Product Type Calibration Estimators of Finite Population Total under Two Stage Sampling

Ankur Biswas, Kaustav Aditya, U.C. Sud and Pradip Basak
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 19 December 2018; Revised 10 October 2019; Accepted 18 October 2019

SUMMARY
The Calibration Approach proposed by Deville and Särndal (1992) is a popular technique to efficiently use auxiliary information in survey sampling. In this study, calibration estimators of the finite population total have been developed under two stage sampling design along with variance of the estimator and the corresponding estimator of variance. It is assumed that the population level complex auxiliary information is available at the second stage of selection and the study variable is inversely related to the available auxiliary information. The proposed calibration estimators were evaluated through a simulation study and it was found that all the proposed product type calibration estimators perform better than the Horvitz-Thompson estimator as well as usual product estimator of the population total under two stage sampling design.

Keywords: Calibration approach; Auxiliary information; Product estimator; Simulation.

1. INTRODUCTION
In sample surveys, auxiliary information on the finite population is often used to increase the precision of estimators of unknown finite population parameters of study variable. In the simplest settings, ratio and regression estimators incorporate known finite population parameters of auxiliary variables in estimation of study variable parameters. The Calibration Approach, proposed by Deville and Särndal (1992), is one of the widely used techniques for incorporation of auxiliary information in estimation stages of survey sampling. In fact, the generalized regression estimator (GREG) (Cassel et al., 1976) is a special case of the calibration estimator choosing the Chi-square distance function (Deville and Särndal, 1992). Calibration technique implies that a set of initial weights (usually the sampling design weights) are transformed into a set of new weights, called calibrated weights, which is the product of its initial weight and a calibration factor. In the past few decades, calibration estimation has gained significant attention not only in the field 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) etc. Kim and Park (2010) and Särndal (2007) provided comprehensive review on calibration approach.

In many medium to large scale surveys, it is very often the case that the sampling frame is often unavailable or it could be too expensive to construct one. Also, the population could be spread over a wide area entailing very high operational expenses for personal interviews and supervisions. Two stage sampling serves as a solution in such situations where groups of elements, called primary stage units (PSU), are selected first and, then, a sample of elements, called secondary stage units (SSU), are selected from each selected PSU. For example, in agricultural surveys, villages can be selected as PSU and farmers as SSU. Estimation of the population parameters in two stage sampling using auxiliary information has been well addressed in survey sampling. Sukhatme et al. (1984) suggested regression estimator...
of the population mean in two-stage sampling. Särndal et al. (1992) considered three different situations with respect to availability of complex auxiliary variable under two stage sampling and discussed extensively on ratio and regression estimators under such situations. Aditya et al. (2016a, 2016b) and Mourya et al. (2016) extended the calibration estimation under different cases of availability of complex auxiliary information under two stage sampling. Sinha et al. (2016) proposed calibration estimators for estimating population mean under stratified sampling and stratified double sampling. Aditya et al. (2017) attempted to use calibration approach for estimation of crop yields at the district level under two-stage sampling. Basak et al. (2017) proposed a calibration estimator of finite population regression coefficient under two-stage sampling design. Veronica et al. (2018) considered computation of calibration weights at both the first and second stages of sample selection for estimation of population mean by assuming the population means of auxiliary variables are known at both the stages of sample selection under equal probability two-stage sampling.

It was observed that most of the work related to calibration estimation for the finite population parameters were mostly restricted with the assumption of linear relationship between the study variable and the auxiliary variable. There may be situations when it can be seen that the study variable is inversely related to the auxiliary variable. For instance, an inverse relation, generally, exists between the age of individuals and hours of sleep (Sud et al., 2014a). Again, in household surveys, it is often the case that marketable surplus is inversely related to family consumption for seed, feed etc. In these situations, the product estimator, proposed by Murthy (1964), is a feasible alternative. In that situation the usual methodology for calibration estimation may not fit in. Sud et al. (2014a, 2014b) studied the calibration approach for estimation of population total when variable of interest and auxiliary information have inverse relation under uni-stage equal probability sampling. However, multi-stage designs are most prevalent in medium to large scale surveys. Therefore, in this present study, an attempt has been made to develop calibration estimators of finite population total under two stage sampling when study variable is inversely related to the auxiliary variable.

In Section 2, proposed product type calibration estimators of finite population total under two stage sampling has been discussed. In order to study the statistical properties of proposed estimators empirically, a simulation study was carried out. Details of simulation study and discussion on simulation results are given in Section 3 and 4 respectively. Section 5 comprises concluding remarks.

2. PROPOSED CALIBRATION ESTIMATORS UNDER TWO STAGE SAMPLING DESIGN

In this section, two different calibration estimates are proposed under two stage sampling design under the assumption that available auxiliary information is inversely related to the study variable. The proposed estimators were developed with the assumption of availability of auxiliary information at SSU level under two stage sampling. Let, $U$ be the finite population under consideration and $Y$ be the character under study. $U$ is grouped into $N$ different PSUs such that $U = \{1, \ldots, N\}$ and $i^{th}$ PSU consists of $M_i$ SSUs such that $U_i = \{1, \ldots, k, \ldots, M_i\}$, $i \in U$. Thus, we have $U = \bigcup_{i=1}^{N} U_i$ and total number of SSUs in the population $U$ is $M_0 = \sum_{i=1}^{N} M_i$. Under two stage sampling, at stage one, a sample of PSUs, $s_i$, of size $n$ PSUs is selected from $U_i$ according to a specified design $p_i(\cdot)$ with $\pi_{i}^{s} = P(i \in s_f)$ and $\pi_{ij}^{s} = P(i, j \in s_f)$ as the inclusion probabilities at the PSU level. Given that the PSU $U_i$ is selected at the first stage, a sample $s_i$ of size $m_i$ SSUs is drawn from $U_i$ according to some specified design $p_i(\cdot)$ with inclusion probabilities $\pi_{k/i}^{s} = P(k \in s_i / i \in s_f)$ and $\pi_{kl/i}^{s} = P(k, l \in s_i / i \in s_f)$ at the PSU level. In the second stage of sampling, invariance and independence property is followed. The entire sample of elements is defined as, $s = \bigcup_{i=1}^{s} s_i$. Let, $y_{ik}$ denotes the observation of the study variable from $k^{th}$ SSU in $i^{th}$ PSU and it is observed for all $k \in s$. 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. An attempt has been made to improve the ordinary Horvitz-Thompson (1952) estimator for population total as given by
\[ \hat{y}_T = \sum_{i=1}^{n} a_{li} \sum_{k=1}^{m_i} a_{k/i} y_{ik} = \sum_{i=1}^{n} m_i a_{ik} y_{ik} \]  

(2.1)

where, the design weights are given as

\[ a_{li} = 1/ \pi_{li}, \quad \forall \ i \in s_I, \quad a_{k/i} = 1/ \pi_{k/i}, \quad \forall \ k \in s_i, \quad i \in s_I \text{ and } a_{ik} = a_{li} a_{k/i}. \]

Under two stage sampling design, the complex auxiliary information may be available for the PSUs as well as the SSUs within the PSUs (Särndal et al., 1992). In the present study, as per availability of complex auxiliary information at the ultimate stage units following two cases have been considered under two stage sampling design

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 PSUs.

2.1 Case 1: Population level complete auxiliary information is available at SSU level

Under this case, it has been assumed that population level complete auxiliary information is available at the unit (SSU) level i.e. the auxiliary information of \( k^{th} \) SSU in \( i^{th} \) PSU, \( x_{ik} \), is known for all elements \( k \in U \).

A correct value of \( \sum_{i=1}^{n} \sum_{k=1}^{m} x_{ik}^{-1} \) is assumed to be known.

In addition, there exist an inverse relationship between the study variable \( Y \) and the auxiliary variable \( X \).

Using the well-known Calibration Approach (Deville and Särndal, 1992), we wish to modify the total design weight of \( k^{th} \) SSU of \( i^{th} \) PSU, i.e. \( a_{ik} = a_{li} a_{k/i} \), as given in the HT estimator of population total in Equation 2.1. The proposed product type calibration estimator of population total under Case 1 is given by

\[ \hat{y}_{CP1} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{ik} y_{ik} \]  

(2.1.1)

where, \( w_{ik} \) is the calibrated weight corresponding to the design weight \( a_{ik} \) under Case 1.

In order to obtain the calibrated weight \( w_{ik} \), we minimized the Chi-square type distance

\[ \sum_{i=1}^{n} \sum_{k=1}^{m_i} \left( \frac{w_{ik} - a_{ik}}{a_{ik} q_{ik}} \right)^2 \]  

subject to the calibration constraint

\[ \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{ik} x_{ik}^{-1} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} x_{ik}^{-1}, \]  

where \( a_{ik} \) are suitably chosen constants. Using the method of Lagrange multiplier, by minimizing

\[ \varphi(w_{ik}, \lambda) = \sum_{i=1}^{n} \sum_{k=1}^{m_i} \left( \frac{w_{ik} - a_{ik}}{a_{ik} q_{ik}} \right)^2 - \lambda \left( \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{ik} x_{ik}^{-1} - \sum_{i=1}^{n} \sum_{k=1}^{m_i} x_{ik}^{-1} \right) \]

the calibrated weights are obtained as given by

\[ w_{ik} = a_{ik} + a_{ik} q_{ik} x_{ik}^{-1} \left[ \sum_{i=1}^{n} \sum_{k=1}^{m_i} x_{ik}^{-1} - \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1} \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} q_{ik} x_{ik}^{-1} \right], \]

\( \forall k = 1, 2, \ldots, m_i \) and \( \forall i = 1, 2, \ldots, n \)

(2.1.2)

Using the results of the Equation (2.1.2) in (2.1.1) considering \( q_{ik} = x_{ik} \), we have therefore proved the following result.

**Theorem 1:** Under Case 1 of two stage sampling, if we consider the calibrated design weights as

\[ w_{ik} = a_{ik} \left( \sum_{i=1}^{n} \sum_{k=1}^{m_i} x_{ik}^{-1} \right) / \left( \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1} \right), \]  

\( \forall k = 1, 2, \ldots, m_i \),

then the proposed product type calibration estimator of population total is given as

\[ \hat{y}_{CP1} = \sum_{i=1}^{n} \sum_{k=1}^{m_i} w_{ik} y_{ik} \]

(2.1.3)

\[ = \left( \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} y_{ik} \right) \left( \sum_{i=1}^{n} \sum_{k=1}^{m_i} x_{ik}^{-1} \right) / \left( \sum_{i=1}^{n} \sum_{k=1}^{m_i} a_{ik} x_{ik}^{-1} \right). \]

**Corollary 1:** Under an equal probability without replacement sampling design (Simple Random Sampling without replacement (SRSWOR)) at both the stages of two stage sampling, the proposed product type calibration estimator under Case 1 reduces to

\[ \hat{y}_{CP1} = \left( \frac{N}{n} \sum_{i=1}^{n} M_i \sum_{k=1}^{m_i} y_{ik} \right) / \left( \sum_{i=1}^{n} M_i \sum_{k=1}^{m_i} x_{ik}^{-1} \right). \]

(2.1.4)

The theoretical bias of the proposed product type calibration estimator \( \hat{y}_{CP1} \) is obtained through Taylor series linearization technique as
Bias(i\text{y}\text{CP}_1) = \left[ t_y \left( \frac{1}{n} - \frac{1}{N} \right) \left( \rho_b C_{by} C_{bx} + c_{bx}^2 \right) + \frac{1}{n} \left( \bar{P}_w C_{wy} C_{wx} + c_{wx}^2 \right) \right] (2.1.6)

where,

\rho_b = \frac{S_{bxy}}{S_{bx} S_{by}}, \quad c_{by}^2 = \frac{S_{by}^2}{\overline{y}_N^2}, \quad c_{bx}^2 = \frac{S_{bx}^2}{\overline{x}_N^2},

\bar{P}_w = \frac{S_{wy}}{S_{wx} S_{wy}}, \quad c_{wy}^2 = \frac{S_{wy}^2}{\overline{y}_N^2}, \quad c_{wx}^2 = \frac{S_{wx}^2}{\overline{x}_N^2},

\overline{x}_N = \frac{1}{N} \sum_{i=1}^{N} M_i \overline{x}_i, \quad \overline{y}_N = \frac{1}{N} \sum_{i=1}^{N} M_i \overline{y}_i,

\overline{x}_i = \frac{1}{M_i} \sum_{k=1}^{M_i} x_{ik}, \quad \overline{y}_i = \frac{1}{M_i} \sum_{k=1}^{M_i} y_{ik},

S_{bxy} = \frac{1}{N-1} \sum_{i=1}^{N} (M_i \overline{y}_i - \overline{y}_N) (M_i \overline{x}_i - \overline{x}_N),

S_{bxy}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (M_i \overline{y}_i - \overline{y}_N)^2,

S_{bx}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (M_i \overline{x}_i - \overline{x}_N)^2,

S_{wxy} = \frac{1}{N} \sum_{i=1}^{N} M_i^2 \left( \frac{1}{m_i} - \frac{1}{M_i} \right) S_{ixy},

S_{wy}^2 = \frac{1}{N} \sum_{i=1}^{N} M_i^2 \left( \frac{1}{m_i} - \frac{1}{M_i} \right) S_{iy},

S_{wx}^2 = \frac{1}{N} \sum_{i=1}^{N} M_i^2 \left( \frac{1}{m_i} - \frac{1}{M_i} \right) S_{ix},

S_{ixy} = \frac{1}{M_i-1} \sum_{k=1}^{M_i} (x_{ik} - \overline{x}_i) (y_{ik} - \overline{y}_i),

\rho = \frac{S_{ixy}}{S_{ix} S_{iy}}, \quad \rho < - \frac{1}{2}, \quad c_{wy}^2 = \frac{S_{wy}^2}{\overline{y}_N^2}, \quad \rho < - \frac{1}{2}.

Usual product estimator under Case 1 of two stage sampling considering SRSWOR at both the stages is given by

\hat{t}_y = \left( \frac{1}{n} - \frac{1}{N} \right) \rho_b C_{by} C_{bx} + \frac{1}{n} \bar{P}_w C_{wy} C_{wx}

(2.1.7)

and, its bias is given as

\text{Bias}(\hat{t}_y) = t_y \left( \frac{1}{n} - \frac{1}{N} \right) \rho_b C_{by} C_{bx} + \frac{1}{n} \bar{P}_w C_{wy} C_{wx}

(2.1.8)

It has been found that under SRSWOR at both the stages of a two stage sampling design under Case 1, product estimator \( \hat{t}_y \) is better than usual HT estimator \( \hat{t}_y \) if \( \rho_b C_{by} < - \frac{1}{2} \) and \( \bar{P}_w C_{wy} < - \frac{1}{2} \). Under these conditions in two stage sampling design, it can be seen that

\[ \text{Bias}(\hat{t}_y) \leq \text{Bias}(\hat{t}_y) \]

Following Deville and Särndal (1992) and Särndal et al. (1992), the approximate variance of the proposed product type calibration estimator under Case 1 by first order Taylor series linearization technique was obtained as

\[ AV(\hat{t}_y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta_{ij} \frac{t_{E_i j} t_{E_i j}}{\pi_i \pi_j} + \sum_{i=1}^{N} \sum_{k=1}^{M_i} \frac{M_i}{\pi_i} K_{i k} \frac{E_{i k i}}{\pi_i \pi_k} \]

(2.1.9)

where,

\[ t_{E_i j} = \sum_{k=1}^{M_i} E_{k i} \pi_i, \quad E_{k i} = y_{ik} - \frac{\sum_{i=1}^{N} M_i \sum_{k=1}^{M_i} y_{ik}}{\sum_{i=1}^{N} M_i \sum_{k=1}^{M_i} x_{ik}} \]

\[ \Delta_{ij} = (\pi_i \pi_j - \pi_i \pi_j), \quad K_{i k} = \pi_i \pi_k - \pi_i \pi_k \]

Under SRSWOR design at both the stages the approximate variance reduces to
\[ AV(\hat{t}_j_{CP1}) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) \left( S^2_{by} + R^2 S^2_{hx} - 2R_i S_{byx}^{-1} \right) + \frac{N}{n} \sum_{i=1}^{N} M_i \left( \frac{1}{m_i} - \frac{1}{M_i} \right) \left( S^2_{ij} + R^2 S^2_{ix} - 2R_i S_{ixy}^{-1} \right) \]

where,
\[ R_i = \left( \frac{\sum_{i=1}^{N} M_i y_{ik}}{\sum_{i=1}^{N} M_i} \right) \bigg/ \left( \frac{\sum_{i=1}^{N} M_i x_{ik}}{\sum_{i=1}^{N} M_i} \right) \]

\[ \bar{X}_{(-1)i} = \frac{1}{M_i} \sum_{i=k=1}^{N} x_{ik}, \quad \bar{X}_{(-1)N} = \frac{1}{N} \sum_{i=1}^{N} M_i \bar{X}_{(-1)i}, \]

\[ S^2_{byx} = \frac{1}{N-i-1} \sum_{i=1}^{N} \left( M_i \bar{Y}_i - \bar{Y}_N \right) \left( M_i \bar{X}_{(-1)i} - \bar{X}_{(-1)N} \right), \]

\[ S^2_{bx} = \frac{1}{N-i-1} \sum_{i=1}^{N} \left( M_i \bar{X}_{(-1)i} - \bar{X}_{(-1)N} \right)^2, \]

\[ S^2_{ixy} = \frac{1}{M_i} \sum_{i=1}^{M_i} \left( y_{ik} - \bar{Y}_i \right) \left( x_{ik} - \bar{X}_{(-1)i} \right), \]

\[ S^2_{ix} = \frac{1}{M_i} \sum_{i=1}^{M_i} \left( x_{ik} - \bar{X}_{(-1)i} \right)^2. \]

Following Särndal et al. (1992), the Yates–Grundy form of estimator of variance (Yates and Grundy, 1953) of the proposed product type calibration estimator under Case 1 is given by

\[ \hat{V}_{YG}(\hat{t}_j_{CP1}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} \left( \hat{E}_{ih} - \hat{E}_{ij} \right)^2 + \frac{1}{2} \sum_{j=1}^{n} \left( \sum_{k=1}^{m_i} \sum_{l=1}^{m_i} d_{kl/i}(w_{ik} e_{k/i} - w_{il} e_{l/i}) \right)^2 \]

where,
\[ \hat{E}_{hi} = \sum_{k=1}^{m_i} e_{k/i}, \]
\[ e_{k/i} = y_{ik} - \left( \sum_{i=1}^{n} a_{ik} y_{ik} / \sum_{i=1}^{n} a_{ik} x_{ik} \right) x_{ik}, \]
\[ d_{ij} = \frac{(\pi_{il}/\pi_{lj} - \pi_{ij})}{\pi_{lij}} \quad \text{and} \quad d_{kl/i} = \frac{\pi_{ikl}/\pi_{i/l} - \pi_{kl/i}}{\pi_{kl/i}}. \]

Under SRSWOR design at both the stages the estimator of variance reduces to

\[ \hat{V}(\hat{t}_j_{CP1}) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) \left( S^2_{by} + R^2 S^2_{bx} - 2R_i S_{byx}^{-1} \right) + \frac{N}{n} \sum_{i=1}^{N} M_i \left( \frac{1}{m_i} - \frac{1}{M_i} \right) \left( S^2_{ij} + R^2 S^2_{ix} - 2R_i S_{ixy}^{-1} \right) \]

where,
\[ R_i = \left( \frac{\sum_{i=1}^{N} M_i y_{ik}}{\sum_{i=1}^{N} M_i} \right) \bigg/ \left( \frac{\sum_{i=1}^{N} M_i x_{ik}}{\sum_{i=1}^{N} M_i} \right) \],

\[ S^2_{byx} = \frac{1}{N-i-1} \sum_{i=1}^{N} \left( M_i \bar{Y}_i - \bar{Y}_N \right) \left( M_i \bar{X}_{(-1)i} - \bar{X}_{(-1)N} \right), \]

\[ S^2_{bx} = \frac{1}{N-i-1} \sum_{i=1}^{N} \left( M_i \bar{X}_{(-1)i} - \bar{X}_{(-1)N} \right)^2, \]

\[ S^2_{ixy} = \frac{1}{M_i} \sum_{i=1}^{M_i} \left( y_{ik} - \bar{Y}_i \right) \left( x_{ik} - \bar{X}_{(-1)i} \right), \]

\[ S^2_{ix} = \frac{1}{M_i} \sum_{i=1}^{M_i} \left( x_{ik} - \bar{X}_{(-1)i} \right)^2. \]

2.2 Case 2: Population level auxiliary information is available only for the selected PSUs

In this case, it has been assumed that the population level auxiliary information is available at the SSU level only for the selected PSUs i.e. the auxiliary information is known for all the SSUs within the PSU \( i \in s_j \). The correct value of \( \sum_{k=1}^{m_i} x_{ik}^{-1} \) is assumed to be available for each \( i^{th} \) sampled PSU. Suppose, there exist inverse
relationship between the study variable \(Y\) and the auxiliary variable \(X\). Using well-known Calibration Approach (Deville and Särndal, 1992), the design weight at the second stage \(a_{k/i}\) has been revised. The proposed product type calibration estimator of population total under Case 2 is given by

\[
\hat{t}_{yCP2} = \sum_{i=1}^{n} a_{li} \sum_{k=1}^{m_i} w_{2ik} y_{ik} \tag{2.2.1}
\]

where, \(w_{2ik}\) is the calibrated weight corresponding to the design weight \(a_{k/i}\).

In this situation, we minimized the Chi-square type distance function \(\sum_{k=1}^{m_i} (w_{2ik} - a_{k/i}/q_{ik})^2\) subject to \(\sum_{k=1}^{m_i} w_{2ik} x_{ik}^{-1} = \sum_{k=1}^{m_i} x_{ik}^{-1}\), where \(q_{ik}\) are suitably chosen constants. Using Lagrange multiplier technique, by minimizing

\[
\varphi(w_{2ik}, \lambda) = \sum_{k=1}^{m_i} (w_{2ik} - a_{k/i}/q_{ik})^2 - \lambda \left( \sum_{k=1}^{m_i} w_{2ik} x_{ik}^{-1} - \sum_{k=1}^{m_i} x_{ik}^{-1} \right),
\]

the new set of calibrated weights is obtained as

\[
w_{2ik} = a_{k/i} + a_{k/i} q_{ik} x_{ik}^{-1}
\]

\(\forall k = 1, 2, \ldots, m_i\) \tag{2.2.2}

Using the results of the Equation (2.2.2) in (2.2.1) considering \(q_{ik} = x_{ik}\), we have therefore proved the following result.

**Theorem 2:** Under Case 2 of two stage sampling, if we consider the calibrated design weights as

\[
w_{2ik} = a_{k/i} \left( \sum_{k=1}^{m_i} x_{ik}^{-1} / \sum_{k=1}^{m_i} a_{k/i} x_{ik}^{-1} \right), \quad \forall k = 1, 2, \ldots, m_i
\]

then the proposed product type calibration estimator of population total is given as

\[
\hat{t}_{yCP2} = \sum_{i=1}^{n} a_{li} \sum_{k=1}^{m_i} w_{2ik} y_{ik}
\]

Using Taylor series linearization technique, its bias is obtained as

\[
\text{Bias}(\hat{t}_{yCP2}) = t_y \left[ \frac{1}{n} (C_{wy} C_{wx} + C_{wx}^2) \right] \tag{2.2.5}
\]

where, the terms are as defined in Case 1 (Eqn. 2.1.6).

Usual product estimator under Case 2 of two stage sampling with SRSWOR at both the stages is given by

\[
\hat{t}_{yP2} = \left( \frac{n}{n} \sum_{i=1}^{n} m_i \sum_{k=1}^{m_i} y_{ik} \left( \frac{M_i}{m_i} \sum_{i=1}^{m_i} x_{ik}^{-1} \right) / \left( \frac{M_i}{m_i} \sum_{i=1}^{m_i} x_{ik}^{-1} \right) \right)
\]

\(\forall k = 1, 2, \ldots, m_i\) \tag{2.2.6}

and, using Taylor series linearization technique, its bias is given by

\[
\text{Bias}(\hat{t}_{yP2}) = t_y \left[ \frac{1}{n} \bar{C}_{wy} C_{wx} \right] \tag{2.2.7}
\]

It has been found that under SRSWOR at both the stages of two stage sampling design under Case 2, product estimator \(\hat{t}_{yP2}\) is better than usual HT estimator \(\hat{t}_{y\pi}\) if \(\bar{C}_{wy} C_{wx} < -\frac{1}{2}\). Under this condition in two stage sampling design, it can be seen that

\[
|\text{Bias}(\hat{t}_{yCP2})| \leq |\text{Bias}(\hat{t}_{yP2})|
\]

Following Särndal et al. (1992) the approximate variance of the proposed product type calibration estimator under Case 2 by first order Taylor series linearization technique was obtained as
Under SRSWOR design at both the stages, it reduces to

\[
AV(\hat{\iota}_{CP2}) = \frac{N}{n} \left\{\sum_{i=1}^{n} \left( \sum_{j=1}^{N} \Delta_{ij} - \frac{1}{n} \right) \pi_{li} \pi_{lj} \right\} + \frac{1}{n} \left( \sum_{i=1}^{n} \left( \frac{1}{M_i} - \frac{1}{n} \right) \left( \hat{R}_{i}^2 \right) \right) \text{,} \quad (2.2.8)
\]

where,

\[
i_{E_{i2}} = \sum_{k=1}^{M_i} E_{k/i} \pi_{k|i}, \quad E_{k/i} = y_{ik} - \hat{R}_{i}^2 x_{ik}^{-1}, \quad R_i = \left( \sum_{k=1}^{M_i} \sum_{i=1}^{n} y_{ik} \pi_{k|i} x_{ik}^{-1} \right) / \left( \sum_{i=1}^{n} \sum_{k=1}^{M_i} \pi_{k|i} x_{ik}^{-1} \right),
\]

Under SRSWOR design at both the stages, it reduces to

\[
AV(\hat{\iota}_{CP2}) = N^2 \left\{\sum_{i=1}^{n} \left( \sum_{j=1}^{N} \Delta_{ij} - \frac{1}{n} \right) \pi_{li} \pi_{lj} \right\} + \frac{1}{n} \left( \sum_{i=1}^{n} \left( \frac{1}{M_i} - \frac{1}{n} \right) \left( \hat{R}_{i}^2 \right) \right) \text{,} \quad (2.2.9)
\]

where, the terms are as defined in Case 1 (Eqn. 2.1.6 and 2.1.10).

3. SIMULATION STUDY

In order to evaluate the statistical performance of proposed product type calibration estimators, a simulation study was carried out. We have considered the case of two stage sampling where sample selection at each stage is governed by SRSWOR for the situation that the size of the PSU and the corresponding SSUs were fixed. For the simulation study, a finite population of 5000 units considering, number of PSU, \(N=50\) and PSU size, \(M_{i}=100\), was generated from \(y_k = \beta x_k^{-1} + e_k\), \(k=1,...,M_{0}\), where \(M_{0} = \sum_{i=1}^{N} M_i\). The auxiliary variable was generated from normal distribution with mean 5 and variance 1 i.e. \(x_k \sim N(5, 1)\) and the errors, \(e_k, \ k=1,...,M_{0}\), from normal distribution with mean 0 and variance \(\sigma^2 x_k^{-1}\) i.e. \(e_k \sim N(0, \sigma^2 x_k^{-1})\). We have fixed the value of \(\beta = 20\) and chosen four different values for \(\sigma^2\) as 0.25, 1.0, 2.0 and 5.0. In this way, we generated four sets of population, denoted as Set 1, Set 2, Set 3 and Set 4, with different correlation coefficient values between study variable \(Y\) and auxiliary variable \(X\) as -0.91, -0.85, -0.78 and -0.64 respectively. The value of left hand side of the Condition 1 and Condition 2 i.e. \(\rho_A C_{by}/C_{bx} < -1/2\) and \(\rho_w C_{wy}/C_{wx} < -1/2\) are lesser than -0.5 in all the population sets which can be seen in the following table:

<table>
<thead>
<tr>
<th>Set</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
<th>Set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>-1.1</td>
<td>-1.17</td>
<td>-1.23</td>
<td>-1.34</td>
</tr>
<tr>
<td>Condition 2</td>
<td>-1.11</td>
<td>-1.1</td>
<td>-1.1</td>
<td>-1.08</td>
</tr>
</tbody>
</table>

Then, from each of the study population sets, we have selected a total of 10000 different samples of following sizes using SRSWOR at both the stages of the two stage sampling design and calculated different estimates of population total under Case 1 and 2:
Developed product type calibration estimators as well as all other usual estimators of population total under two stage sampling 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 \( \theta \) as given by

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

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

### 4. RESULTS AND DISCUSSION

Table 1 shows the %RB of the HT estimators (\( \hat{t}_{yP} \)), product estimators (\( \hat{t}_{yP1} \) and \( \hat{t}_{yP2} \)), ratio estimators (\( \hat{t}_{yR1} \) and \( \hat{t}_{yR2} \)), linear regression estimators (\( \hat{t}_{ylr1} \) and \( \hat{t}_{ylr2} \)) (as in Särndal et al., (1992), pp-323) and proposed product type calibration estimators (\( \hat{t}_{yCP1} \) and \( \hat{t}_{yCP2} \)) of population total under both the Case 1 and 2 respectively when available auxiliary variable is inversely related with the study variable. Table 2 presents comparison of performance of all the estimators for all the population Sets on the basis of %RRMSE.

From Table 1 it can be seen that, the proposed product type calibration estimators of the population total for both the Case 1 and 2 of availability of auxiliary information were giving consistently least amount %RB in all the sets compared to their usual linear regression, product, ratio and HT estimators under two stage sampling design when available auxiliary variable is inversely related with the study variable. It is evident that ratio estimator is not at all suitable for this situation.

A close look of Table 2 reveals that, the product type calibration estimators of the population total developed under two stage sampling design under Case 1 and 2 were always more efficient than the respective linear regression, product, ratio and HT estimators in all the population sets with respect to %RRMSE. The %RRMSE of both the proposed product type calibration estimators of the population total under Case 1 and 2 were decreasing with the increase of sample sizes. With the increase of negative correlation between the study and auxiliary variable, %RRMSE of both the proposed product type calibration estimators of the population total under Case 1 and 2 were decreasing. The proposed product type calibration estimators of the population total developed under Case 1 of two stage sampling design was producing least %RRMSE in all Sets. Therefore, for the situations of availability of population level complete auxiliary information at SSU level i.e. Case 1, performance of the proposed product type calibration estimator is best among all other competitors. On the other hand, for more practical situation of availability of population level auxiliary information only for selected PSUs i.e. Case 2, proposed product type calibration estimator can be preferred over usual HT, product and linear regression estimators of population total.

### 5. CONCLUSIONS

In this study, following the calibration approach (Deville and Särndal, 1992), we proposed product type calibration estimators of the finite population total under two stage sampling design when the available auxiliary variable is inversely related to the study variable. Here, two different cases under two stage sampling viz. “Case 1: population level complete auxiliary information is available at the SSU level” and “Case 2: population level auxiliary information is available only for the selected PSUs” have been considered. In order to study the statistical performance of proposed product type calibration estimators as compared to existing estimators of population total of study variable, a simulation study was carried out. The simulation results show that the proposed product type calibration estimator of the population total were performing better than usual linear regression, product and HT estimators under two stage sampling design when available auxiliary variable is inversely related with the study variable. The proposed product type calibration estimators of the population total developed under Case 1 performs better than that of Case 2, since more auxiliary information was available under Case 1.

### ACKNOWLEDGEMENT

Authors are profoundly thankful to the reviewers for their productive comments for improvements.
REFERENCES


### Table 1. Comparison of all the estimators under Case 1 and 2 with respect to %RB in case of all four population sets when available auxiliary variable is inversely related with the study variable

<table>
<thead>
<tr>
<th>Set</th>
<th>Sample Size ((n_m, n_f))</th>
<th>(\hat{y}_{y\pi})</th>
<th>(\hat{y}_{yP1})</th>
<th>(\hat{y}_{yP1})</th>
<th>(\hat{y}_{yR1})</th>
<th>(\hat{y}_{yCP2})</th>
<th>(\hat{y}_{yP2})</th>
<th>(\hat{y}_{yR2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1 ((\rho=-0.91))</td>
<td>10_20</td>
<td>0.016</td>
<td>0.001</td>
<td>-0.026</td>
<td>-0.057</td>
<td>0.013</td>
<td>-0.005</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>10_25</td>
<td>0.015</td>
<td>-0.001</td>
<td>-0.011</td>
<td>-0.035</td>
<td>0.055</td>
<td>0.000</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>15_25</td>
<td>0.000</td>
<td>0.002</td>
<td>-0.009</td>
<td>-0.024</td>
<td>0.019</td>
<td>0.000</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>15_30</td>
<td>-0.005</td>
<td>0.002</td>
<td>-0.015</td>
<td>-0.026</td>
<td>0.012</td>
<td>0.002</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>20_30</td>
<td>0.013</td>
<td>-0.001</td>
<td>-0.005</td>
<td>-0.016</td>
<td>0.037</td>
<td>0.001</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>20_40</td>
<td>0.010</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.007</td>
<td>0.024</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>25_40</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.008</td>
<td>-0.013</td>
<td>0.009</td>
<td>-0.009</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>25_50</td>
<td>-0.007</td>
<td>-0.002</td>
<td>-0.005</td>
<td>-0.008</td>
<td>-0.007</td>
<td>0.007</td>
<td>-0.009</td>
</tr>
<tr>
<td>Set 2 ((\rho=-0.85))</td>
<td>10_20</td>
<td>-0.004</td>
<td>0.001</td>
<td>-0.024</td>
<td>-0.054</td>
<td>0.034</td>
<td>0.006</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>10_25</td>
<td>0.006</td>
<td>0.000</td>
<td>-0.018</td>
<td>-0.041</td>
<td>0.044</td>
<td>-0.006</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td>15_25</td>
<td>-0.026</td>
<td>-0.011</td>
<td>-0.025</td>
<td>-0.039</td>
<td>-0.018</td>
<td>-0.011</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>15_30</td>
<td>-0.015</td>
<td>-0.004</td>
<td>-0.020</td>
<td>-0.030</td>
<td>-0.002</td>
<td>-0.009</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>20_30</td>
<td>-0.024</td>
<td>-0.006</td>
<td>-0.019</td>
<td>-0.026</td>
<td>-0.024</td>
<td>-0.004</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td>20_40</td>
<td>0.003</td>
<td>0.001</td>
<td>-0.008</td>
<td>-0.014</td>
<td>0.019</td>
<td>-0.003</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>25_40</td>
<td>0.024</td>
<td>0.003</td>
<td>0.001</td>
<td>-0.006</td>
<td>0.050</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>25_50</td>
<td>-0.010</td>
<td>-0.001</td>
<td>-0.004</td>
<td>-0.007</td>
<td>-0.012</td>
<td>-0.002</td>
<td>-0.006</td>
</tr>
</tbody>
</table>
### Table 2. Comparison of all the estimators under Case 1 and 2 with respect to %RRMSE in case of all four population sets of all the estimators when available auxiliary variable is inversely related with the study variable

<table>
<thead>
<tr>
<th>Set</th>
<th>Sample Size (n, m_i)</th>
<th>( \hat{y}_{CP1} )</th>
<th>( \hat{y}_{P1} )</th>
<th>( \hat{y}_{lR1} )</th>
<th>( \hat{y}_{R1} )</th>
<th>( \hat{y}_{CP2} )</th>
<th>( \hat{y}_{P2} )</th>
<th>( \hat{y}_{lR2} )</th>
<th>( \hat{y}_{R2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 1</strong> (( \rho = -0.91 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 20</td>
<td>1.677</td>
<td>0.377</td>
<td>0.726</td>
<td>0.714</td>
<td>2.975</td>
<td>0.690</td>
<td>0.899</td>
<td>0.890</td>
<td>2.841</td>
</tr>
<tr>
<td>10 25</td>
<td>1.469</td>
<td>0.333</td>
<td>0.638</td>
<td>0.630</td>
<td>2.613</td>
<td>0.677</td>
<td>0.827</td>
<td>0.819</td>
<td>2.461</td>
</tr>
<tr>
<td>15 25</td>
<td>1.199</td>
<td>0.271</td>
<td>0.524</td>
<td>0.513</td>
<td>2.112</td>
<td>0.521</td>
<td>0.659</td>
<td>0.651</td>
<td>2.011</td>
</tr>
<tr>
<td>15 30</td>
<td>1.059</td>
<td>0.246</td>
<td>0.466</td>
<td>0.457</td>
<td>1.873</td>
<td>0.516</td>
<td>0.619</td>
<td>0.613</td>
<td>1.750</td>
</tr>
<tr>
<td>20 30</td>
<td>0.921</td>
<td>0.212</td>
<td>0.404</td>
<td>0.395</td>
<td>1.626</td>
<td>0.420</td>
<td>0.522</td>
<td>0.516</td>
<td>1.535</td>
</tr>
<tr>
<td>20 40</td>
<td>0.770</td>
<td>0.176</td>
<td>0.343</td>
<td>0.336</td>
<td>1.355</td>
<td>0.401</td>
<td>0.473</td>
<td>0.469</td>
<td>1.246</td>
</tr>
<tr>
<td>25 40</td>
<td>0.663</td>
<td>0.154</td>
<td>0.299</td>
<td>0.294</td>
<td>1.170</td>
<td>0.332</td>
<td>0.399</td>
<td>0.397</td>
<td>1.091</td>
</tr>
<tr>
<td>25 50</td>
<td>0.576</td>
<td>0.133</td>
<td>0.258</td>
<td>0.253</td>
<td>1.013</td>
<td>0.322</td>
<td>0.369</td>
<td>0.365</td>
<td>0.917</td>
</tr>
<tr>
<td><strong>Set 2</strong> (( \rho = -0.85 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 20</td>
<td>1.781</td>
<td>0.755</td>
<td>0.967</td>
<td>0.958</td>
<td>3.002</td>
<td>0.971</td>
<td>1.116</td>
<td>1.111</td>
<td>2.876</td>
</tr>
<tr>
<td>10 25</td>
<td>1.562</td>
<td>0.668</td>
<td>0.848</td>
<td>0.841</td>
<td>2.639</td>
<td>0.892</td>
<td>1.011</td>
<td>1.006</td>
<td>2.486</td>
</tr>
<tr>
<td>15 25</td>
<td>1.260</td>
<td>0.536</td>
<td>0.682</td>
<td>0.676</td>
<td>2.128</td>
<td>0.708</td>
<td>0.804</td>
<td>0.801</td>
<td>2.018</td>
</tr>
<tr>
<td>15 30</td>
<td>1.139</td>
<td>0.485</td>
<td>0.615</td>
<td>0.610</td>
<td>1.931</td>
<td>0.666</td>
<td>0.746</td>
<td>0.744</td>
<td>1.812</td>
</tr>
<tr>
<td>20 30</td>
<td>0.983</td>
<td>0.420</td>
<td>0.537</td>
<td>0.532</td>
<td>1.656</td>
<td>0.556</td>
<td>0.635</td>
<td>0.633</td>
<td>1.567</td>
</tr>
<tr>
<td>20 40</td>
<td>0.829</td>
<td>0.356</td>
<td>0.454</td>
<td>0.448</td>
<td>1.392</td>
<td>0.518</td>
<td>0.574</td>
<td>0.572</td>
<td>1.279</td>
</tr>
<tr>
<td>25 40</td>
<td>0.716</td>
<td>0.307</td>
<td>0.389</td>
<td>0.385</td>
<td>1.209</td>
<td>0.429</td>
<td>0.478</td>
<td>0.476</td>
<td>1.128</td>
</tr>
<tr>
<td>25 50</td>
<td>0.620</td>
<td>0.267</td>
<td>0.337</td>
<td>0.333</td>
<td>1.045</td>
<td>0.398</td>
<td>0.434</td>
<td>0.432</td>
<td>0.949</td>
</tr>
<tr>
<td><strong>Set 3</strong> (( \rho = -0.78 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 20</td>
<td>1.907</td>
<td>1.081</td>
<td>1.220</td>
<td>1.215</td>
<td>3.048</td>
<td>1.229</td>
<td>1.343</td>
<td>1.341</td>
<td>2.924</td>
</tr>
<tr>
<td>10 25</td>
<td>1.708</td>
<td>0.954</td>
<td>1.076</td>
<td>1.071</td>
<td>2.744</td>
<td>1.131</td>
<td>1.218</td>
<td>1.217</td>
<td>2.601</td>
</tr>
<tr>
<td>15 25</td>
<td>1.372</td>
<td>0.763</td>
<td>0.864</td>
<td>0.860</td>
<td>2.209</td>
<td>0.888</td>
<td>0.964</td>
<td>0.961</td>
<td>2.102</td>
</tr>
<tr>
<td>15 30</td>
<td>1.246</td>
<td>0.700</td>
<td>0.794</td>
<td>0.789</td>
<td>1.990</td>
<td>0.840</td>
<td>0.909</td>
<td>0.908</td>
<td>1.869</td>
</tr>
<tr>
<td>20 30</td>
<td>1.073</td>
<td>0.588</td>
<td>0.667</td>
<td>0.662</td>
<td>1.731</td>
<td>0.696</td>
<td>0.751</td>
<td>0.749</td>
<td>1.644</td>
</tr>
<tr>
<td>20 40</td>
<td>0.894</td>
<td>0.497</td>
<td>0.561</td>
<td>0.557</td>
<td>1.439</td>
<td>0.621</td>
<td>0.664</td>
<td>0.663</td>
<td>1.329</td>
</tr>
<tr>
<td>25 40</td>
<td>0.787</td>
<td>0.435</td>
<td>0.493</td>
<td>0.489</td>
<td>1.263</td>
<td>0.532</td>
<td>0.572</td>
<td>0.570</td>
<td>1.179</td>
</tr>
<tr>
<td>25 50</td>
<td>0.672</td>
<td>0.380</td>
<td>0.427</td>
<td>0.424</td>
<td>1.073</td>
<td>0.484</td>
<td>0.513</td>
<td>0.512</td>
<td>0.979</td>
</tr>
<tr>
<td><strong>Set 4</strong> (( \rho = -0.64 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 20</td>
<td>2.316</td>
<td>1.685</td>
<td>1.767</td>
<td>1.767</td>
<td>3.353</td>
<td>1.807</td>
<td>1.876</td>
<td>1.877</td>
<td>3.222</td>
</tr>
<tr>
<td>10 25</td>
<td>2.063</td>
<td>1.510</td>
<td>1.576</td>
<td>1.574</td>
<td>2.982</td>
<td>1.635</td>
<td>1.690</td>
<td>1.690</td>
<td>2.840</td>
</tr>
<tr>
<td>15 30</td>
<td>1.505</td>
<td>1.090</td>
<td>1.143</td>
<td>1.139</td>
<td>2.177</td>
<td>1.192</td>
<td>1.239</td>
<td>1.237</td>
<td>2.055</td>
</tr>
<tr>
<td>20 30</td>
<td>1.271</td>
<td>0.931</td>
<td>0.975</td>
<td>0.972</td>
<td>1.837</td>
<td>1.007</td>
<td>1.048</td>
<td>1.047</td>
<td>1.749</td>
</tr>
<tr>
<td>20 40</td>
<td>1.072</td>
<td>0.777</td>
<td>0.810</td>
<td>0.807</td>
<td>1.553</td>
<td>0.869</td>
<td>0.895</td>
<td>0.895</td>
<td>1.445</td>
</tr>
<tr>
<td>25 40</td>
<td>0.947</td>
<td>0.697</td>
<td>0.725</td>
<td>0.722</td>
<td>1.364</td>
<td>0.767</td>
<td>0.789</td>
<td>0.788</td>
<td>1.283</td>
</tr>
<tr>
<td>25 50</td>
<td>0.822</td>
<td>0.603</td>
<td>0.626</td>
<td>0.623</td>
<td>1.180</td>
<td>0.679</td>
<td>0.699</td>
<td>0.698</td>
<td>1.084</td>
</tr>
</tbody>
</table>