Journal of Reliability and Statistical Studies (ISSN: 0974-8024)
Vol. 2, Issue 1(2009): 103 - 118

A GENERAL CLASS OF ESTIMATORS OF A FINITE POPULATION MEAN USING MULTI-AUXILIARY INFORMATION UNDER TWO STAGE SAMPLING SCHEME

Meenakshi Srivastava* and Neha Garg
Department of Statistics, Institute of Social Sciences,
Dr. B. R. Ambedkar University, Agra, India
* E.mail: msrivastava_iss@hotmail.com

Abstract

In sample surveys, it is usual to make use of auxiliary information to increase the precision of estimators. A general class of estimators is suggested to estimate the population mean for the variable under study in two stage sampling scheme. Some special cases of this class of estimators are considered and are compared by using a data set.

It turns out that the newly suggested estimators dominate all other well known estimators in terms of mean square error. Finally it is shown, how to extend the class of estimators if multi auxiliary variables are available in the cases of two stage sampling scheme.

Key words: Bias, MSE, Auxiliary variables, Ratio estimator, Two stage sampling.

1. Introduction

It is well known that suitable use of auxiliary information results in considerable reduction in the mean square error of the estimator. In this regard ratio, regression and product estimators are widely used, if the correlation coefficient is high between the auxiliary variable x and the study variable y. Some of the important works in this direction are of Singh (1965), Srivastava (1980), Rao (1991), Kadilar and Cingi (2004) etc. But in large scale surveys, we often collect data on more than one auxiliary variables and some of these may be correlated with Y. Olkin (1958), Raj (1965), Rao and Mudholkar (1967), Srivastava (1971), Singh (1982) etc. have considered some estimators which utilize information on several auxiliary variables which are highly correlated with the variable under study.

Two-stage sampling scheme consists in selecting the first stage unit (fsu) by any of the sampling schemes eg. simple random sampling with replacement, simple random sampling without replacement, systematic sampling, probability proportional to size with replacement, probability proportional to size without replacement etc. and the size of the fsu’s may be equal or unequal. From each selected fsu, a sample of second stage units (ssu) is selected independently by any of the above suitable sampling procedure.

In a socioeconomic survey, for example villages may be considered as fsu’s and households as ssu’s. While preparing lists of households belonging to each selected village, one may collect some information such as type of dwellings, educational standard attained, size of households etc. Such information may be suitably used for drawing the sample at the second stage. By careful exploitation of such information, the efficiency of two stage sampling can be greatly enhanced. In such types of surveys, for example, one may be interested in estimating ratios like yield rates in crop survey, proportion of expenditure on food, clothing etc, sex ratio, birth rates and so on.

The main focus of the present paper is firstly to construct a class of estimators in two-stage sampling for equal fsu and unequal fsu, when the population mean of the auxiliary variable for all fsu's are known i.e. general estimator in two stage sampling scheme secondly to extend the suggested class of estimator using multi-auxiliary variables.

The expressions for bias and mean square error of the usual two stage estimator are given in section 3. Section 4 of the article contains the derivation of the biasness and mean square error for the suggested general class of estimators in two stage sampling while section 5 deals with some special cases of it. The generalization of the proposed class is done in section 6 using multi-auxiliary variables and its special cases are considered in section 7. Section 8 considers the efficiency comparison of the proposed estimator with that of usual two-stage estimator (without auxiliary information). In section 9, all the derived results are numerically supported by database study.

2. Notations

Let fsu's be of unequal size and simple random sampling without replacement be adopted in both the stages. The commonly used notations are as follows:-

N : Total no. of fsu's (clusters) in the population

n : Total no. of fsu's in the sample

M_i : Total no. of ssu's belonging to the ith fsu in the population

M_0 : Total no. of ssu's in the population = $\sum_{i=1}^{N} M_i$

$\overline{M}$: Average size of fsu's = $\dfrac{M_0}{N}$

m_i : Total no. of ssu's selected from ith fsu in the sample

m_0 : Total no. of ssu's in the sample = $\sum_{i=1}^{n} m_i$

Y : Variable under study

Y_{ij} : Observation on j^{th} ssu belonging to the i^{th} fsu in the population $i = 1, 2, ..., N$ and $j = 1, 2, ..., M_i$

$\overline{Y}_{i.}$: Population mean of ssu's in the i^{th} fsu, i.e. $\overline{Y}_{i.} = \dfrac{1}{M_i}\sum_{j=1}^{M_i} Y_{ij}$

$\overline{Y}$: Population mean

$$= \frac{1}{M_0}\sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij} = \frac{1}{N}\sum_{i=1}^{N}\frac{M_i}{\overline{M}}\overline{Y}_{i.}$$

y_{ij} : Observation on j^{th} ssu belonging to the i^{th} fsu in the sample;

$i = 1,2, ..., n$ and $j = 1, 2, ..., m_i$

$\bar{y}_{i.}$: Sample mean of ssu's in ith fsu $= \frac{1}{m_i}\sum_{i=1}^{m_i} y_{ij}$

X_k : kth auxiliary variable; k = 1, 2, …, p.

X_{ijk} : Value of kth auxiliary variable on jth ssu belonging to the ith fsu in the population

$\bar{X}_{ik}$: Population mean of kth auxiliary variable for ssu in ith fsu

x_{ijk} : Value of kth auxiliary variable on jth ssu belonging to the ith fsu in the sample

$\bar{x}_{ik}$: Sample mean of kth auxiliary variable for ssu in ith fsu

α_i : Weight for i^{th} fsu

S_y^2 : Population mean square error of Y variable

$$= \frac{1}{M_0 - 1}\sum_{i=1}^{N}\sum_{j=1}^{M_i}\left(Y_{ij} - \bar{Y}\right)^2$$

S_{yi}^2 : Population mean square error of Y variable for ith fsu

$$= \frac{1}{M_i - 1}\sum_{j=1}^{M_i}\left(Y_{ij} - \bar{Y}_i\right)^2$$

S_{xik}^2 : Population mean square error of kth auxiliary variable for ith fsu

$$= \frac{1}{M_i - 1}\sum_{j=1}^{M_i}\left(X_{ijk} - \bar{X}_{ik}\right)^2$$

C_{yi}^2 : Coefficient of variation of Y for ith fsu $= \frac{S_{yi}^2}{\bar{Y}_{i.}^2}$

C_{xik}^2 : Coefficient of variation of X_k for ith fsu $= \frac{S_{xik}^2}{\bar{X}_{ik.}^2}$

ρ_{ik} : Correlation coefficient between the variables Y and X_k, for ith fsu

ρ_{ikh} : Correlation coefficient between the variables X_k and X_h ($k \neq h$) for ith fsu

b_{ij} : Regression coefficient between Y and X_j for ith fsu for the sample

B_{ij} : Regression coefficient between Y and X_j for ith fsu for the population

$f = \left(\frac{1}{n} - \frac{1}{N}\right)$, $f_i = \left(\frac{1}{m_i} - \frac{1}{M_i}\right)$, Let us define

$e_{io} = \left(\frac{\bar{y} - \bar{Y}}{\bar{Y}}\right)$ and $e_{ij} = \left(\frac{\bar{x}_{ij} - \bar{X}_{ij}}{\bar{X}_{ij}}\right)$; j = $1, 2, \ldots, p$

Such that

$E(e_{i0}) = E(e_{ij}) = 0,$ $\quad j = 1, 2, \ldots, p$

$E(e_{io}^2) = f_i C_{yi}^2,$ $\quad E(e_{ij}^2) = f_i C_{xij}^2,$ $\quad j = 1, 2, \ldots, p$

$E(e_{io} e_{ij}) = f_i \rho_{ij} C_{yi} C_{xij},$ $\quad j = 1, 2, \ldots, p$

$E(e_{ij} e_{ik}) = f_i \rho_{ijk} C_{xij} C_{xik},$ $\quad j = 1, 2, \ldots, p$

3. Estimator and its Mean Square Error

The usual two stage estimator for population mean is given as

$$\bar{y}_{TS} = \frac{1}{n}\sum_{i=1}^{n} \alpha_i \bar{y}_i \tag{1}$$

To the first degree of approximation, the bias and mean square error are given as

$$Bias(\bar{y}_{TS}) = \frac{1}{N\overline{M}}\sum_{i=1}^{N}\left(\alpha_i \overline{M} - M_i\right)\overline{Y}_{i.} \tag{2}$$

$$MSE\left(\bar{y}_{TS}\right) = \frac{f}{N-1}\sum_{i=1}^{N}\left[\alpha_i \overline{Y}_{i.} - \frac{1}{N}\sum_{i=1}^{N}\alpha_i \overline{Y}_{i.}\right]^2 + \frac{1}{nN}\sum_{i=1}^{N} f_i \alpha_i^2 \overline{Y}_{i.}^2 C_{yi}^2 \tag{3}$$

4. Suggested Class of Estimators in Two Stage Sampling

We propose a general class of estimators $\bar{y}_{GTS}$ using two stage sampling when population mean $\overline{X}_i$ is known for every i[th] fsu, as

$$\bar{y}_{GTS} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{ig} \tag{4}$$

where 'GTS' stands for 'General Estimator in Two Stage Sampling Scheme' and $\bar{y}_{ig}$ is a function of $\bar{y}$, $\overline{X}_i$ and $\bar{x}_i$ in i[th] fsu.

4.1 Its Bias and MSE

Theorem 1: The bias of $\bar{y}_{GTS}$ is given as

$$Bias(\bar{y}_{GTS}) = \frac{1}{N\overline{M}}\sum_{i=1}^{N}\left[\alpha_i z_i \overline{M} - M_i \overline{Y}_{i.}\right]$$

Proof.

$$E\left(\bar{y}_{GTS}\right) = E\left[E\left(\bar{y}_{GTS} / i\right)\right]$$

$$= \frac{1}{n} E\sum_{i=1}^{n}\alpha_i z_i \quad , \qquad \text{where } z_i = E\left(\bar{y}_{ig} / i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\alpha_i z_i$$

=> $$E(\bar{y}_{GTS}) - \bar{Y} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i z_i - \frac{1}{N}\sum_{i=1}^{N}\frac{M_i}{\overline{M}}\bar{Y}_{i.} = \frac{1}{N\overline{M}}\sum_{i=1}^{N}\left[\alpha_i z_i \overline{M} - M_i \bar{Y}_{i.}\right] \quad (5)$$

Theorem 2 : The MSE of $\bar{y}_{GTS}$ is given as

$$MSE(\bar{y}_{RTS}) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left[\alpha_i z_i - E(\alpha_i z_i)\right]^2 + \frac{1}{nN}\sum_{i=1}^{N}\alpha_i^2 v_i$$

where $z_i = E(\bar{y}_{ig}/i)$ and $v_i = MSE(\bar{y}_{ig}/i)$

Proof. We have, $MSE(\bar{y}_{GTS}) = MSE[E(\bar{y}_{GTS}/i)] + E[MSE(\bar{y}_{GTS}/i)]$

$$MSE[E(\bar{y}_{GTS}/i)] = MSE\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_i z_i\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left[\alpha_i z_i - E(\alpha_i z_i)\right]^2 \quad (6)$$

where $E(\alpha_i z_i) = \frac{1}{N}\sum_{i=1}^{N}\alpha_i z_i$

$$E[MSE(\bar{y}_{GTS}/i)] = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\alpha_i^2 MSE(\bar{y}_{ig}/i)\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n}\alpha_i^2 v_i\right], \quad \text{where } v_i = MSE(\bar{y}_{ig}/i)$$

$$= \frac{1}{nN}\sum_{i=1}^{N}\alpha_i^2 v_i \quad (7)$$

Adding (6) and (7), we get the final expression.

5. Some Cases for the Class of Estimators

Case 1 : When X is positively correlated with Y for each fsu,, our estimator will convert into separate ratio estimator given as

$$\bar{y}_{RAT.TS} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{i.rat} \quad (8)$$

where $\bar{y}_{i.rat} = \frac{\bar{y}_i}{\bar{x}_i}\bar{X}_i$ is the usual ratio estimator in i[th] fsu

$$z_i = \bar{Y}_{i.}\left[1 + f_i\left(C_{xi}^2 - \rho_i C_{yi} C_{xi}\right)\right]$$

$$v_i = \bar{Y}_{i.}^2 f_i\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right)$$

Case 2 : When X is positively correlated with Y for each fsu, we can also use separate regression estimator given as

$$\bar{y}_{REG.TS} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{i.reg} \qquad (9)$$

where $\bar{y}_{i.reg} = \bar{y}_i + b_i\left(\bar{X}_i - \bar{x}_i\right)$ is the usual regression estimator in ith fsu with

$z_i = \bar{Y}_i$ if b_i is known and

$$v_i = f_i\left[S_{yi}^2 + B_i^2 S_{xi}^2 - 2B_i\rho_i S_{yi} S_{xi}\right]$$

Case 3 : If each X is negatively correlated with Y for each fsu then our estimator will convert into separate product estimator given as

$$\bar{y}_{PROD.TS} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{i.prod} \qquad (10)$$

where $\bar{y}_{i.prod} = \frac{\bar{y}_i}{\bar{X}_i}\bar{x}_i$ is the usual product estimator in ith fsu with

$z_i = \bar{Y}_{i.}\left[1 + f_i\left(C_{xi}^2 + \rho_i C_{yi} C_{xi}\right)\right]$ and

$$v_i = \bar{Y}_{i.}^2 f_i\left(C_{yi}^2 + C_{xi}^2 + 2\rho_i C_{yi} C_{xi}\right)$$

6. Generalization of the Suggested Class

The generalized class of two stage estimators when p auxiliary variables are known for every ith fsu is given by

$$\bar{y}_{GTS.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\bar{y}_{ijg} \qquad (11)$$

where 'GTS.p' stands for 'General Estimator in Two Stage Sampling Scheme when p auxiliary variables are known' with $\sum_{j=1}^{p} w_{ij} = 1$ and $\bar{y}_{ijg}$ is a function of $\bar{y}_i$, $\bar{X}_{ij}$ and $\bar{x}_{ij}$ for jth auxiliary variable in ith fsu.

6.1 Its Bias and MSE

Theorem 3 : The bias of $\bar{y}_{GTS.p}$ is given as

$$Bias\left(\bar{y}_{GTS.p}\right) = \frac{1}{N\bar{M}}\sum_{i=1}^{N}\left[\alpha_i z_i \bar{M} - M_i \bar{Y}_{i.}\right]$$

Proof.

$$E\left(\bar{y}_{GTS.p}\right) = E\left[E\left(\bar{y}_{GTS.p} / i\right)\right]$$

$$= \frac{1}{n}E\left[\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij} E\left(\bar{y}_{ijg} / i\right)\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} \alpha_i z_i\right], \quad \text{where } z_i = \sum w_{ij} E(\bar{y}_{ijg} / i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \alpha_i z_i$$

$$E(\bar{y}_{GTS.p}) - \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i z_i - \frac{1}{N} \sum_{i=1}^{N} \frac{M_i}{\overline{M}} \bar{Y}_{i.}$$

$$= \frac{1}{N\overline{M}} \sum_{i=1}^{N} \left[\alpha_i z_i \overline{M} - M_i \bar{Y}_{i.}\right] \tag{12}$$

Theorem 4 The MSE of $\bar{y}_{GTS.p}$ is given as

$$MSE(\bar{y}_{GTS.p}) = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{N-1} \sum_{i=1}^{N} \left[\alpha_i z_i - E(\alpha_i z_i)\right]^2 + \frac{1}{n^2} E\left[\sum_{i=1}^{n} \alpha_i^2 MSE\left(\sum_{j=1}^{p} w_{ij} \bar{y}_{ijg} / i\right)\right]$$

Proof. $MSE\ (\bar{y}_{GTS.p}) = MSE\ [E\ (\bar{y}_{GTS.p} /i)] + E\ [MSE\ (\bar{y}_{GTS.p} /i)]$

$$MSE[E(\bar{y}_{GTS.p} / i)] = MSE\left[\frac{1}{n} \sum_{i=1}^{n} \alpha_i z_i\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{N-1} \sum_{i=1}^{N} \left[\alpha_i z_i - E(\alpha_i z_i)\right]^2 \tag{13}$$

where $E(\alpha_i z_i) = \frac{1}{N} \sum_{i=1}^{N} \alpha_i z_i$

$$E[MSE(\bar{y}_{GTS.p} / i)] = \frac{1}{n^2} E\left[\sum_{i=1}^{n} \alpha_i^2 MSE\left(\sum_{j=1}^{p} w_{ij} \bar{y}_{ijg} / i\right)\right] \tag{14}$$

$MSE\left(\sum_{j=1}^{p} w_{ij} \bar{y}_{ijg} / i\right)$ can easily be obtained for different values of function $\bar{y}_{ijg}$ by adopting the procedure given by Olkin (1958).

Thus, $$MSE\left(\sum_{j=1}^{p} w_{ij} \bar{y}_{ijg} / i\right) = \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \sum_{j=1}^{p} \sum_{h=1}^{p} w_{ij} w_{ih} v_{ijh}$$

where, $\left(\frac{1}{m_i} - \frac{1}{M_i}\right) v_{ijh} = Cov(\bar{y}_{ijg}, \bar{y}_{ihg})$. In matrix notations,

$$MSE\left(\sum_{j=1}^{p} w_{ij} \bar{y}_{ijg} / i\right) = \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \underset{\sim}{w_i} V_i \underset{\sim}{w_i'}$$

where the matrix $V_i=(v_{ijh})$ and $\underset{\sim}{w_i} = \left(w_{i1}, w_{i2}, ..., w_{ip}\right)$, $\underset{\sim}{w_i^{'}}$ being the transpose of $\underset{\sim}{w_i}$.

Optimum Values of w_{ij} for $j = 1, 2, …, p$

It can be shown easily that the optimum w_{ij} is given by

$$w_{ij} = \frac{Sum\ of\ the\ elements\ of\ the\ j^{th}\ column\ of\ V_i^{-1}}{Sum\ of\ all\ the\ p^2\ elements\ in\ V_i^{-1}}$$

where V_i^{-1} is the matrix inverse to V_i. Using the optimum weights, the mean square error is found to be

$$MSE\left(\sum_{j=1}^{p} w_{ij}\bar{y}_{ijg}\ /\ i\right) = \left(\frac{1}{m_i} - \frac{1}{M_i}\right) \Big/ Sum\ of\ all\ the\ p^2\ elements\ in\ V_i^{-1}$$

Remark (i) To avoid the mathematical complexity in deriving MSE, we will use the above procedure for finding optimum values of w_{ij} for the suggested estimators.

(ii) In deriving the expressions of MSE of all the estimators of the suggested class, the covariance term is taken to be zero because the clusters are independent of each other.

7. Some Cases for the Generalized Class of Estimators

Case 1: Multivariate Ratio Estimator

The combined ratio estimator $\bar{y}_{RAT.TS.p}$, of $\bar{Y}$ in two stage estimators when p auxiliary variables are known for every i[th] fsu is given by

$$\bar{y}_{RAT.TS.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\bar{y}_{ij.rat} \qquad (15)$$

where, $\bar{y}_{ij.rat} = \frac{\bar{y}_i}{\bar{x}_{ij}}\bar{X}_{ij}$ is the usual ratio estimator in i[th] fsu for j[th] auxiliary variable.

$$z_i = \bar{Y}_{i.}\left[1 + f_i\sum_{j=1}^{p} w_{ij}\left(C_{xij}^2 - \rho_{ij}C_{yi}C_{xij}\right)\right]$$

$$v_{ijh.rat} = \bar{Y}_i^2\left[C_{yi}^2 + \rho_{ijh}C_{xij}C_{xih} - \rho_{ij}C_{yi}C_{xij} - \rho_{ih}C_{yi}C_{xih}\right]$$

Case 2: Multivariate Regression Estimator

The combined regression estimator $\bar{y}_{REG.TS.p}$, of $\bar{Y}$ in two stage estimators when p auxiliary variables are known for every i[th] fsu is given by

$$\bar{y}_{REG.TS.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\bar{y}_{ij.reg} \qquad (16)$$

where, $\bar{y}_{ij.reg} = \bar{y}_i + b_{ij}\left(\bar{X}_{ij} - \bar{x}_{ij}\right)$ is the usual ratio estimator in ith fsu for jth auxiliary variable.

$$z_i = \bar{Y}_{i.} - \sum_{j=1}^{p} w_{ij}\, f_i\, cov\left(x_{ij}, b_{ij}\right) \qquad \text{If } b_i \text{ is unknown}$$

$$z_i = \bar{Y}_{i.} \qquad \text{If } b_i \text{ is known}$$

$$v_{ijh.reg} = \left[S_{yi}^2 + B_{ij}B_{ih}\rho_{ijh}S_{xij}S_{xih} - B_{ij}\rho_{ij}S_{yi}S_{xij} - B_{ih}\rho_{ih}S_{yi}S_{xih}\right]$$

Case 3: Multivariate Product Estimator

The combined product estimator $\bar{y}_{PROD.TS.p}$, of $\bar{Y}$ in two stage estimators when p auxiliary variables are known for every ith fsu is given by

$$\bar{y}_{PROD.TS.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\bar{y}_{ij.prod} \tag{17}$$

where, $\bar{y}_{ij.prod} = \frac{\bar{y}_i}{\bar{X}_{ij}}\bar{x}_{ij}$ is the usual product estimator in ith fsu for jth auxiliary variable.

$$z_i = \bar{Y}_{i.}\left[1 + f_i \sum_{j=1}^{p} w_{ij}\rho_{ij}C_{yi}C_{xij}\right]$$

$$v_{ijh.prod} = \bar{Y}_i^2\left[C_{yi}^2 + \rho_{ijh}C_{xij}C_{xih} + \rho_{ij}C_{yi}C_{xij} + \rho_{ih}C_{yi}C_{xih}\right]$$

8. Efficiency Comparison

Theorem 5. $\bar{y}_{RTS}$ will be efficient than $\bar{y}_{TS}$ if $\frac{\sum_{i=1}^{N}\left(N\bar{Y}_{i.} - \sum_{i=1}^{N}\bar{Y}_{i.}\right)^2}{\sum_{i=1}^{N}\bar{Y}_{i.}^2} < A$

where $A = \frac{(N-1)(2\rho - 1)}{nf(1-\rho)\left[2 + f' C^2(1-\rho)\right]}$

Proof: $\bar{y}_{RTS}$ will be more efficient than $\bar{y}_{TS}$ if it satisfies the following condition

$$MSE(\bar{y}_{RTS}) < Var(\bar{y}_{TS})$$

After applying the following approximations,

$$f_i = f_0, \quad C_{yi} = C_{xi} = C, \quad \rho_i = \rho, \quad \alpha_i = \alpha \qquad \forall i$$

efficiency condition reduces to

$$f\left[1 + f' C^2(1-\rho)\right] 2\frac{\alpha^2}{N(N-1)}\sum_{i=1}^{N}\left(N\bar{Y}_{i.} - \sum_{i=1}^{N}\bar{Y}_{i.}\right)^2 + \frac{2f' C^2(1-\rho)\alpha^2}{nN}\sum_{i=1}^{N}\bar{Y}_{i.}^2$$

$$< \frac{f\alpha^2}{N(N-1)}\sum_{i=1}^{N}\left(\bar{Y}_{i.}-\sum_{i=1}^{N}\bar{Y}_{i.}\right)^2+\frac{f' C^2\alpha^2}{nN}\sum_{i=1}^{N}\bar{Y}_{i.}^2$$

$$=> f\, f'C^2(1-\rho)\left[2+f' C^2(1-\rho)\right]\frac{\alpha^2}{N(N-1)}\sum_{i=1}^{N}\left(N\bar{Y}_{i.}-\sum_{i=1}^{N}\bar{Y}_{i.}\right)^2$$

$$<\frac{f' C^2(2\rho-1)\alpha^2}{nN}\sum_{i=1}^{N}\bar{Y}_{i.}^2$$

or $$\frac{\sum_{i=1}^{N}\left(N\bar{Y}_{i.}-\sum_{i=1}^{N}\bar{Y}_{i.}\right)^2}{\sum_{i=1}^{N}\bar{Y}_{i.}^2}<A \qquad (18)$$

Remark : Efficiency conditions for other members of the suggested class can be obtained in similar manner.

9. Numerical Illustrations

For this purpose, we consider the population of N=4 clusters as fsu with equal number of fsu's and another population with unequal number of fsu for comparing the proposed general class of estimators with usual two stage estimator. Suppose a sample of size n=2 clusters is drawn from this population. ssu's can be selected in proportion to M_i, i.e. $m_i=(M_i/\sum_{i=1}^{N}M_i)\times 32$. For unequal fsu's the comparison has been done by taking two values of α_i, i.e. 1 and $M_i/\bar{M}$. Table 1 gives the population parameters for population I (for equal fsu's) and II (for unequal fsu's) given in the Appendix.

10. Discussion and Conclusion

The separate ratio, regression estimators in two stage sampling scheme using multi-auxiliary information have been evaluated for their comparison with usual two stage estimator without using any auxiliary information for equal and unequal fsu considering two different values of α_i i.e. 1 and $M_i/\bar{M}$. The following conclusions can be drawn from this empirical illustrations :

(A) It is clear from Tables 2 and 3 that though the ratio estimator of the suggested class is biased, but the amount of bias is almost negligible in both the cases of equal and unequal fsu’s.

(B) Equal fsu's

(i) When X and Y are positively related, we compare MSE of usual two stage estimator of population mean with the MSE of both the estimators of the suggested class for equal fsu and we find that MSE($\bar{y}_{TS}$) =9.21412, which is significantly higher than MSE($\bar{y}_{RAT.TS.1}$) =3.52483, MSE($\bar{y}_{REG.TS.1}$) =3.28059 (See Table 2).

fsu	Equal				Unequal			
No. of clusters	1	2	3	4	1	2	3	4
M_i	16	16	16	16	18	14	12	20
m_i	8	8	8	8	9	7	6	10
$\overline{Y}_{i.}$	26.20625	24.12313	26.68875	22.11438	25.77722	22.79286	28.43500	23.09050
$\overline{X}_{i1.}$	50.96019	50.35994	62.70413	55.75731	51.06389	46.49700	67.00217	57.11855
$\overline{X}_{i2.}$	35.71519	41.85756	39.68550	48.71470	35.84517	39.49436	39.86467	48.95286
$\overline{X}_{i3.}$	56.48565	47.79563	27.95500	57.78263	52.39391	43.59071	30.69167	55.93210
C_{yi}^2	0.62364	0.33905	0.32637	0.36886	0.58025	0.39297	0.34783	0.31545
C_{xi1}^2	0.47888	0.28038	0.38836	0.49081	0.43322	0.29984	0.41947	0.40689
C_{xi2}^2	0.53798	0.24367	0.38462	0.20182	0.47630	0.26882	0.43302	0.20186
C_{xi3}^2	0.23426	0.27680	0.28155	0.10532	0.29194	0.28803	0.28366	0.15534
ρ_{i1}	0.88451	0.85254	0.84212	0.80242	0.88373	0.83895	0.82425	0.82113
ρ_{i2}	0.79978	0.71317	0.87276	0.79080	0.79943	0.67443	0.90076	0.80311
ρ_{i3}	0.70371	0.74068	0.80029	0.77797	0.66011	0.80597	0.81874	0.61370
ρ_{i12}	0.60065	0.68186	0.64406	0.58869	0.60618	0.61701	0.64034	0.62536
ρ_{i13}	0.62789	0.57917	0.47770	0.67925	0.55943	0.57812	0.45501	0.55727
ρ_{i23}	0.54930	0.54213	0.69703	0.69085	0.49031	0.55708	0.79852	0.52633

Table 1: The population parameters for population I (for equal fsu's) and II (for unequal fsu's) given in Appendix A.

(C) Unequal fsu's

Both the estimators of the suggested class are significantly more efficient than $\bar{y}_{TS}$ in terms of MSE for both the cases i.e. when $\alpha_i = 1$ and $\alpha_i = M_i/\bar{M}$. This is evident when we compare the MSE's.

(a) When X and Y are positively related

(i) MSE($\bar{y}_{TS}$) = 10.22425, MSE($\bar{y}_{RAT.TS.1}$) = 4.57177, MSE($\bar{y}_{REG.TS.1}$) = 4.15115 for $\alpha_i = 1$.

(ii) MSE($\bar{y}_{TS}$) = 13.89066, MSE($\bar{y}_{RAT.TS.1}$) = 8.30537, MSE($\bar{y}_{REG.TS.1}$) = 8.03492 for $\alpha_i = M_i / \bar{M}$ (see Table 3)

No. of used Auxiliary variables	Ratio			Regression		
	|Bias|	MSE	% R .E.	|Bias|	MSE	% R .E.
0	-	9.21412	0	-	9.21412	0
1	0.09331	3.52483	161.41	-	3.28059	180.87
2	0.07319	2.38863	285.75	-	2.51813	265.91
3	0.06935	2.19234	320.29	-	2.39819	284.21

Table 2: The biases and mean square errors for ratio and regression method of estimation with equal fsu's for population data set I (Table 1)

Estimators	No. of Used Auxiliary Variables	$\alpha_i = 1$			$\alpha_i = \frac{M_i}{\bar{M}}$		
		|Bias|	MSE	% R .E.	|Bias|	MSE	% R .E.
Two Stage	0	0.24077	10.22425	0	-	13.89066	0
Ratio	1	0.33658	4.57177	123.64	0.08740	8.30537	67.25
	2	0.32415	3.12462	227.22	0.07299	7.03199	97.54
	3	0.31095	2.81173	263.63	0.06269	6.85332	102.69
Regression	1	0.24077	4.15115	146.30	-	8.03492	72.88
	2	0.24077	3.16832	222.70	-	7.20959	92.67
	3	0.24077	2.96059	245.35	-	7.05727	96.83

Table 3: The biases and mean square errors for ratio and regression method of estimation with unequal fsu for population data set II (table 1)

It is to be noted that as we increase the number of auxiliary variables, the gain in efficiency of all the estimators of the suggested class increases for equal fsu's as well as for unequal fsu's (for both the cases i.e. α_i = 1 and $\alpha_i = M_i / \bar{M}$)

It is important to mention here that this increment for equal fsu's and for data set I is more significant in ratio estimator than regression estimator where it increased from 161 % to 320 %.

For unequal fsu's, relative gain in efficiency is more for α_i = *1* than $\alpha_i = M_i / \bar{M}$ for both the estimators of suggested class. For data set II as we increase the number of auxiliary variables, the % gain in relative efficiency is more for ratio estimator. Where it increased from 123 % to 263 %.

References

1. Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling, Applied mathematics and computation, 151, 893-902.
2. Olkin, I. (1958). Multivariate ratio estimation for finite population, Biometrika, 45, 154-165.
3. Raj, D. (1965). On a method of using multi-auxiliary information in sample survey, Journal of American Statistical Association, 60, 270-277.
4. Rao, P.S.R.S. and Mudholkar, G.S. (1967). Generalized multivariate estimator for the mean of finite population, Journal of American Statistical Association, 62, 1009-1012.
5. Rao, T. J. (1991). On certain methods of improving ratio and regression estimators, Communication in Statistics, Theory and Method, 20(10), 3325-3340.
6. Singh, M. (1982). A note on the use of multi-auxiliary information, Communication in Statistics, Theory and Method, 11(8), 933-939.
7. Singh, M. P. (1965). On the estimation of ratio and product of the population parameters, Sankhya, 27(C), 321-328.
8. Srivastava, S. K. (1971). A generalized estimator for the mean of a finite population using multi-auxiliary information, Journal of American Statistical Association ,June, 66(334), 404-407.
9. Srivastava, S. K. (1980). A class of estimators using auxiliary information in sample surveys, The Canadian Journal of Statistics, 8(2), 253-254.

Appendix A

Equal fsu's (Population Size = 64)
Population Set I

Cluster I					
Y_1	5.58	26.11	11.08	12.66	0.87
	6.40	54.21	3.25		
	37.94	56.92	27.59	45.98	61.21
	14.23	13.59	41.68		
X_{11}	13.18	60.55	22.36	37.18	3.27
	21.62	97.08	7.28		
	50.22	113.44	88.08	92.46	79.77
	56.12	18.07	54.68		
X_{12}	8.91	23.28	12.76	14.24	8.54
	9.48	67.43	15.31		
	41.25	44.84	25.35	37.19	98.41
	54.54	42.05	67.87		
X_{13}	32.51	69.03	42.49	63.37	37.73
	27.47	72.95	25.35		
	38.27	98.41	67.61	55.08	73.57
	29.13	47.48	123.31		
Cluster II					
Y_2	4.84	10.93	11.41	32.52	3.56
	12.52	34.63	35.97		
	47.07	17.69	41.24	15.48	34.35
	16.89	40.76	26.11		
X_{21}	10.92	25.64	35.17	42.78	12.15
	29.30	45.52	82.53		
	61.49	40.48	95.35	50.88	79.51
	39.25	94.25	60.55		
X_{22}	9.64	12.65	18.54	59.37	8.54
	28.88	69.54	51.87		
	41.25	39.56	47.27	61.44	49.83
	54.54	71.19	45.61		
X_{23}	12.52	27.88	43.05	63.37	9.08
	32.41	55.14	40.11		
	98.78	41.56	67.61	55.08	34.55
	29.13	63.27	91.19		
Cluster III					
Y_3	15.21	10.08	4.21	16.92	54.81
	40.05	52.55	29.54		
	19.64	26.24	24.74	47.23	12.18
	28.93	15.15	29.54		
X_{31}	34.77	23.68	9.48	22.50	126.46
	92.62	68.82	67.74		
	84.75	60.14	57.40	155.66	28.51
	66.94	35.35	68.44		
X_{32}	16.08	12.21	9.45	21.62	67.29
	42.78	92.15	30.07		

	20.78	60.23	41.15	64.57	15.08
	67.74	29.54	44.23		
X_{33}	14.24	14.09	9.78	23.44	42.11
	45.75	67.45	35.45		
	24.78	23.44	39.55	28.22	19.58
	20.08	18.77	20.55		
Cluster IV					
Y_4	15.79	11.18	17.41	37.02	23.54
	59.21	37.96	25.28		
	29.11	11.18	9.27	13.47	9.86
	21.70	12.21	19.64		
X_{41}	36.11	26.21	39.84	85.65	54.54
	136.68	61.24	57.94		
	154.58	25.50	21.82	44.25	23.18
	50.31	28.58	45.67		
X_{42}	34.48	65.12	74.23	61.27	45.14
	98.45	78.48	46.36		
	55.47	37.45	24.09	27.48	18.54
	40.89	35.45	36.54		
X_{43}	58.62	45.78	67.46	49.02	71.16
	98.47	79.75	69.17		
	74.15	29.96	54.45	46.72	41.46
	64.74	28.36	45.27		

Unequal fsu's (Population Size = 64)
Population Set II

Cluster I					
Y_1	5.58	26.11	11.08	12.66	0.87
	6.40	54.21	3.25		
	37.94	56.92	27.59	45.98	61.21
	14.23	13.59	41.68		
	15.15	29.54			
X_{11}	13.18	60.55	22.36	37.18	3.27
	21.62	97.08	7.28		
	50.22	113.44	88.08	92.46	79.77
	56.12	18.07	54.68		
	35.35	68.44			
X_{12}	8.91	23.28	12.76	14.24	8.54
	9.48	67.43	15.31		
	41.25	44.84	25.35	37.19	98.41
	54.54	42.05	67.87		
	29.54	44.23			
X_{13}	32.51	69.03	42.49	63.37	37.73
	27.47	72.95	25.35		
	38.27	98.41	67.61	55.08	73.57
	29.13	47.48	123.31		
	18.77	20.55			
Cluster II					
Y_2	4.84	10.93	11.41	32.52	3.56
	12.52	34.63	35.97		

	47.07	17.69	41.24	15.48	34.35
	16.89				
X_{21}	10.92	25.64	35.17	42.78	12.15
	29.30	45.52	82.53		
	61.49	40.48	95.35	50.88	79.51
	39.25				
X_{22}	9.64	12.65	18.54	59.37	8.54
	28.88	69.54	51.87		
	41.25	39.56	47.27	61.44	49.83
	54.54				
X_{23}	12.52	27.88	43.05	63.37	9.08
	32.41	55.14	40.11		
	98.78	41.56	67.61	55.08	34.55
	29.13				
Cluster III					
Y_3	15.21	10.08	4.21	16.92	54.81
	40.05	52.55	29.54		
	19.64	26.24	24.74	47.23	
X_{31}	34.77	23.68	9.48	22.50	126.46
	92.62	68.82	67.74		
	84.75	60.14	57.40	155.66	
X_{32}	16.08	12.21	9.45	21.62	67.29
	42.78	92.15	30.07		
	20.78	60.23	41.15	64.57	
X_{33}	14.24	14.09	9.78	23.44	42.11
	45.75	67.45	35.45		
	24.78	23.44	39.55	28.22	
Cluster IV					
Y_4	15.79	11.18	17.41	37.02	23.54
	59.21	37.96	25.28		
	29.11	11.18	9.27	13.47	9.86
	21.70	12.21	19.64		
	40.76	26.11	12.18	28.93	
X_{41}	36.11	26.21	39.84	85.65	54.54
	136.68	61.24	57.94		
	154.58	25.50	21.82	44.25	23.18
	50.31	28.58	45.67		
	94.25	60.55	28.51	66.94	
X_{42}	34.48	65.12	74.23	61.27	45.14
	98.45	78.48	46.36		
	55.47	37.45	24.09	27.48	18.54
	40.89	35.45	36.54		
	71.19	45.61	15.08	67.74	
X_{43}	58.62	45.78	67.46	49.02	71.16
	98.47	79.75	69.17		
	74.15	29.96	54.45	46.72	41.46
	64.74	28.36	45.27		
	63.27	91.19	19.58	20.08	