International Journal of Scientific Research in Mathematical and Statistical Sciences **Research Paper**
Volume-5, Issue-6, pp.196-207, December (2018) **E-ISSN:** 2348-4519

Double Calibration Approach Based Estimator for Population Mean in Two-Stage Stratified Random Sampling When Auxiliary Information Is Available At Element Level

Dhirendra Singh[1*], B.V.S. Sisodia[2]

Department of Agricultural Statistics, Narendra Deva University of Agriculture & Technology, Kumarganj, Faizabad-224229 (UP), India.

Available online at: www.isroset.org

Received: 03/Sept/2018, Accepted: 11/Oct/2018, Online: 31/Dec/2018

Abstract- Singh *et al.* (2017) have developed calibration estimators of population mean in two-stage stratified random sampling by calibrating integrated sampling design weight when the auxiliary information is available at element (second stage unit) level for the entire population. Obviously, if the auxiliary information is available at the element level for entire population, then population mean/total of the auxiliary variable is also known. In the present paper, double (two-steps) calibration estimators of the population mean have been developed by calibrating integrated sampling design weight at first step and calibrating stratum weight at the second step using known population total/mean of the auxiliary variable. A limited simulation study with real data has been conducted to examine the relative performance of the calibration estimators over the usual estimator of the population mean without using auxiliary information in two-stage stratified random sampling. It has been found from the results of simulation study that double (two-steps) calibration estimator has brought considerable improvement in the precision of the estimate of population mean.

***Keywords*:** Calibration Estimator, Auxiliary Information, Two-Stage Sampling, Stratified Random Sampling.

I. INTRODUCTION

Deville and Särndal (1992) made use of known population totals of auxiliary variables related to the study variate to calibrate sampling design weight for improving Horvitz-Thompson estimator of population total of the variable of interest. Several research workers have made significant contribution in this area. Särndal (2007) and Kim and Park (2010) have presented a comprehensive review of the work in calibration estimation in sample surveys. Calibration approach based estimation of population total has been extended to stratified random sampling by Singh *et al.* (1998), Tracy *et al.* (2003), Kim *et al.* (2007), Singh and Arnab (2011), Sinha *et al.* (2016) etc. Aditya *et al.* (2016) have developed calibration approach based regression type estimator of population total in two-stage sampling when the auxiliary information related to the study variate is available at primary stage unit (psu) level. Mourya *et al.* (2016) have also developed calibration estimator in two-stage sampling when the auxiliary information is available at second stage unit (ssu) level for selected psu(s). Recently, Singh *et al.* (2017) have developed calibration estimators of population mean in two-stage stratified random sampling by calibrating integrated sampling design weight when the auxiliary information is available at element (ssu) level for the entire population. Obviously, if the auxiliary information is available at the element level for entire population, then population mean/total of the auxiliary variable is also known. In the present paper, double (two-steps) calibration estimators of the population mean have been developed by (i) Calibrating integrated sampling design weight at first step, and (ii) calibrating stratum weight at the second step using known population total/mean of the auxiliary variable in section-4., their variances are derived and properties are discussed. Rest of the paper is organized as follows; the usual estimator of population mean in two-stage stratified sampling without using the auxiliary information is described in section-2. Development of one-step calibration estimator in two-stage stratified random sampling is described in section-3. A simulation study with real data has been conducted to examine the relative performance of the estimators in section-5. A concluding remark has been presented in section-6.

II. THE USUAL ESTIMATOR OF POPULATION MEAN IN TWO-STAGE STRATIFIED RANDOM SAMPLING WITHOUT USING AUXILIARY INFORMATION

Let the population of elements $U=(1,2,3.....K,.....N)$ is partitioned into $U_1,U_2,U_3,.......U_i,.......U_{N_I}$ psu's. The population of psu's is denoted by $U_I=(U_1,U_2.......U_i......U_{N_I})$. The size of U_i is denoted by N_i. So, we have $U=\bigcup_{i=1}^{N_I}U_i$ and $N=\sum_{i=1}^{N_I}N_i$. Let the population of psu's U_I is stratified into G strata, i.e $1,2,3,......g,.......G$. The size of the g^{th} stratum is denoted as N_g, i.e g^{th} stratum consists of N_g psu's such that $\sum_{g=1}^{G}N_g=N_I$. Let N_{gi} is the number of ssu of i^{th} psu in g^{th} stratum $(i=1,2,3...........N_g)$, such that $N_{go}=\sum_{i=1}^{N_g}N_{gi}$, the total number of elements in g^{th} stratum . Let the population of N_g psu's in the g^{th} stratum is denoted by $U_g=(U_{g1},U_{g2},.....U_{gi}......U_{gN_g})$.

We further define

$\bar{N}_{go}=\dfrac{N_{go}}{N_g}$, average number of ssu per psu.

t_{ygik} = value of y corresponding to k^{th} element of i^{th} psu in g^{th} stratum.

$t_{ygi}=\sum_{k=1}^{N_{gi}}t_{ygik}$, total of y in i^{th} psu of g^{th} stratum.

$\bar{t}_{ygi}=\dfrac{1}{N_{gi}}\sum_{k=1}^{N_{gi}}t_{ygik}$, mean per ssu in i^{th} psu of g^{th} stratum.

$t_{yg}=\sum_{i=1}^{N_g}\sum_{k=1}^{N_{gi}}t_{ygik}$, total of y in g^{th} stratum.

$\bar{t}_{yg}=\dfrac{t_{yg}}{N_g\bar{N}_{go}}=\dfrac{1}{N_g}\sum_{i=1}^{N_g}\dfrac{N_{gi}}{\bar{N}_{go}}\bar{t}_{ygi}$,the population mean per ssu in g^{th} stratum.

$\tilde{t}_{yg}=\dfrac{1}{N_g}\sum_{i=1}^{N_g}t_{ygi}$,the average total of y per psu.

At-first stage, a random sample s_g of n_g psu's from N_g psu's in g^{th} stratum is drawn according to sampling design $P_g(.)$ with the inclusion probabilities π_{gi} and π_{gij} at psu level.

At-second stage, we draw a random sample s_i of size n_i elements from the selected i^{th} psu in g^{th} stratum $(i=1,2,3.........n_g)$ according to design $P_i(\cdot)$ with inclusion probabilities $\pi_{gk/i}$ and $\pi_{gkl/i}$.

We also define

$$\Delta_{gij}=\pi_{gij}-\pi_{gi}\pi_{gj} \quad \text{with} \quad \tilde{\Delta}_{gij}=\Delta_{gij}\big/\pi_{gij} \quad \text{and}$$

$$\Delta_{gkl/i}=\pi_{gkl/i}-\pi_{gk/i}\pi_{gl/i}, \text{ with } \tilde{\Delta}_{gkl/i}=\frac{\Delta_{gkl/i}}{\pi_{gkl/i}} \tag{2.1}$$

The objective to estimate is the population mean

$$\bar{t}_y=\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\sum_{k=1}^{N_{gi}}t_{ygik}=\sum_{g=1}^{G}\frac{N_{go}}{N}\bar{t}_{yg}=\sum_{g=1}^{G}\Omega_g\bar{t}_{yg} \tag{2.2}$$

where $\Omega_g = \frac{N_{go}}{N}$, stratum weight, such that $\sum_{g=1}^{G} \Omega_g = 1$

The usual Horvitz-Thompson estimator of (HT) $\bar{t}_{yg}$ is given by

$$\hat{\bar{t}}_{yg(HT)} = \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} a_{gik} t_{ygik} \tag{2.3}$$

where $a_{gik} = a_{gi} a_{gk/i}$, $a_{gi} = \frac{1}{\pi_{gi}}$, $a_{gk/i} = \frac{1}{\pi_{gk/i}}$.

The above estimator can be alternatively expressed as

$$\hat{\bar{t}}_{yg(HT)} = \frac{\hat{t}_{yg(HT)}}{N_g \bar{N}_{go}} \tag{2.4}$$

where $\hat{t}_{yg(HT)} = \sum_{s_g} a_{gi} N_{gi} \hat{\bar{t}}_{ygi(HT)}$, and $\hat{\bar{t}}_{ygi(HT)} = \frac{1}{N_{gi}} \sum_{s_i} a_{gk/i} t_{ygik}$ is the HT estimator of $\bar{t}_{ygi}$.

The variance of $\hat{\bar{t}}_{yg(HT)}$ can be written as sum of two components as per Särndal *et al.* (1992)

$$V(\hat{\bar{t}}_{yg(HT)}) = \frac{V_{psu} + V_{ssu}}{N_g^2 \bar{N}_{go}^2} \tag{2.5}$$

With

$$V_{psu} = \sum_{U_g} \sum \Delta_{gij} \frac{t_{ygi}}{\pi_{gi}} \frac{t_{ygj}}{\pi_{gj}}, \quad V_{ssu} = \sum_{U_g} \frac{V_i}{\pi_{gi}} \text{ and } \quad V_i = \sum_{U_{gi}} \sum \Delta_{gkl/i} \frac{t_{ygik}}{\pi_{gk/i}} \frac{t_{ygil}}{\pi_{gl/i}}.$$

The first component V_{psu} is unbiasedly estimated by

$$\hat{V}_{psu} = \sum_{s_g} \sum \tilde{\Delta}_{gij} \frac{\hat{t}_{ygi}}{\pi_{gi}} \frac{\hat{t}_{ygj}}{\pi_{gj}} - \sum_{s_g} \frac{1}{\pi_{gi}} \left(\frac{1}{\pi_{gi}} - 1 \right) \hat{V}_i \tag{2.6}$$

where $\hat{V}_i = \sum_{s_i} \sum \tilde{\Delta}_{gkl/i} \frac{t_{ygik}}{\pi_{gk/i}} \frac{t_{ygil}}{\pi_{gl/i}}$

The second component V_{ssu} is unbiasedly estimated by

$$\hat{V}_{ssu} = \sum_{s_g} \frac{\hat{V}_i}{\pi_{gi}^2} \tag{2.7}$$

Therefore, $\hat{V}(\hat{\bar{t}}_{yg(HT)})$ is given by

$$\hat{V}(\hat{\bar{t}}_{yg(HT)}) = \frac{\hat{V}_{psu} + \hat{V}_{ssu}}{N_g^2 \bar{N}_{go}^2}$$

$$= \frac{1}{N_g^2 \bar{N}_{go}^2} \left(\sum_{s_g} \sum \tilde{\Delta}_{gij} \frac{\hat{t}_{ygi}}{\pi_{gi}} \frac{\hat{t}_{ygj}}{\pi_{gj}} + \sum_{s_g} \frac{\hat{V}_i}{\pi_{gi}} \right) \tag{2.8}$$

Now, the estimator of $\bar{t}_y$ in stratified random sampling is given by

$$\hat{\hat{t}}_y = \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{yg(HT)} \tag{2.9}$$

The variance of $\hat{\hat{t}}_y$ is given by

$$V\left(\hat{\hat{t}}_y\right) = \sum_{g=1}^{G} \Omega_g^2 V(\hat{\hat{t}}_{yg(HT)}), \text{ where } V\left(\hat{\hat{t}}_{yg(HT)}\right) \text{ is given in (2.5).} \tag{2.10}$$

The estimator of variance of $\hat{\hat{t}}_y$ is given by

$$\hat{V}\left(\hat{\hat{t}}_y\right) = \sum_{g=1}^{G} \Omega_g^2 \hat{V}(\hat{\hat{t}}_{yg(HT)}), \text{ where } \hat{V}\left(\hat{\hat{t}}_{yg(HT)}\right) \text{ is given in (2.8).} \tag{2.11}$$

If the sampling design is simple random sampling without replacement (SRSWOR) denoted as SI, the estimator $\hat{\hat{t}}_{yg(SI)}$ under SRSWOR is given by

$$\hat{\hat{t}}_{yg(SI)} = \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{N_{gi}}{\bar{\bar{N}}_{go}} \hat{t}_{ygi}, \text{ where } \hat{t}_{ygi} = \frac{1}{n_i} \sum_{k=1}^{n_i} t_{ygik} \tag{2.12}$$

The variance of $\hat{\hat{t}}_{yg(SI)}$ is given by

$$= \frac{N_g - n_g}{n_g N_g} S_{byg}^2 + \frac{1}{n_g N_g} \sum_{i=1}^{N_g} \left(\frac{N_{gi}}{\bar{\bar{N}}_{go}}\right)^2 \frac{(N_{gi} - n_i)}{n_i N_{gi}} S_{ygi}^2 \tag{2.13}$$

where $S_{byg}^2 = \dfrac{1}{N_g - 1} \displaystyle\sum_{i=1}^{N_g} \left(\frac{N_{gi}}{\bar{\bar{N}}_{go}} \bar{t}_{ygi} - \bar{t}_{yg}\right)^2$ and $S_{ygi}^2 = \dfrac{1}{N_{gi} - 1} \displaystyle\sum_{k=1}^{N_{gi}} \left(t_{ygik} - \bar{t}_{ygi}\right)^2$

The unbiased variance estimator is given by

$$\hat{V}\left(\hat{\hat{t}}_{yg(SI)}\right) = \frac{N_g - n_g}{n_g N_g} s_{byg}^2 + \frac{1}{n_g N_g} \sum_{i=1}^{n_g} \frac{(N_{gi} - n_i)}{n_i N_{gi}} s_{ygi}^2 \tag{2.14}$$

where

$$s_{byg}^2 = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} \left(\frac{N_{gi}}{\bar{\bar{N}}_{go}} \hat{t}_{ygi} - \hat{\hat{t}}_{yg}\right)^2 \text{ and } s_{ygi}^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left(\frac{N_{gi}}{\bar{\bar{N}}_{go}} t_{ygik} - \tilde{t}_{ygi}\right)^2, \ \tilde{t}_{ygi} = \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{N_{gi}}{\bar{\bar{N}}_{go}} t_{ygik}.$$

The estimator $\hat{\hat{t}}_y$ in SRSWOR can be expressed as

$$\hat{\hat{t}}_{y(SI)} = \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{yg(SI)}, \text{ where } \hat{\hat{t}}_{yg(SI)} \text{ is given in above equation (2.12).} \tag{2.15}$$

The variance of $\hat{\hat{t}}_{yg(SI)}$ is given by

$$V\left(\hat{\hat{t}}_{y(SI)}\right) = \sum_{g=1}^{G} \Omega_g^2 V\left(\hat{\hat{t}}_{yg(SI)}\right), \text{ where } V\left(\hat{\hat{t}}_{yg(SI)}\right) \text{ is given in (2.13).} \tag{2.16}$$

The unbiased variance estimator is given by

$$\hat{V}\left(\hat{\hat{t}}_{y(SI)}\right) = \sum_{g=1}^{G} \Omega_g^2 \hat{V}\left(\hat{\hat{t}}_{yg(SI)}\right), \text{ where } \hat{V}\left(\hat{\hat{t}}_{yg(SI)}\right) \text{ is given in (2.14).} \tag{2.17}$$

III. CALIBRATION ESTIMATOR OF POPULATION MEAN IN TWO-STAGE STRATIFIED RANDOM SAMPLING WHEN AUXILIARY INFORMATION IS AVAILABLE AT SSU LEVEL

We follow the notations and definitions as described section-2. Consider that the auxiliary information t_{xgik} related to the study variate t_{ygik} is available at ssu level corresponding to k^{th} elements of i^{th} psu in g^{th} stratum. The psu total of x is therefore automatically obtained, i.e $t_{xgi} = \sum_{k=1}^{N_{gi}} t_{xgik}$ for i^{th} psu. Let $t_{xg} = \sum_{i=1}^{N_g} \sum_{k=1}^{N_{gi}} t_{xgik}$ be the total of x for g^{th} stratum.

The usual Horvitz-Thompson estimator of $\bar{t}_{yg}$ without using auxiliary information given in (2.3) reproduced here is

$$\hat{\bar{t}}_{yg(HT)} = \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} a_{gik} t_{ygik} \tag{3.1}$$

where $a_{gik} = a_{gi} a_{gk/i}$, which is an integrated weight. We want to calibrated a_{gik} .Let w_{gik} be the integrated calibrated weight and therefore, the calibrated estimator of $\bar{t}_{yg}$ is given by

$$\hat{\bar{t}}_{yg}^{c} = \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} w_{gik} t_{ygik} \tag{3.2}$$

We find out the w_{gik} by minimizing the chi-square distance measure

$$\sum_{s_g} \sum_{s_i} \frac{\left(w_{gik} - a_{gik}\right)^2}{q_{gik} a_{gik}}$$

subject to the constraints

$$\frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} w_{gik} t_{xgik} = \bar{t}_{xg} \tag{3.3}$$

Therefore , the following function will be minimized with respect to w_{gik}

$$\phi\left(w_{gik}, \lambda\right) = \sum_{s_g} \sum_{s_i} \frac{\left(w_{gik} - a_{gik}\right)^2}{q_{gik} a_{gik}} - 2\lambda \left(\sum_{s_g} \sum_{s_i} w_{gik} t_{xgik} - N_g \bar{N}_{go} \bar{t}_{xg} \right) \tag{3.4}$$

$$\frac{\partial \phi\left(w_{gik}, \lambda\right)}{\partial\left(w_{gik}\right)} = 0$$, yields

$$w_{gik} = a_{gik} + \frac{q_{gik} a_{gik} t_{xgik}}{\sum_{s_g} \sum_{s_i} q_{gik} a_{gik} t_{xgik}^2} \left(N_g \bar{N}_{go} \bar{t}_{xg} - \sum_{s_g} \sum_{s_i} a_{gik} t_{xgik} \right) \tag{3.5}$$

Putting the value of w_{gik} in above equation (3.2), we get

$$\begin{aligned} \hat{\bar{t}}_{yg}^{c} &= \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} a_{gik} t_{ygik} + \frac{\sum_{s_g} \sum_{s_i} q_{gik} a_{gik} t_{xgik} t_{ygik}}{\sum_{s_g} \sum_{s_i} q_{gik} a_{gik} t_{xgik}^2} \left(\bar{t}_{xg} - \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} a_{gik} t_{xgik} \right) \\ &= \hat{\bar{t}}_{yg(HT)} + \hat{B}_g \left(\bar{t}_{xg} - \hat{\bar{t}}_{xg(HT)} \right) \end{aligned} \tag{3.6}$$

where $\hat{\bar{t}}_{xg(HT)} = \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} a_{gik} t_{xgik}$, $\hat{\bar{t}}_{yg(HT)} = \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} \sum_{s_i} a_{gik} t_{ygik}$ are the Horvitz-Thompson estimators of $\bar{t}_{xg}$ and $\bar{t}_{yg}$, and $\hat{B}_g = \frac{\sum_{s_g} \sum_{s_i} q_{gik} a_{gik} t_{xgik} t_{ygik}}{\sum_{s_g} \sum_{s_i} q_{gik} a_{gik} t_{xgik}^2}$.

Now, the calibration estimator of $\bar{t}_y$ in two-stage stratified random sampling is given by

$$\hat{\bar{t}}_y^c = \sum_{g=1}^{G} \Omega_g \hat{\bar{t}}_{yg}^c$$

$$= \sum_{g=1}^{G} \Omega_g \left[\hat{\bar{t}}_{yg(HT)} + \hat{B}_g \left(\bar{t}_{xg} - \hat{\bar{t}}_{xg(HT)} \right) \right] \tag{3.7}$$

Under SRSWOR (say, SI) and for $q_{gik} = \frac{1}{t_{xgik}}$, we shall show that estimator $\hat{\bar{t}}_y^c$ given in (3.7) reduces to separate ratio estimator in two-stage stratified random sampling. From (3.7), we can write $\hat{\bar{t}}_y^c$ under SRSWOR as

$$\hat{\bar{t}}_{y(SI)}^c = \sum_{g=1}^{G} \Omega_g \hat{\bar{t}}_{yg(SI)}^c$$

$$= \sum_{g=1}^{G} \Omega_g \left[\frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi} + \hat{B}_{g2(SI)} \left(\bar{t}_{xg} - \frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{xgi} \right) \right],$$

where $\hat{B}_{g2(SI)} = \frac{\frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi}}{\frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{xgi}}$

$$= \sum_{g=1}^{G} \Omega_g \frac{\hat{\bar{t}}_{yg}}{\hat{\bar{t}}_{xg}} \bar{t}_{xg}, \text{ where } \hat{\bar{t}}_{yg} = \frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi} \text{ and } \hat{\bar{t}}_{xg} = \frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{xgi}. \tag{3.8}$$

Following the procedure given by Sukhatme *et al.* (1984), the approximate variance of $\hat{\bar{t}}_{y(SI)}^c$ has been derived as

$$V\left(\hat{\bar{t}}_{y(SI)}^c\right) = \sum_{g=1}^{G} \Omega_g^2 V\left(\hat{\bar{t}}_{yg(SI)}^c\right)$$

$$= \sum_{g=1}^{G} \Omega_g^2 \left[\left(\frac{1}{n_g} - \frac{1}{N_g} \right) \left(S_{byg}'^2 - 2R_g S_{bxyg}' + R_g^2 S_{bxg}'^2 \right) + \frac{1}{n_g N_g} \sum_{i=1}^{N_g} u_{gi}^2 \left(\frac{1}{n_i} - \frac{1}{N_{gi}} \right)^2 D_{gi}^2 \right] \tag{3.9}$$

where $D_{gi}^2 = S_{ygi}^2 - 2R_g S_{xygi} + R_g^2 S_g^2$, $R_g = \dfrac{\bar{t}_{yg\bullet\bullet}}{\bar{t}_{xg\bullet\bullet}}$ and $u_{gi}^2 = \dfrac{N_{gi}^2}{\bar{N}_{go}^2}$, $S_{byg}'^2 = \dfrac{1}{N_g - 1}\sum_{i=1}^{N_g}\left(u_{gi}\bar{t}_{ygi\bullet} - \bar{t}_{yg\bullet\bullet}\right)^2$,

$S_{ygi}^2 = \dfrac{1}{N_{gi}-1}\sum_{k=1}^{N_{gi}}\left(t_{ygik} - \bar{t}_{ygi\bullet}\right)^2$, $S_{bxg}'^2 = \dfrac{1}{N_g - 1}\sum_{i=1}^{N_g}\left(u_{gi}\bar{t}_{xgi\bullet} - \bar{t}_{xg\bullet\bullet}\right)^2$, $S_{xgi}^2 = \dfrac{1}{N_{gi}-1}\sum_{k=1}^{N_{gi}}\left(t_{xgik} - \bar{t}_{xgi\bullet}\right)^2$,

$S_{bxyg}' = \dfrac{1}{N_g - 1}\sum_{i=1}^{N_g}\left(u_{gi}\bar{t}_{xgi\bullet} - \bar{t}_{xg\bullet\bullet}\right)\left(u_{gi}\bar{t}_{ygi\bullet} - \bar{t}_{yg\bullet\bullet}\right)$ $S_{xygi} = \dfrac{1}{N_{gi}-1}\sum_{i=1}^{N_{gi}}\left(t_{xgik} - \bar{t}_{xgi\bullet}\right)\left(t_{ygik} - \bar{t}_{ygi\bullet}\right)$.

The approximate variance of $\hat{\bar{t}}_{y(SI)}^{c}$ can alternatively be written as

$$= \sum_{g=1}^{G}\Omega_g^2\left[\left(\frac{1}{n_g} - \frac{1}{N_g}\right)\frac{1}{N_g - 1}\sum_{i=1}^{N_g}u_{gi}^2\left(\bar{y}_{gi\bullet} - R_g\bar{x}_{gi\bullet}\right)^2 + \frac{1}{n_g N_g}\sum_{i=1}^{N_g}u_{gi}^2\left(\frac{1}{n_i} - \frac{1}{N_{gi}}\right)^2 D_{gi}^2\right] \tag{3.10}$$

The approximate estimator of variance of $\hat{\bar{t}}_{y(SI)}^{c}$ is given by

$$\hat{V}\left(\hat{\bar{t}}_{y(SI)}^{c}\right) = \sum_{g=1}^{G}\Omega_g^2\hat{V}\left(\hat{\bar{t}}_{yg(SI)}^{c}\right)$$

$$= \sum_{g=1}^{G}\Omega_g^2\left[\left(\frac{1}{n_g} - \frac{1}{N_g}\right)\left(s_{byg}'^2 - 2R_{g1}s_{bxyg}' + R_{g1}^2 s_{bxg}'^2\right) + \frac{1}{n_g N_g}\sum_{i=1}^{n_g}u_{gi}^2\left(\frac{1}{n_i} - \frac{1}{N_{gi}}\right)^2 d_{gi}^2\right] \tag{3.11}$$

where $d_{gi}^2 = s_{ygi}^2 - 2R_{g1}s_{xygi} + R_{g1}^2 s_g^2$, $R_{g1} = \dfrac{\sum_{i=1}^{n_g}u_{gi}\hat{\bar{t}}_{ygi}/n_g}{\sum_{i=1}^{n_g}u_{gi}\hat{\bar{t}}_{xgi}/n_g}$,

$$s_{bxyg}' = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}\left(u_{gi}\hat{\bar{t}}_{ygi} - \left(\frac{1}{n_g}\sum_{i=1}^{n_g}u_{gi}\hat{\bar{t}}_{ygi}\right)\right)\left(u_{gi}\hat{\bar{t}}_{xgi} - \left(\frac{1}{n_g}\sum_{i=1}^{n_g}u_{gi}\hat{\bar{t}}_{xgi}\right)\right),$$

$s_{byg}'^2 = \dfrac{1}{n_g - 1}\sum_{i=1}^{n_g}\left(u_{gi}\hat{\bar{t}}_{ygi} - \left(\dfrac{1}{n_g}\sum_{i=1}^{n_g}u_{gi}\hat{\bar{t}}_{ygi}\right)\right)^2$ $s_{bxg}'^2 = \dfrac{1}{n_g - 1}\sum_{i=1}^{n_g}\left(u_{gi}\hat{\bar{t}}_{xgi} - \left(\dfrac{1}{n_g}\sum_{i=1}^{n_g}u_{gi}\hat{\bar{t}}_{xgi}\right)\right)^2$,

$s_{xygi} = \dfrac{1}{n_i - 1}\sum_{k=1}^{n_i}\left(t_{xgik} - \hat{\bar{t}}_{xgi}\right)\left(t_{ygik} - \hat{\bar{t}}_{ygi}\right)$, $s_{ygi} = \dfrac{1}{n_i - 1}\sum_{k=1}^{n_i}\left(t_{ygik} - \hat{\bar{t}}_{ygi}\right)^2$, $s_{xgi} = \dfrac{1}{n_i - 1}\sum_{k=1}^{n_i}\left(t_{xgik} - \hat{\bar{t}}_{xgi}\right)^2$

The approximate estimator of variance of $\hat{\bar{t}}_{y(SI)}^{c}$ can alternatively be written as

$$= \sum_{g=1}^{G}\Omega_g^2\left[\left(\frac{1}{n_g} - \frac{1}{N_g}\right)\frac{1}{n_g - 1}\sum_{i=1}^{n_g}u_{gi}^2\left(\bar{y}_{s_i} - R_{g1}\bar{x}_{s_i}\right)^2 + \frac{1}{n_g N_g}\sum_{i=1}^{N_g}u_{gi}^2\left(\frac{1}{n_i} - \frac{1}{N_{gi}}\right)^2 d_{gi}^2\right] \tag{3.12}$$

IV. DEVELOPMENT OF DOUBLE (TWO-STEPS) CALIBRATION ESTIMATOR OF POPULATION MEAN IN TWO-STAGE STRATIFIED RANDOM SAMPLING WHEN BOTH STRATUM WEIGHTS AND DESIGN WEIGHTS AT PSU LEVEL ARE CALIBRATED

The calibrated estimator of $\bar{t}_y$ is given by

$\hat{\hat{t}}_y^c = \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{yg}^c$, where $\Omega_g = \frac{N_{go}}{N}$ = stratum weight and $\hat{\hat{t}}_{yg}^c = \frac{\hat{t}_{ygi}}{N_g \bar{N}_{go}}$. (4.1)

Now, we want to calibrate stratum weight Ω_g . Let Ω'_g be the calibrate weight of Ω_g . Therefore, two-steps calibration estimator is proposed as

$$\hat{\hat{t}}_y^{cc} = \sum_{g=1}^{G} \Omega'_g \hat{\hat{t}}_{yg}^c \tag{4.2}$$

We minimize the chi-square distance measure function $\sum_{g=1}^{G} \left(\Omega'_g - \Omega_g\right)^2 \Big/ q_g \Omega_g$

subject to the calibration constraints

$$\sum_{g=1}^{G} \Omega'_g \hat{\hat{t}}_{xg} = \bar{t}_x \text{ and } \sum_{g=1}^{G} \Omega_g = \sum_{g=1}^{G} \Omega'_g \tag{4.3}$$

The following function is minimized with respect to Ω'_g .

$$\phi\left(\Omega'_g, \lambda\right) = \sum_{g=1}^{G} \left(\Omega'_g - \Omega_g\right)^2 \Big/ q_g \Omega_g + 2\lambda_1 \left(\sum_{g=1}^{G} \Omega'_g \hat{\hat{t}}_{xg} - \bar{t}_x\right) + 2\lambda_2 \left(\sum_{g=1}^{G} \Omega_g - \sum_{g=1}^{G} \Omega'_g\right) \tag{4.4}$$

where λ_1 and λ_2 are Lagrange multiplier. Differentiation equation (4.4) with respect to Ω'_g and equating it to zero, i.e the solution of $\frac{\partial \phi\left(\Omega'_g, \lambda\right)}{\partial\left(\Omega'_g\right)} = 0$, yields

$$\Omega'_g = \Omega_g + \frac{q_g \Omega_g \left[\hat{\hat{t}}_{xg} - \frac{\sum_{g=1}^{G} q_g \Omega_g \hat{\hat{t}}_{xg}}{\sum_{g=1}^{G} q_g \Omega_g}\right]}{\sum_{g=1}^{G} q_g \Omega_g \left(\hat{\hat{t}}_{xg}\right)^2 - \frac{\left(\sum_{g=1}^{G} q_g \Omega_g \hat{\hat{t}}_{xg}\right)^2}{\sum_{g=1}^{G} q_g \Omega_g}} \left(\bar{t}_x - \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{xg}\right) \tag{4.5}$$

Putting the value of Ω'_g in equation (4.2), then the estimator of $\bar{t}_y^{cc}$ reduces as

$$\hat{\hat{t}}_y^{cc} = \hat{\hat{t}}_y^c + \hat{b}_1 \left(\bar{t}_x - \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{xg}\right) \tag{4.6}$$

where $$\hat{b}_1 = \frac{\sum_{g=1}^{G} \Omega_g q_g \hat{\hat{t}}_{yg}^{c} \hat{\hat{t}}_{xg} - \frac{\sum_{g=1}^{G} \Omega_g q_g \hat{\hat{t}}_{yg}^{c} \sum_{g=1}^{G} \Omega_g q_g \hat{\hat{t}}_{xg}}{\sum_{g=1}^{G} \Omega_g q_g}}{\sum_{g=1}^{G} \Omega_g q_g \left(\hat{\hat{t}}_{xg}\right)^2 - \frac{\left(\sum_{g=1}^{G} \Omega_g q_g \hat{\hat{t}}_{xg}\right)^2}{\sum_{g=1}^{G} \Omega_g q_g}}$$

The calibration estimator for $q_g = 1$, reduces as

$$\hat{\hat{t}}_y^{cc} = \hat{\hat{t}}_y^{c} + \hat{b}_1' \left(\bar{t}_x - \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{xg} \right) \tag{4.7}$$

where $$\hat{b}_1' = \frac{\sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{yg}^{c} \hat{\hat{t}}_{xg} - \frac{\sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{yg}^{c} \sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{xg}}{\sum_{g=1}^{G} \Omega_g}}{\sum_{g=1}^{G} \Omega_g \left(\hat{\hat{t}}_{xg}\right)^2 - \frac{\left(\sum_{g=1}^{G} \Omega_g \hat{\hat{t}}_{xg}\right)^2}{\sum_{g=1}^{G} \Omega_g}}$$

The conditional approximate variance of $\hat{\hat{t}}_y^{cc}$ for given Ω_g' is given by

$$V\left(\hat{\hat{t}}_y^{cc}\right) = \sum_{g=1}^{G} \Omega_g'^2 V\left(\hat{\hat{t}}_{yg}^{c}\right) \tag{4.8}$$

Following the Särndal *et al.* (1992) and Aditya *et al.* (2016), the variance of $\hat{\hat{t}}_{yg}^{c}$ can be written as

$$V\left(\hat{\hat{t}}_{yg}^{c}\right) = \frac{1}{N_g^2 \bar{N}_{go}^2} \left[\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \Delta_{gij} \frac{U_{gi} U_{gj}}{\pi_{gi} \pi_{gj}} + \sum_{i=1}^{N_g} \frac{1}{\pi_{gi}} \sum_{k=1}^{N_{gi}} \sum_{l=1}^{N_{gi}} \Delta_{gkl/i} \frac{t_{ygik} t_{ygil}}{\pi_{gk/i} \pi_{gl/i}} \right]$$

where $U_{gi} = t_{ygi} - B t_{xgi}$ and $$B = \frac{\sum_{i=1}^{N_g} a_{gi} t_{ygi} \left(t_{xgi} - \frac{\sum_{i=1}^{N_g} a_{gi} t_{xgi}}{\sum_{i=1}^{N_g} a_{gi}} \right)}{\sum_{i=1}^{N_g} a_{gi} t_{xgi}^2 - \frac{\left(\sum_{i=1}^{N_g} a_{gi} t_{xgi}\right)^2}{\sum_{i=1}^{N_g} a_{gi}}}$$

The conditional approximate estimate of variance of $\hat{\hat{t}}_y^{cc}$ for given Ω_g' is given by

$$\hat{V}\left(\hat{\hat{t}}_{y}^{cc}\right)=\sum_{g=1}^{G}\Omega_{g}'^{2}\hat{V}\left(\hat{\hat{t}}_{yg}^{c}\right) \tag{4.9}$$

where $\hat{V}\left(\hat{\hat{t}}_{yg}^{c}\right)$ is given by

$$\hat{V}\left(\hat{\hat{t}}_{yg}^{c}\right)=\frac{1}{N_g^2\bar{N}_{go}^2}\left[\frac{1}{2}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\left(-\tilde{\Delta}_{gij}\right)\left(w_{gi}u_{gi}-w_{gj}u_{gj}\right)^2+\frac{1}{2}\sum_{i=1}^{n_g}\frac{1}{\pi_{gi}^2}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left(-\tilde{\Delta}_{gkl/i}\right)\left(\frac{t_{ygik}}{\pi_{gk/i}}-\frac{t_{ygil}}{\pi_{gl/i}}\right)^2\right]$$

where $u_{gi}=\hat{t}_{ygi}-\tilde{B}\hat{t}_{xgi}$ and $\tilde{B}=\dfrac{\sum_{i=1}^{n_g}a_{gi}\hat{t}_{ygi(HT)}\left(t_{xgi}-\dfrac{\sum_{i=1}^{n_g}a_{gi}t_{xgi}}{\sum_{i=1}^{n_g}a_{gi}}\right)}{\left[\sum_{i=1}^{n_g}a_{gi}t_{xgi}^2-\dfrac{\left(\sum_{i=1}^{n_g}a_{gi}t_{xgi}\right)^2}{\sum_{i=1}^{n_g}a_{gi}}\right]}$

Under SRSWOR (say, SI), then calibration estimator for $q_g=1$, reduces as

$$\hat{\hat{t}}_{y(SI)}^{cc}=\hat{\hat{t}}_{y}^{c}+\hat{b}_1'\left(\bar{t}_x-\sum_{g=1}^{G}\Omega_g\hat{\hat{t}}_{xg}\right) \tag{4.10}$$

where $\hat{b}_1'=\dfrac{\sum_{g=1}^{G}\Omega_g\hat{\hat{t}}_{yg}^{c}\hat{\hat{t}}_{xg}-\dfrac{\sum_{g=1}^{G}\Omega_g\hat{\hat{t}}_{yg}^{c}\sum_{g=1}^{G}\Omega_g\hat{\hat{t}}_{xg}}{\sum_{g=1}^{G}\Omega_g}}{\sum_{g=1}^{G}\Omega_g\left(\hat{\hat{t}}_{xg}\right)^2-\dfrac{\left(\sum_{g=1}^{G}\Omega_g\hat{\hat{t}}_{xg}\right)^2}{\sum_{g=1}^{G}\Omega_g}}$

The conditional approximate variance of $\hat{\hat{t}}_{y}^{cc}$ for given Ω_g' is given by

$$V\left(\hat{\hat{t}}_{y(SI)}^{cc}\right)=\sum_{g=1}^{G}\Omega_g'^2V\left(\hat{\hat{t}}_{yg(SI)}^{c}\right) \tag{4.11}$$

where $V\left(\hat{\hat{t}}_{yg(SI)}^{c}\right)=\dfrac{1}{N_g^2\bar{N}_{go}^2}\left[\dfrac{\left(N_g-n_g\right)}{n_g\left(N_g-1\right)}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(-U_{gi}U_{gj}\right)-\dfrac{N_g}{n_g}\sum_{i=1}^{N_g}\dfrac{\left(N_{gi}-n_i\right)}{n_i\left(N_{gi}-1\right)}\sum_{k=1}^{N_{gi}}\sum_{l=1}^{N_{gi}}t_{ygik}t_{ygil}\right]$

$U_{gi}=t_{ygi}-B_{(SI)}t_{xgi}$ and $B_{(SI)}=\dfrac{\sum_{i=1}^{N_g'}t_{ygi}\left(t_{xgi}-\sum_{i=1}^{N_g}t_{xgi}\Big/N_g\right)}{\sum_{i=1}^{N_g}t_{xgi}^2-\left(\sum_{i=1}^{N_g}t_{xgi}\right)^2\Big/N_g}$

The conditional approximate estimate of variance of $\hat{\hat{t}}_{y}^{cc}$ for given Ω'_g is given by

$$\hat{V}\left(\hat{\hat{t}}_{y}^{cc}\right)=\sum_{g=1}^{G}\Omega_g'^2\hat{V}\left(\hat{\hat{t}}_{yg}^{c}\right) \tag{4.12}$$

where

$$V\left(\hat{\hat{t}}_{yg(SI)}^{c}\right)=\frac{1}{N_g^2\bar{N}_{go}^2}\left[\frac{1}{2}\frac{\left(N_g-n_g\right)}{N_g\left(n_g-1\right)}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g}\left(w_{gi}u_{gi}-w_{gj}u_{gj}\right)+\frac{1}{2}\frac{N_g^2}{n_g^2}\sum_{i=1}^{n_g}\frac{N_{gi}\left(N_{gi}-n_i\right)}{n_i^2\left(n_i-1\right)}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left(t_{ygik}-t_{ygil}\right)^2\right]$$

$$u_{gi}=\hat{t}_{ygi}-\hat{B}_{(SI)}\hat{t}_{xgi} \text{ and } \hat{B}_{(SI)}=\frac{\sum_{i=1}^{n_g}\hat{t}_{ygi}\left(\hat{t}_{xgi}-\sum_{i=1}^{n_g}\hat{t}_{xgi}\Big/n_g\right)}{\sum_{i=1}^{n_g}\hat{t}_{xgi}^2-\left(\sum_{i=1}^{n_g}\hat{t}_{xgi}\right)^2\Big/n_g}.$$

V. SIMULATION STUDY

A limited simulation study has been carried out with real data. The population MU284 given in Appendix-C of Särndal *et al.* (1992) have been used. There are 50 psu's of varying size. The variable under study (y) is population of 1985 and an auxiliary variable (x) is the population of 1975. The 50 psu's are stratified into 4 strata considering the value of x in ascending order. The stratum I consists of 13 psu's, stratum II consists of 14 psu's, stratum III consists of 12 psu's, stratum IV consists of 11 psu's respectively. The samples of size 4 psu's were drawn by SRSWOR independently from strata 1 to 4, respectively. This process has been repeated 300 times independently. That means, we obtained 300 samples of size 4 psu's from each stratum. Sub samples of size 3 ssu's are drawn by SRSWOR from each sample of psu's in each stratum. The values of y and x in sub samples were used to compute the population mean. In this process, we get 300 estimates of $\hat{\hat{t}}_{yg(SI)}$, $\hat{\hat{t}}_{yg(SI)}^{c}$ from 300 sub samples in each stratum. We compute the values of $\hat{T}_i$ based on usual estimator $\hat{\hat{t}}_{y(SI)}$ without using auxiliary information and calibration estimators, $\hat{\hat{t}}_{y(SI)}^{c}$, $\hat{\hat{t}}_{y(SI)}^{cc}$ from 1200 samples. The true populations mean of y has also been computed i.e 29.363. The following two criteria were used for assessing the relative performance of these estimators:

(i) The percent absolute relative bias (%RB) defined as,

$$\%RB\left(\hat{T}\right)=\frac{1}{S}\left(\sum_{i=1}^{S}\left|\frac{\hat{T}_i-T}{T}\right|\right)\times 100$$

(ii) The percent relative root mean square error (%RRMSE) defined as,

$$\%RRMSE(\hat{\theta})=\sqrt{\frac{1}{S}\sum_{i=1}^{S}\left(\frac{\hat{T}_i-T}{T}\right)^2}\times 100$$

Where S is the number of simulation.

The percent relative bias (%RB) and the percent relative root mean square error (%RRMSE) has been computed for each $\hat{T}_i$. Their values are presented in the table-5.1.

Table-5.1: Percent relative bias (%RB) and percent relative root mean square error (%RRMSE) of the estimators.

Estimators	%RB	%RRMSE
$\hat{\hat{t}}_{y(SI)}$	*	**7.013**
$\hat{\hat{t}}_{y(SI)}^{c}$	**0.189**	**0.362**

$\hat{\hat{t}}^{cc}_{y(SI)}$	**0.146**	**0.281**

[*] **Unbiased estimator**

It can be observed from the results of the Table-5.1 that calibration approach for estimation of the population mean of y has drastically decreased the percent relative root mean square error (%RRMSE) to about 0.3 percent from 7.0 percent when usual estimator without using auxiliary information was applied. Among the calibration estimators, two-steps calibration estimator $\left(\hat{\hat{t}}^{cc}_{y(SI)}\right)$ was found to be the best as it has lowest %RRMSE of 0.281 percent. The percent relative bias has been found to be within the range of below one percent for all the calibrated estimators. The result shows that the two-steps calibration approach of estimating population mean in two-stage stratified random sampling has brought considerable improvement in the precision of the estimates.

VI. CONCLUDING REMARKS

If the auxiliary information is available at element (second stage unit) level for the entire population, then the calibration approach based calibration estimator have brought significant improvement in the precision of the estimate of population mean in two-stage stratified random sampling. It may be mentioned that the two-steps calibration estimator $\hat{\hat{t}}^{cc}_{y(SI)}$ has been found better performance than the other one-step calibration estimator.

REFERENCES

[1]. Aditya, Kaustav., Sud, U.C., and Chandra, H. (2016). Calibration based regression type estimator of the population total under two-stage sampling design. Journal of the Indian Society of Agricultural Statistics, **70 (1)**, 19-24.

[2]. Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87,** 376–382.

[3]. Estevao, V.M. and Särndal, C.E. (2009). A new face on two-phase sampling with calibration estimators. Survey Methodology, **35(1)**, 3-4.

[4]. Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663-685.

[5]. Kim, J.M., Sunger, E.A. and Heo, T.Y. (2007). Calibration approach estimators in stratified sampling. Statistics and Probability Letters, **77,** 99-103.

[6]. Kim, J.K. and Park, M. (2010). Calibration estimation in survey sampling. *International Statistical Review*, **78,** 21-39.

[7]. Mourya, K.K., Sisodia, B.V.S. and Chandra, H. (2016). Calibration approach for estimating finite population parameter in two-stage sampling, Journal of Statistical Theory and Practice, (Online, 19 May 2016).

[8]. Mourya, K.K., Sisodia, B.V.S., Singh, Amar and Rai, V.N. (2016). Calibration estimators in cluster sampling when cluster are of unequal size. Int. J. Agricult. and Statistical Sci., **12(2),** 351-358.

[9]. Särndal, C.E., Swensson, B. and Wretman, J. (1992). Model-assisted Survey Sampling. New York: Springer-Verlag.

[10]. Särndal, C.E. (2007). The calibration approach in survey theory and practice. Survey Methodology, **33(2)**, 99-119.

[11]. Singh, Dhirendra., Sisodia, B.V.S., Kumar, Sunil., and Kumar, Sandeep. (2017). Some calibration estimators for finite population mean in two-stage stratified random sampling. Communication in Statistics- Theory and Methods, (submitted).

[12]. Singh, Dhirendra., Sisodia, B.V.S., Rai, V.N., and Kumar, Sandeep. (2017). A calibration approach based regression and ratio type estimators of finite population mean in two-stage stratified random sampling. Journal of the Indian Society of Agricultural Statistics, (submitted).

[13]. Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of the general regression estimators: Higher level calibration approach. *Survey methodology*, **24(1)**, 41-50.

[14]. Sinha, N., Sisodia, B.V.S., Singh, S. and K. Singh, Sanjay. (2016). Calibration approach estimation of mean in stratified sampling and stratified double sampling. Communication in Statistics- Theory and Methods, (Online, 27 May, 2016).

[15]. Singh, S. and Arnab, R. (2011). On calibration of design weights. *Metron*, **69(2),** 185-205.

[16]. Tracy, D.S., Singh, S. and Arnab, R. (2003). Note on calibration estimators in stratified and double sampling. *Survey Methodology*, **29**, 99-106.