Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 6, Issue 1 (2013): 115-132

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

[1]Meenakshi Srivastava and [2]Neha Garg

[1]Department of Statistics, Institute of Social Sciences, Dr. B. R. Ambedkar University, Agra (U.P.), India.
[2]School of Sciences, Indira Gandhi National Open University, New Delhi, India.
Email: [1]msrivastava_iss@hotmail.com, [2]neha1garg@gmail.com

(Received November 01, 2012)

Abstract

When the population mean $\overline{X}$ of the auxiliary variate x is not known, it is well known that two-phase sampling is of significant use in practice. Keeping this in view, a general class of estimators is suggested to estimate the population mean for the variable under study using auxiliary variable in two stage double sampling scheme. Some special cases of this class of estimators are considered and compared by using a data set. Finally, it is shown, how to extend the class of estimators if multi auxiliary variables are available in the case of two stage double sampling scheme. The approximate expressions for bias and mean square error of the suggested estimators have also been derived and theoretical results are numerically supported.

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

1. Introduction

One of the major developments in sample surveys over the last five decades is the use of auxiliary variable x, correlated with the study variable y, in order to obtain the estimates of the population total or mean of the study variable. In large scale surveys, we often collect data on more than one auxiliary variable and some of these may be correlated with y. Olkin (1958), Srivastava (1971), Singh (1982), Diana and Perri (2007) etc. considered some estimators which utilize information on several auxiliary variables which are positively correlated with variable under study.

In sample surveys, the information on an auxiliary variable is required many times either at the estimation stage or at the selection stage, for increasing the efficiency of the estimator. Various estimation procedures in sample surveys need advance knowledge of some auxiliary variable x_i, which is then used to increase the precision of the estimates. If such information is lacking it is sometimes advantageous to take a large preliminary sample to observe the auxiliary variable and further, information on the character under study is collected on the basis of a sub-sample. This technique is known as double sampling. It is profitable only if the gain in precision is substantial as compared to the increase in the cost due to collection of information on the auxiliary variable for the larger sample.

For example, the classical ratio, regression and product estimators require the advance knowledge of population mean $\overline{X}$ of the auxiliary variable x. Some times the population mean $\overline{X}$ is unknown, then it is usual practice to estimate it from a large preliminary sample on which only the auxiliary characteristic x is observed. The value of $\overline{X}$ in the estimator is then replaced by its estimate say $\overline{x}'$. In the second phase the variate of interest y is then observed on a subsample. It is especially appropriate if the x_i values are easily accessible and much cheaper to collect than the y_i values. Some important works in this direction have discussed by several authors. (See : Dalabehara and Sahoo (2000), Chandra and Singh (2003), Diana and Tommasi (2003, 04), Roy (2003), Singh et. al. (2004), Singh and Espejo (2007), Singh et. al. (2010), Singh et. al. (2012), Das and Mishra (2011) etc.).

Generally the large scale surveys specially socio- economic or crop yield surveys are conducted in different stages. Some times such types of surveys also contain information on auxiliary variables which are related with the study variable y. For such a set up Srivastava and Garg (2009) have considered a general class of estimators for estimating the finite population mean using multi auxiliary information under two stage sampling scheme. But practically it happens that the population means for auxiliary variables are not known, then in this case the double sampling scheme is employed. For example in a crop surveys for estimating yield of a crop in a district, a block may be considered a primary sampling unit, the villages the second stage units, the crop fields the third stage units and a plot of fixed size the ultimate unit of sampling. This is termed as multi-stage sampling. The yield of the crop may depend on number of other factors like average rainfall, fertility of the soil, area under crop etc. It may happen that population means of such auxiliary variables may not be known. Deriving motivation from such type of situations, in this paper, two stage double sampling scheme using multi-auxiliary information is used for estimating population mean.

This paper suggests a general class of estimators in two-stage double sampling for equal and unequal first stage unit (fsu), when the population means of the auxiliary variables for all fsu's are unknown. Then the suggested general class of estimators is extended for the multivariate case, i.e. when p auxiliary variables are used. The proposed class of estimators dominates the usual two-stage estimator in terms of efficiency with more practical utility.

2. Notations

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

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

n : Total no. of fsu in the sample

M_i : Total no. of second stage unit (ssu) belonging to the i^{th} fsu in the population

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

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

m_i : Total no. of ssu selected from i^{th} fsu in the sample

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

Y : Variable under study

Y_{ij} : Observation on the variable under study for 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 variable under study for ssu in the i^{th} fsu,

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

$\overline{Y}$: 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 variable under study for j^{th} ssu belonging to the i^{th} fsu in the sample; $i = 1,2, ..., n$ and $j = 1, 2, ..., m_i$

$\overline{y}_{i.}$: Sample mean of variable under study of ssu in i^{th} fsu = $\frac{1}{m_i}\sum_{i=1}^{m_i} y_{ij}$

X_k : k^{th} auxiliary variable; $k = 1, 2, …, p$.

X_{ijk} :Value of k^{th} auxiliary variable on j^{th} ssu belonging to the i^{th} fsu in the population

$\overline{X}_{ik}$: Population mean of k^{th} auxiliary variable for ssu in i^{th} fsu

x_{ijk} : Value of k^{th} auxiliary variable on j^{th} ssu belonging to the i^{th} fsu in the sample

$\overline{x}_{ik}$: Sample mean of k^{th} auxiliary variable for ssu in i^{th} fsu

α_i : Weight for i^{th} fsu

S_y^2 : Population mean square error of y variable

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

S_{yi}^2 : Population mean square error of y variable for i^{th} fsu

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

S_{xik}^2 : Population mean square error of k^{th} auxiliary variable for i^{th} fsu

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

C_{yi}^2 : Coefficient of variation of the variable under study Y for i^{th} fsu $= \frac{S_{yi}^2}{\overline{Y}_{i.}^2}$

C_{xik}^2 : Coefficient of variation of the k^{th} auxiliary variable for i^{th} fsu $= \frac{S_{xik}^2}{\overline{X}_{ik.}^2}$

ρ_{ik} : Correlation coefficient between the variables *Y*, variable under study and X_k, k[th] auxiliary variable for i[th] fsu

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

b_{ij} : Regression coefficient between the variables *Y*, variable under study and X_j, j[th] auxiliary variable for i[th] fsu for the sample

B_{ij} : Regression coefficient between the variables *Y*, variable under study and X_j, j[th] auxiliary variable for i[th] fsu for the population

m'_i : Total no. of ssu selected from i[th] fsu in the sample; $i = 1,2, ..., n$

Where $(m'_i > m_i)$

$\bar{x}'_{ik}$: Sample mean of k[th] auxiliary variable for ssu in i[th] fsu based on preliminary sample of size m'_i

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

$$f'_i = \left(\frac{1}{m_i} - \frac{1}{m'_i}\right), \qquad f''_i = \left(\frac{1}{m'_i} - \frac{1}{M_i}\right)$$

Define

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

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

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

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

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

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

$$E(e_{i0}) = E(e'_{ij}) = 0, \qquad j = 1, 2, \dots, p$$

$$E(e_{io}^2) = f_i C_{yi}^2, \quad E(e'^2_{ij}) = f''_i C_{xij}^2, \; j = 1, 2, \dots, p$$

$$E(e_{io} e'_{ij}) = f''_i \rho_{ij} C_{yi} C_{xij}, \; j = 1, 2, \dots, p$$

$$E(e'_{ij} e_{ik}) = E(e_{ij} e'_{ik}) = E(e'_{ij} e'_{ik}) = f''_i \rho_{ijk} C_{xij} C_{xik}, \quad j = 1, 2, \dots, p$$

3. Estimator and its Mean Square Error

The usual ratio estimator in double sampling for population mean is given as

$$\bar{y}_{rat.d} = \frac{\bar{y}}{\bar{x}} \bar{x}' \tag{1}$$

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

$$\text{Bias}(\bar{y}_{rat.d}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \bar{Y} \left(C_x^2 - \rho \; C_y C_x\right) \tag{2}$$

$$MSE(\bar{y}_{rat.d}) = \bar{Y}^2\left\{\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left(C_x^2 - 2\rho\ C_y C_x\right)\right\} \tag{3}$$

The usual regression estimator in double sampling for population mean is given as

$$\bar{y}_{reg.d} = \bar{y} + b(\bar{x}' - \bar{x}) \tag{4}$$

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

$$Bias(\bar{y}_{reg,d}) = -B\ \left(\frac{1}{n}-\frac{1}{n'}\right)\left(\frac{\mu_{21}}{S_{xy}} - \frac{\mu_{30}}{S_x^2}\right), \text{ If b is unknown} \tag{5}$$

If b_i is known i.e. $b_i = B_i$,

$$Bias(\bar{y}_{reg,d}) = 0$$

$$MSE(\bar{y}_{reg.d}) = \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left[B^2 S_x^2 - 2B\rho\ S_y S_x\right] \tag{6}$$

The usual product estimator in double sampling for population mean is given as

$$\bar{y}_{prod.d} = \frac{\bar{y}}{\bar{x}'}\bar{x} \tag{7}$$

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

$$Bias(\bar{y}_{prod.d}) = \left(\frac{1}{n}-\frac{1}{n'}\right)\bar{Y}\rho\ C_y C_x \tag{8}$$

$$MSE(\bar{y}_{prod.d}) = \bar{Y}^2\left\{\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left(C_x^2 + 2\rho\ C_y C_x\right)\right\} \tag{9}$$

The usual two stage estimator for population mean is given as

$$\bar{y}_{TS} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_i \tag{10}$$

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

$$Bias(\bar{y}_{TS}) = \frac{1}{N\bar{M}}\sum_{i=1}^{N}(\alpha_i \bar{M} - M_i)\bar{Y}_{i.} \tag{11}$$

$$MSE(\bar{y}_{TS}) = \frac{f}{N-1}\sum_{i=1}^{N}\left[\alpha_i \bar{Y}_{i.} - \frac{1}{N}\sum_{i=1}^{N}\alpha_i \bar{Y}_{i.}\right]^2 + \frac{1}{nN}\sum_{i=1}^{N} f_i \alpha_i^2 \bar{Y}_{i.}^2 C_{yi}^2 \tag{12}$$

4. Suggested Class of Estimators in Two Stage Double Sampling

When $\bar{X}_i$ is unknown, we use double sampling in two stage sampling scheme and proposed the estimator of population mean is defined as

$$\bar{y}_{GTSD} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{igd} \tag{13}$$

Where ‘GTSD’ stands for ‘General Estimator in Two Stage Double Sampling Scheme’ and $\bar{y}_{igd}$ is a function of $\bar{y}$, $\bar{x}_i'$ and $\bar{x}_i$ in i^{th} fsu

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

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

Proof. $E(\bar{y}_{GTSD}) = E[E(\bar{y}_{GTSD}/i)]$

$$= \frac{1}{n} E \sum_{i=1}^{n} \alpha_i z_{id} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i z_{id}$$

where $z_{id} = E(\bar{y}_{igd}/i)$

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

$$= \frac{1}{N\overline{M}} \sum_{i=1}^{N} \left[\alpha_i z_{id} \overline{M} - M_i \overline{Y}_{i.}\right] \tag{14}$$

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

$$MSE(\bar{y}_{RTSD}) = \left(\frac{1}{n} - \frac{1}{N}\right) \cdot \frac{1}{N-1} \sum_{i=1}^{N} \left[\alpha_i z_{id} - E(\alpha_i z_{id})\right]^2 +$$

$$\frac{1}{nN} \sum_{i=1}^{N} \alpha_i^2 v_{id}$$

where $z_{id} = E(\bar{y}_{igd}/i)$ and $v_{id} = MSE(\bar{y}_{igd}/i)$

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

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

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \cdot \frac{1}{N-1} \sum_{i=1}^{N} \left[\alpha_i z_{id} - E(\alpha_i z_{id})\right]^2 \tag{15}$$

Where $E(\alpha_i z_{id}) = \frac{1}{N}\sum_{i=1}^{N} \alpha_i z_{id}$

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

$$= \frac{1}{n^2}\sum_{i=1}^{n} \alpha_i^2 E(v_{igd})$$

Where $v_{id} = MSE(\bar{y}_{igd}/i)$

$$= \frac{1}{nN}\sum_{i=1}^{N} \alpha_i^2 v_{id} \tag{16}$$

Adding (15) and (16), we get the final expression.

5. Special 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

$$\bar{y}_{RAT.TSD} = \frac{1}{n}\sum_{i=1}^{n} \alpha_i \bar{y}_{id.rat} \tag{17}$$

Where $\bar{y}_{id.\,rat} = \frac{\bar{y}_i}{\bar{x}_i}\bar{x}'_i$ is the usual ratio estimator in i^{th} fsu with

$$z_{id} = \bar{Y}_{i.}\left[1 + f'_i\left(C^2_{xi} - \rho_i C_{yi} C_{xi}\right)\right]$$

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

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

$$\bar{y}_{REG.\,TSD} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{id.\,reg} \tag{18}$$

Where $\bar{y}_{id.\,reg} = \bar{y}_i + b_i\left(\bar{x}'_i - \bar{x}_i\right)$ is the usual regression estimator in i^{th} fsu with

If b_i is unknown,

$$z_{id} = \bar{Y}_i - B_i f'_i\left(\frac{\mu_{21.i}}{S_{xy.i}} - \frac{\mu_{30.i}}{S^2_{xi}}\right)$$

This is a biased estimate. biasness is of the order of the linear regression estimator. Biasness will be negligible when sample size is sufficiently large and regression coefficient b_i is known.

If bi is known i.e. $b_i = B_i$, $z_i = \bar{Y}_i$

$$v_i = f_i S^2_{yi} + f'_i\left[B^2_i S^2_{xi} - 2B_i \rho_i S_{yi} S_{xi}\right]$$

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

$$\bar{y}_{PROD.\,TSD} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \bar{y}_{id.\,prod} \tag{19}$$

Where $\bar{y}_{id.\,prod} = \frac{\bar{y}_i}{\bar{x}'_i}\bar{x}_i$ is the usual product estimator in i^{th} fsu with

$$z_{id} = \bar{Y}_{i.}\left[1 + f'_i \rho_i C_{yi} C_{xi}\right]$$

and $$v_i = \bar{Y}^2_{i.}\left\{f_i \; C^2_{yi} + f'_i\left(C^2_{xi} + 2\rho_i C_{yi} C_{xi}\right)\right\}$$

6. Generalization of the Suggested Class of Estimators

When $\bar{X}_i$ is unknown, the generalized class of two stage double sampled estimators using p auxiliary variables for every i^{th} fsu is given by

$$\bar{y}_{GTSD.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\bar{y}_{ijgd} \tag{20}$$

Where 'GTSD.p' stands for 'General estimator in two stage double sampling scheme for p auxiliary variables ' with $\sum_{j=1}^{p} w_{ij} = 1$ and $\bar{y}_{ijgd}$ is a function of $\bar{y}_i$, $\bar{x}'_{ij}$ and $\bar{x}_{ij}$ for j^th auxiliary variable in i^th fsu.

6.1 Bias and MSE of the Suggested Class of Estimators

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

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

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

$$= \frac{1}{n}E\left[\sum_{i=1}^{n}\alpha_i z_{id}\right] = \frac{1}{N}\sum_{i=1}^{N}\alpha_i z_{id}$$

Where $z_{id} = \sum_{j=1}^{p} w_{ij} E\left(\bar{y}_{ijgd} / i\right)$

$$E\left(\bar{y}_{GTSD.p}\right) - \bar{Y} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i z_{id} - \frac{1}{N}\sum_{i=1}^{N}\frac{M_i}{\bar{\bar{M}}}\bar{Y}_{i.}$$

$$= \frac{1}{N\bar{\bar{M}}}\sum_{i=1}^{N}\left[\alpha_i z_{id}\bar{M} - M_i \bar{Y}_{i.}\right] \tag{21}$$

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

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

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

$$= \left(\frac{1}{n} - \frac{1}{N}\right).\frac{1}{N-1}\sum_{i=1}^{N}\left[\alpha_i z_{id} - E\left(\alpha_i z_{id}\right)\right]^2 \tag{22}$$

Where $E\left(\alpha_i z_{id}\right) = \frac{1}{N}\sum_{i=1}^{N}\alpha_i z_{id}$

$$E\left[MSE\left(\bar{y}_{GTSD.p} / i\right)\right] = \frac{1}{n^2}E\left[\sum_{i=1}^{n}\alpha_i^2 MSE\left(\sum_{j=1}^{p} w_{ij}\bar{y}_{ijgd} / i\right)\right] \tag{23}$$

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

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

Where

$$\left(\frac{1}{m_i}-\frac{1}{M_i}\right)v_{ijhd} = Cov(\bar{y}_{ijgd}, \bar{y}_{ihgd})$$

In matrix notation,

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

Where the matrix $V_{id}=(v_{ijhd})$ and $\underset{\sim}{w_i} = (w_{i1}, w_{i2}, \ldots\ldots\ldots\ldots, w_{ip})$, $\underset{\sim}{w_i'}$ being the transpose of $\underset{\sim}{w_i}$.

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

It is fairly simple to establish that the optimum w_{ij} is given by

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

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

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

Adding (22) and (23), we get the final expression.

Remark (i) To avoid the mathematical complexity in deriving $MSE\left(\sum_{j=1}^{p} w_{ij}\bar{y}_{ijgd} / i\right)$ for different values of function $\bar{y}_{ijg}$, we will use the procedure given in section (6.1) 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 others

7. Special Cases for the Generalized Class of Estimators

Case 1: Multivariate Ratio Estimator

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

$$\bar{y}_{RAT.TSD.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\bar{y}_{ijd.rat} \tag{24}$$

Where, $\bar{y}_{ijd.rat} = \frac{\bar{y}_i}{\bar{x}_{ij}}\bar{x}'_{ij}$ is the usual ratio estimator in i^{th} fsu for j^{th} auxiliary variable.

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

$$v_{ijh.rat.d} = \overline{Y}_i^2\left[f_iC_{yi}^2 + f_i'\left(\rho_{ijh}C_{xij}C_{xih} - \rho_{ij}C_{yi}C_{xij} - \rho_{ih}C_{yi}C_{xih}\right)\right]$$

Case 2: Multivariate Regression Estimator

The combined regression estimator $\overline{y}_{REG.TSD.p}$, of $\overline{Y}$ in two stage estimators when p auxiliary variables are known for every ith fsu is given by

$$\overline{y}_{REG.TSD.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\overline{y}_{ijd.reg} \tag{25}$$

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

The bias and MSE of $\overline{y}_{REG.TS.p}$ have been computed for known b_{ij}

If b_{ij} is known i.e. $b_{ij} = B_{ij}$ then $z_{id} = \overline{Y}_{i.}$

For $\alpha_i = \frac{M_i}{\overline{M}}$, this estimator will be unbiased

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

Case 3: Multivariate Product Estimator

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

$$\overline{y}_{PROD.TSD.p} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \sum_{j=1}^{p} w_{ij}\overline{y}_{ijd.prod} \tag{26}$$

Where $\overline{y}_{ijd.prod} = \frac{\overline{y}_i}{\overline{x}_{ij}'}\overline{x}_{ij}$ is the usual product estimator in ith fsu for jth auxiliary variable.

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

$$v_{ijh.prod.d} = \overline{Y}_i^2\left[f_iC_{yi}^2 + f_i'\left(\rho_{ijh}C_{xij}C_{xih} + \rho_{ij}C_{yi}C_{xij} + \rho_{ih}C_{yi}C_{xih}\right)\right]$$

8. Numerical Illustration

For this purpose, we consider a simulated data of N=4 clusters as fsu with equal number of fsu and other populations 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 can be selected in proportion to

M_i, i.e. $m_i = \frac{M_i}{\sum_{i=1}^{N} M_i} \times 32$. Preliminary sample of size 52 in ssu can be selected in proportion to M_i, i.e. $m_i^{'} = \frac{M_i}{\sum_{i=1}^{N} M_i} \times 52$

fsu	Equal				Unequal			
No. of clusters	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
M_i	16	16	16	16	18	14	12	20
$m_1^{'}$	13	13	13	13	15	11	10	16
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) and II (for unequal fsu) given in Appendix A.

No. of used Auxiliary variable	Ratio			Regression		
	|Bias|	*MSE*	*% R .E.*	*|Bias|*	*MSE*	*% R .E.*
0	-	9.21412	0	-	9.21412	0
1	0.07178	4.83729	90.48	-	4.64987	98.16
2	0.05630	3.96365	132.47	-	4.06338	126.76
3	0.05335	3.81270	141.67	-	3.97110	132.03

Table 2: The biases and mean square errors for ratio and regression estimators with equal fsu for population data set I (table 1)

Estimators	*Used Auxiliary Variable*	$\alpha_i = 1$			$\alpha_i = \frac{M_i}{\overline{M}}$		
		|Bias|	*MSE*	*% R .E.*	*|Bias|*	*MSE*	*% R .E.*
Two Stage	0	0.24077	10.22425	0	-	13.89066	0
Ratio	1	0.31544	5.82769	75.44	0.06758	9.53597	45.67
	2	0.30610	4.68669	118.16	0.05692	8.54441	62.57
	3	0.29573	4.44895	129.81	0.04885	8.41089	65.15
Regression	1	0.24077	5.49452	86.08	-	9.33288	48.84
	2	0.24077	4.71722	116.74	-	8.68406	59.96
	3	0.24077	4.56347	124.05	-	8.57176	62.05

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

9. Discussion and Conclusion

(A) It is clear from tables 2 and 3 that though the ratio estimators of the constructed class are biased but the amount of bias is not significantly high for equal and unequal fsu with $\alpha_i = 1$ and $\alpha_i = \frac{M_i}{\overline{M}}$. The regression estimators of the suggested class are biased with very small amount for unequal fsu, for $\alpha_i = 1$ and the amount of bias remains same as we increase the number of auxiliary variables as the expression of bias is independent of the number of auxiliary variables. For equal fsu and unequal fsu with $\alpha_i = \frac{M_i}{\overline{M}}$ the regression estimators are unbiased. Though for both the cases, they have been derived for known value of b_{ij}

(B) EQUAL FSU

(i) It is to be noted from the table 2, the MSE of usual two stage estimator of population mean MSE($\bar{y}_{TS}$) =9.21412, is substantially higher than MSE($\bar{y}_{RAT.TSD.1}$) =4.83729, MSE($\bar{y}_{REG.TSD.1}$) =4.64987.

(C) UNEQUAL FSU

It is important to note that all the estimators of the suggested class dominate over $\bar{y}_{TS}$ in terms of MSE's (see table 3) for both the cases i.e. when $\alpha_i = 1$ and $\alpha_i = \frac{M_i}{\overline{\overline{M}}}$ for the data set taken.

(i) MSE($\bar{y}_{TS}$)=10.22425, which is substantially higher than MSE $\bar{y}_{RAT.TSD.1}$)=5.82769, MSE($\bar{y}_{REG.TSD.1}$)=5.49452for $\alpha_i = 1$.

(i) MSE($\bar{y}_{TS}$)=13.89066, which too is substantially higher than MSE($\bar{y}_{RAT.TSD.1}$)=9.53597, MSE($\bar{y}_{REG.TSD.1}$)=9.33288 for $\alpha_i = \frac{M_i}{\overline{\overline{M}}}$ (see table 3)

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 as well as for unequal fsu (for both the cases i.e. $\alpha_i = 1$ and $\alpha_i = \frac{M_i}{\overline{\overline{M}}}$)

It is important to mention here that this increment for equal fsu and for data set I, is more significant in ratio estimator, where it increased from 90 % to 141 %.

For unequal fsu, relative gain in efficiency is more for $\alpha_i = 1$ than $\alpha_i = \frac{M_i}{\overline{\overline{M}}}$ for all 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. It has increased from 75 % to 129%.

References

1. Chandra, P. and Singh, H.P. (2003). A family of unbiased estimators in two phase sampling using two auxiliary variables, Statistics in Transition, 6, 1, p. 131-141.
2. Dalabehara, M. and Sahoo, L.N.(2000). An Unbiased estimator in two phase sampling using two auxiliary variables, Jour. Ind. Soc. Ag. Statistics, 53, 2, p. 134-140.
3. Dash, P. R. and Mishra, G. (2011). An Improved Class of Estimators in Two-Phase Sampling Using Two Auxiliary Variables, Communications in Statistics - Theory and Methods, 40, 24, p. 4347-4352.
4. Diana, G. and Perri, P.F. (2007). Estimation of finite population mean using multi-auxiliary information, Metron-International Journal of Statistics, Vol. LXV, 1, p. 99-112.
5. Diana, G. and Tommasi, C.(2003). Optimal estimation for finite population mean in two phase sampling, Statistical Methods and applications, 12, 1, p. 41-48.

6. Diana, G. and Tommasi, C. (2004). Optimal use of two auxiliary variables in double sampling, Statistical Methods and applications, 13, 3, p. 275-284.
7. Olkin, I. (1958). Multivariate ratio estimation for finite population, Biometrika. 45, p. 154-165.
8. Roy, D.C. (2003). A regression type estimator in two phase sampling using two auxiliary variables, Pakistan Journal of Statistics, 19, 3, p. 281-290.
9. Singh, H.P. and Espejo, M.R. (2007). Double Sampling Ratio-product Estimator of a Finite Population Mean in Sample Surveys, Journal of Applied Statistics, 34(1), p. 71-85.
10. Singh, H.P., Tailor, R. and Tailor, R. (2012). Estimation of finite population mean in two - phase sampling with known coefficient of variation of an auxiliary character, Statistica, 72(1), p. 111-126.
11. Singh, H.P., Singh, S. and Kim, J.M. (2010). Efficient use of auxiliary variables in estimating finite population variance in two-phase sampling. Communications of the Korean Statistical Society, 17(2), p. 165–181.
12. Singh, H.P., Upadhyaya, L.N. and Chandra, P. (2004). A general family of estimators for estimating population mean using two auxiliary variables in two-phase sampling, Statistics in Transition, 6(7), p. 1055-1077.
13. Singh, M. (1982). A note on the use of multi-auxiliary information, Communication in Statistics, Theory and Method, 11(8), p. 933-939.
14. Srivastava, M. and Garg, N. (2009). A general class of estimators of a finite population mean using multi-auxiliary information under two stage sampling scheme, Journal of Reliability and Statistical Studies, 2(1), p. 103-118.
15. Srivastava, S. K. (1971). A generalized estimator for the mean of a finite population using multi-auxiliary information, Journal of American Statistical Association, 66(334), p. 404-407.

Appendix A

Equal FSU (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	7.45	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 (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