Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 77(2) 2023 171–176

# Product Type Calibration Estimator with inversely related PSU level auxiliary variable under Two Stage Sampling

**Ankur Biswas, Kaustav Aditya and U.C. Sud**

*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

*Received 23 June 2021; Revised 24 January 2023; Accepted 04 March 2023*

## SUMMARY

In survey sampling, auxiliary information is often used to increase the efficiency of estimators of finite population parameters. The Calibration Approach is one of the popular techniques for such purposes. In this current study, product type calibration estimator of the finite population total has been proposed following well-known Calibration Approach for the situations of availability of inversely related auxiliary information at PSU level under two stage sampling design framework. Statistical properties of proposed product type calibration estimator of population total were studied through a simulation study. The simulation results suggest that the proposed product type calibration estimator is performing better than usual Horvitz-Thompson and linear regression estimators of the population total under two stage sampling design.

*Keywords:* Auxiliary information; Calibration; Design weights; Product estimator; Simulation.

## 1. INTRODUCTION

Calibration (Deville and Särndal (1992) is one of the popular techniques in surveys to efficiently incorporate ancillary information to improve the efficiency of the estimators of population parameter e.g. population total, mean, proportions etc. During past two decades or so, calibration approach became an important topic in survey research and vast literature has been dedicated to it. It has gained substantial attention not only in the area of survey methodology, but also in survey practice. Following Deville and Särndal (1992), a lot work has been carried out in calibration estimation i.e. Singh *et al.* (1998, 1999), Wu and Sitter (2001), Sitter and Wu (2002), Kott (2006), Estevao and Sarndal (2006), Kim and Park (2010), Särndal (2007) etc. In recent past decent amount of research has been carried in extension of the calibration estimation under more complex survey designs like stratified sampling, two stage sampling, double sampling, stratified double sampling as per availability of complex auxiliary information (Tracy *et al.*, 2003; Aditya *et al.*, 2016a, 2016b; Mourya *et al.*, 2016; Sinha *et al.*, 2016; Aditya *et al.*, 2017; Basak *et al.*, 2017; Salinas *et al.*, 2018; Biswas *et al.*, 2020).

It was observed that most literature in Calibration Approach for estimation of population parameters were limited to linear relationship assumption between the study and the auxiliary variable. In contrary, there may be situations where the characteristics under study is inversely related to auxiliary variable and, in that situation, the earlier developed methodologies for calibration estimation does not fit in as it is. As for example, in household surveys, marketable surplus is inversely related to family consumption for seed, feed, waste etc. Sud *et al.* (2014a, 2014b) extended the calibration approach when variable of interest and auxiliary information have inverse relation under uni-stage equal probability sampling design. But, multi-stage designs are most prevalent in medium to large scale surveys. In case of two stage sampling, auxiliary information may be existing for the primary stage units (PSU) as well as the secondary stage units (SSU) within the PSUs (Särndal *et al.*, 1992). Biswas *et al.* (2020) developed product type calibration estimators of finite

*Corresponding author*: Ankur Biswas
*E-mail address*: ankur.biswas@icar.gov.in

population total of study variable inversely related with available auxiliary information at SSU level under two stage sampling. But, according to Särndal *et al.* (1992), there may be situation where auxiliary information is existing for all population level PSUs. As for example, for national surveys for certain establishments, say hospitals or other administrative offices, under each tehsil in which tehsils are considered as PSU (Särndal *et al.*, 1992). Information on number of hospitals in each tehsil (PSU) can be easily obtained and used as auxiliary variable at PSU level. In this current study, attempt has been made to propose calibration methodology for estimation of finite population total when information on auxiliary variable is available at the PSU level under two stage sampling design and it is inversely related to the study variable.

## 2. PROPOSED CALIBRATION ESTIMATOR BASED ON INVERSELY RELATED PSU LEVEL AUXILIARY VARIABLE UNDER TWO STAGE SAMPLING DESIGN

In this study, under two stage sampling design framework, calibration estimator is developed with assumption that that there exists information on inversely related PSU level auxiliary variable. General notations for two stage sampling has been followed from Särndal *et al.* (1992). Let, $\Omega=\{1,\ldots,k,\ldots,N\}$ be the finite population under consideration and $Y$ be the character under study. $\Omega$ is grouped into $N_I$ different clusters or primary stage units (PSUs) such that $\Omega_I=\{1,...,i,\ \ldots,\ N_I\}$ and $i^{th}$ PSU consists $N_i$ secondary stage units (SSUs) such that $\Omega_i=\{1,...,k,...,N_i\}$, $i\in\Omega_I$. Thus, we have $\Omega=\bigcup_{i=1}^{N_I}\Omega_i$ and $N=\sum_{i=1}^{N_I}N_i$. Under two stage sampling, at the first stage, let, a sample of PSUs $s_I$ of size $n_I$ PSUs is selected from $\Omega_I$ according to a specified design $p_I(.)$ with the inclusion probabilities $\pi_{Ii}$ and $\pi_{Iij}$ at the PSU level. Given that the $i^{\text{th}}$ PSU is selected at the first stage, a sample $s_i$ of size $n_i$ units is drawn from $\Omega_i$ according to some specified design $p_i(.)$ with inclusion probabilities $\pi_{k/i}$ and $\pi_{kl/i}$. $s=\bigcup_{i=1}^{s_I}s_i$ and $n_s=\sum_{i=1}^{n_I}n_i$ define respective total sample of SSUs and its size.

Let, $y_k$ defines the study variable which is observed for all the sample element $k\in s$. In this study, the parameter of interest is the population total which is defined as

$$t_y=\sum_{k=1}^{N}y_k=\sum_{i=1}^{N_I}t_{yi}\,, \tag{2.1}$$

where $t_{yi}=\sum_{k=1}^{N_i}y_k=i^{th}$ PSU total.

Usual Horvitz-Thompson (HT) estimator for population total under the case concerned is defined as

$$\hat{t}_{y\pi}=\sum_{i=1}^{n_I}a_{Ii}\sum_{k=1}^{n_i}\left(a_{k/i}y_k\right)=\sum_{i=1}^{n_I}a_{Ii}\hat{t}_{yi\pi}\,, \tag{2.2}$$

where design weights under two stage sampling are given as

$a_{Ii}=1/\pi_{Ii}$, $\forall\, i\in s_I$ and $a_{k/i}=1/\pi_{k/i}$ $\forall\, k\in s_i,\, i\in s_I$.

Suppose, there exist a PSU level auxiliary variable $U$ having an inverse relationship with cluster totals of the study variable $Y$, i.e. $t_{yi}$. Again, assume that, value of $u_i^{-1}$ is observed for all the sampled PSUs $i\in s_I$ and a correct value of $\sum_{i=1}^{N_I}u_i^{-1}$ is available. Using the well-known Calibration Approach (Deville and Särndal, 1992), we wish for modifying the sampling weight of PSU level, $a_{Ii}$, useful for obtaining the HT estimator as in Equation (2.2). A chi-square type distance function $\sum_{i=1}^{n_I}\frac{\left(w_{Ii}-a_{Ii}\right)^2}{a_{Ii}q_{Ii}}$ has been minimized subject to the constraints $\sum_{i=1}^{n_I}w_{Ii}\,u_i^{-1}=\sum_{i=1}^{N_I}u_i^{-1}$, where $q_{Ii}$ are suitably chosen constants. We minimize following objective function using the method of Lagrange multiplier

$$\varphi(w_{Ii},\lambda)=\sum_{i=1}^{n_I}\frac{\left(w_{Ii}-a_{Ii}\right)^2}{a_{Ii}q_{Ii}}-\lambda\left[\sum_{i=1}^{n_I}w_{Ii}\,u_i^{-1}-\sum_{i=1}^{N_I}u_i^{-1}\right]. \tag{2.3}$$

The new set of calibration weights obtained by minimization of above objective function is given by

$$w_{Ii}=a_{Ii}+a_{Ii}q_{Ii}u_i^{-1}\left[\frac{\sum_{i=1}^{N_I}u_i^{-1}-\sum_{i=1}^{n_I}a_{Ii}\,u_i^{-1}}{\sum_{i=1}^{n_I}a_{Ii}q_{Ii}u_i^{-2}}\right],\quad \forall\, i=1,2,...,n_I\,. \tag{2.4}$$

When we choose $q_{Ii}=u_i$, the revised weights simplify to

$$w_{Ii}=a_{Ii}\left(\sum_{i=1}^{N_I}u_i^{-1}\Big/\sum_{i=1}^{n_I}a_{Ii}u_i^{-1}\right),\quad \forall\, i=1,2,...,n_I. \tag{2.5}$$

The proposed product type calibration estimator under two stage sampling based on the calibration weights $w_{Ii}$ is given by

$$\hat{t}_{yCP}=\sum_{i=1}^{n_I}w_{Ii}\hat{t}_{yi\pi}=\left(\sum_{i=1}^{n_I}a_{Ii}\hat{t}_{yi\pi}\right)\left(\sum_{i=1}^{N_I}u_i^{-1}\right)\Big/\left(\sum_{i=1}^{n_I}a_{Ii}u_i^{-1}\right). \tag{2.6}$$

Using Simple Random Sampling Without Replacement (SRSWOR) at both the stages, proposed product type calibration estimator reduces to

$$\hat{t}_{yCP,SI} = \left(\frac{N_I}{n_I}\sum_{i=1}^{n_I}\frac{N_i}{n_i}\sum_{k=1}^{n_i} y_k\right)\left(\sum_{i=1}^{N_I} u_i^{-1}\right)\Big/\left(\frac{N_I}{n_I}\sum_{i=1}^{n_I} u_i^{-1}\right). \quad (2.7)$$

The theoretical bias of the proposed product type calibration estimator $\hat{t}_{yCP}$ through Taylor series linearization technique is obtained as

$$Bias\left(\hat{t}_{yCP}\right) = \frac{\sum_{i=1}^{N_I} t_{yi}}{\sum_{i=1}^{N_I} u_i^{-1}}\left[\frac{Cov\left(\sum_{i=1}^{n_I} a_{Ii}u_i^{-1},\ \sum_{i=1}^{n_I} a_{Ii}t_{yi}\right)}{\sum_{i=1}^{N_I} t_{yi}} + \frac{V\left(\sum_{i=1}^{n_I} a_{Ii}u_i^{-1}\right)}{\sum_{i=1}^{N_I} u_i^{-1}}\right]. \quad (2.8)$$

Under SRSWOR at both the stages, we obtain the bias through Taylor series linearization technique as

$$Bias(\hat{t}_{yCP,SI}) = \left(\frac{1}{n_I} - \frac{1}{N_I}\right) t_y \left(\rho_A C_{by} C_u + C_u^2\right), \quad (2.9)$$

where

$$\rho_A = \frac{S_{uy}}{S_u S_{yb}}, \quad C_{yb}^2 = \frac{S_{yb}^2}{\bar{Y}_{N_I.}^2}, \quad C_u^2 = \frac{S_u^2}{\bar{U}_{N_I.}^2},$$

$$S_{yb}^2 = \frac{1}{N_I - 1}\sum_{i=1}^{N_I}(N_i\bar{Y}_i - \bar{Y}_{N_I.})^2,$$

$$S_{uy} = \frac{1}{N_I - 1}\sum_{i=1}^{N_I}(N_i\bar{Y}_i - \bar{Y}_{N_I.})(U_i - \bar{U}_{N_I.}) \text{ and}$$

$$S_u^2 = \frac{1}{N_I - 1}\sum_{i=1}^{N_I}(U_i - \bar{U}_{N_I.})^2.$$

Usual product estimator under SRSWOR is given by

$$\hat{t}_{yP} = \left(\frac{N_I}{n_I}\sum_{i=1}^{n_I}\frac{N_i}{n_i}\sum_{k=1}^{n_i} y_k\right)\left(\frac{N_I}{n_I}\sum_{i=1}^{n_I} u_i\right)\Big/\left(\sum_{i=1}^{N_I} u_i\right) \quad (2.10)$$

and its bias through Taylor series linearization technique is given as

$$Bias(\hat{t}_{yP}) = \left(\frac{1}{n_I} - \frac{1}{N_I}\right) t_y \left(\rho_A C_{yb} C_u\right). \quad (2.11)$$

Approximate sampling variance of the proposed product type calibration estimator $(\hat{t}_{yCP})$ was obtained as

$$AV(\hat{t}_{yCP}) = \sum_{i=1}^{N_I}\sum_{j=1}^{N_I}\Delta_{Iij}\frac{Z_{Ii}}{\pi_{Ii}}\frac{Z_{Ij}}{\pi_{Ij}} + \sum_{i=1}^{N_I}\frac{1}{\pi_{Ii}}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i}\Delta_{kl/i}\frac{y_k}{\pi_{k/i}}\frac{y_l}{\pi_{l/i}} \quad (2.12)$$

where

$$Z_{Ii} = t_{yi} - \left(\frac{\sum_{i=1}^{N_I} t_{yi}}{\sum_{i=1}^{N_I} u_i^{-1}}\right)u_i^{-1}, \ t_{yi} = \sum_{k=1}^{N_i} y_k,$$

$$\Delta_{Iij} = \pi_{Iij} - \pi_{Ii}\pi_{Ij}, \ \Delta_{kl/i} = \pi_{kl/i} - \pi_{k/i}\pi_{l/i}.$$

Under SRSWOR at both the stages, it reduces to

$$AV(\hat{t}_{yCP,SI}) = N_I^2\left(\frac{1}{n_I} - \frac{1}{N_I}\right)\left(S_{yb}^2 + R^2 S_{u^{-1}}^2 - 2RS_{yu^{-1}}\right) + \frac{N_I}{n_I}\sum_{i=1}^{N_I} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{yi}^2, \quad (2.13)$$

where

$$R = \left(\sum_{i=1}^{N_I} t_{y_i}\Big/\sum_{i=1}^{N_I} u_i^{-1}\right), \ \bar{Y}_{N_I.} = \frac{1}{N_I}\sum_{i=1}^{N_I} t_{y_i} = \frac{1}{N_I}\sum_{i=1}^{N_I} N_i\bar{Y}_i,$$

$$\bar{U}_{(-1)N_I.} = \frac{1}{N_I}\sum_{i=1}^{N_I} u_i^{-1}, \ S_{yb}^2 = \frac{1}{N_I - 1}\sum_{i=1}^{N_I}\left(N_i\bar{Y}_i - \bar{Y}_{N_I.}\right)^2,$$

$$S_{u^{-1}}^2 = \frac{1}{N_I - 1}\sum_{i=1}^{N_I}\left(u_i^{-1} - \bar{U}_{(-1)N_I.}\right)^2,$$

$$S_{yu^{-1}} = \frac{1}{N_I - 1}\sum_{i=1}^{N_I}\left(N_i\bar{Y}_i - \bar{Y}_{N_I.}\right)\left(u_i^{-1} - \bar{U}_{(-1)N_I.}\right),$$

$$S_{yi}^2 = \frac{1}{N_i - 1}\sum_{k=1}^{N_i}\left(y_k - \bar{Y}_{\text{i}}\right)^2.$$

Following Särndal *et al.* (1992), the Yates–Grundy estimator of the approximate sampling variance of the proposed product type calibration estimator is given by

$$\hat{V}_{YG}(\hat{t}_{yCP}) = -\frac{1}{2}\sum_{i=1}^{n_I}\sum_{j=1}^{n_I}\breve{\Delta}_{Iij}(w_{Ii}\, z_{Ii} - w_{Ij}\, z_{Ij})^2 - \frac{1}{2}\sum_{i=1}^{n_I}\frac{1}{\pi_{Ii}}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\breve{\Delta}_{kl/i}\left(\frac{y_k}{\pi_{k/i}} - \frac{y_l}{\pi_{l/i}}\right)^2, \quad (2.14)$$

where

$$z_{Ii} = \hat{t}_{yi\pi} - \left[\sum_{i=1}^{n_I} a_{Ii}\hat{t}_{yi\pi}\Big/\sum_{i=1}^{n_I} a_{Ii}u_i^{-1}\right]u_i^{-1}, \quad \hat{t}_{yi\pi} = \sum_{k=1}^{n_i}\frac{y_k}{\pi_{k/i}},$$

$$\breve{\Delta}_{Iij} = \frac{(\pi_{Iij} - \pi_{Ii}\pi_{Ij})}{\pi_{Iij}} \text{ and } \breve{\Delta}_{kl/i} = \frac{\pi_{kl/i} - \pi_{k/i}\pi_{l/i}}{\pi_{kl/i}}.$$

Under SRSWOR at both the stages, the Yates–Grundy estimator of the approximate variance of the proposed product type calibration estimator reduces to

$$\hat{V}_{YG}(\hat{t}_{yCP,SI}) = \left( \frac{\sum_{i=1}^{N_I} u_i^{-1}}{\frac{1}{n_I}\sum_{i=1}^{n_I} u_i^{-1}} \right)^2 \left( \frac{1}{n_I} - \frac{1}{N_I} \right) \Big( s_{yb}^2 + \hat{R}^2 s_{u^{-1}}^2 - 2\hat{R}\, s_{yu^{-1}} \Big) + \frac{N_I}{n_I} \sum_{i=1}^{n_I} N_i^2 \left( \frac{1}{n_i} - \frac{1}{N_i} \right) s_{yi}^2, \quad (2.15)$$

where

$$\hat{R} = \left( \sum_{i=1}^{n_I} N_i \bar{y}_i \Big/ \sum_{i=1}^{n_I} u_i^{-1} \right), \; \hat{\bar{Y}}_{N.} = \frac{1}{n_I} \sum_{i=1}^{n_I} N_i \bar{y}_i,$$

$$\bar{y}_{\mathrm{i}} = \frac{1}{n_i} \sum_{k=1}^{n_i} y_k, \; \hat{\bar{U}}_{(-1)\mathrm{N.}} = \frac{1}{n_I} \sum_{i=1}^{n_I} u_i^{-1}, \; s_{yb}^2 = \frac{1}{n_I - 1} \sum_{i=1}^{n_I} \Big( N_i \bar{y}_i - \hat{\bar{Y}}_{N.} \Big)^2,$$

$$s_{u^{-1}}^2 = \frac{1}{n_I - 1} \sum_{i=1}^{n_I} \Big( u_i^{-1} - \hat{\bar{U}}_{(-1)\mathrm{N.}} \Big)^2,$$

$$S_{yu^{-1}} = \frac{1}{n_I - 1} \sum_{i=1}^{n_I} \Big( N_i \bar{y}_i - \hat{\bar{Y}}_{N.} \Big)\Big( u_i^{-1} - \hat{\bar{U}}_{(-1)\mathrm{N.}} \Big),$$

$$s_{yi}^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left( y_k - \bar{y}_{\mathrm{i}} \right)^2.$$

Under this scenario another popular estimator is the regression estimator. From Särndal *et al.* (1992), the regression estimator of the population total under two stage sampling with PSU level auxiliary variable along with its approximate variance and estimator of variance are given by

$$\hat{t}_{ylr} = \hat{t}_{y\pi} + \hat{B}_I \left( \sum_{i=1}^{N_I} u_i - \sum_{i=1}^{n_I} a_{Ii} u_i \right), \quad (2.16)$$

$$AV(\hat{t}_{ylr}) = \sum_{i=1}^{N_I} \sum_{j=1}^{N_I} \Delta_{Iij} \frac{E_{Ii}}{\pi_{Ii}} \frac{E_{Ij}}{\pi_{Ij}} + \sum_{i=1}^{N_I} \frac{1}{\pi_{Ii}} \sum_{k=1}^{N_i} \sum_{l=1}^{N_i} \Delta_{kl/i} \frac{y_k}{\pi_{k/i}} \frac{y_l}{\pi_{l/i}}$$

and

$$\hat{V}_{YG}(\hat{t}_{ylr}) = -\frac{1}{2} \sum_{i=1}^{n_I} \sum_{j=1}^{n_I} \breve{\Delta}_{Iij} \Big( \frac{\hat{E}_{Ii}}{\pi_{Ii}} - \frac{\hat{E}_{Ij}}{\pi_{Ij}} \Big)^2 - \frac{1}{2} \sum_{i=1}^{n_I} \frac{1}{\pi_{Ii}} \sum_{k=1}^{n_i} \sum_{l=1}^{n_i} \breve{\Delta}_{kl/i} \Big( \frac{y_k}{\pi_{k/i}} - \frac{y_l}{\pi_{l/i}} \Big)^2,$$

where

$$E_{Ii} = t_{yi} - B_I u_i, \; \hat{E}_{Ii} = \hat{t}_{yi\pi} - \hat{B}_I u_i, \; B_I = \left( \sum_{i=1}^{N_I} t_{yi} u_i \right) \Big/ \left( \sum_{i=1}^{N_I} u_i^2 \right),$$

$$\hat{B}_I = \left( \sum_{i=1}^{N_I} \frac{t_{yi} u_i}{\pi_{Ii}} \right) \Big/ \left( \sum_{i=1}^{N_I} \frac{u_i^2}{\pi_{Ii}} \right).$$

## 3. SIMULATION STUDY

To study the statistical properties of the product type calibration estimator $(\hat{t}_{yCP})$, a simulation study was carried out considering information on inversely related PSU level auxiliary variable is available. Performance of proposed product type calibration estimators were compared with that of usual HT estimator and regression estimator. Sample selection at each stage of two stage sampling is done by SRSWOR. The sizes of the PSU and the corresponding SSUs were considered to be fixed. In the simulation study, three sets of finite population were generated considering availability of three different types of PSU level auxiliary variable $U$. At first, a finite population of $X$ (auxiliary variable) and $Y$ (study variable) consisting of $N = 5000$ units was generated under the model shown below:

$$y_k = \beta x_k^{-1} + e_k \; ; \; k = 1,\ldots,N \quad (3.1)$$

where, $x_k \sim N(5,\ 1)$ and the errors, $e_k \sim N(0,\ \sigma^2 x_k^{-1})$. We have fixed $\sigma^2 = 0.25$ and $\beta = 10$.

Under the population **Set 1**, PSU level auxiliary variable $U$ is taken as the sum of $X$ variables under a specific PSU i.e. $u_i = t_{xi} = \sum_{k=1}^{N_i} x_k$ ; $i=1,\ldots,N_I$. Under the population **Set 2**, same set of $y$ values from the population **Set 1** are considered for study variable $Y$. Then, PSU level auxiliary variable $U$ was generated under the model

$$u_i = \beta_o t_{yi}^{-1} + e_{oi} \; ; i=1,\ldots,N_I \quad (3.2)$$

where, $t_{yi} = \sum_{k=1}^{N_i} y_k$ and the errors, $e_{oi} \sim N(0,\ \sigma_o^2 t_{yi}^{-1})$.

We have fixed $\sigma^2 = 0.25$ and $\beta = 500$. Under the population **Set 3**, study variable $Y$ is generated under the normal distribution i.e. $y_k \sim N(2,\ 1)$. Then, PSU level auxiliary variable $U$ was generated under the same model as given in Equation (3.2). The correlation coefficient between PSU total of study variable ($t_{yi}$) and the auxiliary variable ($u_i$) are as -0.88, -0.90 and -0.95 respectively under Set 1, 2 and 3. Under **Set 1**, the value of left-hand side of the condition in Equation 2.12 i.e. $\rho\, C_{by} / C_u$ was -1.2 which is quite lower than -0.5. Similarly, under **Set 2**, the value was -0.8. Under **Set 3**, the value was -0.8. These values are lower than -0.5.

From the generated study populations, we have selected a total of 10000 independent samples of different sample sizes using two stage sampling design. For this, at the first stage, 10000 independent samples of PSUs of different sample sizes have been drawn using

SRSWOR scheme and, at the second stage, samples of SSUs of different sample sizes have been drawn from each of the selected PSUs using SRSWOR scheme that resulting in 10000 independent samples of SSUs. From each of these two stage samples, estimates of the proposed product type calibration estimators as well as usual HT and regression estimators of population total were calculated. Different combinations of sample sizes at first stage ($n_I$) and second stage ($n_i$) are as given below:

| | | | |
|---|---|---|---|
| $n_I$ =25, $n_i$=50 | $n_I$ =20, $n_i$=40 | $n_I$ =15, $n_i$ =25 | $n_I$ =10, $n_i$ =25 |
| $n_I$ =25, $n_i$=40 | $n_I$ =20, $n_i$=30 | $n_I$ =15, $n_i$ =30 | $n_I$ =10, $n_i$ =20 |

Developed product type calibration estimators as well as usual HT, product and regression estimators of population total were evaluated on the basis of two measures viz. percentage Relative Bias (%RB) and percentage Relative Root Mean Squared Error (%RRMSE) of any estimator of the population parameter θ as given by

$$RB(\hat{\theta}) = \frac{1}{S}\sum_{i=1}^{S}\left(\frac{\hat{\theta}_i - \theta}{\theta}\right)\times 100 \text{ and}$$

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

where, $\hat{\theta}_i$ are the value of the estimator of population parameter $\theta$ for the character under study obtained at $i^{\text{th}}$ sample in the simulation study and $\theta$ is the overall population total.

## 4. RESULTS AND DISCUSSION

The following tables contain the simulation results obtained for each of the cases considered for evaluation of the proposed product type calibration estimators. Table 1, 2 and 3 contain the %RB and %RRMSE of all the estimators for three different population sets respectively. Three different estimators of population total of the study variable under two stage sampling i.e. the Horvitz-Thompson estimator ($\hat{t}_{y\pi}$), regression estimator ($\hat{t}_{ylr}$) and the proposed product type calibration estimator under SRSWOR ($\hat{t}_{yCP,SI}$) are compared in the following tables.

Table 1 shows %RB of the proposed product type calibration estimators of the population total are very less and nearer to zero. Similar trend can be seen in Table 2 and 3. It can also be seen that, with respect to % RB, the proposed product type calibration estimators were performing comparably with usual HT and regression estimator for the situations of availability of inversely related auxiliary information at PSU level under two stage sampling design. By comparing the estimators with respect to % RRMSE, Table 1 shows that the

**Table 1.** Comparison of all the estimators under two stage sampling on the basis of % RB and % RRMSE under Population **Set 1**

| **Sample Sizes** ($n_I$ _ $n_i$) | **%RB** | | | **%RRMSE** | | |
|---|---|---|---|---|---|---|
| | $\hat{t}_{y\pi}$ | $\hat{t}_{ylr}$ | $\hat{t}_{yCP,SI}$ | $\hat{t}_{y\pi}$ | $\hat{t}_{ylr}$ | $\hat{t}_{yCP,SI}$ |
| 10_20 | -0.006 | -0.016 | 0.007 | 1.790 | 1.752 | 1.665 |
| 10_25 | 0.002 | -0.023 | 0.027 | 1.606 | 1.519 | 1.448 |
| 15_25 | 0.011 | 0.007 | 0.010 | 1.325 | 1.252 | 1.225 |
| 15_30 | 0.021 | 0.016 | 0.023 | 1.185 | 1.100 | 1.080 |
| 20_30 | -0.008 | -0.017 | 0.000 | 1.005 | 0.929 | 0.919 |
| 20_40 | 0.002 | -0.009 | 0.012 | 0.848 | 0.749 | 0.742 |
| 25_40 | 0.004 | -0.002 | 0.009 | 0.741 | 0.671 | 0.669 |
| 25_50 | 0.003 | -0.007 | 0.016 | 0.644 | 0.550 | 0.549 |

**Table 2.** Comparison of all the estimators under two stage sampling on the basis of % RB and % RRMSE under Population **Set 2**

| **Sample Sizes** ($n_I$ _ $n_i$) | **%RB** | | | **%RRMSE** | | |
|---|---|---|---|---|---|---|
| | $\hat{t}_{y\pi}$ | $\hat{t}_{ylr}$ | $\hat{t}_{yCP,SI}$ | $\hat{t}_{y\pi}$ | $\hat{t}_{ylr}$ | $\hat{t}_{yCP,SI}$ |
| 10_20 | -0.006 | -0.025 | 0.004 | 1.790 | 1.750 | 1.660 |
| 10_25 | 0.002 | -0.020 | 0.017 | 1.606 | 1.521 | 1.452 |
| 15_25 | 0.011 | 0.007 | 0.010 | 1.325 | 1.255 | 1.222 |
| 15_30 | 0.021 | 0.014 | 0.026 | 1.185 | 1.108 | 1.078 |
| 20_30 | -0.008 | -0.020 | -0.002 | 1.005 | 0.929 | 0.915 |
| 20_40 | 0.002 | -0.007 | 0.008 | 0.848 | 0.750 | 0.738 |
| 25_40 | 0.004 | -0.002 | 0.008 | 0.741 | 0.670 | 0.663 |
| 25_50 | 0.003 | -0.005 | 0.010 | 0.644 | 0.558 | 0.551 |

**Table 3.** Comparison of all the estimators under two stage sampling on the basis of % RB and % RRMSE under Population **Set 3**

| **Sample Sizes** ($n_I$ _ $n_i$) | **%RB** | | | **%RRMSE** | | |
|---|---|---|---|---|---|---|
| | $\hat{t}_{y\pi}$ | $\hat{t}_{ylr}$ | $\hat{t}_{yCP,SI}$ | $\hat{t}_{y\pi}$ | $\hat{t}_{ylr}$ | $\hat{t}_{yCP,SI}$ |
| 10_20 | 0.025 | 0.038 | 0.058 | 3.412 | 3.381 | 3.198 |
| 10_25 | 0.002 | -0.030 | 0.059 | 3.027 | 2.973 | 2.812 |
| 15_25 | 0.000 | 0.015 | 0.000 | 2.433 | 2.325 | 2.266 |
| 15_30 | -0.002 | -0.006 | 0.021 | 2.212 | 2.081 | 2.029 |
| 20_30 | -0.017 | -0.015 | -0.007 | 1.855 | 1.738 | 1.707 |
| 20_40 | 0.013 | 0.007 | 0.032 | 1.564 | 1.414 | 1.395 |
| 25_40 | -0.002 | -0.002 | 0.009 | 1.376 | 1.253 | 1.243 |
| 25_50 | 0.012 | 0.004 | 0.031 | 1.160 | 1.023 | 1.018 |

proposed product type calibration estimator gives lesser % RRMSE than the usual HT and regression estimator when available PSU level auxiliary variable is inversely related with the character under study. Similar encouraging results can also be seen in Table 2 and 3. With the increase of negative correlation in the study and auxiliary variable, % RRMSE of the proposed product type calibration estimators is decreasing.

## 5. CONCLUSIONS

In this study, under finite population two stage sampling design framework, product type calibration estimator of the population total has been proposed following well-known Calibration Approach for the situations of availability of inversely related auxiliary information at PSU level. In addition, it is assumed that the value of $u_i^{-1}$ is observed for all the sampled PSUs and a correct value of $\sum_{i=1}^{N_I} u_i^{-1}$ is available. To study the statistical performance of the proposed product type calibration estimator as compared to existing estimators of population total, a simulation study was conducted. The simulation results show that the proposed product type calibration estimator of the population total is performing better than usual linear regression and Horvitz-Thompson (HT) estimators under two stage sampling design when available auxiliary variable is inversely related with the study variable.

## ACKNOWLEDGEMENTS

The authors are grateful to the editors and the reviewers for their constructive suggestions/comments that helped to improve the quality of the manuscript.

## REFERENCES

Aditya, K., Sud, U.C., Chandra, H. and Biswas, A. (2016a). Calibration Based Regression Type Estimator of the population total under two stage sampling design. *Journal of Indian Society of Agricultural Statistics*, **70 (1)**, 19-24.

Aditya, K., Sud, U.C. and Chandra, H. (2016b). Calibration approach based estimation of finite population total under two stage sampling. *Journal of the Indian Society of Agricultural Statistics*, **70(3)**, 219-226.

Aditya K., Biswas A., Gupta, A.K. and Chandra, H. (2017). District-level crop yield estimation using calibration approach. *Current Science*, **112(9)**, 1927-1931.

Basak P., Sud, U.C. and Chandra, H. (2017). Calibration estimation of regression coefficient for two-stage sampling design. *Journal of the Indian Society of Agricultural Statistics*, **71(1)**, 1-6.

Biswas, A., Aditya, K., Sud U.C., and Basak, P. (2020). Product type calibration estimation of finite population total under two stage sampling. *Journal of the Indian Society of Agricultural Statistics,* **74(1)**, 23-32.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd Edition. New York: John Wiley & Sons, Inc.

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

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

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.

Kim, J., Sungur, E. and Heo, T. (2007). Calibration approach estimators in stratified sampling. *Statistics and Probability Letters,* **77(1)**, 99-103.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, **32**, 133142.

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.

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

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

Salinas, V.I., Sedory, S.A. and Singh, S. (2018). Calibrated estimators in two-stage sampling. Communications in Statistics - Theory and Methods, **48**(6), 1449-1469. DOI: 10.1080/03610926.2018.1433850.

Sinha N., Sisodia, B.V.S., Singh, S. and Singh, S.K. (2016). Calibration approach estimation of mean in stratified sampling and stratified double sampling, *Communications in Statistics - Theory and Methods*, **46(10)**, 4932-4942. DOI: 10.1080/03610926.2015.1091083.

Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of the general regression estimator: Higher level calibration approach. *Survey Methodology*, **24**, 41-50.

Singh, S., Horn, S., Choudhury, S. and Yu, F. (1999). Calibration of the estimators of variance, *Australian and New Zealand Journal of Statistics,* **41(2)**, 199-212.

Sitter, R.R. and Wu, C. (2002). Efficient estimation of quadratic finite population functions. *Journal of the American Statistical Association*, **97**, 535-543.

Sud, U.C., Chandra, H. and Gupta, V.K. (2014a). Calibration based product estimator in single and two phase sampling. *Journal of Statistical theory and Practice,* **8(1)**, 1-11.

Sud, U.C., Chandra, H. and Gupta, V.K. (2014b). Calibration approach based regression type estimator for inverse relationship between study and auxiliary variable. *Journal of Statistical theory and Practice,* **8(4)**, 707-721.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok. C. (1984). *Sampling Theory with Applications*. Indian Society of Agricultural Statistics, New Delhi.

Tracy, D.S., Singh, S. and Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology,* **29**, 99-104.

Wu, C. and Sitter, R.R. (2001). A model calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, **96**, 185-193.