Estimation of Population Mean in Two Stage Design using Double Sampling for Stratification and Multiauxiliary Information

Monika Saini
Research Scholar Department of Statistics
M.D. University, Rohtak

Shashi Bahl
Prof. & Head Department of Statistics
M.D. University, Rohtak

ABSTRACT
For estimating the population mean of the heterogeneous population, when sampling frame in strata and the strata weight are not defined, we have proposed new difference and ratio estimators using double sampling for stratification and auxiliary information in two stage design. The expression for bias and variances of the proposed estimators have been obtained based on double sampling for stratification approach and comparison have been made with estimator under unstratified double sampling. In addition, we support these theoretical results by an empirical study.

General Terms
Sampling Theory

Keywords
Double Sampling for Stratification; Two Stage Design; Auxiliary Information; Difference (D) and Ratio (R) estimators.

1. INTRODUCTION
Information on variables correlated with the main variable under study is popularly known as auxiliary information which may be fruitfully utilized either at planning stage or at designing stage or at the estimation stage to arrive at improved estimator compared to those, not utilizing auxiliary information. If the required information is not available in advance for the entire population, it might be advantageous to collect certain items for large preliminary sample and then use this information at the estimation stage to improve the estimators. This technique is generally referred to as double sampling.
Stratification is one of the most widely used techniques in sample survey design serving the dual purpose of providing samples that are representative of major sub-groups of the population and improving the precision of estimators. Stratified sampling presupposes the knowledge of strata size as well as the availability of a frame for drawing a sample in each stratum. However application of this technique presupposes the knowledge of strata size and the availability of sampling frames within strata. But in many practical situations, when we want to estimate the population mean $\bar{\boldsymbol{Y}}$ of a variate y and consider it desirable to stratify the population, consisting of N units on the basis of the value of auxiliary character x but the information of x is unknown, the sampling frame for various strata and the strata weight $\boldsymbol{W_h}=(\boldsymbol{N_h}|\boldsymbol{N})$ (h=1,2,……L) are not known although the strata may be fixed in advance. In such situation the method of double sampling for stratification (DSS) can be used. Many authors including, Rao [3], Tripathi [7], Sukhatme et. al. [6] Ige and Tripathi [1], Tripathi and Bahl [8], Singh and Vishwakarma [5] and Shabbir and Gupta [4] contributed well in the field of stratification.
In this paper we propose difference and ratio estimators in two stage design using DSS and multi-auxiliary information.

2. SUGGESTED PROCEDURE
Let a finite (survey) population U be partitioned into N fsu (first stage units) denoted by ($U_1,..,U_i,..,U_N$) such that the number of ssu (second stage units) in U_i be M_i . Suppose we want to estimate the population mean $\bar{Y}$ of study variate y and consider it desirable to stratify the population consisting of N units on the basis of one of values of the auxiliary characters $x_k^{'s}$ (k= 1, 2,….,p) but the frequency distribution of $x_k^{'s}$ is unknown. The sampling frame for various strata and the strata weight W_h =($N_h|N$), (h=1, 2 …L), are not known although the strata may be fixed in advance. We use the technique of double sampling for stratification (DSS) in two stage design at fsu level; the sampling procedure consists of the following steps:

Step (1): We select a preliminary large sample of n' fsus rather inexpensively from a population of N fsus with simple random sampling without replacement and observe the auxiliary character $x_1, x_2, \dots\dots, x_p$ and then classify sample n' into L strata on the basis of the information obtained through n' for one or more $x_k^{'s}$. Let n_h' denote the number of units falling into stratum h (h=1,2,….,L) i.e. n_1', n_2'……..n_L' respectively, where $\sum_h^L n_h'=n'$.

Step (2): Randomly select n_h i.e no. of fsu`s in the sample, in stratum h out of n_h' without replacement ,where $n_h= v_h n_h'$, 0< $v_h < 1$ (h=1,2…..L)v_h being predetermined for each h are selected from strata independently and n=$\sum n_h$; n=(n_1,…..n_L).

Step (3): Randomly select m_{hi} i.e no. of ssu`s to be selected from each sampled first stage units out of M_{hi} in stratum h without replacement (h=1,2,....,L) and character of main interest y is observed.

Using above sampling scheme we propose difference and ratio estimators for the estimation of population mean.

3. PROPOSED ESTIMATORS AND THEIR PROPERTIES

Utilizing the information collected on $x_1, x_2, \dots\dots, x_p$, through the preliminary large sample n' on x_k-variates at fsu. We define multivariate difference and ratio estimators in two stage design using DSS at fsu level for the estimation of population mean and let us denote that estimator by T_{DSS} for difference and ratio estimators i.e. $(T_{DSS})_D$ and $(T_{DSS})_R$ as

$$T_{DSS} = \sum a_k t_k \quad \dots(3.1)$$

as $t_k = \begin{cases} (t_k)_D = \bar{y}_{(3)} - \lambda_k\,(\bar{x}_{k(2)} - \bar{x}_{k(1)}) & \dots(3.2) \\ (t_k)_R = (\frac{\bar{y}_{(3)}}{\bar{x}_{k(2)}})\,\bar{x}_{k(1)} & \dots(3.3) \end{cases}$

where $a = a_1, a_2, \dots\dots, a_p$,with $\sum_k^p a_k = 1$ is a weighting function and λ_k' s are suitably chosen constants.

In proposed estimator we note that

$$\bar{y}_{(3)} = \sum_h^L w_h' \frac{1}{n_h} \sum_i^{n_h} M_{hi}\,\bar{y}_{hi}$$

$$\bar{x}_{k(2)} = \sum_h^L w_h'\,\bar{x}_{kh}$$

$$\bar{x}_{k(1)} = \sum_h^L w_h'\,\bar{x}_{kh}'$$

such that $\bar{y}_{hi} = \frac{1}{m_{hi}} \sum_j^{m_{hi}} y_{hij}$

$$\bar{x}_{kh} = \frac{1}{n_h} \sum_i^{n_h} x_{khi}$$

$$\bar{x}_{kh}' = \frac{1}{n_h'} \sum_i^{n_h'} x_{khi}$$

where $\bar{y}_{(3)}$, $\bar{x}_{k(2)}$ and $\bar{x}_{k(1)}$ are unbiased estimator of population mean $\bar{Y}$ and $\bar{X}_k$. Note that $w_h' = \frac{n_h'}{n'}$ is an unbiased estimator of strata weights $W_h = \frac{N_h}{N}$. Throughout we assume that n' is large enough so that Pr (n_h'= 0) = 0 for all h. These estimators can shown to be unbiased by taking conditional expectation as

$$E(\bar{y}_{(3)}) = E_1 E_2 E_3 (\bar{y}_{(3)}) = \bar{Y}$$

$$E(\bar{x}_{k(2)}) = E_1 E_2\,(\bar{x}_{k(2)}) = \bar{X}_k$$

$$E(\bar{x}_{k(1)}) = E_1 (\bar{x}_{k(1)}) = \bar{X}_k$$

i.e. $\bar{x}_{k(1)} = E_2\,(\bar{x}_{k(2)})$

Theorem (3.1) : The multivariate difference estimator T_{DSS} is unbiased for $\bar{Y}$ and the exact expression for its variance is given by

$$V(T_{DSS})_D = (\frac{1}{n'} - \frac{1}{N})S_y^2 + \frac{1}{n'} \sum_h W_h (\frac{1}{v_h} - 1) a' B_h a + \frac{1}{N} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 (\frac{1}{m_{hi}} - \frac{1}{M_{hi}}) S_{hyi}^2$$

with $B_h = (b_{hkl})$ h=1, 2,…….,L ; k ,l=1,2,……,p

$$(b_{hkl}) = (S^2{}_{hy} - \lambda_k S_{hyk} - \lambda_l S_{hyl} - \lambda_k \lambda_l S_{hkl})$$

Further for large samples, the approximate expression for the bias and mean square error of $(T_{DSS})_R$ is given by

$$B(T_{DSS})_R = \frac{1}{n'} \sum_k \frac{a_k}{\bar{X}_k} \sum_h W_h\,(\frac{1}{v_h} - 1)\,(R_k\,S_{hk}^2 - S_{hyk})$$

$$MSE(T_{DSS})_R = V(T_{DSS})_D \quad \lambda_k = R_k$$

$$\lambda_l = R_l \qquad (\text{k, l} = 1,2,\dots\dots,\text{p})$$

Proof: Since $T_{DSS} = \sum a_k t_k$

$$(t_k)_D = \bar{y}_{(3)} - \lambda_k\,(\bar{x}_{k(2)} - \bar{x}_{k(1)})$$

$$E(T_{DSS})_D = E_1 E_2 E_3 (T_{DSS})_D$$

$$E_2 E_3 (t_k) = \bar{y}'$$

$$E(T_{DSS})_D = E_1 E_2 E_3 (T_{DSS})_D = \bar{Y}$$

The multivariate difference estimator T_{DSS} for $\bar{Y}$ is unbiased.

The exact expression for its variance is given as

$$V(T_{DSS})_D = \sum_k \sum_l a_k a_l \,\text{Cov}\,(t_k, t_l)$$

$$\begin{aligned} \text{Cov}(t_k t_l) = {} & V(\bar{y}_{(3)}) - \lambda_l \text{Cov}(\bar{y}_{(3)}, \bar{x}_{l(2)}) \\ & + \lambda_k \text{Cov}(\bar{y}_{(3)}, \bar{x}_{k(1)}) - \lambda_k \text{Cov}(\bar{y}_{(3)}, \bar{x}_{k(2)}) \\ & + \lambda_l \text{Cov}(\bar{y}_{(3)}, \bar{x}_{l(1)}) + \lambda_k \lambda_l [\,\text{Cov}\,(\bar{x}_{k(2)}, \bar{x}_{l(2)}) \\ & - \text{Cov}\,(\bar{x}_{k(2)}, \bar{x}_{l(1)}) - \text{Cov}\,(\bar{x}_{k(1)}, \bar{x}_{l(2)}) + \text{Cov}(\bar{x}_{k(1)}, \bar{x}_{l(1)}) \end{aligned}$$

$$V(T_{DSS})_D = (\frac{1}{n'} - \frac{1}{N})S_y^2 + \frac{1}{n'} \sum_h W_h (\frac{1}{v_h} - 1) a' B_h a + \frac{1}{N} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 (\frac{1}{m_{hi}} - \frac{1}{M_{hi}}) S_{ihy}^2 \quad \dots(3.4)$$

with $B_h = (b_{hkl})$ h=1, 2….,L ; k ,l=1,2,……,p

$$(b_{hkl}) = (S^2{}_{hy} - \lambda_k S_{hyk} - \lambda_l S_{hyl} - \lambda_k \lambda_l S_{hkl})$$

where $S_{hkl} = \frac{1}{N_{h-1}} \sum_i^{N_h} (x_{hki} - \bar{X}_{hk})(x_{hli} - \bar{X}_{hl})$

$$S^2{}_{hyi} = \frac{1}{M_{hi-1}} \sum_j^{M_{ih}} (y_{hij} - \bar{Y}_{ih})^2$$

To obtain the bias and mean square error of multivariate ratio estimator $(T_{DSS})_R$ we use the delta notation

$$e_o = \frac{\bar{y}_{(3)} - \bar{Y}}{\bar{Y}} \qquad e_1 = \frac{\bar{x}_{k(2)} - \bar{X}_k}{\bar{X}_k} \qquad e_2 = \frac{\bar{x}_{k(1)} - \bar{X}_k}{\bar{X}_k}$$

$$\bar{y}_{(3)} = \bar{Y}\,[1+e_o] \qquad \bar{x}_{k(2)} = \bar{X}_k [1+e_1] \qquad \bar{x}_{k(1)} = \bar{X}_k [1+e_2]$$

now equation (3.3) becomes

$$(T_{DSS})_R = \sum a_k\,\bar{Y}\,[1+e_o]\,[1+e_1]^{-1}\,[1+e_2]$$

Further for large samples, neglecting terms of higher powers, the approximate expression for its bias and mean square error is given by

$B(T_{DSS})_R = \frac{1}{n'} \sum_k \frac{a_k}{\bar{X}_k} \sum_h W_h \left(\frac{1}{v_h} - 1\right) (R_k S_{hk}^2 - S_{hyk})$

$M(T_{DSS})_R = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \frac{1}{n'} \sum_h W_h \left(\frac{1}{v_h} - 1\right) a' B_h a$

$+ \frac{1}{N} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{ihy}^2$ .... (3.5)

with $B_h = (b_{hkl})$ h=1, 2 …, L; k ,l =1,2,…….p

$(b_{hkl}) = (S^2_{hy} - R_k S_{hyk} - R_l S_{hyl} - R_k R_l S_{hkl})$

where $S_{hkl} = \frac{1}{N_h - 1} \sum_i^{N_h} (x_{hki} - \bar{X}_{hk})(x_{hli} - \bar{X}_{hl})$

$S^2_{hyi} = \frac{1}{M_{hi} - 1} \sum_j^{M_{ih}} (y_{hij} - \bar{Y}_{ih})^2$

Remark: From equation (3.4) and (3.5) is can easily be proved that for large sample

$M(T_{DSS})_R = V(T_{DSS})_D$ for $\lambda_k = R_k$, $\lambda_l = R_l$

4. OPTIMUM ESTIMATORS

Let $\beta_{ok} = \sum_h c_{kh} \beta_{okh} / \sum_h c_{kh}$ with $c_{kh} = W_h \left(\frac{1}{v_h} - 1\right) S^2_{hk}$ be the weighted average of the strata population regression coefficients $\beta_{okh} = S_{hyk} / S^2_{hk}$ of y on x_k and

$$\rho_{kl} = \frac{\sum_h W_h \left(\frac{1}{v_h} - 1\right) \rho_{hkl} S_{hk} S_{hl}}{\left[\sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hk}^2 \ \sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hl}^2\right]^{1/2}}$$

where $\rho_{hkl} = S_{hkl} / S_{hk} S_{hl}$ is the correlation coefficient between x_k and x_l in stratum h.

For p=1, when information on only x_k, is used, following Ige and Tripathi [1] the optimum value of λ_k is given by

$$\lambda_{ok} = \frac{\sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hyk}}{\sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hk}^2} = \beta_{ok}$$

When the choices for each $\lambda_k = \beta_{ok}$ are made for each k, the resulting variance is given by

$V(T_{DSS})_{\lambda_k = \beta_{ok}} = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2$

$+ \frac{1}{n'} \sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hy}^2 (1 - \rho_{yk}^2 - \rho_{yl}^2 + \rho_{yk} \rho_{yl} \rho_{kl})$

$+ \frac{1}{N} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{yhi}^2$ .... (4.1)

Further, when optimum weight vector

$a_o = \frac{B^{-1} g}{g' B^{-1} g}$ $g = (1, 1, \ldots, 1)'$

is used , we obtain

$V(T_{DSS})_{\substack{\lambda_k = \beta_{ok} \\ a = a_o}} = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2$

$+ \frac{1}{n'} (g' B^{-1} g)^{-1} \sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hy}^2$

$+ \frac{1}{n'} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{yhi}^2$

In practice, when exact value of $\lambda_{ok} = \beta_{ok}$ is not available, it may be estimated through the sample data

$$\hat{\beta}_{ok} = \frac{\sum_h w'_h \left(\frac{1}{v_h} - 1\right) s_{hyk}}{\sum_h w'_h \left(\frac{1}{v_h} - 1\right) s_{hk}^2}$$

where $s_{hyk} = \frac{1}{n_h - 1} \sum_i^{n_h} (y_{ih} - \bar{y}_h)(x_{hki} - \bar{x}_{kh})$

$s_{hk}^2 = s_{hkk}$

are unbiased estimator for S_{hyk} and S_{hk}^2 respectively.

Using the estimated optimum values, we may define a multiple regression estimator for $\bar{Y}$ in DSS by

$$(T_{DSS})_{reg}^1 = \bar{y}_{(3)} - \sum_k a_k \hat{\beta}_{ok} (\bar{x}_{k(2)} - \bar{x}_{k(1)})$$

for large samples, $M(T_{DSS})_{reg}$ would again be given by (4.1).

One may in fact obtain simultaneous optimum value of $T_k = a_k \lambda_k$ (k = 1,2, ..., p) as follows.

Let $S^* = (S_{kl}^*)$ $Q = (Q_1, Q_2, \ldots, Q_p)'$

where $S_{kl}^* = \sum_h W_h \left(\frac{1}{v_h} - 1\right) S_{hkl}$

$Q_k = S_{yk}^*$ k,l=1,2,…,p

then $V(T_{DSS}) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \frac{1}{n'} (S_y^{*2} - 2T'Q + T'S^*T)$

$+ \frac{1}{N} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{yhi}^2$

which gives $T_{opt} = T_o = S^{*-1} Q$

$V(T_{DSS})_{opt} = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \frac{1}{n'} S_y^{*2} (1 - R^2)$

$+ \frac{1}{N} \sum_h \frac{W_h}{n_h} \sum_i M_{hi}^2 \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{yhi}^2$ ….(4.2)

where $R^2 = \frac{Q' S^{*-1} Q}{S_y^{*2}}$ and R being the multiple is the correlation coefficient between $\bar{y}_{(3)}$ and $(\bar{x}_{k(2)} - \bar{x}_{k(1)})$.

The optimum value of T may be estimated by

$\hat{T} = \hat{S}^{*-1} \hat{Q}$ $\hat{S}^* = (\hat{S}_{kl}^*)$ $\hat{Q} = (\hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}_p)'$

where $\hat{S}_{kl}^* = \sum_h w'_h \left(\frac{1}{v_h} - 1\right) s_{hkl}$ $\hat{Q}_k = \hat{S}_{yk}^*$

Using these estimated values, we may define a multiple regression estimator based on DSS for $\bar{Y}$ as

$$(T_{DSS})_{reg} = \bar{y}_{(3)} - \sum_k \hat{T}_k^* (\bar{x}_{k(2)} - \bar{x}_{k(1)})$$

whose variance for large samples is given by (4.2).

5. COMPARISION

In this section we compare the proposed multivariate difference and ratio estimators based on DSS with Prabha [2] **i**n Unstratified Double Sampling (USDS) where second sample is a simple random at size n from the first random sample n'instead of being selected in the form of stratified sub-samples.

The multivariate difference and ratio estimator for the difference and ratio estimator for the population mean in USDS for the population mean in USDS are defined by:

$$T = \sum_k^p a_k d_k$$

where $\quad (d_k)_D = \bar{\bar{y}}_w - \lambda_k(\bar{x}_w - \bar{x}'_w)$

$$(d_k)_R = \frac{\bar{\bar{y}}_w}{\bar{x}_w}\bar{x}'_w \qquad \text{k=1,2,……,p}$$

where the weights a_k (k=1,2,….p) add ud to unity and λ_k are suitably chosen constants with

$$V(T)_D = (\frac{1}{n'} - \frac{1}{N})S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)a'Ba$$

$$+\frac{1}{nN}\sum_i^N M_i^2(\frac{1}{m_i} - \frac{1}{M_i})S_{iy}^2 \qquad \ldots(5.1)$$

with B=(b_{kl})

$$(b_{kl}) = (S_y^2 - \lambda_k S_{yk} - \lambda_l S_{yl} + \lambda_k\lambda_l S_{kl})$$

$$M(T)_R = (\frac{1}{n'} - \frac{1}{N})S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)a'Ba$$

$$+\frac{1}{nN}\sum_i^N M_i^2(\frac{1}{m_i} - \frac{1}{M_i})S_{iy}^2 \qquad \ldots(5.2)$$

with B= (b_{kl})

$$(b_{kl}) = (S_y^2 - R_k S_{yk} - R_l S_{yl} + R_k R_l S_{kl})$$

$$M(T)_R = V(T)_D$$

We note that both of $(T_{DSS})_D$ and $(\text{T})_D$ are unbiased and both have exact expression for their variances for all sample sizes while $(T_{DSS})_R$ and $(\text{T})_R$ are biased and the expression for their MSE is approximated (for large samples). Further, the information on $x_1, x_2, \ldots, x_p$ is basic for all the above estimators, it is being used both for stratification as well as for constructing estimators in case of DSS while only for defining estimators in case of USDS.

Using

$$S_{kl} = \sum_h W_h S_{hkl} + \sum_h W_h(\bar{x}_{kh} - \bar{X}_k)(\bar{x}_{lh} - \bar{X}_l)$$

$$S_{yk} = \sum_h W_h S_{hyk} + \sum_h W_h(\bar{Y}_h - \bar{Y})(\bar{x}_{kh} - \bar{X}_k)$$

we get $\quad b_{kl} = (S_y^2 - \lambda_k S_{yk} - \lambda_l S_{yl} + \lambda_k\lambda_l S_{kl})$

$$= \sum_h W_h(S_{hy}^2 - \lambda_k S_{hyk} - \lambda_l S_{hyl} + \lambda_k\lambda_l S_{hkl})$$

$$+\sum_h W_h(\bar{Y}_h - \bar{Y})^2 - \lambda_k(\bar{Y}_h - \bar{Y})(\bar{x}_{kh} - \bar{X}_k)$$

$$-\lambda_l(\bar{Y}_h - \bar{Y})(\bar{x}_{lh} - \bar{X}_l) + \lambda_k\lambda_l(\bar{x}_{kh} - \bar{X}_k)(\bar{x}_{lh} - \bar{X}_l)$$

For comparisons assuming

$$\frac{n_h}{n'_h} = \frac{n}{n'} = \nu_h$$

$$V(T_{DSS}) = (\frac{1}{n'} - \frac{1}{N})S_y^2 +$$

$$(\frac{1}{n} - \frac{1}{n'})\sum\sum a_k a_l \sum_h W_h(S_{hy}^2 - \lambda_k S_{hyk} - \lambda_l S_{hyl} + \lambda_k\lambda_l S_{hkl})$$
$$+\frac{1}{n'}\sum_h \frac{W_h}{n_h}\sum_i M_{hi}^2(\frac{1}{m_{hi}} - \frac{1}{M_{hi}})S_{hyi}^2$$

Then $\quad V(T)_D - V(T_{DSS})_D = (\frac{1}{n} - \frac{1}{n'})\sum\sum a_k a_l \sum_h W_h\, d_{hkl}^{(1)}$

where $d_{hkl}^{(1)} = (\bar{Y}_h - \bar{Y})^2 - \lambda_k(\bar{Y}_h - \bar{Y})(\bar{x}_{kh} - \bar{X}_k)$

$$-\lambda_l(\bar{Y}_h - \bar{Y})(\bar{x}_{lh} - \bar{X}_l) + \lambda_k\lambda_l(\bar{x}_{kh} - \bar{X}_k)(\bar{x}_{lh} - \bar{X}_l)$$

$$= ((\bar{Y}_h - \bar{Y}) - \lambda_k(\bar{x}_{kh} - \bar{X}_k))((\bar{Y}_h - \bar{Y}) - \lambda_l(\bar{x}_{lh} - \bar{X}_l))$$

$$= (\frac{1}{n} - \frac{1}{n'})\sum_h W_h\, a' D_h^{(1)} a$$

$$D_h^{(1)} = d_{hkl}^{(1)}$$

$$M(T)_R - M(T_{DSS})_R = (\frac{1}{n} - \frac{1}{n'})\sum\sum a_k a_l \sum_h W_h\, d_{hkl}^{(2)}$$

$$D_h^{(2)} = d_{hkl}^{(2)}$$

$$d_{hkl}^{(2)} = [(\bar{Y}_h - R_k\bar{X}_{kh}][(\bar{Y}_h - R_l\bar{X}_{lh}] \qquad k, l = 1,2,\ldots\ldots,p$$

It is to be noted that $D_h^{(1)}$ and $D_h^{(2)}$ are positive definite matrices which proves that under proportional allocation of the second sample, the multivariate difference and ratio estimators in DSS are always better than corresponding estimators in UDSS.

6. EMPIRICAL STUDY

To show the usefulness of the proposed estimators based on DSS compared to estimators based on UDSS suggested by Prabha (1992) in two stage design. It can be checked by comparing their variances and percent relative efficiency, numerically, an empirical study has been carried out.

Population: For this we consider the 2001 census data which is relates to the total number of agricultural labourers, total area, total population, and total no. of cultivators of 444 villages of Bhiwani district of Haryana. We take number of agricultural labourers in villages as y ,the total area of villages asx_1 , the total population of villages as x_2 ,and the total no. of cultivators as x_3.The whole population of Bhiwani district (444 villages) is divided into villages at fsu level and household at ssu level.

The whole population of 444 village's fsus is stratified into three strata according to area. So the strata become as under:

Strata	Area in Hectare
I	1-800 (227 villages)
II	800- 1660 (135 villages)
III	1660-2400 & above (82 villages)

The numerical values of the estimate of the population mean,its variances are worked out under each of the proposed estimators from population values. Subsequently, the percent relative efficiency of each of the estimator given in this paper is calculated. The result are presented in the form of Table II given below

Table I
Description of Population

Source	Village wise information of Bhiwani District of Haryana (2001 census data)
y	study variable (agricultural laborers)
x_1	auxiliary information (total area)
x_2	auxiliary information (total population)
x_3	auxiliary information (total number of cultivators)

Table II

(i)	(ii)	(iii)	(iv)
Estimators	**Auxiliary variable used**	**Variance** $\times \mathbf{10^7}$	**R.E** % **w.r.t.** $(\mathbf{T})_{D}$
$(T)_D$	x_1, x_2 and x_3	2.27	100
$(T_{DSS})_D$	x_1, x_2 and x_3	1.78	127
			R.E % **w.r.t.** $(\mathbf{T})_{R}$
$(T)_R$	x_1, x_2 and x_3	2.24	100
$(T_{DSS})_R$	x_1, x_2 and x_3	1.72	130

7. CONCLUSION

From table II, column III, we interpret that the estimators based on DSS give less variance as compare to USDS and in column IV estimators based on DSS have more relative efficiency as compare to USDS.

From the perusal of above table we conclude that estimators based on DSS at fsu level in two stage design as compare to the estimators based on USDS. Thus the estimator $(T_{DSS})_D$ and $(T_{DSS})_R$ are recommended for used in practice for the estimation of population mean

8. REFERENCES

[1] Ige, A.F. and Tripathi, T.P. 1987. On double sampling for stratification and use of auxiliary information. Jour. *Ind. Soc. Ag. Stat. 39*, 191-201.

[2] Prabha, S. 1992. On Some Estimation Problems in Sampling Using Multiauxiliary Information. *Ph.D Thesis, M.D. University, Rohtak*

[3] Rao, J.N.K. 1973. On double sampling for stratification and analytical surveys. *Biometrika 60*, 125-133.

[4] Shabbir, J. and Gupta S. 2011. On Estimating Finite Population Mean in Simple and Stratified Random Sampling. Commun. in Stat. 40, 199-212.

[5] Singh, H.P. and Vishwakarma, G.K. 2010. A general procedure for estimating the population mean in Stratified sampling using auxiliary information, Metron, vol LXVIII, no.1, 47-65.

[6] Sukhatme, P. V., Sukhatme, B.V. Sukhatme, S. and Ashok, C. 1984.*Sampling Theory of Surveys with Applications*, Iowa State University Press, Ames, Iowa.

[7] Tripathi, T.P. 1976. On double sampling for multi--variate ratio and difference methods of estimation. *Jour. Ind. Soc. Agr. Stat*. *38*, 33-34.

[8] Tripathi, T. P. and Bahl, S. 1991. Estimation of mean using double Sampling for Stratification and Multi auxiliary Information, Common. *Statist. Theory Meth.*. 20, 8, 2589-2602.

.