Statistics and Applications {ISSN 2454-7395 (online)}
Volume 21, No. 1, 2023 (New Series), pp 11-22

Calibration Estimator in Two Stage Sampling Using Double Sampling Approach when Study Variable is Inversely Related to Auxiliary Variable

Ankur Biswas[1], Kaustav Aditya[1], U.C. Sud[2] and Pradip Basak[3]
[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[2]*Former Director of ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[3]*Department of Agricultural Statistics, Uttar Banga Krishi Vishwavidyalaya, West Bengal*

Received: 05 April 2021; Revised: 08 December 2021; Accepted: 20 January 2022

Abstract

The calibration approach is a popular technique for incorporating auxiliary information for estimation of population parameters in survey sampling. In general, the Calibration Approach assumes the availability of population-level auxiliary information. On the contrary, in large scale surveys, it is often the case that population-level data on auxiliary variable is not available, but it is relatively inexpensive to collect. In the present article, in case of non-availability of population-level relatively inexpensive data on auxiliary variable under two stage sampling, we developed product type calibration estimator of the finite population total using double sampling approach along with the sampling variance and variance estimator. The study variable is assumed to be inversely related with the auxiliary variable. Proposed product type calibration estimator was evaluated through a simulation study which showed that the proposed product type calibration estimator was performing efficiently over traditional Narain-Horvitz-Thompson type expansion estimator as well as product estimator of the finite population total in case of two stage sampling involving two phases at both the stages.

Key words: Auxiliary information; Calibration; Design weights; Product estimator; Simulation; Double sampling.

1. Introduction

The calibration approach was originally suggested by Deville and Särndal (1992). It is a most widely used techniques combining auxiliary information for estimation of unknown finite population parameters of the character under study efficiently. In calibration approach, initial design weights would be converted to calibrated weights which is product of a calibration factor with the initial design weight. Following Deville and Särndal (1992), plenty of work has been carried out in the calibration estimation *i.e.* Singh *et al.* (1998, 1999), Wu and Sitter (2001), Sitter and Wu (2002), Kott (2006), Estevao and Särndal (2006), *etc*. (see Kim and Park (2010) and Särndal (2007) for comprehensive review of calibration approach).

In various medium to large scale surveys, two stage sampling is followed since at most situations it is very often the case that the sampling frame is often unavailable or it

Corresponding Author: Ankur Biswas
Email: ankur.biswas@icar.gov.in

could be too expensive to construct one. Under this sampling design, first, groups of elements are selected which are called as primary stage units (PSU) and, then, a sample of basic elements which are called as secondary stage units (SSU) are selected from each selected PSU. For example, in agricultural surveys, villages can be selected as PSU and farmers can be selected as SSU. Sukhatme *et al.* (1984) suggested several estimators of the finite population parameters using auxiliary information in two stage sampling. Särndal *et al.* (1992) considered three different situations concerning availability of complex auxiliary information under two stage sampling and discussed extensively on ratio and regression estimators under such situations. The calibration estimation under availability of complex auxiliary information under two stage sampling has been discussed by several authors such as Aditya *et al.* (2016a, 2016b), Mourya *et al.* (2016), Aditya *et al.* (2017), Basak *et al.* (2017), Salinas *et al.* (2018) and Biswas *et al.* (2020) etc.

In surveys, it is often the case that there exist certain auxiliary variables which are inversely related to the character under study. For example, in household based surveys, the marketable surplus is inversely related to family consumption of seed, feed etc. In the past, the product estimator (Murthy, 1964) was used as an efficient alternative to the traditional estimators. In such a situation, the usual methodology for calibration estimation may not fit in. Sud *et al.* (2014 a, b) and Biswas *et al.* (2020) proposed calibration estimation procedures for finite population total under uni-stage equal probability sampling and two stage sampling respectively, when a character under study is inversely related to the available auxiliary variable.

Generally, in calibration approach, it is assumed that population-level auxiliary information is available. On the contrary, population-level data on auxiliary variable is not available in practice, but relatively inexpensive to collect. Under this scenario, double sampling approach serves as a feasible solution for the estimation of finite population parameters. Double sampling has generated extensive research interests. For example Rao (1973), Hidiroglou *et al.* (2009), Haziza *et al.* (2011), Sinha *et al.* (2016), Arnab (2017), *etc*. In this present study, in case of non-availability of population-level relatively inexpensive data on auxiliary variable under two stage sampling, an attempt has been made to develop calibration estimation procedure for estimation of finite population total using double sampling approach when character under study is inversely related to the available auxiliary variable. In Section 2, we give a brief of the product type calibration estimators of finite population total under two stage sampling as proposed by Biswas *et al.* (2020). In Section 3, calibration estimators have been proposed in case of two stage sampling using double sampling approach when there was unavailability of population level auxiliary information at the SSU level and the character under study is inversely related to auxiliary variable. The statistical properties of the proposed estimators are studied empirically through a simulation study. Section 4 provides the technical details of the simulation study and simulation results. Concluding remarks are given in Section 5.

2. Calibration estimators under two stage sampling when character under study is inversely related to available auxiliary information

In this section, first, we briefly describe two stage sampling design along with two different calibration estimators under two stage sampling under the assumption that the character under study is inversely related to available auxiliary information as proposed by Biswas *et al.* (2020).

Let, the finite population under consideration and the corresponding character under study is denoted by U and Y. Population U is grouped into N different PSUs such that $U_I = \{1,...,i,...,N\}$ and i^{th} PSU consists of M_i SSUs such that $U_i = \{1,...,k,...,M_i\}$, $i \in U_I$. Thus, we have $U = \bigcup_{i=1}^{N} U_i$ and total number of SSUs in U is $M_0 = \sum_{i=1}^{N} M_i$. Under two stage sampling, a sample of n PSUs (s_I) is drawn from U_I at stage one. First and second order inclusion probabilities at the PSU level are $\pi_{Ii} = P(i \in s_I)$ and $\pi_{Iij} = P(i, j \in s_I)$. A second stage sample (s_i) of m_i SSUs is drawn from U_i provided, at the first stage, the i^{th} PSU (U_i) is selected. First and second order inclusion probabilities at the SSU level are $\pi_{k/i} = P(k \in s_i / i \in s_I)$ and $\pi_{kl/i} = P(k,l \in s_i / i \in s_I)$. In the second stage of sampling, invariance and independence property is followed. The final sample of SSUs is denoted as, $s = \bigcup_{i=1}^{s_I} s_i$. Let, y_{ik} denotes the observation of the study variable from k^{th} SSU in i^{th} PSU and it is observed for all the sampled SSUs. The parameter of interest is the population total $t_y = \sum_{i=1}^{N}\sum_{k=1}^{M_i} y_{ik} = \sum_{i=1}^{N} t_{yi}$, where $t_{yi} = \sum_{k=1}^{M_i} y_{ik}$ = i^{th} PSU total. Usual Narain-Horvitz-Thompson estimator for population total is given by

$$\hat{t}_{y\pi} = \sum_{i=1}^{n} a_{Ii} \sum_{k=1}^{m_i} \left(a_{k/i} y_{ik}\right) = \sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik} y_{ik} \tag{1}$$

where, the design weights are given as

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

Biswas *et al.* (2020) proposed product type calibration estimators of population total for two situations under two stage sampling design as per Särndal *et al.* (1992) as mentioned below:

Case 1: Population level complete auxiliary information is available at the SSU level.
Case 2: Population level auxiliary information is available only for the selected Primary Stage Units (PSU).

The product type calibration estimators of population total (Biswas *et al.*, 2020) under these two cases of two stage sampling as stated above were given by

$$\hat{t}_{yCP1} = \sum_{i=1}^{n}\sum_{k=1}^{m_i} w_{1ik} y_{ik} = \left(\sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik} y_{ik}\right)\left(\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}^{-1}\right)\Bigg/\left(\sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1}\right) \text{ and}$$

$$\hat{t}_{yCP2} = \sum_{i=1}^{n} a_{Ii} \sum_{k=1}^{m_i} w_{2ik} y_{ik} = \sum_{i=1}^{n}\left[a_{Ii}\left(\sum_{k=1}^{m_i} a_{k/i} y_{ik}\right)\left(\sum_{i=1}^{M_i} x_{ik}^{-1}\right)\Bigg/\left(\sum_{i=1}^{m_i} a_{k/i} x_{ik}^{-1}\right)\right]$$

where,

$$w_{1ik} = a_{ik}\left(\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}^{-1} \Bigg/ \sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1}\right),\; \forall k = 1,2,...,m_i \;\text{ and}$$

$$w_{2ik} = a_{k/i}\left(\sum_{k=1}^{M_i} x_{ik}^{-1} \Bigg/ \sum_{k=1}^{m_i} a_{k/i}\, x_{ik}^{-1}\right), \forall k = 1,2,...,m_i \,.$$

3. Proposed calibration estimator using double sampling approach in both stages of two stage sampling

The double sampling was first given by Neyman (1938) which is generally used when the information on auxiliary variable is lacking, but comparatively low-cost to obtain. Information on auxiliary variable shall be obtained by selecting a larger preliminary sample. Further, sub-sample is taken to observe the charater under study. In the present study, we have developed calibration estimators using double sampling approach in two stage sampling for the situations when of population level auxiliary information (x_{ik}) was unavailabile at SSU level.

Biswas *et al.* (2020) developed product type calibration estimator under two stage sampling assuming the SSU level auxiliary variable is inversely related to the characteristics under study and $\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}^{-1}$ is already known. Under the present situation, it is assumed that a correct value of $\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}^{-1}$ is unavailable since there was there was unavailability of population level auxiliary information (x_{ik}). We consider double sampling approach under this scenario. First, a large first phase sample (s'_I) of n' PSUs is selected from the population of N PSUs (U_I) following a sampling design $p'_I(.)$. The design weight for i^{th} PSU is given by $a'_{Ii} = 1/\pi'_{Ii}$, where $\pi'_{Ii} = P(i \in s'_I)$ is the known first phase first order inclusion probability of i^{th} PSU. Under SRSWOR, $\pi'_{Ii} = n'/N$. In the second stage of the first phase sampling, from each of the i^{th} selected PSU, $i \in s'_I$, a sub-sample (s'_i) of m'_i SSUs is selected from M_i population SSUs (U_i) by a sampling design $p_{k/s'_I}(.)$. The design weight for k^{th} SSU provided i^{th} PSU is already selected can be given as $a'_{k/i} = 1/\pi'_{k/i}$, where $\pi'_{k/i} = P(k \in s'_i / i \in s'_I)$ is first phase inclusion probability of k^{th} SSU and under SRSWOR it is given by $\pi'_{k/i} = m'_i / M_i$. The observation on auxiliary variable x_{ik} is taken from the k^{th} SSU in i^{th} PSU.

In the second phase, a smaller sub-sample (s_I) of n PSUs is drawn from s'_I by a sampling design $p_I(.)$. The design weight of the i^{th} PSU is $a_{Ii/s'_I} = 1/\pi_{Ii/s'_I}$, where $\pi_{Ii/s'_I} = P(i \in s_I / s'_I)$ is the second phase conditional inclusion probability of i^{th} PSU, given s'_I, and under SRSWOR, $\pi_{Ii/s'_I} = n/n'$. In the second stage of second phase sampling, from the i^{th} selected PSU, $i \in s_I$, a smaller sub-sample s_i of size m_i SSUs is selected from m'_i first phase SSUs by a sampling design $p_{k/s_I}(.)$. The sampling weight for k^{th} SSU is given by

$a_{k/s'_i,i} = 1/\pi_{k/s'_i,i}$, where $\pi_{k/s'_i,i} = P(k \in s_i / k \in s'_i, i \in s_I)$ is the second phase conditional inclusion probability of k^{th} SSU and under SRSWOR it is given by $\pi_{k/s'_i,i} = m_i / m'_i$. The observations on the character under study, y_{ik}, and auxiliary variable, x_{ik}, are taken from the k^{th} sampled SSUs in i^{th} selected PSU.

In this study, an attempt has been made to improve the traditional design weighted Narain-Horvitz-Thompson (NHT) (Narain, 1951; Horvitz and Thompson, 1952) type expansion estimator for population total (t_y) under two stage sampling following double sampling at both the stages which is given by

$$\hat{t}_{y\pi} = \sum_{i=1}^{n} a_{Ii} \sum_{k=1}^{m_i} a_{k/i} y_{ik} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} y_{ik} \tag{2}$$

where, $a_{ik} = a_{Ii}\, a_{k/i} = \left(a'_{Ii}\, a_{i/s'_I}\right)\left(a'_{k/i}\, a_{k/s'_i i}\right)$ is the total sampling weight of k^{th} SSU in i^{th} selected PSU in the second phase sample, which reduces to $a_{ik} = \left(\frac{N}{n'}\frac{n'}{n}\right)\left(\frac{M_i}{m'_i}\frac{m'_i}{m_i}\right) = \frac{NM_i}{nm_i}$ under SRSWOR at all stages and phases.

Proposed calibration estimator of the population total (t_y) in case of two stage sampling following double sampling approach is given by

$$\hat{t}_{yCPd} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{ikd}\, y_{ik} \tag{3}$$

where w_{ikd} is the calibration weight under double sampling corresponding to the total sampling weight a_{ik}.

We obtained calibration weights w_{ikd} by minimizing the Chi-square type distance function $\sum_{i=1}^{n} \sum_{k=1}^{m_i} \frac{\left(w_{ikd} - a_{ik}\right)^2}{a_{ik} q_{ik}}$ subject to the constraint $\sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{ikd}\, x_{ik}^{-1} = \sum_{i=1}^{n'} \sum_{k=1}^{m'_i} a'_{ik}\, x_{ik}^{-1}$, where, $a'_{ik} = a'_{Ii}\, a'_{k/i}$. Using Lagrangian multiplier technique, the new calibrated weight is given by

$$w_{ikd} = a_{ik} + a_{ik} q_{ik} x_{ik}^{-1} \left[\frac{\sum_{i=1}^{n'} \sum_{k=1}^{m'_i} a'_{ik}\, x_{ik}^{-1} - \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik}\, x_{ik}^{-1}}{\sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} q_{ik} x_{ik}^{-2}} \right] \quad \forall\, k = 1,2,\ldots,m_i,\ i \in s_I . \tag{4}$$

Using the results of the Equation (4) in (3) and considering $q_{ik} = x_{ik}$, we have, therefore, proved the following result.

Theorem 1: Following double sampling approach under two stage sampling, the proposed product type calibration estimator of population total is given as

$$\hat{t}_{yCPd} = \sum_{i=1}^{n}\sum_{k=1}^{m_i} w_{ikd}\, y_{ik} = \frac{\left(\sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik}\, y_{ik}\right)\left(\sum_{i=1}^{n'}\sum_{k=1}^{m'_i} a'_{ik}\, x_{ik}^{-1}\right)}{\left(\sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik}\, x_{ik}^{-1}\right)}, \tag{5}$$

where, proposed calibration weights corresponding to respective design weights are

$$w_{ikd} = a_{ik}\left(\sum_{i=1}^{n'}\sum_{k=1}^{m'_i} a'_{ik}\, x_{ik}^{-1} \Bigg/ \sum_{i=1}^{n}\sum_{k=1}^{m_i} a_{ik}\, x_{ik}^{-1}\right),\ \forall\, k = 1,2,\ldots,m_i,\ i \in s_I\ .$$

Corollary 1: Under SRSWOR at both the stages of two stage sampling, the proposed product type calibration estimator reduces to

$$\hat{t}_{yCPd} = \left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i} y_{ik}\right)\left(\frac{N}{n'}\sum_{k=1}^{n'}\frac{M_i}{m'_i}\sum_{k=1}^{m'_i} x_{ik}^{-1}\right)\Bigg/\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i} x_{ik}^{-1}\right). \tag{6}$$

Usual product estimator of two stage sampling using double sampling approach is given by

$$\hat{t}_{yPd} = \left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i} y_{ik}\right)\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{n_i}\sum_{k=1}^{n_i} x_{ik}\right)\Bigg/\left(\frac{N}{n'}\sum_{i=1}^{n'}\frac{M_i}{m'_i}\sum_{k=1}^{m'_i} x_{ik}\right). \tag{7}$$

The approximate sampling variance of the proposed estimator $\hat{t}_{yCPd}$ is given by

$$\begin{aligned} AV(\hat{t}_{yCPd}) = {} & \sum_{i=1}^{N}\sum_{j=1}^{N} \Delta_{Iij}\frac{t_{E_{i1}}}{\pi_{Ii}}\frac{t_{E_{j1}}}{\pi_{Ij}} + \sum_{i=1}^{N}\frac{1}{\pi_{Ii}}\sum_{k=1}^{M_i}\sum_{l=1}^{M_i}\Delta_{kl/i}\frac{E_{k/i}}{\pi_{k/i}}\frac{E_{l/i}}{\pi_{l/i}} \\ & + R_1^2\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta'_{Iij}\frac{t_{x^{-1}i}}{\pi'_{Ii}}\frac{t_{x^{-1}j}}{\pi'_{Ij}} + \sum_{i=1}^{N}\frac{1}{\pi'_{Ii}}\sum_{k=1}^{M_i}\sum_{l=1}^{M_i}\Delta'_{kl/i}\frac{x_{ik}^{-1}}{\pi'_{k/i}}\frac{x_{il}^{-1}}{\pi'_{l/i}}\right] \end{aligned} \tag{8}$$

where, $\Delta'_{Iij} = (\pi'_{Iij} - \pi'_{Ii}\pi'_{Ij})$, $\Delta'_{kl/i} = \pi'_{kl/i} - \pi'_{k/i}\pi'_{l/i}$, $t_{x^{-1}i} = \sum_{k=1}^{M_i} x_{ik}^{-1}$ and

$$R_1 = \left(\sum_{i=1}^{N}\sum_{k=1}^{M_i} y_{ik}\right)\Bigg/\left(\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}^{-1}\right).$$

Under SRSWOR design at all the stages and phases, approximate variance of the proposed product type calibration estimator reduces to

$$\begin{aligned} AV(\hat{t}_{yCPd}) = N^2 \Bigg[& \left(\frac{1}{n'} - \frac{1}{N}\right)S_{by}^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{S_{by}^2 + R_1^2 S_{bx^{-1}}^2 - 2R_1 S_{byx^{-1}}\right\} \\ & + \frac{1}{nN}\sum_{i=1}^{N} M_i^2\left\{\left(\frac{1}{m'_i} - \frac{1}{M_i}\right)S_{iy}^2 + \left(\frac{1}{m_i} - \frac{1}{m'}\right)\left(S_{iy}^2 + R_1^2 S_{ix^{-1}}^2 - 2R_1 S_{iyx^{-1}}\right)\right\}\Bigg] \end{aligned} \tag{9}$$

where,

$$\bar{Y}_{N.}=\frac{1}{N}\sum_{i=1}^{N}M_i\bar{Y}_{i.},\ \bar{Y}_{i.}=\frac{1}{M_i}\sum_{k=1}^{M_i}y_{ik},\ \bar{X}_{(-1)\mathrm{N}.}=\frac{1}{N}\sum_{i=1}^{N}M_i\bar{X}_{(-1)i.},\ \bar{X}_{(-1)i.}=\frac{1}{M_i}\sum_{k=1}^{M_i}x_{ik}^{-1},$$

$$S_{by}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{Y}_{i.}-\bar{Y}_{N.}\right)^2,\quad S_{bx^{-1}}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{X}_{(-1)i.}-\bar{X}_{(-1)\mathrm{N}.}\right)^2,$$

$$S_{byx^{-1}}=\frac{1}{N-1}\sum_{i=1}^{N}\left(M_i\bar{Y}_{i.}-\bar{Y}_{N.}\right)\left(M_i\bar{X}_{(-1)i.}-\bar{X}_{(-1)\mathrm{N}.}\right),\ S_{iy}^2=\frac{1}{M_i-1}\sum_{k=1}^{M_i}\left(y_{ik}-\bar{Y}_{i.}\right)^2,$$

$$S_{ix^{-1}}^2=\frac{1}{M_i-1}\sum_{i=1}^{M_i}\left(x_{ik}^{-1}-\bar{X}_{(-1)i.}\right)^2 \text{ and } S_{iyx^{-1}}=\frac{1}{M_i-1}\sum_{i=1}^{M_i}\left(y_{ik}-\bar{Y}_{i.}\right)\left(x_{ik}^{-1}-\bar{X}_{(-1)i.}\right).$$

Following Särndal *et al.* (1992), the estimator of variance of proposed product type calibration estimator can be written as

$$\hat{V}(\hat{t}_{yCPd})=\sum_{i=1}^{n}\sum_{j=1}^{n}d_{Iij}\frac{\hat{t}_{Ei1}}{\pi_{Ii}}\frac{\hat{t}_{Ej1}}{\pi_{Ij}}+\sum_{i=1}^{n}\frac{1}{\pi_{Ii}}\sum_{k=1}^{m_i}\sum_{l=1}^{m_i}d_{kl/i}\frac{e_{k/i}}{\pi_{k/i}}\frac{e_{l/i}}{\pi_{l/i}}$$
$$+\hat{R}_1^2\left[\sum_{i=1}^{n'}\sum_{j=1}^{n'}d'_{Iij}\frac{\hat{t}_{x^{-1}i}}{\pi'_{Ii}}\frac{\hat{t}_{x^{-1}j}}{\pi'_{Ij}}+\sum_{i=1}^{n'}\frac{1}{\pi'_{Ii}}\sum_{k=1}^{m'_i}\sum_{l=1}^{m'_i}d'_{kl/i}\frac{x_{ik}^{-1}}{\pi'_{k/i}}\frac{x_{il}^{-1}}{\pi'_{l/i}}\right] \tag{10}$$

where, $\hat{R}_1=\left(\sum_{i=1}^{n}a_{Ii}\sum_{k=1}^{m_i}a_{k/i}\,y_{ik}\right)\Big/\left(\sum_{i=1}^{n}a_{Ii}\sum_{k=1}^{m_i}a_{k/i}\,x_{ik}^{-1}\right),\ \hat{t}_{x^{-1}i}=\sum_{k=1}^{m_i}\frac{x_{ik}^{-1}}{\pi_{k/i}}$,

$$d'_{Iij}=\frac{(\pi'_{Iij}-\pi'_{Ii}\pi'_{Ij})}{\pi'_{Iij}} \text{ and } d'_{kl/i}=\frac{\pi'_{kl/i}-\pi'_{k/i}\pi'_{l/i}}{\pi'_{kl/i}}.$$

Under SRSWOR design at all the stages and phases, it reduces to

$$\hat{V}(\hat{t}_{yCPd})=N^2\left[\left(\frac{1}{n'}-\frac{1}{N}\right)\hat{S}_{by}^2+\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{\hat{S}_{by}^2+\hat{R}_1^2\hat{S}_{bx^{-1}}^2-2\hat{R}_1\hat{S}_{byx^{-1}}\right\}\right.$$
$$\left.+\frac{1}{nN}\sum_{i=1}^{n}M_i^2\left\{\left(\frac{1}{m'_i}-\frac{1}{M_i}\right)\hat{S}_{iy}^2+\left(\frac{1}{n_i}-\frac{1}{m'_i}\right)\left(\hat{S}_{iy}^2+\hat{R}_1^2\hat{S}_{ix^{-1}}^2-2\hat{R}_1\hat{S}_{iyx^{-1}}\right)\right\}\right] \tag{11}$$

where, $\hat{R}_1=\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i}y_{ik}\right)\Big/\left(\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{k=1}^{m_i}x_{ik}^{-1}\right),\quad \hat{S}_{by}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{y}_{i.}-\bar{y}_{n.}\right)^2,$

$$\hat{S}_{bx^{-1}}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{x}_{(-1)i.}-\bar{x}_{(-1)n.}\right)^2,\quad \hat{S}_{byx^{-1}}=\frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{y}_{i.}-\bar{y}_{n.}\right)\left(M_i\bar{x}_{(-1)i.}-\bar{x}_{(-1)n.}\right),$$

$$\hat{S}_{iy}^2=\frac{1}{m_i-1}\sum_{i=1}^{m_i}\left(y_{ik}-\bar{y}_{i.}\right)^2,\ \hat{S}_{ix^{-1}}^2=\frac{1}{m_i-1}\sum_{i=1}^{m_i}\left(x_{ik}^{-1}-\bar{x}_{(-1)i.}\right)^2,$$

$$\hat{S}_{iyx^{-1}}=\frac{1}{m_i-1}\sum_{i=1}^{m_i}\left(y_{ik}-\bar{y}_{i.}\right)\left(x_{ik}^{-1}-\bar{x}_{(-1)i.}\right),\ \bar{y}_{i.}=\frac{1}{m_i}\sum_{k=1}^{m_i}y_{ik},\quad \bar{y}_{n.}=\frac{1}{n}\sum_{i=1}^{n}M_i\bar{y}_{i.},$$

$$\bar{x}_{(-1)i.}=\frac{1}{m_i}\sum_{k=1}^{m_i}x_{ik}^{-1},\quad \bar{x}_{(-1)n.}=\frac{1}{n}\sum_{i=1}^{n}M_i\bar{x}_{(-1)i.}.$$

4. Simulation study

In order to evaluate the statistical performance of proposed product type calibration estimators, a simulation study was carried out following double sampling approach in two stage sampling. SRSWOR is used for sample selection at both stages and the size of the PSU and the corresponding SSUs were assumed to be fixed. First, a finite population of 5000 units were generated in the similar way of Biswas *et al.* (2020). Let, number of PSU, N=50 and PSU size, M_i=100. The finite population was generated from the following model as

$$y_k = \beta x_k^{-1} + e_k,\ k = 1,...,M_0 \tag{12}$$

where, $M_0 = \sum_{i=1}^{N} M_i$.

The distribution of auxiliary variable was considered as normal distribution as $x_k \sim N(5,\ 1)$ and the random errors, e_k, $k = 1,...,M_0$, are taken from normal distribution as $e_k \sim N(0,\ \sigma^2 x_k^{-1})$. The value of β has been fixed as 20. Four different values for σ^2 as 0.25, 1.0, 2.0 and 5.0 are taken. In this way, four sets of population have been generated denoted by **Set 1**, **2**, **3** and **4**, with different values of correlation coefficient between Y and X as -0.91, -0.85, -0.78 and -0.64 respectively. Then, from each of the study population sets, we have selected a total of 10000 different samples of different sizes of double samples under two stage sampling were drawn from the populations sets as given below

n' =10, m'_i =20, n =4, m_i =6	n' =10, m'_i =25, n =4, m_i =8
n' =15, m'_i =25, n =6, m_i =8	n' =15, m'_i =30, n =6, m_i =10
n' =20, m'_i =30, n =8, m_i =10	n' =20, m'_i =40, n =8, m_i =12
n' =25, m'_i =40, n =10, m_i =12	n' =25, m'_i =50, n =10, m_i=15

Developed product type calibration estimators as well as all other usual estimators of population total using double sampling approach under two stage sampling were evaluated based on 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

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

where, $\hat{\theta}_i$ are the estimates of population parameter θ for the character under study obtained at i^{th} sample in the simulation study.

Table 1, 2, 3 and 4 present the results of the simulation study under population Set 1, 2, 3 & 4 in terms of %RB and %RRMSE of the proposed product type calibration estimator ($\hat{t}_{yCPd}$) and usual NHT estimator ($\hat{t}_{y\pi}$) and product estimator ($\hat{t}_{yPd}$) of the population total under two stage sampling design using double sampling approach. These estimators have been calculated assuming that the complete auxiliary information (x_{ik}) was not available at the SSU level in the population and auxiliary variable is inversely related to the character under study.

Table 1: Comparison of all the estimators with respect to %RB and %RRMSE under two stage sampling in case of population Set 1 having correlation coefficient (ρ) as -0.91 using double sampling approach

Sample size ($n'\_m'_i\_n\_m_i$)	% RB			% RRMSE		
	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$
10_20_4_6	0.014	-0.010	-0.152	4.896	1.963	2.559
10_25_4_8	0.052	0.001	-0.092	4.261	1.723	2.256
15_25_6_8	0.011	0.004	-0.083	3.532	1.402	1.839
15_30_6_10	-0.005	-0.003	-0.058	3.101	1.248	1.639
20_30_8_10	-0.005	0.007	-0.039	2.676	1.083	1.432
20_40_8_12	0.019	0.009	-0.037	2.434	0.933	1.258
25_40_10_12	-0.004	0.003	-0.027	2.178	0.811	1.117
25_50_10_15	-0.022	-0.004	-0.033	1.950	0.716	0.996

Table 2: Comparison of all the estimators with respect to %RB and %RRMSE under two stage sampling in case of population Set 2 having correlation coefficient (ρ) as -0.85 using double sampling approach

Sample size ($n'\_m'_i\_n\_m_i$)	% RB			% RRMSE		
	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$
10_20_4_6	0.058	0.026	-0.123	5.286	2.756	3.195
10_25_4_8	0.109	0.048	-0.056	4.499	2.406	2.785
15_25_6_8	-0.046	-0.047	-0.122	3.694	1.939	2.248
15_30_6_10	-0.030	-0.011	-0.081	3.297	1.730	2.014
20_30_8_10	-0.003	-0.027	-0.061	2.847	1.490	1.739
20_40_8_12	0.021	0.035	-0.005	2.585	1.346	1.570
25_40_10_12	-0.008	0.016	-0.017	2.313	1.173	1.386
25_50_10_15	-0.041	-0.021	-0.045	2.065	1.040	1.219

Table 3: Comparison of all the estimators with respect to %RB and %RRMSE under two stage sampling in case of population Set 3 having correlation coefficient (ρ) as -0.78 using double sampling approach

Sample size ($n'\_m'_i\_n\_m_i$)	% RB			% RRMSE		
	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$
10_20_4_6	-0.095	-0.008	-0.167	5.699	3.569	3.892
10_25_4_8	-0.018	-0.002	-0.135	4.952	3.127	3.410
15_25_6_8	0.002	-0.013	-0.087	4.012	2.506	2.743
15_30_6_10	-0.020	0.021	-0.047	3.612	2.231	2.448
20_30_8_10	0.025	-0.002	-0.042	3.095	1.945	2.137
20_40_8_12	-0.031	-0.001	-0.054	2.853	1.762	1.930
25_40_10_12	-0.016	0.001	-0.040	2.537	1.543	1.687
25_50_10_15	0.012	0.006	-0.020	2.243	1.359	1.509

Table 4: Comparison of all the estimators with respect to %RB and %RRMSE under two stage sampling in case of population Set 4 having correlation coefficient (ρ) as -0.64 using double sampling approach

Sample size $(n'\_m'_i\_n\_m_i)$	% RB			% RRMSE		
	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$	$\hat{t}_{y\pi d}$	$\hat{t}_{yCPd}$	$\hat{t}_{yPd}$
10_20_4_6	0.048	0.080	-0.094	6.791	5.230	5.425
10_25_4_8	-0.015	0.052	-0.083	5.971	4.612	4.791
15_25_6_8	-0.019	-0.012	-0.100	4.786	3.679	3.804
15_30_6_10	0.003	0.032	-0.030	4.228	3.293	3.422
20_30_8_10	-0.020	-0.020	-0.083	3.745	2.884	2.988
20_40_8_12	-0.031	0.001	-0.056	3.372	2.565	2.671
25_40_10_12	0.001	-0.016	-0.049	2.991	2.275	2.367
25_50_10_15	0.005	-0.009	-0.035	2.694	2.051	2.129

From Table 1 it is notable that the proposed product type calibration estimator of the finite population total was giving consistently least amount %RB compared to their usual NHT and product estimator using double sampling approach for the Poulation Set 1 where correlation coefficient (ρ) was -0.91. Here, it is assumed that the auxiliary information was unavailable at SSU level and auxiliary variable is inversely related with the character under study. It was also seen that the proposed product type calibration estimator of the population total is always more efficient than the NHT and product estimators, since %RRMSE of the proposed product type calibration estimator is always least at different sample size combinations. It can also be seen that the %RRMSE of the proposed product type calibration estimator was decreasing with increase of sample sizes, thus, it provides a consistent estimator of the finite population total. Similar trend in simulation results can be observed in Table 2, 3 and 4, where Population Set 2, 3 and 4 are considered for simulation in which correlation coefficient (ρ) were -0.85, -0.78 and -0.64 respectively. Close look of Table 2, 3 and 4 reveals that %RRMSE of the proposed product type calibration estimator was decreasing with the increase in the amount of negative correlation.

5. Conclusions and way forward

In general, the Calibration Approach assumes the availability of population-level auxiliary information. On the contrary, in large scale surveys involving two stage sampling, it is often the case that population-level data on auxiliary variable is not available in practice, but relatively inexpensive to collect. In the present article, in case of non-availability of population-level relatively inexpensive data on auxiliary variable in two stage sampling, product type calibration estimator (Equation 5) of the finite population total has been proposed using double sampling approach when there exist inverse relation between auxiliary variable and chracter under study. In order to study the statistical performance of proposed product type calibration estimator as compared to existing estimators of population total of character under study, a simulation study was conducted. The simulation results also suggests that the proposed product type calibration estimators using double sampling approach in two stage sampling, performs better than usual Narain-Horvitz-Thompson estimator (Equation 2) and product estimator (Equation 7) of the finite population total with respect to %RB and %RRMSE. In future, investigation may be carried out to extent the work in case of different type of varying probability sampling schemes in oder to improve several well-known estimators *viz.* Narain-Horvitz-Thompson estimator, Rao, Hartley and Cochran (1962) estimator etc.

Acknowledgement

Authors are profoundly thankful to the editors and the anonymous reviewer for their productive comments which led to improvements in the article.

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**, 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**, 219–226.

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

Arnab, R. (2017). *Survey Sampling Theory and Applications*, Academic Press, Oxford.

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**, 1–6.

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**, 23–32.

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.

Haziza, D., Hidiroglou, M. A. and Rao, J. N. K. (2011). Comparison of variance estimators in two-phase sampling: An empirical investigation. *Pakistan Journal of Statistics*, **27**, 477-492.

Hidiroglou, M. A., Rao, J. N. K. and Haziza, D. (2009). Variance estimation in two-phase sampling. *Australian and New Zealand Journal of Statistics*, **51**, 127-141.

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

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

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**, 550-562.

Murthy, M. N. (1964). Product method of estimation. *Sankhya*, **26A**, 69–74.

Narain, R. D. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Aricultural Statistics*, **3**, 169-174.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, **33**, 101-116.

Rao, J. N. K. (1973). On double sampling for stratification and analytical surveys. *Biometrika*, **60**, 125- 133.

Rao, J. N. K., Hartley, H. O. and Cochran, W. G. (1962). On a simp,e procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society*, Series B, **24**, 482-491.

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

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

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

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**, 199-212.

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**, DOI: 10.1080/03610926.2015.1091083.

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**, 1-11.

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**, 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.

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.