https://doi.org/10.56093/JISAS.V80I1.1
Available online at https://isas.live/newsite
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 80(1) 2026 1–8

Calibration Estimators of Finite Population Mean in Multi-Stage Stratified Random Sampling

B.V.S. Sisodia[1] and Dhirendra Singh[2]

[1]*Acharya Narendra Deva University of Agriculture and Technology, Ayodhya*
[2]*Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi*

Received 09 February 2026 ; Revised 12 March 2026; Accepted 14 March 2026

SUMMARY

In present paper, an attempt has been made to develop calibration estimators of population mean in two-stage stratified random sampling when the auxiliary information is available at element (secondary stage unit) level for the entire population by calibrating integrated design weight. Some simulation studies with real data have been conducted to examine the relative performance of the 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 the simulation studies that there has been significant improvement in precision of the estimate by using proposed calibration estimators.

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

1. INTRODUCTION

Calibration approach based estimation of finite population mean / total in survey sampling has received a lot of attention by survey statisticians after the initial work of Deville and Särndal (1992). Since then a lot of contributions have been made on calibration estimators by various research workers. Notably among them are Singh *et al.* (1998), (1999), Kott (2006), Farrel and Singh (2005), Estevao and Särndal (2000), (2006), (2009), Kim and Park (2010) etc. Tracy *et al.* (2003), Kim *et al.* (2007), Singh and Arnab (2011), Sinha *et al.* (2016) etc. have extended calibration approach based estimation to stratified random sampling. Kaustav *et al.* (2016) have developed calibration estimator of population mean in two-stage sampling when auxiliary information is available at primary stage unit (psu) level. Mourya *et al.* (2016), (2016) have also dealt with calibration estimators in two-stage sampling and cluster sampling when auxiliary information is available at second stage unit (ssu) level for the selected psu(s). In fact, in most of the sample surveys, multi-stage stratified random sampling designs are generally used. In recent paper by Singh *et al.* (2017), some calibration estimators of finite population mean in two-stage stratified random sampling have been developed when the auxiliary information is available at ssu level for the selected psu(s). Singh *et al.* (2020) have also developed calibration estimators in two-stage stratified random sampling when information on auxiliary is available at all psu level. However, there could be some situations where the auxiliary information is available at the element (ssu) level for the entire population. For example, in survey for crop yield estimation at district level, operation land holding sizes of households (element level) are easily available on web site of the government. This can be considered as the auxiliary information at element level (Household level) in two-stage stratified random sampling where Tehsil (Taluka) are strata, villages are psu(s) and households are ssu(s). Similarly, in health surveys for a given district, Tehsils can be considered as strata, the number of public and

Corresponding author: B.V.S. Sisodia
E-mail: bvssisodia@gmail.com & dhirendra.teotia@gmail.com

private hospitals can be considered as psu(s) within strata and number of indoor patients suffering with different ailments can be considered as ssu(s)/ elements within hospital(psu). These number of indoor patients are therefore available auxiliary information at elements level.

Therefore, an attempt has been made in the present paper to develop calibration estimators of population mean in two-stage stratified random sampling when the auxiliary information is available at the element (ssu) level for the entire population by calibrating integrated design weight (see Särndal, (2017)). The usual estimator of population mean in two-stage stratified random sampling without using auxiliary information is briefly described in section-2. The section-3 dealt with development of calibration estimators as proposed above. Different version of proposed calibration estimators under simple random sampling without replacement (SRSWOR) have been described in section-4. Some simulation studies with real data have been conducted to examine the relative performance of the estimators in section-5. A concluding remark has been given in section-6.

2. 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 $U=\sum_{i=1}^{N_I}U_i$ 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 in the g^{th} stratum.

t_{ygik} = value of y corresponding to k^{th} element of i^{th} psu in the 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 the 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 mean of y 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

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

$$\Delta_{gkl/i}=\pi_{gkl/i}-\pi_{gk/i}\pi_{gl/i},\ \text{with}\ \tilde{\Delta}_{gkl/i}=\frac{\Delta_{gkl/i}}{\pi_{gkl/i}} \tag{1}$$

The objective to estimate is the population mean

$$\bar{t}_y=\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\sum_{k=1}^{N_{gi}}t_{ygik}=\sum_{g=1}^{G}\frac{N_{go}}{N}\bar{t}_{yg}=\sum_{g=1}^{G}\Omega_g\bar{t}_{yg} \tag{2}$$

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

The Horvitz-Thompson estimator (20) of $\bar{t}_{yg}$ is given by

$$\hat{\bar{t}}_{yg(HT)} = \frac{1}{N_g \bar{N}_{go}} \sum_{i=1}^{n_g} \sum_{k=1}^{n_i} a_{gi} a_{gk/i} t_{ygik} = \frac{1}{N_g \bar{N}_{go}} \sum_{i=1}^{n_g} a_{gi} \hat{t}_{ygi(HT)}$$

$$= \frac{1}{N_g} \sum_{i=1}^{n_g} a_{gi} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi(HT)}$$

$$= \frac{\hat{t}_{yg(HT)}}{N_g \bar{N}_{go}} \tag{3}$$

where $\hat{t}_{yg(HT)} = \sum_{i=1}^{n_g} a_{gi} \hat{t}_{ygi(HT)}$ and $\hat{\bar{t}}_{ygi(HT)} = \frac{1}{N_{gi}} \sum_{i=1}^{n_i} a_{gk/i} t_{ygik}$ are the Horvitz-Thompson estimator of t_{yg} and $\bar{t}_{ygi}$ respectively, $a_{gi} = \frac{1}{\pi_{gi}}$ and $a_{gk/i} = \frac{1}{\pi_{gk/i}}$.

The variance of $\hat{\bar{t}}_{yg(HT)}$ can be written as sum of two components as per Särndal *et al.* (14)

$$V(\hat{\bar{t}}_{yg(HT)}) = \frac{V_{psu} + V_{ssu}}{N_g^2 \bar{N}_{go}^2} \tag{4}$$

where

$V_{psu} = \sum\sum_{U_g} \Delta_{gij} \frac{t_{ygi}}{\pi_{gi}} \frac{t_{ygj}}{\pi_{gj}}$, $\quad V_{ssu} = \sum_{U_g} \frac{V_i}{\pi_{gi}}$ and

$$V_i = \sum\sum_{U_{gi}} \Delta_{gkl/i} \frac{t_{ygik}}{\pi_{gk/i}} \frac{t_{ygil}}{\pi_{gl/i}}.$$

The first component V_{psu} is unbiasedly estimated by

$$\hat{V}_{psu} = \sum\sum_{s_g} \tilde{\Delta}_{gij} \frac{\hat{t}_{ygi}}{\pi_{gi}} \frac{\hat{t}_{ygj}}{\pi_{gj}} - \sum_{s_g} \frac{1}{\pi_{gi}} \left(\frac{1}{\pi_{gi}} - 1 \right) \hat{V}_i \tag{5}$$

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

The second component V_{ssu} is unbiasedly estimated by

$$\hat{V}_{ssu} = \sum_{s_g} \frac{\hat{V}_i}{\pi_{gi}^2} \tag{6}$$

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

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

$$= \frac{1}{N_g^2 \bar{N}_{go}^2} \left(\sum\sum_{s_g} \tilde{\Delta}_{gij} \frac{\hat{t}_{ygi}}{\pi_{gi}} \frac{\hat{t}_{ygj}}{\pi_{gj}} + \sum_{s_g} \frac{\hat{V}_i}{\pi_{gi}} \right) \tag{7}$$

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

$$\hat{\bar{t}}_y = \sum_{g=1}^{G} \Omega_g \hat{\bar{t}}_{yg(HT)} \tag{8}$$

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

$V\left(\hat{\bar{t}}_y\right) = \sum_{g=1}^{G} \Omega_g^2 V(\hat{\bar{t}}_{yg(HT)})$, where $V\left(\hat{\bar{t}}_{yg(HT)}\right)$ is given in (4) (9)

and the estimator of variance of $\hat{\bar{t}}_y$ is given by

$\hat{V}\left(\hat{\bar{t}}_y\right) = \sum_{g=1}^{G} \Omega_g^2 \hat{V}(\hat{\bar{t}}_{yg(HT)})$, where $\hat{V}\left(\hat{\bar{t}}_{yg(HT)}\right)$ is given in (7). (10)

Let SRSWOR is denoted as SI. Therefore, the estimator $\hat{\bar{t}}_{yg(SI)}$ under SRSWOR is given by

$$\hat{\bar{t}}_{yg(SI)} = \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi}, \text{ where } \hat{\bar{t}}_{ygi} = \frac{1}{n_i} \sum_{k=1}^{n_i} t_{ygik} \tag{11}$$

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

$$V\left(\hat{\bar{t}}_{yg(SI)}\right) = \frac{N_g - n_g}{n_g N_g} S_{byg}^2 + \frac{1}{n_g N_g} \sum_{i=1}^{N_g} \left(\frac{N_{gi}}{\bar{N}_{go}} \right)^2 \frac{\left(N_{gi} - n_i\right)}{n_i N_{gi}} S_{ygi}^2 \tag{12}$$

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

$$S_{ygi}^2 = \frac{1}{N_{gi} - 1} \sum_{k=1}^{N_{gi}} \left(t_{ygik} - \bar{t}_{ygi} \right)^2$$

The unbiased variance estimator is given by

$$\hat{V}\left(\hat{\bar{t}}_{yg(SI)}\right) = \frac{N_g - n_g}{n_g N_g} s_{byg}^2 + \frac{1}{n_g N_g} \sum_{i=1}^{n_g} \frac{\left(N_{gi} - n_i\right)}{n_i N_{gi}} s_{ygi}^2 \tag{13}$$

where

$s_{byg}^2 = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} \left(\frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi} - \hat{\bar{t}}_{yg} \right)^2$ and

$$s_{ygi}^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left(\frac{N_{gi}}{\bar{N}_{go}} t_{ygik} - \tilde{t}_{ygi} \right)^2, \quad \tilde{t}_{ygi} = \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{N_{gi}}{\bar{N}_{go}} t_{ygik}.$$

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

$\hat{\bar{t}}_{y(SI)} = \sum_{g=1}^{G} \Omega_g \hat{\bar{t}}_{yg(SI)}$, where $\hat{\bar{t}}_{yg(SI)}$ is given in above equation (11). (14)

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

$V\left(\hat{\bar{t}}_{y(SI)}\right)=\sum_{g=1}^{G}\Omega_g^2 V\left(\hat{\bar{t}}_{yg(SI)}\right)$, where $V\left(\hat{\bar{t}}_{yg(SI)}\right)$ is given in above equation (12). (15)

The unbiased variance estimator is given by

$\hat{V}\left(\hat{\bar{t}}_{y(SI)}\right)=\sum_{g=1}^{G}\Omega_g^2 \hat{V}\left(\hat{\bar{t}}_{yg(SI)}\right)$, where $\hat{V}\left(\hat{\bar{t}}_{yg(SI)}\right)$ is given in above equation (13). (16)

3. 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, i.e. 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 (3) is reproduced here

$$\hat{\bar{t}}_{yg(HT)}=\frac{1}{N_g\bar{N}_{go}}\sum_{s_g}\sum_{s_i}a_{gik}t_{ygik} \tag{17}$$

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

$$\hat{\bar{t}}_{yg}^{c}=\frac{1}{N_g\bar{N}_{go}}\sum_{s_g}\sum_{s_i}w_{gik}t_{ygik} \tag{18}$$

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

$\sum_{s_g}\sum_{s_i}\frac{\left(w_{gik}-a_{gik}\right)^2}{q_{gik}a_{gik}}$, where q_{gik} is some scalar quantity

Subject to the constraints

$$\frac{1}{N_g\bar{N}_{go}}\sum_{s_g}\sum_{s_i}w_{gik}t_{xgik}=\bar{t}_{xg} \tag{19}$$

Therefore, minimizing the following function with respect to w_{gik}

$$\varphi\left(w_{gik},\lambda\right)=\sum_{s_g}\sum_{s_i}\frac{\left(w_{gik}-a_{gik}\right)^2}{q_{gik}a_{gik}}-2\lambda\left(\sum_{s_g}\sum_{s_i}w_{gik}t_{xgik}-N_g\bar{N}_{go}\bar{t}_{xg}\right) \tag{20}$$

where, λ is Lagrange multiplier, yields

$$w_{gik}=a_{gik}+\frac{q_{gik}a_{gik}t_{xgik}}{\sum_{s_g}\sum_{s_i}q_{gik}a_{gik}t_{xgik}^2}\left(N_g\bar{N}_{go}\bar{t}_{xg}-\sum_{s_g}\sum_{s_i}a_{gik}t_{xgik}\right) \tag{21}$$

Putting the value of w_{gik} in above equation (18), we get

$$\hat{\bar{t}}_{yg}^{c}=\frac{1}{N_g\bar{N}_{go}}\sum_{s_g}\sum_{s_i}a_{gik}t_{ygik}+\frac{\sum_{s_g}\sum_{s_i}q_{gik}a_{gik}t_{xgik}t_{ygik}}{\sum_{s_g}\sum_{s_i}q_{gik}a_{gik}t_{xgik}^2}\left(\bar{t}_{xg}-\frac{1}{N_g\bar{N}_{go}}\sum_{s_g}\sum_{s_i}a_{gik}t_{xgik}\right)$$

$$=\hat{\bar{t}}_{yg(HT)}+\hat{B}_g\left(\bar{t}_{xg}-\hat{\bar{t}}_{xg(HT)}\right) \tag{22}$$

where $\hat{\bar{t}}_{xg(HT)}=\frac{1}{N_g\bar{N}_{go}}\sum_{s_g}\sum_{s_i}a_{gik}t_{xgik}$, is the Horvitz-Thompson estimator of $\bar{t}_{xg}$ 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

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

$$=\sum_{g=1}^{G}\Omega_g\left[\hat{\bar{t}}_{yg(HT)}+\hat{B}_g\left(\bar{t}_{xg}-\hat{\bar{t}}_{xg(HT)}\right)\right] \tag{23}$$

Under SRSWOR, say SI, the estimator $\hat{\bar{t}}_{y}^{c}$ can be written as

$$\hat{\bar{t}}_{y(SI)}^{c}=\sum_{g=1}^{G}\Omega_g\hat{\bar{t}}_{yg(SI)}=\sum_{g=1}^{G}\Omega_g\left[\hat{\bar{t}}_{yg(SI)}+\hat{B}_{g(SI)}\left(\bar{t}_{xg}-\hat{\bar{t}}_{xg(SI)}\right)\right] \tag{24}$$

where $\hat{\bar{t}}_{yg(SI)}=\frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\hat{\bar{t}}_{ygi}$, $\hat{\bar{t}}_{yg(SI)}=\frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\hat{\bar{t}}_{ygi}$, $\hat{\bar{t}}_{ygi}=\frac{1}{n_i}\sum_{s_i}t_{ygik}$, $\hat{\bar{t}}_{xgi}=\frac{1}{n_i}\sum_{s_i}t_{xgik}$ and

$$\hat{B}_{g(SI)}=\frac{\sum_{s_g}\sum_{s_i}q_{gik}t_{ygik}t_{xgik}}{\sum_{s_g}\sum_{s_i}q_{gik}t_{xgik}^2}.$$

4. EXACT EXPRESSION OF $\hat{t}^c_{yg}$ FOR DIFFERENT CHOICES OF q_{gik} UNDER SRSWOR

The exact expression of estimator $\hat{t}^c_{yg}$ would depend on the choices of q_{gik} in the formula of $\hat{B}_g$. We shall derive the exact expression for $\hat{t}^c_{yg}$ along with variances and estimator of variances under two plausible choice of q_{gik}.

Choice-I For $q_{gik}=1$ and under SRSWOR, we have

$$\hat{t}^{c1}_{y(SI)} = \sum_{g=1}^{G}\Omega_g \hat{t}^{c1}_{yg(SI)}$$

$$= \sum_{g=1}^{G}\Omega_g\left[\frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\hat{t}_{ygi} + \hat{B}_{g1(SI)}\left(\bar{t}_{xg} - \frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\hat{t}_{xgi}\right)\right]$$

$$= \sum_{g=1}^{G}\Omega_g\left[\hat{\bar{t}}_{yg(SI)} + \hat{B}_{g1(SI)}\left(\bar{t}_{xg} - \hat{\bar{t}}_{xg(SI)}\right)\right],$$ where

$$\hat{B}_{g1(SI)} = \frac{\sum_{s_g}\sum_{s_i} N_{gi}t_{xgik}t_{ygik}/n_i}{\sum_{s_g}\sum_{s_i} N_{gi}t^2_{xgik}/n_i} \tag{25}$$

Now, the above calibration estimator can alternatively be written as

$$\hat{t}^{c1}_{y(SI)} = \sum_{g=1}^{G}\Omega_g\left[\frac{1}{n_g}\sum_{s_g}u_{gi}\hat{t}_{ygi} + \hat{B}_{g1(SI)}\left(\bar{t}_{xg} - \frac{1}{n_g}\sum_{s_g}u_{gi}\hat{t}_{xgi}\right)\right],$$

where $u_{gi} = \dfrac{N_{gi}}{\bar{N}_{go}}$ (26)

The approximate variance of $\hat{t}^{c1}_{yg(SI)}$ has been obtained as

$$V\left(\hat{t}^{c1}_{yg(SI)}\right) = \left(\frac{1}{n_g} - \frac{1}{N_g}\right)\left(S'^2_{byg} - 2B_{g1(SI)}S'_{bxyg} + B^2_{g1(SI)}S'^2_{bxg}\right)$$
$$+\frac{1}{n_gN_g}\sum_{i=1}^{N_g}u^2_{gi}\left(\frac{1}{n_i} - \frac{1}{N_{gi}}\right)^2\left(S^2_{ygi} - 2B_{g1(SI)}S_{xygi} + B^2_{g1(SI)}S^2_{xgi}\right) \tag{27}$$

where $S'^2_{byg} = \dfrac{1}{N_g - 1}\sum_{i=1}^{N_g}\left(u_{gi}\bar{t}_{ygi} - \bar{t}_{yg}\right)^2$,

$S^2_{ygi} = \dfrac{1}{N_{gi}-1}\sum_{k=1}^{N_{gi}}\left(t_{ygik} - \bar{t}_{ygi}\right)^2$, $S'^2_{bxg} = \dfrac{1}{N_g-1}\sum_{i=1}^{N_g}\left(u_{gi}\bar{x}_{gi} - \bar{x}_g\right)^2$,

$S^2_{ygi} = \dfrac{1}{N_{gi}-1}\sum_{k=1}^{N_{gi}}\left(t_{xgik} - \bar{t}_{xgi}\right)^2$,

$S'_{bxyg} = \dfrac{1}{N_g-1}\sum_{i=1}^{N_g}\left(u_{gi}\bar{t}_{xgi} - \bar{t}_{xg}\right)\left(u_{gi}\bar{t}_{ygi} - \bar{t}_{yg}\right)$,

$S_{xygi} = \dfrac{1}{N_{gi}-1}\sum_{i=1}^{N_{gi}}\left(t_{xgik} - \bar{t}_{xgi}\right)\left(t_{ygik} - \bar{t}_{ygi}\right)$

The approximate estimator of variance of $\hat{t}^{c1}_{yg(SI)}$ is given by

$$\hat{V}\left(\hat{t}^{c1}_{yg(SI)}\right) = \left[\left(\frac{1}{n_g} - \frac{1}{N_g}\right)\left(s'^2_{byg} - 2\hat{B}_{g1(SI)}s'_{bxyg} + \hat{B}^2_{g1(SI)}s'^2_{bxg}\right)\right.$$
$$\left.+\frac{1}{n_gN_g}\sum_{i=1}^{n_g}u^2_{gi}\left(\frac{1}{n_i} - \frac{1}{N_{gi}}\right)^2\left(s^2_{ygi} - 2\hat{B}_{g1(SI)}s_{xygi} + \hat{B}^2_{g1(SI)}s^2_{xgi}\right)\right] \tag{28}$$

where

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

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

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

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

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

The approximate variance of $\hat{t}^{c1}_{y(SI)}$ is given by

$V\left(\hat{t}^{c1}_{y(SI)}\right) = \sum_{g=1}^{G}\Omega^2_g\left(\hat{t}^{c1}_{yg(SI)}\right)$, where variance of $\hat{t}^{c1}_{yg(SI)}$ is given in the above equation (27). (29)

The approximate estimator of variance of $\hat{t}^{c1}_{y(SI)}$ is given by

$\hat{V}\left(\hat{t}^{c1}_{y(SI)}\right) = \sum_{g=1}^{G}\Omega^2_g\hat{V}\left(\hat{t}^{c1}_{yg(SI)}\right)$, where the variance of $\hat{t}^{c1}_{yg(SI)}$ is given in above equation (28). (30)

Choice-II For $q_{gik} = \dfrac{1}{t_{xgik}}$ and under SRSWOR, we have

$$\hat{t}^{c2}_{y(SI)} = \sum_{g=1}^{G} \Omega_g \hat{t}^{c2}_{yg(SI)}$$

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

where $$\hat{B}_{g2(SI)} = \frac{\frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{t}_{ygi}}{\frac{1}{n_g} \sum_{s_g} \frac{N_{gi}}{\bar{N}_{go}} \hat{t}_{xgi}} \quad (31)$$

Following the procedure given by Sukhatme *et al.* (1984), the approximate variance of $\hat{t}^{c2}_{y(SI)}$ has been derived as

$$V\left(\hat{t}^{c2}_{y(SI)}\right) = \sum_{g=1}^{G} \Omega_g^2 V\left(\hat{t}^{c2}_{yg(SI)}\right)$$

$$= \sum_{g=1}^{G} \Omega_g^2 \left[\left(\frac{1}{n_g} - \frac{1}{N_g} \right) \left(S'^2_{byg} - 2R_g S'_{bxyg} + R_g^2 S'^2_{bxg} \right) + \frac{1}{n_g N_g} \sum_{i=1}^{N_g} u_{gi}^2 \left(\frac{1}{n_i} - \frac{1}{N_{gi}} \right)^2 D_{gi}^2 \right] \quad (32)$$

where $D_{gi}^2 = S_{ygi}^2 - 2R_g S_{xygi} + R_g^2 S_g^2$, $R_g = \frac{\bar{t}_{yg}}{\bar{t}_{xg}}$ and the value of S'^2_{byg}, S'^2_{bxg}, S'_{bxyg}, S^2_{ygi}, S^2_{xgi}, S_{xygi}, u^2_{gi} are given in the equation (27).

The approximate variance of $\hat{t}^{c2}_{y(SI)}$ can alternatively be written as

$$= \sum_{g=1}^{G} \Omega_g^2 \left[\left(\frac{1}{n_g} - \frac{1}{N_g} \right) \frac{1}{N_g - 1} \sum_{i=1}^{N_g} u_{gi}^2 \left(\bar{y}_{gi\bullet} - R_g \bar{x}_{gi\bullet} \right)^2 + \frac{1}{n_g N_g} \sum_{i=1}^{N_g} u_{gi}^2 \left(\frac{1}{n_i} - \frac{1}{N_{gi}} \right)^2 D_{gi}^2 \right] \quad (33)$$

where $\bar{y}_{gi} = \frac{1}{N_{gi}} \sum_{k=1}^{N_{gi}} t_{ygik}$ and $\bar{x}_{gi} = \frac{1}{N_{gi}} \sum_{k=1}^{N_{gi}} t_{xgik}$

The approximate estimator of variance of $\hat{t}^{c2}_{y(SI)}$ has been derived as

$$\hat{V}\left(\hat{t}^{c2}_{y(SI)}\right) = \sum_{g=1}^{G} \Omega_g^2 \hat{V}\left(\hat{t}^{c2}_{yg(SI)}\right)$$

$$= \sum_{g=1}^{G} \Omega_g^2 \left[\left(\frac{1}{n_g} - \frac{1}{N_g} \right) \left(s'^2_{byg} - 2R_{g1} s'_{bxyg} + R_{g1}^2 s'^2_{bxg} \right) + \frac{1}{n_g N_g} \sum_{i=1}^{n_g} u_{gi}^2 \left(\frac{1}{n_i} - \frac{1}{N_{gi}} \right)^2 d_{gi}^2 \right] \quad (34)$$

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

and the value of s'^2_{byg}, s'^2_{bxg}, s'_{bxyg}, s^2_{ygi}, s^2_{xgi}, s_{xygi}, are given in above equations (28).

The approximate estimator of variance of $\hat{t}^{c2}_{y(SI)}$ can alternatively be written as

$$\hat{V}\left(\hat{t}^{c2}_{y(SI)}\right) = \sum_{g=1}^{G} \Omega_g^2 \left[\left(\frac{1}{n_g} - \frac{1}{N_g} \right) \frac{1}{n_g - 1} \sum_{i=1}^{n_g} u_{gi}^2 \left(\bar{y}_{s_i} - R_{g1} \bar{x}_{s_i} \right)^2 + \frac{1}{n_g N_g} \sum_{i=1}^{N_g} u_{gi}^2 \left(\frac{1}{n_i} - \frac{1}{N_{gi}} \right)^2 d_{gi}^2 \right] \quad (35)$$

where $\bar{y}_{si} = \frac{1}{n_i} \sum_{k=1}^{n_i} t_{ygik}$ and $\bar{x}_{si} = \frac{1}{n_i} \sum_{k=1}^{n_i} t_{xgik}$

5. SIMULATION STUDY

A simulation study has been carried out with real data. The data of MU284 given in Appendix- B and C of Särndal *et al.* (1992) have been used. There are 284 municipalities. The data on various variables for each municipality are given in Appendix-B. These municipalities have been classified into 50 psu's of varying size and corresponding psu totals of that variables are given in Annexure-C. We have considered three populations of MU284 data. The description of the population is given in theTable-1.

Table 1. Description of the populations considered for simulation study

Populations	Study variable (Y)	Auxiliary variable (X)
I	**P85:** Population of 1985	**P75:** Population of 1975
II	**REV84:** Real estate values according to 1984 assessment (in millions of kronor)	**ME84:** Number of municipal employees in 1984
III	**RMT85:** Revenues from 1985 municipal taxation (in millions of kronor)	**REV84:** Real estate values according to 1984 assessment (in millions of kronor)

The 50 psu's are stratified into 4 strata considering the psu total for the auxiliary variable X in ascending order. Consequently, the stratum I consists of 13 psu's, stratum II consists of 14 psu's, stratum III consists of

12 psu's and stratum IV consists of 11 psu's. The samples of size 4 psu's were drawn by SRSWOR independently in each stratum. This process has been repeated 300 times independently in each stratum. 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 $\hat{t}_{yg(SI)}$, $\hat{t}^{c1}_{yg(SI)}$ and $\hat{t}^{c2}_{yg(SI)}$ from 300 sub samples in each stratum. Let $\hat{T}_i$ (i=1,2,3) be the estimate of the population mean based on usual estimator $\hat{t}_{y(SI)}$ without using auxiliary information and calibration estimators $\hat{t}^{c1}_{y(SI)}$ and $\hat{t}^{c2}_{y(SI)}$ from 1200 sample estimates of strata. The true means of y for the populations- I, II and III have been computed to be 29.363, 3077.525 and 244.993 respectively. The following two criteria were used for assessing the relative performance of these estimators:

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

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

(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. R-software was used for simulation study.

The coefficient of correlation between y and x for each population using 284 municipal data has been computed. The results are as follows:

Population I: r_{yx} = 0.9984, **Population II:** r_{yx} = 0.9401, **Population III:** r_{yx} = 0.9358

The percent absolute relative bias (%ARB) and the percent relative root mean square error (%RRMSE) of estimators have been computed for each of the population. The results are presented in the Table 2.

Table 2. Percent absolute relative bias (%ARB) and percent relative root mean square error (%RRMSE) of estimators in different populations

Estimators	Population-I		Population-II		Population-III	
	%ARB	%RRMSE	%ARB	%RRMSE	%ARB	%RRMSE
$\hat{t}_{y(SI)}$	*	7.013	*	9.873	*	10.509
$\hat{t}^{c1}_{y(SI)}$	0.345	0.521	1.114	1.371	1.629	1.918
$\hat{t}^{c2}_{y(SI)}$	0.189	0.362	0.916	1.015	1.224	1.703

*** Unbiased estimator.**

It is evident from the results of the Table-2 that calibration approach based estimation of the population mean of y has outperformed the usual estimator without using the auxiliary information in terms of the percent relative root mean square error (%RRMSE) in all the populations. Among the calibration estimators, $\hat{t}^{c2}_{y(SI)}$ was found to be the best as it has lowest % RRMSE of 0.362, 1.015 and 1.703 percent, respectively, in the three populations. The percent relative bias has been found to be within the range of below about 1.63 percent for all the calibrated estimators in all the populations. The results of the Table 2 also reveal interesting findings that the larger is the correlation of coefficient between y and x, the larger is the precision of the calibration estimators in terms of % RRMSE. Therefore, the overall results of the simulation studies show that the calibration approach based estimation of population mean in two-stage stratified random sampling has brought out considerable improvement in the precision of the estimates.

6. CONCLUDING REMARKS

If the auxiliary information is available at the element level, then the calibration approach based estimation can bring out significant improvement in the precision of the estimate of population mean in two-stage stratified random sampling. It may be mentioned that the calibration estimator $\hat{t}^{c2}_{y(SI)}$ has been found better than the calibration estimator $\hat{t}^{c1}_{y(SI)}$.

REFERENCES

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

Estevao, V.M. and Särndal, C.E. (2000). A functional form approach to calibration.*Journal of Official Statistics*, **16**, 379-399.

Estevao, V.M. and Särndal, C.E. (2006). Survey estimates by calibration on complex auxiliary information. *International Statistical Review*, **74**, 127-147.

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

Farrell, P.J. and Singh, S. (2005). Model-assisted higher order calibration of estimators of variance. *Austr.& New Zealand J. Statist.*, **47**(3)**,** 375-383.

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.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, **32**, 133-142.

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

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

Kaustav. Aditya, 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.

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, **10**(3), 550-562.

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.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). Sampling theory of surveys with applications.*Iowa State Univ. Press, Ames, Iowa and Indian Soc. Agric. Statist.*, New Delhi.

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

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.

Singh, S., Horn, S., Chodhury, S. and Yu, F. (1999). Calibration of the estimators of variance. *Austr.& New Zealand J. Statist.*, **41**(2), 199-122.

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

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,* **46**(10), 4932-4942.

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

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,* **71**(3)**,** 217-224

Singh, Dhirendra, Sisodia, B. V. S., Nidhi and Sandeep Pundir (2020). Some calibration estimators for finite population mean in two-stage stratified random sampling. Communication in Statistics-Theory and Methods, Vol.(49), Issue17, 4234-4247

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