



Thailand Statistician
April 2023; 21(2): 370-382
<http://statassoc.or.th>
Contributed paper

Estimation of Finite Population Mean by Using Median Type Estimators in Two-Stage Sampling

Waqar Hafeez* [a], Javid Shabbir [b] and Nazrina Aziz [a, c]

[a] School of Quantitative Sciences, University Utara Malaysia, Malaysia.

[b] Department of Statistics, Quaid-i-Azam University Islamabad, Pakistan.

[c] Institute of Strategic Industrial Decision Modelling (ISIDM), Universiti Utara Malaysia, Malaysia.

*Corresponding author; e-mail: waqarhafeez78601@yahoo.com

Received: 12 October 2020

Revised: 29 April 2021

Accepted: 30 May 2021

Abstract

This paper suggests some median based estimators in two-stage sampling (TSS) with unequal first and second-stage sampling units. Expressions for bias and mean square error (MSE) are derived up to the first-order approximation. In an empirical study, the proposed estimators are compared with existing estimators. The results explain that the proposed class of estimators performs much better than all other considered estimators.

Keywords: Efficiency, mean estimation, median estimator, mean square error, two-stage.

1. Introduction

Most statistical studies have used sample means of study and the auxiliary variables to estimate the population mean. In some cases, when outliers exist in the data sets, sample means may yield inefficient results. Noughabi and Jarrahiferiz (2019) proposed a new estimator for Shannon entropy. Many authors used the auxiliary information to improve the efficiency of the estimators (see Sharma and Kumar 2020, Hafeez and Shabbir 2015). If the correlation between the study variable and the auxiliary variable is very low, using the sample mean may produce inefficient results (Hafeez et al. 2020). Mahalanobis et al. (1940) were the first who initiated the idea of two-stage sampling (TSS). When clusters have many sampling units, estimates for population totals, means, and proportions are usually derived from the sub-sample units (see Ponkaew and Lawson, 2019). Saxena et al. (1984) proposed a method for estimating population totals using a TSS design with unequal first-stage sampling (FSS) units chosen with equal probabilities. The situation can occur when more than one frame is available at each stage of selection and given the uses of multi-frame surveys for two-stage designs. When the auxiliary information for the FSS units in TSS is established, Sahoo (1987) proposed a regression type estimator to estimate the population total and population mean. Later it was also used through controlled sampling by (Nayak and Sahoo 2016). Sahoo and Panda (1999) proposed a general class of estimators using a TSS scheme with two auxiliary variables. Hossain and

Ahmed (2001) proposed a class of ratio, regression, and product type estimators using one auxiliary variable in TSS for unequal FSS units. In various cases, expressions for MSEs are derived up to the first-order approximation (Qureshi, Riaz and Hanif 2020).

In the case of known overall mean and means of FSS units, Sahoo et al. (2009) proposed a general class of estimators to estimate the finite population mean in TSS. Olufadi and Yunusa (2010) investigated a class of ratio-cum-product estimators in TSS, using unequal sub-sampling fractions to obtain MSEs. Theoretically and numerically, it was demonstrated that this class produces more effective results than Singh (1967) estimators. Sahoo et al. (2011) suggested a general class of estimators in TSS to calculate finite population totals. A simulation analysis discovered that the proposed estimators outperform the ratio and regression estimators under different sample sizes at the second stage. Using the auxiliary information, Meenakshi and Neha (2013) presented a general class of estimators for estimating the population mean in a two-stage double sampling scheme. It was demonstrated how the proposed class of estimators could be expanded to consider several known auxiliary variables. Singh et al. (2013) investigated the use of ratio, product, and linear regression estimators in TSS when the population mean of the auxiliary variable was unknown. Through numerical analysis, it was confirmed that the proposed estimator outperforms the usual ratio, regression, and product estimators under certain conditions.

When conducting a large survey with an incomplete sampling frame or more than one level of sampling is needed, TSS is used, which is an extension of cluster sampling. In TSS, the population is divided into several groups and select a sample of some groups known as FSS units (FSUs). Again, draw a sub-sample from the selected first stage units, called second-stage sampling units (SSUs). Similarly, it can extend to three-stage, four-stage, and multi-stage sampling.

Consider a finite population $U = \{u_1, u_2, \dots, u_N\}$ having N first-stage units (FSU), and M_i be the number of second-stage units (SSU). Let y_i and x_i be the characteristics of the study variable (y) and the auxiliary variable (x) respectively, for j^{th} SSU in unit u_i , ($i = 1, 2, \dots, N$) and ($j = 1, 2, \dots, M_i$). This study assumes that the auxiliary variable (x) and the study variable (y) are positively correlated. This research is limited to estimating the population mean $\bar{Y}$ based on median information.

Draw a sample of size n , FSU's from U and then again, draw a sample of size m_i SSU's from selected i^{th} FSU U_i from the population by using SRSWOR sampling scheme. Define $\bar{P}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} P_{ij}$ and $\bar{P} = \frac{1}{N} \sum_{i=1}^N \bar{P}_i$, where $P = Y, X$.

Also define: $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$, $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$, $S_{yx} = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})$,

$f = \frac{n}{N}$ and $\lambda = \frac{1-f}{n}$. Let under TSS when the FSU are unequal in size;

$$V(\bar{y}) = \lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2, \quad V(\bar{x}) = \lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2, \quad \text{Cov}(\bar{y}, \bar{x}) = \lambda S_{1yx} + \sum_{i=1}^N \lambda_{2i} S_{2yxi},$$

$$f_{2i} = \frac{m_i}{M_i}, \quad \lambda_{2i} = \frac{M_i^2 (1-f_{2i})}{n N M_i}, \quad S_{1y}^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{Y}_i - \bar{Y})^2, \quad S_{2yi}^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_{ij} - \bar{Y}_i)^2,$$

$$S_{1x}^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{X}_i - \bar{X})^2, \quad S_{2xi}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_{ij} - \bar{X}_i)^2, \quad S_{1yx} = \frac{1}{N-1} \sum_{i=1}^N (\bar{Y}_i - \bar{Y})(\bar{X}_i - \bar{X})$$

$$\text{and } S_{2yxi} = \frac{1}{N-1} \sum_{i=1}^N (Y_{ij} - \bar{Y}_i)(X_{ij} - \bar{X}_i).$$

In the absence of the auxiliary variable under TSS, the variance of $\bar{y}$, is given by

$$V(\bar{y}_2) = \lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2. \quad (1)$$

The classical ratio estimator suggested by Cochran (1940) under TSS is given by

$$\hat{\bar{Y}}_{R2} = \bar{y}_2 \left(\frac{\bar{X}}{\bar{x}_2} \right), \quad (2)$$

where $\bar{X}$ is the known population mean of x . The bias and MSE of $\hat{\bar{Y}}_{R2}$ are given by

$$B(\hat{\bar{Y}}_{R2}) \cong \frac{1}{\bar{X}} \left[\lambda [S_{1x}^2 - S_{1yx}] + \sum_{i=1}^N \lambda_{2i} [R_1 S_{2xi}^2 - S_{2yxi}] \right] \quad (3)$$

and

$$MSE(\hat{\bar{Y}}_{R2}) \cong \lambda [S_{1y}^2 + R_1^2 S_{1x}^2 - 2R_1 S_{1yx}] + \sum_{i=1}^N \lambda_{2i} [S_{2yi}^2 + R_1^2 S_{2xi}^2 - 2R_1 S_{2yxi}]. \quad (4)$$

The usual regression estimator suggested by Watson (1937) under TSS is given by

$$\hat{\bar{Y}}_{lr2} = \bar{y}_2 + b_{yx} (\bar{X} - \bar{x}_2), \quad (5)$$

where $b_{yx} = \frac{S_{yx}}{S_x^2}$ is the least square estimate of $\beta_{yx} = \frac{S_{yx}}{S_x^2}$. The MSE of the regression estimator is given by

$$MSE(\hat{\bar{Y}}_{lr2}) = \lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 - \frac{\left(\lambda S_{1yx} + \sum_{i=1}^N \lambda_{2i} S_{2yxi} \right)^2}{\lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2} \quad (6)$$

or

$$MSE(\hat{\bar{Y}}_{lr2}) = \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2), \quad (7)$$

where

$$\rho_2 = \frac{\lambda S_{1yx} + \sum_{i=1}^N \lambda_{2i} S_{2yxi}}{\sqrt{\left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] \left[\lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2 \right]}}.$$

Bahl and Tuteja (1991) defined the following estimator under TSS:

$$\hat{\bar{Y}}_{BT2} = \bar{y}_2 \exp \left(\frac{\bar{X} - \bar{x}_2}{\bar{X} + \bar{x}_2} \right). \quad (8)$$

The bias and MSE of $\hat{\bar{Y}}_{BT2}$ are

$$B\left(\hat{Y}_{BT2}\right) \cong \frac{\lambda}{2\bar{X}}\left[\frac{3}{4}R_1S_{1x}^2 - S_{1yx}\right] - \frac{1}{2\bar{X}}\sum_{i=1}^N\lambda_{2i}\left[\frac{3}{4}R_1S_{2xi}^2 - S_{2yxi}\right] \quad (9)$$

and

$$MSE\left(\hat{Y}_{BT2}\right) \cong \lambda\left[S_{1y}^2 + \frac{1}{4}R_1^2S_{1x}^2 - R_1S_{1yx}\right] + \sum_{i=1}^N\lambda_{2i}\left[S_{2yi}^2 + \frac{1}{4}R_1^2S_{2xi}^2 - R_1S_{2yxi}\right]. \quad (10)$$

Rao (1991) defined the following estimator under TSS which is given by

$$\hat{Y}_{Rao2} = c_1\bar{y}_2 + c_2(\bar{X} - \bar{x}_2), \quad (11)$$

where c_1 and c_2 are unknown constants. The bias and MSE of $\hat{Y}_{Rao2}$ are given below

$$B\left(\hat{Y}_{Rao2}\right) = (c_1 - 1)\bar{Y} \quad (12)$$

and

$$MSE\left(\hat{Y}_{Rao2}\right) = \frac{\bar{Y}^2\left[\lambda S_{1y}^2 + \sum_{i=1}^N\lambda_{2i}S_{2yi}^2\right](1-\rho_2^2)}{\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N\lambda_{2i}S_{2yi}^2\right](1-\rho_2^2)}. \quad (13)$$

Grover and Kaur (2011) defined the following estimator under TSS:

$$\hat{Y}_{GK2} = \left[c_3\bar{y}_2 + c_4(\bar{X} - \bar{x}_2)\right]\exp\left(\frac{\bar{X} - \bar{x}_2}{\bar{X} + \bar{x}_2}\right), \quad (14)$$

where c_3 and c_4 are unknown constants. The bias and MSE of $\hat{Y}_{GK2}$ are given by

$$B\left(\hat{Y}_{GK2}\right) \cong \bar{Y}(c_3 - 1) + \frac{1}{2\bar{X}}\left[\frac{3}{4}c_3R_1 + c_4\right]\left[\lambda S_{1x}^2 + \sum_{i=1}^N\lambda_{2i}S_{2xi}^2\right] - \frac{c_3}{2\bar{X}}\left[\lambda S_{1yx} + \sum_{i=1}^N\lambda_{2i}S_{2yxi}\right]$$

and

$$MSE\left(\hat{Y}_{GK2}\right) \cong \frac{\bar{Y}^2\left[\lambda S_{1y}^2 + \sum_{i=1}^N\lambda_{2i}S_{2yi}^2\right](1-\rho_2^2)}{\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N\lambda_{2i}S_{2yi}^2\right](1-\rho_2^2)} - \frac{\left[\lambda S_{1x}^2 + \sum_{i=1}^N\lambda_{2i}S_{2xi}^2\right]\left[4\left[\lambda S_{1y}^2 + \sum_{i=1}^N\lambda_{2i}S_{2yi}^2\right](1-\rho_2^2) + \frac{R^2}{4}\left[\lambda S_{1x}^2 + \sum_{i=1}^N\lambda_{2i}S_{2xi}^2\right]\right]}{16\bar{X}^2\left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N\lambda_{2i}S_{2yi}^2\right](1-\rho_2^2)\right]}. \quad (15)$$

2. Proposed Estimators in Tow-Stage Sampling

On the lines of Bahl and Tuteja (1991), this study proposes the following estimator to estimate the finite population mean ($\bar{Y}$) under TSS:

$$\hat{Y}_{BT2(m)} = \bar{y}_2 \exp\left(\frac{\bar{M} - m_2}{\bar{M} + m_2}\right), \quad (16)$$

where $\bar{M}$ is the known population median and m_2 be the sample median in TSS. To obtain the bias and MSE of $\hat{Y}_{BT2(m)}$, need to define the following error terms.

Let $e_0 = \frac{\bar{y}_2 - \bar{Y}}{\bar{Y}}$ and $e_1 = \frac{m_2 - \bar{M}}{\bar{M}}$, such that $E(e_0) = E(e_1) = 0$, $E(e_0^2) = \frac{V(\bar{y})}{\bar{Y}^2}$, $E(e_1^2) = \frac{V(m)}{\bar{M}^2}$ and $E(e_0 e_1) = \frac{Cov(\bar{y}, m)}{\bar{Y} \bar{M}}$. When FSU are unequal in size, then under TSS:

$$V(\bar{y}) = \lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2, \quad V(m) = \lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2, \quad Cov(\bar{y}, m) = \lambda S_{1ym} + \sum_{i=1}^N \lambda_{2i} S_{2ymi}^-,$$

$$S_{1y}^2 = \frac{1}{\binom{N}{n} - 1} \sum_{i=1}^{\binom{N}{n}} (\bar{Y}_i - \bar{Y})^2, \quad S_{2yi}^2 = \frac{1}{\binom{N}{n} - 1} \sum_{i=1}^{\binom{N}{n}} (\bar{y}_{ij} - \bar{Y}_i)^2, \quad S_{1m}^2 = \frac{1}{\binom{N}{n} - 1} \sum_{i=1}^{\binom{N}{n}} (\bar{M}_i - \bar{M})^2,$$

$$S_{2mi}^2 = \frac{1}{\binom{N}{n} - 1} \sum_{i=1}^{\binom{N}{n}} (m_{ij} - \bar{M}_i)^2, \quad S_{1ym} = \frac{1}{\binom{N}{n} - 1} \sum_{i=1}^{\binom{N}{n}} (\bar{Y}_i - \bar{Y})(\bar{M}_i - \bar{M}),$$

$$S_{2ymi}^- = \frac{1}{\binom{N}{n} - 1} \sum_{i=1}^{\binom{N}{n}} (\bar{y}_{ij} - \bar{Y}_i)(m_{ij} - \bar{M}_i).$$

In term of errors, (16) can be written as $\hat{\bar{Y}}_{BT2(m)} = \bar{Y} (1 + e_0) \exp \left(-\frac{e_1}{2} \left(1 + \frac{e_1}{2} \right)^{-1} \right)$. Up to first-order approximation, expand the term $\left(1 + \frac{e_1}{2} \right)^{-1}$ as

$$\hat{\bar{Y}}_{BT2(m)} = \bar{Y} (1 + e_0) \exp \left(\frac{e_1^2}{4} - \frac{e_1}{2} \right), \quad (17)$$

$$\hat{\bar{Y}}_{BT2(m)} - \bar{Y} \cong \bar{Y} \left(e_0 - \frac{1}{2} e_1 + \frac{3}{8} e_1^2 - \frac{1}{2} e_0 e_1 \right). \quad (18)$$

Using (18), the bias of $\hat{\bar{Y}}_{BT2(m)}$ under TSS is

$$B(\hat{\bar{Y}}_{BT2(m)}) \cong \frac{\lambda}{2\bar{M}} \left[\frac{3}{4} R_2 S_{1m}^2 - S_{1ym} \right] - \frac{1}{2\bar{M}} \sum_{i=1}^N \lambda_{2i} \left[\frac{3}{4} R_2 S_{2mi}^2 - S_{2ymi}^- \right]. \quad (19)$$

To get MSE of $\hat{\bar{Y}}_{BT2(m)}$ squaring on both sides of (18), and after taking expectation, one gets

$$MSE(\hat{\bar{Y}}_{BT2(m)}) \cong \bar{Y}^2 E \left(e_0 - \frac{1}{2} e_1 \right)^2 \cong \bar{Y}^2 E \left(e_0^2 + \frac{1}{4} e_1^2 - e_0 e_1 \right) \cong V(\bar{y}) + \frac{1}{4} R_2^2 V(m) - R_2 Cov(\bar{y}, m).$$

MSE of $\hat{\bar{Y}}_{BT2(m)}$ under TSS is given by

$$MSE(\hat{\bar{Y}}_{BT2(m)}) \cong \lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 + \frac{1}{4} R_2^2 \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] - R_2 \left[\lambda S_{1ym} + \sum_{i=1}^N \lambda_{2i} S_{2ymi}^- \right], \quad (20)$$

$$MSE(\hat{\bar{Y}}_{BT2(m)}) \cong \lambda \left[S_{1y}^2 + \frac{1}{4} R_2^2 S_{1m}^2 - R_2 S_{1ym} \right] + \sum_{i=1}^N \lambda_{2i} \left[S_{2yi}^2 + \frac{1}{4} R_2^2 S_{2mi}^2 - R_2 S_{2ymi}^- \right]. \quad (21)$$

Propose another unbiased difference type median based estimator to estimate the finite population mean using the median of samples. The proposed estimator under TSS is given by

$$\hat{Y}_{d2(m)} = \bar{y}_2 + k^* (\bar{M} - m_2), \quad (22)$$

where k^* is unknown constant. In terms of errors, (22) can be written as

$$\hat{Y}_{d2(m)} - \bar{Y} = \bar{Y}e_0 + k^* \bar{M}e_1. \quad (23)$$

For MSE taking square on both sides of (23) and after taking expectation, one gets

$$\begin{aligned} MSE(\hat{Y}_{d2(m)}) &= E(\bar{Y}e_0 + k^* \bar{M}e_1)^2 = \bar{Y}^2 E(e_0^2) + k^{*2} \bar{M}^2 E(e_1^2) - 2k^* \bar{Y}\bar{M}E(e_0e_1) \\ &= V(\bar{y}) + k^{*2}V(m) - 2k^*Cov(\bar{y}, m). \end{aligned}$$

Under TSS, the MSE of $\hat{Y}_{d2(m)}$ is given by

$$MSE(\hat{Y}_{d2(m)}) = \lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 + k^{*2} \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] - 2k^* \left[\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi} \right]. \quad (24)$$

The minimum MSE of $\hat{Y}_{d2(m)}$ at the optimum value of k^* ; $k_{(opt)}^* = \frac{\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi}}{\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2}$ is given by

$$MSE(\hat{Y}_{d2(m)})_{(min)} = \lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 - \frac{\left[\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi} \right]^2}{\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2} \quad (25)$$

or

$$MSE(\hat{Y}_{d2(m)})_{(min)} = \left[\lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 \right] (1 - \rho_2^{*2}), \quad (26)$$

where

$$\rho_2^* = \frac{\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi}}{\sqrt{\left[\lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 \right] \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right]}}.$$

On the lines of Rao (1991), the following median based estimator is proposed to estimate the finite population mean under TSS

$$\hat{Y}_{Rao2(m)} = k_1^* \bar{y}_2 + k_2^* (\bar{M} - m_2), \quad (27)$$

where k_1^* and k_2^* are unknown constants. In terms of errors (27) can be written as

$$\hat{Y}_{Rao2(m)} - \bar{Y} = (k_1^* - 1)\bar{Y} + k_1^* \bar{Y}e_0 - k_2^* \bar{M}e_1. \quad (28)$$

By using (28), the bias of $\hat{Y}_{Rao2(m)}$ is given by

$$B(\hat{Y}_{Rao2(m)}) = (k_1^* - 1)\bar{Y}. \quad (29)$$

By squaring on both sides of (28) and then solve up to first-order approximation. The MSE of the estimator $\hat{Y}_{Rao2(m)}$ is

$$MSE\left(\hat{\bar{Y}}_{Rao2(m)}\right) = \left(k_1^* - 1\right)^2 \bar{Y}^2 + k_1^{*2} V(\bar{y}) + k_2^{*2} V(m) - 2k_1^* k_2^* Cov(\bar{y}, m). \quad (30)$$

Under TSS, MSE of $\hat{\bar{Y}}_{Rao2(m)}$ is given by

$$MSE\left(\hat{\bar{Y}}_{Rao2(m)}\right) = \left(k_1^* - 1\right)^2 \bar{Y}^2 + k_1^{*2} \left[\lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 \right] + k_2^{*2} \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] - 2k_1^* k_2^* \left[\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi} \right].$$

The optimum values of k_1^* and k_2^* are given by

$$k_{1(opt)}^* = \frac{\bar{Y}^2}{\bar{Y}^2 + \left[\lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 \right] (1 - \rho_2^{*2})} \quad \text{and} \quad k_{2(opt)}^* = k_{1(opt)}^* \frac{\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi}}{\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2}.$$

Now, the minimum MSE of $\hat{\bar{Y}}_{Rao2(m)}$ is

$$MSE\left(\hat{\bar{Y}}_{Rao2(m)}\right)_{(min)} = \frac{\bar{Y}^2 \left[\lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 \right] (1 - \rho_2^{*2})}{\bar{Y}^2 + \left[\lambda S_{1\bar{y}}^2 + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}i}^2 \right] (1 - \rho_2^{*2})} \quad (31)$$

On the lines of Grover and Kaur (2011), the following median based estimator is proposed to estimate the finite population mean under TSS

$$\hat{\bar{Y}}_{P2(m)} = \left[k_3^* \bar{y}_2 + k_4^* (\bar{M} - m_2) \right] \exp\left(\frac{\bar{M} - m_2}{\bar{M} + m_2}\right), \quad (32)$$

where k_3^* and k_4^* are unknown constants. To find bias and MSE of proposed estimator, (32) can be written in terms of errors as

$$\hat{\bar{Y}}_{P2(m)} = \left[k_3^* \bar{Y} (1 + e_0) - k_4^* \bar{M} e_1 \right] \exp\left[-\frac{e_1}{2} \left(1 + \frac{e_1}{2}\right)^{-1}\right]. \quad (33)$$

Expand the term $\left(1 + \frac{e_1}{2}\right)^{-1}$ up to the first degree of e 's,

$$\hat{\bar{Y}}_{P2(m)} \cong \left[k_3^* \bar{Y} (1 + e_0) - k_4^* \bar{M} e_1 \right] \left(1 - \frac{1}{2} e_1 + \frac{3}{8} e_1^2\right), \quad (34)$$

$$\hat{\bar{Y}}_{P2(m)} - \bar{Y} \cong \bar{Y} \left[(k_3^* - 1) + k_3^* e_0 - \frac{1}{2} k_3^* (e_1 + e_0 e_1) + \frac{3}{8} k_3^* e_1^2 \right] + k_4^* \bar{M} \left(\frac{1}{2} e_1^2 - e_1 \right). \quad (35)$$

After taking the expectation of (35), the bias of $\hat{\bar{Y}}_{P2(m)}$ is

$$B\left(\hat{\bar{Y}}_{P2(m)}\right) \cong \bar{Y} \left[(k_3^* - 1) + \frac{3}{8} k_3^* \frac{Var(m)}{\bar{M}^2} - \frac{1}{2} k_3^* \frac{Cov(\bar{y}, m)}{\bar{Y} \bar{M}} \right] + \frac{1}{2} k_4^* \frac{Var(m)}{\bar{M}}. \quad (36)$$

Under TSS (36), is given as

$$B\left(\hat{\bar{Y}}_{P2(m)}\right) \cong \bar{Y} (k_3^* - 1) + \frac{1}{2\bar{M}} \left(\frac{3}{4} k_3^* R_2 + k_4^* \right) \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] - \frac{1}{2} k_3^* R_2 \left[\lambda S_{1\bar{y}m} + \sum_{i=1}^N \lambda_{2i} S_{2\bar{y}mi} \right] \quad (37)$$

Squaring on both sides of (35), one gets

$$\begin{aligned} \left(\hat{Y}_{P2(m)} - \bar{Y} \right)^2 &\cong \bar{Y}^2 \left[\left(k_3^* - 1 \right)^2 + k_3^{*2} \left(e_0^2 + \frac{1}{4} e_1^2 \right) - k_3^* \left(2k_3^* - 1 \right) e_0 e_1 + \frac{3}{4} k_3^* \left(k_3^* - 1 \right) e_1^2 \right] \\ &\quad + k_4^{*2} \bar{M}^2 e_1^2 + 2k_4^* \bar{Y} \bar{M} \left[k_3^* \left(\frac{1}{2} e_1^2 - e_0 e_1 \right) + \frac{1}{2} \left(k_3^* - 1 \right) e_1^2 \right]. \end{aligned} \quad (38)$$

Using (38), the MSE of the proposed estimator $\hat{Y}_{P2(m)}$ up to first-order approximation is given as

$$\begin{aligned} MSE \left(\hat{Y}_{P2(m)} \right) &\cong \bar{Y}^2 \left(k_3^* - 1 \right)^2 + k_3^{*2} \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] + \left[\left(k_3^* R_2 + k_4^* \right)^2 - \left(\frac{3}{4} k_3^* R_2 + k_4^* \right) R_2 \right] \\ &\quad \times \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] + k_3^* \left(R_2 - 2k_3^* R_2 - 2k_4^* \right) \left[\lambda S_{1ym}^2 + \sum_{i=1}^N \lambda_{2i} S_{2ymi}^2 \right]. \end{aligned} \quad (39)$$

The optimum values of k_3^* and k_4^* are

$$k_{3(opt)}^* = R_2^2 \frac{\bar{M}^2 - \frac{1}{8} \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right]}{\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})} \quad \text{and} \quad k_{4(opt)}^* = k_{3(opt)}^* \left[\frac{\lambda S_{1ym}^2 + \sum_{i=1}^N \lambda_{2i} S_{2ymi}^2}{\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2} - R_2 \right] + \frac{R_2}{2}.$$

The minimum value of MSE of $\hat{Y}_{P2(m)}$, is given by

$$\begin{aligned} MSE \left(\hat{Y}_{P2(m)} \right)_{(min)} &\cong \frac{\bar{Y}^2 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})}{\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})} \\ &\quad - \frac{\left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] \left[4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) + \frac{R_2^2}{4} \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] \right]}{16\bar{M}^2 \left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) \right]} \end{aligned} \quad (40)$$

3. Efficiency Conditions

In this section, a proposed median based estimator is compared in terms of MSE with other estimators in TSS.

Let

$$\begin{aligned} A &= \frac{\left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] \left[4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) + \frac{R_2^2}{4} \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] \right]}{16\bar{M}^2 \left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) \right]} \geq 0 \\ B &= \frac{\bar{Y}^2 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 + \bar{Y}^2 \right] \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})}{\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})} \geq 0 \end{aligned}$$

$$C = \frac{\left[\lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2 \right] \left[4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2) + \frac{R_1^2}{4} \left[\lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2 \right] \right]}{16 \bar{X}^2 \left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2) \right]} \geq 0$$

and

$$D = \frac{\left[\left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] + \bar{Y}^2 \right] \left[\left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) \right]}{\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})} \geq 0..$$

Condition 1: Using (1) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq V(\bar{y}_2)$ if $A + B \geq 0$.

Condition 2: Using (4) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{\bar{Y}}_{R2}\right)$ if

$$A + B + R_1^2 \left[\lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2 \right] - 2R_1 \left[\lambda S_{1yx} + \sum_{i=1}^N \lambda_{2i} S_{2yxi} \right] \geq 0.$$

Condition 3: Using (6) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq V\left(\hat{\bar{Y}}_{lr2}\right)$ if

$$A + B - \frac{\left[\lambda S_{1yx} + \sum_{i=1}^N \lambda_{2i} S_{2yxi} \right]^2}{\lambda S_{1x}^2 + \sum_{i=1}^N \lambda_{2i} S_{2xi}^2} \geq 0.$$

Condition 4: Using (10) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{\bar{Y}}_{BT2}\right)$ if

$$A + B + \lambda \left[\frac{1}{4} R_1 S_{1x}^2 - S_{1yx} \right] R_1 + \sum_{i=1}^N \lambda_{2i} \left[\frac{1}{4} R_1 S_{2xi}^2 - S_{2yxi} \right] R_1 \geq 0.$$

Condition 5: Using (13) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{\bar{Y}}_{Rao2}\right)$ if

$$A + \frac{\bar{Y}^4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2) - \bar{Y}^4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})}{\left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2) \right] \left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) \right]} \geq 0.$$

Condition 6: Using (15) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{\bar{Y}}_{GK2}\right)_{(\min)}$ if

$$A - C + \frac{\bar{Y}^4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2) - \bar{Y}^4 \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2})}{\left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^2) \right] \left[\bar{Y}^2 + \left[\lambda S_{1y}^2 + \sum_{i=1}^N \lambda_{2i} S_{2yi}^2 \right] (1 - \rho_2^{*2}) \right]} \geq 0.$$

Condition 7: Using (21) and (40), $MSE\left(\hat{\bar{Y}}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{\bar{Y}}_{BT2(m)}\right)$ if

$$A + D + \frac{1}{4} R_2^2 \left[\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2 \right] - R_2 \left[\lambda S_{1ym} + \sum_{i=1}^N \lambda_{2i} S_{2ymi} \right] \geq 0.$$

Condition 8: Using (25) and (40), $MSE\left(\hat{Y}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{Y}_{d2(m)}\right)$ if

$$A + D - \frac{\left[\lambda S_{1ym} + \sum_{i=1}^N \lambda_{2i} S_{2ymi} \right]^2}{\lambda S_{1m}^2 + \sum_{i=1}^N \lambda_{2i} S_{2mi}^2} \geq 0.$$

Condition 9: Using (31) and (40), $MSE\left(\hat{Y}_{P2(m)}\right)_{(\min)} \leq MSE\left(\hat{Y}_{Rao2(m)}\right)_{(\min)}$ if $A \geq 0$.

Conditions 1 and 9 are always true.

4. Empirical Study

In this section, the MSE's of the proposed class of estimators is compared with simple mean, ratio, linear regression, and all existing estimators under the TSS scheme.

Data set (Sarndal, 2003, p.652):

According to this data set, Sweden is divided into 284 municipalities. A municipality consists of a town and its surrounding areas. The revenues from 1985 municipal taxation (in millions of kronor) are considered a study variable, and the population of 1975 is considered an auxiliary variable in the first data set. The population in 1985 is considered as an auxiliary variable in the second data set. The summary of data is showing in Table 1. Under TSS for the sample size $n = 15$ and $n = 20$, Table 2 demonstrates the MSE and Table 3 demonstrates the PRE for both populations.

Table 1 Summary statistics for both populations

Parameters	Population 1	Population 2
N	50	50
$\bar{Y}$	245.088	245.088
$\bar{X}$	28.810	29.363
R_1	8.507	8.347
S_y^2	355612.5	355612.5
S_x^2	2795.59	2658.09
S_{yx}	30482.96	29563.62
ρ_{yx}	0.9668	0.9607

5. Conclusion

This research proposed a class of median type estimators to estimate the finite population mean under TSS. In Table 2, it can be observed that the proposed class of estimators performs much better than the simple mean, usual ratio, linear regression, Bahl and Tuteja (1991), Rao (1991) and Grover and Kaur (2011) estimators based on the auxiliary information. Theoretically, the proposed estimator has been compared with all other estimators and efficiency conditions have been obtained. In real-life phenomena, the use of auxiliary information supports the estimators to increase the precision of the

estimators. In Table 2, the difference between $\hat{Y}_{BT2(m)}$ and $\hat{Y}_{d2(m)}$ is very small but $\hat{Y}_{Rao2(m)}$ and $\hat{Y}_{P2(m)}$ gives more precise results.

Table 2 Variance/MSE of different estimators for different sample size in TSS

Estimator	Population 1		Population 2	
	$n = 15$	$n = 20$	$n = 15$	$n = 20$
$\bar{y}_2$	96808.19	72166.44	96808.19	72166.44
Using the auxiliary variable				
$\hat{Y}_{R2}$	81839.29	61001.75	80723.09	60171.09
$\hat{Y}_{lr2}$	38149.29	28429.64	40179.02	29880.73
$\hat{Y}_{BT2}$	39486.04	29440.25	41730.27	31113.83
$\hat{Y}_{Rao2}$	57863.14	57146.00	58208.4	57599.63
$\hat{Y}_{GK2}$	33487.50	26769.37	31848.85	25493.19
Using the median of study variable				
$\hat{Y}_{BT2(m)}$	15335.99	11172.82	15335.99	11172.82
$\hat{Y}_{d2(m)}$	15248.02	11121.28	15248.02	11121.28
$\hat{Y}_{Rao2(m)}$	12161.00	9383.903	12161.00	9383.903
$\hat{Y}_{P2(m)}$	11705.25	9124.767	11705.25	9124.767

Table 3 PRE of estimators with respect to $\bar{y}$ in TSS

Estimator	Population 1		Population 2	
	$n = 15$	$n = 20$	$n = 15$	$n = 20$
$\bar{y}_2$	100	100	100	100
Using the auxiliary variable				
$\hat{Y}_{R2}$	118.2906	118.3022	119.9263	119.9350
$\hat{Y}_{lr2}$	253.7614	253.8423	240.9421	241.5149
$\hat{Y}_{BT2}$	245.1707	245.1285	231.9855	231.9433
$\hat{Y}_{Rao2}$	167.3055	126.2843	166.3131	125.2898
$\hat{Y}_{GK2}$	289.0875	269.5859	303.9613	283.0812
Using the median of study variable				
$\hat{Y}_{BT2(m)}$	631.2484	645.9107	631.2484	645.9107
$\hat{Y}_{d2(m)}$	634.8902	648.9041	634.8902	648.9041
$\hat{Y}_{Rao2(m)}$	796.0545	769.045	796.0545	769.045
$\hat{Y}_{P2(m)}$	827.0493	790.8853	827.0493	790.8853

Acknowledgments

All authors are grateful to the reviewers for their valuable suggestions, which helped to improve the manuscript.

References

- Bahl S, Tuteja R. Ratio and product type exponential estimators. *J Info Optim Sci.* 1991; 12(1): 159-164.
- Cochran WG. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *J Agric Sci.* 1940; 30(2): 262-275.
- Grover LK, Kaur P. An improved estimator of the finite population mean in simple random sampling. *Model Assist Stat Appl.* 2011; 6(1): 47-55.
- Hafeez W, Shabbir J. Estimation of the finite population mean, using median based estimators in stratified random sampling. *J Stat Appl Prob.* 2015; 4(3): 367-374.
- Hafeez W, Shabbir J, Shah MT, Ahmed S. Some median type estimators to estimate the finite population mean. *Asian J Prob Stat.* 2020; 7(4): 48-58.
- Hossain MI, Ahmed MS. A class of predictive estimators in two stage sampling using auxiliary information. *Int J Info Manag Sci.* 2001; 12(1): 49-56.
- Mahalanobis PC, Mukherjea R, Ghosh A. A sample survey of after-effects of the Bengal famine of 1943. *Sankhya.* 1946; 337-400.
- Nayak R, Sahoo L. On the performance of some regression-type estimators obtained through controlled splitting of sample. *Thail Stat.* 2016; 14(2): 203-18.
- Noughabi HA, Jarrahiferiz J. A new estimator for Shannon entropy. *Thail Stat.* 2019; 17(2): 190-197.
- Srivastava M, Garg N. A general class of estimators of a finite population mean using multi-auxiliary information under two-stage double sampling scheme. *J Reliab Stat Stud.* 2013; 115-132.
- Olufadi Y. On the estimation of ratio-cum-product estimators using two stage sampling. *Stat Transit.* 2010; 11(2): 253-265.
- Ponkaew C, Lawson N. New population total estimators under stratified sampling design in the presence of nonresponse. *Thail Stat.* 2019; 17(2): 198-211.
- Qureshi MN, Riaz N, Hanif M. Joint influence of exponential ratio and exponential product estimators for the estimation of clustered population variance in adaptive cluster sampling. *Thail Stat.* 2020; 18(1): 65-76.
- Rao TJ. On certain methods of improving ratio and regression estimators. *Commun Stat-Theory Methods.* 1991; 20(10): 3325-3340.
- Sahoo LN. A regression-type estimator in two stage sampling. *Calcutta Stat Assoc Bull.* 1987; 36(1-2): 97-100.
- Sahoo LN, Mahapatra N, Senapati SC. A generalized method of estimation for two stage sampling using two auxiliary variables. *J Stat Theory Pract.* 2009; 3(4): 831-839.
- Sahoo LN, Sahoo RK, Senapati SC, Mangaraj AK. A general class of estimators in two stage sampling with two auxiliary variables. *Hacettepe J Math Stat.* 2011; 40(5): 757-765.
- Sahoo LN, Panda P. Theory and methods: a class of estimators using auxiliary information in two-stage sampling. *Aust N Z J Stat.* 1999; 41(4): 405-410.
- Sarndal CE, Swensson B, Wretman J. Model assisted survey sampling. Berlin: Springer Science & Business Media; 2003.
- Saxena BC, Narain P, Srivastava AK. Multiple frame surveys in two stage sampling. *Sankhya.* 1984; 75-82.

- Sharma V, Kumar S. Simulation study of ratio type estimators in stratified random sampling using multi-auxiliary information. *Thail Stat.* 2020; 18(3): 281-289.
- Singh MP. Ratio cum product method of estimation. *Metrika.* 1967; 12(1): 34-42.
- Singh R, Vishwakarma GK, Gupta PC, Pareek S. An alternative approach to estimation of population mean in two stage sampling. *Math Theory Model.* 2013; 3(13): 48-53.
- Watson DJ. The estimation of leaf area in field crops. *J Agric Sci.* 1937; 27(3): 474-483.