Aligarh Journal of Statistics
Vol. 29 (2009), 51-73

A CLASS OF ESTIMATORS OF FINITE POPULATION MEAN USING INCOMPLETE MULTI-AUXILIARY INFORMATION WHEN FRAME IS UNKNOWN

Meenakshi Srivastava and Neha Garg

ABSTRACT

Frequently, there may arise situations where we have information on several auxiliary variables only for some part of the population. The maximum utilization of incomplete multi-auxiliary information is carried out in such cases by stratifying the population on the basis of available multi-auxiliary information at hand. In this paper a class of estimators is considered for estimating the mean of the finite population utilizing available incomplete multi-auxiliary information when frame is unknown for each stratum. Some special cases of this class of estimators are considered. The approximate expressions for bias and mean square error of the suggested estimators have also been derived and theoretical results are numerically supported.

1. INTRODUCTION

In most applications of stratified sampling, the prior knowledge of strata sizes, strata frames and possible variability within stratum are essential requirements. In practical situations, strata sizes are known but lists of stratum units are hard to get. Moreover, stratum frames may be incomplete. So one cannot apply stratified sampling for estimating the population parameters. When such type of situation occurs it is advised to use post-stratification technique. Post stratification means stratification after selection of the sample. The technique consists in selecting a random sample from the entire population and classifying units later according to their representation from different stratums.

The role of post stratification is well recognized as an effective method for obtaining more accurate estimates of population quantities in the context of survey sampling when frame in each stratum is not known. [See: Holt and Smith (1979), Jager *et.al.* (1985), Jager (1986), Smith (1991), Agarwal and Panda (1993, 1995), Shukla and Trivedi (2003) etc.].

The use of auxiliary information for improving the precision of the estimators is well known when the variable under study, Y and the auxiliary variable X are correlated. 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 positively correlated with the variable under study.

In many situations, we may have information on several auxiliary variables but each variable may not be known for each population unit. Singh (1977) has considered the concept of stratification for weighting the given incomplete auxiliary information.

The aim of the present paper is to develop a general class of weighted estimators based on incomplete multi- auxiliary information under post stratified setup. This is done in order to show that it is always better to use additional auxiliary variables, which are correlated with Y . The stratification of the units is done on the basis of incomplete multi-auxiliary information. It is seen that the given method is capable of giving more precise results than simple sample mean per unit.

Generally, the stratification is done on the basis of heterogeneity in the population with respect to the study variable Y . But we view stratification in the other way. Here, the heterogeneity of the population is considered with respect to the unequal number of auxiliary variables. We have stratified the population in terms of the information provided by the p auxiliary variable to make the homogeneous strata in terms of numbers of auxiliary variables. Thus, there be a stratum for which no auxiliary variable is known, p strata for which only one auxiliary variable out of p auxiliary variables is known. Similarly there will be ${}^{p}C_{2}$ strata for which the two auxiliary variables are known, ${}^{p}C_{3}$ strata for which three auxiliary variables are known, and so on. Ultimately, we will have a stratum for which all the p auxiliary variables are known. 2^{p} is the maximum possible number of strata. The number of strata can be less than 2^{p} also depending upon the auxiliary variables on which we have complete information.

In section 3, the bias and mean square error of the suggested general class of estimators have been derived. Section 4 deals with the some special cases of it. In section 5, an empirical study is carried out on three data sets.

2. NOTATIONS

Let us consider a finite population $U = (U_1, U_2, \ldots, U_N)$ of N identifiable units taking values on a study variable Y and p auxiliary variables $X_1, X_2, \ldots, X_p$, which are correlated with Y . Auxiliary variables $X_1, X_2, \ldots, X_p$ are known for total $M_1, M_2, \ldots, M_p$ units of the population respectively. For the maximum utilization of available incomplete auxiliary information, the population is divided into different strata according to the known number of auxiliary variables and a random sample of n units is drawn from these groups with simple random sampling without replacement.

${}^{p}C_{j}$: Number of strata for which j auxiliary variables are known; $j = 0, 1, 2, \ldots, p$.

N : Population size

n : Sample size

N_0 : Size of the stratum for which no auxiliary variable is known

N_{ii} : Size of the stratum for which 1 auxiliary variable X_i is known; $i = 1, 2, \ldots, p$

N_{ij} : Size of the stratum for which 2 auxiliary variable X_i and X_j are known; $i < j$; $i, j = 1, 2, \ldots, p$

N_{ijk} : Size of the stratum for which 3 auxiliary variable X_i, X_j and X_k are known; $i < j < k$; $i, j, k = 1, 2, \ldots, p$

$N_{1,2,\ldots,p}$: Size of the strata for which all p auxiliary variable $X_1, X_2, \ldots, X_p$ is known; $i < j < k$; $i, j, k = 1, 2, \ldots, p$, where

$$N_{ii} + N_{ij} + N_{ijk} + \ldots + N_{1,2,\ldots,i,\ldots,p} = M_i$$

2^p : Total number of strata, i.e. $\sum_{i=0}^{p} {}^pC_i = 2^p$

N_i : Population size of the $i-$th stratum; $i = 1, 2, \ldots, 2^p$, such that $\sum_{i=1}^{2^p} N_i = N$

n_i : Sample size of the $i-$th stratum; $i = 1, 2, \ldots, 2^p$, such that $\sum_{i=1}^{2^p} n_i = n$

Y_{ik} : Value of the $k-$th observation on variable under study in $i-$th stratum; $i = 1, 2, \ldots, {}^pC_j$; $j = 0, 1, 2, \ldots, p$; $k = 0, 1, 2, \ldots, N_i$

X_{ijk} : Value of the $k-$th observation on $j-$th auxiliary variable in $i-$th stratum; $i = 1, 2, \ldots, {}^pC_j$; $j = 0, 1, 2, \ldots, p$; $k = 0, 1, 2, \ldots, N_i$

$\bar{Y}_i$: Population mean of the Y variable in $i-$th stratum

$\bar{y}_i$: Sample mean of the Y variable in $i-$th stratum

$\bar{X}_{ij}$: Population mean of the $j-$th auxiliary variable in $i-$th stratum

$\bar{x}_{ij}$: Sample mean of the $j-$th auxiliary variable in $i-$th stratum

S_i^2 : Population mean square error of Y variable in $i-$th stratum

S_{ij}^2 : Population mean square error of $j-$th auxiliary variable in $i-$th stratum

C_i^2 : Coefficient of variation of the variable under study Y in $i-$th stratum, i.e. $C_i^2 = \frac{S_i^2}{\bar{Y}_i^2}$

C_{ij}^2 : Coefficient of variation of the $j-$th auxiliary variable in $i-$th stratum, i.e. $C_{ij}^2 = \frac{S_{ij}^2}{\bar{X}_{ij}^2}$

ρ_{ij} : Correlation coefficient between the variables Y, variable under study and X_{ij}, $j-$th auxiliary variable in $i-$th stratum

ρ_{ijh} : Correlation coefficient between the variables X_i and X_h $(j \neq h)$ in $i-$th stratum

b_{ij} : Regression coefficient of the variables Y, variable under study and X_{ij}, $j-$th auxiliary variable in $i-$th stratum

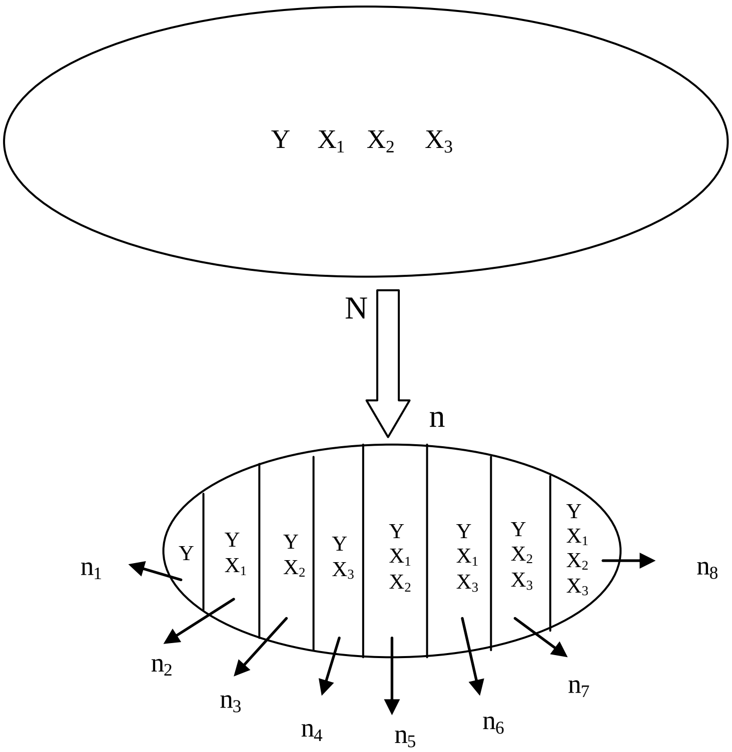
W_i : Proportion of units in the $i-$th stratum, i.e. $W_i = \frac{N_i}{N}$

f : Sampling fraction, i.e. $f = \frac{n}{N}$

f_i : Sampling fraction in the $i-$th stratum, i.e. $f_i = \frac{n_i}{N_i}$

α_{ij} : Weights, attached to the $j-$th auxiliary variable in $i-$th stratum adding up to unity i.e. $\sum_{j=1}^{p} \alpha_{ij} = 1$

The following figure shows the construction of strata $(2^3 = 8)$ on the basis of available incomplete three-auxiliary information $(X_1, X_2$ and $X_3)$ after the selection of sample:-



In this example, suppose we have complete information on X_1 then there will be $2^{3-1} = 4$ strata. In general, out of p auxiliary variables if the complete information on q auxiliary variables is available then the total number of strata will be 2^{p-q}.

3. SUGGESTED GENERAL CLASS OF ESTIMATORS

The general class of estimators using post stratification on the basis of incomplete multi-auxiliary information can be defined as:-

$$\bar{y}'_{pr} = \sum_{i=1}^{2^p} W_i g_i(y_i, x_i), \tag{3.1}$$

where $g_i(y_i, x_i) = \sum_{j=1}^{p} \alpha_{ij} g_{ij}(y_i, x_{ij})$ and $g_{11}(y_1, x_{11}) = \bar{y}_1$

$g_{ij}(y_i, x_{ij})$ is the function of $\underline{y}_i = (y_{ik}; k = 1, 2, \ldots, N_i)$ and $\underline{x}_{ij} = (x_{ijk};$ $k = 1, 2, \ldots, N_i)$.

Bias and *MSE* :

The conditional Bias and MSE of $\bar{y}'_{pr}$ given n_i are as follows:

$$E(\bar{y}'_{pr}) = E[E(\bar{y}'_{pr} \,|\, n_i)]$$

$$= \sum_{i=1}^{2^p} W_i \sum_{j=1}^{p} \alpha_{ij} \; E[E\, g_{ij}\,(y_i, x_{ij}) | n_i] \qquad (3.2)$$

$$Bias(\bar{y}'_{pr}) = E(\bar{y}'_{pr}) - \bar{Y}$$

The conditional MSE is a sum of two components.

$$MSE(\bar{y}'_{pr}) = E[MSE(\bar{y}'_{pr} \,|\, n_i)] + MSE[E(\bar{y}'_{pr} \,|\, n_i)], \qquad (3.3)$$

where

$$MSE(\bar{y}'_{pr} \,|\, n_i) = \sum_{i=1}^{2^p} W_i^2 \; MSE[g_i\,(y_i, x_i)].$$

Note: $MSE[g_i\,(y_i, x_i)]$ can easily be obtained for different values of function g by generalizing the procedure used by Olkin (1958). Thus,

$$MSE[g_i\,(y_i, x_i)] = \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \sum_{j=1}^{p} \sum_{h=1}^{p} \alpha_{ij}\; \alpha_{ih}\; v_{ijh} \;, \qquad (3.4)$$

where

$$\left(\frac{1}{n_i} - \frac{1}{N_i}\right) v_{ijh} \; = Cov(g_{ij}, g_{ih}).$$

In matrix notation

$$MSE[g_i\,(y_i, x_i)] = \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \underline{\alpha}_i \, V_i \, \underline{\alpha}'_i \,,$$

where the matrix $V_i = (v_{ijh})$ and $\underline{\alpha} = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ip})$, $\underline{\alpha}_i'$ being the transpose of $\underline{\alpha}_i$.

Optimum Values of $\boldsymbol{\alpha_{ij}}$ for $\boldsymbol{j} = 1, 2, \ldots, \boldsymbol{p}$:

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

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

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

$$MSE[g_i\,(y_i,x_i)]=\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\frac{1}{Sum of\ all\ the\ p^2\ elements in\ V_i^{-1}}\,.$$

Remark 3.1: It is to be noted that $\bar{y}'_{pr}$ could be used if the parameters b, C_y, C_x and ρ_{yx} are known. As remarked by Murthy (1967, p-96), Sahai and Sahai (1985) and Tracy and Singh (1999), these parameters are stable quantities, therefore, these can be known either from the past studies or from the experience gathered in due course of time.

To evaluate the expressions of bias and MSE, we have used the following standard results

$$E\left(\frac{1}{n_i}\right)=\frac{1}{nW_i}+\frac{(N-n)(1-W_i)}{(N-1)\,n^2W_i^2} \tag{3.5}$$

$$MSE\left(\frac{1}{n_i}\right)=\frac{(N-n)(1-W_i)}{(N-1)\,n^3\,W_i^3} \tag{3.6}$$

$$Cov\left(\frac{1}{n_i},\frac{1}{n_j}\right)=-\frac{(N-n)}{(N-1)\,n^2\,W_i\,W_j}\ . \tag{3.7}$$

4. SPECIAL CASES FOR THE CLASS OF ESTIMATORS

Case 1: When each X_j, $j=1,2,\ldots,p$ is positively correlated with Y in each stratum, our estimator will convert into weighted ratio estimator, given as

$$\bar{y}'_{pr.rat}=\sum_{i=1}^{2^p}W_i\sum_{j=1}^{p}\alpha_{ij}\,g_{ij.rat}\,(y_i,x_{ij})\,, \tag{4.1}$$

where $g_{ij.rat}\,(y_i,x_{ij})=\dfrac{\bar{y}_i}{\bar{x}_{ij}}\bar{X}_{ij}$.

To the first-degree approximation, the bias and MSE are respectively given by

$$E[\bar{y}'_{pr.rat}]=E[E(\bar{y}'_{pr}\,|n_i)]$$

$$=\bar{Y}+\sum_{i=1}^{2^p}W_i\bar{Y}_i\left[\left\{\frac{1}{nW_i}+\frac{(N-n)(1-W_i)}{(N-1)\,n^2\,W_i^2}\right\}-\frac{1}{N_i}\right]\left[\sum_{j=1}^{p}\alpha_{ij}\,C_{ij}^2\,(1-K_{ij})\right]$$

from equation (3.5).

The conditional MSE is a sum of two components.

$$MSE(\bar{y}'_{pr.rat})=E[MSE(\bar{y}'_{pr.rat}\,|n_i)]+MSE[E(\bar{y}'_{pr.rat}\,|n_i)]$$

Consider the first component

$$E[MSE(\bar{y}'_{pr.rat}\,|n_i) = E\left[\sum_{i=1}^{2^p} W_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\sum_{j=1}^{p}\sum_{h=1}^{p}\alpha_{ij}\,\alpha_{ih}\,v_{ijh.rat}\right],$$

where

$$v_{ijh.rat} = \bar{Y}_i^2\,[C_i^2 + \rho_{ijh}C_{ij}C_{ih} - \rho_{ij}C_iC_{ij} - \rho_{ih}C_iC_{ih}]$$

$$= \sum_{i=1}^{2^p} W_i^2\left[\left(\frac{1}{nW_i}+\frac{(N-n)(1-W_i)}{(N-1)n^2W_i^2}\right)-\frac{1}{N_i}\right]\underline{\alpha}'\,V_{i.rat}\,\underline{\alpha} \tag{4.2}$$

The second component provides

$$MSE[E(\bar{y}'_{pr.rat}\,|n_i) = MSE\left[\sum_{i=1}^{2^p} W_i\,\bar{Y}_i\left\{1+\left(\frac{1}{n_i}-\frac{1}{N_i}\right)z_{i.rat}\right\}\right],$$

where

$$z_{i.rat} = \sum_{j=1}^{p}\alpha_{ij}C_{ij}^2\,(1-K_{ij})$$

$$= \sum_{i=1}^{2^p}\sum_{j=1}^{2^p} A_{i.rat}\,A_{j.rat}\,n_{ij}\,,$$

where $A_{i.rat} = W_i\,\bar{Y}_i\,z_{i.rat}$

$$n_{ij} = MSE\left(\frac{1}{n_i}\right) \qquad \forall\ i=j \text{ by using (3.6)}$$

$$n_{ij} = Cov\left(\frac{1}{n_i},\frac{1}{n_j}\right) \qquad \forall\ i\neq j \text{ by using (3.7)}$$

$$= \underline{A}'_{rat}\,N\,\underline{A}_{rat}\,, \tag{4.3}$$

where, $\underline{A}'_{rat} = (A_{1rat}, A_{2rat},\ldots,A_{2^p rat})$ and $N = (n_{ij})_{2^p\times 2^p}$

By adding the first and second components, we get the required result.

Case 2: When each X_j, $j=1,2,\ldots,p$ is positively correlated with Y in each stratum, we can also use weighted regression estimator given as

$$\bar{y}'_{pr.reg} = \sum_{i=1}^{2^p} W_i \sum_{j=1}^{p} \alpha_{ij} g_{ij.reg}(y_i, x_{ij}), \tag{4.4}$$

where $g_{ij.reg}(y_i, x_{ij}) = \bar{y}_i + b_{ij}(\bar{X}_{ij} - \bar{x}_{ij})$ and b_{ij}'s are pre-assigned constants.

To the first-degree approximation, the bias and mean square error are respectively given by

$$E(\bar{y}'_{pr.reg}) = E[E(\bar{y}'_{pr.reg} \mid n_i)]$$

$$= E(\bar{Y}) = \bar{Y}. \tag{4.5}$$

$\because$ Bias $= 0$

$\therefore$ Variance $= MSE$.

The conditional MSE is a sum of two components.

$$MSE(\bar{y}'_{pr.reg}) = E[MSE(\bar{y}'_{pr.reg} \mid n_i)] + MSE[E(\bar{y}'_{pr.reg} \mid n_i)]$$

Consider the first component

$$E[MSE(\bar{y}'_{pr.reg} \mid n_i)] = E\left[\sum_{i=1}^{2^p} W_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \sum_{j=1}^{p} \sum_{h=1}^{p} \alpha_{ij}\, \alpha_{ih}\, v_{ijh.reg}\right]$$

$$= \sum_{i=1}^{2^p} W_i^2 \left(\left(\frac{1}{nW_i} + \frac{(N-n)(1-W_i)}{(N-1)n^2 W_i^2}\right) - \frac{1}{N_i}\right) \underline{\alpha}'\, V_{i.reg}\, \underline{\alpha}, \tag{4.6}$$

where

$$v_{ijh.reg} = [S_i^2 + b_{ij}\, b_{ih}\, \rho_{ijh}\, S_{ij}\, S_{ih} - b_{ij}\, \rho_{ij}\, S_i\, S_{ij} - b_{ih}\, \rho_{ih}\, S_i\, S_{ih}]$$

The second component provides

$$MSE[E(\bar{y}'_{pr.reg} \mid n_i)] = MSE[\bar{Y}] = 0 \tag{4.7}$$

By the first component, we get the required result.

Case 3: If each X_j, $j = 1, 2, \ldots, p$ is negatively correlated with Y in each stratum, then our estimator will convert into weighted product estimator given as

$$\bar{y}'_{pr.prod} = \sum_{i=1}^{2^p} W_i \sum_{j=1}^{p} \alpha_{ij} g_{ij.prod}(y_i, x_{ij}), \tag{4.8}$$

where

$$g_{ij.prod}(y_i, x_{ij}) = \frac{\bar{y}_i}{\bar{X}_{ij}} \bar{x}_{ij}.$$

To the first-degree approximation, the bias and MSE are respectively given by

$$E(\bar{y}'_{pr.prod}) = E[E(\bar{y}'_{pr.prod} \mid n_i)]$$

$$= \bar{Y} + \sum_{i=1}^{2^p} W_i \bar{Y}_i \left(\left(\frac{1}{nW_i} + \frac{(N-n)(1-W_i)}{(N-1)n^2 W_i^2} \right) - \frac{1}{N_i} \right) \sum_{j=1}^{p} \alpha_{ij} C_{ij}^2 K_{ij} \tag{4.9}$$

The conditional MSE is a sum of two components.

$$MSE(\bar{y}'_{pr.prod}) = E[MSE(\bar{y}'_{pr.prod} \mid n_i)] + MSE[E(\bar{y}'_{pr.prod} \mid n_i)]$$

Consider the first component

$$E[MSE(\bar{y}'_{pr.prod} \mid n_i)] = E\left[\sum_{i=1}^{2^p} W_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) \sum_{j=1}^{p} \sum_{h=1}^{p} \alpha_{ij} \alpha_{ih} v_{ijh.prod} \right]$$

$$= \sum_{i=1}^{2^p} W_i^2 \left(\left(\frac{1}{nW_i} + \frac{(N-n)(1-W_i)}{(N-1)n^2 W_i^2} \right) - \frac{1}{N_i} \right) \underline{\alpha}' V_{i.prod} \underline{\alpha} , \tag{4.10}$$

where

$$v_{ijh.prod} = \bar{Y}_i^2 [C_i^2 + \rho_{ijh} C_{ij} C_{ih} + \rho_{ij} C_i C_{ij} + \rho_{ih} C_i C_{ih}].$$

The second component provides

$$MSE[E(\bar{y}'_{pr.rat} \mid n_i) = MSE\left[\sum_{i=1}^{2^p} W_i \bar{Y}_i \left\{ 1 + \left(\frac{1}{n_i} - \frac{1}{N_i} \right) z_{i.prod} \right\} \right],$$

where

$$z_{i.prod} = \sum_{j=1}^{p} \alpha_{ij} C_{ij}^2 K_{ij}$$

$$= \underline{A}'_{prod} N \underline{A}_{prod} \tag{4.11}$$

By adding the first and second components, we get the required result.

Case 4: If some of the ρ_{ij}'s are positive (say, for $j = 1, 2, \ldots, p'$) and some ρ_{ij}'s are negative (say, for $j = p'+1, \ldots, p$) $\forall i$ then our estimator will be weighted ratio cum product type or regression cum product type, given by

$$\bar{y}'_{pr.r\,cum\,p} = \sum_{i=1}^{2^p} W_i \left[\sum_{j=1}^{p'} \alpha_{ij} g_{ij.rat\,or\,reg}(y_i, x_{ij}) + \sum_{j=p'+1}^{p} \alpha_{ij} g_{ij.prod}(y_i, x_{ij}) \right] \tag{4.12}$$

5. COST ANALYSIS

The total cost over 2^p strata is given as

$$T_c = C_0 + \sum_{i=1}^{2^p} C_{pi}\, n_i \ , \tag{5.1}$$

where C_0 is the overhead cost and C_{pi} is the cost of collecting, editing and processing per n_i unit in the $i-$th strata. It varies from stratum to stratum and it depends on the auxiliary variables in the stratum. The expected cost

$$E(T_c) = \left(\frac{n}{N}\right)\left\{C_0 + \sum_{i=1}^{2^p} C_{pi} N_i\right\} \tag{5.2}$$

To get optimum n, define a function δ with Lagrange multiplier λ and pre-fixed level of variance V_0

$$\delta = E(T_c) + \lambda[MSE(\bar{y}'_{pr}) - V_0]$$

$$= \left(\frac{n}{N}\right)\left[C_0 + \sum_{i=1}^{2^p} C_{pi}\, N_i\right] + \lambda\left[n + \frac{A}{n} + \frac{B}{n^2} + \frac{C}{n^3} - D\right], \tag{5.3}$$

where

$$A = \frac{1}{N-1}\left[\sum_{i=1}^{2^p}(1-W_i^2)\underline{\alpha}'_i V_i \underline{\alpha}_i - \sum_{i=1}^{2^p}\sum_{\substack{j=1 \\ i\neq j}}^{2^p} \frac{A_i A_j}{W_i W_j}\right]$$

$$B = \frac{1}{N-1}\left[N\sum_{i=1}^{2^p}(1-W_i^2)\underline{\alpha}'_i V_i \underline{\alpha}_i - \sum_{i=1}^{2^p} A_i^2 \frac{(1-W_i)}{W_i^3} - N\sum_{i=1}^{2^p}\sum_{\substack{j=1 \\ i\neq j}}^{2^p} \frac{A_i A_j}{W_i W_j}\right]$$

$$C = \frac{N}{N-1}\sum_{i=1}^{2^p} A_i^2 \frac{(1-W_i)}{W_i^3}$$

$$D = \frac{1}{N}\sum_{i=1}^{2^p} W_i\, \underline{\alpha}'_i V_i \underline{\alpha}_i + V_0 .$$

On differentiating (5.3) with respect to λ and n and equating to zero, we get two equations.

Differentiating with respect to λ, we get

$$n+\frac{A}{n}+\frac{B}{n^2}+\frac{C}{n^3}=D \tag{5.4}$$

Secondly, differentiating with respect to n, we get

$$\left(\frac{1}{N}\right)\left[C_0+\sum_{i=1}^{2^p} C_{pi}\, N_i\right]+\lambda\left[1-\frac{A}{n^2}-\frac{2B}{n^3}-\frac{3C}{n^4}\right]=0$$

$$\lambda=\frac{\left(C_0+\sum_{i=1}^{2^p} C_{pi}\, N_i\right)}{N\left(\frac{A}{n^2}+\frac{2B}{n^3}+\frac{3C}{n^4}-1\right)}\ . \tag{5.5}$$

Equations (5.4) and (5.5) give the values of n and λ respectively.

Equation (5.4) can be solved by adopting standard Iterative procedures (Newton-Raphson method) given by

$$n_{k+1}=n_k-\frac{f\,(n_k)}{f'(n_k)},$$

where

$$f\,(n_k)=n+\frac{A}{n}+\frac{B}{n^2}+\frac{C}{n^3}-D$$

and $f'(n_k)$ is the derivative of $f\,(n_k)$, we begin with a first guess n_0.

After substituting the value of n in equation (5.5), we will get the value of λ.

6. AN EMPIRICAL STUDY

The above theoretical developments are applied on three artificially constructed populations. In data sets first and second, the population consisting of $N=70$ observations. For example yield rates of the crop can be taken as a study variable Y and no of shoots / canes, average height of shoots / cane and average width of 3rd leaf three auxiliary variables can be considered as auxiliary variables X_1, X_2 and X_3 respectively. Suppose a sample of size $n=25$ is drawn by $SRSWOR$ from these populations. A sample of $n=43$ units has been drawn by $SRSWOR$ from the third data set of size $N=90$. All the three populations are given in the appendix. In first and third data sets, the information on all the three auxiliary variables were missed for different population unit and in second data set, information was missed on two auxiliary variables X_2 and X_3. In data sets first and third, there will be 8 strata and in data set second, there will be only 4 strata. The computed sample size for each stratum by proportional allocation and the observed statistics about the population in case of incomplete auxiliary information are given in the following tables.

Data Set I:

Table 1: The computed sample size for each stratum by proportional allocation

Strata	I	II	III	IV	V	VI	VII	VIII	Total
N_i	4	9	13	6	11	6	4	17	70
n_i	1	3	5	2	4	2	1	6	25

Table 2: Population parameters

Without Stratification	$\bar{Y}=33.95903$ $S_y^2=4.543188$		
Stratum			
I	$\bar{Y}_1=30.76000$ $S_1^2=8.12500$		
II	$\bar{Y}_2=36.27778$ $S_2^2=124.50694$ $S_{21}^2=399.50000$	$\bar{X}_{21}=63.66667$ $\rho_{21}=0.79522$	$K_{21}=0.77911$ $b_{21}=0.44394$
III	$\bar{Y}_3=37.63846$ $S_3^2=206.57090$ $S_{32}^2=0.06008$	$\bar{X}_{32}=1.13538$ $\rho_{32}=0.70908$	$K_{32}=1.25426$ $b_{32}=41.57907$
IV	$\bar{Y}_4=43.25000$ $S_4^2=181.87500$ $S_{43}^2=0.24272$	$\bar{X}_{43}=3.70000$ $\rho_{43}=0.96175$	$K_{43}=2.25222$ $b_{43}=26.32663$
V	$\bar{Y}_5=26.49382$ $S_5^2=88.52406$ $S_{51}^2=602.00000$ $S_{52}^2=0.12349$ $b_{52}=16.24694$	$\bar{X}_{51}=55.00000$ $\rho_{51}=0.85455$ $\rho_{52}=0.60681$ $K_{52}=0.68348$	$\bar{X}_{52}=1.11455$ $\rho_{512}=0.65101$ $K_{51}=0.68028$ $b_{51}=0.32770$

VI	$\bar{Y}_6 = 45.83333$ $S_6^2 = 32.96667$ $S_{61}^2 = 236.96667$ $S_{63}^2 = 0.05499$ $b_{63} = 20.38070$	$\bar{X}_{61} = 87.83333$ $\rho_{61} = 0.79500$ $\rho_{63} = 0.83236$ $K_{63} = 1.66899$	$\bar{X}_{63} = 3.75333$ $\rho_{613} = 0.63680$ $K_{61} = 0.56825$ $b_{61} = 0.29653$
VII	$\bar{Y}_7 = 19.50000$ $S_7^2 = 12.16667$ $S_{72}^2 = 0.03497$ $S_{73}^2 = 0.21823$ $b_{73} = 3.46724$	$\bar{X}_{72} = 0.91500$ $\rho_{72} = 0.69503$ $\rho_{73} = 0.46436$ $K_{73} = 0.54320$	$\bar{X}_{73} = 3.05500$ $\rho_{723} = 0.95205$ $K_{72} = 0.60834$ $b_{72} = 12.96473$
VIII	$\bar{Y}_8 = 33.99412$ $\bar{X}_{83} = 3.65176$ $\rho_{812} = 0.63381$ $\rho_{82} = 0.54901$ $S_{82}^2 = 0.04304$ $K_{83} = 0.87872$ $b_{82} = 40.87506$	$\bar{X}_{81} = 60.82353$ $S_8^2 = 238.57309$ $\rho_{813} = 0.14220$ $\rho_{832} = 0.52914$ $K_{81} = 0.89667$ $S_{83}^2 = 0.17490$ $b_{83} = 8.17996$	$\bar{X}_{82} = 1.18529$ $\rho_{81} = 0.82364$ $S_{81}^2 = 644.40441$ $\rho_{83} = 0.22148$ $K_{82} = 1.42522$ $b_{81} = 0.50115$

Table 3: The biases, mean square errors and the relative efficiencies of the estimators of the suggested class

Estimators	| Bias |	*MSE*	% Relative Efficiency
$\bar{y}$	-	4.54319	100.00
$\bar{y}'_{pr.rat}$	1.08896	1.76330	257.65
$\bar{y}'_{pr.reg}$	-	1.52358	298.19

Data Set II:

Table 4: The computed sample size for each stratum by proportional allocation

Strata	I	II	III	IV	Total
N_i	13	24	12	21	70
n_i	5	9	4	8	25

Table 5: Population parameters

Without Stratification	$\bar{Y} = 33.95903$ $S_y^2 = 4.543188$		
Stratum			
I	$\bar{Y}_1 = 30.13846$ $S_1^2 = 99.93465$ $S_{11}^2 = 395.10256$	$\bar{X}_{11} = 57.46154$ $\rho_{11} = 0.90371$	$K_{11} = 0.86653$ $b_{11} = 0.45450$
II	$\bar{Y}_2 = 34.49300$ $S_2^2 = 231.99382$ $S_{21}^2 = 630.99819$ $S_{22}^2 = 0.07319$ $b_{22} = 31.35968$	$\bar{X}_{21} = 68.70833$ $\rho_{21} = 0.73409$ $\rho_{22} = 0.55701$ $K_{22} = 1.06751$	$\bar{X}_{22} = 1.17417$ $\rho_{212} = 0.53933$ $K_{21} = 0.88665$ $b_{21} = 0.44512$
III	$\bar{Y}_3 = 40.29167$ $S_3^2 = 157.06629$ $S_{31}^2 = 716.08333$ $S_{33}^2 = 716.08333$ $b_{33} = 15.06138$	$\bar{X}_{31} = 78.91667$ $\rho_{31} = 0.86981$ $\rho_{33} = 0.47167$ $K_{33} = 1.36814$	$\bar{X}_{33} = 3.66000$ $\rho_{313} = 0.16533$ $K_{31} = 0.79788$ $b_{31} = 0.40737$
IV	$\bar{Y}_4 = 32.09524$ $\bar{X}_{43} = 3.72810$ $\rho_{412} = 0.34947$ $\rho_{42} = 0.49505$ $S_{42}^2 = 0.04866$ $K_{43} = 1.34101$ $B_{42} = 28.28672$	$\bar{X}_{41} = 53.85714$ $S_4^2 = 158.86548$ $\rho_{413} = 0.49809$ $\rho_{432} = 0.45124$ $K_{41} = 1.21734$ $S_{43}^2 = 0.19578$ $b_{43} = 11.54477$	$\bar{X}_{42} = 1.05095$ $\rho_{41} = 0.87180$ $S_{41}^2 = 229.42857$ $\rho_{43} = 0.40528$ $K_{42} = 0.92624$ $b_{41} = 0.72545$

Table 6: The biases, mean square errors and the relative efficiencies of the estimators of the suggested class

Estimators	| Bias |	MSE	% Relative Efficiency
$\bar{y}$	-	4.54319	100.00
$\bar{y}_{rat}$	0.66430	1.74367	260.55
$\bar{y}_{reg}$	-	1.65869	273.90
$\bar{y}'_{pr.rat}$	0.43734	1.48505	305.93
$\bar{y}'_{pr.reg}$	-	1.50979	300.00

Data Set III:

Table 7: The computed sample size for each stratum by proportional allocation

Strata	I	II	III	IV	V	VI	VII	VIII	Total
N_i	16	13	7	10	12	8	9	15	90
n_i	8	6	3	5	6	4	4	7	43

Table 8: Population parameters for the numerically simulated population (given in Appendix)

Without Stratification	$\bar{Y} = 48.14444$ $S_y^2 = 228.01261$		
Stratum			
I	$\bar{Y}_1 = 45.81250$ $S_1^2 = 207.22917$		
II	$\bar{Y}_2 = 48.53846$ $S_2^2 = 284.26923$ $S_{21}^2 = 721.39744$	$\bar{X}_{21} = 51.69231$ $\rho_{21} = 0.85959$	$K_{21} = 0.57466$ $b_{21} = 0.53960$
III	$\bar{Y}_3 = 47.85714$ $S_3^2 = 285.14286$ $S_{32}^2 = 605.00000$	$\bar{X}_{32} = 53.00000$ $\rho_{32} = 0.83425$	$K_{32} = 0.63427$ $b_{32} = 0.57273$
IV	$\bar{Y}_4 = 50.30000$ $S_4^2 = 312.23333$ $S_{43}^2 = 558.48889$	$\bar{X}_{43} = 46.60000$ $\rho_{43} = 0.93639$	$K_{43} = 0.64864$ $b_{43} = 0.70014$

V	$\bar{Y}_5 = 50.16667$ $S_5^2 = 191.60606$ $S_{51}^2 = 619.72727$ $S_{52}^2 = 247.53788$ $b_{52} = 0.58032$	$\bar{X}_{51} = 56.50000$ $\rho_{51} = 0.85239$ $\rho_{52} = 0.65961$ $K_{52} = 0.53694$	$\bar{X}_{52} = 46.41667$ $\rho_{512} = 0.54232$ $K_{51} = 0.53380$ $b_{51} = 0.47396$
VI	$\bar{Y}_6 = 41.00000$ $S_6^2 = 208.0000$ $S_{61}^2 = 399.98214$ $S_{63}^2 = 496.98214$ $b_{63} = 0.59358$	$\bar{X}_{61} = 46.62500$ $\rho_{61} = 0.77808$ $\rho_{63} = 0.91763$ $K_{63} = 0.57730$	$\bar{X}_{63} = 39.87500$ $\rho_{613} = 0.71440$ $K_{61} = 0.63808$ $b_{61} = 0.56110$
VII	$\bar{Y}_7 = 47.88889$ $S_7^2 = 252.61111$ $S_{72}^2 = 647.94444$ $S_{73}^2 = 384.50000$ $b_{73} = 0.75033$	$\bar{X}_{72} = 50.77778$ $\rho_{72} = 0.86535$ $\rho_{73} = 0.92570$ $K_{73} = 0.68939$	$\bar{X}_{73} = 44.00000$ $\rho_{723} = 0.87401$ $K_{72} = 0.57291$ $b_{72} = 0.54032$
VIII	$\bar{Y}_8 = 51.33333$ $\bar{X}_{83} = 27.40000$ $\rho_{812} = 0.58511$ $\rho_{82} = 0.71264$ $S_{82}^2 = 361.80952$ $K_{83} = 0.53627$ $b_{82} = 0.54120$	$\bar{X}_{81} = 50.66667$ $S_8^2 = 208.66667$ $\rho_{813} = 0.56521$ $\rho_{832} = 0.60341$ $K_{81} = 0.51478$ $S_{83}^2 = 66.97143$ $b_{83} = 1.00469$	$\bar{X}_{82} = 51.33333$ $\rho_{81} = 0.89878$ $S_{81}^2 = 619.66667$ $\rho_{83} = 0.56918$ $K_{82} = 0.54120$ $b_{81} = 0.52156$

Table 9: The biases, mean square errors and the relative efficiencies of the estimators of the suggested class

Estimators	| Bias |	MSE	% Relative Efficiency
$\bar{y}$	-	2.76915	100.00
$\bar{y}'_{pr.rat}$	0.41774	1.51334	182.98
$\bar{y}'_{pr.reg}$	-	1.05536	262.39

7. DISCUSSION AND CONCLUSION

The suggested class of estimators using incomplete multi-auxiliary information under post stratified set up has been compared with simple mean per unit estimator in which auxiliary information has not been used and estimators using complete auxiliary information. It is to be noted that the members of the class dominate over mean per unit estimator in terms of relative efficiency for the three data set taken.

i) The results of the table 3, 6 and 9 indicate that, though the suggested ratio estimator of the suggested class is biased, the amount of bias is not significantly high.

ii) MSE of both the estimators of the proposed class is substantially lower than usual mean per unit estimators without using any auxiliary information for data set I and III. (See table 3, 6 and 9).

iii) It is clear from table 6; MSE of both the suggested estimators using incomplete information on two auxiliary variables is significantly lower than usual mean per unit estimators without using any auxiliary information and the traditional ratio and regression estimator having complete information on an auxiliary variable.

iv) The trend becomes more clear when we compare the % relative efficiency from table 3 and 9 i.e. the gain in % relative efficiency of the estimators of suggested class is substantially more higher as compared to usual per unit estimator (see table 6).

v) For data sets 2, the gain in % relative efficiency is considerably higher as compared to usual per unit estimator and usual ratio and regression estimator using an auxiliary variable.

Thus, we infer that an optimum use of incomplete multi-auxiliary information can be made by the procedure suggested, when frame in each stratum is not known. And we can construct different types of estimators for estimating population mean which are more efficient than mean per unit estimator in simple random sampling.

Appendix

Data Set I:

Stratum I										
Y	28.8	32.3	34.0	28.0						
Stratum II										
Y	50	29	26.5	46	25.5	38	45	46.5	20	
X_1	87	56	50	56	36	60	96	82	50	
Stratum III										
Y	39	26.5	13	66	53.5	58	28.5	24.7	36.3	36.4
	34.5	34.4	38.5							
X_2	1.01	1.02	0.61	1.35	1.55	1.29	1.21	1.04	0.87	1.12
	1.10	1.43	1.16							
Stratum IV										
Y	16.4	43.5	48	52	47.5	52				
X_3	2.78	3.78	3.64	4.14	3.76	4.1				
Stratum V										
Y	20	27	26	28.5	34.1	35.1	23.5	19	39.5	33
	5.7									
X_1	42	48	46	64	104	69	39	32	86	56
	19									
X_2	0.85	0.75	0.87	0.74	1.57	1.64	0.89	1.01	1.45	1.51
	0.98									
Stratum VI										
Y	35.5	48.5	43.5	48	52	47.5				
X_1	69	110	77	89	101	81				
X_3	3.42	3.78	3.78	3.64	4.14	3.76				
Stratum VII										
Y	18	24.5	19	16.5						
X_2	1.04	1.11	0.76	0.75						
X_3	3.48	3.42	2.54	2.78						

Stratum VIII										
Y	64.5	10	37.3	36.8	33.5	31	16.5	12.8	43.5	32.5
	25	62	32.5	51	38.5	32.5	18			
X_1	87	22	67	72	68	59	34	26	60	59
	44	76	56	129	72	67	36			
X_2	1.18	0.7	1.25	1.24	1.35	1.47	0.86	0.94	0.98	1.36
	1.22	1.36	1.13	1.33	1.37	1.27	1.14			
X_3	3.74	3.8	3.98	3.98	4.08	3.9	2.92	2.66	3.34	4.12
	3.8	3.78	3.46	3.12	3.72	3.98	3.7			

Data Set II:

Stratum I										
Y	33.5	31	16.5	12.8	25.5	38	45	46.5	20	28.75
	32.25	34	28							
X_1	68	59	34	26	36	60	96	82	50	46
	56	76	58							
Stratum II										
Y	64.5	10	37.3	36.8	34.091	35.091	23.5	19	39.5	33
	5.75	39	26.5	13	66	53.5	58	28.5	24.7	36.3
	36.4	34.5	34.4	38.5						
X_1	87	22	67	72	104	69	39	32	86	56
	19	86	56	46	109	102	60	96	59	98
	76	74	69	65						
X_2	1.18	0.7	1.25	1.24	1.57	1.64	0.89	1.01	1.45	1.51
	0.98	1.01	1.02	0.61	1.35	1.55	1.29	1.21	1.04	0.87
	1.12	1.1	1.43	1.16						
Stratum III										
Y	35.5	48.5	51	38.5	32.5	18	16.5	43.5	48	52
	47.5	52								
X_1	69	110	129	72	67	36	38	77	89	101
	81	78								
X_3	3.42	3.78	3.12	3.72	3.98	3.7	2.78	3.78	3.64	4.14
	3.76	4.1								

Stratum IV										
Y	20	27	26	28.5	50	29	26.5	46	43.5	18
	24.5	19	35	45	12	24	48	32.5	25	62
	32.5									
X_1	42	48	46	64	87	56	50	56	60	36
	40	37	69	72	23	46	64	59	44	76
	56									
X_2	0.85	0.75	0.87	0.74	0.94	1.03	0.92	0.96	0.98	1.04
	1.11	0.76	1.21	1.34	1.05	0.92	1.53	1.36	1.22	1.36
	1.13									
X_3	3.48	3.62	3.45	3.98	4.22	3.36	3.48	4.16	3.34	3.48
	3.42	2.54	4.6	4.12	4.06	3.78	4.04	4.12	3.8	3.78
	3.46									

Data Set III:

Stratum I										
Y	36	72	54	35	47	67	38	45	57	30
	26	44	49	67	27	39				
Stratum II										
Y	30	26	44	72	40	67	72	43	36	72
	54	35	40							
X_1	18	16	39	95	30	69	80	50	53	97
	28	48	49							
Stratum III										
Y	43	36	72	54	35	28	67			
X_2	25	55	97	53	42	30	69			
Stratum IV										
Y	47	67	38	30	26	44	72	40	67	72
X_3	29	77	28	20	19	54	65	30	69	75

Stratum V										
Y	72	43	36	72	54	35	40	34	48	52
	49	67								
X_1	75	19	37	97	53	26	49	44	53	56
	72	92								
X_2	80	41	46	47	27	31	33	27	55	55
	57	58								
Stratum VI										
Y	27	39	57	28	36	67	44	30		
X_1	18	53	60	34	65	69	54	20		
X_3	23	29	60	12	46	82	34	33		
Stratum VII										
Y	28	57	39	27	67	49	52	40	72	
X_2	25	60	28	15	55	72	56	49	97	
X_3	12	62	30	23	58	57	55	33	66	
Stratum VIII										
Y	47	67	38	29	57	44	72	40	63	69
	43	36	72	54	39					
X_1	48	51	24	19	42	54	95	38	69	75
	31	25	97	63	29					
X_2	47	79	53	31	29	34	75	31	79	75
	41	46	66	31	53					
X_3	29	29	16	18	18	20	34	18	31	39
	32	21	35	31	40					

REFERENCES

Agarwal, M.C. and Panda, K.B. (1993): An efficient estimator in post-stratification. *Metron*, **LI**, 179-188.

Agarwal, M.C. and Panda, K.B. (1995): On efficient estimation in post-stratification. *Metron*, **LIII**, 107-115.

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

Holt, D. and Smith, T.M.F. (1979): Post-stratification. *J. Roy. Statist. Soc. Ser. A*, **142**, 33-36.

Jager, P. (1986): Post-stratification against bias in sampling. *Internat. Statist. Rev.*, **54**, 159-167.

Jager, P., Oden, A. and Truslson, L. (1985): Post-stratification and ratio estimation. *Internat. Statist. Rev.*, **53**, 221-238.

Murthy, M.N. (1967): *Sampling Theory and Methods*. Calcutta Statistical Publishing Society, Calcutta.

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

Raj, D. (1965): On a method of using multi-auxiliary information in sample survey. *J. Amer. Statist. Assoc.*, **60**, 270-277.

Rao, P.S.R.S. and Mudholkar, G.S. (1967): Generalized multivariate estimator for the mean of finite population. *J. Amer. Statist. Assoc.*, **62**, 1009-1012.

Sahai, A. and Sahai, A. (1985): On efficient use of auxiliary information. *J. Statist. Plann. Inference*, **12**, 203 – 212.

Sukhatme P.V. and Sukhatme, B.V. (1997): *Sampling Theory of Surveys with Applications*. Iowa State University Press, Iowa.

Singh, R. (1977): A note on the use of incomplete multi-auxiliary information in sample surveys. *Aust. J. Statist.*, **19**, 105-107.

Singh, M. (1982): A note on the use of multiauxiliary information. *Comm. Statist. Theory Methods*, **11**, 933-939.

Smith, T.M.F. (1991): Post-stratification. *The Statistician*, **40**, 323.

Srivastava, S.K. (1971): A generalized estimator for the mean of a finite population using multi-auxiliary information. *J. Amer. Statist. Assoc.*, **66**, 404-407.

Tracy. D.S. and Singh, H.P. (1999): An improved class of estimators for finite population mean in sample surveys. *Biom. J.*, **41**, 891-895.

Received : 10-03-2008

Revised : 27-10-2008

Department of Statistics, Institute of Social Sciences,
Dr. B.R. Ambedkar University, Agra, India.
e-mail: msrivastava_iss@hotmail.com,
neha1garg@gmail.com