International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
www.chemijournal.com
IJCS 2020; 8(5): 1360-1371

Received: 25-06-2020
Accepted: 06-08-2020

Ajay Kumar Gautam
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture and Technology, Ayodhya, Uttar Pradesh, India

Manoj Kumar Sharma
Department of Statistics, Mathematics and Computer Science, SKN College of Agriculture, Sri Karan Narendra Agriculture University, Jobner, Jaipur, Rajasthan, India

BVS Sisodia
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture and Technology, Ayodhya, Uttar Pradesh, India

Corresponding Author:
Manoj Kumar Sharma
Department of Statistics, Mathematics and Computer Science, SKN College of Agriculture, Sri Karan Narendra Agriculture University, Jobner, Jaipur, Rajasthan, India

Development of calibration estimator of population mean under non-response

Ajay Kumar Gautam, Manoj Kumar Sharma and BVS Sisodia

DOI: https://doi.org/10.22271/chemi.2020.v8.i5s.10491

Abstract
Using the calibration approach, the Hanson and Hurwitz (1946) technique-based estimator is developed for the situation where the information on auxiliary variable is assumed known for the entire sample units. Expressions for the estimator of population mean/total, its variance, and variance estimator were developed. The theoretical results are illustrated with the help of empirical studies. Empirical results showed that proposed calibration approach-based estimator outperforms the Hansen and Hurwitz technique.

Keywords: Calibration approach, Hansen and Hurwitz estimator, non-response, population mean

Introduction

In many human surveys, it generally is not possible to obtain information from all the units in the surveyed population. The problem of non-response persists even after call backs. The estimates obtained from incomplete data may be biased particularly when the respondents differ from the non-respondents. To address the problem of bias, Hansen and Hurwitz (1946) [2] proposed a technique essentially to adjust for non-response. The technique consists of selecting a sample from the population, identifying the non-respondents in the sample and selecting a sub sample of non-respondent. When the auxiliary information on auxiliary variable related to study variate in available, Deville and Särndal (1992) have developed calibration approach based estimator of the finite population mean/total of the study variate by calibrating sampling design weight using certain calibration equation which involve known population mean/total of the Auxiliary variable. Methods of estimation of finite population mean under non-response in sample surveys using auxiliary information have been developed by various research workers in the past which have been explained in the previous chapter.

Raman *et al.* (2013 a, 2013 b) have developed calibration estimator of the finite population total under non-response depending upon the different situation of availability of the auxiliary information. However, before we present a brief account of their work, we will first describe the Hansen and Hurwitz (1946) [2] estimator under general sampling design.

It may be mentioned that the weighting and imputation procedures aim at elimination of bias caused by non-response. However, these procedures are based on certain assumptions on the response mechanism. When these assumptions do not hold good the resulting estimate may be seriously biased. Further, when the non-response is confounded, i.e. the response probability is dependent on the survey character. It becomes difficult to eliminate the bias entirely.

Theoretical developments

Consider that the finite population U=(1,2,….k,…,N) consists of N identifiable sampling units. Consider that a sample s_a of size n_a is drawn by sampling design P(.) from U with first and second order inclusion probabilities π_{ak} and π_{akl}, $K \neq l$ we also define

$$\Delta_{akl} = \pi_{akl} - \pi_{ak}\pi_{al}, \; k \neq l \subset U \qquad \ldots(1.1)$$

Suppose that the responding sample s_{a1} is of size n_{a1} and non-responding sample s_{a2} of size n_{a2}, Such that $n_{a1}+n_{a2}=n_a$ .A subsample s_2 of size n_2 is drawn from s_{a2} by design P($\cdot/s_{a2}$), with first and second order inclusion properties $\pi_{k/s_{a2}}$ and $\pi_{kl/s_{a2}}$ and

$$\Delta_{kl/s_{a2}} = \pi_{kl/s_{a2}} - \pi_{k/s_{a2}}\pi_{l/s_{a2}} \quad k \neq l \in s_{a2} \qquad \text{…(1.2)}$$

Hansen &Hurwitz (HH) estimator for the total $T=\sum_{i=1}^{N} Y_k$ is given by

$$\hat{T}_\pi = \sum_s y_k/\pi_k^*$$, with

$$\pi_k^* = \begin{cases} \pi_{ak} & \text{if } k \in s_{a1} \\ \pi_{ak}\pi_{k/s_{a2}} & \text{if } k \in_{s2} \end{cases} \qquad \text{…(1.3)}$$

The above estimator can also be expressed as

$$\hat{T}_\pi = \sum_{s_{a1}} y_k/\pi_{ak} + \sum_{s_2} y_k/\pi_{ak}\pi_k/s_{a2} \qquad \text{…(1.4)}$$

$$\hat{T}_\pi = \sum_{s_{a1}} d_{ak}y_k + \sum_{s_2} d_k^* y_k \qquad \text{…(1.5)}$$

Where $d_{ak} = \frac{1}{\pi_{ak}}$, $d_k^* = \frac{1}{\pi_{ak}}\frac{1}{\pi_k/s_{a2}}$

with $E_{RD}(./s_a)$ denoting the expectation with respect to (unknown) Random Distribution (RD), given s_a, E_2 denoting the expectation of all possible samples of size n_2 from n_{a2}, with $E_2(\hat{T}_\pi/s_a, s_{a2}) = \hat{T}_\pi/s_a$, and E_1 denoting the expectation of all possible samples of size n_a from N. That is, the estimator $\hat{T}_{\pi 1}$ given by (1.5) is unbiased for population total T.
Under simple random sampling without replacement (SRSWOR, hereafter SI), the estimation is given by

$$\hat{T}_{SI} = N\left(w_{a1}\bar{y}_{n_{a1}} + w_{a2}\bar{y}_{n2}\right) \qquad \text{…(1.6)}$$

where $w_{a1}=n_{a1}n_a^{-1}$, $w_{a2}=n_{a2}n_a^{-1}$, $\bar{y}_{n_{a1}} = n_{a1}^{-1}\sum_{k=1}^{n_{a1}} y_k$, and $\bar{y}_{n_{a2}} = n_2^{-1}\sum_{k=1}^{n_2} y_k$

In this case, the variance of estimator for the population total (1.5) is given by (Särndal *et al.* 1992).

$$V(\hat{T}_\pi) = \sum_{k=1}^{N_1}\sum_{l=1}^{N_1}\Delta_{akl}\frac{y_k}{\pi_{ak}}\frac{y_l}{\pi_{al}} + \sum_{k=1}^{N_2}\sum_{l=1}^{N_2}\Delta_{akl}\frac{y_k}{\pi_{ak}}\frac{y_l}{\pi_{al}} + 2\sum_{k=1}^{N_1}\sum_{l=1}^{N_2}\Delta_{akl}\frac{y_k}{\pi_{ak}}\frac{y_l}{\pi_{al}} + E_1E_{RD}\left\{\sum_{k=1}^{s_{a2}}\sum_{l=1}^{s_{a2}}\Delta_{kl/s_{a2}}\frac{y_k}{\pi_{ak}\,\pi_k/s_{a2}}\frac{y_l}{\pi_{ak}\,\pi_k/s_{a2}}\right\} \qquad \text{…(1.7)}$$

Under SRSWOR, the variance expression reduces to

$$V(\hat{T}_{SI}) = N(f_a-1)S_y^2 + f_aN_2(f_2-1)S_{2y}^2 \qquad \text{…(1.8)}$$

where $f_a = \frac{N}{n_a}$, $f_2 = \frac{n_{a2}}{n_2}$, $S_y^2 = \frac{1}{(N-1)}\sum_{k=1}^{N}(y_k-\bar{Y}_N)^2$, $S_{2y}^2 = \frac{1}{(N-1)}\sum_{k=1}^{N_2}(y_k-\bar{Y}_{N_2})^2$

An unbiased variance estimator of variance (1.8) is

$$\hat{V}(T_{\pi 1})=\sum_{k=1}^{s_{a1}}\sum_{l=1}^{s_{a1}}\frac{\Delta_{akl}}{\pi^*_{kl}}\frac{y_k}{\pi_{ak}}\frac{y_l}{\pi_{al}}+\sum_{k=1}^{s_{a2}}\sum_{l=1}^{s_{a2}}\frac{\Delta_{akl}}{\pi^*_{kl}}\frac{y_k}{\pi_{ak}}\frac{y_l}{\pi_{al}}+2\sum_{k=1}^{s_{a1}}\sum_{l=1}^{s_{a2}}\frac{\Delta_{akl}}{\pi^*_{kl}}\frac{y_k}{\pi_{ak}}\frac{y_l}{\pi_{al}}+\left\{\sum_{k=1}^{s_2}\sum_{l=1}^{s_2}\frac{\Delta_{kl/s_{a2}}}{\pi_{kl/s_{a2}}}\frac{y_k}{\pi_{ak}\pi_{k/s_{a2}}}\frac{y_l}{\pi_{al}\pi_{l/s_{a2}}}\right\} \quad \ldots(1.9)$$

$$\pi^*_k = \begin{Bmatrix} \pi_{ak}\pi_{kl/s_{a2}} & \text{if } k,l\in s_{a2} \\ \pi_{akl}\pi_{k/s_{a2}} & \text{if } k\in s_{a2},\ l\in s_{a1} \\ \pi_{akl}\pi_{l/s_{a2}} & \text{if } k\in s_{a1},\ l\in s_{a2} \\ \pi_{akl} & \text{if } k,l\in s_{a1} \end{Bmatrix}$$

Under SRSWOR, the estimator of variance (1.9) is

$$V(\hat{T}_{SI})=\frac{N(N-1)n_a}{n_a-1}\left[\frac{(N-n_a)}{n_a(N-1)}\left(\overline{G}_w-\bar{y}_w^{-2}\right)+(f_2-1)\frac{w_{a2}}{n_a}s^2_{n2}\right] \quad \ldots(1.10)$$

Where

$$\overline{G}_w=\frac{1}{n_a}\left(\sum_{k=1}^{n_{a1}}y_k^2+\frac{n_{a2}}{n_2}\sum_{k=1}^{n_2}y_k^2\right) \quad \bar{y}_w=\frac{1}{n_a}\left(n_{a1}\bar{y}_{n_{a1}}+n_{a2}\bar{y}_{n2}\right)$$

$$s^2_{n_2}=\frac{1}{(n_2-1)}\left(\sum_{k=1}^{n_2}y_k^2-n_2\bar{y}^2_{n_2}\right)$$

When auxiliary variable x and study variable y is positively correlated, Cochran (1977) suggested the ratio type extension of HH estimator of population total of form

$$\hat{T}^*_r=N\left(\frac{\hat{\bar{y}}^*}{\bar{x}^*}\overline{X}\right)=N\left(r^*\overline{X}\right) \quad \ldots(1.11)$$

where $r^*=\frac{\bar{y}^*}{\bar{x}^*}$, where $\bar{x}^*$ and $\bar{y}^*$ are the HH estimators for population mean $\overline{X}$ and $\overline{Y}$, respectively.

A large sample first-order approximation to the variance of estimator $\hat{T}^*_r$ given in (1.11), obtained by using the Taylor linearization, is given by

$$V(\hat{T}^*_r)=N^2\left\{\left(\frac{1}{n_a}-\frac{1}{N}\right)S_r^2+\frac{N_2}{N}\frac{(f_2-1)}{n_a}S^2_{2r}\right\} \quad \ldots(1.12)$$

Here $S_r^2=S_y^2+R^2S_x^2-2RS_{xy}$, $S^2_{2r}=S^2_{2y}+R^2S^2_{2x}-2RS_{2xy}$, $S_x^2=\frac{1}{(N-1)}\sum_{k=1}^{N}\left(x_k-\overline{X}_N\right)^2$, and $S^2_{2x}=\frac{1}{(N_2-1)}\sum_{k=1}^{N}\left(x_k-\overline{X}_{N_2}\right)^2$, where R is the population ratio of $\overline{Y}$ to $\overline{X}$. Again S_u^2 and S^2_{2u} are respectively, the variance for the whole population and population variance for the stratum of non-respondents of the variable u. Similarly S_{xy} and S_{2xy} are the covariance for the whole population and the population of non-respondents, respectively.

An approximate estimate of variance of ratio estimator $\hat{T}^*_r$ (1.12) is given by

$$\hat{V}(T^*_r)=N^2\left\{\left(\frac{1}{n_a}-\frac{1}{N}\right)s_r^2+w_{a2}\frac{(f_2-1)}{n_a}s^2_{2r}\right\} \quad \ldots(1.13)$$

where $s_r^2=\frac{1}{(n_a-1)}\sum_{k=1}^{n_{a1}}\left(y_k-r^*x_k\right)^2+f_2\sum_{k=1}^{n_2}\left(y_k-r^*x_k\right)^2$ and $s_{2r}^2=\frac{1}{(n_a-1)}\sum_{k=1}^{n_2}\left(y_k-r^*x_k\right)^2$

are the sample estimates of respectively.

Raman *et al.* (2013) proposed a calibration estimator when auxiliary information is available for the sample. In this case, the variance x is known for both the subset s_{a1} and s_{a2} because non- response is assumed absent for the auxiliary variable. Let us define $\hat{X}_{s_{a2}}=\sum_{s_{a2}}d_{ak}x_k$ with $d_{ak}=\frac{1}{n_{ak}}$. The calibration estimator of population total T given by Raman *et al.* (2013) is defined as

$$\hat{T}_{cal}=\sum_{s_{a1}}d_{ak}y_k+\sum_{s_2}w_ky_k \qquad \ldots(1.14)$$

where $w_k=d_{ak}\,d_k/s_{a2}+\left(\hat{X}_{s_{a2}}-\sum_{s_2}x_kd_{ak}\,d_k/s_{a2}\right)\frac{d_{ak}\,d_k/s_{a2}q_kx_k}{\sum_{s_2}d_{ak}\,d_k/s_{a2}q_kx_k^2}$ are the calibrated weights and $d_{ak}\,d_k/s_{a2}$ are the original weights with $d_{k/s_{a2}}=\frac{1}{\pi_k/s_{a2}}$, for $q_k=\frac{1}{x_k}$, the estimator (1.14) as

$$\hat{T}_{cal}=\sum_{s_{a1}}y_kd_{ak}+\frac{\sum_{s_2}y_kd_{ak}d_{k/s_{a2}}}{\sum_{s_2}x_kd_{ak}d_{k/s_{a2}}}\hat{X}_{s_{a2}} \qquad \ldots(1.15)$$

Note that the estimator $\hat{T}_{cal}$ given in (1.15) has from of Hajek type estimator, and therefore it is based. Under SRSWOR, the estimator $\hat{T}_{cal}$ reduces to

$$\hat{T}_{cal,SI}=N\left(w_{a1}\bar{y}_{n_{a1}}+w_{a2}\frac{\bar{y}_{n_2}}{\bar{x}_{n2}}\bar{x}_{n_{a2}}\right) \text{ with } \bar{x}_{n_{a2}}=\frac{1}{n_{a2}}\sum_{k=1}^{n_{a2}}x_k \qquad \ldots(1.16)$$

Raman *et al.* (2013) [3, 4] have considered simple constraint, i.e. $\sum_{s_2}w_kx_k=\hat{X}_{s_{a2}}$ in developing calibration estimator of finite population total. However, we should have another calibration constraint, i.e. $\sum_{s_2}w_k=\sum_{s_2}d_k^*$. Singh & Sedery (2013) have considered calibrated weight w_k proportional to the design weight d_k, i.e.

$$w_k\propto d_k^* \qquad \ldots(1.17)$$

or $w_k=g_kd_k^*$

The leads to some other calibration constraints, i.e.

$$\sum_{s_2}w_k=\sum_{s_2}g_kd_k^* \qquad \ldots(1.18)$$

Motivated by these works, an attempt has been made in the present paper to develop calibration estimators of finite population total of the study variate y depending upon the different situations of availability of information on the auxiliary variable related to y. The properties of the calibration estimators are derived and are discussed. Empirical studies have been conducted to study relative performance of the estimators.

2. Development calibration estimator under non-response when auxiliary information is available only on sampled units.

Recall from equation (1.5) that the HH estimator of T under general sampling design $P(.)$ is given by

$$T_{\pi} = \sum_{s_{a1}} d_{ak} y_k + \sum_{s_2} d_k^* y_k \quad \ldots(2.1)$$

The calibration estimator of t can be expressed as

$$\hat{T}_{cal.(1)} = \sum_{s_{a1}} d_{ak} y_k + \sum_{s_2} W_k y_k \quad \ldots(2.2)$$

where W_k is calibration weight of sampling design weight d_k^* obtained by minimizing the distance measure $\sum_{s_2} \left(W_k - d_k^*\right)^2 \Big/ q_k d_k^*$,subject to the following calibration equations

$$\sum_{s_2} W_k x_k = \hat{X}_{s_{a2}} \quad \ldots(2.3)$$

$$\sum_{s_2} W_k = \sum_{s_2} d_k^* \quad \ldots(2.4)$$

The following function is minimized with regression w_k

$$\varphi = \sum_{s_2} \frac{\left(W_k - d_k^*\right)^2}{q_k d_k^*} + 2\lambda_1 \left(\hat{X}_{s_{a2}} - \sum_{s_2} W_k x_k \right) + 2\lambda_2 \left(\sum_{s_2} d_k^* - \sum_{s_2} W_k \right) \quad \ldots(2.5)$$

Differentiating φ with respect to W_k, and equating it to zero and solving for W_k ,we get

$$W_k = d_k^* + \lambda_1 x_k q_k d_k^* + \lambda_2 q_k d_k^* \quad \ldots(2.6)$$

Substituting w_k from equation (2.6) into equation (2.3) and equation (2.4), we get the following two equations

$$\sum_{s_2} d_k^* x_k + \lambda_1 \sum_{s_2} q_k d_k^* x_k^2 + \lambda_2 \sum_{s_2} q_k d_k^* x_k = \hat{X}_{s_{a2}} \quad \ldots(2.7)$$

$$\sum_{s_2} d_k^* + \lambda_1 \sum_{s_2} q_k d_k^* x_k + \lambda_2 \sum_{s_2} q_k d_k^* = \sum_{s_2} d_k^* \quad \ldots(2.8)$$

From equation (2.8), we get λ_2 as

$$\lambda_2 = -\frac{\lambda_1 \sum_{s_2} q_k d_k^* x_k}{\sum_{s_2} q_k d_k^*} \quad \ldots(2.9)$$

Substituting λ_2 from above into equation (2.7), we get

$$\sum_{s_2} d_k^* x_k + \lambda_1 \sum_{s_2} q_k d_k^* x_k^2 - \lambda_1 \frac{\left(\sum_{s_2} q_k d_k^* x_k\right)^2}{\sum_{s_2} q_k d_k^*} = \hat{X}_{s_{a2}}$$

or

$$\lambda_1 = \frac{\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right)}{\sum_{s_2} q_k d_k^* x_k^2 - \left(\sum_{s_2} q_k d_k^* x_k\right)^2 \Big/ \sum_{s_2} q_k d_k^*} \qquad \ldots(2.10)$$

From equation (2.9) and equation (2.10), we get

$$\lambda_2 = -\frac{\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right)\sum_{s_2} q_k d_k^* x_k}{A\sum_{s_2} q_k d_k^*} \qquad \ldots(2.11)$$

where

$$A = \sum_{s_2} q_k d_k^* x_k^2 - \frac{\left(\sum_{s_2} q_k d_k^* x_k\right)^2}{\sum_{s_2} q_k d_k^*}$$

From equation (2.6), (2.10) and (2.11), we get the calibration weight W_k as follows

$$W_k = d_k^* + \frac{\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right) x_k q_k d_k^*}{A} - \frac{\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right)\sum_{s_2} q_k d_k^* x_k}{A\sum_{s_2} q_k d_{ki}^*} q_k d_k^*$$

$$W_k = d_k^* + \left[x_k q_k d_k^* - \frac{q_k d_k^* \sum_{s_2} q_k d_k^* x_k}{\sum_{s_2} q_k d_k^*}\right]\frac{\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right)}{A} \qquad \ldots(2.12)$$

Now, calibration estimator given in (2.2) can be expressed as

$$\hat{T}_{cal_{(1)}} = \sum_{s_{a1}} d_{ak} y_k + \sum_{s_2} W_k y_k$$

$$\hat{T}_{cal_{(1)}} = \sum_{s_{a1}} d_{ak} y_k + \sum_{s_2} d_k^* y_k + \sum_{s_2}\left[q_k d_k^* x_k y_k - \frac{q_k d_k^* y_k \sum_{s_2} q_k d_k^* x_k}{\sum_{s_2} q_k d_k^*}\right]\frac{\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right)}{A}$$

$$\hat{T}_{cal_{(1)}} = \sum_{s_{a1}} d_{ak} y_k + \sum_{s_2} d_k^* y_k + \hat{B}\left(\hat{X}_{s_{a2}} - \sum_{s_2} d_k^* x_k\right) \qquad \ldots(2.13)$$

$$\hat{B}=\frac{\sum_{s_2} q_k d_k^* x_k y_k-\frac{\left(\sum_{s_2} q_k d_k^* x_k\right)\left(\sum_{s_2} q_k d_k^* y_k\right)}{\sum_{s_2} q_k d_k^*}}{\sum_{s_2} q_k d_k^* x_k^2-\frac{\left(\sum_{s_2} q_k d_k^* x_k\right)^2}{\sum_{s_2} q_k d_k^*}}$$

where

If the sampling design is sample random sampling without replacement (SRSWOR, saySI), than and

$$d_{ak}=\frac{1}{\pi_{ak}}=\frac{N}{n_a}, \quad d_k^*=d_{ak}\, d_k / s_{a2}=\frac{1}{\pi_{ak}}\frac{1}{\pi_{ak}/s_{a2}}=\frac{N}{n_a}\frac{n_{a2}}{n_2}$$

For $q_k=1$, the calibration estimator $\hat{T}_{cal}$ under SRSWOR is deduced to

$$\hat{T}_{cal(1)(SI)}=N\left[\frac{n_{a1}\bar{y}_{s_{a1}}+n_{a2}\bar{y}_{s_2}}{n_a}\right]+\frac{\sum_{s_2} y_k x_k-\sum_{s_2} y_k \sum_{s_2} x_k / n_2}{\sum_{s_2} x_k^2-\left(\sum_{s_2} x_k\right)^2 / n_2}\left(\hat{X}_{s_{a2}}-\sum_{s_2} d_k^* x_k\right)$$

$$=N\left[\left(W_1\bar{y}_{s_{a1}}+W_2\bar{y}_{s_2}\right)+W_2\hat{\beta}_2\left(\bar{x}_{s_{a2}}-\bar{x}_{s_2}\right)\right] \qquad \ldots(2.14)$$

where $\bar{y}_{s_{a1}}=\frac{1}{n_{a1}}\sum_{s_{a1}} y_k \quad \bar{y}_{s_2}=\frac{1}{n_2}\sum_{s_2} y_k$

$$\bar{x}_{s_{a2}}=\frac{1}{n_{a2}}\sum_{s_{a2}} x_k \quad \bar{x}_{s_2}=\frac{1}{n_2}\sum_{s_2} x_k$$

$W_1=\frac{n_{a1}}{n_a}$, $W_2=\frac{n_{a2}}{n_a}$ and $\hat{\beta}_2=\frac{\sum_{s_2} y_k x_k-\sum_{s_2} y_k \sum_{s_2} x_k / n_2}{\sum_{s_2} x_k^2-\left(\sum_{s_2} x_k\right)^2 / n_2}$

Infect, $\hat{\beta}_2$ is estimator of population regression coefficient of Y and X in non-response class based on sub- sample s_2 from square s_{a2}.

The estimator $T_{cal(1)(SI)}$ in equation (2.14) can also be expressed as

$$\hat{T}_{cal(1)(SI)}=N\bar{y}^*+NW_2\hat{\beta}_2\left(\bar{x}_{s_{a2}}-\bar{x}_{s_2}\right) \qquad \ldots(2.15)$$

The $\bar{y}^*$ is the usual unbiased estimator of the population mean under non-response due to Hansen & Hurwitz (1946) [2].

The second part of the estimator $\hat{T}_{cal(1)(SI)}$ make its biased estimator of population total. In the following sub-section, we derive the bias and mean square error (MSE) of $\hat{T}_{cal(1)(SI)}$.

4.2.1 Bias and MSE of $\hat{T}_{cal(1)(SI)}$

The bias of $\hat{T}_{cal(1)(SI)}$ is given by

$$B\left[\hat{T}_{cal(1)(SI)}\right] = E\left[\hat{T}_{cal(1)(SI)}\right] - T$$

$$= NE_{RD}\left(W_2\right)\left[E_2\left(\frac{s_{yx(2)}}{s^2_{x_2}}\right)\left(\bar{x}_{s_{a_2}} - \bar{x}_{s_2}\right)\right] \quad \ldots(4.2.1.1)$$

where, $s_{xy_{(2)}} = \frac{1}{n_2 - 1}\sum_{s_2}\left(x_k - \bar{x}_{s_2}\right)\left(y_k - \bar{y}_{s_2}\right)$, $s^2_{x_2} = \frac{1}{n_2 - 1}\sum_{s_2}\left(x_k - \bar{x}_{s_2}\right)^2$

Now, we first derive the second part of the expectation of (2.1.1).
Then, we have
Assume that

$$\bar{x}_{s_2} = \bar{X}_2 + \varepsilon_1$$
$$\bar{x}_{s_{a2}} = \bar{X}_2 + \varepsilon_2$$
$$s_{yx(2)} = S_{yx_2} + \eta_1$$
$$s^2_{x_2} = S^2_{x_2} + \eta_2$$

Then, we have

$$E_2\left(\frac{s_{yx(2)}}{s^2_{x_2}}\right)\left(\bar{x}_{s_{a_2}} - \bar{x}_{s_2}\right) = E_2\left[\frac{S_{yx_2}}{S^2_{x_2}}\left(1 + \frac{\eta_1}{S_{yx_2}}\right)\left(1 + \frac{\eta_2}{S^2_{x_2}}\right)^{-1}\left(\varepsilon_2 - \varepsilon_1\right)\right]$$

$$= \beta_2 E_2\left[\left(1 - \frac{\eta_2}{S^2_{x_2}} + \frac{\eta_2^2}{S^4_{x_2}} - \frac{\eta_1\eta_2}{S_{yx_2}S^2_{x_2}} + \frac{\eta_1}{S_{yx_2}} + \frac{\eta_1\eta_2^2}{S_{yx_2}S^4_{x_2}}\right)\left(\varepsilon_2 - \varepsilon_1\right)\right]$$

Ignoring the higher term of order η_1^2, η_2^2, and their product more than 2. we get

$$E_2\left[\left(\frac{s_{yx(2)}}{s^2_{x_2}}\right)\left(\bar{x}_{s_{a_2}} - \bar{x}_{s_2}\right)\right] = \beta_2 E_2\left[\left(1 - \frac{\eta_2}{S^2_{x_2}} + \frac{\eta_1}{S_{yx_2}}\right)\left(\varepsilon_2 - \varepsilon_1\right)\right]$$

$$= \beta_2\left[\frac{1}{S^2_{x_2}}\left[E_2\left(\varepsilon_1\eta_2\right) - E_2\left(\varepsilon_2\eta_2\right)\right] + \frac{1}{S_{yx_2}}\left[E_2\left(\varepsilon_2\eta_1\right) - E\left(\varepsilon_1 \cdot \eta_1\right)\right]\right] \quad \ldots(2.1.2)$$

$$E\left(\varepsilon_1\eta_2\right) = \mathrm{cov}\left(\bar{x}_{s_2}, s^2_{x_2}\right)$$
$$= \left(\frac{h-1}{N_2} + \frac{k-1}{n_{a_2}}\right)\mu_{30(2)}, \quad \ldots(2.1.3)$$

where $h = \frac{N_2}{n_{a2}}$, $k = \frac{n_{a2}}{n_2}$ and $\mu_{rv_{(2)}} = \frac{1}{N_{2-1}}\sum_{k=1}^{N_2}\left(x_k - \bar{X}_2\right)^r\left(y_k - \bar{Y}_2\right)^v$

$$E(\varepsilon_2\eta_2) = \mathrm{cov}\left(\bar{x}_{s_{a2}}, s^2_{x_2}\right)$$

$$= \frac{h-1}{N_2} \mu_{30(2)} \quad \ldots(2.1.4)$$

$$E(\varepsilon_2\eta_1) = \text{cov}\left[\bar{x}_{s_{a2}}, s_{yx(2)}\right]$$
$$= \frac{h-1}{N_2} \mu_{21(2)} \quad \ldots(2.1.5)$$

$$E(\varepsilon_1\eta_1) = \text{cov}\left[\bar{x}_{s2}, s_{yx2}\right]$$
$$= \left(\frac{h-1}{N_2} + \frac{k-1}{n_{a2}}\right)\mu_{21} \quad \ldots(2.1.6)$$

Now, we have from the equation (2.1.3), (2.1.4), (2.1.5), and (2.1.6), putting these values in equation (2.1.2). We get

$$E_2\left[\hat{\beta}_2\left(\bar{x}_{s_{a2}} - \bar{x}_{s_2}\right)\right] = \frac{k-1}{n_{a2}} \beta_2 \left[\frac{\mu_{30(2)}}{\mu_{20(2)}} - \frac{\mu_{21(2)}}{\mu_{11(2)}}\right] \quad \ldots(2.1.7)$$

Now, from the equation (2.1.7) putting these value in equation (2.1.2), we get

$$B\left[\hat{T}_{cal(1)(SI)}\right] = NE_{RD}\left(W_2\right)\frac{k-1}{n_{a2}} \beta_2 \left[\frac{\mu_{30(2)}}{\mu_{20(2)}} - \frac{\mu_{21(2)}}{\mu_{11(2)}}\right]$$
$$B\left[\hat{T}_{cal(1)(SI)}\right] = N_2 \frac{k-1}{n_{a_2}} \beta_2 \left[\frac{\mu_{30(2)}}{\mu_{20(2)}} - \frac{\mu_{21(2)}}{\mu_{11(2)}}\right] \quad \ldots(2.1.8)$$

where, $E_{RD}\left(\frac{n_{a2}}{n_a}\right) = \frac{N_2}{N}$

Mean square error (MSE) of $\hat{T}_{cal(SI)}$ is derived as fallows

$$MSE\left[\hat{T}_{cal(1)(SI)}\right] = E\left[\hat{T}_{cal(1)(SI)} - T\right]^2$$
$$= E\left[N\left(\bar{y}_W - \bar{Y}\right) + NW_2\hat{\beta}_2\left(\bar{x}_{s_{a_2}} - \bar{x}_{s_2}\right)\right]^2$$

Where, $T = N\bar{Y}$, $\bar{Y}$ = the population mean of Y.

$$= N^2\left[V\left(\bar{y}_W\right) + \beta_2^2 E\left(W_2^2\left(\varepsilon_2^2 - \varepsilon_1^2 - 2\varepsilon_1\varepsilon_2\right)\right) + 2\beta_2 E\left(W_2\left(\bar{y}_W - \bar{Y}\right)\left(\varepsilon_2 - \varepsilon_1\right)\right)\right] \quad \ldots(2.1.9)$$

Now, we drive the first various expectations as fallows.

$$E\left(\varepsilon_1^2\right) = V\left(\bar{x}_{s_2}\right) = V\left[E\left(\bar{x}_{s_2}/s_{a2}\right)\right] + E\left[V\left(\bar{x}_{s_2}/s_{a2}\right)\right]$$

$$= \left(\frac{h-1}{N_2} + \frac{k-1}{n_{a_2}}\right) S_{x_2}^2 \quad \ldots(2.1.10)$$

$$E\left(\varepsilon_2^2\right) = V\left(\bar{x}_{s_{a2}}\right) = V\left[E\left(s_{a2}/\bar{x}_{s_a}\right)\right] + E\left[V\left(\bar{x}_{s_{a2}}/s_{a2}\right)\right]$$

$$= \frac{N_2 - n_{a2}}{N_2 n_{a2}} S_{x_2}^2$$
$$= \frac{h-1}{N_2} S_{x_2}^2$$
…(2.1.11)

$$E(\varepsilon_1 \varepsilon_2) = cov(\bar{x}_{s_{a2}} \bar{x}_{s_2}) = cov[E(\bar{x}_{s_{a2}} / s_{a2}), E(\bar{x}_{s_2} / s_{a2})] + E(cov[\bar{x}_{s_{a2}}, \bar{x}_{s_{a2}} / s_{a2}])$$
$$= cov[\bar{x}_{s_{a2}}, \bar{x}_{s_{a2}}]$$
$$= V(\bar{x}_{s_{a2}})$$
$$= \frac{h-1}{N_2} S_{x_2}^2$$
…(2.1.12)

Substituting the value of $E(\varepsilon_1^2)$, $E(\varepsilon_1^2)$ and $E(\varepsilon_1 \varepsilon_2)$ from the equation (2.1.10), (2.1.11) and (2.1.12) in to the equation (2.1.9). we have

$$MSE[\hat{T}_{cal(1)(SI)}] = N^2 \left[V(\bar{y}_W) + \beta_2^2 E_{RD}(W_2^2) \left(\frac{k-1}{n_{a2}} S_{x_2}^2 \right) + 2\beta_2 E_{RD}(W_2) E_2 (\bar{y}_w - \bar{Y})(\varepsilon_2 - \varepsilon_1) \right]$$
…(2.1.13)

We now drive the expectation of third term in equation (2.1.13) as fallows

$$E_2(\bar{y}_w - \bar{Y})(\varepsilon_2 - \varepsilon_1) = E_2[(W_1 \bar{y}_{s_{a1}} + W_2 \bar{y}_{s_2} - \bar{Y})(\varepsilon_2 - \varepsilon_1)]$$

and get

$$= W_2 E_2 \bar{y}_{s_2} (\varepsilon_2 - \varepsilon_1)$$
…(2.1.14)

Now, we derive the expectation of equation (2.1.14).
Assume that

$$\bar{y}_{s_2} = \left(1 + \frac{\varepsilon_0}{\bar{Y}_2}\right) \bar{Y}_2$$
$$= \bar{Y}_2 E(W_2) E\left[\frac{\varepsilon_0 \varepsilon_2}{\bar{Y}_2} - \frac{\varepsilon_0 \varepsilon_1}{\bar{Y}_2} \right]$$
…(2.1.15)

$$E(\varepsilon_0 \varepsilon_2) = cov(\bar{y}_{s_2} \bar{x}_{s_{a2}})$$
$$= \frac{h-1}{N_2} S_{xy_2}$$
…(2.1.16)

$$E(\varepsilon_0 \varepsilon_1) = cov(\bar{y}_{s_2}, \bar{x}_{s_2})$$
$$= \left[\frac{h-1}{N_2} + \frac{k-1}{n_{a2}} \right] S_{xy}$$
…(2.1.17)

Now, we have from (2.1.15), (2.1.16) and (2.1.17) putting these values in equation (2.1.14). we get

$$EW_2 E(\bar{y}_w - \bar{Y})(\varepsilon_2 - \varepsilon_1) = \bar{Y}_2 E(W_2) \left[\frac{h-1}{N_2 \bar{Y}_2} - \frac{h-1}{\bar{Y}_2 N_2} - \frac{k-1}{n_{a2} \bar{Y}_2} \right] S_{xy_2}$$
$$= -E(W_2) \frac{k-1}{n_{a2}} S_{xy_2}$$

$$E\left(W_2^2\right) = E\left(\frac{n_{a2}^2}{n_a^2}\right)\frac{1}{n_{a2}}$$

$$= E\left(\frac{n_{a2}}{n_a}\right)\frac{1}{n_a}$$

$$= \frac{N_2}{N}\frac{k-1}{n_a}S_{x_2}^2 \qquad \ldots(2.1.18)$$

$$E\left(W_2\right) = E\frac{n_{a2}^2}{n_a^2}\frac{k-1}{n_{a2}}S_{xy_2}$$

$$= E\left(\frac{n_{a2}}{n_a}\right)\frac{k-1}{n_a}S_{xy_2}$$

$$= \frac{N_2}{N}\frac{k-1}{n_a}S_{xy_2} \qquad \ldots(2.1.19)$$

$$\beta_2^2 = \left(\frac{S_{xy_2}}{S_{x_2}^2}\right)^2$$

$$= \frac{\rho_2^2 S_{x_2}^2 S_{y_2}^2}{S_{x_2}^2}$$

$$= \rho_2^2 S_{y_2}^2 \qquad \ldots(2.1.20)$$

From the equation of (2.1.18) , (2.1.19) and (2.1.20), we get

$$MSE\left[\hat{T}_{cal(1)(SI)}\right] = N^2\left[V(\bar{y}_W) + \beta_2^2 E_{RD}\left(W_2^2\right)\left(\frac{k-1}{n_{a2}}S_{x_2}^2\right) - 2\beta_2 E_{RD}\left(W_2^2\right)\frac{k-1}{n_{a2}}S_{xy_2}\right]$$

$$V\left[\hat{T}_{cal(1)(SI)}\right] = N^2\left[\frac{N - n_a}{Nn_a}S_y^2 + \frac{k-1}{n_a}W_2\left[1-\rho_2^2\right]S_{y_2}^2\right] \qquad \ldots(2.1.21)$$

4.3 An empirical study

An empirical study has been conducted with real data to examine the relative efficiency of the proposed and existing estimators in comparison to Hansen & Hurwitz (1946) [2] estimator. We have considered the municipality level data of two populations from the Appendix B (MU284 population) of Sarndal *et al.* (2003), the details of which are given below.

Population	Study variate	Auxiliary variable
(I)	RMT85 Revenues from the 1985 municipal taxation (in millions of kronor).	REV 84 Real estate values according to 1984 assessment (in millions of kronor).

The population consists of N=284 municipalities. It is assumed that the last 25 percent municipalities fall under non-response. That means response class consists of first 213 (N_1) municipalities and non-response class consist of last 71 (N_2) municipalities. The various population parameters of the both populations have been computed and are presented in the Table 1. It is also considered that a sample size of n=57 was drawn by SRSWOR from the population of size N=284. It is also assumed that 42 units respondent while 15 units did not respond while conducting the surveys. A sub –sample of size h_2 =9 from non-respondent units (15) was drawn by SRSWOR, and efforts were made to get response from there 9 units.

Here $k = \frac{n_2}{h_2} = 1.67$, i.e. k-1=0.67 and $f = \frac{n}{N} = 0.20$.

We have considered the some real population -1 as described below. We have also considered the following value of the sample size in reference to the Table 1 of the below

$n_a = 57$, $n_{a1} = 42$, $n_{a2} = 15$ and $n_2 = 9$

The variance of HH estimator of the population total and MSE of calibration estimator $\hat{T}_{cal(1)(SI)}$ using auxiliary information on only sample units belonging to non-response class have been computed for the population. The relative efficiency of the $\hat{T}_{cal(1)(SI)}$ over HH estimator has also been computed to the 115.22 and percent for the population -1.

Table 1: Results of population parameters

Population parameters	Population
S_y^2	355612.50
$S_{y_2}^2$	31736.44
S_x^2	22521693.33
$S_{x_2}^2$	5698777.027
S_{xy}	2648372.57
S_{xy2}	360961.938
ρ	0.936
ρ_2	0.848
$\beta_{2(x)}$	85.0496
β	0.117592
β_2	0.063340
C_x	1.54
C_{x_2}	0.9131
C_y	2.433
C_{y_2}	1.0727
$\overline{X}$	3074.355
$\overline{Y}$	245.088

This shows that if partial auxiliary information is available, then calibration approach can be applied to get more precise estimate of population mean/ total as compared to HH estimator.

References

1. Aditya Kaustav, Sud UC, Chandra H. Calibration based regression type estimator of the population total under two-stage sampling design. Journal of the Indian Society of Agricultural Statistics. 2016; 70(1):19-24.
2. Hansen MH, Hurwitz WN. The problem of non-response in sample surveys. Journal of the American Statistical Association. 1946; 41:517-529.
3. Raman *et al*. Calibration approach for estimating population total with subsampling of non respondents under single- and two-phase sampling_ Communications in Statistics - Theory and Methods, 2016, 45(10).
4. Raman RK, Sud UC, Chandra H. Calibration approach for estimating population total with sub-sampling of non-respondents under single and two-phase sampling. Communication in statistics theory & method, (Accepted), 2013.
5. Rao PSRS. Ratio estimation with sub sampling the non-respondents. Survey Methodology. 1986; 12:217-230.
6. Rao TJ, Sengupta S. Calibration linear unbiased estimators in finite population sampling. Journal of Statistical Planning and Inference. 2010; 140:652-658.
7. Mourya KK, Sisodia BVS, Chandra H. Calibration approach for estimating finite population parameter in two-stage sampling, Journal of Statistical Theory and Practice, (Online, 19 May 2016).
8. Mourya KK, Sisodia BVS, Kumar Sunil, Singh Amar. Calibrated estimators of finite population parameters in equal cluster sampling. Int. J Agricult. Sci. 2016; 12:351-358.