JIRSS (2022)
Vol. 21, No. 01, pp 37-54
DOI: 10.22034/JIRSS.2022.704624

Two-Step Calibration Estimator with Double Use of Auxiliary Variable: Method and Application

Singh Alka [1], Piyush Kant Rai [2], Muhammad Qasim [3]

[1] Department of Mathematics and Statistics, Banasthali University, Rajasthan, 304022, India.
[2] Department of Statistics, Banaras Hindu University, UP, 221005, India.
[3] Department of Economics, Finance and Statistics, Jonkoping University, Jonkoping, 55111, Sweden.

Received: 01/08/2020, Accepted: 21/12/2022, Published online: 03/05/2023

Abstract. This article introduces a two-step calibration technique for the inverse relationship between study variable and auxiliary variable along with the double use of the auxiliary variable. In the first step, the calibration weights and design weights are set proportional to each other for a given sample. While in the second step, the constant of proportionality is to be obtained on the basis of some different objectives of the investigation viz. bias reduction or minimum Mean Squared Error (MSE) of the proposed estimator. Many estimators based on inverse relationship between x and y have been already developed and are considered to be special cases of the proposed estimator. Properties of the proposed estimator is discussed in details. Moreover, a simulation study has also been conducted to compare the performance of the proposed estimator under Simple Random Sampling Without Replacement (SRSWOR) and Lahiri-Midzuno (L-M) sampling design in terms of percent relative bias and MSE.

Corresponding Author: Singh Alka (singhalka2889@gmail.com)
Piyush Kant Rai (raipiyush5@gmail.com)
Muhammad Qasim (muhammad.qasim@ju.se)

The benefits of two-step calibration estimator are also demonstrated using real life data.

Keywords. Auxiliary Information, Calibration Technique, Distance Functions, Mean Squared Error, Ranks.

MSC: 62Dxx, 62Pxx.

1 Introduction

The prime objective of finite sampling theory is to develop such methodology which provides reliable estimates of the population parameters viz. population mean, total, ratio etc. by incorporating valid and proper additional informations. These additional information are also known as auxiliary information and are used to obtain the improved estimator of population mean or total by means of ratio method of estimation. In this method of estimation, information on auxiliary variable is available which is linearly related to the variable under study and is utilized to estimate the parameter (s). Whereas, in some practical situations, the variable under study and its associated auxiliary variable are negatively correlated. For example, a negative correlation exists between the age of individuals and their sleeping hours in general, height of sea level and temperature etc. In these situations, ratio estimator does not perform well and thus the product estimator, developed by (Murthy, 1964), seems a good alternative.

Deville and Sarndal (1992) considered a new method of calibration for designing weights, by incorporating additional information in the (Horvitz and Thompson, 1952) estimator to improve the parameter(s) estimates. Based on their method, an estimator is developed by (Sud et al., 2014) for the cases with negatively correlated study and auxiliary variables, which outperforms product estimator in terms of bias and mean squared error criterion. The available literature is enriched with a class of estimators based on their proposed calibration methods. (Singh, 2004), (Singh, 2006), (Singh, 2012), (Sud et al., 2014), (Farrell and Singh, 2002), (Farrell and Singh, 2005), (Wu and Sitter, 2001), (Estevao and Sarndal, 2003), (Kott, 2003), (Montanari and Ranalli, 2005), (Rueda et al., 2006), (Rai et al., 2018), (Raiet al., 2020), (Alka et al., 2021), (Alka et al., 2021) derived different efficient calibrated estimators. The two-step technique is proposed by (Singh and Sedory, 2016) for the calibration of design weights for linear relationship between auxiliary and study variables. Further, (Alka et al., 2019) used the same two-step calibration technique using two auxiliary variable in sample surveys. (Alam and Shabbir, 2020) enhanced the accuracy of the estimator of the finite population mean using auxillary information in the ranks of the auxiliary variable in stratified random sampling design.

Now, as far as the negative correlation between the study and auxiliary variables is concerned, the available literature indicates that very few studies have been carried out on a two-step calibration technique for the inverse relationship. Therefore, the present literature encouraged us to introduce a two-step regression-type calibration based estimator under the assumption that a negative correlation exists between the study and auxiliary variables. Furthermore, the Mean Square Error (MSE) properties are derived for the proposed estimator and values of the proportionality constant are obtained theoretically. In order to judge the performance of the proposed estimator, the percent Relative Bias (RB%) and Percent Relative Mean Square Error (RMSE%) are evaluated by using a simulation study under SRSWOR and Lahiri-Midzuno (L-M) sampling designs. In addition, an application of the proposed estimator is shown using Sweden revenue data (Särndal et al., 1992), where the effect of the number of municipal employees on the total number of seat revenue for 284 different municipalities is considered. The very basic properties of such calibration approach based estimators are already developed in the literature of survey sampling. The fundamental problem of such estimators are related with their optimality condition for Bias and MSE's and their real-life applications under such situation are questionable regarding these issues. Here, in the present article, such issues are also considered and an attempt is made to obtain better solution.

2 Proposed Estimator

Let us consider a case where information on one auxiliary variable is available. Suppose there is a finite population $\Omega = \{1, 2, ..., N\}$ from which a probability sample $s(s \in \Omega)$ of size n is drawn following a sampling design denoted by p(.). The first- and second-order inclusion probabilities $\pi_i = P(i \in s)$ and $\pi_{ij} = P(i, j \in s)$ are assumed to be strictly positive, $\forall\, i, j \in \Omega$. The study and the auxiliary variables are denoted as y and x, respectively, and further let (y_i, x_i) denote the values taken by the i^{th} unit in the population by both y and x, $i \in \Omega$. Let the population total of the auxiliary variable $X = \sum_{i=1}^{N} x_i$ be known. One more assumption can be made that the investigator has access to the information on the unit-level x_i, if not then at least the population total of the inverse values of x_i, i.e., $X' = \sum_{i=1}^{N}(1/x_i)$ is known.

The purpose is to estimate the total $Y = \sum_{i=1}^{N} y_i$ of population. The (Horvitz and Thompson, 1952) estimator for population total Y is expressed as $\hat{Y}_{HT} = \sum_{i=1}^{n} d_i y_i$ and that for the population total of inverse values of x_i is $\hat{X}'_{HT} = \sum_{i=1}^{n} \frac{d_i}{x_i}$, where, for $d_i = 1/\pi_i$ are design weights. (Deville and Sarndal, 1992) proposed the calibrated estimator of the population total Y as:

$$\hat{Y}_{cal} = \sum_{i=1}^{n} w_i y_i \,, \tag{2.1}$$

where w_i are modified calibrated weights obtained by minimizing the chi-square distance function $\frac{1}{2}\sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i q_i}$ subject to the constraints

$$\sum_{i=1}^{n} \frac{w_i}{x_i} = \sum_{i=1}^{N} \frac{1}{x_i} \,, \tag{2.2}$$

and $q_i's$ are known positive weights and are usually considered as 1, but (Deville and Sarndal, 1992) also motivated to use unequal weights for the purpose.

2.1 Two- Step Calibration based Estimator

Under this technique, the calibration weights w_i are set proportional to the design weights d_i, that is

$$w_i \propto d_i, \tag{2.3}$$

or equivalently

$$w_i = c_i d_i \implies \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} c_i d_i \,, \tag{2.4}$$

where c_i are constants of proportionality, and can be determined on the basis of different options considered by an investigator. Now, we use the method of Lagrange's multiplier for minimization of function given below

$$L_1 = \frac{1}{2}\sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i} - \lambda_1\left(\sum_{i=1}^{n} \frac{w_i}{x_i} - \sum_{i=1}^{N} \frac{1}{x_i}\right) - \lambda_2\left(\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} c_i d_i\right), \tag{2.5}$$

where λ_1 and λ_2 are the Lagrange multipliers. The value of q_i is considered as 1. Differentiating the above equation with respect to w_i and setting equal to 0, we have

$$w_i = d_i\left(1 + \lambda_1 \frac{1}{x_i} + \lambda_2\right). \tag{2.6}$$

Substituting the value of w_i in Eq.(2.2) and Eq.(2.4), we have

$$\lambda_1 \sum_{i=1}^{n} \frac{d_i}{x_i^2} + \lambda_2 \sum_{i=1}^{n} \frac{d_i}{x_i} = \sum_{i=1}^{N} \frac{1}{x_i} - \sum_{i=1}^{n} \frac{d_i}{x_i}\,. \tag{2.7}$$

and

$$\lambda_1 \sum_{i=1}^{n} \frac{d_i}{x_i} + \lambda_2 \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} c_i d_i - \sum_{i=1}^{n} d_i\,. \tag{2.8}$$

On solving further and replacing the values of λ_1 and λ_2 in w_i, the proposed calibrated estimator based on modified weight is obtained as

$$\hat{Y}_{cal,1} = \hat{Y}_{HT} + \hat{\beta}_1 \left(X' - \hat{X}'_{HT}\right) + \hat{\beta}_2 \sum_{i=1}^{n} (c_i - 1) d_i\,, \tag{2.9}$$

where

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} d_i \sum_{i=1}^{n} \frac{d_i y_i}{x_i} - \sum_{i=1}^{n} d_i y_i \sum_{i=1}^{n} \frac{d_i}{x_i}}{\sum_{i=1}^{n} \frac{d_i}{x_i^2} \sum_{i=1}^{n} d_i - \left(\sum_{i=1}^{n} \frac{d_i}{x_i}\right)^2},$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} d_i y_i}{\sum_{i=1}^{n} d_i} + \left\{ \frac{\frac{\sum_{i=1}^{n} d_i y_i}{\sum_{i=1}^{n} d_i} \left(\sum_{i=1}^{n} \frac{d_i}{x_i}\right)^2 - \sum_{i=1}^{n} \frac{d_i y_i}{x_i} \sum_{i=1}^{n} \frac{d_i}{x_i}}{\sum_{i=1}^{n} \frac{d_i}{x_i^2} \sum_{i=1}^{n} d_i - \left(\sum_{i=1}^{n} \frac{d_i}{x_i}\right)^2} \right\}.$$

- **Special Cases**

  1. If $c_i = 1$, for all $i \in s$, the proposed estimator in Eq.(2.9) reduces to the estimator derived by (Sud et al., 2014).
  2. If $c_i = \frac{N}{n}\pi_i$, for all $i \in s$, the proposed estimator in Eq.(2.9) reduces to

$$\hat{Y}_{cal,1} = \hat{Y}_{HT} + \hat{\beta}_1 \left(X' - \hat{X}'_{HT}\right) + \hat{\beta}_2 N \left(1 - \frac{1}{n} \sum_{i=1}^{n} d_i\right). \tag{2.10}$$

- **Some Properties**

Let us define some notations to study different properties of the proposed estimator.

$$\epsilon_1 = \frac{\hat{Y}_{HT}}{Y} - 1, \epsilon_2 = \frac{\hat{X}'_{HT}}{X'} - 1, \delta_1 = \frac{\hat{\beta}_1}{\beta_1} - 1, \delta_2 = \frac{\hat{\beta}_2}{\beta_2} - 1, \eta = \frac{\sum_{i=1}^{n}(c_i - 1)d_i}{\sum_{i=1}^{N}(c_i - 1)} - 1\,,$$

such that $E(\epsilon_1) = E(\epsilon_2) = E(\delta_1) = E(\delta_2) = E(\eta) = 0$. Rewriting Eq.(2.9) by using above notations, we have

$$\hat{Y}_{cal,1} = Y + Y\epsilon_1 - \beta_1\epsilon_2 X' - \beta_1\delta_2\epsilon_2 X' + \beta_2\left[\sum_{i=1}^{N} c_i - N\right](1 + \eta + \delta_2 + \eta\delta_2).$$

If $c_i = c$, this estimator will be simplified to the following form:

$$\hat{Y}_{cal,1} = Y + Y\epsilon_1 - \beta_1\epsilon_2 X' - \beta_1\delta_2\epsilon_2 X' + \beta_2 N\left[c - 1\right](1 + \eta + \delta_2 + \eta\delta_2).$$

Taking expectation on both sides and setting the bias of the estimator equal to 0, we have

$$c = 1 + \frac{\beta_1 E(\delta_1\epsilon_2)X'}{N\beta_2(1 + E(\eta\delta_2))}\,. \tag{2.11}$$

The above Eq.(2.11) gives the obvious choice of proportionality constant c in order to obtain an exactly unbiased estimator i.e. it should be a constant other than unity. Now, the expression of MSE is obtained as

$$MSE(\hat{Y}_{cal,1}) = E\left[Y\epsilon_1 - \beta_1\epsilon_2 X' - \beta_1\delta_2\epsilon_2 X' + \beta_2 N(c - 1)(1 + \eta + \delta_2 + \eta\delta_2)\right]^2.$$

By taking expected values and neglecting higher order terms, we get

$$\begin{aligned} MSE(\hat{Y}_{cal,1}) =& V(\hat{Y}_{HT}) + \beta_1^2 V(\hat{X}'_{HT}) - 2\beta_1 Cov(\hat{Y}_{HT}, \hat{X}'_{HT}) \\ &+ (c-1)^2\left[N^2\beta_2^2 + N^2 V(\hat{\beta}_2) + \beta_2^2 V(\hat{N}) + 4N\beta_2 Cov(\hat{\beta}_2, \hat{N})\right] \\ &- 2(c-1)\Big[N\beta_2 Cov(\hat{X}'_{HT}, \hat{\beta}_1) + N\beta_1 Cov(\hat{X}'_{HT}, \hat{\beta}_2) \\ &+ \beta_1\beta_2 Cov(\hat{X}'_{HT}, \hat{N}) - NCov(\hat{Y}_{HT}, \hat{\beta}_2) - \beta_2 Cov(\hat{Y}_{HT}, \hat{N})\Big]. \end{aligned} \tag{2.12}$$

On setting $\dfrac{MSE(\hat{Y}_{cal,1})}{(c-1)} = 0$ and keeping the same constants β_1 and β_2 we have,

$$c = 1 + \frac{N\beta_2 Cov(\hat{X}'_{HT}, \hat{\beta}_1) + N\beta_1 Cov(\hat{X}'_{HT}, \hat{\beta}_2) + \beta_1\beta_2 Cov(\hat{X}'_{HT}, \hat{N}) - NCov(\hat{Y}_{HT}, \hat{\beta}_2) - \beta_2 Cov(\hat{Y}_{HT}, \hat{N})}{N^2\beta_2^2 + N^2 V(\hat{\beta}_2) + \beta_2^2 V(\hat{N}) + 4N\beta_2 Cov(\hat{\beta}_2, \hat{N})}.$$

2.2 Calibration-based Estimator with Double Use of Auxiliary Variable

Now, to use auxiliary information in another way by means of their ranks. Let $R_i(x),\ i = 1, 2, \cdots, N$ be the rank of i^{th} unit of auxiliary variable (according to their magnitude) in the population. Along with the two constraints defined in Eq.(2.2) and Eq.(2.4), let us define another rank-based constraint [(Alam and Shabbir, 2020)] defined by

$$\sum_{i=1}^{n} w_i r_i = \sum_{i=1}^{N} R_i\,, \tag{2.13}$$

where $r_i(x),\ i = 1, 2, \cdots, n$ is the rank of i^{th} unit of X in the sample. Again this is the problem of optimization thus, using the method of Lagrange's multiplier for minimization of a function,

$$L_2 = \frac{1}{2}\sum_{i=1}^{n}\frac{(w_i - d_i)^2}{d_i q_i} - \lambda_3\left(\sum_{i=1}^{n}\frac{w_i}{x_i} - \sum_{i=1}^{N}\frac{1}{x_i}\right) - \lambda_4\left(\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} c_i d_i\right) - \lambda_5\left(\sum_{i=1}^{n} w_i r_i - \sum_{i=1}^{N} R_i\right), \tag{2.14}$$

where λ_3, λ_4 and λ_5 are the Lagrange multipliers. Differentiating the above equation with respect to w_i and setting it equal to 0, we have

$$w_i = d_i + d_i q_i\left(\lambda_3\frac{1}{x_i} + \lambda_4 + \lambda_5 r_i\right). \tag{2.15}$$

The calibtration estimator based on the double use of the auxiliary variables (see Appendix), is obtained as

$$\hat{Y}_{cal,2} = \hat{Y}_{HT} + \hat{\beta}_3\left(\sum_{i=1}^{N}\frac{1}{x_i} - \sum_{i=1}^{n}\frac{d_i}{x_i}\right) + \hat{\beta}_4\left(\sum_{i=1}^{n} c_i d_i - \sum_{i=1}^{n} d_i\right) + \hat{\beta}_5\left(\sum_{i=1}^{N} R_i - \sum_{i=1}^{n} d_i r_i\right), \tag{2.16}$$

where

$$\beta_3 = \frac{\left(df - e^2\right)\sum_{i=1}^{N}\dfrac{d_i q_i y_i}{x_i} + (ce - bf)\sum_{i=1}^{n} d_i q_i y_i + (be - cd)\sum_{i=1}^{n} d_i q_i y_i r_i}{det(A)},$$

$$\beta_4 = \frac{(ce - bf)\sum_{i=1}^{N}\dfrac{d_i q_i y_i}{x_i} + \left(af - c^2\right)\sum_{i=1}^{n} d_i q_i y_i + (bc - ae)\sum_{i=1}^{n} d_i q_i y_i r_i}{det(A)},$$

$$\beta_5 = \frac{(be - cd)\sum_{i=1}^{N}\dfrac{d_i q_i y_i}{x_i} + (bc - ae)\sum_{i=1}^{n} d_i q_i y_i + \left(ad - b^2\right)\sum_{i=1}^{n} d_i q_i y_i r_i}{det(A)}.$$

3 Simulation Study

This section presents the design of simulation study to check the performance of the proposed estimators and their results. The aim of this study is to introduce a two-step calibration technique for inverse relationship between x and y. In addition, we compared the efficiency of proposed estimator under SRSWOR and L-M designs using percentage RB (%) and percentage Root MSE (RMSE) (%) as performance criteria. We follow the simulation scheme of Sud et al. (2014). The values of the estimators are replicated M=50000 times using R-software. The percent RB, RMSE and RE of proposed estimator $\hat{Y}_{prop}$ can be given as

$$\mathrm{RB}(\hat{Y}_p) = \frac{\frac{1}{M}(\sum_{i=1}^{M} \hat{Y}_p) - Y}{Y} \times 100\%, \tag{3.1}$$

$$\mathrm{RMSE}(\hat{Y}_p) = \frac{\sqrt{\frac{1}{M}\sum_{i=1}^{M}(\hat{Y}_p - Y)^2}}{Y} \times 100\%, \tag{3.2}$$

$$\mathrm{PRE}(\hat{Y}_p) = \frac{MSE(\hat{Y}_e)}{MSE(\hat{Y}_p)} \times 100\%, \tag{3.3}$$

where $\hat{Y}_p$ is the proposed estimators derived in Eqs. (2.9) and (2.16) and $\hat{Y}_e$ is the existing calibration estimator. The model $y_i = \beta_1 + \beta_2 x_i^{-1} + \epsilon_i\, ;\, i = 1, 2, ..., N$ is considered, where N refers to be population size, ϵ_i is generated from normal distribution with mean $\mu = 0$ and $\sigma_\epsilon^2 = 0.25, 1.00, 1.50, 2.00$, and auxiliary variable x_i is generated from normal distribution with mean $\mu = 5$ and $\sigma_x^2 = 0.50, 1.00, 2.00$, and fixed the values of $\beta_j, j = 1, 2$. The description of the simulation parameter is given in Table 1, where $N = 1000$ and twelve different population datasets $(A_1, A_2, A_3, ...)$ are generated. The inverse relationship between x and y is given in Table 1. Furthermore, in order to check the effect of sample sizes on the proposed two-step calibration estimator, six different combinations of sample size such as $n = 50, 100, 150, 200, 250, 300$ are considered.

4 Real Data based Application

To check the efficiency of proposed two-step inverse calibration estimatorThe Sweden data among its (N=284) municipalities is considered from appendix B, (Särndal et al., 1992) to check the efficiency of proposed two-step inverse (for details see appendix B, (Särndal et al., 1992)). In this dataset, two variables are used i.e. total number of seats revenue as y and the number of municipal employees in 1984 as x. The objective of this application is to estimate the population total Y with known $\sum_{i=1}^{N} \frac{1}{x_i}$. Samples of proportion 5%, 10%, 15%, 20%, 25% and 30% of total population are drawn under two sampling designs i.e. SRSWOR and L-M scheme.

Moreover, to check the efficacy of the proposed two-step inverse calibration estimator with double use of auxiliary variable, the Iris dataset used by (Fisher, 1936) is considered. This data set consists of 3 classes with 50 instances each, where each class refers to a type of iris plant. The study variable is taken as a sepal width (cm) and the auxiliary variable as sepal length (cm) with a correlation of -0.10937 between them. The proposed estimator is compared with the product estimator and the estimator developed by (Sud et al., 2014) as defined below, in terms of their PRE's.

$$\hat{Y}_{prod} = \bar{y}\,\frac{\bar{x}}{\bar{X}}, \tag{4.1}$$

$$\hat{Y}_{Sud} = \hat{Y}_{HT} + \hat{\beta}_{\pi}\left(\sum_{i=1}^{N}\frac{1}{x_i} - \sum_{i=1}^{n}\frac{d_i}{x_i}\right), \tag{4.2}$$

where

$$\hat{\beta_\pi} = \frac{\sum_{i=1}^{n} d_i \sum_{i=1}^{n} \frac{d_i y_i}{x_i} - \sum_{i=1}^{n} d_i y_i \sum_{i=1}^{n} \frac{d_i}{x_i}}{\sum_{i=1}^{n} \frac{d_i}{x_i^2} \sum_{i=1}^{n} d_i - \left(\sum_{i=1}^{n} \frac{d_i}{x_i}\right)^2},$$

Table 1: Simulation Parameter Combination

Parameter Combination	σ_x^2	σ_e^2	$Cor(y, x)$	$Cor(y, \frac{1}{x})$
A_1	0.50	0.25	-0.65011	0.65359
A_2	0.50	1.00	-0.17835	0.18215
A_3	0.50	1.50	-0.10882	0.10389
A_4	0.50	2.00	-0.08566	0.08103
B_1	1.00	0.25	-0.83495	0.88126
B_2	1.00	1.00	-0.47283	0.48141
B_3	1.00	1.50	-0.31475	0.32417
B_4	1.00	2.00	-0.21301	0.22195
C_1	2.00	0.25	-0.19111	0.99858
C_2	2.00	1.00	-0.18766	0.97949
C_3	2.00	1.50	-0.18766	0.97949
C_4	2.00	2.00	-0.16228	0.92429

Table 2: RB(%) and RMSE(%) of proposed estimator $\hat{Y}_{cal,1}$ under SRSWOR design

Sample Proportion (%)	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE
	A_1		A_2		A_3		A_4	
5	0.00038	0.06881	-0.00178	0.26432	0.00223	0.40479	0.00762	0.53503
10	0.00009	0.04715	-0.00066	0.18162	0.00239	0.27638	0.00189	0.36604
15	0.00018	0.03718	-0.00050	0.14186	0.00084	0.21963	0.00343	0.29190
20	0.00031	0.03126	0.00006	0.11956	-0.00027	0.18434	-0.00050	0.24409
25	0.00014	0.02700	-0.00078	0.10374	0.00045	0.15939	0.00079	0.21080
30	0.00014	0.02393	0.00025	0.09140	0.00101	0.14056	0.00133	0.18702
	B_1		B_2		B_3		B_4	
5	-0.00016	0.06830	0.00205	0.26352	0.00010	0.40609	0.00013	0.53543
10	-0.00053	0.04705	0.00063	0.18070	0.00126	0.27555	0.00235	0.36684
15	-0.00011	0.03716	0.00072	0.14215	0.00260	0.21988	0.00139	0.28874
20	-0.00022	0.03112	0.00065	0.12071	0.00067	0.18323	0.00058	0.24276
25	0.00010	0.02689	0.00019	0.10361	0.00075	0.15920	0.00032	0.21128
30	-0.00001	0.02368	0.00142	0.09155	0.00076	0.14062	0.00002	0.18631
	C_1		C_2		C_3		C_4	
5	0.00002	0.06845	0.00683	0.26118	-0.01778	0.39919	0.00977	0.53094
10	0.00019	0.04703	0.00386	0.17877	-0.00971	0.27429	0.00751	0.36569
15	0.00033	0.03679	0.00346	0.14134	-0.00990	0.21689	0.00635	0.28949
20	-0.00003	0.03104	0.00206	0.11904	-0.00653	0.18242	0.00506	0.24248
25	0.00004	0.02689	0.00250	0.10355	-0.00487	0.15788	0.00364	0.20964
30	0.00011	0.02361	0.00198	0.09077	-0.00457	0.13923	0.00299	0.18448

Table 3: RB(%) and RMSE(%) of proposed estimator $\hat{Y}_{cal,1}$ under L-M design

Sample Proportion (%)	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE
	A_1		A_2		A_3		A_4	
5	-0.01692	2.80011	0.00020	2.83086	-0.04339	2.84582	-0.01841	2.85918
10	-0.00536	0.88070	-0.00488	0.89370	-0.00222	0.91894	-0.00356	0.94794
15	-0.00504	0.43372	-0.00238	0.45550	0.00098	0.48578	-0.00688	0.52304
20	-0.00373	0.25726	0.00183	0.28041	-0.00060	0.31601	0.00025	0.35325
25	0.00023	0.16920	-0.00119	0.19478	-0.00118	0.22894	0.00121	0.26554
30	-0.00045	0.11575	-0.00037	0.14547	0.00064	0.17919	0.00081	0.21747
	B_1		B_2		B_3		B_4	
5	-0.03433	1.44462	-0.03362	1.46907	-0.04539	1.49526	-0.02535	1.54558
10	-0.01077	0.45528	-0.01125	0.49047	-0.01101	0.53140	-0.00837	0.58788
15	-0.00552	0.22678	-0.00551	0.26501	-0.00200	0.31340	-0.00723	0.36783
20	-0.00285	0.13660	-0.00288	0.17798	-0.00332	0.22749	-0.00392	0.27854
25	-0.00188	0.09106	-0.00153	0.13487	-0.00110	0.18020	-0.00110	0.22893
30	-0.00127	0.06472	-0.00135	0.10896	-0.00114	0.15175	0.00032	0.19550
	C_1		C_2		C_3		C_4	
5	-0.04409	1.46436	-0.03697	1.47113	-0.04062	1.52060	-0.04198	1.55579
10	-0.00773	0.88371	-0.00505	0.89370	-0.01337	0.93985	0.00003	0.95816
15	0.00035	0.68633	-0.00291	0.69584	-0.00059	0.73563	-0.00237	0.75268
20	0.00157	0.57416	-0.00237	0.58364	-0.00414	0.61489	0.00138	0.62561
25	-0.00096	0.49533	0.00132	0.50229	-0.00569	0.52679	0.00403	0.54282
30	0.00430	0.43489	0.00216	0.44501	-0.00081	0.46569	0.00128	0.47660

Table 4: RB (%) and RMSE (%) of proposed estimator $\hat{Y}_{cal,1}$ under SRSWOR design

Sample Proportion (%)	RB	RMSE
5	9.49	34.87
10	5.06	26.20
15	4.64	20.43
20	3.91	18.54
25	1.09	16.83
30	1.36	14.77

Table 5: RB (%) and RMSE (%) of proposed estimator $\hat{Y}_{cal,1}$ under L-M design

Sample Proportion (%)	RB	RMSE
5	12.55	60.10
10	6.30	39.20
15	3.69	31.03
20	2.66	26.45
25	2.02	23.79
30	1.43	21.16

Probability samples of various proportion sizes at an increment of 5% are drawn using SRSWOR sampling scheme. . The PRE values are calculated for the proposed estimators ($\hat{Y}_{cal,1}$ and $\hat{Y}_{cal,2}$) with respect to the existing estimators which are shown in Table 6. The whole procedure is replicated 500 times using R-software.

Table 6: PRE of Proposed Calibration Estimators ($\hat{Y}_{cal,1}$ and $\hat{Y}_{cal,2}$)

Sample Proportion (%)	$\hat{Y}_{cal,1}$		$\hat{Y}_{cal,2}$	
	$\hat{Y}_{prod}$	$\hat{Y}_{Sud}$	$\hat{Y}_{prod}$	$\hat{Y}_{Sud}$
5	127.92	6539.54	151.85	7762.51
10	163.14	6840.45	181.64	7616.49
15	137.57	1525.81	143.05	4996.99
20	151.48	5282.29	159.58	5564.55
25	146.91	1624.04	154.46	3551.96
30	173.08	3210.93	161.51	2996.23

5 Results and Discussion

Tables 2 and 3 show the result of RB (%) and RMSE (%) of population total, Y, under SRSWOR and L-M sampling designs, respectively. It is found that, as the inverse relationship or the degree of correlation decreased, the percent RB and RMSE is increased. When the sample size increased, the percent relative bias and root mean squared error decreased. For instance, the minimum RMSE is obtained for 30 % sample proportion for B_1 population set (0.02361) and maximum for 5 % sample proportions for B_4 population set (0.53543) under SRSWOR sampling design. Whereas, under L-M sampling scheme, the minimum and maxmum RMSE values are obtained for B_1 population set

(0.06472) using 30 % sample proportion and A_4 population set (2.85918) using 5 % sample proportion respectively. It is noted that the sample size (or proportion) has a significant effect on the percent relative root mean squared error. The two-step calibration method always performed quite well under SRSWOR design instead of the L-M design for the simulation study.

From the Tables 4 and 5, it can be concluded empirically that the proposed estimator under SRSWOR design performed consistently better than for L-M design in terms of percent relative bias and percent relative root mean squared error. On increasing sample proportions from 5 % to 30 %, a decrease of 57.64 % and 64.79 % in RMSE values is observed under SRSWOR and L-M sampling designs respectively. Further, a decrease of 85.66 % and 88.60 % in RB values is obtained for same sampling proportions increasing patterns under SRSWOR and L-M sampling designs respectively. This reveals that the suggested estimator shows better efficiency as compared to existing dataset for real life application also.

The PRE values of the proposed calibration estimators ($\hat{Y}_{cal,1}$ and $\hat{Y}_{cal,2}$) with respect to the existing calibration estimators ($\hat{Y}_{prod}$ and $\hat{Y}_{Sud}$) is presented in Table 6. The maximum PRE is obtained for 10 % and 5 % sample proportions for $\hat{Y}_{cal,1}$ (6840.45) and $\hat{Y}_{cal,2}$ (7762.51) respectively. It can be observed that, both proposed estimators are considerably efficient compared to the (Sud et al., 2014) estimator and product estimator available in the literature.

6 Concluding Remarks

In this article, a new calibrated estimator under the two-step calibration technique given by (Singh and Sedory, 2016) is proposed for the inverse relationship between x and y. Moreover, the properties of the proposed estimator have been studied. The result of the proposed two-step calibtration estimator is also extended for the case of double use of auxiliary variable. The simulated and real application results show that the proposed estimator performs quite well under basic SRSWOR design. Since large sample sizes (e.g., when sampling fraction was 0.25 or more) show a visible gain in efficiency of the proposed estimator with respect to RB(%) and RMSE(%) under both sampling designs. Also, the calibrated estimator with double use of auxiliary variable is found to perform better than the simple two-step inverse calibration estimator, product estimator and (Sud et al., 2014) estimator in terms of PRE.

Acknowledgements

The authors are grateful to Banasthali Vidyapith, Rajasthan, Banaras Hindu University, Varanasi, India and Jonkoping University, Sweden for support in present research.

References

Alam, Sh., and Shabbir, J. (2020), Calibration Estimation of Mean by Using Double Use of Auxiliary Information. *Communications in Statistics - Simulation and Computation*, 10.1080/03610918.2020.1749660, 1-19.

Alka, S., Rai P. K., and Qasim, M. (2019), Two-Step Calibration of Design Weights under Two Auxiliary Variables in Sample Survey *Journal of Statistical Computation and Simulation*, **89**(12), 2316-2327.

Alka, S., Rai P. K., and Qasim, M. (2021), Calibration-Based Estimators using Different Distance Measures under Two Auxiliary Variables: A Comparative Study *Journal of Modern Applied Statistical Methods*, **19**, 3003.

Alka, S., Emmanuel J. E., Rai P. K., and Tiwari, Sh. (2021), Calibration Estimator of Population Mean under Stratified Systematic Sampling Design. *International Journal of Agricultural and Statistical Sciences*, **17**, 2353-2361.

Deville, J. C., and Sarndal, C. E. (1992), Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, **87**(418), 376-382.

Estevao, V. M., and Särndal, C. E. (2003), *Proceedings of Joint Statistical Meeting-Section on Survey Research Methods: A New Perspective to Calibration Estimators*, 1346-1356.

Farrell, P. J., and Singh, S. (2002), *Proceedings of Joint Statistical Meeting-Section on Survey Research Methods: Penalized Chi-Square Distance Function in Survey Sampling*, New York, 963-968.

Farrell, P. J., and Singh, S. (2005), Model Assisted Higher Order Calibration of Estimators of Variance. *Australian and New Zealand Journal of Statistics*, **47**(3), 375-383.

Fisher, R. A. (1936), The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7**(2), 179-188.

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**(260), 663-685.

Kott, P. S. (2003), *Proceedings of Joint Statistical Meetings - Section on Survey Research Methods: An Overview of Calibration Weighting*, Wiley Online Library, 2241-2252.

Montanari, G. E, and Ranalli, M. G. (2005), Nonparametric Model Calibration Estimation in Survey Sampling. *Journal of the American Statistical Association*, **100**(472), 1429-1442.

Murthy, M. N. (1964), Product Method of Estimation. *Sankhyā: The Indian Journal of Statistics, Series A*, **26**(1), 69-74.

Rai, P. K., Tikkiwal G. C., and Alka (2018), Calibration approach-based estimator under minimum entropy distance Function VIS-A-VIS T-2 class of estimator. *International Journal of Applied Engineering Research*, **13**, 15329–15342.

Rai P. K., Tikkiwal, G. C., and Alka (2020), *Statistical Methods and Applications in Forestry and Environmental Sciences: A Joint Calibration Estimator of Population Total Under Minimum Entropy Distance Function Based on Dual Frame Surveys*, Springer, 125-150.

Rueda M., Martínez , S., Martínez, H., and Arcos, A. (2006), Mean Estimation with Calibration Techniques in Presence of Missing Data. *Computational Statistics and Data Analysis*, **50**(11), 3263-3277.

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

Singh, S. (2004), *Proceedings of the American Statistical Association, Survey Method Section: Golden and Silver Jubilee Year-2003 of the Linear Regression Estimators*, 4382-4389.

Singh, S. (2006), Survey Statisticians Celebrate Golden Jubilee Year-2003 of the Linear Regression Estimator. *Metrika*, **63**, 1–18.

Singh, S. (2012), On the Calibration of Design Weights Using a Displacement Function. *Metrika*, **75**(1), 85–107.

Singh, S., and Sedory, S. A. (2016), Two-Step Calibration of Design Weights in Survey Sampling. *Communications in Statistics - Theory and Methods*, **45**(12), 3510-3523.

Sud, U. C., Chandra, H., and Gupta, V. K. (2014), 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. (2014), Calibration Approach-Based Regression-Type Estimator for Inverse Relationship Between Study and Auxiliary Variable. *Journal of Statistical Theory and Practice*, **8**(4), 707-721.

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**(453), 185-193.

A Appendix

On substituting the value of w_i from Eq.(2.15) in Eqs.(2.2), (2.4) and (2.13), we get a system of linear equations in lambdas, whose matrix form is given by

$$\begin{bmatrix} \sum_{i=1}^{n} \frac{d_i q_i}{x_i^2} & \sum_{i=1}^{n} \frac{d_i q_i}{x_i} & \sum_{i=1}^{n} \frac{d_i q_i r_i}{x_i} \\ \sum_{i=1}^{n} \frac{d_i q_i}{x_i} & \sum_{i=1}^{n} d_i q_i & \sum_{i=1}^{n} d_i q_i r_i \\ \sum_{i=1}^{n} \frac{d_i q_i r_i}{x_i} & \sum_{i=1}^{n} d_i q_i r_i & \sum_{i=1}^{n} d_i q_i r_i^2 \end{bmatrix} \begin{bmatrix} \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \frac{1}{x_i} - \sum_{i=1}^{n} \frac{d_i}{x_i} \\ \sum_{i=1}^{n} c_i d_i - \sum_{i=1}^{n} d_i \\ \sum_{i=1}^{N} R_i - \sum_{i=1}^{n} d_i r_i \end{bmatrix}, \tag{A.1}$$

where

$$\begin{bmatrix} \sum_{i=1}^{n} \frac{d_i q_i}{x_i^2} & \sum_{i=1}^{n} \frac{d_i q_i}{x_i} & \sum_{i=1}^{n} \frac{d_i q_i r_i}{x_i} \\ \sum_{i=1}^{n} \frac{d_i q_i}{x_i} & \sum_{i=1}^{n} d_i q_i & \sum_{i=1}^{n} d_i q_i r_i \\ \sum_{i=1}^{n} \frac{d_i q_i r_i}{x_i} & \sum_{i=1}^{n} d_i q_i r_i & \sum_{i=1}^{n} d_i q_i r_i^2 \end{bmatrix} = A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

After solving the system of linear equations, we get

$$\lambda_3 = \frac{(df-e^2)\left(\sum_{i=1}^{N}\dfrac{1}{x_i}-\sum_{i=1}^{n}\dfrac{d_i}{x_i}\right)+(ce-bf)\left(\sum_{i=1}^{n}c_id_i-\sum_{i=1}^{n}d_i\right)+(be-cd)\left(\sum_{i=1}^{N}R_i-\sum_{i=1}^{n}d_ir_i\right)}{det(A)},$$

$$\lambda_4 = \frac{(ce-bf)\left(\sum_{i=1}^{N}\dfrac{1}{x_i}-\sum_{i=1}^{n}\dfrac{d_i}{x_i}\right)+(af-c^2)\left(\sum_{i=1}^{n}c_id_i-\sum_{i=1}^{n}d_i\right)+(bc-ae)\left(\sum_{i=1}^{N}R_i-\sum_{i=1}^{n}d_ir_i\right)}{det(A)},$$

$$\lambda_5 = \frac{(be-cd)\left(\sum_{i=1}^{N}\dfrac{1}{x_i}-\sum_{i=1}^{n}\dfrac{d_i}{x_i}\right)+(bc-ae)\left(\sum_{i=1}^{n}c_id_i-\sum_{i=1}^{n}d_i\right)+(ad-b^2)\left(\sum_{i=1}^{N}R_i-\sum_{i=1}^{n}d_ir_i\right)}{det(A)},$$

where $det(A) = adf-ae^2-b^2f+2bce-dc^2$. On putting these values of λ_3, λ_4 and λ_5 in Eq. (2.15), the optimum calibrated weights are obtained which are used for the derivation of final form of calibration estimator in Eq.(2.16).