A.2.1) to (A.2.6).

All entries were obtained with the use of the IBM 360 computer.

TABLE A.1

Ratio of m.s.e. of $\hat{T}_1'$ to m.s.e. of $\hat{T}_1^*$ for

Model I

Group 1

V_a \ V_t	V_1	V_2	V_3
V_1	1.07	1.27	1.56
V_2	1.07	1.26	1.56
V_3	1.08	1.27	1.55

TABLE A.2

Ratio of m.s.e. of $\hat{T}_1'$ to m.s.e. of $\hat{T}_1^*$ for

Model I

Group 2

V_a \ V_t	V_1	V_2	V_3
V_1	1.07	1.27	1.56
V_2	1.07	1.26	1.56
V_3	1.08	1.27	1.55

TABLE A.3

Ratio of m.s.e. of $\hat{T}_1'$ to m.s.e. of $\hat{T}_1^*$ for

Model I

Group 3

V_a \ V_t	V_1	V_2	V_3
V_1	1.07	1.27	1.56
V_2	1.07	1.26	1.56
V_3	1.08	1.27	1.55

TABLE A.4

Ratio of m.s.e. of $\hat{T}_1'$ to m.s.e. of $\hat{T}_1^*$ for

Model II

Group 1

V_a \ V_t	V_1	V_2	V_3
V_1	0.43	0.45	0.48
V_2	0.43	0.45	0.48
V_3	0.43	0.45	0.48

TABLE A.5

Ratio of m.s.e. of $\hat{T}_1'$ to m.s.e. of $\hat{T}_1^*$ for

Model II

Group 2

$V_a \backslash V_t$	V_1	V_2	V_3
V_1	0.43	0.45	0.48
V_2	0.43	0.45	0.48
V_3	0.43	0.45	0.48

TABLE A.6

Ratio of m.s.e. of $\hat{T}_1'$ to m.s.e. of $\hat{T}_1^*$ for

Model II

Group 3

$V_a \backslash V_t$	V_1	V_2	V_3
V_1	0.43	0.45	0.48
V_2	0.43	0.45	0.48
V_3	0.43	0.45	0.48

Tables A.1 - A.3 show that for Model I and each of the three groups, the univariate estimator $\hat{T}_1^*$ has a smaller m.s.e. than the estimator $\hat{T}_1'$. Tables A.4 - A.6 imply that for Model II and each of the three groups, the univariate estimator has a larger m.s.e. than the estimator $\hat{T}_1'$.

Hence, (i) a univariate estimator may out-perform a multivariate one if the assumed covariance structure is incorrect. (ii) Even if the assumed covariance structure is correct; that is, $V_a = V_t$, a univariate estimator may be better <u>if different units are selected</u>.

Obviously, much additional work needs to be done before more conclusive results regarding the relative performances of univariate and multivariate estimators, when one is unsure of the covariance structure, can be reached.

A.3 A Second Problem Relating to the Model

Thus far in this chapter comparisons between multivariate and univariate estimators when one is uncertain of the covariance structure have been made. These comparisons were made under the assumption that the regression part of the model is correct. The situation can obviously be reversed. That is, one may be willing to accept a particular covariance structure but is unsure of the proper regression. Then it is of interest to compare the m.s.e. of the multivariate estimator derived under the <u>assumed</u> regression to the m.s.e. of a simpler univariate estimator when the <u>true</u> regression possibly differs from that assumed.

<u>A.3.1 An Approach to the Solution</u>. The problem can be illustrated by considering the following example: Let a sample of two units be selected and for each unit two characteristics and one auxiliary variable are observed. <u>Assume</u> the characteristics are generated according to the model

$$\left.\begin{aligned} Y_{i1} &= \beta_1 x_i + \varepsilon_{i1} \quad ; \quad i = 1,\ 2 \\ Y_{i2} &= \beta_2 x_i + \varepsilon_{i2} \quad ; \quad i = 1,\ 2 \end{aligned}\right\} \quad , \tag{A.3.1}$$

where $E(\varepsilon_{i1}) = E(\varepsilon_{i2}) = 0$; $i = 1, 2$. Let the covariance matrix V be

$$V(S) = \begin{pmatrix} \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} & \begin{pmatrix} x_1-1 & 0 \\ 0 & x_2-1 \end{pmatrix} \\ \begin{pmatrix} x_1-1 & 0 \\ 0 & x_2-1 \end{pmatrix} & \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} .$$

In the notation of Section A.1 we have

$$p = 2,\ q = 1,\ r = 1, \text{ and}$$

$$\underline{x}_i = x_i \ ; \ i = 1,\ 2$$

$$z_1(\underline{x}_i) = x_i \ ; \ i = 1,\ 2,$$

and

$$\underline{\beta}_j = \beta_j \ ; \ j = 1,\ 2 \ .$$

It is not too difficult to show that V(S) is positive definite for $x_1, x_2 > 2/3$. Considerable algebra gives the following matrices:

$$V^{-1}(S) = \begin{pmatrix} \begin{pmatrix} x_1/(2x_1-1) & 0 \\ 0 & x_2/(2x_2-1) \end{pmatrix} & \begin{pmatrix} (1-x_1)/(2x_1-1) & 0 \\ & (1-x_2)/(2x_2-1) \end{pmatrix} \\ \begin{pmatrix} (1-x_1)/(2x_1-1) & 0 \\ 0 & (1-x_2)(2x_2-1) \end{pmatrix} & \begin{pmatrix} x_1/(2x_1-1) & 0 \\ 0 & x_2/(2x_2-1) \end{pmatrix} \end{pmatrix}$$

and

$$[Z^*(S)'V^{-1}(S)Z^*(S)]^{-1} = \frac{1}{D}\begin{bmatrix} \sum_1^2 \frac{x_i^3}{2x_i-1} & \sum_1^2 \frac{x_i^2(x_i-1)}{2x_i-1} \\ \sum_1^2 \frac{x_i^2(x_i-1)}{2x_i-1} & \sum_1^2 \frac{x_i^3}{2x_i-1} \end{bmatrix}$$

where

$$D = 2\sum_1^2 \frac{x_i^2}{2x_i-1} \sum_1^2 \frac{x_i^3}{2x_i-1} - \left(\sum \frac{x_i^2}{2x_i-1}\right)^2 .$$

The expression for $\hat{\beta}_1$ is found from (A.1.2) to be

$$\hat{\beta}_1 = \frac{1}{D}\left\{ \left[\sum_1^2 \frac{x_i^3}{2x_i-1} \sum_1^2 \frac{y_{i1}x_i}{2x_i-1} + \sum_1^2 \frac{x_i^2}{2x_i-1} \sum_1^2 \frac{y_{i1}x_i^2}{2x_i-1} \right.\right.$$

$$\left. - \sum_1^2 \frac{x_i^2}{2x_i-1} \sum_1^2 \frac{y_{i1}x_i}{2x_i-1}\right] + \left[\sum_1^2 \frac{x_i^3}{2x_i-1} \sum_1^2 \frac{y_{i2}x_i}{2x_i-1}\right.$$

$$\left.\left. - \sum_1^2 \frac{x_i^2}{2x_i-1} \sum_1^2 \frac{y_{i2}x_i^2}{2x_i-1}\right]\right\} .$$

From (A.1.1) we have that the b.l.u.e. of T_1 is given by

$$\hat{T}_1 = \sum_1^2 y_{i1} + \hat{\beta}_1 \sum_{i\varepsilon\bar{S}} x_i .$$

We see that $\hat{T}_1$ is a multivariate estimator since it depends on y_{i1} and y_{i2}; $i = 1, 2$.

Now, suppose the true model is actually

$$\left.\begin{aligned} Y_{i1} &= \beta_1 x_i + \varepsilon_{i1} \; ; \; i = 1, 2 \\ Y_{i2} &= \beta_0 + \beta_2 x_i + \varepsilon_{i2} \; ; \; i = 1, 2 \end{aligned}\right\} \qquad \text{(A.3.2)}$$

with the same covariance structure. We then have that

$$E(\hat{\beta}_1) = \frac{\beta_0}{D}\left[\sum_1^2 \frac{x_i^3}{2x_i-1} \sum_1^2 \frac{x_i}{2x_i-1} - \left(\sum_1^2 \frac{x_i^2}{2x_i-1}\right)^2\right] + \beta_1$$

$$= \beta_0 A/D + \beta_1 .$$

So,

$$E(\hat{T}_1 - T_1) = E(\hat{\beta}_1 \sum_{i\in\bar{S}} x_i - \sum_{i\in\bar{S}} y_{i1})$$

$$= (\beta_0 A/D + \beta_1) \sum_{i\in\bar{S}} x_i - \beta_1 \sum_{i\in\bar{S}} x_i$$

$$= \beta_0 \sum_{i\in\bar{S}} x_i A/D .$$

Thus, $\hat{T}_1$ is biased for T_1 if the true regression model is (A.3.2).

In order to calculate the m.s.e. of T_1 under model (A.3.2), we use the general relation

$$E(\hat{T}_1 - T_1)^2 = [E(\hat{T}_1 - T_1)]^2 + \mathrm{Var}(\hat{T}_1 - T_1) .$$

Since the units in the sample are independent of those not in the sample, we have

$$E(\hat{T}_1 - T_1)^2 = [E(\hat{T}_1 - T_1)]^2 + \mathrm{Var}[(\sum_1^2 y_{i1} + \hat{\beta}_1 \sum_{i\in\bar{S}} x_i) - (\sum_1^2 y_{i1} + \sum_{i\in\bar{S}} y_{i1})]$$

$$= [E(\hat{T}_1 - T_1)]^2 + (\sum_{i\in\bar{S}} x_i)^2 \mathrm{Var}\, \hat{\beta}_1 + \mathrm{Var}(\sum_{i\in\bar{S}} y_{i1}) .$$

The value of Var $\hat{\beta}_1$ is given as the (1,1) element of $[Z^*(S)'V^{-1}(S)Z^*(S)]^{-1}$. Hence, the m.s.e. is

$$E(\hat{T}_1-T_1)^2 = \beta_0^2(A/D)^2(\sum_{i\varepsilon\bar{S}} x_i)^2 + (\sum_{i\varepsilon\bar{S}} x_i)^2 \sum_1^2 \frac{x_i^3}{2x_i-1}/D$$

$$+ \sum_{i\varepsilon\bar{S}} x_i \; . \qquad (A.3.3)$$

A simple univariate estimator of T_1 is the ratio estimator given by

$$\hat{T}_1' = (\sum_1^N x_i/\sum_1^2 x_i) \sum_1^2 y_{i1} \; .$$

We can write the estimator in the form

$$\hat{T}_1' = \sum_1^2 y_{i1} + \sum_1^2 y_{i1} \sum_{i\varepsilon\bar{S}} x_i/\sum_1^2 x_i \quad .$$

The bias is given by (with respect to model (A.3.2))

$$E(\hat{T}_1'-T_1) = E[\sum_1^2 y_{i1} \sum_{i\varepsilon\bar{S}} x_i/\sum_1^2 x_i - \sum_{i\varepsilon\bar{S}} y_{i1}]$$

$$= \beta_1[\sum_1^2 x_i \sum_{i\varepsilon\bar{S}} x_i/\sum_1^2 x_i - \sum_{i\varepsilon\bar{S}} x_i]$$

$$= 0 \quad .$$

Hence, $\hat{T}_1'$ is unbiased for T_1 under model (A.3.2).

The m.s.e. of $\hat{T}_1'$ is given by

$$E(\hat{T}_1'-T_1)^2 = Var(\hat{T}_1'-T_1) = Var[\sum_1^2 y_{i1} \sum_{i\varepsilon\bar{S}} x_i/(\sum_1^2 x_i) - \sum_{i\varepsilon\bar{S}} y_{i1}]$$

$$= \sum_1^2 x_i(\sum_{i\varepsilon\bar{S}} x_i/\sum_1^2 x_i)^2 + \sum_{i\varepsilon\bar{S}} x_i$$

$$= (\sum_{i\in\bar{S}} x_i)^2 / \sum_1^2 x_i + \sum_{i\in\bar{S}} x_i \quad . \tag{A.3.4}$$

Obviously, (A.3.3) could exceed (A.3.4) for certain values of β_0.

Now, suppose the <u>true</u> model is

$$\left.\begin{aligned} Y_{i1} &= \beta_0 + \beta_1 x_i + \varepsilon_i \; ; \; i = 1, 2 \\ Y_{i2} &= \qquad \beta_2 x_i + \varepsilon_i \; ; \; i = 1, 2 \end{aligned}\right\} \tag{A.3.5}$$

with the same covariance structure. We then have that

$$E(\hat{\beta}_1) = \frac{\beta_0}{D} \left[\sum_1^2 \frac{x_i^3}{2x_i-1} \sum_1^2 \frac{x_i}{2x_i-1} + \left(\sum_1^2 \frac{x_i^2}{2x_i-1}\right)^2 - \sum_1^2 \frac{x_i^2}{2x_i-1} \sum_1^2 \frac{x_i}{2x_i-1}\right] + \beta_1 \; .$$

$$= \beta_0 \, B/D + \beta_1 \quad .$$

So,

$$E(\hat{T}_1 - T_1) = E(\hat{\beta}_1 \sum_{i\in\bar{S}} x_i - \sum_{i\in\bar{S}} y_{i1})$$

$$= (\beta_0 \, B/D + \beta_1) \sum_{i\in\bar{S}} x_i - \sum_{i\in\bar{S}} (\beta_0 + \beta_1 \, x_i)$$

$$= \beta_0 \, [B \sum_{i\in\bar{S}} x_i/D - (N-2)] \; .$$

The m.s.e. of $\hat{T}_1$ under model (A.3.5) is given by

$$E(\hat{T}_1 - T_1)^2 = [E(\hat{T}_1 - T_1)]^2 + \mathrm{Var}(\hat{T}_1 - T_1)$$

$$= \beta_0^2 [B \sum_{i\in\bar{S}} x_i/D - (N-2)]^2 + (\sum_{i\in\bar{S}} x_i)^2 \sum_1^2 \frac{x_i^3}{2x_i-1} / D$$

$$+ \sum_{i\in\bar{S}} x_i \quad . \tag{A.3.6}$$

The bias of the ratio estimator $\hat{T}_1'$ under model (A.3.5) is

$$\begin{aligned} E(\hat{T}_1'-T_1) &= \beta_0[2 \sum_{i\varepsilon\bar{S}} x_i/\sum_1^2 x_i - (N-2)] \\ &+ \beta_1[\sum_1^2 x_i \sum_{i\varepsilon\bar{S}} x_i/\sum_1^2 x_i - \sum_{i\varepsilon\bar{S}} x_i] \\ &= \beta_0 \, [2 \sum_{i\varepsilon\bar{S}} x_i - (N-2) \sum_1^2 x_i]/\sum_1^2 x_i \\ &= \beta_0[2 \sum_1^N x_i - N \sum_1^2 x_i]/\sum_1^2 x_i \, . \end{aligned}$$

Thus, $\hat{T}_1'$ is unbiased under model (A.3.5) if

$$\sum_1^N x_i/N = \sum_1^2 x_i/2 \quad ,$$

or, in other words, if the sample of the two units is balanced on the x's.

Assuming the two sample units are balanced,the m.s.e. of $\hat{T}_1'$ becomes

$$E(\hat{T}_1'-T_1)^2 = \mathrm{Var}(\hat{T}_1'-T_1) = (\sum_{i\varepsilon\bar{S}} x_i)^2/\sum_1^2 x_i + \sum_{i\varepsilon\bar{S}} x_i \, . \qquad \text{(A.3.7)}$$

Clearly, for certain values of β_0 (A.3.6) could exceed (A.3.7).

In summary, the example illustrates the fact that a univariate estimator may out-perform a multivariate one if the assumed regression is in error. Also, the univariate ratio estimator can be made unbiased for alternative regression models by balancing on the x's. Actually, this latter result is stated in a more general form in the strictly univariate case in Royall and Herson, [15]. They examine the behavior of the expansion estimator, $\hat{T} = N \sum_S y_i/n$, for a whole class of regression models. It is easy to see that in the case of balancing the ratio estimator $\hat{T}_1'$ reduces to the expansion estimator.

As in the case of uncertainty of the covariance matrix, there is a need for further research in the comparisons of multivariate and univariate estimators when one is uncertain of the proper regression model.

A.4 A General Multivariate Estimator for a Particular Model

Let

$$Y_{ij} = \beta_j x_i + \varepsilon_{ij} \; ; \; i = 1, \ldots, N$$
$$j = 1, \ldots, p \; ,$$

where $E(\varepsilon_{ij}) = 0$ for all i,j. A sample of n units is selected, and let the covariance structure of the characteristics in the sample be as follows:

$$V(S) = \begin{pmatrix} V_{11} & V_{12} & \cdots & V_{1p} \\ & V_{22} & \cdots & V_{2p} \\ \text{symm.} & & \ddots & \\ & & & V_{pp} \end{pmatrix}$$

where the (n×n) submatrices are defined to be

$$V_{jj}(\ell,m) = \text{diag}(x_\ell); \; \ell = 1, \ldots, n$$
$$j = 1, \ldots, p$$

and

$$V_{jk}(\ell,m) = \text{diag}(1); \; \ell = 1, \ldots, n$$
$$j,k = 1, \ldots, p \; ; \; j \neq k \; .$$

Note that the correlation of any two characteristics in the i^{th} unit is given by $\rho_i = 1/x_i$; $i = 1, \ldots, n$. So, the correlation decreases as x increases. Greenstreet [5] mentions a practical situation in which this type of behavior for ρ_i may be feasible.

After much algebra the inverse of V(S) is found to be

$$V^{-1}(S) = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ & A_{22} & \cdots & A_{2p} \\ & & \ddots & \\ & & & A_{pp} \end{pmatrix}$$

where each submatrix is of dimension (n×n). Letting

$$[x_\ell] = (p - 2 + x_\ell)\, x_\ell - p + 1 \; ,$$

the submatrices are given by

$$A_{jj}(\ell,m) = \mathrm{diag}[(p-2+x_\ell)/[x_\ell]]; \; j = 1, \ldots, p \qquad \ell = 1, \ldots, n$$

and

$$A_{jk}(\ell,m) = \mathrm{diag}[-1/[x_\ell]]; \; j, k = 1, \ldots, p; \; j \neq k \qquad \ell = 1, \ldots, n \; .$$

Again after much algebra, the inverse of the (p×p) matrix $Z^*(S)'V^{-1}(S)Z^*(S)$ is found to be

$$[Z^*(S)'V^{-1}(S)Z^*(S)]^{-1} = \frac{1}{D} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ & b_{22} & \cdots & b_{2p} \\ & & \ddots & \\ & & & b_{pp} \end{pmatrix}$$

where

$$b_{jj} = \sum_{i\varepsilon S} x_i^3/[x_i]; \; j = 1, \ldots, p \;,$$

$$b_{jk} = \sum_{i\varepsilon S} x_i^2/[x_i]; \; j,\, k = 1, \ldots, p,\, j \neq k \;,$$

and

$$D = [(\sum_{i\varepsilon S} x_i^3/[x_i])^2 + (p-2) \sum_{i\varepsilon S} x_i^2/[x_i] \sum_{i\varepsilon S} x_i^3/[x_i]$$

$$- (p-1)(\sum_{i\varepsilon S} x_i^2/[x_i])^2] \;.$$

The estimator $\hat{\beta}_1$ is derived from (A.1.2) and is given by

$$D\hat{\beta}_1 = \sum_{j\varepsilon S} y_{j1} \{\frac{x_j(p-2+x_j)}{[x_j]} \sum_{i\varepsilon S} x_i^3/[x_i] - \frac{x_j(p-1)}{[x_j]} \sum_{i\varepsilon S} x_i^2/[x_i]\}$$

$$+ \sum_{j\varepsilon S} \sum_{k=2}^{p} y_{jk} \{-\frac{x_j}{[x_j]} \sum_{i\varepsilon S} x_i^3/[x_i] + \frac{x_j^2}{[x_j]} \sum_{i\varepsilon S} x_i^2/[x_i]\} \;.$$

The optimum estimator $\hat{T}_1$ is then

$$\hat{T}_1 = \sum_{i\varepsilon S} y_{i1} + \hat{\beta}_1 \sum_{i\varepsilon \bar{S}} x_i \;.$$

The vita has been removed from the scanned document

THE USE OF AUXILIARY INFORMATION IN THE LINEAR LEAST-SQUARES PREDICTION APPROACH TO CLUSTER SAMPLING IN A FINITE POPULATION

by

Ragan Burt Madden

(ABSTRACT)

Linear least-squares prediction methods are applied to cluster (two-stage) sampling problems in a finite population where auxiliary information is available. Two regression models which describe the behavior of the second-stage units and which utilize the auxiliary information are considered. For one model the optimum estimator of the total of the second-stage units and its mean square error (m.s.e.) are derived. The selection of clusters which minimize the m.s.e. are determined for certain cases. For both models a conventional estimator of the total is analyzed in the prediction theory framework. Optimum sampling designs for the conventional estimator are obtained for certain parameter configurations. A computer implemented study to compare the performances of the estimators for a wide range of parameter values is done. A practical problem is analyzed.