STUDIES ON SOME ESTIMATORS IN DOUBLE SAMPLING

GITASREE DAS

DEPARTMENT OF ECONOMICS

SCHOOL OF SOCIAL SCIENCES

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENT OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

NORTH-EASTERN HILL UNIVERSITY

SHILLONG, INDIA

AUGUST, 1991

PHONE :
GRAMS : NEHU

North-Eastern Hill University

Mayurbhanj Complex, Nongthymmai, Shillong-793014

Department of Economics

It is my pleasure to certify that MS. GITASREE DAS has done original work on her thesis entitled "STUDIES ON SOME ESTIMATORS IN DOUBLE SAMPLING" for the Degree of Doctor of Philosophy of the North-Eastern Hill University, Shillong, which has neither been submitted anywhere for publication nor has been submitted for any Degree of any other University. I recommend her thesis for the award of Ph.D. degree.

Date : August 20, 1991.

Place : Shillong.

(Prof. K. Bez)

Supervisor

ACKNOWLEDGEMENT

I express my deep sense of gratitude to Professor K. Bez, for his guidance and encouragement given during the course of this work.

I am grateful to the librarians and other staff of Delhi University Library, Indian Statistical Institute Library, Calcutta, Gauhati University Library and Dibrugarh University Library, for their kind cooperation in collecting literatures from their library.

I am thankful to Dr. Asit Baran Chakraborty, Lecturer, St. Anthony's College, Shillong, for his suggestions and encouragement during the tenure of this work. My thanks are also due to Dr. Ashesh Kumar Das, Research Associate, UGC-DRS programme, Department of Botany, NEHU, Shillong, for introducing me with various quantitative work in their field which gave me a new insight into applications of sampling techniques. I am also thankful to all my colleagues at the Centre for Applied Statistics and Department of Economics, NEHU, Shillong, for their help and encouragement.

My sincere thanks are also due to my revered teachers at the Department of Statistics, University of Delhi, whose sincere contribution in my training in the field has led me to take up this work.

I owe my deep sense of gratitude to my husband Mr. M.H. Barbhuiya, Lecturer, Centre for Creative Arts, NEHU, Shillong and my daughter Ms. Rikhia, whose constant encouragement and sincere cooperation has helped me to complete this work. Mr. Barbhuiya has also helped in drawing the figures. I am also thankful to my parents, sisters and brothers for their encouragement to complete this work.

Finally, I wish to place on record my sincere gratitude to Mr. Cliff Robert Dohling, for typing the manuscript with great patience and sincere effort.

Gitasree Das
GITASREE DAS

Date : August, 1991.
Place : Shillong.

C O N T E N T S

....

CHAPTER ONE

Introduction

1.1 The double sampling

One of the most important developments in sampling theory, which has wide scope of applicability in various statistical enquiries and investigations, has been to study the various possible ways of utilizing auxiliary information for estimating a finite population parameter. There exist number of sampling techniques, wherein availability of advance information on one or more auxiliary variables leads to greater precision. To name a few, ratio and regression estimates require a knowledge of the population mean $\bar{X}$. If it is desired to stratify the population according to the values of x_i, their frequency distribution must be known. However, when such information is lacking it is sometimes relatively inexpensive to take a large preliminary sample in which the auxiliary variable X alone is measured. The purpose of this sample is to furnish a good estimate of $\bar{X}$, the population mean of X or the frequency distribution of X. This technique is known as double sampling or two-phase sampling.

1.2 Double sampling for regression estimation

In some applications of double sampling the auxiliary variate x_i has been used to make a regression estimate of $\bar{Y}$. In the first (large) sample of size n', we measure only x_i; in the second a sub sample of size n, we measure both x_i and y_i. The estimate of $\bar{Y}$ is

$$\bar{y}_{lr} = \bar{y} + b\,(\bar{x}' - \bar{x})$$

where $\bar{x}'$, $\bar{x}$ are the means of x_i in the first and second sample, $\bar{y}$ is the mean of y_i and b is the least squares regression coefficient of y_i on x_i, computed from the second sample. Assuming random sampling and $\frac{1}{n}$ and $\frac{1}{n'}$ negligible with respect to 1,

$$V(\bar{y}_{lr}) = \frac{S_y^2\,(1 - \rho^2)}{n} + \frac{\rho^2\, S_y^2}{n'} - \frac{S_y^2}{N}$$

[From Cochran (1977)]

Royall (1970) considered the problems of estimating totals in finite populations, when auxiliary information regarding variate values is available, under some linear regression, 'super population' models. Optimal strater-gies involving linear estimators are derived under certain variance assumptions and compared under various assumptions.

Searls (1964) after utilising prior information about the coefficient of variation of the character under study, suggested an estimator for the population mean which has smaller mean square error than the conventional unbiased estimator, in case of simple random sampling with replacement.

Das and Tripathi (1980) have suggested an estimator similar to one suggested by Searls (1964) which is based on the prior knowledge of the coefficient of kurtosis of the character under study or its guess value obtained from the past experience and have shown that their estimator under symmetrical population model with SRSWOR is superior to the Searls (1964) estimator if the value of coefficient of kurtosis lies within certain limits.

Bedi and Hajela (1984) have attempted to improve over Searls estimator after utilizing the information about the auxiliary variate at the estimation stage by making use of (i) SRSWOR and (ii) Two-phase sampling.

We find an early note on double sampling by Bose (1943). She discussed there that the sample mean value of the character Y as an estimate of the population mean value can be estimated with some margin of error from the knowledge of a second character X provided there is significant correlation between the two characters X and Y. This problem was looked upon as a procedure of double sampling involving (1) the estimation of the regression equation from the first or 'exploratory stage' of sampling and (2) the estimation of the mean value of the character X from the second or 'survey stage'.

Tripathi (1969) presented a regression type estimator for Y in pps wr sampling when information on Z is readily available and the population total of a character X is known. Also he (1970, 1973) developed a unified technique of double sampling for regression method of estimation and extended the above results to the situations:

(a) information on Z is available but population total of X is not known, and

(b) neither the information on Z is readily available nor the population total of X is known.

Tamhane (1952) considered the problem of hypothesis testing using regression estimator in double sampling. Test procedures are provided when the covariance matrix between the primary and auxiliary variables is either partially known or completely unknown. For the latter case a new 'stutentized' version of the regression estimator is proposed as a test statistic. The exact null distribution of this statistic is derived in a special case. Khatri, Bhargava and Shah (1974) also derived the exact distribution of a certain 'studentized' version of the regression estimator. However, the exact distribution derived by Khatri et al is very complicated and depends on ρ, the correlation coefficient between X and Y.

1.3 Double sampling for ratio estimation

If the first sample is used to obtain $\bar{x}'$ as an estimate of $\bar{X}$ in a ratio estimate of $\bar{Y}$, the estimator of $\bar{Y}$ is

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{x}',$$

and

$$V(\bar{y}_R) = \frac{1}{n} (S_y^2 - 2 R S_{yx} + R^2 S_x^2)$$

$$+ \frac{1}{n'} (2 R S_{yx} - R^2 S_x^2) - \frac{S_y^2}{N}$$

[From Cochran (1977)]

Rao (1975) has suggested three unbiased ratio estimators of the Hartley - Ross (1954) type for the case in which samples at the two phases are drawn without replacement and independent of each other. The estimators are formed by estimating the population mean of the auxiliary characteristic by the mean of the first sample $\bar{x}_1$, the mean of the distinct units in the two samples $\bar{x}_w$ and by pooling the means of the two samples

$$\bar{x}^* = a\bar{x}_1 + b\bar{x}, \quad a + b = 1$$

They are as follows :

$$t = \bar{r} (\bar{x}_1 - \bar{x}) + M (\bar{y} - \bar{r}\,\bar{x}) + \bar{y}$$

where

$r_i = y_i/x_i$, $\bar{r} = \sum r_i/n$ and $M = (N - n)/N(n-1)$

$$t_w = \bar{r}\,(\bar{x}_w - \bar{x}) + M_w(\bar{y} - \bar{r}\,\bar{x}) + \bar{y}$$

where $M_w = (w - n)/w(n - 1)$

$$t^* = \bar{r}\,(\bar{x}^* - \bar{x}) + a\,M\,(\bar{y} - \bar{r}\,\bar{x}) + \bar{y}$$

It is shown that the latter two methods result in gains in efficiency for the ratio estimator. Sizes of the population and samples are considered to be finite.

Sukhatme (1962) also constructed a Hartley-Ross (1954) type of estimator and compared it with the classical estimator.

Rao (1972) dealing with one auxiliary variable has suggested two estimates i.e. $\bar{x}^*$ (the best linear combination of the two independent samples) and $\bar{x}_w$ (mean based on w distinct units in two independent samples) for $\bar{X}$. He has further shown that the efficiency of the regression estimator of $\bar{Y}$ will increase when $\bar{X}$ is estimated by $\bar{x}_w$ instead of $\bar{x}^*$ or $\bar{x}'$.

Rao (1975a) proposed modified ratio estimators for the case in which the first - and second - phase samples are drawn without replacement and independent of each other. They are given as follows:

$$r_w = (\bar{y}\ /\ \bar{x})\ \bar{x}_w$$

and $$r^* = (\bar{y}\ /\ \bar{x})\ \bar{x}^*$$

where $\bar{x}_w$ is the mean of the w distinct units in the two independent samples and $\bar{x}^* = (a\bar{x}_1 + b\bar{x})$, $a + b = 1$, $\bar{x}_1$, $\bar{x}$ being the means of first and second samples respectively.

Conditions for these estimators being more efficient than the classical estimators were investigated. Two methods for evaluating the expectation of the reciprocal of the number of distinct units were also presented.

Rao (1981) presented nine two-phase ratio estimators for the mean of a finite population. Assuming a linear model, exact expressions for their biases and mean squared errors are derived. Then the nine estimators are compared as to bias and mean squared error. The list of estimators includes the Jack knife estimator and four modifications; two of the proposed modifications are implemented by first expressing the classical estimator in the form of a regression estimator.

Srivastava (1970) considered a two-phase sampling estimator of the ratio-type for estimating the mean $\bar{Y}$ of a finite population, where the value of $\rho C_y/C_x$ can be guessed or estimated in advance. Here C_y and C_x denote respectively the coefficients of variation of the characteristic under study, Y, and the auxiliary characteristic X and ρ denotes the coefficient of correlation between Y and X. When the value of $\rho C_y/C_x$ is guessed or estimated exactly, the estimator has a smaller large-sample variance compared with either an ordinary ratio or linear regression estimator in two-phase sampling in the case where the first-phase sample is drawn independently from the second-phase sample. If the sample at the second phase is a subsample of the first-phase sample, the estimator has variance equal to that of the linear regression estimator. The largest value of the difference between the assumed value and the actual value of $\rho C_y/C_x$ has been obtained so as not to result in the variance of the estimator being larger than the variances of either an ordinary ratio estimator or an ordinary linear regression estimator.

Singh (1982) has proposed generalized estimators for the estimation of ratio and product of population parameters, of which the estimators given by Singh (1965) are special cases. Following are the proposed generalized estimators in double sampling for estimating R and P:

? to be defined?

$$\hat{R}^*_{3d} = \hat{R}\ f\ (u) \qquad \text{and}$$

$$\hat{P}^*_{3d} = \hat{P}\ f\ (u)$$

respectively, where

$$u = \frac{\bar{x}_2}{\bar{x}'_2}$$

and f(u) is a function of u such that f(1) = 1, satisfying the following conditions:

1. Whatever be the sample chosen, u assumes values in a bounded, closed interval I of the real line con-

taining the point unity.

2. In I, the function f(u) is continuous and bounded.

3. The first, second and third partial derivatives of f(u) exist and are continuous and bounded in I.

Bias and mean square error of the proposed estimators were found and a comparative study has been made with the double sampling estimators considered by Singh (1965).

The regression estimator and the ratio estimator are commonly used in survey practice. In the past, more attention has been given to the ratio estimator because of its computational ease and applicability for general sampling design. The ratio estimator is appropriate for populations whose regression line passes close to the origin. If the intercept of the regression line is significantly non zero, however, it is much less efficient than the regression estimator (Deng 1984). In general, apart from n^{-2} terms, the mean squared error of the former is bigger than the latter (Cochran 1977. p.196). Given the present computing capacity, the computational advantage of the ratio estimator should be less of a concern and the regression estimator will gain wider popularity. Deng and Wu (1987) provided a theoritical and empirical comparison of several variance estimators for the regression estimator in simple random sampling without replacement. Under comparison are several design-based and model-based estimators and a new class of estimators. Their second-order expressions and biases are derived and compared.

1.4 Double sampling for two auxiliary variables

Kiregyera (1980) proposed a chain ratio-type estimator using two auxiliary variables in double sampling. The performance of the constructed estimator was compared with the simple mean, ratio type estimate based on double sampling and with ratio type estimator given by Chand (1975).

Kiregyera (1984) used two auxiliary variables x and z to construct two regression-type estimators for the population mean of the study variable Y.

Ratio-in-Regression Estimator

$$t'_{21r} = \bar{y}_n + b_2 \left[\frac{\bar{x}_{n'}}{\bar{z}_{n'}} \bar{z}_N - \bar{x}_n \right]$$

where b_2 is the estimate of the regression coefficient of y on x.

The approximate expressions for bias and MSE of t'_{21r} to $O(1/n)$ and for N large are as follows :

$$\text{Bias}\ (t'_{21r}) = \bar{y}_N\ \alpha_1 \left[\frac{1}{n} \alpha_2 + \frac{1}{n'} \alpha_3 \right]$$

and

$$MSE\ (t'_{21r}) = \bar{y}_N^2 \left[\frac{1}{n} \alpha_4 + \frac{1}{n'} \alpha_5 \right]$$

where

$$\alpha_1 = C_{110} / C_{200}$$

$$\alpha_2 = C_{300} / C_{200} - C_{210} / C_{110}$$

$$\alpha_3 = C_{210} / C_{110} - C_{111} / C_{110} - C_{300} / C_{200}$$

$$+ C_{201} / C_{200} + C_{002} - C_{101}$$

$$\alpha_4 = C_{020} - C_{110}^2 / C_{200}$$

$$\alpha_5 = \alpha_1 \ (C_{110}^2 C_{002} / C_{200} - 2 C_{110} C_{011})$$

$$C_{rst} = \frac{E\ (x_i - \bar{x}_N)^r \ (y_i - \bar{y}_N)^s \ (z_i - \bar{z}_N)^t}{\bar{x}_N^r \ \bar{y}_N^s \ \bar{z}_N^t}$$

$$r \geqslant 0,\ s \geqslant 0,\ t \geqslant 0$$

Regression-in-Regression Estimator

$$t_{2RR} = \bar{y}_n + b_2 \left[(\bar{x}_{n'} - \bar{x}_n) - b_1 \ (\bar{z}_{n'} - \bar{z}_N) \right]$$

where b_1 is the estimate of the regression coefficient of x on z.

The approximate expressions for bias and MSE of t_{2RR} to $O(1/n)$ and for large N are :

$$\text{Bias }(t_{2RR}) = \bar{y}_N \; \alpha_1 \left[K \alpha_2 + \frac{1}{n'} \alpha_8 \; \alpha_9 \right]$$

$$\text{MSE }(t_{2RR}) = \bar{y}_N^2 \left[\frac{1}{n} C_{020} - K \alpha_1 \, C_{110} + \frac{1}{n'} \alpha_7 \, \alpha_8 \right]$$

where

$$K = \frac{n' - n}{n \; n'}$$

$$\alpha_8 = C_{101} \;/\; C_{002}$$

$$\alpha_9 = C_{003} \;/\; C_{200} + C_{201} \;/\; C_{200} - C_{012} \;/\; C_{001} - C_{102} \;/\; C_{101}$$

The efficiency of the proposed estimators is investigated under a super-population model. A numerical study is done to demonstrate the practical use of different estimation formulae and emprically demonstrate the performance of the constructed estimators.

Singh (1984) studied the various modes of sampling in double-sampling with two auxiliary characters and examined the efficiency of these selection procedures when two auxiliary variables are used. The three possible sampling schemes considered are :

1. A preliminary sample of size n' is selected by SRSWOR for observing the auxiliary characters x_1, x_2 and a

smaller sub-sample of size n for observing y is sellected from n'.

2. The preliminary sample is selected as in 1 above but the smaller of a size n is selected independently.

3. Many times the auxiliary information may be collected by two different agencies and hence two independent preliminary samples of size n_1' and n_2' are selected for observing x_1 and x_2 and the small sample of size n is also selected independently from the population by SRSWOR.

Mukherjee, Rao and Vijayan (1987) have considered a practical situation where information on two auxiliary variables related to the study variable is available at different levels. Following Kiregyera (1980, 1984) they studied several estimators and compared them under mean square error criterion. They have also extended these results to the case when multiple auxiliary information ia available.

1.5 Double sampling for multiple auxiliary variables

Double sampling with regression has been extended by Khan and Tripathi (1967) to the case where p auxiliary X variables are measured in the second sample, $\overline{Y}$ being estimated by the multiple linear regression of Y on these variables. With the second sample a random subsample of the first and with multivariate normality assumed for Y and the X's, the extension for $p > 1$ gives for the average variance

$$V(\bar{y}_{1r}) = \frac{S_y^2 (1 - R^2)}{n} \left[1 + \frac{n' - n}{n'} \frac{p}{(n - p - 2)} \right] + \frac{R^2 S_y^2}{n'} - \frac{S_y^2}{N}$$

where R is the multiple correlation coefficient between Y and the X's.

Tripathi (1968) has proposed multivariate ratio and difference estimators when information on Z is readily available and the population totals $(X_1, X_2, \ldots\ldots, X_p)$ are known. Further the author (1976) considered the problem of estimating the population total of Y, in case the units are to be selected with pps wr and the information on yet a p-dimensional vector $X = (X_1, \ldots\ldots, X_p)$ of auxiliary characters is to be used to form multivariate difference-type and ratio-type estimators. A general double sampling scheme is developed for multivariate difference and ratio-type estimators and then the general results are used to derive mean and mean square errors of these estimators in two particular sampling schemes.

(i) First sample is a PPSWR sample but second a SRSWOR (the situation where information on the selection variable z is available but $\underset{\sim}{X}$ is unknown) and

(ii) first sample is a SRSWOR but second a PPSWR (the situation where information on neither of Z and $\underset{\sim}{X}$ is available) taking motivation from Raj (1964, 1965) and Singh and Singh (1965).

Expressions for optimum sizes of first phase and second phase samples and the resulting optimum mean square errors are obtained and comparision of the proposed estimators is made with other estimators.

Srivastava (1981) has proposed a generalized two-phase sampling estimator. It has been shown that the asymptotic minimum variance for any estimator of the class is equal to that of the conventional linear regression estimator for the case of two-phase sampling when the second phase sample is a subsample of the first phase sample. For the case when the two samples are drawn independently, an explanation is given for the lower value of the minimum variance of the proposed class of estimators than that of the conventional linear regression estimator, as also obtained by Srivastava (1970) and Gupta (1978).

Bedi (1985) has shown that, in two-phase sampling when the two samples are drawn independently, the suggested multivariate regression estimator and generalized two-phase estimator have smaller mean square error than the corresponding usual multivariate regression estimator and Srivastava's (1981) estimator. When the coefficients of the proposed estimators are estimated, expected mean square error under a suitable model are also derived.

Singh and Namjoshi (1988) proposed a class of multivariate regression estimators in two-phase sampling for estimating population mean $\bar{Y}$ of the study character Y. The exact expression for its mean squared error (MSE) is obtained and optimum estimator of the class is also investigated. It is shown that the minimum MSE of the proposed class is smaller than the one reported by Bedi (1985).

Tripathi (1988) gives a general theory of double sampling for estimating a parameter θ_0 utilising the information on a supplementary parameter θ_1, obtained inexpensively through a preliminary large sample. A wide class of estimators is discussed, an asymptotically optimum subclass is identified and the estimators based on two phase sampling are compared, with usual unbiased estimator of σ_y^2 in case of single phase sampling, under a linear cost function. In particular the results are derived for bivariate normal populations.

Tripathi (1989) obtained a class of estimators, based on general double sampling schemes for estimating the population mean $\bar{Y}$ in case information on an auxiliary character X, expected to increase precision of estimation, is not available a priori. It is shown that the usual ratio, regression and product estimators in double sampling may always be improved in case correlation between Y and X^2 is appreciable and/or marginal distribution of X is skewed.

1.6 Double sampling for stratification

Let the first sample be a random sample be of size n'. Let

$W_h = N_h / N$ = proportion of population falling in stratum h.

$w_h = n'_h / n$ = proportion of first sample falling in stratum h.

Then w_h is an unbiased estimate of W_h

The second sample is a stratified random sample of size n in which the y_{hi} are measured: n_h units are drawn from stratum h. Usually the second sample in stratum h is a random subsample from the n'_h in the stratum. The objective of the first sample is to estimate the strata weights; that of the second sample is to estimate the strata means $\bar{Y}_h$.

The population mean $\bar{Y} = \sum W_h \bar{Y}_h$. As an estimate we use

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h$$

and

$$V(\bar{y}_{st}) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \sum_{h=1}^{L} \frac{W_h S_h^2}{n'}\left(\frac{n'_h}{n_h} - 1\right)$$

where S^2 is the population variance

[From Cochran (1977)]

Singh and Singh (1965) discussed about double sampling for stratification. Suppose a population of size N is to be stratified according to the values of an auxiliary variable X, but the frequency distribution of X is unknown. A large sample $s(n')$ of size n' was taken by simple random sampling without replacement and only X was observed. The selected units were classified into L strata according to X. Let n'_h denote the number of units in $s(n')$ falling into stratum h ($h = 1, \ldots L$; $\sum n'_h = n'$). A subsample $s(n_h)$ of size n_h is drawn from $s(n'_h)$ by simple random sampling without replacement independently for each h, and the character of interest y is observed.

Further, Singh and Singh (1965) pointed out that n_h is bounded above by the random variable n'_h, which varies from 0 to $\min.(n', N_h)$, where N_h is the total number of units in stratum h. They proposed three procedures which are free of inconsistency :

(i) the $s(n_h)$ are selected with replacement, and all units are used in the estimators;

(ii) as in (i), but with only distinct units used;

(iii) sub-sampling is without replacement, the size being $\min (n'_h, n_h)$.

Rao (1973) proposed a simple method of double sampling for stratification and ~~the~~ classical non-response theory is obtained as a special case. The method leads to simple solutions for the optimal design of analytical surveys involving comparison of group means, when the groups are not identifiable in advance.

1.7 Other uses of double sampling

Raj (1964) suggested the use of double sampling for pps estimation. He considered the problem of selecting the sample with probability proportional to X, but information on X is not available. This information was then collected from an initial sample (simple random) of size n' from which a sub-sample of size n is selected with replacement with pp to x. He also considered the situation when the first sample is used solely for X. Then an independent sample of size n is selected with pps using a procedure due to Lahiri (1951), in which it is not necessary to know X.

Sedransk (1965) has proposed a double sampling scheme in analytical sample surveys. It was assumed that the sub groups of the population which are to be compared form a one-way classification, and are not identifiable in advance of drawing a simple random sample from the population. With double sampling, a large sample is selected, and the group to which each element belongs is identified. Then a sub-sample is selected within each of the groups according to a sampling rule specified in advance. Those values of the preliminary and main sample sizes were considered which will maximise some precision statement subject to a given budget. He also gave approximate procedures to secure quick and reliable solutions. The validity of the approximations are also investigated.

The application of double sampling is also being made in successive sampling on two occasions by Sen (1972, 1973) and the theory developed so far aims at providing the optimum estimate by combining

(i) a double sampling regression estimate from the matched portion of the sample and

(ii) a sample mean based on a ~~random~~ sample from the unmatched portion of the sample on the second occasion. Sen (1972) has generalized the theory by using a double sampling multivariate ratio estimate using p auxiliary variates ($p \geqslant 1$) from the matched portion of the sample.

The author (1973a) has also generalized the theory to provide optimum estimate by combining a double sampling multivariate ratio or regression estimate using p auxiliary variables ($p \geqslant 1$) from the matched portion of the sample with a mean per unit estimate from the unmatched portion of the sample. Results have been presented for the more general and practical case when the samples on the two occasions are of unequal size.

Sedransk and Singh (1974) have suggested the use of double sampling to estimate parameters of sub-populations which are not "identifiable in advance". That is, one selects a large, preliminary ("first phase") sample and identifies the sub-population to which each element belongs. Then, for sub-population j, a subsample is selected from the elements identified in the first phase sample as being members of j. Finally, the variable of interest is measured for each element in this "second phase" sample. Both simple random and stratified random sampling was considered at the first phase while (in each case) simple random sampling is assumed at the second phase. They investigated two types of sample size allocation problem :

1. determination of the sample size (s) at the first phase and
2. determination of the second phase sample sizes given the results of the first phase sample.

In this work they have also discussed about the estimation and comparison of demain means. The procedures used here also given in somewhat greater detail by deGraft-Johnson and Sedransk (1973).

1.3 Preliminary test estimator in double sampling

Han (1973) suggested the use of preliminary test estimator in double sampling. When the mean of the auxiliary variable is completely unknown, double sampling techniques can be adopted. If the experimenter has partial information about the mean, he may perform a preliminary test and construct a preliminary test estimator.

Preliminary test estimators are studied by Bancroft (1944, 1964), Kitagawa (1963), Mosteller (1948) and others. The bias, mean square error and relative efficiency were obtained for the preliminary test estimator by Han (1973). He has also given the optimum allocation of the sample sizes.

Bock, Yancey and Judge (1973) have studied the statistical consequences of preliminary test estimators in regression. The study is concerned with deriving the properties of the preliminary test estimator for the general linear normal regression model, ascertaining the characteristics of the risk functions over the parameter space, and determining the conditions necessary for the risk of this estimator to exceed or be less than the conventional one under squared error loss. A test procedure and the problem of choosing an optimum level of significance for the test are discussed.

Sukhatme and Tang (1975) discussed allocation in stratified sampling subsequent to preliminary test of significance. A procedure for allocation of sample sizes to different strata consists of drawing a preliminary sample of fixed size from each stratum, to estimate the strata variances and test their homgeneity. If the strata variances are found homogeneous, the sample sizes to be drawn from different strata are allocated according to proportional allocation; otherwise they are allocated according to Neyman allocation using estimated variances. The efficiency of the proposed allocation, based on preliminary test of significance with respect to proportional allocation and modified Neyman allocation is investigated.

If data on an auxiliary variable X correlated with the variable Y under study are available, regression-type estimators are often used to estimate the population mean μ_y. An estimator based on a preliminary test of significance was suggested by Grimes and Sukhatme (1980), that chooses between the difference estimator and the regression estimator. They have also investigated the efficiency of the proposed regression-type estimator with respect to other regression-type estimators.

A preliminary test estimator for the population mean on the current occasion in case of sampling over two occassions is built up by Sisodia (1981) which depends on the outcome of the preliminary test. Both the cases are considered when variance-covariance of the variables on both the occas. ion is known and unknown. In both the cases, the preliminary test estimators are found to be better than usual estimators for large value of ρ , depending upon the proper choice of α and q .

An alternative to the usual regression estimator for a population mean in double sampling was suggested by Sisodia and Srivastava (1982), on the basis of a preliminary test of a simple hypothesis about the auxiliary variate-mean. Two phase sampling is assumed from a bivariate normal population. Gain in efficiency is investigated theoritically and empirically.

Esimai and Han (1982) have proposed a linear regression preliminary test estimator for estimating the population mean μ_y of Y with X, a p X 1 ($p \geqslant 1$) vector, as the auxiliary variable. The bias, mean square error and relative efficiency were obtained. They have also given the optimum allocation of the sample sizes.

1.9 Objective of the study

One of the main objectives of the thesis is to suggest alternative estimators of population mean in double sampling. Bias function is being obtained for these suggested estimators. In order to compare the efficiency of the suggested estimators with that of the other existing estimators, the mean square error (MSE) was used as a useful criterion.

It is a well known fact that there exist number of sampling techniques wherein availability of advance information about an auxiliary variable leads to greater precision. Some authors have shown that estimators using advance information about two auxiliary variables leads to even greater precision when compared with that of using only one auxiliary variable.

In the present investigation an attempt is being made to construct preliminary test estimators in double sampling using two auxiliary variables. Comparisons are made to see whether or not the inclusion of second auxiliary variable leads to higher efficiency when compared with preliminary test estimator using one auxiliary variable. Attempt is also being made to determine the conditions under which the preliminary test estimator using two auxiliary variables is better than the other known estimators using two auxiliary variables.

In order to determine the optimum MSE of the suggested estimators cost function approach was followed. This is done by suggesting a suitable linear cost function C and then conditions are determined to minimise MSE for specified C.

The determination of bias, mean square error etc. for the preliminary test estimators involves conditional expectations. These are derived under certain assumptions about the joint distribution of the parent populations of the variables from where the samples are drawn.

Finally, the thesis aims at doing some emperical studies to show the applications of the suggested estimators under practical situations and also to demonstrate the performance of these vis-a-vis other existing estimators.

1.10 Plan of the thesis

Here, we are trying to give a brief account of the chapter plan of the thesis, from next chapter onwards.

Chapter Two : Double sampling with two auxiliary variables with partial information on one auxiliary variable.

In this chapter an attempt is being made to suggest a preliminary test estimator using two auxiliary variables. Generally in double sampling it is assumed that mean of the auxiliary variable is completely unknown. However, here it is being assumed that the experimenter has partial information about the mean of one auxiliary variable. And the mean of the other auxiliary variable is assumed to be totally unknown. In order to utilise the partial information about the mean of one of the auxiliary variables, a preliminary test is being performed. Then an estimator based on this preliminary test is being defined. The derivation of the bias function of the estimator involves conditional expectations, the conditions being acceptance or rejection of the hypothesis considered in the preliminary test. To determine this function assumptions are being made about the joint distribution of the parent populations of the variables from where the samples are drawn. Behaviour of the bias function is being studied under different levels of the preliminary test.

Chapter Three : Relative efficiemcy and optimum allocation of preliminary test estimators in double sampling with two auxiliary variables with partial information on one auxiliary variable.

In order to judge the relative efficiency of the preliminary test estimator in relation with other existing estimators in double sampling MSE of the estimator is being obtained. This involves conditional expectations, the conditions being acceptance or rejection of the hypothesis considered in the preliminary test. To derive this, joint distribution of the parent population of the variables is assumed to be known. To determine the optimum allocation of the sample sizes, a linear cost function is being formed. The optimum conditions of the suggested estimators are being compared with that of the other existing estimators. Behaviour of the relative efficiency function is also being studied in this chapter.

Chapter Four : Double sampling with two auxiliary variables with partial information on both the auxiliary variables.

Generally in double sampling with more than one auxiliary variables, it is assumed that the mean of the auxiliary variables are completely unknown. However, here it is being assumed that the experimenter has partial information about the mean of both the auxiliary variables. In order to utilise these partial informations, preliminary tests are being performed, and an estimator based on these tests is being defined. The bias function was computed under the assumption about the joint distribution of the variables concerned. A study of the bias function was done under certain assumptions about the level of the preliminary test.

Chapter Five : Relative efficiency and optimum allocation of preliminary test estimators in double sampling with two auxiliary variables with partial information on both the auxiliary variables.

The mean square error of the preliminary test estimator with partial information on two auxiliary variables, which involved conditional expectations of order two, is being computed under certain assumptions about the joint distribution of the variables. A linear cost function is being formed to determine the optimum allocation of the sample sizes. The optimum conditions so obtained are being compared with that of other commonly used estimators. Behaviour of the relative efficiency function is also being studied.

Chapter Six : A generalised study of the preliminary test estimators in double sampling.

In this chapter, as the statement of the problem suggests, a more generalised study of the preliminary test estimators in double sampling is being done. Estimators along with their bias, MSE etc. are obtained under the situations when the assumptions about the sample sizes atc. are being violated. Behaviour of the bias function and relative efficiency function are being studied under the present situation.

Chapter Seven : Conclusion

In this chapter a general discussion is being given about the estimators suggested in the previous chapters, reflecting the main benefits of the whole study. Emprical studies were made to show the applications and performance of the suggested estimators as compared with the other existing estimators.

CHAPTER TWO

Double sampling with two auxiliary variables with partial information on one auxiliary variable.

2.1 Introduction

It is a well known fact that for estimating the population mean μ_y of the random variable Y, precision of the estimator can be increased when information on an auxiliary variable X, highly correlated with Y is readily available on all the units of the population, incorporating the knowledge of μ_x, the population mean of X. When the relationship between Y and X is found to be approximately linear but the line does not go through the origin, linear regression estimate may be used. To use the linear regression estimator it is usually assumed that the population mean μ_x is known. However, in certain practical situations μ_x is not known a priori, in which case the technique of double sampling is applied. Here one may take a preliminary sample to estimate it. In the first sample of size n', we measure only x_i; in the second, a random sub sample of size n ($<$ n'), we measure both x_i and y_i. The estimate of μ_y is defined as

$$t_1 = \bar{y}_n + b_{yx} (\bar{x}_{n'} - \bar{x}_n) \quad \ldots\ldots\ldots(2.1)$$

here $\bar{x}_{n'}$ is the mean of the x_i in the first sample, $\bar{y}_n$, $\bar{x}_n$ are the means of the y_i and x_i in the second sample and b_{yx} is the least squares regression coefficient of y_i on x_i computed from the second sample.

2.2 Preliminary test estimator

In certain situations, the experimenter may have partial information about μ_x. Han (1973) has suggested the use of double sampling with partial information on the auxiliary variable. In order to utilise the partial information one can perform a preliminary ~~test~~ about the hypothesis that $\mu_x = \mu_o$ where μ_o is the value obtained from the partial information. After the preliminary sample is obtained, he can test H_o: $\mu_x = \mu_o$ against H_1 : $\mu_x \neq \mu_o$. If H_o is accepted, μ_o will be used in the regression estimator; if H_o is rejected, the sample mean based on the preliminary sample is used. This estimator is usually called the preliminary test estimator.

Let (X, Y) have a bivariate normal distribution with mean (μ_x, μ_y) and covariance matrix Σ in which the variances are denoted by σ_x^2 and σ_y^2 and the correlation coefficient by ρ_{yx}. X can be readily observed, while it is more expensive to observe the pair (X, Y). The problem is to estimate μ_y. Let (x_i, y_i), $i = 1, 2, \ldots, n$; be n independent observations on the pair (X, Y) which is supplemented by observing m more independent observations on X. In practice, the sample consisting of n observations may be a sub sample from the sample consisting of $n' = n + m$ observations. Define

$$\bar{x}_{n'} = \frac{1}{n'} \sum_{i=1}^{n'} x_i \qquad \ldots\ldots(2.2)$$

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad ....(2.3)$$

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad ....(2.4)$$

The regression estimator depends on whether the covariance matrix is known or not. If Σ is known, we may let $\sigma_x^2 = \sigma_y^2 = 1$ without loss of generality. The joint distribution of $(\bar{x}_n, \bar{y}_n)$ is normal with mean (μ_x, μ_y) and covariance matrix

$$\Sigma = \frac{1}{n} \begin{bmatrix} 1 & \rho_{yx} \\ \rho_{yx} & 1 \end{bmatrix}$$

When μ_x is unknown and the experimenter has partial information about it, he can employ a preliminary test for $H_o: \mu_x = 0$ (letting $\mu_o = 0$ without loss of generality). The preliminary test estimator is defined as

$$t_2 = \begin{cases} \bar{y}_n - \rho_{yx}\, \bar{x}_n & \text{if } |\bar{x}_{n'}| \leqslant Z_\alpha/\sqrt{n'} \\ \bar{y}_n + \rho_{yx}\, (\bar{x}_{n'} - \bar{x}_n) & \text{if } |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'} \end{cases}$$

$$.....(2.5)$$

Where Z_α is the 100 $(1 - \alpha/2)$% point of $N(0, 1)$ and α is the level of the preliminary test.

2.3 Regression estimator using two auxiliary variables in double sampling

In estimating the population mean μ_y of the random variable Y, suppose that in addition to information on an auxiliary variable X, information on yet another auxiliary variable Z is available. When μ_x is not known, we can take a preliminary sample to estimate it, as done earlier to define t_1 in (2.1). Again if μ_z is also not known, assume that Z is known over another large sample, also of size n'. In such a situation an estimator using X and Z is being suggested by Mukherjee et al (1987) as follows :

$$t_3 = \bar{y}_n + b_{yx}(\bar{x}_{n'} - \bar{x}_n) + b_{yz}(\bar{z}_{n'} - \bar{z}_n) \quad ....(2.6)$$

Where $\bar{x}_{n'}$, $\bar{x}_n$, $\bar{y}_n$ are given in (2.2), (2.3) and (2.4) respectively

$$\bar{z}_{n'} = \frac{1}{n'} \sum_{i=1}^{n'} z_i \quad ......(2.7)$$

$$\bar{z}_n = \frac{1}{n} \sum_{i=1}^{n} z_i \quad ......(2.8)$$

and b_{yx}, b_{yz} are regression estimators of y_i on x_i and z_i based on the smaller samples.

2.4 Suggested preliminary test estimator using two auxiliary variables

Suppose, while considering regression estimators with two auxiliary variables in double sampling, partial information about only one of the variables say μ_z is available. In order to utilise the partial information, one can perform a preliminary test about the hypothesis that H_o: $\mu_z = \mu_o$ where μ_o is the value obtained from the partial information. After the preliminary sample is obtained, H_o: $\mu_z = \mu_o$ can be tested against H_1: $\mu_z \neq \mu_o$. If H_o is accepted, μ_o will be used in the regression estimator; if H_o is rejected, the sample mean based on the preliminary sample is used.

Now, we proceed to construct a preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable. Let (X, Y, Z) have a trivariate normal distribution with mean (μ_x, μ_y, μ_z) and covariance matrix $\sum$ in which the variances are denoted by σ_x^2, σ_y^2 and σ_z^2 and the correlation coefficients by ρ_{yx}, ρ_{yz} and ρ_{xz}. X and Z can be readily observed, while it is more expensive to observe the triplet (X, Y, Z). The problem is to estimate μ_y. Let (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, be n independent observations on the triplet (X, Y, Z) which is supplemented by m more independent observations on X and another m independent observations on Z, where $n + m = n'$.

If $\sum$ is known, we may let $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ without loss of generality. The joint distribution of $(\bar{x}_n, \bar{y}_n, \bar{z}_n)$ is normal with mean (μ_x, μ_y, μ_z) and covariance matrix

$$\frac{1}{n}\begin{bmatrix} 1 & \rho_{yx} & \rho_{xz} \\ \rho_{yx} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{bmatrix}$$

When μ_z is unknown and the experimenter has partial information about it, a preliminary test H_o: $\mu_z = 0$ (letting $\mu_o = 0$, without loss of generality) can be employed. If H_o is accepted μ_o will be used in the regression estimator; if H_o is rejected, the sample mean $\bar{x}_{n'}$ based on the preliminary sample consisting of n' independent observations on Z is used. Since μ_x is totally unknown, it is estimated from another preliminary sample also of size n'. The preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable is defined as

$$t_4 = \begin{cases} \bar{y}_n + \underline{\underline{B}}_{yx}(\bar{x}_{n'} - \bar{x}_n) - \underline{B}_{yz}\,\bar{z}_n & \text{if } |\bar{z}_{n'}| \leq z_\alpha/\sqrt{n'} \\ \bar{y}_n + \underline{B}_{yx}(\bar{x}_{n'} - \bar{x}_n) + \underline{\underline{B}}_{yz}(\bar{z}_{n'} - \bar{z}_n) & \text{if } |\bar{z}_{n'}| > z_\alpha/\sqrt{n'} \end{cases} \quad \ldots(2.9)$$

Where

$$B_{yx} = \frac{\rho_{yx} - \rho_{yz}\ \rho_{xz}}{1 - \rho_{xz}^{2}}$$

and $$B_{yz} = \frac{\rho_{yz} - \rho_{yx}\ \rho_{xz}}{1 - \rho_{xz}^{2}}$$

are the population regression coefficients of Y on X and Z respectively, assumed to be known since $\sum$ is known. And Z_{α} is the 100 $(1 - \alpha/2)\%$ point of N (0, 1), α is the level of the preliminary test.

2.5 Bias of the preliminary test estimator

To evaluate the bias of t_4, we require the joint distribution of $(\bar{x}_{n'}, \bar{x}_n, \bar{z}_{n'}, \bar{z}_n, \bar{y}_n)$. It can be easily verified that the joint distribution of these is nothing but a multivariate normal with mean $(\mu_x, \mu_x, \mu_z, \mu_z, \mu_y)$ and variance covariance matrix.

$$\begin{bmatrix} \frac{1}{n'} & \frac{1}{n'} & \frac{\rho_{xz}}{n'} & \frac{\rho_{xz}}{n'} & \frac{\rho_{yx}}{n'} \\ \frac{1}{n'} & \frac{1}{n} & \frac{\rho_{xz}}{n'} & \frac{\rho_{xz}}{n} & \frac{\rho_{yx}}{n} \\ \frac{\rho_{xz}}{n'} & \frac{\rho_{xz}}{n'} & \frac{1}{n'} & \frac{1}{n'} & \frac{\rho_{yz}}{n'} \\ \frac{\rho_{xz}}{n'} & \frac{\rho_{xz}}{n} & \frac{1}{n'} & \frac{1}{n} & \frac{\rho_{yz}}{n} \\ \frac{\rho_{yx}}{n'} & \frac{\rho_{yx}}{n} & \frac{\rho_{yz}}{n'} & \frac{\rho_{yz}}{n} & \frac{1}{n} \end{bmatrix} \quad \ldots\ldots (2.10)$$

The deriviation of the bias function of the estimator involves conditional expectations, the conditions being acceptance or rejection of the hypothesis considered in the preliminary test. To obtain the expected value of t_4, we proceed as follows :

$$E\ (t_4) = E\ (t_4 \,\big|\, |\bar{z}_{n'}| \leqslant z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| \leqslant z_\alpha/\sqrt{n'}\ )$$

$$+ E\ (t_4 \,\big|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\ )$$

$$= E\ (\bar{y}_n + B_{yx}\ (\bar{x}_{n'} - \bar{x}_n) - B_{yz}\ \bar{z}_n \,\big|\, |\bar{z}_{n'}| \leqslant z_\alpha/\sqrt{n'}\ )$$

$$X\ P\ (|\bar{z}_{n'}| \leqslant z_\alpha/\sqrt{n'}\ )$$

$$+ E\ (\ \bar{y}_n + B_{yx}\ (\bar{x}_{n'} - \bar{x}_n) + B_{yz}(\bar{z}_{n'} - \bar{z}_n) \,\big|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$X\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= E\ (\bar{y}_n + B_{yx}\ (\bar{x}_{n'} - \bar{x}_n) - B_{yz}\ \bar{z}_n)$$

$$+ B_{yz}\ E\ (\bar{z}_{n'} \,\big|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\ )$$

$$= \mu_y - B_{yz}\,\mu_z + B_{yz}\left[\ \int_{z_\alpha/\sqrt{n'}}^{\infty} \bar{z}_{n'}\ f\ (\bar{z}_{n'})\ d\bar{z}_{n'}\right.$$

$$\left.\int_{-\infty}^{-z_\alpha/\sqrt{n'}} \bar{z}_{n'}\ f\ (\bar{z}_{n'})\ d\ \bar{z}_{n'}\ \right] \qquad .....(2.11)$$

Now, since marginal distribution of a multivariate normal is also normal, therefore,

$$\bar{z}_{n'} \sim N\left(\mu_z, \frac{1}{\sqrt{n'}}\right)$$

Hence, (2.11) reduces to

$$E(t_4) = \mu_y - B_{yz}\mu_z\left(\Phi(A) - \Phi(B)\right) + \frac{B_{yz}}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right)$$

(derivation given in Appendix A)

...(2.12)

Where $\Phi(.)$ is the cumulative distribution function of $N(0, 1)$, $\phi(.)$ is its density function and

$$A = Z_\alpha - \sqrt{n'}\,\mu_z,$$

$$B = -Z_\alpha - \sqrt{n'}\,\mu_z .$$

Therefore, from (2.12)

$$B_4 = \text{Bias}(t_4) = \frac{B_{yz}}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right)$$

$$- B_{yz}\mu_z\left(\Phi(A) - \Phi(B)\right) \quad .....(2.13)$$

2.6 Behaviour of the bias function

As partial checks it can be seen the $B_4 = -B_{yz}\mu_z$ when $\alpha = 0$, i.e., when we always accept H_o. Also $B_4 = 0$

when $\alpha = 1$.

Further, the value of B_4 for $\mu_z < 0$ is the negative of the value for $\mu_z > 0$. Hence we need to consider only the behaviour of B_4 when $\mu_z \geqslant 0$. To give an idea of the behaviour of the bias with respect to μ_z, we computed the values of B_4 (in absolute value) for a set of values of n', α and B_{yz} and are presented in the following tables, (table 2.1 - 2.4).

We notice that $B_4 = 0$ when $\mu_z = 0$. Also when μ_z increases from 0, B_4 increases to a maximum, then decreases to zero. The bias is very close to zero at $\mu_z = 1$. The bias found here is quite small almost in all cases. The general behaviour of B_4 with respect to μ_z is given in fig. 2.1.

Table 2.1 Behaviour of Bias (t_4) w.r.t. μ_z

For $n' = 30$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_4|$

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	0	0.040	0.066	0.070	0.055	0.033	0.015	0.005	0.001	0	0
0.10	0	0.030	0.047	0.046	0.032	0.017	0.007	0.002	0	0	0
0.25	0	0.015	0.021	0.018	0.010	0.004	0.001	0	0	0	0

Table 2.2 Behaviour of Bias (t_4) w.r.t. μ_z

For $n' = 30$, $\rho_{xz} = 0.7$, $\rho_{yx} = 0.8$, $\rho_{yz} = 0.9$

Values of $|B_4|$

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	0	0.045	0.074	0.079	0.062	0.037	0.017	0.006	0.002	0	0
0.10	0	0.034	0.053	0.100	0.064	0.019	0.007	0.002	0	0	0
0.25	0	0.016	0.023	0.020	0.012	0.005	0.002	0	0	0	0

Table 2.3 Behaviour of Bias (t_4) w.r.t. μ_z

For $n' = 50$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_4|$

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	0	0.038	0.055	0.045	0.023	0.007	0.002	0	0	0	0
0.10	0	0.029	0.038	0.026	0.011	0.003	0.001	0	0	0	0
0.25	0	0.013	0.015	0.009	0.003	0.001	0	0	0	0	0

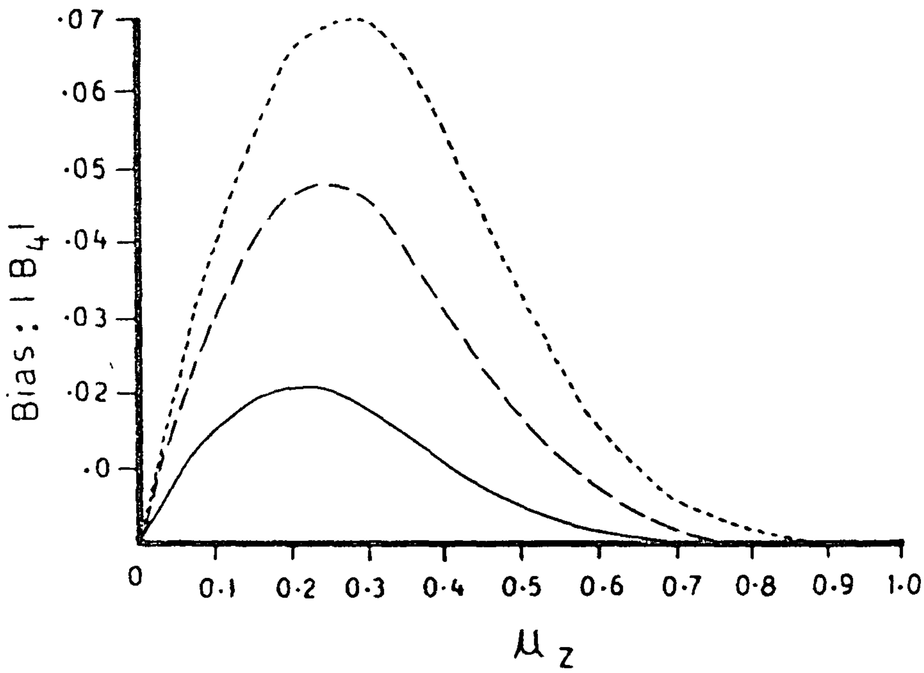
Table 2.4 Behaviour of Bias (t_4) w.r.t. μ_z

For $n' = 50$, $\rho_{xz} = 0.7$, $\rho_{yx} = 0.8$, $\rho_{yz} = 0.9$

Values of $|B_4|$

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	0	0.043	0.062	0.050	0.025	0.018	0.002	0.001	0	0	0
0.10	0	0.032	0.042	0.030	0.013	0.003	0.001	0	0	0	0
0.25	0	0.015	0.017	0.010	0.003	0.001	0	0	0	0	0

Fig. 2·1



CHAPTER THREE

Relative efficiency and optimum allocation of PTE in double sampling with two auxiliary variables with partial information on one auxiliary variable.

3.1 Introduction

The precision, or a measure of the closeness of the sample estimates to the census count taken under identical conditions is judged in sampling theory by the variance of the estimators concerned. Here reliance is placed on the fact that with a small variance the probability of large deviations from the census count will be small. The general principle is to use estimators which give the highest concentration of the sample estimates (in the sense of probability) around the value aimed for. With unbiased estimators the method used for judging the degree of concentration is the variance of the estimators.

It may happen sometime that the degree of concentration of the sample estimates around the value aimed at is higher for the distribution of a biased estimator than for an unbiased one. In such a situation the biased estimator is preferable to the unbiased one. However in order to compare a biased estimator with an unbiased estimator, or two estimators with different amounts of bias, variance is not a satisfactory criterion, since it measures deviation from the expected value of the estimator, which is not the same as the population value. A useful criterion is the mean square error (MSE) of the estimate, measured from the population value that is being estimated. Formally,

$$\text{MSE}(t) = \text{Variance of } t + (\text{bias of } t)^2 \quad \ldots\ldots(3.1)$$

obviously, if t is unbiased, the variance and the mean square error would coincide.

Next, the problem of optimum allocation of the sample sizes is considered. This is done by defining a linear cost function C. Then, the value of the sample sizes are obtained by minimising the MSE for a specified C or minimising C for a specified MSE.

3.2 Mean square error of t_4

As discussed in the earlier chapter, the preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable, is a biased estimator. The bias, though, found to be quite small almost in all cases. In order to compare the relative efficiency of the preliminary test estimator in relation with other existing estimators in double sampling, MSE of the estimator is to be obtained.

From (2.9) the estimator t_4 is defined as

$$t_4 = \begin{cases} \bar{y}_n + B_{yx}(\bar{x}_{n'} - \bar{x}_n) - B_{yz}\bar{z}_n & \text{if } |\bar{z}_{n'}| \leqslant Z_\alpha/\sqrt{n'} \\ \bar{y}_n + B_{yx}(\bar{x}_{n'} - \bar{x}_n) + B_{yz}(\bar{z}_{n'} - \bar{z}_n) & \text{if } |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'} \end{cases}$$

The components viz. B_{yx}, B_{yz}, $\bar{x}_{n'}$, $\bar{z}_n$ etc. have the same definition as in chapter 2. The assumptions about the distribution of (X, Y, Z) is also same as before.

The deriviation of MSE of the estimator involves conditional expectations of products, the conditions being acceptance or rejection of the hypothesis. To obtain MSE of t_4, we notice that

$$M_4 = MSE\ (t_4) = E\ (t_4^2) - (E\ (t_4)\ )^2 + B_4^2 \qquad \ldots\ldots(3.2)$$

Now,

$$E\ (t_4^2) = E\ (t_4^2 \,\big|\, |\bar{z}_{n'}| \leq Z_\alpha/\sqrt{n'}\,)\ P\ (|\bar{z}_{n'}| \leq Z_\alpha/\sqrt{n'}\,)$$

$$+\ E\ (t_4^2 \,\big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\,)\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\,)$$

$$= E\left[(\bar{y}_n + B_{yx}(\bar{x}_{n'} - \bar{x}_n) - B_{yz}\ \bar{z}_n)^2\right]$$

$$+\ E\left[B_{yz}^2\ \bar{z}_{n'}^2 + 2\ B_{yz}\bar{z}_{n'}(\bar{y}_n + B_{yx}(\bar{x}_{n'} - \bar{x}_n) - B_{yz}\bar{z}_n)\,\Big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right]$$

$$X\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\,)$$

$$= (\mu_y^2 + \tfrac{1}{n}) + B_{yx}^2\ (\mu_x^2 + \tfrac{1}{n'}) + B_{yx}^2\ (\mu_x^2 + \tfrac{1}{n})$$

$$+\ B_{yz}^2(\mu_z^2 + \tfrac{1}{n}) - 2\ B_{yx}^2(\mu_x^2 + \tfrac{1}{n'}) + 2\ B_{yx}(\mu_x\mu_y + \tfrac{\rho_{yx}}{n'}\ \ )$$

$$- 2 B_{yx}(\mu_x \mu_y + \frac{\rho_{yx}}{n}) - 2 B_{yz} (\mu_y \mu_z + \frac{\rho_{yz}}{n})$$

$$- 2 B_{yx} B_{yz} (\mu_x \mu_z + \frac{\rho_{xz}}{n'}) + 2 B_{yx} B_{yz} (\mu_x \mu_z + \frac{\rho_{xz}}{n})$$

$$+ E \Big[B_{yz}^2 \bar{z}_{n'}^2 + 2 B_{yz} \bar{z}_{n'} \bar{y}_n + 2 B_{yz} B_{yx} \bar{z}_{n'} \bar{x}_{n'}$$

$$- 2 B_{yz} B_{yx} \bar{z}_{n'} \bar{x}_n - 2 B_{yz}^2 \bar{z}_{n'} \bar{z}_n \Big| |\bar{z}_{n'}| > z_\alpha/\sqrt{n'} \Big]$$

$$X\ P (|\bar{z}_{n'}| > z_\alpha /\sqrt{n'}) \qquad \ldots\ldots.(3.3)$$

Again,

$$E (\bar{z}_{n'}^2 \Big| |\bar{z}_{n'}| > z_\alpha/\sqrt{n'}) P (|\bar{z}_{n'}| > z_\alpha /\sqrt{n'})$$

$$= (\mu_z^2 + \frac{1}{n'}) (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{2\mu_z}{\sqrt{n'}} (\phi(A) - \phi(B))$$

$$+ \frac{1}{n'} (A \phi(A) - B \phi(B)) \qquad \ldots..(3.4)$$

Where $\Phi(.)$ is the cumulative distribution function of $N (0, 1)$, $\phi(.)$ is its density function and $A = z_\alpha - \sqrt{n'}\mu_z$, $b = - z_\alpha - \sqrt{n'}\mu_z$

Also,

$$E (\bar{z}_{n'} \bar{z}_n \Big| |\bar{z}_{n'}| > z_\alpha/\sqrt{n'}) P (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= \left(\mu_z^2 + \frac{1}{n'}\right)\left(1 - \Phi(A) + \Phi(B)\right)$$

$$+ \frac{2\mu_z}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right)$$

$$+ \frac{1}{n'}\left(A\,\phi(A) - B\,\phi(B)\right) \quad \ldots\ldots\ldots\ldots\ldots(3.5)$$

$$E\left(\bar{z}_{n'}\,\bar{y}_n \,\middle|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_y\,\mu_z + \frac{\rho_{yz}}{n'}\right)\left(1 - \Phi(A) - \Phi(B)\right)$$

$$+ \frac{1}{\sqrt{n'}}\left(\mu_y + \rho_{yz}\,\mu_z\right)\left(\phi(A) - \phi(B)\right)$$

$$+ \frac{\rho_{yz}}{n'}\left(A\,\phi(A) - B\,\phi(B)\right) \quad \ldots\ldots\ldots(3.6)$$

$$E\left(\bar{z}_{n'}\,\bar{x}_n \,\middle|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x\mu_z + \frac{\rho_{xz}}{n'}\right)\left(1 - \Phi(A) + \Phi(B)\right)$$

$$+ \frac{1}{\sqrt{n'}}\left(\mu_x + \rho_{xz}\,\mu_z\right)\left(\phi(A) - \phi(B)\right)$$

$$+ \frac{\rho_{xz}}{n'}\left(A\,\phi(A) - B\,\phi(B)\right)$$

$$\ldots\ldots\ldots\ldots(3.7)$$

$$E\left(\bar{z}_{n'}\,\bar{x}_{n'}\,\middle|\,|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x \mu_z + \frac{\rho_{xz}}{n'}\right)\left(1 - \Phi(A) + \Phi(B)\right)$$

$$+ \frac{1}{\sqrt{n'}}\left(\mu_x + \rho_{xz}\mu_z\right)\left(\phi(A) - \phi(B)\right)$$

$$+ \frac{\rho_{xz}}{n'}\left(A\,\phi(A) - B\,\phi(B)\right) \qquad \ldots\ldots\ldots\ldots(3.8)$$

(Derivations of (3.4) - (3.8) are given in Appendix B, C, D and E)

Now, substituting (3.4) - (3.8) in (3.3) and simplifying we obtain,

$$E\left(t_4^2\right) = \left(\mu_y^2 + \frac{1}{n}\right) - B_{yx}^2\left(\mu_x^2 + \frac{1}{n'}\right)$$

$$+ B_{yx}^2\left(\mu_x^2 + \frac{1}{n}\right) + B_{yz}^2\left(\mu_z^2 + \frac{1}{n}\right)$$

$$- 2\,B_{yx}\,\rho_{yx}\left(\frac{1}{n} - \frac{1}{n'}\right) - 2\,B_{yz}\left(\mu_y\mu_z + \frac{\rho_{yz}}{n}\right)$$

$$+ 2\,B_{yx}\,B_{yz}\,\rho_{xz}\left(\frac{1}{n} - \frac{1}{n'}\right)$$

$$- B_{yz}^2\Big[\left(\mu_z^2 + \frac{1}{n'}\right)\left(1 - \Phi(A) + \Phi(B)\right)$$

$$+ \frac{2\mu_z}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right) + \frac{1}{n'}\left(A\,\phi(A) - B\,\phi(B)\right)\Big]$$

$$+ 2\,B_{yz}\Big[\left(\mu_y\mu_z + \frac{\rho_{yz}}{n'}\right)\left(1 - \Phi(A) + \Phi(B)\right)$$

$$+ \frac{1}{\sqrt{n'}} \left(\mu_y + \rho_{yz} \mu_z \right) \left(\phi(A) - \phi(B) \right)$$

$$+ \frac{\rho_{yz}}{n'} \left(A\,\phi(A) - B\,\phi(B) \right) \Big]$$

$$= \left(\mu_y^2 + \frac{1}{n} \right) + B_{yx}^2 \left(\frac{1}{n} - \frac{1}{n'} \right) + B_{yz}^2 \left(\frac{1}{n} - \frac{1}{n'} \right)$$

$$- 2\,B_{yx}\,\rho_{yx} \left(\frac{1}{n} - \frac{1}{n'} \right) - 2\,B_{yz}\,\rho_{yz} \left(\frac{1}{n} - \frac{1}{n'} \right)$$

$$+ 2\,B_{yx}\,B_{yz}\,\rho_{xz} \left(\frac{1}{n} - \frac{1}{n'} \right)$$

$$+ \left(\Phi(A) - \Phi(B) \right) \left[B_{yz}^2 \left(\mu_z^2 + \frac{1}{n'} \right) - 2\,B_{yz} \left(\mu_y \mu_z + \frac{\rho_{yz}}{n'} \right) \right]$$

$$+ \frac{1}{\sqrt{n'}} \left(\phi(A) - \phi(B) \right) \left[-2\,B_{yz}^2 \mu_z + 2\,B_{yz} \left(\mu_y + \rho_{yz} \mu_z \right) \right]$$

$$+ \frac{1}{n'} \left(A\,\phi(A) - B\,\phi(B) \right) \left[-B_{yz}^2 + 2\,B_{yz}\,\rho_{yz} \right] \qquad \ldots\ldots(3.9)$$

Now, from (2.12), (2.13) and (3.2), it follows that

$$M_4 = E\left(t_4^2\right) - \mu_y^2 - 2\,\mu_y\,B_4 \quad \ldots\ldots\ldots\ldots\ldots(3.10)$$

Substituting (3.9) and (2.13) in (3.10), we get

$$M_4 = \frac{1}{n} \left[1 + B_{yx}^2 + B_{yz}^2 - 2\,B_{yx}\,\rho_{yx} - 2\,B_{yz}\,\rho_{yz} + 2\,B_{yx}\,B_{yz}\,\rho_{xz} \right]$$

$$- \frac{1}{n'} \left[B_{yx}^2 + B_{yz}^2 - 2\,B_{yx}\,\rho_{yx} - 2\,B_{yz}\,\rho_{yz} + 2\,B_{yx}\,B_{yz}\,\rho_{xz} \right]$$

$$+ \left(\Phi(A) - \Phi(B) \right) \left[B_{yz}^2 \left(\mu_z^2 + \frac{1}{n'} \right) - 2 B_{yz} \frac{\rho_{yz}}{n'} \right]$$

$$+ \frac{1}{\sqrt{n'}} \left(\phi(A) - \phi(B) \right) \left[- 2 B_{yz}^2 \mu_z + 2 B_{yz} \rho_{yz} \mu_z \right]$$

$$+ \frac{1}{n'} \left(A \phi(A) - B \phi(B) \right) \left[- B_{yz}^2 + 2 B_{yz} \rho_{yz} \right]$$

$$= g_4 + h_4 \qquad \qquad \ldots\ldots\ldots\ldots(3.11)$$

Where $g_4 = \frac{1}{n} (1 - \rho^2_{y.xz}) + \frac{1}{n'} \rho^2_{y.xz}$ ....(3.12)

and $h_4 = \left(\Phi(A) - \Phi(B) \right) \left[B_{yz}^2 \left(\mu_z^2 + \frac{1}{n'} \right) - 2 B_{yz} \frac{\rho_{yz}}{n'} \right]$

$$- \frac{1}{\sqrt{n'}} \left(\phi(A) - \phi(B) \right) \left[2 B_{yz}^2 \mu_z - 2 B_{yz} \rho_{yz} \mu_z \right]$$

$$- \frac{1}{n'} \left(A \phi(A) - B \phi(B) \right) \left[B_{yz}^2 - 2 B_{yz} \rho_{yz} \right] \qquad \ldots.(3.13)$$

3.3 Relative efficiency of t_4

The quantity g_4 in (3.12) is the variance of the estimator t_3 (Mukherjee et al 1987), the linear regression estimator using two auxiliary variables in double sampling, under the assumption that Σ is known and $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$. The relative efficiency of t_4 to t_3 is defined as

$$e_4 = \frac{MSE\ (t3)}{MSE\ (t_4)} = \frac{g_4}{g_4 + h_4} \qquad \ldots.(3.14)$$

The values of e_4 can be easily computed for different values of μ_z. We notice that e_4 is symmetric about $\mu_z = 0$, hence we need to consider only $\mu_z \geqslant 0$. To give an idea about the behaviour of the relative efficiency function with respect to μ_z, e_4 was computed for a set of values of n, n', α and B_{yz} and are presented in the following tables (Table 3.1 - 3.4).

In general e_4 has a maximum at $\mu_z = 0$, when μ_z increases e_4 decreases to a minimum and then increases to unity. It is found that e_4 is very close to 1 at $\mu_z = 1$. The general behaviour of e_4 is given in fig 3.1.

3.4 Optimum allocation of sample sizes

We next consider the following sample design problem For a given cost function, what is the optimum allocation of the sample sizes n' and n ? Let the cost function be of the form

$$C = nc_1 + n'c_2 + n'c_3$$

where c_1, c_2 and c_3 are the costs of observing Y, X and Z respectively. Or, equivalently,

$$C = nc_1 + n'c_1' \quad ....(3.15)$$

where $c_1' = c_2 + c_3$.

Table 3.1 Behaviour of relative efficiency 'e_4' w.r.t μ_z

For $n' = 30$, $n = 10$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of e_4

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	1.378	1.219	0.951	0.778	0.722	0.753	0.841	0.924	0.975	0.994	0.999
0.10	1.270	1.147	0.943	0.822	0.801	0.848	0.918	0.968	0.991	0.998	1.000
0.25	1.118	1.059	0.960	0.908	0.915	0.950	0.980	0.994	0.999	1.000	1.000

Table 3.2 Behaviour of relative efficiency 'e_4' w.r.t μ_z

For n' = 30, n = 10, ρ_{xz} = 0.7, ρ_{yx} = 0.8, ρ_{yz} = 0.9

Values of e_4

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	1.754	1.393	0.926	0.692	0.625	0.661	0.772	0.886	0.962	0.991	0.998
0.10	1.499	1.252	0.915	0.849	0.620	0.781	0.877	0.951	0.987	0.997	1.000
0.25	1.197	1.096	0.940	0.863	0.873	0.924	0.968	0.991	0.998	1.000	1.000

Table 3.3 Behaviour of relative efficiency 'e_4' w.r.t. μ_z

For $n' = 50$, $n = 10$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of e_4

α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	1.252	1.098	0.873	0.781	0.818	0.911	0.975	0.996	1.000	1.000	1.000
0.10	1.185	1.061	0.893	0.844	0.893	0.959	0.991	0.999	1.000	1.000	1.000
0.25	1.084	1.021	0.941	0.934	0.967	0.991	0.999	1.000	1.000	1.000	1.000

Table 3.4 Behaviour of relative efficiency 'e_4' w.r.t. μ_z

For n' = 50, n = 10, ρ_{xz} = 0.7, ρ_{yx} = 0.8, ρ_{yz} = 0.9

Values of e_4

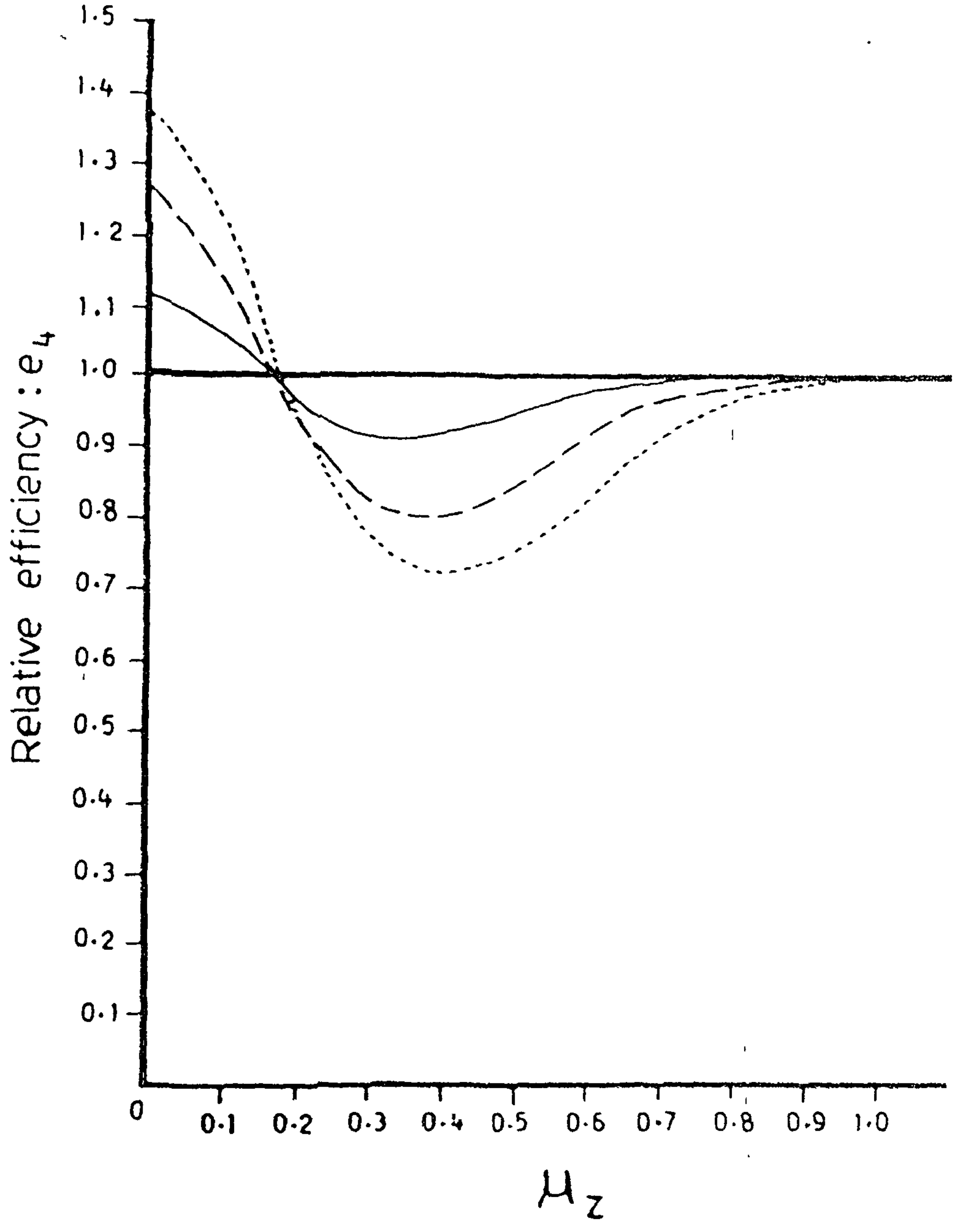
α \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	1.551	1.187	0.797	0.670	0.719	0.854	0.956	0.993	1.000	1.000	1.000
0.10	1.379	1.113	0.825	0.755	0.826	0.930	0.984	0.998	1.000	1.000	1.000
0.25	1.158	1.033	0.901	0.889	0.819	0.984	0.997	1.000	1.000	1.000	1.000

Fig 3.1: Behaviour of relative efficiency e_4 w.r.t. μ_z

For $n' = 30$, $n = 10$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

....., $\alpha = 0.05$; ---, $\alpha = 0.10$, ———, $\alpha = 0.25$

Fig. 3·1



1·5
1·4
1·3
1·2
1.1
1·0
0·9
0·8
0·7
0·6
0·5
0·4
0·3
0·2
0·1
Relative efficiency : e_4
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
μ_z

The values of n and n' are obtained by minimising MSE (t_4) subject to the cost constraint (3.15). In general, the values of μ_z are unknown, the experimenter has partial information about μ_z and believes that μ_z is close to zero. When $\mu_z = 0$, the relative efficiency of t_4 is the largest. Thus, it would reasonable to let $\mu_z = 0$ in MSE (t_4) and obtain the values of n' and n under the optimum situation.

Now, when $\mu_z = 0$, then

$$A = Z_\alpha - \sqrt{n'}\ \mu_z = Z_\alpha \quad \text{and}$$

$$B = -Z_\alpha - \sqrt{n'}\ \mu_z = -Z_\alpha$$

Which further implies,

$$\left.\begin{aligned} \Phi(A) &= \Phi(Z_\alpha) = 1 - \frac{\alpha}{2} \\ \text{and} \qquad \Phi(B) &= \Phi(-Z_\alpha) = \frac{\alpha}{2} \end{aligned}\right\} \qquad \ldots\ldots(3.16)$$

Substituting (3.16) in (3.13) we get,

$$M_4 = \frac{1}{n}\left(1 - \rho^2_{y.xz}\right) + \frac{1}{n'}\ \rho^2_{y.xz}$$

$$+ \frac{1-\alpha}{n'}\ \left(B^2_{yz} - 2\ B_{yz}\ \rho_{yz}\right)$$

$$- \frac{2 Z_\alpha \phi(Z_\alpha)}{n'} (B_{yz}^2 - 2 B_{yz} \rho_{yz})$$

$$= \frac{1}{n} (1 - \rho_{y.xz}^2) + \frac{1}{n'} \rho_{y.xz}^2$$

$$+ \frac{1 - \alpha - 2 Z_\alpha \phi(Z_\alpha)}{n' (1 - \rho_{xz}^2)} (\rho_{y.xz}^2 \rho_{xz}^2 - \rho_{yz}^2)$$

$$= \frac{1}{n} (1 - \rho_{y.xz}^2) + \frac{1}{n'(1 - \rho_{xz}^2)} (\rho_{y.xz}^2 - \rho_{yz}^2)$$

$$+ \frac{\alpha + 2 Z_\alpha \phi(Z_\alpha)}{n' (1 - \rho_{xz}^2)} (\rho_{yz}^2 - \rho_{y.xz}^2 \rho_{xz}^2)$$

$$= \frac{k}{n} + \frac{k'}{n'} \qquad \text{.............................(3.17)}$$

where

$k = 1 - \rho_{y.xz}^2$ and

$$k' = \frac{1}{1 - \rho_{xz}^2} \left[\rho_{y.xz}^2 - \rho_{yz}^2 + (\alpha + 2 Z_\alpha \phi(Z_\alpha)) (\rho_{yz}^2 - \rho_{y.xz}^2 \rho_{xz}^2) \right]$$

In order to minimise (3.17) subject to (3.15) we have to minimise M_4C where

$$M_4C = \left(\frac{k}{n} + \frac{k'}{n'} \right) (nc_1 + n'c_1') \qquad \text{........(3.18)}$$

which by Cauchy - Schwarz inequality is minimised when

$$\frac{k}{n^2 c_1} = \frac{k'}{n'^2 c_1'}$$

or $$n = \frac{C\sqrt{k}}{\sqrt{c_1}\,(\sqrt{kc_1} + \sqrt{k'c_1'})} \quad .....(3.19)$$

and $$n' = \frac{C\sqrt{k'}}{\sqrt{c_1'}\,(\sqrt{kc_1} + \sqrt{k'c_1'})} \quad ....(3.20)$$

Substituting (3.19) and (3.20) in (3.17) we get,

$$M_{4',opt} = \frac{(\sqrt{kc_1} + \sqrt{k'c_1'})^2}{C} \quad .....(3.21)$$

3.5 Comparison of the suggested estimator with other existing estimators

3.5.1 Comparison of $M_{4',opt}$ and $V_{opt}(t_3)$

We shall first compare $M_{4',opt}$ with the minimum variance of the regression estimator with two auxiliary variables under double sampling. The variance of t_3 from (3.12) is

$$V(t_3) = \frac{1}{n}(1 - \rho^2_{y.xz}) + \frac{1}{n'}\,\rho^2_{y.xz}$$

under the assumption that $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ and let as before $C = nc_1 + n'c_1'$. Again by Cauchy - Schwarz inequality, VC is minimised when

$$\frac{1 - \rho^2_{y.xz}}{n^2 c_1} = \frac{\rho^2_{y.xz}}{n'^2 c_1'}$$

$$\text{or } n = \frac{C\sqrt{1 - \rho^2_{y.xz}}}{\sqrt{c_1}\,(\sqrt{c_1'}\,\rho_{y.xz} + \sqrt{c_1\,(1 - \rho^2_{y.xz})})} \qquad \ldots\ldots(3.22)$$

$$\text{and } n' = \frac{C\,\rho_{y.xz}}{\sqrt{c_1'}\,(\sqrt{c_1'}\,\rho_{y.xz} + \sqrt{c_1\,(1 - \rho^2_{y.xz})}\,)} \qquad \ldots\ldots(3.23)$$

Substituting (3.22) and (3.23) in $V(t_3)$, we obtain

$$V_{opt}(t_3) = \frac{(\sqrt{c_1'}\,\rho_{y.xz} + c_1\sqrt{(1 - \rho^2_{y.xz})}\,)^2}{C} \qquad \ldots(3.24)$$

In order to compare (3.21) and (3.24), we observe that, $\alpha + 2\,Z_\alpha\,\phi(Z_\alpha)$ is a decreasing function of Z_α with a maximum equal to unity at $Z_\alpha = 0$. Therefore, we conclude that

$$M_{4,opt} \leqslant V_{opt}(t_3) \text{ provided}$$

$$\rho^2_{yz} \geqslant \rho^2_{y.xz}\;\rho^2_{xz}\,,$$

with equality holding for $Z_\alpha = 0$, which is the case when the two estimators coincide.

3.6 Discussion

In the earlier section we have proved that under certain conditions, mean square error of a preliminary test estimator in double sampling with two auxiliary variables is smaller than the mean square error of the usual regression estimator in double sampling with two auxiliary variables. Therefore under the stated assumptions, preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable is more efficient than the regression estimator in double sampling with two auxiliary variables.

CHAPTER FOUR

Double sampling with two auxiliary variables with partial information on both the auxiliary variables.

4.1 Introduction

In Chapter 2 we have discussed the situation when in double sampling with two auxiliary variables, partial information on one of the auxiliary variables is available. In order to utilise this information, need for preliminary test estimator was felt and accordingly we suggested a preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable. In this chapter, attempt is being made to consider the situation when partial information on two auxiliary variables related to the study variable is available.

Suppose we are interested in estimating the population mean μ_y of a study variable Y. When information on two auxiliary variables X and Z are available, and μ_x, μ_z both being unknown, Mukherjee et al (1987) have suggested an estimator using X and Z given in (2.6). Suppose, further we have partial information about both μ_x and μ_z. In order to utilise this information we can perform preliminary tests for the hypotheses

$$H_{o1} : \mu_x = \mu_{ox}$$

and $$H_{o2} : \mu_z = \mu_{oz}$$

where μ_{ox}, μ_{oz} are the values obtained from the partial informations. If H_{o1} is accepted, μ_{ox} will be used in the regression estimator; if H_{o1} is rejected, the sample mean based on the preliminary sample for X is used. Similar arguments can be given for H_{o2} as well.

4.2 Suggested preliminary test estimator

Now, we proceed to construct a preliminary test estimator in double sampling with two auxiliary variables having partial information on both the auxiliary variables. Let, as before, (X, Y, Z) have a trivariate normal distribution with mean (μ_x, μ_y, μ_z) and covariance matrix Σ, in which the variances are denoted by σ_x^2, σ_y^2 and σ_z^2 and the correlation coefficients by ρ_{yx}, ρ_{yz} and ρ_{xz}. X and Z can be readily observed, while it is more expensive to observe the triplet (X, Y, Z). The problem is to estimate μ_y, Let (x_i, y_i, z_i), i = 1, 2 ...., n, be n independent observations on the triplet (X, Y, Z) which is supplemented by m more independent observations on X and another m independent observations on Z, where n + m = n'

If Σ is known, we may let $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ without loss of generality. The joint distribution of $(\bar{x}_n, \bar{y}_n, \bar{z}_n)$ is normal with mean (μ_x, μ_y, μ_z) and covariance matrix

$$\Sigma = \frac{1}{n} \begin{bmatrix} 1 & \rho_{yx} & \rho_{xz} \\ \rho_{yx} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{bmatrix}$$

When μ_x, μ_z are unknown and the experimenter has partial information about them, he can employ preliminary tests for

$$H_{o1} : \mu_x = 0 \quad \text{and} \quad H_{o2} : \mu_z = 0$$

(letting $\mu_{ox} = \mu_{oz} = 0$ without loss of generality). If H_{o1} is accepted, μ_{ox} will be used in the regression estimator. Similarly if H_{o2} is accepted μ_{oz} will be used. However, if H_{o1} is rejected the sample mean $\bar{x}_{n'}$, based on the preliminary sample consisting of n' independent observation on X is used. And if H_{o2} is rejected the sample mean $\bar{z}_{n'}$, based on another preliminary sample consisting of n' independent observations on Z is used. The preliminary test estimator in double sampling with two auxiliary variables having partial information on both the auxiliary variables is defined as

$$t_5=\begin{cases}\bar{y}_n - B_{yx}\bar{x}_n - B_{yz}\bar{z}_n & \text{if } |\bar{x}_{n'}|\leqslant Z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}|\leqslant Z_\alpha/\sqrt{n'}\\ \bar{y}_n + B_{yx}(\bar{x}_{n'}-\bar{x}_n) - B_{yz}\bar{z}_n & \text{if } |\bar{x}_{n'}|> Z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}|\leqslant Z_\alpha/\sqrt{n'}\\ \bar{y}_n - B_{yx}\bar{x}_n + B_{yz}(\bar{z}_{n'}-\bar{z}_n) & \text{if } |\bar{x}_{n'}|\leqslant Z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}|> Z_\alpha/\sqrt{n'}\\ \bar{y}_n + B_{yx}(\bar{x}_{n'}-\bar{x}_n) + B_{yz}(\bar{z}_{n'}-\bar{z}_n) & \text{if } |\bar{x}_{n'}|> Z_\alpha/\sqrt{n'},\\ & \quad |\bar{z}_{n'}|> Z_\alpha/\sqrt{n'}\end{cases}$$

.....(4.1)

The components B_{yx}, B_{yz}, $\bar{x}_{n'}$, $\bar{z}_{n'}$, Z_α etc. have the same definition as in Chapter 2

4.3 Bias of the preliminary test estimator

To evaluate the bias of t_5, we required the joint distribution of $(\bar{x}_{n'}, \bar{x}_n, \bar{z}_{n'}, \bar{z}_n, \bar{y}_n)$. It can be easily verified that the joint distribution of these is a multivariate normal distribution with mean $(\mu_x, \mu_x, \mu_z, \mu_z, \mu_y)$ and covariance matrix as given in (2.10).

The derivation of the bias function of the estimator involves conditional expectations, the conditions being acceptance and rejection of the hypotheses H_{01} and H_{02}. To obtain the expected value of t_5 we proceed as follows :

$$E\ (t_5) = (t_5 \Big| |\bar{x}_{n'}| \leqslant Z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}| \leqslant Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| \leqslant Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| \leqslant Z_\alpha/\sqrt{n'})|$$

$$+\ E\ (t_5 \Big| |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| \leqslant Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| \leqslant Z_\alpha/\sqrt{n'}\ )$$

$$+\ E\ (t_5 \Big| |\bar{x}_{n'}| \leqslant Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| \leqslant Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$+\ E\ (t_5 \Big| |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$= E\ (\bar{y}_n - B_{yx}\ \bar{x}_n - B_{yz}\ \bar{z}_n)$$

$$+\ E\ (B_{yx}\ \bar{x}_{n'} \Big| |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'})\ P\ (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$+\ E\ (B_{yz}\ \bar{z}_{n'} \Big| |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$= \mu_y - B_{yx}\ \mu_x - B_{yz}\ \mu_z$$

$$+ B_{yx} \left[\int_{Z_\alpha/\sqrt{n'}}^{\infty} \bar{x}_{n'} \, f(\bar{x}_{n'}) \, d\bar{x}_{n'} + \int_{-\infty}^{-Z_\alpha/\sqrt{n'}} \bar{x}_{n'} \, f(\bar{x}_{n'}) \, d\bar{x}_{n'} \right]$$

$$+ B_{yz} \left[\int_{Z_\alpha/\sqrt{n'}}^{\infty} \bar{z}_{n'} \, f(\bar{z}_{n'}) \, d\bar{z}_{n'} + \int_{-\infty}^{-Z_\alpha/\sqrt{n'}} \bar{z}_{n'} \, f(\bar{z}_{n'}) \, d\bar{z}_{n'} \right]$$

......(4.2)

Now, since marginal distribution of a multivariate normal is also normal, therefore,

$\bar{x}_{n'} \sim N(\mu_x, 1/\sqrt{n'})$ and

$\bar{z}_{n'} \sim N(\mu_z, 1/\sqrt{n'})$

Hence from (4.2) we obtain

$$E(t_5) = \mu_y - B_{yx} \mu_x (\Phi(a) - \Phi(b)) + \frac{B_{yx}}{\sqrt{n'}} (\phi(a) - \phi(b))$$

$$- B_{yz} \mu_z (\Phi(A) - \Phi(B)) + \frac{B_{yz}}{\sqrt{n'}} (\phi(A) - \phi(B))$$

....(4.3)

(refer Appendix A for derivation).

where $\Phi(.)$ is the cumulative distribution function of $N(0, 1)$; $\phi(.)$ is its density function and

$$a = Z_\alpha - \sqrt{n'}\,\mu_x,$$

$$b = -Z_\alpha - \sqrt{n'}\,\mu_x,$$

$$A = Z_\alpha - \sqrt{n'}\,\mu_z,$$

and $$B = -Z_\alpha - \sqrt{n'}\,\mu_z,$$

From (4.3)

$$B_5 = \text{Bias}\,(t_5) = \frac{B_{yx}}{\sqrt{n'}}\left(\phi(a) - \phi(b)\right) - B_{yx}\mu_x\left(\Phi(a) - \Phi(b)\right)$$

$$+ \frac{B_{yz}}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right) - B_{yz}\mu_z\left(\Phi(A) - \Phi(B)\right)$$

...(4.4)

4.4 Behaviour of the bias function

As partial checks it can be seen that

$$B_5 = -B_{yx}\,\mu_x - B_{yz}\,\mu_z \text{ when } \alpha = 0,$$

i.e. when we always accept H_o; and $B_5 = 0$ when $\alpha = 1$.

Further, to give an idea of the bias with respect μ_x and μ_z, we computed the values of B_5 (in absolute values) for a set of values of n', α , B_{yx} and B_{yz}

which are presented in the following tables, Table 4.1 - 4.6.

We notice that $B_5 = 0$ when $\mu_x = \mu_z = 0$. Also, when μ_x, μ_z increases from 0, B_5 increases to a maximum, then decreases to zero. The bias is very close to zero at $\mu_x = \mu_z = 1$. The bias found here is quite small almost in all cases. The general behaviour B_5 with respect to μ_x and μ_z is shown in Fig 4.1.

For n' = 30, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_5|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.040	.066	.070	.055	.033	.015	.005	.001	0	0
0.1	.023	.063	.089	.093	.078	.056	.038	.028	.024	.023	.023
0.2	.038	.078	.104	.108	.093	.071	.053	.043	.039	.038	.038
0.3	.041	.081	.107	.111	.096	.074	.056	.046	.042	.041	.041
0.4	.032	.072	.098	.102	.087	.065	.047	.037	.033	.032	.032
0.5	.019	.059	.035	.089	.074	.052	.034	.024	.020	.019	.019
0.6	.009	.049	.075	.079	.064	.042	.024	.014	.010	.009	.009
0.7	.003	.043	.069	.073	.058	.036	.018	.008	.004	.003	.003
0.8	.001	.041	.067	.071	.056	.034	.016	.006	.002	.001	.001
0.9	0	.040	.066	.070	.055	.033	.015	.005	.001	0	0
1.0	0	.040	.066	.070	.055	.033	.015	.005	.001	0	0

Table 4.2 Behaviour of Bias (t_5) w.r.t. μ_x and μ_z

For $n' = 30$, $\alpha = 0.10$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_5|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.030	.047	.046	.032	.017	.007	.002	0	0	0
0.1	.017	.047	.064	.063	.049	.034	.024	.019	.017	.017	.017
0.2	.027	.057	.074	.073	.059	.044	.034	.029	.027	.027	.027
0.3	.027	.057	.074	.073	.059	.044	.034	.029	.027	.027	.027
0.4	.019	.049	.066	.065	.051	.036	.026	.021	.019	.019	.019
0.5	.009	.039	.056	.055	.041	.026	.016	.011	.009	.009	.009
0.6	.004	.034	.051	.050	.036	.021	.011	.006	.004	.004	.004
0.7	.001	.031	.048	.047	.033	.018	.008	.003	.001	.001	.001
0.8	0	.030	.047	.046	.032	.017	.007	.002	0	0	0
0.9	0	.030	.047	.046	.032	.017	.007	.002	0	0	0
1.0	0	.030	.047	.046	.032	.017	.007	.002	0	0	0

Table 4.3 Behaviour of Bias (t_5) w.r.t. μ_x and μ_z

For $n' = 30$, $\alpha = 0.25$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_5|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.015	.021	.018	.010	.004	.001	0	0	0	0
0.1	.009	.024	.030	.027	.019	.013	.010	.009	.009	.009	.009
0.2	.012	.027	.033	.030	.022	.016	.013	.012	.012	.012	.012
0.3	.010	.025	.031	.028	.020	.014	.011	.010	.010	.010	.010
0.4	.006	.021	.027	.024	.016	.010	.007	.006	.006	.006	.006
0.5	.002	.017	.023	.020	.012	.006	.003	.002	.002	.002	.002
0.6	.001	.016	.022	.019	.011	.005	.002	.001	.001	.001	.001
0.7	0	.015	.021	.018	.010	.004	.001	0	0	0	0
0.8	0	.015	.021	.018	.010	.004	.001	0	0	0	0
0.9	0	.015	.021	.018	.010	.004	.001	0	0	0	0
1.0	0	.015	.021	.018	.010	.004	.001	0	0	0	0

Table 4.4 Behaviour of Bias (t_5) w.r.t. μ_x and μ_z

For $n' = 50$, $\alpha = 0.05$, $\rho_{xz} = 0.7$, $\rho_{yx} = 0.8$, $\rho_{yz} = 0.9$

Values of $|B_5|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.043	.062	.050	.025	.018	.002	.001	0	0	0
0.1	.021	.064	.083	.071	.046	.039	.023	.022	.021	.021	.021
0.2	.031	.074	.093	.081	.056	.049	.033	.032	.031	.031	.031
0.3	.025	.068	.087	.075	.050	.043	.027	.026	.025	.025	.025
0.4	.012	.055	.074	.062	.037	.030	.014	.013	.012	.012	.012
0.5	.009	.052	.071	.059	.034	.027	.011	.010	.009	.009	.009
0.6	.001	.044	.063	.051	.026	.019	.003	.002	.001	.001	.001
0.7	0	.043	.062	.050	.025	.018	.002	.001	0	0	0
0.8	0	.043	.062	.050	.025	.018	.002	.001	0	0	0
0.9	0	.043	.062	.050	.025	.018	.002	.001	0	0	0
1.0	0	.043	.062	.050	.025	.018	.002	.001	0	0	0

Table 4.5 Behaviour of Bias (t_5) w.r.t. μ_x and μ_z

For $n' = 50$, $\alpha = 0.10$, $\rho_{xz} = 0.7$, $\rho_{yx} = 0.8$, $\rho_{yz} = 0.9$

Values of $|B_5|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.032	.042	.030	.013	.003	.001	0	0	0	0
0.1	.016	.048	.058	.046	.029	.019	.017	.016	.016	.016	.016
0.2	.021	.053	.063	.051	.034	.024	.022	.021	.021	.021	.021
0.3	.015	.047	.057	.045	.028	.018	.016	.015	.015	.015	.015
0.4	.006	.038	.048	.036	.019	.009	.007	.006	.006	.006	.006
0.5	.001	.033	.043	.031	.014	.004	.002	.001	.001	.001	.001
0.6	0	.032	.042	.030	.013	.003	.001	0	0	0	0
0.7	0	.032	.042	.030	.013	.003	.001	0	0	0	0
0.8	0	.032	.042	.030	.013	.003	.001	0	0	0	0
0.9	0	.032	.042	.030	.013	.003	.001	0	0	0	0
1.0	0	.032	.042	.030	.013	.003	.001	0	0	0	0

Table 4.6 Behaviour of Bias (t_5) w.r.t. μ_x and μ_z

For $n' = 50$, $\alpha = 0.25$, $\rho_{xz} = 0.7$, $\rho_{yx} = 0.8$, $\rho_{yz} = 0.9$

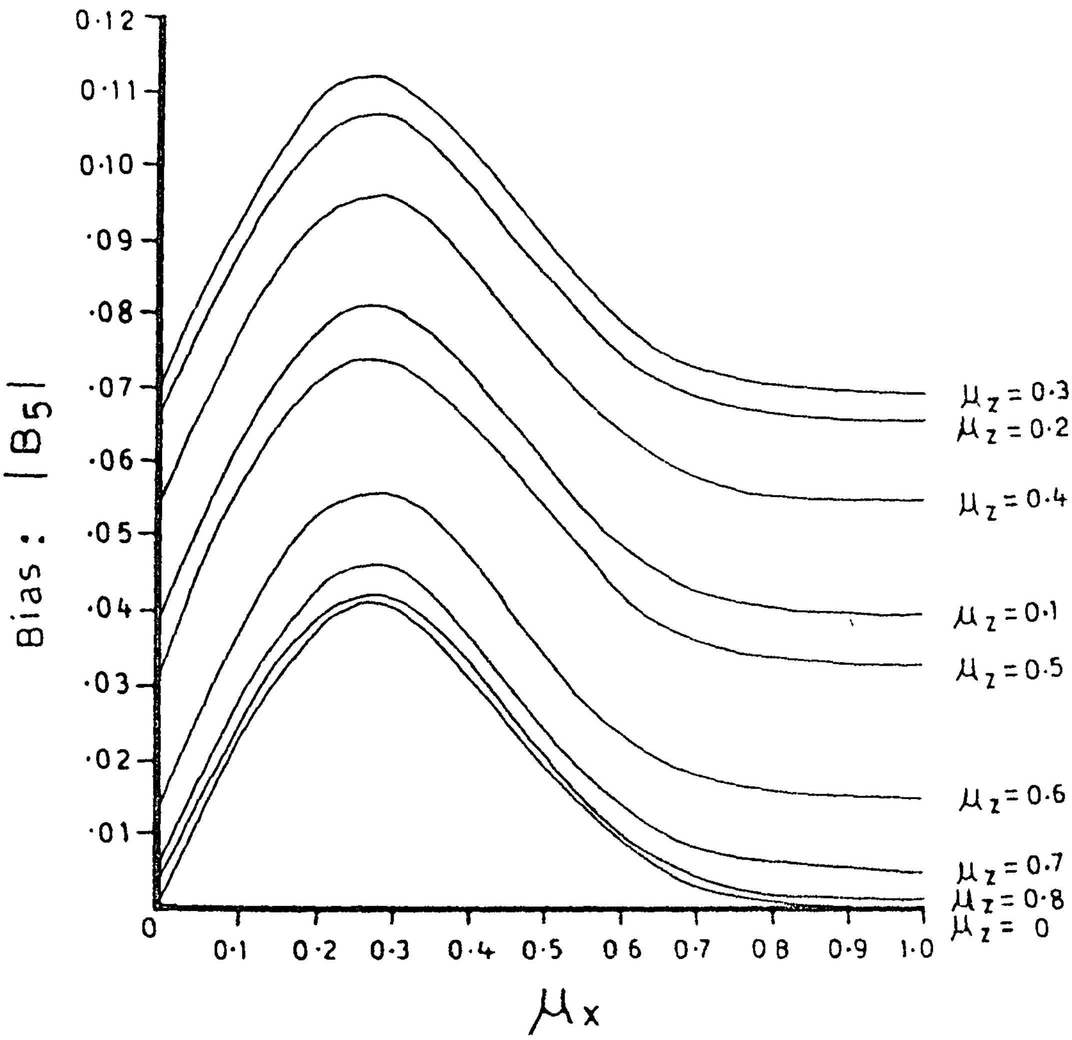
Values of $|B_5|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.015	.017	.010	.003	.001	0	0	0	0	0
0.1	.007	.022	.024	.017	.010	.008	.007	.007	.007	.007	.007
0.2	.008	.023	.025	.018	.011	.009	.008	.008	.008	.008	.008
0.3	.005	.020	.022	.015	.008	.006	.005	.005	.005	.005	.005
0.4	.001	.016	.018	.011	.004	.002	.001	.001	.001	.001	.001
0.5	0	.015	.017	.010	.003	.001	0	0	0	0	0
0.6	0	.015	.017	.010	.003	.001	0	0	0	0	0
0.7	0	.015	.017	.010	.003	.001	0	0	0	0	0
0.8	0	.015	.017	.010	.003	.001	0	0	0	0	0
0.9	0	.015	.017	.010	.003	.001	0	0	0	0	0
1.0	0	.015	.017	.010	.003	.001	0	0	0	0	0

Fig 4.1. Behaviour of Bias (t_5) w.r.t. μ_x and μ_z

For $n' = 30$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$.

Fig 4 1



0·12
0·11
0·10
·09
·08
·07
·06
·05
·04
·03
·02
·01
0
0·1
0·2
0·3
0·4
0·5
0 6
0·7
0 8
0·9
1·0
Bias : |B5|
μx
μz = 0·3
μz = 0·2
μz = 0·4
μz = 0·1
μz = 0·5
μz = 0·6
μz = 0·7
μz = 0·8
μz = 0

CHAPTER FIVE

Relative efficiency and optimum allocation of PTE in double sampling with two auxiliary variables having partial information on both the auxiliary variables.

5.1 Introduction

As discussed earlier, in Chapter 3, we know that in order to compare a biased estimator with an unbiased estimator, or two estimators with different amounts of bias, a useful criterion is the mean square error (MSE) of the estimate, measured from the population value that is being estimated. In this chapter an attempt will be made to derive the MSE of the preliminary test estimator in double sampling with two auxiliary variables having partial information on both the auxiliary variables and then to compare it with the other estimators. The problem of optimum allocation of sample sizes for the suggested estimator will also be considered here. For this, a linear cost function C will be considered, and the value of sample sizes will be obtained by minimising MSE for a specified C or by minimising C for a specified MSE.

5.2 Mean square error of t_5

As discussed in Chapter 4, the preliminary test estimator in double sampling with two auxiliary variables having partial information on both the auxiliary variables, is a biased estimator, the bias being very small in most of the cases. Thus, in order to compare this suggested estimator with other existing estimator, MSE of the said estimator is to be obtained.

To obtain MSE of t_5, we notice that

$$M_5 = MSE\ (t_5) = E\ (t_5^2) - (E\ (t_5)\ )^2 + B_5^2 \qquad \ldots.(5.1)$$

The derivation of MSE of the estimator involves conditional expectation of products, the conditions eing acceptance or rejection of the hypotheses.

Now, from (4.1)

$$E(t_5^2) = E(t_5^2 \mid |\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| \leq Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}| \leq Z_\alpha/\sqrt{n'}\ )$$

$$+\ E\ (t_5^2 \mid |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| \leq Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| \leq Z_\alpha/\sqrt{n'}\ )$$

$$+\ E\ (t_5^2 \mid |\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$+\ E\ (t_5^2 \mid |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$\times\ P\ (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\ ,\ |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\ )$$

$$= E\ (\bar{y}_n^2 + B_{yx}^{\ 2}\ \bar{x}_n^2 + B_{yz}^{\ 2}\ \bar{z}_n^2 - 2\ B_{yx}\ \bar{y}_n\bar{x}_n - 2\ B_{yz}\ \bar{y}_n\ \bar{z}_n$$

$$+ 2 B_{yx} B_{yz} \bar{x}_n \bar{z}_n)$$

$$+ E (B_{yx}^2 \bar{x}_{n'}^2 - 2 B_{yx}^2 \bar{x}_{n'} \bar{x}_n + 2 B_{yx} \bar{x}_{n'} \bar{y}_n - 2B_{yx} B_{yz} \bar{x}_{n'} \bar{z}_n$$

$$\Big| |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}) P (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$+ E (B_{yz}^2 \bar{z}_{n'}^2 - 2 B_{yz}^2 \bar{z}_{n'} \bar{z}_n + 2 B_{yz} \bar{z}_{n'} \bar{y}_n - 2 B_{yx} B_{yz} \bar{z}_{n'} \bar{x}_n$$

$$\Big| |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}) P (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$+ E (2 B_{yx} B_{yz} \bar{x}_{n'} \bar{z}_{n'} \Big| |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$X P (|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'} , |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

......(5.2)

Now,

$$E (\bar{y}_n^2 + B_{yx}^2 \bar{x}_n^2 + B_{yz}^2 \bar{z}_n^2 - 2 B_{yx} \bar{y}_n \bar{x}_n - 2 B_{yz} \bar{y}_n \bar{z}_n$$

$$+ 2 B_{yx} B_{yz} \bar{x}_n \bar{z}_n)$$

$$= (\mu_y^2 + \frac{1}{n}) + B_{yx}^2 (\mu_x^2 + \frac{1}{n}) + B_{yz}^2 (\mu_z^2 + \frac{1}{n})$$

$$- 2 B_{yx}(\mu_x \mu_y + \frac{\rho_{yx}}{n}) - 2 B_{yz}(\mu_y \mu_z + \frac{\rho_{yz}}{n})$$

$$+ 2 B_{yx} B_{yz} (\mu_x \mu_z + \frac{\rho_{xz}}{n})$$

......(5.3)

Also,

$$E\left(\bar{x}_{n'}^2 \,\middle|\, |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x^2 + \frac{1}{n'}\right)\left(1 - \Phi(a) + \Phi(b)\right)$$

$$+ \frac{2\mu_x}{\sqrt{n'}}\left(\phi(a) - \phi(b)\right)$$

$$+ \frac{1}{n'}\left(a\,\phi(a) - b\,\phi(b)\right) \qquad \text{......(5.4)}$$

Where $\Phi(.)$ is the cumulative distribution function of N (0, 1), $\phi(.)$ is its density function and $a = Z_\alpha - \sqrt{n'}\,\mu_x$, $b = -Z_\alpha - \sqrt{n'}\,\mu_x$.

$$E\left(\bar{x}_{n'}\,\bar{x}_n \,\middle|\, |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x^2 + \frac{1}{n'}\right)\left(1 - \Phi(a) + \Phi(b)\right)$$

$$+ \frac{2\mu_x}{\sqrt{n'}}\left(\phi(a) - \phi(b)\right)$$

$$+ \frac{1}{n'}\left(a\,\phi(a) - b\,\phi(b)\right) \qquad \text{........(5.5)}$$

$$E\left(\bar{x}_{n'}\,\bar{y}_n \,\middle|\, |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x\,\mu_y + \frac{\rho_{yx}}{n'}\right)\left(1 - \Phi(a) + \Phi(b)\right)$$

$$+ \frac{1}{\sqrt{n'}} \left(\mu_y + \rho_{yx} \mu_x \right) \left(\phi(a) - \phi(b) \right)$$

$$+ \frac{\rho_{yx}}{n'} \left(a\,\phi(a) - b\,\phi(b) \right) \qquad \ldots\ldots\ldots(5.6)$$

$$E\left(\bar{x}_{n'}\, \bar{z}_n \,\middle|\, |\bar{x}_{n'}| > z_\alpha/\sqrt{n'}\right) P\left(|\bar{x}_{n'}| > z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x \mu_z + \frac{\rho_{xz}}{n'} \right) \left(1 - \Phi(a) + \Phi(b) \right)$$

$$+ \frac{1}{\sqrt{n'}} \left(\mu_z + \rho_{xz} \mu_x \right) \left(\phi(a) - \phi(b) \right)$$

$$+ \frac{\rho_{xz}}{n'} \left(a\,\phi(a) - b\,\phi(b) \right) \qquad \ldots\ldots(5.7)$$

$$E\left(2\,\bar{x}_{n'}\, \bar{z}_{n'} \,\middle|\, |\bar{x}_{n'}| > z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\right)$$

$$P\left(|\bar{x}_{n'}| > z_\alpha/\sqrt{n'},\ |\bar{z}_{n'}| > z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_x \mu_z + \frac{\rho_{xz}}{n'} \right) \left(1 - \Phi(A) + \Phi(B) \right)$$

$$+ \frac{1}{\sqrt{n'}} \left(\mu_x + \rho_{xz} \mu_z \right) \left(\phi(A) - \phi(B) \right)$$

$$+ \frac{\rho_{xz}}{n'} \left(A\,\phi(A) - B\,\phi(B) \right))$$

$$+ \left(\mu_x \mu_z + \frac{\rho_{xz}}{n'} \right) \left(1 - \Phi(a) + \Phi(b) \right)$$

$$+ \frac{1}{\sqrt{n'}} \left(\mu_z + \rho_{xz} \mu_x \right) \left(\phi(a) - \phi(b) \right)$$

$$+ \frac{\rho_{xz}}{n'} \left(a\, \phi(a) - b\, \phi(b) \right) \qquad \text{......(5.8)}$$

(refer to Appendix B, C, D and E for the derivation of (5.4) - (5.8))

Now substituting (5.3) - (5.8) and (3.4) - (3.7) in (5.2) we have

$$
\begin{aligned}
E(t_4^2) = {} & \left(\mu_y^2 + \frac{1}{n} \right) + B_{yx}^2 \left(\mu_x^2 + \frac{1}{n} \right) + B_{yz}^2 \left(\mu_z^2 + \frac{1}{n} \right) \\
& - 2 B_{yx} \left(\mu_x \mu_y + \frac{\rho_{yx}}{n} \right) - 2 B_{yz} \left(\mu_y \mu_z + \frac{\rho_{yz}}{n} \right) \\
& + 2 B_{yx} B_{yz} \left(\mu_x \mu_z + \frac{\rho_{xz}}{n} \right) \\
& - B_{yx}^2 \Big[\left(\mu_x^2 + \frac{1}{n'} \right) \left(1 - \Phi(a) + \Phi(b) \right) \\
& \qquad + \frac{2 \mu_x}{\sqrt{n'}} \left(\phi(a) - \phi(b) \right) \\
& \qquad + \frac{1}{n'} \left(a\, \phi(a) - b\, \phi(b) \right) \Big] \\
& + 2 B_{yx} \Big[\left(\mu_x \mu_y + \frac{\rho_{yx}}{n'} \right) \left(1 - \Phi(a) + \Phi(b) \right) \\
& \qquad + \frac{1}{\sqrt{n'}} \left(\mu_y + \rho_{yx} \mu_x \right) \left(\phi(a) - \phi(b) \right) \\
& \qquad + \frac{\rho_{yx}}{n'} \left(a\, \phi(a) - b\, \phi(b) \right) \Big]
\end{aligned}
$$

$$- B_{yx} B_{yz} \Big[(\mu_x \mu_z + \frac{\rho_{xz}}{n'}) (1 - \Phi(a) + \Phi(b))$$

$$+ \frac{1}{\sqrt{n'}} (\mu_z + \rho_{xz} \mu_x) (\phi(a) - \phi(b))$$

$$+ \frac{\rho_{xz}}{n'} (a \phi(a) - b \phi(b)) \Big]$$

$$- B_{yz}^2 \Big[(\mu_z^2 + \frac{1}{n'}) (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{2 \mu_z}{\sqrt{n'}} (\phi(A) - \phi(B))$$

$$+ \frac{1}{n'} (A \phi(A) - B \phi(B)) \Big]$$

$$+ 2 B_{yz} \Big[(\mu_y \mu_z + \frac{\rho_{yz}}{n'}) (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{1}{\sqrt{n'}} (\mu_y + \rho_{yz} \mu_z) (\phi(A) - \phi(B))$$

$$+ \frac{\rho_{yz}}{n'} (A \phi(A) - B \phi(B)) \Big]$$

$$- B_{yx} B_{yz} \Big[(\mu_x \mu_z + \frac{\rho_{xz}}{n'}) (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{1}{\sqrt{n'}} (\mu_x + \rho_{xz} \mu_z) (\phi(A) - \phi(B))$$

$$+ \frac{\rho_{xz}}{n'} (A \phi(A) - B \phi(B)) \Big] \qquad ....(5.9)$$

Now, from (5.1), (4.3) and (4.4) it follows that

$$M_5 = E\,(t_5^2) - \mu_y^2 - 2\,\mu_y\,B_5 \qquad \ldots\ldots\ldots\ldots(5.10)$$

Therefore, substituting (5.9) and (4.4) in (5.10) we get

$$M_5 = g_5 + h_5 \qquad \ldots\ldots\ldots(5.11)$$

where

$$g_5 = \frac{1}{n}\,(1 - \rho^2_{y.xz}) + \frac{1}{n'}\,\rho^2_{y.xz} \qquad \ldots\ldots\ldots(5.12)$$

and

$$h_5 = (\,\Phi(a) - \Phi(b)\,)\Big[B_{yx}^2\,(\mu_x^2 + \tfrac{1}{n'}) - 2\,B_{yx}\,\frac{\rho_{yx}}{n'} + B_{yx}\,B_{yz}\,(\mu_x\mu_z + \frac{\rho_{xz}}{n'})\Big]$$

$$+\frac{1}{\sqrt{n'}}\,(\,\phi(a) - \phi(b)\,)\Big[-2\,\mu_x\,B_{yx}^2 + 2\,\mu_x\,B_{yx}\,\rho_{yx} - B_{yx}\,B_{yz}\,(\mu_z + \rho_{xz}\,\mu_x)\Big]$$

$$+\frac{1}{n'}\,(a\,\phi(a) - b\,\phi(b)\,)\Big[-B_{yx}^2 + 2\,B_{yx}\,\rho_{yx} - B_{yx}\,B_{yz}\,\rho_{xz}\Big]$$

$$+ (\,\Phi(A) - \Phi(B)\,)\Big[B_{yz}^2\,(\mu_z^2 + \tfrac{1}{n'}) - 2\,B_{yz}\,\frac{\rho_{yz}}{n'} + B_{yx}\,B_{yz}\,(\mu_x\,\mu_z + \frac{\rho_{xz}}{n'})\Big]$$

$$+\frac{1}{\sqrt{n'}}\,(\,\phi(A) - \phi(B)\,)\Big[-2\mu_z\,B_{yz}^2 + 2\,B_{yz}\,\rho_{yz}\,\mu_z$$

$$- B_{yx} B_{yz} (\mu_x + \rho_{xz} \mu_z) \Big]$$

$$+ \frac{1}{n'} (A \phi(A) - B \phi(B)) \Big[- B_{yz}^2 + 2 B_{yz} \rho_{yz} - B_{yx} B_{yz} \rho_{xz} \Big] \qquad ...(5.13)$$

5.3 Relative efficiency of t_5

The quantity g_5 in (5.12) is the variance of the estimator t_3 (Mukherjee et al 1987), the linear regression estimator using two auxiliary variables in double sampling, under the assumption that Σ is known and $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$. The relative efficiency of t_5 to t_3 is defined as

$$e_5 = \frac{\text{MSE } (t_3)}{\text{MSE } (t_5)} = \frac{g_5}{g_5 + h_5} \qquad .....(5.14)$$

The values of e_5 can be easily computed for different values of μ_x and μ_z. In general, e_5 has a maximum at $\mu_x = \mu_z = 0$. To give an idea about the behaviour of the relative efficiency function with respect to μ_x and μ_z, e_5 was computed for a set of values of n, n', α , B_{yx} and B_{yz}. These are presented in the following tables, Table 5.1 - 5.2.

Table 5.1 Behaviour of relative efficiency e_5 w.r.t. μ_x and μ_z

For $n = 10$, $n' = 30$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of e_5

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	1.490	1.323	1.040	0.858	0.802	0.841	0.943	1.038	1.096	1.118	1.123
0.1	1.416	1.185	0.915	0.754	0.703	0.726	0.791	0.843	0.863	0.857	0.841
0.2	1.262	1.028	0.796	0.663	0.620	0.637	0.684	0.717	0.723	0.711	0.692
0.3	1.136	0.917	0.718	0.605	0.572	0.590	0.634	0.664	0.669	0.658	0.640
0.4	1.087	0.873	0.687	0.586	0.561	0.586	0.639	0.676	0.688	0.681	0.666
0.5	1.110	0.878	0.689	0.605	0.573	0.611	0.680	0.733	0.756	0.756	0.747

Table 5.2 Behaviour of relative efficiency e_5 w.r.t. μ_x and μ_z

For $n = 10$, $n' = 30$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of e_5

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.6	1.175	0.906	0.704	0.610	0.595	0.646	0.737	0.812	0.851	0.860	0.856
0.7	1.232	0.922	0.708	0.610	0.605	0.667	0.775	0.868	0.920	0.937	0.938
0.8	1.264	0.919	0.699	0.603	0.603	0.672	0.791	0.895	0.955	0.976	0.981
0.9	1.276	0.904	0.683	0.590	0.593	0.667	0.793	0.902	0.967	0.990	0.995
1.0	1.279	0.884	0.664	0.575	0.582	0.658	0.788	0.902	0.969	0.993	0.998
1.1	1.279	0.865	0.646	0.560	0.570	0.649	0.782	0.900	0.968	0.993	0.999

Fig 5.1: Behaviour of relative efficiency e_5 w.r.t. μ_x and μ_z

For $n' = 30$, $n = 10$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$.

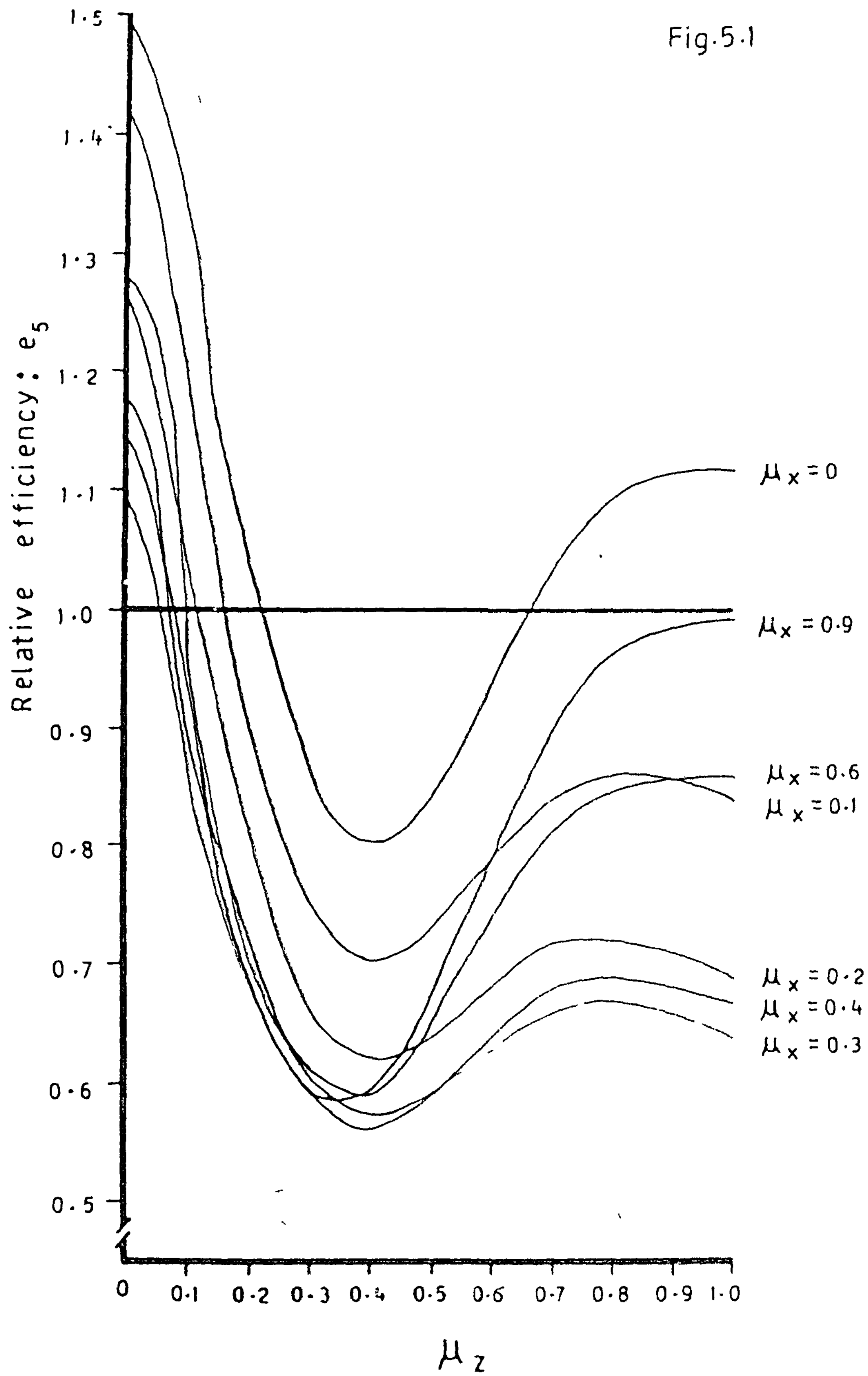


Fig.5.1
Relative efficiency: e_5
1·5
1·4
1·3
1·2
1·1
1·0
0·9
0·8
0·7
0·6
0·5
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
μ_z
$\mu_x = 0$
$\mu_x = 0{\cdot}9$
$\mu_x = 0{\cdot}6$
$\mu_x = 0{\cdot}1$
$\mu_x = 0{\cdot}2$
$\mu_x = 0{\cdot}4$
$\mu_x = 0{\cdot}3$

5.4 Optimum allocation of sample sizes

We next consider the following sample design problem. For a given cost function, what is the optimum allocation of the sample sizes n' and n ? Let the cost function be of the form

$$C = nc_1 + n'c_2 + n'c_3$$

where c_1, c_2, c_3 are the costs of observing y, x and z respectively. Or equivalently.

$$C = nc_1 + n'c_1' \quad \ldots\ldots.(5.15)$$

where $c_1' = c_2 + c_3$

The values of n and n' are obtained by minimising MSE (t_5) subject to the cost constraint (5.15). In general, the values of μ_x and μ_z are unknown, the experimenter has partial information about these and believes that both μ_x and μ_z are close to zero. When $\mu_x = \mu_z = 0$, the relative efficiency of t_5 is the largest. It would thus be reasonable to let $\mu_x = \mu_z = 0$ in MSE (t_5) and obtain the values of n' and n under the optimum situation.

Now, when $\mu_x = \mu_z = 0$, then

$a = A = Z_\alpha$ and

$b = B = - Z_\alpha$

which further implies that

$$\left.\begin{aligned} \Phi(a) &= \Phi(A) = 1 - \frac{\alpha}{2} \\ \text{and} \quad \Phi(b) &= \Phi(B) = \frac{\alpha}{2} \end{aligned}\right\} \qquad \ldots.(5.16)$$

Substituting (5.16) in (5.13) and simplifying we get,

$$M_5 = \frac{1}{n}(1 - \rho^2_{y.xz}) + \frac{1}{n'} \rho^2_{y.xz}$$

$$+ \frac{1-\alpha}{n'}(B^2_{yx} - 2 B_{yx} \rho_{yx} + B_{yx} B_{yz} \rho_{xz})$$

$$- \frac{2 Z_\alpha \phi(Z_\alpha)}{n'}(B^2_{yx} - 2 B_{yx} \rho_{yx} + B_{yx} B_{yz} \rho_{xz})$$

$$+ \frac{1-\alpha}{n'}(B^2_{yz} - 2 B_{yz} \rho_{yz} + B_{yx} B_{yz} \rho_{xz})$$

$$- \frac{2 Z_\alpha \phi(Z_\alpha)}{n'}(B^2_{yz} - 2 B_{yz} \rho_{yz} + B_{yx} B_{yz} \rho_{xz})$$

$$= \frac{1}{n}(1 - \rho^2_{y.xz}) + \frac{1}{n'} \rho^2_{y.xz}$$

$$- \frac{1-\alpha}{n'} \rho^2_{y.xz} + \frac{2 Z_\alpha \phi(Z_\alpha)}{n'} \rho^2_{y.xz}$$

$$= \frac{k}{n} + \frac{k'}{n'} \qquad \ldots\ldots\ldots\ldots(5.17)$$

Where $k = 1 - \rho^2_{y.xz}$

and $k' = (\alpha + 2 Z_\alpha \phi(Z_\alpha))\ \rho^2_{y.xz}$

In order to minimise (5.17) subject to (5.15) we have to minimise $M_5' C$ where

$$M_5 C = \left(\frac{k}{n} + \frac{k'}{n'} \right) (nc_1 + n'c_1') \qquad \text{......(5.18)}$$

which by Cauchy - Schwarz inequality is minimised when

$$\frac{k}{n^2 c_1} = \frac{k'}{n'^2 c_1'}$$

or

$$n = \frac{C\sqrt{k}}{\sqrt{c_1}\,(\sqrt{kc_1} + \sqrt{k'c_1'})} \qquad \text{......(5.19)}$$

and

$$n' = \frac{C\sqrt{k'}}{\sqrt{c_1'}\,(\sqrt{kc_1} + \sqrt{k'c_1'})} \qquad \text{........(5.20)}$$

Substituting (5.19) and (5.20) in (5.17) we have

$$M_{5',opt} = \frac{(\sqrt{kc_1} + \sqrt{k'c_1'})^2}{C} \qquad \text{.....(5.21)}$$

5.5 Comparison of the suggested estimator with other existing estimators

5.5.1 Comparison of $M_{5,opt}$ and $V_{opt}(t_3)$

We shall now first compare $M_{5,opt}$ with the minimum variance of the regression estimator with two auxiliary variables under double sampling. The variance of t_3 from (5.12) is

$$V(t_3) = \frac{1 - \rho^2_{y.xz}}{n} + \frac{\rho^2_{y.xz}}{n'}$$

under the assumption $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ and let as before $C = nc_1 + n'c_1'$. Again by Cauchy - Schwarz inequality VC is minimised when

$$\frac{1 - \rho^2_{y.xz}}{n^2 c_1} = \frac{\rho^2_{y.xz}}{n'^2 c_1'}$$

$$\text{or} \quad n = \frac{C\sqrt{1 - \rho^2_{y.xz}}}{\sqrt{c_1}\left(\sqrt{c_1'}\,\rho_{y.xz} + \sqrt{c_1(1 - \rho^2_{y.xz})}\right)} \qquad \ldots\ldots(5.22)$$

$$\text{and} \quad n' = \frac{C\,\rho_{y.xz}}{\sqrt{c_1'}\left(\sqrt{c_1'}\,\rho_{y.xz} + \sqrt{c_1(1 - \rho^2_{y.xz})}\right)} \qquad \ldots\ldots(5.23)$$

Substituting (5.22) and (5.23) in $V(t_3)$ we have

$$V_{opt}(t_3) = \frac{(\sqrt{c_1'}\ \rho_{y.xz} + \sqrt{c_1\ (1 - \rho^2_{y.xz})}\ )^2}{C} \quad \text{.....(5.24)}$$

In order to compare (5.21) and (5.24), we find that, since $\alpha + 2\ Z_\alpha\ \phi\ (Z_\alpha)$ is a decreasing function of Z_α with a maximum equal to unity at $Z_\alpha = 0$, the V_{opt} (t_3) is atleast as large as that of $M_{5,opt}$. Thus, we conclude that $M_{5,opt} \leq V_{opt}(t_3)$ with equality holding for $Z_\alpha = 0$, which is the case when the two estimators coincide.

5.5.2 Comparison of MSE (t_5) and MSE (t_2)

Next, we shall compare MSE (t_5) with MSE (t_2), the mean square error of the preliminary test estimator in double sampling with one auxiliary variable.

Now, from Han (1973)

$$M_2 = \text{MSE}\ (t_2) = \frac{1}{n}\ (1 - \rho^2_{yx}) + \frac{1}{n'}\ \rho^2_{yx}$$
$$+ \frac{\rho^2_{yx}}{n'}\ (a\ \phi\ (a) - b\ \phi\ (b)\ )$$
$$- \rho^2_{yx}\ (\frac{1}{n'} - \mu^2_x)\ (\ \Phi\ (a) - \Phi (b)\ ) \quad \text{...(5.25)}$$

where

$$a = Z_\alpha - \sqrt{n'}\ \mu_x \text{ and } b = -Z_\alpha - \sqrt{n'}\ \mu_x$$

In order to compare (5.17) with MSE (t_2) we must put $\mu_x = 0$ in (5.25) since relative efficiency of t_2 w.r.t. t_1 (with $\sigma_x^2 = \sigma_y^2 = 1$) is maximum at that point.

Therefore we have

$$M_2 = \frac{1}{n}(1 - \rho_{yx}^2) + \frac{\rho_{yx}^2}{n'} (\alpha + 2 Z_\alpha \phi(Z_\alpha)) \quad \text{....(5.26)}$$

Now, from (5.17) and (5.26) we get

MSE (t_5) - MSE (t_2)

$$= (\rho_{yx}^2 - \rho_{y.xz}^2) \left[\frac{1}{n} - \frac{1}{n'} (\alpha + 2 Z_\alpha \phi(Z_\alpha))\right] \quad \text{...(5.27)}$$

Since $\alpha + 2 Z_\alpha \phi(Z_\alpha)$ is a decreasing function of Z_α with maximum equal to unity at $Z_\alpha = 0$. Therefore,

$$\frac{1}{n} - \frac{1}{n'} (\alpha + 2 Z_\alpha \phi(Z_\alpha)) > 0$$

and $\rho_{yx}^2 \leqslant \rho_{y.xz}^2$

Thus from (5.27)

$$\text{MSE}(t_5) \leqslant \text{MSE}(t_2) \quad \text{.....(5.28)}$$

5.5.3 Comparison of MSE (t_5) and MSE (t_4)

Next we shall compare MSE (t_5) with the mean square error of t_4, the preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable.

From (5.17) and (3.17), we have

MSE (t_4) - MSE (t_5)

$$= \frac{(\rho^2_{y.xz} - \rho^2_{yz})}{n' \, (1 - \rho^2_{xz})} \left[1 - (\alpha + 2 \, Z_\alpha \, \phi(Z_\alpha)) \right] \qquad \text{.....(5.29)}$$

Now, since $\alpha + 2 \, Z_\alpha \, \phi(Z_\alpha)$ is a decreasing function of Z_α with a maximum equal to unity at $Z_\alpha = 0$ and $\rho^2_{yz} \leqslant \rho^2_{y.xz}$, therefore from (5.29) we get

$$\text{MSE } (t_5) \leqslant \text{MSE } (t_4) \qquad \text{.....(5.30)}$$

5.6 Discussion

From (5.28) we find that the efficiency of the preliminary test estimator in double sampling with partial information on two auxiliary variables, increases by utilising z - values in addition to x - values. This is also true for linear regression estimator in double sampling without using preliminary test. In the earlier sub - section i.e. in 5.5.1, we have also proved that under the optimum conditions mean square error of a preliminary test estimator in double sampling with two auxiliary variables is lesser than the mean square error of usual regression estimator in double sampling with two auxiliary variables. Again from (5.30) we find that the preliminary test estimator in double sampling with two auxiliary variables, having partial information on both the auxiliary variables is even better than the preliminary test estimator in double sampling with two auxiliary variables having partial information on only one auxiliary variable. Therefore, under the stated assumptions, preliminary test estimator in double sampling, with two auxiliary variables, having partial information on both the auxiliary variables is more efficient.

CHAPTER SIX

A generalized study of the preliminary test estimators in double sampling.

6.1 Introduction

In earlier chapters we have discussed about preliminary test estimators in double sampling with two auxiliary variables and their bias, relative efficiency, optimum allocation etc. under different assumptions. However, in all the cases, preliminary samples for estimating μ_x and μ_z were, even though independent, but were assumed to be of the same size n' ($> n$). In the present chapter an attempt will be made to consider the situation when the sizes of the preliminary samples for estimating μ_x and μ_z are different.

Suppose we are interested in estimating the population mean μ_y of a study variable Y. When information on an auxiliary variable X highly correlated with Y is readily available on all the units of the population, it is well known that ratio or regression-type estimators could be used for increased efficiency, incorporating the knowledge of μ_x. However, in certain practical situation μ_x is not known a priori in which case the technique of double sampling can be fruitfully applied. The values of X are assumed to be known over a large sample of size n' ($> n$). Now suppose that information on yet another variable Z is available. Again if μ_z is not known, assume that Z is known over another large sample of size $n''(>n')$. In such a situation an estimator using X and Z is being suggested by Mukherjee et al (1987) as follows :

$$t_6 = \bar{y}_n + b_{yx}(\bar{x}_{n'} - \bar{x}) + b_{yz}(\bar{z}_{n''} - \bar{z}_n) \quad \ldots.(6.1)$$

with

$$MSE\,(t_6) = V(t_6) = \frac{\sigma_y^{\,2}}{n}(1 - \rho^2_{y.xz}) + \frac{B^2_{yx}\,\sigma_x^{\,2}}{n'}$$

$$+ \frac{1}{n''}(2\,B_{yz}\,\sigma_{yz} - B^2_{yz}\,\sigma_z^{\,2}) \quad \ldots\ldots(6.2)$$

$$\text{where } \bar{z}_{n''} = \frac{1}{n''}\sum_{i=1}^{n''} z_i \quad \ldots\ldots(6.3)$$

and b_{yx}, B_{yx}, $\rho_{y.xz}$ etc. have the same definition as in earlier chapters.

6.2 Suggested preliminary test estimator

Now suppose we have partial information about μ_x and μ_z, then we may perform preliminary tests to construct preliminary test estimator. Let (X, Y, Z) have a trivariate normal distribution with mean (μ_x, μ_y, μ_z) and covariance matrix Σ in which the variances are denoted by $\sigma_x^{\,2}$, $\sigma_y^{\,2}$ and $\sigma_z^{\,2}$ and the correlation coefficients by ρ_{yx}, ρ_{yz} and ρ_{xz}. X and Z can be readily observed, while it is more expensive to observe the triplet (X, Y, Z). The problem is to estimate μ_y. Let (x_i, y_i, z_i), $i = 1, 2, \ldots, n$ be n independent observations on the triplet (X, Y, Z) which is supplemented by m more independent observations on X and another m' ($> m$) independent observations on Z where $n + m = n'$ and $n + m' = n''$.

If $\sum$ is known we may let $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ without loss of generality. The joint distribution of $(\bar{x}_n, \bar{y}_n, \bar{z}_n)$ is normal with mean (μ_x, μ_y, μ_z) and covariance matrix.

$$\sum = \frac{1}{n} \begin{bmatrix} 1 & \rho_{yx} & \rho_{xz} \\ \rho_{yx} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{bmatrix}$$

When μ_x and μ_z are unknown and the experimenter has partial information about them, he can employ preliminary tests for

$$H_{01} : \mu_x = 0 \text{ and } H_{02} : \mu_z = 0$$

(taking $\mu_{ox} = \mu_{oz} = 0$ without loss of generality)

The preliminary test estimator utilising partial information about two auxiliary variables X and Z is defined as

$$t_7=\begin{cases}\bar{y}_n - B_{yx}\bar{x}_n - B_{yz}\bar{z}_n & \text{if } |\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'}, |\bar{z}_{n''}| \leq Z_\alpha/\sqrt{n''} \\ \bar{y}_n + B_{yx}(\bar{x}_{n'} - \bar{x}_n) - B_{yz}\bar{z}_n & \text{if } |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}, |\bar{z}_{n''}| \leq Z_\alpha/\sqrt{n''} \\ \bar{y}_n - B_{yx}\bar{x}_n + B_{yz}(\bar{z}_{n''} - \bar{z}_n) & \text{if } |\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'}, |\bar{z}_{n''}| > Z_\alpha/\sqrt{n''} \\ \bar{y}_n + B_{yx}(\bar{x}_{n'} - \bar{x}_n) + B_{yz}(\bar{z}_{n''} - \bar{z}_n) & \text{if } |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'}, \\ & |\bar{z}_{n''}| > Z_\alpha/\sqrt{n''} \end{cases} \qquad ....(6.4)$$

where the components B_{yx}, B_{yz}, Z_α etc. have the same definition as in chapter 2.

6.3 Bias of the suggested estimator

To evaluate the bias of t_7, we require the joint distribution of $(\bar{x}_{n'}, \bar{x}_n, \bar{z}_{n''}, \bar{z}_n, \bar{y}_n)$. It can be easily verified that the joint distribution of these is a multivariate normal distribution with mean $(\mu_x, \mu_x, \mu_z, \mu_z, \mu_y)$ and covariance matrix

$$\begin{bmatrix} \frac{1}{n'} & \frac{1}{n'} & \frac{\rho_{xz}}{n'''} & \frac{\rho_{xz}}{n'} & \frac{\rho_{yx}}{n'} \\ \frac{1}{n'} & \frac{1}{n} & \frac{\rho_{xz}}{n'''} & \frac{\rho_{xz}}{n} & \frac{\rho_{yx}}{n} \\ \frac{\rho_{xz}}{n'''} & \frac{\rho_{xz}}{n'''} & \frac{1}{n''} & \frac{1}{n''} & \frac{\rho_{yz}}{n'''} \\ \frac{\rho_{xz}}{n'} & \frac{\rho_{xz}}{n} & \frac{1}{n''} & \frac{1}{n} & \frac{\rho_{yz}}{n} \\ \frac{\rho_{yx}}{n'} & \frac{\rho_{yx}}{n} & \frac{\rho_{yz}}{n'''} & \frac{\rho_{yz}}{n} & \frac{1}{n} \end{bmatrix} -$$

The derivation of the bias function of the estimator involves conditional expectations, the conditions being acceptance and rejection of the hypotheses H_{o1} and H_{o2}. Therefore,

$$
\begin{aligned}
E(t_7) = {} & E\left(t_7 \,\middle|\, |\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| \leq Z_\alpha/\sqrt{n''}\right) \\
& \times P\left(|\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| \leq Z_\alpha/\sqrt{n''}\right) \\
& + E\left(t_7 \,\middle|\, |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| \leq Z_\alpha/\sqrt{n''}\right) \\
& \times P\left(|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| \leq Z_\alpha/\sqrt{n''}\right) \\
& + E\left(t_7 \,\middle|\, |\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| > Z_\alpha/\sqrt{n''}\right) \\
& \times P\left(|\bar{x}_{n'}| \leq Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| > Z_\alpha/\sqrt{n''}\right) \\
& + E\left(t_7 \,\middle|\, |\bar{x}_{n'}| > Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| > Z_\alpha/\sqrt{n''}\right) \\
& \times P\left(|\bar{x}_{n'}| > Z_\alpha/\sqrt{n'},\ |\bar{z}_{n''}| > Z_\alpha/\sqrt{n''}\right)
\end{aligned}
$$

$$= \mu_y + \text{Bias}(t_7) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.5)$$

where

$$
\begin{aligned}
\text{Bias}(t_7) = B_7 = {} & \frac{B_{yx}}{\sqrt{n'}}\left(\phi(a) - \phi(b)\right) - B_{yx}\mu_x\left(\Phi(a) - \Phi(b)\right) \\
& + \frac{B_{yz}}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right) - B_{yz}\mu_z\left(\Phi(A) - \Phi(B)\right)
\end{aligned}
$$

$$\ldots.(6.6)$$

where $a = Z_{\alpha} - \sqrt{n'}\,\mu_x$, $b = -Z_{\alpha} - \sqrt{n'}\,\mu_x$

and $A = Z_{\alpha} - \sqrt{n''}\,\mu_z$, $B = -Z_{\alpha} - \sqrt{n''}\,\mu_z$

(for derivation refer Appendix A)

6.4 Mean square error of t_7

We know that,

$$M_7 = MSE\ (t_7) = E\ (t_7^2) - (E\ (t_7)\ )^2 + B_7^2 \qquad \ldots\ldots(6.7)$$

The derivation of MSE of the estimator involves conditional expectations of products, the conditions being acceptance or rejection of the hypotheses H_{o1} and H_{o2}.

Now, from (6.4)

$$E(t_7^2) = (\mu_y^2 + \tfrac{1}{n}) + B_{yx}^2\ (\mu_x^2 + \tfrac{1}{n}) + B_{yz}^2\ (\mu_z^2 + \tfrac{1}{n})$$

$$- 2\ B_{yx}(\mu_x \mu_y + \frac{\rho_{yx}}{n}) - 2\ B_{yz}(\mu_y \mu_z + \frac{\rho_{yz}}{n})$$

$$+ 2\ B_{yx}\ B_{yz}\ (\mu_x\ \mu_z + \frac{\rho_{xz}}{n})$$

$$- B_{yx}^2 \Big[(\mu_x^2 + \tfrac{1}{n'}) (1 - \Phi(a) + \Phi(b))$$

$$+ \frac{2\,\mu_x}{\sqrt{n'}}\ (\phi(a) - \phi(b))$$

$$+ \frac{1}{n'}\ (a\ \phi(a) - b\ \phi(b)) \Big]$$

$$
\begin{aligned}
&+ 2\, B_{yx} \Big[\big(\mu_y\, \mu_x + \frac{\rho_{yx}}{n'} \big) \big(1 - \Phi(a) + \Phi(b)\big) \\
&\qquad + \frac{1}{\sqrt{n'}} \big(\mu_y + \rho_{yx}\, \mu_x \big) \big(\phi(a) - \phi(b) \big) \\
&\qquad + \frac{\rho_{yx}}{n'} \big(a\, \phi(a) - b\, \phi(b) \big) \Big] \\
&- 2\, B_{yx}\, B_{yz} \Big[\big(\mu_x \mu_z + \frac{\rho_{xz}}{n'} \big) \big(1 - \Phi(a) + \Phi(b)\big) \\
&\qquad + \frac{1}{\sqrt{n'}} \big(\mu_z + \rho_{xz}\, \mu_x \big) \big(\phi(a) - \phi(b) \big) \\
&\qquad + \frac{\rho_{xz}}{n'} \big(a\, \phi(a) - b\, \phi(b) \big) \Big] \\
&- B_{yz}^2 \Big[\big(\mu_z^2 + \frac{1}{n''} \big) \big(1 - \Phi(A) + \Phi(B)\big) \\
&\qquad + \frac{2\, \mu_z}{\sqrt{n''}} \big(\phi(A) - \phi(B) \big) \\
&\qquad + \frac{1}{n''} \big(A\, \phi(A) - B\, \phi(B) \big) \Big] \\
&+ 2\, B_{yz} \Big[\big(\mu_y\, \mu_z + \frac{\rho_{yz}}{n''} \big) \big(1 - \Phi(A) + \Phi(B)\big) \\
&\qquad + \frac{1}{\sqrt{n''}} \big(\mu_y + \rho_{yz}\, \mu_z \big) \big(\phi(A) - \phi(B) \big) \\
&\qquad + \frac{\rho_{yz}}{n''} \big(A\, \phi(A) - B\, \phi(B) \big) \Big] \qquad \ldots.(6.8)
\end{aligned}
$$

Now, from (6.5) and (6.7)

$$
MSE(t_7) = E(t_7^2) - \mu_y^2 - 2\, \mu_y\, B_7 \qquad \ldots\ldots.(6.9)
$$

Substituting (6.6) and (6.8) in (6.9) and simplifying we get

$$MSE(t_7) = M_7 = g_7 + h_7 \qquad \ldots\ldots\ldots(6.10)$$

where

$$g_7 = \frac{1}{n}\left(1 - \rho^2_{y.xz}\right) + \frac{B^2_{yx}}{n'} + \frac{1}{n''}\left(2\, B_{yz}\, \rho_{yz} - B^2_{yz}\right)$$

$$\ldots\ldots(6.11)$$

and

$$h_7 = \left(\Phi(a) - \Phi(b)\right)\Big[B^2_{yx}\left(\mu^2_x + \frac{1}{n'}\right) - 2\, B_{yx}\frac{\rho_{yx}}{n'} + 2\, B_{yx}\, B_{yz}\left(\mu_x\mu_z + \frac{\rho_{xz}}{n'}\right)\Big]$$

$$+ \frac{1}{\sqrt{n'}}\left(\phi(a) - \phi(b)\right)\Big[-2\, B^2_{yx}\mu_x + 2\, B_{yx}\, \rho_{yx}\mu_x - 2\, B_{yx}\, B_{yz}\left(\mu_z + \rho_{xz}\mu_x\right)\Big]$$

$$+ \frac{1}{n'}\left(a\,\phi(a) - b\,\phi(b)\right)\Big[-B^2_{yx} + 2\, B_{yx}\, \rho_{yx} - 2\, B_{yx}\, B_{yz}\, \rho_{xz}\Big]$$

$$+ \left(\Phi(A) - \Phi(B)\right)\Big[B^2_{yz}\left(\mu^2_z + \frac{1}{n''}\right) - 2\, B_{yz}\frac{\rho_{yz}}{n''}\Big]$$

$$+ \frac{1}{\sqrt{n''}}\left(\phi(A) - \phi(B)\right)\Big[-2\, B^2_{yz}\mu_z + 2\, B_{yz}\, \rho_{yz}\, \mu_z\Big]$$

$$+ \frac{1}{n''}\left(A\,\phi(A) - B\,\phi(B)\right)\Big[-B^2_{yz} + 2\, B_{yz}\, \rho_{yz}\Big] \quad \ldots.(6.12)$$

6.5 Relative efficiency of t_7

The quality g_7 in (6.11) is the variance of t_6 given in (6.2), the linear regression estimator in double sampling, with $n' \neq n''$ under the assumption that Σ is known and $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$. The relative efficiency of t_7 to t_6 is defined as

$$e_7 = \frac{MSE\ (t_6)}{MSE\ (t_7)} = \frac{g_7}{g_7 + h_7} \quad ....(6.13)$$

The values of e_7 can be easily computed for different values of μ_x and μ_z. In general, e_7 has a maximum at $\mu_x = \mu_z = 0$. To give an idea about the behaviour of the relative efficiency function with respect to μ_x and μ_z, e_7 was computed for a set of values of n, n', n'', α , B_{yx} and B_{yz}. These are presented in the following tables, 6.4 and 6.5.

6.6 Optimum allocation of sample sizes

We next consider the following sample design problem. For a given cost function, what is the optimum allocation of the sample sizes n, n' and n'' ? Let the cost function be of the form

$$C = nc + n'c' + n''c'' \quad ....(6.14)$$

where c, c' and c" are the costs of observing Y, X and Z respectively.

The values of n, n' and n" are obtained by minimising $MSE(t_7)$ subject to the cost constraint (6.14). In general, the values of μ_x and μ_z are unknown, the experimenter has partial information about these and believes that both μ_x and μ_z are close to zero. When $\mu_x = \mu_z = 0$, the relative efficiency of t_7 is the largest. It would be reasonable to let $\mu_x = \mu_z = 0$ in $MSE(t_7)$ and obtain the values of n, n' and n" under the optimum situation.

Now when $\mu_x = \mu_z = 0$, then

$$\left.\begin{aligned} \Phi(a) &= \Phi(A) = 1 - \frac{\alpha}{2} \\ \text{and } \Phi(b) &= \Phi(B) = \frac{\alpha}{2} \end{aligned}\right\} \qquad \ldots\ldots(6.15)$$

Substituting (6.15) in (6.12) and simplifying we get

$$M_7 = \frac{1}{n}(1 - \rho^2_{y.xz}) + \frac{B_{yx}^2}{n'} + \frac{1}{n''}(2\,B_{yz}\,\rho_{yz} - B_{yz}^2)$$

$$- \frac{(1-\alpha)}{n'}\,B_{yx}^2 + \frac{1}{n'}\,2\,Z_\alpha\,\phi(Z_\alpha)\,B_{yx}^2$$

$$+ \frac{(1-\alpha)}{n''}(B_{yz}^2 - 2\,B_{yz}\,\rho_{yz}) - \frac{1}{n''}\,2\,Z_\alpha\,\phi(Z_\alpha)\,(B_{yz}^2 - 2B_{yz}\,\rho_{yz})$$

$$= \frac{k}{n} + \frac{k'}{n'} + \frac{k''}{n''} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.16)$$

where $k = 1 - \rho^2_{y.xz}$

$$k' = (\alpha + 2 Z_\alpha \Phi(Z_\alpha)) B^2_{yx}$$

and $$k'' = (\alpha + 2 Z_\alpha \Phi(Z_\alpha)) (2 B_{yz} \rho_{yz} - B^2_{yz})$$

In order to minimise (6.16) subject to (6.14), we have to minimise $M_\gamma C$ where

$$M_\gamma C = \left(\frac{k}{n} + \frac{k'}{n'} + \frac{k''}{n''} \right) (nc + n'c' + n''c'') \quad .....(6.17)$$

which by Cauchy - Schwarz inequality is minimised when

$$\frac{k}{n^2 c} = \frac{k'}{n'^2 c'} = \frac{k''}{n''^2 c''}$$

or $$n = \frac{C \sqrt{k}}{\sqrt{c} \, (\sqrt{kc} + \sqrt{k'c'} + \sqrt{k''c''})} \quad ......(6.18)$$

$$n' = \frac{C \sqrt{k'}}{\sqrt{c'} \, (\sqrt{kc} + \sqrt{k'c'} + \sqrt{k''c''})} \quad ....(6.19)$$

and $$n'' = \frac{C \sqrt{k''}}{\sqrt{c''} \, (\sqrt{kc} + \sqrt{k'c'} + \sqrt{k''c''})} \quad ......(6.20)$$

Substituting (6.18) - (6.20) in (6.16) we have

$$M_{7',opt} = \frac{(\sqrt{kc} + \sqrt{k'c'} + \sqrt{k''c''})^2}{C} \qquad \text{........(6.21)}$$

6.7 Comparison of the suggested estimator with other existing estimator

We shall now compare $M_{7',opt}$ with minimum variance of the regression estimator t_6 given in (6.1). The variance of t_6 from (6.2) is

$$V(t_6) = \frac{1}{n}(1 - \rho^2_{y.xz}) + \frac{B_{yx}^2}{n'} + \frac{1}{n''}(2\, B_{yz}\, \rho_{yz} - B_{yz}^2)$$

$$\text{......(6.22)}$$

under the assumption that $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ and let as before $C = nc + n'c' + n''c''$. Again by Cauchy - Schwarz inequality VC is minimised when

$$\frac{1 - \rho^2_{y.xz}}{n^2 c} = \frac{B_{yx}^2}{n'^2 c'} = \frac{2\, B_{yz}\, \rho_{yz} - B_{yz}^2}{n''^2 c''}$$

Therefore,

$$n = \frac{C\sqrt{1 - \rho^2_{y.xz}}}{\sqrt{c}\quad W} \qquad \text{.......(6.23)}$$

$$n' = \frac{C\, B_{yx}}{\sqrt{c'}\quad W} \qquad \text{......(6.24)}$$

and $$n'' = \frac{C\sqrt{2\, B_{yz}\, \rho_{yz} - B_{yz}^2}}{\sqrt{c''}\quad W} \qquad \text{.....(6.25)}$$

where

$$W = \sqrt{(1 - \rho^2_{y.xz})\, c} + \sqrt{B_{yx}^2\, c'} + \sqrt{(2\, B_{yz}\, \rho_{yz} - B_{yz}^2)\, c''}$$

Substituting (6.23) - (6.25) in (6.22) we get

$$V_{opt}(t_6) = \frac{\left(\sqrt{(1 - \rho^2_{y.xz})c} + \sqrt{B_{yx}^2\, c'} + \sqrt{(2\, B_{yz}\, \rho_{yz} - B_{yz}^2)c''}\right)^2}{C}$$

.....(6.26)

In order to compare (6.21) with (6.26) we find that, since $\alpha + 2\, Z_\alpha\, \phi(Z_\alpha)$ is a decreasing function of Z_α with a maximum equal to unity at $Z_\alpha = 0$, the $V_{opt}(t_6)$ is atleast as large as that of $M_{7,opt}$. Thus, we conclude that $M_{7,opt} \leqslant V_{opt}(t_6)$ with equality holding for $Z_\alpha = 0$, which is the case when the two estimators coincide.

6.8 Discussion

As partial checks it can be seen from (6.6) that

$$B_7 = - B_{yx}\, \mu_x - B_{yz}\, \mu_z \quad \text{when} \quad \alpha = 0$$

i.e. when we always accept H_o; and $B_7 = 0$ when $\alpha = 1$.

Further to give an idea about the bias with respect to μ_x and μ_z, we computed the values of B_7 (in absolute values) for a set of values of n', α, B_{yx}, B_{yz}, which are presented in the following tables, Table 6.1 - 6.3.

We notice that $B_7 = 0$, when $\mu_x = \mu_z = 0$. Also when μ_x, μ_z increases from 0, B_7 increases to a maximum, then decreases to zero. The bias is very close to zero at $\mu_x = \mu_z = 1$. The bias found here is quite small almost in all cases. The general behaviour of B_7 with respect to μ_x and μ_z is shown in Fig. 6.1

In section 5.7 we have proved that under the optimum conditions mean square error of a preliminary test estimator in double sampling with two auxiliary variables is lesser than the mean square error of usual regression estimator in double sampling with two auxiliary variables. Therefore, under the stated assumptions, preliminary test estimator is more efficient even when the preliminary sample sizes for the two auxiliary variables are different.

Table 6.1 Behaviour of Bias (t_7) w.r.t. μ_x and μ_z

For $n' = 30$, $n'' = 50$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yz} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_7|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.038	.055	.045	.023	.007	.002	0	0	0	0
0.1	.023	.061	.078	.068	.046	.030	.025	.023	.023	.023	.023
0.2	.038	.076	.093	.083	.061	.045	.040	.038	.038	.038	.038
0.3	.041	.079	.096	.086	.064	.048	.043	.041	.041	.041	.041
0.4	.032	.070	.087	.077	.055	.039	.034	.032	.032	.032	.032
0.5	.019	.057	.074	.064	.042	.026	.021	.019	.019	.019	.019
0.6	.009	.047	.064	.054	.032	.016	.011	.009	.009	.009	.009
0.7	.003	.041	.058	.048	.026	.010	.005	.003	.003	.003	.003
0.8	.001	.039	.056	.046	.024	.008	.003	.001	.001	.001	.001
0.9	0	.038	.055	.045	.023	.007	.002	0	0	0	0
1.0	0	.038	.055	.045	.023	.007	.002	0	0	0	0

Table 6.2 Behaviour of Bias (t_7) w.r.t. μ_x and μ_z

For $n' = 30$, $n'' = 50$, $\alpha = 0.10$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_7|$

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.029	.038	.026	.011	.003	.001	0	0	0	0
0.1	.017	.046	.055	.043	.028	.020	.018	.017	.017	.017	.017
0.2	.027	.056	.065	.053	.038	.030	.028	.027	.027	.027	.027
0.3	.027	.056	.065	.053	.038	.030	.028	.027	.027	.027	.027
0.4	.019	.048	.057	.045	.030	.022	.020	.019	.019	.019	.019
0.5	.009	.038	.047	.035	.020	.012	.010	.009	.009	.009	.009
0.6	.004	.033	.042	.030	.015	.007	.005	.004	.004	.004	.004
0.7	.001	.030	.039	.027	.012	.004	.002	.001	.001	.001	.001
0.8	0	.029	.038	.026	.011	.003	.001	0	0	0	0
0.9	0	.029	.038	.026	.011	.003	.001	0	0	0	0
1.0	0	.029	.038	.026	.011	.003	.001	0	0	0	0

Table 6.3 Behaviour of Bias (t_7) w.r.t. μ_x and μ_z

For $n' = 30$, $n'' = 50$, $\alpha = 0.25$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of $|B_7|$

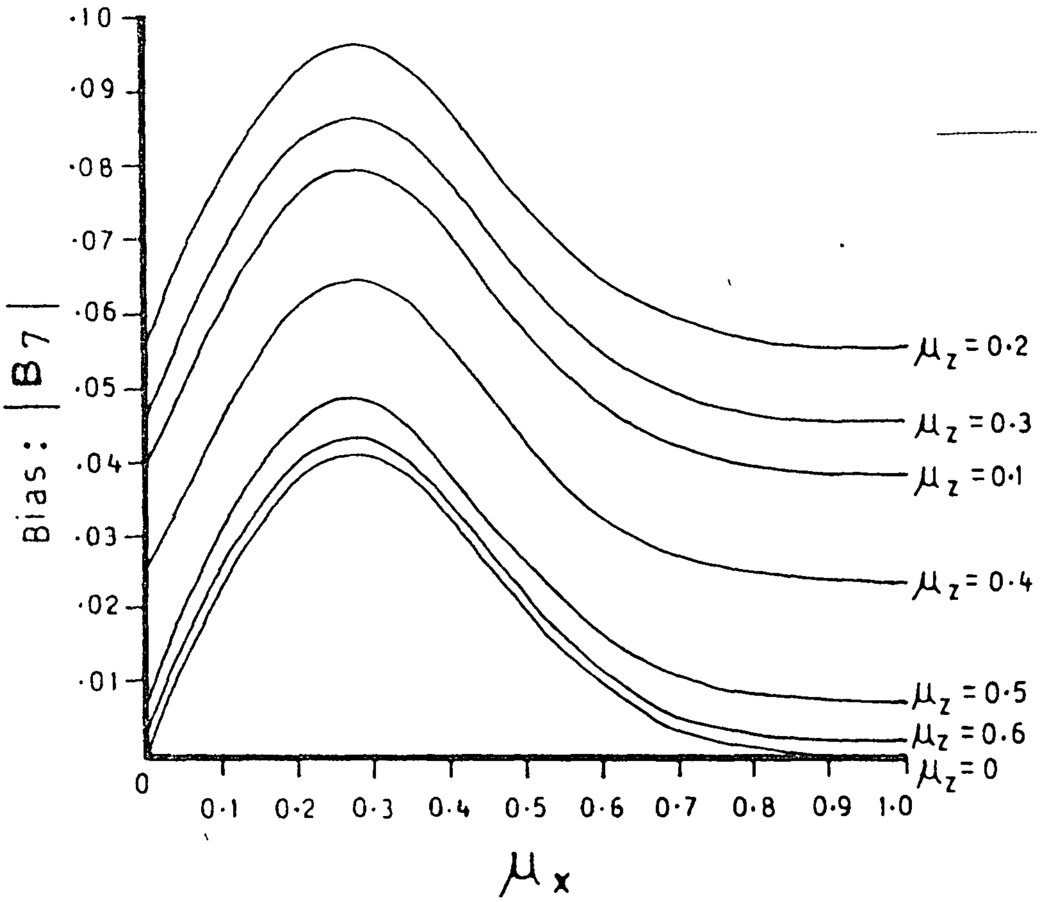
μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	.013	.015	.009	.003	.001	0	0	0	0	0
0.1	.009	.022	.024	.018	.012	.010	.009	.009	.009	.009	.009
0.2	.012	.025	.027	.021	.015	.013	.012	.012	.012	.012	.012
0.3	.010	.023	.025	.019	.013	.011	.010	.010	.010	.010	.010
0.4	.006	.019	.021	.015	.009	.007	.006	.006	.006	.006	.006
0.5	.002	.015	.017	.011	.005	.003	.002	.002	.002	.002	.002
0.6	.001	.014	.016	.010	.004	.002	.001	.001	.001	.001	.001
0.7	0	.013	.015	.009	.003	.001	0	0	0	0	0
0.8	0	.013	.015	.009	.003	.001	0	0	0	0	0
0.9	0	.013	.015	.009	.003	.001	0	0	0	0	0
1.0	0	.013	.015	.009	.003	.001	0	0	0	0	0

Fig 6.1: Behaviour of Bias (t_7) w.r.t. μ_x and μ_z

For $n' = 30$, $n'' = 50$, $\alpha = 0.05$, $\rho_{xz} = 0.6$,

$\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$.

Fig.6.1



·10
·09
·08
·07
·06
·05
·04
·03
·02
·01
Bias : |B7|
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
μ_x
$\mu_z = 0·2$
$\mu_z = 0·3$
$\mu_z = 0·1$
$\mu_z = 0·4$
$\mu_z = 0·5$
$\mu_z = 0·6$
$\mu_z = 0$

Table 6.4 Behaviour of relative efficiency e_7 w.r.t. μ_x and μ_z

For $n = 10$, $n' = 30$, $n'' = 50$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of e_7

μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	1.346	1.177	0.930	0.829	0.870	0.971	1.041	1.064	1.068	1.068	1.068
0.1	1.297	1.062	0.813	0.703	0.700	0.729	0.733	0.712	0.683	0.655	0.630
0.2	1.188	0.950	0.724	0.620	0.602	0.609	0.597	0.569	0.539	0.511	0.485
0.3	1.098	0.886	0.683	0.587	0.569	0.573	0.560	0.534	0.505	0.479	0.455
0.4	1.069	0.885	0.694	0.604	0.593	0.606	0.600	0.578	0.552	0.527	0.504
0.5	1.096	0.932	0.740	0.653	0.654	0.685	0.693	0.679	0.657	0.636	0.616

Table 6.5 Behaviour of relative efficiency e_7 w.r.t. μ_x and μ_z

For $n = 10$, $n' = 30$, $n'' = 50$, $\alpha = 0.05$, $\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$

Values of e_7

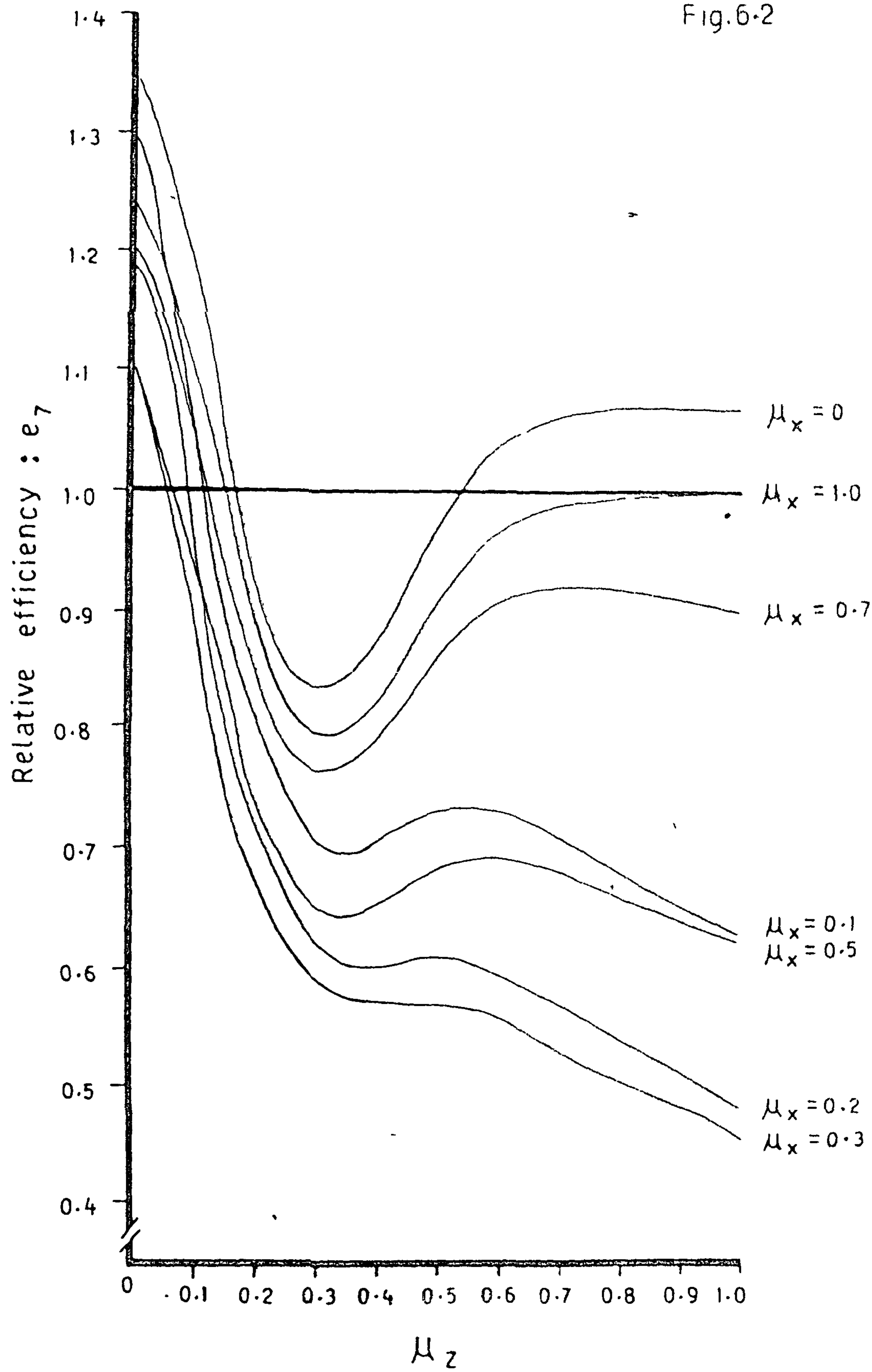
μ_x \ μ_z	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.6	1.155	1.003	0.802	0.715	0.712	0.789	0.817	0.816	0.804	0.789	0.775
0.7	1.203	1.055	0.846	0.757	0.786	0.862	0.909	0.920	0.917	0.910	0.903
0.8	1.229	1.082	0.868	0.779	0.813	0.899	0.956	0.974	0.976	0.974	0.972
0.9	1.238	1.092	0.875	0.785	0.821	0.911	0.972	0.991	0.995	0.994	0.994
1.0	1.241	1.094	0.877	0.787	0.823	0.914	0.975	0.995	0.999	0.999	0.999
1.1	1.241	1.094	0.877	0.787	0.824	0.914	0.976	0.996	1.000	1.000	1.000

Fig 6.2: Behaviour of relative efficiency e_7 w.r.t. μ_x and μ_z

For $n = 10$, $n' = 30$, $n'' = 50$, $\alpha = 0.05$,

$\rho_{xz} = 0.6$, $\rho_{yx} = 0.7$, $\rho_{yz} = 0.8$.

Fig.6·2



1·4
1·3
1·2
1·1
1·0
0·9
0·8
0·7
0·6
0·5
0·4
Relative efficiency : e_7
$\mu_x = 0$
$\mu_x = 1{\cdot}0$
$\mu_x = 0{\cdot}7$
$\mu_x = 0{\cdot}1$
$\mu_x = 0{\cdot}5$
$\mu_x = 0{\cdot}2$
$\mu_x = 0{\cdot}3$
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
μ_z

CHAPTER SEVEN

Conclusion

7.1 Introduction

In earlier Chapters we suggested a number of preliminary test estimators in double sampling under varied situations. The efficiency of these suggested estimators as compared to other commonly used estimators is also being studied. In the present chapter a comparative study will be done on these suggested estimators, theoritically as well as empirically. The empirical studies are done mainly to show the practical applications of the suggested estimators and to demonstrate the performance of these estimators as compared with other existing estimators.

7.2 The estimator t_4

As one may recall we have defined the estimator t_4 in chapter 2, as a preliminary test estimator with two auxiliary variables having partial information on only one auxiliary variable. The estimator was found to be a biased one. From the Tables 2.1 - 2.4 and Fig. 2.1, it follows that in all the cases Bias $(t_4) = 0$ when $\mu_z = 0$. As μ_z increases Bias (t_4) increases to a maximum, then decreases to zero. Bias $(t_4) = 0$ when μ_z is very close to 1. We may also notice from Fig. 2.1 that with larger values of α Bias (t_4) has comparatively lower peak.

Next the mean square error M_4 of t_4 was obtained and relative efficiency e_4 of t_4 with respect to t_3, the linear regression estimator in double sampling with two auxiliary variables, was computed. We notice from Tables 3.1 - 3.4 and also from Fig. 3.1 that e_4 is maximum at $\mu_z = 0$. As μ_z increases, e_4 decreases to a minimum and then increases to unity. Also it is found that e_4 is very close to 1 at $\mu_z = 1$. Further, at $\mu_z = 0$, where e_4 is maximum, $M_4 \leq V(t_3)$ provided $\rho_{yz}^2 > \rho_{y.xz}^2 \; \rho_{xz}^2$. That is the preliminary test estimator in double sampling with two auxiliary variables having partial information on one auxiliary variable is more efficient than the regression estimator in double sampling with two auxiliary variables, but with no partial information.

7.3 The estimator t_5

This estimator as suggested in chapter 4, is a preliminary test estimator with two auxiliary variables having partial information on both the auxiliary variables. We have noticed that like t_4, t_5 is also a biased estimator. Bias (t_5) is a function of μ_x and μ_z. From Tables 4.1 - 4.6 and Fig. 4.1 it follows that when $\mu_x = \mu_z = 0$, Bias (t_5) = 0. As μ_x, μ_z increase, Bias (t_5) increases to a maximum, the decreases to zero. Again Bias (t_5) = 0 when μ_x and μ_z are very close to 1. Thus we observe that the behaviour of Bias (t_5) is very much similar to that of Bias (t_4). The bias is very negligible in both the cases. It is also observed that, as in the case of Bias (t_4), Bias (t_5) also decreases with an increases in α

Next the mean square error M_5 of t_5 was obtained from which e_5, the relative efficiency of t_5 was computed. Further we notice from Tables 5.1 - 5.2 and also Fig. 5.1 that e_5 is maximum at $\mu_x = \mu_z = 0$. As μ_x, μ_z increase e_5 decreases to a minimum and then increases to unity. Also e_5 is very close to 1 at $\mu_x = \mu_z = 1$. Thus, the general behaviour of e_4 and e_5 are very much similar. We have also shown that for $\mu_x = \mu_z = 0$, the following inequalities hold good.

$$M_5 \leqslant V(t_3)$$

$$M_5 \leqslant M_2$$

$$M_5 \leqslant M_4$$

Thus, the preliminary test estimator in double sampling with two auxiliary variables having partial information on both is more efficient than the regression estimator in double sampling with two auxiliary variables without any partial information. It is also more efficient than preliminary test estimator in double sampling with one auxiliary variable. This is also true incase of usual regression estimators that inclusion of an additional variable leads to more efficient estimator. Finally t_5

is also better than t_4 where we have partial information on one of the two auxiliary variables.

7.4 <u>The estimator t_7</u>

This is also a preliminary test estimator with two auxiliary variables with partial information on both the auxiliary variables. However, unlike before, here the sizes of the preliminary samples for the two auxiliary variables differ from each other. In this case we notice that behaviour of the Bias (t_7) is very much similar to that of Bias (t_5). As is seen in Table 6.1 - 6.3 and Fig. 6.1 Bias (t_7) = 0 when $\mu_x = \mu_z = 0$. As μ_x, μ_z increase, Bias (t_7) increases to a maximum and again decreases to zero. Bias (t_7) = 0 when μ_x and μ_z are very close to 1.

Relative efficiency e_7 of t_7 was obtained with respect to t_6, the linear regression estimator in double sampling, with two auxiliary variables, with preliminary samples of unequal sizes. Behaviour of e_7 is being shown in Tables 6.4 - 6.5 and Fig. 6.2. Here also e_7 is maximum at $\mu_x = \mu_z = 0$. As μ_x, μ_z increase, e_7 decreases to a minimum and then increases to unity. Also e_7 is very close to 1 at $\mu_x = \mu_z = 1$. Further, we have shown that at $\mu_x = \mu_z = 0$, $M_7 \leqslant V(t_6)$. That is t_7 is more efficient than t_6.

7.5 Empirical studies

Many authors have done empirical studies to show the application of the estimators in double sampling. They have also demonstrated the performance of these estimators as compared to other existing estimators. We also have made some such studies to demonstrate the performance of the preliminary test estimators suggested by us.

7.5.1 Empirical studies for double sampling

Som (1973) has shown the application of double sampling for estimating the average number of cattle per farm by using a regression estimator, taking area per farm as an auxiliary variable. He has obtained the average area per farm from the first phase sample. He has further shown that the estimated standard error for the regression estimator in double sampling is 0.1891 whereas if no account is taken of the information in the first phase sample then the estimated standard error for the average number of cattle per farm is 0.2548.

7.5.2 Empirical studies for double sampling with two auxiliary variables

Shukla (1966) has shown application of double sampling for two auxiliary variables. In order to estimate the mean yield of the jute fibre per plant, he considered two auxiliary variables namely height and base diameter, both of which are correlated with the yield of the fibre. The data was obtained from Jute Agricultural Research Institute Farm, Barrackpore, from plants sown in the year 1962 - 63. Using a regression estimator with two auxiliary variables in double sampling, he has shown that the suggested method is superior to Olkin's (1958) ratio method.

Mukherjee et al (1987) have shown a number of applications of double sampling with two auxiliary variables. They have used the following Data sets for comparing some regression-type estimators.

Data sets 1 - 3 : These are based on data sets 1 - 3 in Kiregyera (1984) relating to the 1959 and 1964 censuses of agriculture for the state of Iowa, U.S.A.

Data set 1 :

y : acres of corn harvested for grain, 1964,

x : acres under corn, 1964,

z : acres under corn, 1959.

Data set 2 :

y : bushels of corn harvested, 1964,

x : acres under corn, 1964,

z : acres of corn harvested for grain, 1959.

Data set 3 :

y : bushels of soy beans harvested for beans, 1964,

x : acres of soybeans harvested for beans, 1964,

z : acres for soybeans harvested for beans, 1959.

Data set 4 : This is based on data from 1961 and 1971 censuses on 63 cities and urban agglomerations with population 200,000 and above. Here

y : number of workers, 1971,

x : population, 1971, z : population, 1961.

Data set 5 : This is based on data on 80 Indian factories as recorded by Murthy (1967). Here

y : output, x : number of workers,

z : fixed capital.

7.5.3 Empirical studies for preliminary test estimators

In this subsection we have considered the following data sets to show the practical applications of the suggested estimators and also to compare their performance as compared to other estimators.

Data set 1 :

This constitutes the data collected by us from a local forest, representing social forestry plantation, for estimating the leaf area of a broad leaved fodder tree species Exbucklandia populnea (Griff.) R.W.Br. Two hundred leaves were collected and for measuring the leaf area a very popular method was followed. Firstly, the leaves were drawn on graph sheets and the number of squares covered were counted to the nearest 0.5 mm^2. The length and the width of the leaves were also measured from these figures, and were considered as two auxiliary variables. Both these variables have high correlation with leaf area and are easy to measure as compared to leaf area. In fact, one can measure these two characters even without ~~plucking~~ the leaves from the trees. Leaf area estimation with auxiliary characters such as length, width etc. is thus very common in practice. Here we take

y : leaf area, x : leaf length,

z : leaf width.

Data set 2 : This is based on data on 80 Indian factories as recorded by Murthy (1967). Here

y : output, x : number of workers,

z : fixed capital.

In order to show the application of the various preliminary test estimators suggested in the earlier chapters and to judge their performance, M_4, M_5, M_7, M_2 at $\mu_x = \mu_z = 0$ and $V(t_3)$, $V(t_6)$ etc. were computed from the above two data sets and are recorded in Table 7.1.

Table 7.1 Comparative study of the mean square errors of the suggested estimators and other estimators based on some empirical studies.

Data Set	ρ_{yx}	ρ_{yz}	ρ_{xz}	M_4	M_5	M_7	M_2	$V(t_3)$	$V(t_6)$
1	.948	.981	.954	.0130	.0126	.0104	.0127	.0357	.0278
2	.915	.941	.988	.0301	.0187	.0174	.0197	.0402	.0358

As we may see from Table 7.1 that for both the data sets

$$\rho^2_{yz} > \rho^2_{y.xz} \quad \rho^2_{xz}$$

Therefore, for both the data sets

$$M_4 < V(t_3) \quad \ldots\ldots\ldots\ldots(7.1)$$

Again, we find from table 7.1 that in both the cases following are true

$$M_5 < V(t_3) \quad \ldots\ldots\ldots(7.2)$$

$$M_5 < M_2 \quad \ldots\ldots\ldots\ldots(7.3)$$

and $$M_5 < M_4 \quad \ldots\ldots\ldots\ldots(7.4)$$

Finally we notice that for both the situations,

$$M_7 < V(t_6) \quad \ldots\ldots\ldots\ldots(7.5)$$

Summarising (7.1) - (7.5) we can say that except (7.1), rest of the inequalities are true under all conditions at $\mu_x = \mu_z = 0$. Again from (7.4) we have $M_5 < M_4$, suggesting that amongst the preliminary test estimators with two auxiliary variables, the one with partial information on both the auxiliary variables is more efficient. Finally when we have partial information on both the auxiliary variables and preliminary samples are of different sizes then one can use t_7 which is more efficient than t_5 as is clear from (7.5). Thus we can conclude that preliminary test estimators are more preferred than the usual regression estimators in double sampling whenever we have partial information on one or more auxiliary variables.

References

Bancroft, T.A. (1944) : On biases in estimation due to the use of preliminary test of significance Ann. Math. Statist, 15, 190 - 204.

Bancroft, T.A. (1964) : Analysis and inference for incompletely specified models involving the use of preliminary test (s) of significance, Biometrics, 20, 427 - 442.

Bedi, P.K. and Hajela, D.(1984) : An estimator for mean utilizing known coefficient of variation and auxiliary variate, J. Statist Res, 18, 29 - 33.

Bedi, P.K. (1985) : On two-phase multivariate sampling estimator, JISAS, 37, 158 - 162.

Bock, M.E.; Yancey, T.A. and Judge, G.G.(1973) : The statistical consequences of preliminary test estimators in regression, JASA, 68, 109 - 116.

Bose, C.(1943) : Note on the sampling error in the method of double sampling, Sankhya, 6, 329 - 30.

Chand, L. (1975) : Some ratio-type estimators based on two or more auxiliary variables. Unpublished Ph.D. dissertation, Iowa State University, Ames, Iowa.

Cochran, W.G. (1977) : Sampling Techniques (3rd ed.) New York, John Wiley.

Das, A.K. and Tripathi, T.P. (1980) : Sampling stratergies for population mean when the coefficient of variation of an auxiliary character is known, Sankhya, 42 (c), 76 - 86.

deGraft - Johnson, K.T. and Sedransk, J. (1973) : Estimation of domain means using two-phase sampling. Biometrika, 60, 387 - 393.

Deng, L.Y. (1984) : Statistical inference in finite population sampling when auxiliary information is available, unpublished Ph.D. thesis, University of Wisconsin, Madison, Department of Statistics.

Deng, L.Y. and Wu, C.F.J. (1987) : Estimation of variance of the regression estimator, JASA, 82, 568 - 576.

Esimai, G.O. and Han, C.P. (1982) : Regression estimation with double sample and partial information, JISA, 20, 1 - 7.

Grimes, J,E. and Sukhatme, B.V. (1980) : A regression-type estimator based on preliminary test of significance, JASA, 75, 957 - 962.

Gupta. P.C. (1978) : On some quadratic and higher degree ratio and product estimators, JISAS, 30, 71 - 80.

Han, C.P. (1973) : Double sampling with partial information on auxiliary variables, JASA, 68, 914 - 918.

Hartley, H.O. and Ross, A. (1954) : Unbiased ratio estimates, Nature, 174, 270 - 271.

Khan, S. and Tripathi, T.P. (1967) : The use of multivariate auxiliary information in double sampling, JISA, 5, 43-48.

Khatri, C.G., Bhargava, R.P. and Shah, K.R. (1974) : Distribution of regression estimate in double ~~sampling,~~ Sankhya, 36 (c), 3 - 22.

Kiregyera, B.(1980) : A chain ratio-type estimator in finite population double sampling using two auxiliary variables. Metrika, 27, 217 - 223.

Kiregyera, B.(1984) : Regression-type estimators using two auxiliary variables and the model of double sampling from finite population, Metrika, 31, 215 - 226.

Kitagawa, T.(1963) : Estimation after preliminary tests of significance, University of California, Publications in Statistics, 3, 147 - 186.

Lahiri, D.B.(1951) : A method of sample selection providing unbiased ratio estimates, Intern. Stat. Inst. Bull, 33, 133 - 140.

Mosteller, F.(1948) : On pooling data, JASA, 43, 231 - 242.

Mukherjee, R. ; Rao, T.J. and Vijayan, K. (1987) : Regression-type estimators using multiple auxiliary information, Austral. J. Statist., 29, 244 - 254.

Murthy, M.N. (1977) : 'Sampling Theory and Methods', Statistical Publishing Society, Calcutta.

Olkin, I. (1958) : Multivariate ratio estimates for finite population, Biometrika, 45, 154 - 165.

Raj, D. (1964) : On double sampling for pps estimation. Ann. Math. Statist., 35, 900 - 902.

Raj, D. (1965) : On sampling over two occassions with pps, Ann. Math. Statist., 36, 327 - 330.

Raj, D. (1968) : 'Sampling Theory', Mc Graw-Hill, TMH edition.

Rao, J.N.K. (1973) : On double sampling for stratification and analytical surveys, Biometrika, 60, 125 - 133.

Rao, P.S.R.S. (1972) : On two-phase regression estimator, Sankhya, 34 A, 473 - 476.

Rao, P.S.R.S. (1975) : Hartley-Ross type estimators with two phase sampling, Sankhya, 37 (c), 140 - 146.

Rao, P.S.R.S. (1975 a) : On the two-phase ratio estimator in finite populations, JASA, 70, 839 - 845.

Rao, P.S.R.S. (1981) : Efficiencies of nine two-phase ratio estimators for the mean, JASA, 76, 434 - 442.

Royall, R.M. (1970) : On finite population sampling theory under certain linear regression models, Biometrika, 57, 377 - 387.

Searls, D.T. (1964) : The utilization of known coefficient of variation in the examination procedure, JASA, 59, 1225 - 26.

Sedransk, J. (1965) : A double sampling scheme for analytical surveys, JASA, 60, 985 - 1004.

Sedransk, J. and Singh, B. (1974) : Sample size allocation in two-phase sampling. Comm. Statist., 3 (11), 1025 - 1040.

Sen, A.R. (1972) : Successive sampling with p ($p \geqslant 1$) auxiliary variables, Ann. Math. Statist., 43, 2031 - 2034.

Sen, A.R. (1973) : Theory and application of sampling on repeated occasions with several auxiliary variables, Biometrics, 29, 381 - 385.

Sen, A.R. (1973 a) : Some theory of sampling on successive occasions, Aust. J. Statist., 15, 105 - 110.

Shukla, G.K. (1966) : Multivariate regression estimate, JISA, 4, 202 - 211.

Singh, B.D. and Singh, D. (1965) : Some remarks on double sampling for stratification, Biometrika, 52, 587 - 590.

Singh, D. and Singh, B.D. (1965) : Some contributions to two-phase sampling, Austral. J. Statist., 7, 45 - 47.

Singh, H.P. and Namjoshi, U.D. (1988) : A class of multivariate regression estimators in two-phase sampling. Assam Statistical Review, 2, 91 - 97.

Singh, M.P. (1965) : On the estimation of ratio and product of the population parameters, Sankhya, 27 (B), 321 - 328.

Singh, R. (1984) : Double sampling for two auxiliary characters. Calcutta Statist. Assoc. Bull., 33, 193 - 197.

Singh, R.K. (1982) : Generalized double sampling estimators for the ratio and product of population parameters, JISA, 20, 39 - 49.

Sisodia, B.V.S. (1981) : On use of preliminary tests of significance in repeated surveys, JISAS, 33, 47 - 66.

Sisodia, B.V.S. and Srivastava, A.K. (1982) : Modifying regression estimators with a preliminary test in double sampling, Sankhya, 44 (B), 295 - 303.

Som, R.K. (1973) : 'A Manual for Sampling Techniques' Heinemann Educational Books Ltd., London.

Srivastava, S.K. (1970) : A two-phase sampling estimator in sample surveys. Aust. Jour. Statist., 12, 23 - 27.

Srivastava, S.K. (1970) : A two-phase sampling estimator in sample surveys, Aust. J. Statist., 12, 23 - 27.

Srivastava, S.K. (1981) : A generalized two-phase sampling estimator, JISAS, 33, 38 - 46.

Sukhatme, B.V. (1962) : Some ratio-type estimators in two-phase sampling, JASA, 57, 628 - 632.

Sukhatme , B.V., and Tang Victor K.T. (1975) : Allocation in stratified sampling subsequent to preliminary test of significance, JASA, 70, 175 - 179.

Tamhane, A.C. (1978) : Inference based on regression estimator in double sampling, Biometrika, 65, 419 - 427.

Tripathi, T.P. (1968) : The use of multivariate auxiliary information in sample surveys when the units are selected with pps and with replacement. Presented at the 55th annual session of 'Indian Science Congress' held at Varanasi in early January 1968.

Tripathi, T.P. (1969) : A regression-type estimator in sampling with replacement, Austral J. Statist., 11, 140-148.

Tripathi, T.P. (1970) : Double sampling for regression method of estimation using pps wr selection. A portion of chapter II of author's Ph.D. thesis, 'Some Contributions to the Sampling Theory Using Multivariate Information' submitted to Punjabi University, Patiala, India.

Tripathi, T.P. (1973) : Double sampling for inclusion probabilities and regression method of estimation, JISA, 10, 33 - 46.

Tripathi, T.P. (1976) : On double sampling for multivariate ratio and difference methods of estimation, JISAS, 28, 33 - 54.

Tripathi, T.P. (1988) : A generalised method of estimation in double sampling, JISA, 26, 91 - 101.

Tripathi, T.P. (1989) : Improved estimators for population mean based on double sampling, JISA, 1989.

Appendix A

Derivation of

$$E\left(\bar{z}_{n'} \,\Big|\, |\bar{z}_{n'}| > Z_{\alpha}/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_{\alpha}/\sqrt{n'}\right)$$

Let $I = E\left(\bar{z}_{n'} \,\Big|\, |\bar{z}_{n'}| > Z_{\alpha}/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_{\alpha}/\sqrt{n'}\right)$

$$= E\left(\bar{z}_{n'} \,\Big|\, \bar{z}_{n'} > Z_{\alpha}/\sqrt{n'}\right) P\left(\bar{z}_{n'} > Z_{\alpha}/\sqrt{n'}\right)$$

$$+ E\left(\bar{z}_{n'} \,\Big|\, \bar{z}_{n'} \leqslant -Z_{\alpha}/\sqrt{n'}\right) P\left(\bar{z}_{n'} \leqslant -Z_{\alpha}/\sqrt{n'}\right)$$

$$= \int_{Z_{\alpha}/\sqrt{n'}}^{\infty} \bar{z}_{n'}\, f(\bar{z}_{n'})\, d\bar{z}_{n'} + \int_{-\infty}^{-Z_{\alpha}/\sqrt{n'}} \bar{z}_{n'}\, f(\bar{z}_{n'})\, d\bar{z}_{n'}$$

Now, we know that

$$\bar{z}_{n'} \sim N\left(\mu_z, 1/\sqrt{n'}\right)$$

Therefore,

$$I = \frac{\sqrt{n'}}{\sqrt{2\pi}} \left[\int_{Z_{\alpha}/\sqrt{n'}}^{\infty} \bar{z}_{n'} \exp(-1/2) \left[\frac{z_{n'} - \mu_z}{1/\sqrt{n'}}\right]^2 d\bar{z}_{n'} \right.$$

$$\left. + \int_{-\infty}^{-Z_{\alpha}/\sqrt{n'}} \bar{z}_{n'} \exp(-1/2) \left[\frac{z_{n'} - \mu_z}{1/\sqrt{n'}}\right]^2 d\bar{z}_{n'} \right]$$

Putting $w = \dfrac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}$, which implies

$dw = \sqrt{n'}\ d\bar{z}_{n'}$, we have

$$I = \frac{1}{\sqrt{2\pi}} \left[\int_A^\infty (\mu_z + w/\sqrt{n'})\, e^{-\frac{1}{2}w^2}\, dw \right.$$

$$\left. + \int_{-\infty}^{B} (\mu_z + w/\sqrt{n'})\, e^{-\frac{1}{2}w^2}\, dw \right]$$

where $A = Z_\alpha - \sqrt{n'}\,\mu_z$

and $B = -Z_\alpha - \sqrt{n'}\,\mu_z$

Therefore,

$$I = \frac{\mu_z}{\sqrt{2\pi}} \left[\int_A^\infty e^{-\frac{1}{2}w^2}\, dw + \int_{-\infty}^{B} e^{-\frac{1}{2}w^2}\, dw \right]$$

$$+ \frac{1}{\sqrt{2\pi n'}} \left[\int_A^\infty w\, e^{-\frac{1}{2}w^2}\, dw + \int_{-\infty}^{B} w\, e^{-\frac{1}{2}w^2}\, dw \right]$$

Again putting $\frac{w^2}{2} = t \Rightarrow w\, dw = d\,t$, we have

$$I = \mu_z \,(1 - \Phi(A) + \Phi(B))$$

$$+ \frac{1}{\sqrt{2\pi n'}} \left[\int_{A^2/2}^{\infty} e^{-t}\, dt + \int_{\infty}^{B^2/2} e^{-t}\, dt \right]$$

$$= \mu_z \,(1 - \Phi(A) + \Phi(B)) + \frac{1}{\sqrt{2\pi n'}} \,(e^{-A^2/2} - e^{-B^2/2})$$

$$= \mu_z \,(1 - \Phi(A) + \Phi(B)) + \frac{1}{\sqrt{n'}} \,(\phi(A) - \phi(B))$$

Appendix B

Derivation of

$$E\left(\bar{z}_{n'}^2 \middle| |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

Let $I = E\left(\bar{z}_{n'}^2 \middle| |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$

$$= E\left(\bar{z}_{n'}^2 \middle| \bar{z}_{n'} > Z_\alpha/\sqrt{n'}\right) P\left(\bar{z}_{n'} > Z_\alpha/\sqrt{n'}\right)$$

$$+ E\left(\bar{z}_{n'}^2 \middle| \bar{z}_{n'} \leqslant - Z_\alpha/\sqrt{n'}\right) P\left(\bar{z}_{n'} \leqslant - Z_\alpha/\sqrt{n'}\right)$$

$$= \int_{Z_\alpha/\sqrt{n'}}^{\infty} \bar{z}_{n'}^2 f(\bar{z}_{n'})\, d\bar{z}_{n'} + \int_{-\infty}^{-Z_\alpha/\sqrt{n'}} \bar{z}_{n'}^2 f(\bar{z}_{n'})\, d\bar{z}_{n'}$$

Now, since

$$\bar{z}_{n'} \sim N\left(\mu_z, 1/\sqrt{n'}\right)$$

Therefore,

$$I = \frac{\sqrt{n'}}{\sqrt{2\pi}} \left[\int_{Z_\alpha/\sqrt{n'}}^{\infty} \bar{z}_{n'}^2 \exp(-1/2) \left[\frac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}\right]^2 d\bar{z}_{n'} \right.$$

$$\left. + \int_{-\infty}^{-Z_\alpha/\sqrt{n'}} \bar{z}_{n'}^2 \exp(-1/2) \left[\frac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}\right]^2 d\bar{z}_{n'} \right]$$

Putting $w = \dfrac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}$, which implies

$dw = \sqrt{n'}\ d\bar{z}_{n'}$, we have

$$I = \frac{1}{\sqrt{2\pi}} \left[\int_A^{\infty} (\mu_z + w/\sqrt{n'})^2 \, e^{-\frac{1}{2} w^2} \, dw \right.$$

$$\left. + \int_{-\infty}^{B} (\mu_z + w/\sqrt{n'})^2 \, e^{-\frac{1}{2} w^2} \, dw \right]$$

(where A and B are same as in Appendix A)

$$= \mu_z^2 \, (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{2\mu_z}{\sqrt{2\pi n'}} \left[\int_A^{\infty} w \, e^{-\frac{1}{2} w^2} \, dw + \int_{-\infty}^{B} w \, e^{-\frac{1}{2} w^2} \, dw \right]$$

$$+ \frac{1}{n'\sqrt{2\pi}} \left[\int_A^{\infty} w^2 \, e^{-\frac{1}{2} w^2} \, dw + \int_{-\infty}^{B} w^2 \, e^{-\frac{1}{2} w^2} \, dw \right]$$

Again putting $\frac{w^2}{2} = t \Rightarrow w\, dw = dt$, we have

$$I = \mu_z^2 \, (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{2\mu_z}{\sqrt{2\pi n'}} \left[\int_{A^2/2}^{\infty} e^{-t}\, dt + \int_{\infty}^{B^2/2} e^{-t}\, dt \right]$$

$$+ \frac{1}{n'\sqrt{2\pi}} \left[\int_{A^2/2}^{\infty} (2t)^{1/2} e^{-t}\, dt + \int_{\infty}^{B^2/2} (2t)^{1/2} e^{-t}\, dt \right]$$

$$= (\mu_z^2 + \frac{1}{n'}) (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{2\mu_z}{\sqrt{n'}} (\phi(A) - \phi(B))$$

$$+ \frac{1}{n'} (A\,\phi(A) - B\,\phi(B))$$

Appendix C

Derivation of

$$E\left(\bar{z}_{n'}\bar{z}_n \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

We know that

$$E\left(\bar{z}_{n'}\bar{z}_n \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \frac{\partial}{\partial t_1}\left[\frac{\partial}{\partial t_2} E\left(e^{t_1\bar{z}_n + t_2\bar{z}_{n'}} \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)\right.$$

$$\left. X\ P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)\right] \text{ at } t_1 = t_2 = 0$$

$$\text{.......}(1)$$

Now, let

$$I = E\left(e^{t_1\bar{z}_n + t_2\bar{z}_{n'}} \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= E\left(e^{t_1\bar{z}_n + t_2\bar{z}_{n'}} \,\middle|\, \bar{z}_{n'} > Z_\alpha/\sqrt{n'}\right) P\left(\bar{z}_{n'} > Z_\alpha/\sqrt{n'}\right)$$

$$+ E\left(e^{t_1\bar{z}_n + t_2\bar{z}_{n'}} \,\middle|\, \bar{z}_{n'} \leqslant -Z_\alpha/\sqrt{n'}\right) P\left(\bar{z}_{n'} \leqslant -Z_\alpha/\sqrt{n'}\right)$$

$$= \int_{\bar{z}_n:-\infty}^{\infty} \int_{\bar{z}_{n'}:Z_\alpha/\sqrt{n'}}^{\infty} e^{t_1\bar{z}_n + t_2\bar{z}_{n'}} f_1(\bar{z}_n, \bar{z}_{n'})\, d\bar{z}_n\, d\bar{z}_{n'}$$

$$+ \int_{\bar{z}_n:-\infty}^{\infty} \int_{\bar{z}_{n'}:-\infty}^{-Z_\alpha/\sqrt{n'}} e^{t_1\bar{z}_n + t_2\bar{z}_{n'}} f_1(\bar{z}_n, \bar{z}_{n'})\, d\bar{z}_n\, d\bar{z}_{n'}$$

where $f_1(\bar{z}_n, \bar{z}_{n'})$ is a bivariate normal p.d.f. with mean (μ_z, μ_z) and covariance matrix

$$\begin{bmatrix} \frac{1}{n} & \frac{1}{n'} \\ \frac{1}{n'} & \frac{1}{n'} \end{bmatrix}$$

Therefore,

$$f_1(\bar{z}_n, \bar{z}_{n'}) = \frac{1}{2\pi(\frac{1}{nn'} - \frac{1}{n'^2})^{1/2}} \exp\left(-\frac{1}{2}\right)\left(\frac{n'}{n'-n}\right)$$

$$X \left[n(\bar{z}_n - \mu_z)^2 + n'(\bar{z}_{n'} - \mu_z)^2 - 2n(\bar{z}_n - \mu_z)(\bar{z}_{n'} - \mu_z) \right]$$

Now, putting $z_1 = \frac{\bar{z}_n - \mu_z}{1/\sqrt{n}}$ and $z_2 = \frac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}$,

we have

$$I = \frac{\sqrt{n'}}{2\pi(n'-n)^{1/2}} \exp(t_1 + t_2)\mu_z$$

$$X \left[\int_{z_1:-\infty}^{\infty} \int_{z_2:A}^{\infty} f_2(z_1, z_2)\, dz_1\, dz_2 + \int_{z_1:-\infty}^{\infty} \int_{z_2:-\infty}^{B} f_2(z_1, z_2)\, dz_1\, dz_2 \right]$$

where

$$f_2(z_1, z_2) = \exp\left(-\frac{1}{2}\right)\left(\frac{n'}{n'-n}\right)$$

$$X\left[z_1^2 + z_2^2 - 2\frac{\sqrt{n}}{\sqrt{n'}} z_1 z_2 - 2\left(\frac{n'-n}{n'}\right)\left(\frac{t_1 z_1}{\sqrt{n}} + \frac{t_2 z_2}{\sqrt{n'}}\right)\right]$$

and A and B are same as in Appendix A.

Now, put

$$z_1 - \frac{\sqrt{n}}{\sqrt{n'}} z_2 - \left(\frac{n'-n}{n'}\right)\frac{t_1}{\sqrt{n}} = \left(\frac{n'-n}{n'}\right)^{1/2} u$$

and $z_2 - \frac{t_1}{\sqrt{n'}} - \frac{t_2}{\sqrt{n'}} = v$

Then, if we define

$J = \frac{\partial(u, v)}{\partial(z_1, z_2)} = \left(\frac{n'-n}{n'}\right)^{-1/2}$, we obtain

$$du\, dv = \left(\frac{n'-n}{n'}\right)^{-1/2} dz_1\, dz_2$$

Therefore, using the above substitution in I, we have

$$I = \frac{\exp(t_1 + t_2)\mu_z}{2\pi}\left[\int_{u:-\infty}^{\infty}\int_{v:A'}^{\infty} f_3(u, v)\, du\, dv\right.$$

$$+ \int_{u:-\infty}^{\infty} \int_{v:-\infty}^{B'} f_3(u, v)\, du\, dv \Big]$$

where

$$f_3(u, v) = \exp\left(-\frac{1}{2}\right)\left(\frac{n'}{n' - n}\right)\left[\left(\frac{n' - n}{n'}\right)(u^2 + v^2)\right.$$

$$\left. - \left(\frac{n' - n}{n'}\right)\left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2\, t_1\, t_2}{n'}\right)\right]$$

and

$$A' = A - (t_1 + t_2)/\sqrt{n'}$$

$$B' = B - (t_1 + t_2)/\sqrt{n'}$$

This is because

$$z_1^2 + z_2^2 - \frac{2\sqrt{n}}{\sqrt{n'}}\, z_1\, z_2 - 2\left(\frac{n' - n}{n'}\right)\left(\frac{t_1 z_1}{\sqrt{n}} + \frac{t_2 z_2}{\sqrt{n'}}\right)$$

$$= \left(z_1 - \frac{\sqrt{n}}{\sqrt{n'}}\, z_2 - \left(\frac{n' - n}{n'}\right)\frac{t_1}{\sqrt{n}}\right)^2$$

$$+ \left(\frac{n' - n}{n'}\right)\left(z_2 - \frac{t_1}{\sqrt{n'}} - \frac{t_2}{\sqrt{n'}}\right)^2$$

$$- \left(\frac{n' - n}{n'}\right)\left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2\, t_1 t_2}{n'}\right)$$

Thus,

$$I = \exp\left[(t_1 + t_2)\mu_z + \frac{1}{2}\left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2\, t_1\, t_2}{n'} \right)\right]$$

$$X\ (1 - \Phi(A') + \Phi(B')) \quad \ldots\ldots\ldots(2)$$

Now, substituting (2) in (1) we have,

$$E\ (\bar{z}_n,\ \bar{z}_n \,\big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$= \frac{\partial}{\partial t_1}\left[\frac{\partial}{\partial t_2} \exp\left[(t_1 + t_2)\mu_z + \frac{1}{2}\left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2\, t_1 t_2}{n'} \right)\right] \right.$$

$$\left. X\ (1 - \Phi(A') + \Phi(B'))\right] \text{ at } t_1 = t_2 = 0$$

Here, we shall use the formula for differentiation under the integral sign. That is if

$$\xi(y) = \int_{g(y)}^{h(y)} f\ (x,\ y)\ dx\ , \quad \text{then}$$

$$\xi'(y) = \int_{g(y)}^{h(y)} f_y(x,\ y)\ dx + h'(y)\ f\ (h\ (y),\ y)$$

$$-\ g'(y)\ f\ (g\ (y),\ y)$$

Hence,

$$E\left(\bar{z}_{n'}\bar{z}_n \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \left(\mu_z^2 + \frac{1}{n'}\right)\left(1 - \Phi(A) + \Phi(B)\right)$$

$$+ \frac{2\mu_z}{\sqrt{n'}}\left(\phi(A) - \phi(B)\right)$$

$$+ \frac{1}{n'}\left(A\,\phi(A) - B\,\phi(B)\right)$$

Appendix D

Derivation of

$$E\ (\bar{z}_{n'}\bar{y}_n \,\big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

Now, we know that

$$E\ (\bar{z}_{n'}\bar{y}_n \,\big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$= \frac{\partial}{\partial t_1}\left[\frac{\partial}{\partial t_2}\ E\ (e^{t_1\bar{y}_n + t_2\bar{z}_{n'}} \,\big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})\right.$$

$$\left. X\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})\right] \text{ at } t_1 = t_2 = 0 \qquad .....(1)$$

Now, let

$$I = E\ (e^{t_1\bar{y}_n + t_2\bar{z}_{n'}} \,\big|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'})$$

$$= E\ (e^{t_1\bar{y}_n + t_2\bar{z}_{n'}} \,\big|\, \bar{z}_{n'} > Z_\alpha/\sqrt{n'})\ P\ (\bar{z}_{n'} > Z_\alpha/\sqrt{n'})$$

$$+ E\ (e^{t_1\bar{y}_n + t_2\bar{z}_{n'}} \,\big|\, \bar{z}_{n'} \leqslant -Z_\alpha/\sqrt{n'})\ P\ (\bar{z}_{n'} \leqslant -Z_\alpha/\sqrt{n'})$$

$$= \int_{\bar{y}_n:-\infty}^{\infty} \int_{\bar{z}_{n'}:\, Z_\alpha/\sqrt{n'}}^{\infty} e^{t_1\bar{y}_n + t_2\bar{z}_{n'}}\ f_1(\bar{y}_n,\ \bar{z}_{n'})\ d\bar{y}_n\ d\bar{z}_{n'}$$

$$+ \int_{\bar{y}_n=-\infty}^{\infty} \int_{\bar{z}_{n'}=-\infty}^{-z_\alpha/\sqrt{n'}} e^{t_1\bar{y}_n + t_2\bar{z}_{n'}} f_1(\bar{y}_n, \bar{z}_{n'})\, d\bar{y}_n\, d\bar{z}_{n'}$$

where $f_1(\bar{y}_n, \bar{z}_{n'})$ is a bivariate normal p.d.f. with mean (μ_y, μ_z) and covariance matrix

$$\begin{bmatrix} \frac{1}{n} & \frac{\rho_{yz}}{n'} \\ \frac{\rho_{yz}}{n'} & \frac{1}{n'} \end{bmatrix}$$

Therefore,

$$f_1(\bar{y}_n, \bar{z}_{n'}) = \frac{1}{2\pi\left(\frac{1}{nn'} - \frac{\rho_{yz}^2}{n'^2}\right)^{1/2}} \exp\left(-\tfrac{1}{2}\right)\left(\frac{n'}{n' - \rho_{yz}^2 n}\right)$$

$$X \left[n(\bar{y}_n - \mu_y)^2 + n'(\bar{z}_{n'} - \mu_z)^2 - 2\rho_{yz}\, n(\bar{y}_n - \mu_y)(\bar{z}_{n'} - \mu_z)\right]$$

Now, putting $z_1 = \dfrac{\bar{y}_n - \mu_y}{1/\sqrt{n}}$ and $z_2 = \dfrac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}$, we have

$$I = \frac{\sqrt{n'}}{2\pi(n' - \rho_{yz}^2 n)^{1/2}} \exp(t_1\mu_y + t_2\mu_z)$$

$$X \left[\int_{z_1:-\infty}^{\infty} \int_{z_2:A}^{\infty} f_2(z_1, z_2)\, dz_1\, dz_2 \right.$$

$$\left. + \int_{z_1:-\infty}^{\infty} \int_{z_2:-\infty}^{B} f_2\, (z_1, z_2)\, dz_1\, dz_2 \right]$$

where $f_2(z_1, z_2) = \exp\,(-\frac{1}{2})\,\left(\frac{n'}{n' - \rho_{yz}^2\, n}\right)$

$$X \left[z_1^2 + z_2^2 - 2\,\rho_{yz}\,\frac{\sqrt{n}}{\sqrt{n'}}\, z_1 z_2 - 2\,\left(\frac{n' - \rho_{yz}^2\, n}{n'}\right)\left(\frac{t_1 z_1}{\sqrt{n}} + \frac{t_2 z_2}{\sqrt{n'}}\right) \right]$$

and A and B are same as in Appendix A.

Now, put

$$z_1 - \rho_{yz}\,\frac{\sqrt{n}}{\sqrt{n'}}\, z_2 - \left(\frac{n' - \rho_{yz}^2\, n}{n'}\right)\frac{t_1}{\sqrt{n}} = \left(\frac{n' - \rho_{yz}^2\, n}{n'}\right)^{1/2} u$$

$$z_2 - \rho_{yz}\,\frac{t_1}{\sqrt{n'}} - \frac{t_2}{\sqrt{n'}} = v$$

Then, if we define

$J = \dfrac{\partial(u, v)}{\partial(z_1, z_2)} = \left(\dfrac{n' - \rho_{yz}^2\, n}{n'}\right)^{-1/2}$, we obtain

$$du\, dv = \left(\frac{n' - \rho_{yz}^2\, n}{n'}\right)^{-1/2} dz_1\, dz_2$$

Therefore, using the above substitution in I, we have

$$I = \frac{\exp\,(t_1\mu_y + t_2\mu_z)}{2\pi} \cdot \left[\int_{u:-\infty}^{\infty} \int_{v:A'}^{\infty} f_3(u,\, v)\; du\; dv \right.$$

$$\left. + \int_{u:-\infty}^{\infty} \int_{v:-\infty}^{B'} f_3(u,\, v)\; du\; dv \right].$$

where

$$f_3(u,\, v) = \exp\,(-\tfrac{1}{2})\,\left(\frac{n'}{n' - \rho_{yz}^2\, n}\right) \left[\left(\frac{n' - \rho_{yz}^2\, n}{n'} \right) (u^2 + v^2) \right.$$

$$\left. - \left(\frac{n' - \rho_{yz}^2\, n}{n'} \right) \left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2\,\rho_{yz}\, t_1 t_2}{n'} \right) \right]$$

and $A' = A - \frac{1}{\sqrt{n'}}\,(\rho_{yz}\, t_1 + t_2)$

$$B' = B - \frac{1}{\sqrt{n'}}\,(\rho_{yz}\, t_1 + t_2)$$

This is because

$$z_1^2 + z_2^2 - 2\,\rho_{yz}\frac{\sqrt{n}}{\sqrt{n'}}\, z_1 z_2 - 2\left(\frac{n' - \rho_{yz}^2\, n}{n'} \right) \left(\frac{t_1 z_1}{\sqrt{n}} + \frac{t_2 z_2}{\sqrt{n'}} \right)$$

$$= \left[z_1 - \rho_{yz} \frac{\sqrt{n}}{\sqrt{n'}},\ z_2 - \left(\frac{n' - \rho_{yz}^2 n}{n'} \right) \frac{t_1}{\sqrt{n'}} \right]^2$$

$$+ \left(\frac{n' - \rho_{yz}^2 n}{n'} \right) \left[z_2 - \rho_{yz} \frac{t_1}{\sqrt{n'}} - \frac{t_2}{\sqrt{n'}} \right]^2$$

$$- \left(\frac{n' - \rho_{yz}^2 n}{n'} \right) \left[\frac{t_1^2}{n} + \frac{t_2^2}{n'} + 2 \rho_{yz} \frac{t_1 t_2}{n'} \right]$$

Thus,

$$I = \exp \left[t_1 \mu_y + t_2 \mu_z + \frac{1}{2} \left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2 \rho_{yz} t_1 t_2}{n'} \right) \right]$$

$$X\ (1 - \Phi(A') + \Phi(B')) \quad \ldots\ldots\ldots\ldots (2)$$

Now, substituting (2) in (1) we have

$$E\ (\bar{z}_n, \bar{y}_n \,\big|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= \frac{\partial}{\partial t_1} \left[\frac{\partial}{\partial t_2} \exp \left[t_1 \mu_y + t_2 \mu_z + \frac{1}{2} \left(\frac{t_1^2}{n} + \frac{t_2^2}{n'} + \frac{2 \rho_{yz} t_1 t_2}{n'} \right) \right] \right.$$

$$\left. X\ (1 - \Phi(A') + \Phi(B')) \right] \text{ at } t_1 = t_2 = 0$$

which again by the formula for differentiation under the integral sign as in Appendix C, leads to

$$E\ (\bar{z}_{n'}\bar{y}_n \mid |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= (\mu_y \mu_z + \frac{\rho_{yz}}{n'})\ (1 - \Phi(A) + \Phi(B))$$

$$\frac{1}{\sqrt{n'}}\ (\mu_y + \rho_{yz}\ \mu_z)\ (\phi(A) - \phi(B))$$

$$+ \frac{\rho_{yz}}{n'}\ (A\ \phi(A) - B\ \phi(B))$$

In the same manner we can obtain

$$E\ (\bar{z}_{n'}\bar{x}_n \mid |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= (\mu_x \mu_z + \frac{\rho_{xz}}{n'})\ (1 - \Phi(A) + \Phi(B))$$

$$+ \frac{1}{\sqrt{n'}}\ (\mu_x + \rho_{xz}\ \mu_z)\ (\phi(A) - \phi(B))$$

$$+ \frac{\rho_{xz}}{n'}\ (A\ \phi(A) - B\ \phi(B))$$

Appendix E

Derivation of

$$E\left(\bar{x}_{n'}\bar{z}_{n'} \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

Now, we know that

$$E\left(\bar{x}_{n'}\bar{z}_{n'} \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= \frac{\partial}{\partial t_1}\left[\frac{\partial}{\partial t_2} E\left(e^{t_1\bar{x}_{n'} + t_2\bar{z}_{n'}} \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)\right.$$

$$\left. \times P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)\right] \text{ at } t_1 = t_2 = 0 \quad \ldots.(1)$$

Now, let

$$I = E\left(e^{t_1\bar{x}_{n'} + t_2\bar{z}_{n'}} \,\middle|\, |\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right) P\left(|\bar{z}_{n'}| > Z_\alpha/\sqrt{n'}\right)$$

$$= E\left(e^{t_1\bar{x}_{n'} + t_2\bar{z}_{n'}} \,\middle|\, \bar{z}_{n'} > Z_\alpha/\sqrt{n'}\right) P\left(\bar{z}_{n'} > Z_\alpha/\sqrt{n'}\right)$$

$$+ E\left(e^{t_1\bar{x}_{n'} + t_2\bar{z}_{n'}} \,\middle|\, \bar{z}_{n'} \leqslant -Z_\alpha/\sqrt{n'}\right) P\left(\bar{z}_{n'} \leqslant -Z_\alpha/\sqrt{n'}\right)$$

$$= \int_{\bar{x}_{n'}:-\infty}^{\infty} \int_{\bar{z}_{n'}:\,Z_\alpha/\sqrt{n'}}^{\infty} e^{t_1\bar{x}_{n'} + t_2\bar{z}_{n'}} f_1(\bar{x}_{n'}, \bar{z}_{n'})\, d\bar{x}_{n'}\, d\bar{z}_{n'}$$

$$+ \int_{\bar{x}_{n'}:-\infty}^{\infty} \int_{\bar{z}_{n'}:-\infty}^{-Z_\alpha/\sqrt{n'}} e^{t_1\bar{x}_{n'}+ t_2\bar{z}_{n'}} f_1(\bar{x}_{n'}, \bar{z}_{n'})\, d\bar{x}_{n'}\, d\bar{z}_{n'}$$

where $f_1(\bar{x}_{n'}, \bar{z}_{n'})$ is a bivariate normal p.d.f. with mean (μ_x, μ_z) and covariance matrix

$$\begin{bmatrix} \frac{1}{n'} & \frac{\rho_{xz}}{n'} \\ \frac{\rho_{xz}}{n'} & \frac{1}{n'} \end{bmatrix}$$

Therefore,

$$f_1(\bar{x}_{n'}, \bar{z}_{n'}) = \frac{n'}{2\pi(1 - \rho_{xz}^2)^{1/2}} \exp\left(-\tfrac{1}{2}\right)\left(\frac{n'}{1 - \rho_{xz}^2}\right)$$

$$x \left[(\bar{x}_{n'} - \mu_x)^2 + (\bar{z}_{n'} - \mu_z)^2 - 2\rho_{xz}(\bar{x}_{n'} - \mu_x)(\bar{z}_{n'} - \mu_z)\right]$$

On putting $z_1 = \dfrac{\bar{x}_{n'} - \mu_x}{1/\sqrt{n'}}$ and $z_2 = \dfrac{\bar{z}_{n'} - \mu_z}{1/\sqrt{n'}}$

we have

$$I = \frac{\exp(t_1\mu_x + t_2\mu_z)}{2\pi(1 - \rho_{xz}^2)^{1/2}} \left[\int_{z_1:-\infty}^{\infty} \int_{z_2:A}^{\infty} f_2(z_1, z_2)\, dz_1\, dz_2 \right.$$

$$+ \int_{z_1:-\infty}^{\infty} \int_{z_2:-\infty}^{B} f_2(z_1, z_2)\, dz_1\, dz_2 \Big]$$

where $f_2\,(z_1, z_2) = \exp\,(-\frac{1}{2})\,(\frac{1}{1-\rho_{xz}^2})$

$$X\left[z_1^2 + z_2^2 - 2\,\rho_{xz}\, z_1\, z_2 - \frac{2(1-\rho_{xz}^2)}{\sqrt{n'}}\,(t_1 z_1 + t_2 z_2)\right]$$

where A and B are same as in Appendix A.

Now, put

$$z_1 - \rho_{xz}\, z_2 - (1-\rho_{xz}^2)\,\frac{t_1}{\sqrt{n'}} = (1-\rho_{xz}^2)^{1/2}\, u$$

$$z_2 - \rho_{xz}\frac{t_1}{\sqrt{n'}} - \frac{t_2}{\sqrt{n'}} = v$$

Then, if we define

$$J = \frac{\partial\,(u, v)}{\partial(z_1, z_2)} = (1-\rho_{xz}^2)^{-1/2}\text{ , we obtain}$$

$$du\, dv = (1-\rho_{xz}^2)^{-1/2}\, dz_1\, dz_2$$

Therefore, using the above substitution in I, we have

$$I = \frac{\exp(t_1\,\mu_x + t_2\,\mu_z)}{2\pi}\left[\int_{u:-\infty}^{\infty}\int_{v:A'}^{\infty} f_3(u, v)\, du,\, dv\right.$$

$$+ \int_{u:-\infty}^{\infty} \int_{v:-\infty}^{B'} f_3(u, v)\, du\, dv \Big]$$

where

$$f_3(u, v) = \exp\left(-\tfrac{1}{2}\right)\left(\frac{1}{1 - \rho_{xz}^2}\right)\Big[(1 - \rho_{xz}^2)(u^2 + v^2)$$

$$- (1 - \rho_{xz}^2)\left(\frac{t_1^2}{n'} + \frac{t_2^2}{n'} + \frac{2\rho_{xz} t_1 t_2}{n'}\right)\Big]$$

and $A' = A - \frac{1}{\sqrt{n'}}(\rho_{xz} t_1 + t_2)$

$B' = B - \frac{1}{\sqrt{n'}}(\rho_{xz} t_1 + t_2)$

This is because

$$z_1^2 + z_2^2 - 2\rho_{xz} z_1 z_2 - \frac{2(1 - \rho_{xz}^2)}{\sqrt{n'}}(t_1 z_1 + t_2 z_2)$$

$$= \left[z_1 - \rho_{xz} z_2 - (1 - \rho_{xz}^2)\frac{t_1}{\sqrt{n'}}\right]^2$$

$$+ (1 - \rho_{xz}^2)\left[z_2 - \rho_{xz}\frac{t_1}{\sqrt{n'}} - \frac{t_2}{\sqrt{n'}}\right]^2 .$$

$$- (1 - \rho_{xz}^2)\left[\frac{t_1^2}{n'} + \frac{t_2^2}{n'} + \frac{2\rho_{xz} t_1 t_2}{n'}\right]$$

Thus,

$$I = \exp\left[t_1 \mu_x + t_2 \mu_z + \frac{1}{2\, n'} (t_1^2 + t_2^2 + 2\, \rho_{xz}\, t_1 t_2) \right]$$

$$X\ (1 - \Phi(A') + \Phi(B')\ ) \ldots\ldots\ldots (2)$$

Now, substituting (2) in (1) we have

$$E\ (\bar{x}_{n'} \bar{z}_{n'} \,\big|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= \frac{\partial}{\partial t_1} \left[\frac{\partial}{\partial t_2} \exp\left[t_1 \mu_x + t_2 \mu_z + \frac{1}{2n'} (t_1^2 + t_2^2 + 2\, \rho_{xz}\, t_1 t_2) \right] \right.$$

$$\left. X\ (1 - \Phi(A') + \Phi(B')\ ) \right] \text{ at } t_1 = t_2 = 0$$

which again by the formula for differentiation under the integral sign, as in Appendix C leads to

$$E\ (\bar{x}_{n'} \bar{z}_{n'} \,\big|\, |\bar{z}_{n'}| > z_\alpha/\sqrt{n'})\ P\ (|\bar{z}_{n'}| > z_\alpha/\sqrt{n'})$$

$$= (\ \mu_x \mu_z + \frac{\rho_{xz}}{n'}\ )\ (1 - \Phi(A) + \Phi(B)\ )$$

$$+ \frac{1}{\sqrt{n'}}\ (\ \mu_x + \rho_{xz}\, \mu_z)\ (\ \varphi(A) - \varphi(B)\ )$$

$$+ \frac{\rho_{xz}}{n'}\ (A\ \varphi(A) - B\ \varphi(B)\ )$$