Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 71(3) 2017 217–224

A Calibration Approach based Regression and Ratio Type Estimators of Finite Population Mean in Two-stage Stratified Random Sampling

Dhirendra Singh, B.V.S. Sisodia, V.N. Rai and Sandeep Kumar
Narendra Deva University of Agriculture & Technology, Faizabad

Received 26 May 2017; Accepted 27 June 2017

SUMMARY

Most common methods of estimation where the auxiliary information are used ratio, regression and product methods of estimation. Calibration estimation has developed into an important field of research in survey sampling where the auxiliary information plays an important role. In the present paper an efforts has been made to develop calibration estimator in under two-stage stratified random sampling design when auxiliary information relative to variable under study are available at element level for selected primary stage unit (psu) i.e auxiliary information is available at secondary stage unit (ssu) for selected primary stage unit (psu). A simulation study has been conducted to investigate the relative performance of calibration estimator over the usual estimator of the population mean without using auxiliary information in two-stage stratified random sampling.

Keywords: Calibration approach, Calibration estimator, Auxiliary information, Two-stage sampling, Stratified random sampling.

1. INTRODUCTION

The aim of finite population survey sampling is to provide estimate of population parameters such as population mean, population total etc. of the items under study. The use of the auxiliary information is well known in sample surveys to improve the precision of the estimate. Most common methods of estimation where the auxiliary information are used are ratio, regression and product methods of estimation. Calibration estimation has developed into an important field of research in survey sampling where the auxiliary information plays an important role. A method to calibrate the design weights in Horvitz-Thompson (1952) estimator was considered by Deville and Särndal (1992) by making use of auxiliary information. Recently, Mourya *et al.* (2016) and Kaustav *et al.* (2016) have developed calibration based estimator of the population total under two-stage sampling design. Mourya *et al.* (2016) have also developed calibration estimator of population total in cluster sampling. Calibration approach based estimation has been extended to stratified random sampling by Singh *et al.* (1998), Kim *et al.* (2007), Singh and Arnab (2011), Sinha *et al.* (2016) etc. Most of the large scale surveys are conducted under multi-stage stratified random sampling. Recently, Singh *et al.* (2017) have developed calibration estimators of population in two-stage stratified random sampling when auxiliary information is available at primary stage unit (psu) level. In the present paper, an effort has been made to develop calibration estimator under two-stage stratified random sampling design when auxiliary information relative to variable under study are available at secondary stage unit (ssu) for selected psu(s) in section-3. The usual estimator of population mean in two-stage stratified random sampling without using auxiliary information has been described in section-2. Calibration estimators under simple random sampling without replacement (SRSWOR) have been derived in section-4. The properties of the developed estimators are derived in-terms of design based approximate variance and approximate consistent design based estimator of the variance. A simulation study has been conducted to investigate the relative

Corresponding author: B.V.S. Sisodia
E-mail address: bvssisodia@gmail.com

performance of calibration estimator over the usual estimator of the population mean without using auxiliary information in two-stage stratified random sampling 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 $U = \overset{N_I}{\underset{i=1}{U}} U_i$ 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_g 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

$\bar{N}_{go} = \frac{N_{go}}{N_g}$, 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.

$\bar{t}_{ygi} = \frac{1}{N_{gi}} \sum_{k=1}^{N_{gi}} t_{ygik}$, 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.

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

$\tilde{t}_{yg} = \frac{1}{N_g} \sum_{i=1}^{N_g} t_{ygi}$, 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 $P_i(.)$ 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

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

The objective to estimate is the population mean

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

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

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

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

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

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

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

where $\hat{t}_{yg(HT)} = \sum_{i=1}^{n_g}\sum_{k=1}^{n_i} a_{gi}\hat{t}_{ygi(HT)}$ 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 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.* (1992)

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

With

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

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

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)

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)

If the sampling design is simple random sampling without replacement (SRSWOR) denoted as SI, 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

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

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

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

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

$$\tilde{t}_{ygi} = \frac{1}{n_i}\sum_{k=1}^{n_i}\frac{N_{gi}}{\bar{N}_{go}}t_{ygik}.$$

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. DEVELOPMENT OF CALIBRATION ESTIMATOR OF POPULATION MEAN IN TWO-STAGE STRATIFIED RANDOM SAMPLING WHEN THE AUXILIARY INFORMATION IS AVAILABLE AT SSU LEVEL FOR SELECTED PSU

Consider that the auxiliary information x_k related to the study variate y is available at ssu's level for selected psu(s) corresponding to i^{th} psu $\left(i = 1,2,3...n_g\right)$ in g^{th} stratum. Let $t_{ygi} = \sum_{k=1}^{N_{gi}} t_{ygik}$ be the total of y for i^{th} psu $\left(i = 1,2,3,....N_g\right)$, and $t_{yg} = \sum_{i=1}^{N_g} t_{ygi}$ in the g^{th} stratum. Let $t_{xgi} = \sum_{k=1}^{N_{gi}} t_{xgik}$ be the total of x for i^{th} selected psu in g^{th} stratum, and $t_{xg} = \sum_{k=1}^{N_g} t_{xgi}$, where N_{gi} is the number of elements in i^{th} psu in g^{th} stratum, and $N_I = \sum_{g=1}^{G} N_g$, the total numbers of psu's. The objective is to estimate $\bar{t}_y = \frac{1}{N}\sum_{g=1}^{G} t_{yg}$, where $N = \sum_{g=1}^{G} N_{go}$, the total number ssu(s) in the population.

The Horvitz-Thompson estimator of $\bar{t}_{ygi}$ is given by

$$\hat{\bar{t}}_{ygi(HT)} = \frac{1}{N_{gi}} \sum_{s_i} a_{gk/i} t_{ygik} \tag{17}$$

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

$$\hat{\bar{t}}_{yg(HT)} = \frac{1}{N_g \bar{N}_{go}} \sum_{s_g} a_{gi} \sum_{s_i} a_{gk/i} t_{ygik} \tag{18}$$

The variance and variance estimator are given in the (4) and (7)

We want to calibrate $a_{gk/i}$. Let $w_{gk/i}$ be a calibrated weight, and therefore the calibrated estimator of $\bar{t}_{yg}$ is given by

$$\hat{\bar{t}}_{yg}^{c} = \frac{1}{N_g \bar{N}_{gi}} \sum_{s_g} a_{gi} N_{gi} \hat{\bar{t}}_{ygi}^{c}$$

$$= \frac{1}{N_g} \sum_{s_g} a_{gi} \frac{N_{gi}}{\bar{N}_{go}} \hat{\bar{t}}_{ygi}^{c} z \tag{19}$$

where $\hat{\bar{t}}_{ygi}^{c} = \frac{1}{N_{gi}} \sum_{s_i} w_{gk/i} t_{ygik}$ is a calibration estimator of $\bar{t}_{ygi}$; $i \in s_g$.

We find out the $w_{gk/i}$ by minimizing the chi-square distance measure

$$\sum_{s_i} \frac{\left(w_{gk/i} - a_{gk/i}\right)^2}{q_{gk} a_{gk/i}}$$

subject to the constraints

$$\frac{1}{N_{gi}} \sum_{s_i} w_{gk/i} t_{xgik} = \bar{t}_{xgi} \tag{20}$$

Therefore, the following function will be minimized with respect to $w_{gk/i}$

$$\phi\left(w_{gk/i}, \lambda\right) = \sum_{s_i} \frac{\left(w_{gk/i} - a_{gk/i}\right)^2}{q_{gk} a_{gk/i}} - 2\lambda\left(\frac{1}{N_{gi}} \sum_{s_i} w_{gk/i} t_{xgik} - \bar{t}_{xgi}\right) \tag{21}$$

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

$$w_{gk/i} = a_{gk/i} + \frac{q_{gk} a_{gk/i} t_{xgik}}{\frac{1}{N_{gi}} \sum_{s_i} q_{gk} a_{gk/i} t_{xgik}^2} \left(\bar{t}_{xgi} - \frac{1}{N_{gi}} \sum_{s_i} a_{gk/i} t_{xgik}\right) \tag{22}$$

Therefore, calibration estimator of $\bar{t}_{ygi}$ is given by

$$\hat{\bar{t}}^{c}_{ygi} = \hat{\bar{t}}_{ygi(HT)} + \hat{B}_{gi}\left(\bar{t}_{xgi} - \hat{\bar{t}}_{xgi(HT)}\right) \tag{23}$$

where $\hat{\bar{t}}_{ygi(HT)} = \frac{1}{N_{gi}}\sum_{s_i} a_{gk/i} t_{ygik}$ and $\hat{\bar{t}}_{xgi(HT)} = \frac{1}{N_{gi}}\sum_{s_i} a_{gk/i} t_{xgik}$ are the Horvitz-Thompson estimators of $\bar{t}_{ygi}$ and $\bar{t}_{xgi}$ respectively, and $\hat{B}_{gi} = \frac{\sum_{s_i} q_{gk} a_{gk/i} t_{xgik} t_{ygik}}{\sum_{s_i} q_{gk} a_{gk/i} t^2_{xgik}}$.

The calibrated estimator of $\bar{t}_{yg}$ is now defined as

$$\hat{\bar{t}}^{c}_{yg} = \frac{1}{N_g \bar{N}_{go}}\sum_{s_g} a_{gi} \hat{t}^{c}_{ygi}, \text{ where } \hat{t}^{c}_{ygi} = N_{gi}\hat{\bar{t}}^{c}_{ygi}$$

$$= \frac{1}{N_g}\sum_{s_g} a_{gi}\frac{N_{gi}}{\bar{N}_{go}}\left[\hat{\bar{t}}_{ygi(HT)} + \hat{B}_{gi}\left(\bar{t}_{xgi} - \hat{\bar{t}}_{xgi(HT)}\right)\right] \tag{24}$$

Following, Särndal *et al.* (1992), the approximate variance of $\hat{\bar{t}}^{c}_{yg}$ is obtained as

$$V\left(\hat{\bar{t}}^{c}_{yg}\right) = \frac{1}{N_g^2 \bar{N}_{go}^2}\left[\sum\sum_{U_g} \Delta_{gij}\frac{t_{ygi}}{\pi_{gi}}\frac{t_{ygj}}{\pi_{gj}} + \sum_{U_g}\frac{1}{\pi^2_{gi}}\sum\sum_{U_{gi}}\left(-\Delta_{gkl/i}\right)\frac{E_{gk}}{\pi_{gk/i}}\frac{E_{gl}}{\pi_{gl/i}}\right] \tag{25}$$

where $E_{gk} = t_{ygik} - B_{gi} t_{xgik}$, and

$$B_{gi} = \frac{\sum_{k=1}^{N_{gi}} a_{gk/i} q_{gk} t_{ygik} t_{xgik}}{\sum_{k=1}^{N_{gi}} a_{gk/i} q_{gk} t^2_{xgik}}$$

An approximate unbiased estimator of $V\left(\hat{\bar{t}}^{c}_{yg}\right)$ is obtained as

$$\hat{V}\left(\hat{\bar{t}}^{c}_{yg}\right) = \frac{1}{N_g^2 \bar{N}_{go}^2}\left[\sum\sum_{s_g} \tilde{\Delta}_{gij}\frac{\hat{t}^{c}_{ygi}}{\pi_{gi}}\frac{\hat{t}^{c}_{ygj}}{\pi_{gj}} + \frac{1}{2}\sum_{s_g}\frac{1}{\pi_{gi}}\sum\sum_{s_i}\left(-\tilde{\Delta}_{gkl/i}\right)\left(w_{gk/i}e_{gk} - w_{gl/i}e_{gl}\right)^2\right] \tag{26}$$

where $e_{gk} = t_{ygik} - \hat{B}_{gi} t_{xgik}$, and

$$\hat{B}_{gi} = \frac{\sum_{k=1}^{s_i} a_{gk/i} q_{gk} t_{ygik} t_{xgik}}{\sum_{k=1}^{s_i} a_{gk/i} q_{gk} t^2_{xgik}}$$

Now, calibration estimator of $\bar{t}_y$ is given by

$$\hat{\bar{t}}^{c}_{y} = \sum_{g=1}^{G}\Omega_g \hat{\bar{t}}^{c}_{yg}, \text{ where } \Omega_g = \frac{N_{go}}{N} \tag{27}$$

The variance of $\hat{\bar{t}}^{c}_{y}$ are given by

$V\left(\hat{\bar{t}}^{c}_{y}\right) = \sum_{g=1}^{G}\Omega_g^2 V\left(\hat{\bar{t}}^{c}_{yg}\right)$, where $V\left(\hat{\bar{t}}^{c}_{yg}\right)$ is given in equation (25). (28)

The variance estimator of $\hat{\bar{t}}^{c}_{y}$ are given by

$\hat{V}\left(\hat{\bar{t}}^{c}_{y}\right) = \sum_{g=1}^{G}\Omega_g^2 \hat{V}\left(\hat{\bar{t}}^{c}_{yg}\right)$, where $\hat{V}\left(\hat{\bar{t}}^{c}_{yg}\right)$ is given in equation (26). (29)

4. CALIBRATION ESTIMATORS OF POPULATION MEAN IN SRSWOR (SAY, SI), FOR DIFFERENT CHOICE OF q_{gk}

For $q_{gk} = 1$, the estimator in (24) reduces to regression estimator

$$\hat{\bar{t}}^{c-reg}_{yg(SI)} = \frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\hat{\bar{t}}^{c-reg}_{ygi(SI)} \tag{30}$$

where $\hat{B}'_{gi} = \frac{\sum_{s_i} t_{ygik} t_{xgik}}{\sum_{s_i} t^2_{xgik}}$, $\hat{\bar{t}}_{ygi(SI)} = \frac{1}{n_i}\sum_{s_i} t_{ygik}$,

$\hat{\bar{t}}_{xgi(SI)} = \frac{1}{n_i}\sum_{s_i} t_{xgik}$ and $\hat{\bar{t}}^{c-reg}_{ygi(SI)} = \hat{\bar{t}}_{ygi(SI)} + \hat{B}'_{gi}\left(\bar{t}_{xg} - \hat{\bar{t}}_{xgi}\right)$

Evidently, the estimator $\hat{\bar{t}}^{c-reg}_{yg(SI)}$ is a regression-type estimator.

Therefore, a regression-type calibration estimator of $\bar{t}_{y(SI)}$ is given by

$$\hat{\bar{t}}^{c-reg}_{y(SI)} = \sum_{g=1}^{G}\Omega_g \hat{\bar{t}}^{c-reg}_{yg(SI)} \tag{31}$$

For $q_{gk} = \frac{1}{t_{xgik}}$ **,** the estimator in (24) reduces to a ratio-type estimator

$$\hat{\bar{t}}_{yg(SI)}^{c-r} = \frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\frac{\hat{\bar{t}}_{ygi(SI)}}{\hat{\bar{t}}_{xgi(SI)}}\bar{t}_{xgi}$$

$$= \frac{1}{n_g}\sum_{s_g}\frac{N_{gi}}{\bar{N}_{go}}\hat{\bar{t}}_{ygi(SI)}^{c-r},$$

where $\hat{\bar{t}}_{ygi(SI)}^{c-r} = \frac{\hat{\bar{t}}_{ygi(SI)}}{\hat{\bar{t}}_{xgi(SI)}}\bar{t}_{xgi}$ (32)

Similarly, a ratio-type calibration estimator of $\bar{t}_{y(SI)}$ is given by

$$\hat{\bar{t}}_{y(SI)}^{c-r} = \sum_{g=1}^{G}\Omega_g\hat{\bar{t}}_{yg(SI)}^{c-r} \quad (33)$$

The approximate variance and the estimate of variance of $\hat{\bar{t}}_{yg(SI)}^{c-reg}$ and $\hat{\bar{t}}_{y(SI)}^{c-reg}$ for large sample are obtained as

$$V\left(\hat{\bar{t}}_{yg(SI)}^{c-reg}\right) = \left(\frac{1}{n_g}-\frac{1}{N_g}\right)S_{bg}^2 + \frac{1}{n_gN_g}\sum_{i=1}^{N_g}\frac{N_{gi}^2}{\bar{N}_{go}^2}\left(\frac{1}{ni}-\frac{1}{N_i}\right)\left(1-\rho_{gi}^2\right)S_{ygi}^2 \quad (34)$$

where

$$S_{bg}^2 = \frac{1}{N_g-1}\sum_{i=1}^{N_g}\left(\bar{U}_i-\bar{t}_{yg}\right)^2 = \frac{1}{N_g-1}\sum_{i=1}^{N_g}\left(\frac{N_{gi}}{\bar{N}_{go}}-\bar{t}_{yg}\right)^2 \text{ and}$$

$S_{ygi}^2 = \frac{1}{N_{gi}-1}\sum_{k=1}^{N_{gi}}\left(t_{ygik}-\bar{t}_{ygi}\right)^2$ and ρ_{gi} is the correlation coefficient between t_{ygik} and t_{xgik} in the i^{th} psu of g^{th} stratum.

An approximate estimator of variance of $\hat{\bar{t}}_{ygi(SI)}^{c-reg}$ is

$$\hat{V}\left(\hat{\bar{t}}_{yg(SI)}^{c-reg}\right) = \left(\frac{1}{n_g}-\frac{1}{N_g}\right)s_{bg}^2 + \frac{1}{n_gN_g}\sum_{s_g}\frac{N_{gi}^2}{\bar{N}_{go}^2}\left(\frac{1}{ni}-\frac{1}{N_{gi}}\right)\left(1-r_{gi}^2\right)s_{ygi}^2 \quad (35)$$

where

$$s_{bg}^2 = \frac{1}{n_g-1}\sum_{i=1}^{n_g}\left(\bar{u}_i^{reg}-\bar{\bar{u}}^{reg}\right)^2 = \frac{1}{n_g-1}\sum_{i=1}^{n_g}\left(\frac{N_{gi}}{\bar{N}_{go}}\hat{\bar{t}}_{ygi(SI)}^{c-reg}-\bar{\bar{u}}^{reg}\right)^2,$$

where $\bar{\bar{u}}^{reg} = \frac{1}{n_g}\sum_{i=1}^{n_g}\bar{u}_i^{reg}$, $s_{ygi}^2 = \frac{1}{n_i-1}\sum_{k=1}^{n_i}\left(t_{ygik}-\hat{\bar{t}}_{ygi}\right)^2$, $s_{xgi}^2 = \frac{1}{n_i-1}\sum_{k=1}^{n_i}\left(t_{xgik}-\hat{\bar{t}}_{xgi}\right)^2$, $r_{gi} = \frac{s_{yxgi}}{s_{ygi}s_{xgi}}$ the correlation coefficient between t_{ygik} and t_{xgik} computed from observation from sample s_i.

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

$$V\left(\hat{\bar{t}}_{y(SI)}^{c-reg}\right) = \sum_{g=1}^{G}\Omega_g^2V\left(\hat{\bar{t}}_{yg(SI)}^{c-reg}\right) \quad (36)$$

and its approximate variance estimator is given by

$$\hat{V}\left(\hat{\bar{t}}_{y(SI)}^{c-reg}\right) = \sum_{g=1}^{G}\Omega_g^2\hat{V}\left(\hat{\bar{t}}_{yg(SI)}^{c-reg}\right) \quad (37)$$

The approximate variance and the estimate of variance of $\hat{\bar{t}}_{yg(SI)}^{c-r}$ and $\hat{\bar{t}}_{y(SI)}^{c-r}$ for large sample are obtained as

$$V\left(\hat{\bar{t}}_{yg(SI)}^{c-r}\right) = \left(\frac{1}{n_g}-\frac{1}{N_g}\right)S_{bg}^2 + \frac{1}{n_gN_g}\sum_{i=1}^{N_g}\frac{N_{gi}^2}{\bar{N}_{go}^2}\left(\frac{1}{n_i}-\frac{1}{N_{gi}}\right)S_{gi}^2 \quad (38)$$

where

$$S_{gi}^2 = S_{ygi}^2 + R_g^2S_{xgi}^2 - 2R_gS_{yxgi},$$

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

$$S_{xgi}^2 = \frac{1}{N_{gi}-1}\sum_{k=1}^{N_{gi}}\left(t_{xgik}-\bar{t}_{xgi}\right)^2, \; R_g = \frac{\bar{t}_{yg}}{\bar{t}_{xg}},$$

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

An approximate estimator of variance of $\hat{\bar{t}}_{ygi(SI)}^{c-r}$ is

$$\hat{V}\left(\hat{\bar{t}}_{yg(SI)}^{c-r}\right) = \left(\frac{1}{n_g}-\frac{1}{N_g}\right)s_{bg}^2 + \frac{1}{n_gN_g}\sum_{s_g}\frac{N_{gi}^2}{\bar{N}_{go}^2}\left(\frac{1}{n_i}-\frac{1}{N_{gi}}\right)s_{gi}^2$$

where

$$s_{bg}^2 = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}\left(\bar{u}_i^r - \bar{\bar{u}}^r\right)^2 = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}\left(\frac{N_{gi}}{\bar{N}_{go}}\hat{t}_{ygi(SI)}^{c-r} - \bar{\bar{u}}^r\right)^2,$$

$$\bar{\bar{u}}^r = \frac{1}{n_g}\sum_{i=1}^{n_g}\bar{u}_i^r \ , \ s_{ygi}^2 = \frac{1}{n_i - 1}\sum_{k=1}^{n_i}\left(t_{ygik} - \hat{\bar{t}}_{ygi}\right)^2,$$

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

$$s_{yxgi}^2 = \frac{1}{n_i - 1}\sum_{i=1}^{n_i}\left(t_{ygik} - \hat{t}_{ygi}\right)\left(t_{xgik} - \hat{t}_{xgi}\right), \ \hat{R}_g = \frac{\hat{t}_{yx}}{\hat{t}_{xg}}.$$

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

$$V\left(\hat{t}_{y(SI)}^{c-r}\right) = \sum_{g=1}^{G}\Omega_g^2 V\left(\hat{t}_{yg(SI)}^{c-r}\right) \tag{39}$$

and its approximate variance estimator is given by

$$\hat{V}\left(\hat{t}_{y(SI)}^{c-r}\right) = \sum_{g=1}^{G}\Omega_g^2 \hat{V}\left(\hat{t}_{yg(SI)}^{c-r}\right) \tag{40}$$

5. A LIMITED SIMULATION STUDY

A limited simulation study has been carried out with real data. The population MU284 given in Appendix-C of Särndal *et al.* (2003) 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, 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 $\hat{t}_{yg(SI)}$, $\hat{t}_{yg(SI)}^{c-reg}$, $\hat{t}_{yg(SI)}^{c-r}$ from 300 sub samples in each stratum. We compute the values of $\hat{T}_i$ based on usual estimator $\hat{t}_{y(SI)}$ without using auxiliary information in the section 3.2 and calibration estimators $\hat{t}_{y(SI)}^{c-reg}$, $\hat{t}_{y(SI)}^{c-r}$ 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 given Table 1.

Table 1. Percent relative bias (%RB) and percent relative root mean square error (%RRMSE) over usual estimator.

Estimators	%RB	%RRMSE
$\hat{t}_{y(SI)}$	-	7.013
$\hat{t}_{y(SI)}^{c-reg}$	0.397	0.632
$\hat{t}_{y(SI)}^{c-r}$	0.796	0.942

It can be observed from the results of the Table 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 0.6 percent from 7.0 percent when usual estimator without using auxiliary information was applied. Among the calibrated estimators, $\hat{t}_{y(SI)}^{c-reg}$ was found to be the best as it has lowest %RRMSE of 0.632 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 calibration approach of estimating population mean in two-stage stratified random sampling has brought considerable improvement in the precision of the estimates.

6. CONCLUDING REMARK

Finally, it can be concluded that 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 here that if the auxiliary information are available at ssu's level for selected psu level, then the regression type calibration estimator have improved the precision of the estimates in comparison to the estimates obtained by other calibration estimators.

REFERENCES

Aditya, Kaustav, Sud, U.C. and Chandra, H. (2016). Calibration based regression type estimator of the population total under two-stage sampling design. *J. Ind. Soc. Agril. Statist.*, **70(1)**, 19-24.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, **87,** 376–382.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47**, 663-685.

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

Mourya, K.K., Sisodia, B.V.S. and Chandra, H. (2016). Calibration approach for estimating finite population parameter in two-stage sampling, *J. Statist. Theo. Prac.*, (Online, 19 May 2016).

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. Agric. Sci*., **12(2)**, 351-358.

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

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. *Comm. Statist.-Theory. Methods* (Submitted).

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.

Sinha, N., Sisodia, B.V.S., Singh, S. and Singh, Sanjay K. (2016). Calibration approach estimation of mean in stratified sampling and stratified double sampling. *Comm. Statist.-Theory. Methods*, (online).

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

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.