

Research Article

Calibration Approach for Variance Estimation of Small Domain

Ashutosh Ashutosh ¹, Usman Shahzad ^{2,3}, Ishfaq Ahmad,² Muhammad Ahmad Raza,⁴ and Troon J. Benedict ⁵

¹Department of Statistics, Faculty of Science & Technology, Mahatma Gandhi Kashi Vidyapith, Varanasi, India

²Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan

³Department of Mathematics and Statistics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan

⁴Department of Statistics, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan

⁵Department of Economics, Maasai Mara University, Narok, Kenya

Correspondence should be addressed to Troon J. Benedict; troon@mmarau.ac.ke

Received 13 July 2022; Accepted 12 August 2022; Published 28 August 2022

Academic Editor: Tahir Mehmood

Copyright © 2022 Ashutosh Ashutosh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The population parameters are estimated using sample surveys. However, the current investigation's goal is to estimate subpopulations parameters such as total through a calibration approach. The proposed estimator's properties have been given under simple random sampling. In addition, the class of estimators for subpopulation total has been discussed. The proposed estimator has been evaluated theoretically and empirically in a comparative study. The results demonstrated that a high-level estimate outperforms a low-level estimate for subpopulation total using a calibration approach.

1. Introduction

Generally, the sample surveys are applied to estimate the population parameters. However, this study is interested in estimating the subpopulations total. If these subpopulations are classified through some characteristics, like socio-economic, geographical regions may also term to be the domains. The estimate of these subpopulations (domains) total has become a very popular and effective tool when framing the program and policies of the government and private sectors. Hence, subpopulations demand has been growing acceleratory for a couple of years. Purcell and Kish [1] have classified the subpopulation due to the size with respect to population. The simple classification of the subpopulations is as follows:

Major subpopulation: it comprises 1/10 of the population or more. For example, the major geographical regions like (north, east, west, south, central), 10-year age group, or major classes like occupations.

Minor subpopulation: it comprises between 1/10 to 1/100 of the population, for example, state population, single-

year age, two-fold classification like education and occupations.

Mini subpopulation: it comprises between 1/100 to 1/10000 of the population, for example, the population of the counties (more than 3000 in the U.S.A) or three fold classifications like age, education, and occupation.

Rare subpopulation: it comprises less than 1/10000 of the population. For example, health services regions are classified into local regions of residence.

The subpopulation total estimate can be used to estimate the social exclusion and well-being levels. Also, some of the environmental and epidemiological issues can be solved through subpopulation estimates. The consequence of the domain estimate with the popular direct and indirect methods has been explained in the literature (see Rao [2] and Singh [3]). Rahman [4] has been given the direct and indirect estimation with the model-based ideas. The availability of units in the study subpopulation depends on the method. Hence, in such a situation we used the indirect method. In the indirect method, the sample selected from the population (which consists of the subpopulations) than the

subpopulation. The surrounding units are utilized when the units in the study subpopulation may be low or near zero for some subpopulations. This is due to the restriction of the units in the study subpopulation so, utilized the surrounding subpopulations units. If the surrounding units are similar in nature to the subpopulation, then the estimate gives a precise result that is acceptable at the desired level. The indirect estimates for subpopulations parameters like total have been explained through an auxiliary variable by Rao and Molina [5]. Singh et al. [6] have estimated the poverty indicators, many socio-economic indicators, and food insecurity for subpopulations total. In the indirect method, the availability of the auxiliary variable for the subpopulation has significant importance to Tikkiwal et al. [7]. The ratio and regression estimators through the calibration approach for population total have been discussed by numerous authors, such as Särndal [8], Singh and Mohl [9], Singh et al. [10], Wu and Sitter [11]. The extension of the calibration-based mean estimation through an auxiliary variable has been discussed by Koyuncu and Kadilar [12] and Koyuncu [13]. Furthermore, the estimation of the subpopulation parameter by Khare et al. [14]. The calibrate estimator was initiated by Horvitz and Thompson [15] on the study variable Y , and estimate the population total through simple random sampling without replacement designs. Deville and Särndal [16] have incorporated the auxiliary variable under the restriction of minimum chi-square distance. Särndal [8] has estimated the variance of the ratio and regression estimators for the population parameter. The mean square error using model based approach has been given by Slud and Maiti [17]. Whenever, recently classes of the estimators for variance estimation have been discussed by Bhushan et al. [18, 19]. We are motivated by Sarndal [8] and hence propose the calibration estimate for the subpopulation total as well as their variance. The rest of the article is constructed as follows: firstly, the Methodology employed is explained. Also, evaluate the variance of the proposed estimator using a calibration approach with the low and high levels, respectively. Then, a class of calibration estimators is presented. Furthermore, theoretical and empirical comparisons along with Concluding Results are presented and summarized. Finally, Recommendations and Applications are outlined.

2. Methodology

Consider the a^{th} subpopulation $\Omega_1 + \Omega_2 + \dots + \Omega_A = \sum_{a=1}^A \Omega_a = \Omega$ with the sizes $N_a: \{a = 1, \dots, A\}$. Select a sample from the population in which the selected sampling units of a^{th} domains are $S_a \{S_a \subset \Omega_a\}$. The overall sample of size n is $\sum_{a=1}^A n_a = n$. A sample is selected by the population rather than the subpopulation. The auxiliary variable of the subpopulation should be known. Hence, the auxiliary information for the proposed estimator is utilized. In the current work, both ideas of Deville and Särndal [16], and Tikkiwal et al. [7] are used in the same context. The proposed indirect generalized regression (GREG) estimator of a^{th} subpopulation total is written as follows:

$$T_{PD,a} = \sum_{i=1}^n t_{i,a} y_i, \quad (1)$$

where $t_i; i = 1, \dots, n$ represent the updated weight which is close to the design weight d_i . The utilization of the auxiliary variable can be a good option in the calibration estimator of the a^{th} domain. Also, a^{th} subpopulation total of the auxiliary variable is equal to the sum of the n^{th} units equal to the subpopulation total of auxiliary variable X . The auxiliary equation of the subpopulation total of X is

$$\sum_{i=1}^n t_{i,a} x_i = X_a. \quad (2)$$

The minimum chi-square distance function is

$$\sum_{i=1}^n \frac{(t_{i,a} - d_i)^2}{d_i q_i}, \quad (3)$$

where q_i is a chosen constant. Utilize (2) and (3), the calibration equation can be written as follows:

$$f(t_{i,a}, \delta) = \sum_{i=1}^n \frac{(t_{i,a} - d_i)^2}{d_i q_i} + 2\delta \left(X_a - \sum_{i=1}^n t_{i,a} x_i \right). \quad (4)$$

Partially differentiate (4) w.r. to new weights $t_{i,a}$,

$$\frac{\partial f(t_{i,a}, \delta)}{\partial t_{i,a}} = \sum_{i=1}^n \frac{2(t_{i,a} - d_i)}{d_i q_i} - 2\delta \sum_{i=1}^n x_i = 0. \quad (5)$$

Simplifying (5), then

$$t_{i,a} = d_i + \delta d_i q_i x_i. \quad (6)$$

Substitute $t_{i,a}$ in the auxiliary equation (2), we obtain the value of δ which further substitute in (5), we obtain the new weight as follows:

$$t_{i,a} = d_i + \frac{d_i q_i x_i}{\sum_{i=1}^n d_i q_i x_i^2} \left(X_a - \sum_{i=1}^n d_i x_i \right). \quad (7)$$

The new weight $t_{i,a}$ substitute in (1), the proposed estimator will be

$$T_{PD,a} = \sum_{i=1}^n d_i y_i + \frac{\sum_{i=1}^n d_i q_i x_i y_i}{\sum_{i=1}^n d_i q_i x_i^2} \left(X_a - \sum_{i=1}^n d_i x_i \right). \quad (8)$$

3. Variance of $T_{PD,a}$ through Low-Level Calibration Approach

Särndal et al. [20] provided the variance of the GREG estimator. Zaman and Bulut [21] recently discussed the variance of ratio estimators. The proposed estimator is like a regression estimator. For obtaining the variance of the population's total estimator of Deville and Särndal [16], the sample is taken based on Yates and Grundy [22]

$$V(T_{DS}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} (d_i y_i - d_j y_j)^2, \quad (9)$$

where $D_{ij} = \pi_i \pi_j - \pi_{ij} / \pi_{ij}$; $i \neq j$, $e_i = y_i - \hat{\beta} x_i$, $\hat{\beta} = \sum_{i=1}^n d_i y_i / \sum_{i=1}^n d_i x_i$.

The proposed estimator is an indirect estimator for subpopulation hence we take the asymptotic concept of an unbiased estimator. The variance of the subpopulation total is written as follows:

$$V(T_{PD,DS,a}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} e_{i,a}^2, \quad (10)$$

where $e_{i,a} = y_{i,a} - \hat{\beta} X_a$, $\hat{\beta}$ and D_{ij} as defined previously.

The members of the proposed estimator are as follows:

- (I) For $q_i = 1$, the proposed estimator in (8) reduced to the GREG estimator, as follows:

$$T_{YG,1,a} = \sum_{i=1}^n d_i y_i + \frac{\sum_{i=1}^n d_i x_i y_i}{\sum_{i=1}^n d_i x_i^2} \left(X_a - \sum_{i=1}^n d_i x_i \right). \quad (11)$$

- (II) For $q_i = 1/x_i$, the proposed estimator in (8) will be ratio estimator, as follows:

$$T_{YG,2,a} = \frac{\sum_{i=1}^n d_i y_i}{\sum_{i=1}^n d_i x_i} X_a. \quad (12)$$

The variance of $T_{PD,DS,a}$ can be obtain with the help of (9). The variance of the proposed estimator

$$V(T_{PD,DS,a}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} (t_{i,a} e_{i,a} - t_{j,a} e_{j,a})^2, \quad (13)$$

where

$$D_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}}; i \neq j, e_{i,a} = y_{i,a} - \hat{\beta} X_a. \quad (14)$$

The variance of the ratio estimator

$$V(T_{PD,DS,a}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) (d_i e_{i,a} - d_j e_{j,a})^2 + \hat{\xi}_{a1} \left(X_a - \sum_{i=1}^n d_i x_i \right) + \hat{\xi}_{a2} \left(X_a - \sum_{i=1}^n d_i x_i \right)^2, \quad (15)$$

and

$$\hat{\xi}_{a1} = \frac{\sum_{i=1}^n \sum_{j=1}^n D_{ij} (d_i e_{i,a} - d_j e_{j,a}) (d_i q_i x_i e_{i,a} - d_j q_j x_j e_{j,a})}{\sum_{i=1}^n d_i q_i x_i^2},$$

$$\hat{\xi}_{a2} = \frac{\sum_{i=1}^n \sum_{j=1}^n D_{ij} (d_i e_{i,a} - d_j e_{j,a}) (d_i q_i x_i e_{i,a} - d_j q_j x_j e_{j,a})}{2 \left(\sum_{i=1}^n d_i q_i x_i^2 \right)^2}. \quad (16)$$

The value of GREG is under the simple random sampling without replacement. The probability of i^{th} unit is selected

$\pi_i = n/N$, the probability of j^{th} unit is selected $\pi_j = n/N$ and the probability of both i^{th} and j^{th} units are selected $\pi_{ij} = n(n-1)/N(N-1)$. The variance of the low-level calibration of the GREG estimator of the subpopulation total.

$$V_L(T_{YG,1,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2, \quad (17)$$

where $e_{i,a} = y_{i,a} - \hat{\beta} X_a$, $f = n/N$ and the variance of the ratio estimator that given in (15) is

$$V_L(T_{YG,2,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 \left(\frac{X_a}{\bar{X}} \right)^2, \quad (18)$$

where $e_{i,a} = y_{i,a} - \bar{y}/\bar{x} X_a$.

The indirect regression estimator is $T_{YG,1,a} = \sum_{i=1}^n d_i y_i + \sum_{i=1}^n d_i x_i y_i / \sum_{i=1}^n d_i x_i^2 (X_a - \sum_{i=1}^n d_i x_i)$ and $\hat{X} = N/n \sum_{i=1}^n x_i$.

We used the idea of Deng and Wu [23] and estimated the variance of the subpopulation total. The variance of the low-level calibration of ratio estimator $V_L(T_{YG,2,a})$ can be estimated for the subpopulation using equation in (15).

$$V_L(T_{YG,3,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 \left(\frac{X_a}{\bar{X}} \right)^h, \quad (19)$$

where $e_{i,a} = y_{i,a} - \hat{\beta} X_a$ and write $(X_a/\bar{X})^h$ up to second order, neglect higher order due to a small value

$$\left(\frac{X_a}{\bar{X}} \right)^h = \left[1 + h \left(\frac{X_a}{\bar{X}} - 1 \right) + h(h-1) \left(\frac{X_a}{\bar{X}} - 1 \right)^2 + \dots \right]. \quad (20)$$

If substitute $h = 2$ in (19) the then variance of the ratio estimator.

We obtain the variance of the linear regression estimator which is a special case of the estimator of the variance of ratio estimator of domain total. If the subpopulation total is equivalent to the population total that is $(X_a/\bar{X}) \cong 1$, then it reduces into the linear form of the class of estimators of (19). The variance of the regression estimator is written as follows:

$$V_L(T_{YG,4,a}) \cong \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2. \quad (21)$$

The variance of low-level calibration $\hat{V}_L(T_{YG,1,a})$ of the regression estimator is more efficient than the ratio estimator when $(X_a/\bar{X}) > 1$. However, for $(X_a/\bar{X}) < 1$ always variance ratio estimator of the low-level calibration approach is lower than the value of the regression estimator of the low-level calibration approach. The variance of the GREG estimator is equal to the class of estimates of Deng and Wu [23].

4. Variance of $T_{PD,a}$ through High Level Calibration Approach

The high-level calibration is the adjustment of the weight function of the selected units of i^{th} unit, j^{th} unit, and both i^{th} and j^{th} units. We estimate the variance and checked the variance of the high-level calibration approach of ratio and

regression estimators. We consider the function Ω_{ij} instead of the D_{ij} then the variance of the proposed estimator

$$V(T_{PD,DS,a}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Omega_{ij} (t_{i,a} e_{i,a} - t_{j,a} e_{j,a})^2, \quad (22)$$

where Ω_{ij} is the new weight which is very close to D_{ij} . The simple calibration equation

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Omega_{ij} (d_i e_{i,a} - d_j e_{j,a})^2 = \widehat{V}_{YG}(\widehat{X}_{HT,a}), \quad (23)$$

where $V_{YG}(\widehat{X}_{HT,a}) \cong 1/2 \sum_{i=1}^N \sum_{j=1}^N (\pi_i \pi_j - \pi_{ij}) (d_i x_{i,a} - d_j x_{j,a})^2$ is known variance with the auxiliary information for each subpopulation total should be known $X_a = \sum_{i=1}^N x_i$. This is an auxiliary equation (see (3)) of the Horvitz–Thompson which is written as $\widehat{X}_{HT} = \sum_{i=1}^n d_i x_i$. We estimated the value of the GREG estimator for the subpopulation total. Utilize the information of the auxiliary variable of the census registers, previous survey value, or administrative registers. The estimation of the variance of regression and ratio estimators have been given by Singh and Srivastava [24], Srivastava and Jhaji [25], Isaki [26], and Wu [27]. Fuller [28] has given the adjustment the weight for the regression estimator. The minimum chi-square distance function of the new weights Ω_{ij} and weight of i^{th} and j^{th} units D_{ij} ($i, j=1, 2, \dots, n$) from the population

$$D = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{(\Omega_{ij} - D_{ij})^2}{D_{ij} L_{ij}}, \quad (24)$$

where L_{ij} is the chosen constant. The estimated value of D under the restriction of (23) is

$$\Omega_{ij} = D_{ij} + P_a \left[V_{YG}(\widehat{X}_{HT,a}) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} (d_i x_{i,a} - d_j x_{j,a})^2 \right], \quad (25)$$

where $P_a = (d_i x_{i,a} - d_j x_{j,a})^2 / 1/2 \sum_{i=1}^n \sum_{j=1}^n D_{ij} L_{ij} (d_i x_{i,a} - d_j x_{j,a})^4$.

The optimum value of the Lagrange's multiplier λ_a with the help of (23) and (24) is written as follows:

$$\lambda_a = \frac{V_{YG}(\widehat{X}_{HT,a}) - 1/2 \sum_{i=1}^n \sum_{j=1}^n D_{ij} (d_i x_{i,a} - d_j x_{j,a})^2}{1/2 \sum_{i=1}^n \sum_{j=1}^n D_{ij} L_{ij} (d_i x_{i,a} - d_j x_{j,a})^4}. \quad (26)$$

Substitute the value of Ω_{ij} from (26) in (22). The GREG estimator is

$$V_{SS}(T_{GREG,a}) = V_{YG}(T_{DS,a}) + \phi_{1,a} [V_{YG}(\widehat{X}_{HT,a}) - \widehat{V}_{YG}(\widehat{X}_{HT,a})], \quad (27)$$

where

$$\phi_{1,a} = \frac{\sum_{i=1}^n \sum_{j=1}^n D_{ij} L_{ij} (d_i x_{i,a} - d_j x_{j,a})^2 (t_{i,a} e_{i,a} - t_{j,a} e_{j,a})^2}{1/2 \sum_{i=1}^n \sum_{j=1}^n D_{ij} L_{ij} (d_i x_{i,a} - d_j x_{j,a})^4} = \dots \quad (28)$$

The members of calibration based approach of higher-level calibration approach are

- (I) If we put the weights $L_{ij,a} = 1/(d_i x_{i,a} - d_j x_{j,a})^2$ within (28), then the variance of the ratio estimator through high-level is

$$V_H(T_{YG,1,a}) \cong \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 \left(\frac{X_a}{\widehat{X}} \right)^2 \left(\frac{S_x^2}{s_x^2} \right), \quad (29)$$

where $s_x^2 = 1/n - 1 \sum_{i=1}^n (x_i - \bar{x})^2$ and $S_x^2 = 1/N - 1 \sum_{i=1}^N (x_i - \bar{X})^2$, the sample variance is an asymptotic unbiased estimate of the population variance.

- (II) For $L_{ij,a} = 1$, the equation (28) will be the variance of the regression estimator

$$V_H(T_{YG,2,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 + \widehat{\delta}_{1,a} (X_a - \widehat{X}) + \widehat{\delta}_{2,a} (X_a - \widehat{X})^2 + \widehat{\delta}_{3,a} (S_x^2 - s_x^2), \quad (30)$$

where

$$\widehat{\delta}_{1,a} = \frac{\sum_{i=1}^n \sum_{j=1}^n D_{ij,a} (d_i e_{i,a} - d_j e_{j,a}) (d_i q_i x_{i,a} e_{i,a} - d_j q_j x_{j,a} e_{j,a})}{\sum_{i=1}^n d_i q_i x_i^2}$$

$$\left(X_a - \sum_{i=1}^n d_i x_i \right),$$

$$\widehat{\delta}_{2,a} = \frac{\sum_{i=1}^n \sum_{j=1}^n D_{ij,a} (d_i q_i x_i e_{i,a} - d_j q_j x_j e_{j,a})^2}{2 \left(\sum_{i=1}^n d_i q_i x_i^2 \right)^2} \left(X_a - \sum_{i=1}^n d_i x_i \right)^2,$$

$$\widehat{\delta}_{3,a} = \frac{N^2(1-f)}{n \sum_{i=1}^n (x_i - x_j)^4} \left[(x_i - x_j) (e_{i,a} - e_{j,a}) \frac{(X_a - \widehat{X})(x_i - x_j)^2}{\sum_{i=1}^n x_i^2} \right]^2. \quad (31)$$

Equation (31) shows that the high-level calibration estimator is different from Deng and Wu [23]. The GREG and ratio estimators are members of the class of estimators of Srivastava and Jhaji [25]

$$V_{SJ}(T_{YG,2,a}) \cong \left(\frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 \right) F \left(\frac{X_a}{\widehat{X}}, \frac{S_x^2}{s_x^2} \right), \quad (32)$$

where $F(\dots)$ is the function such that $F(1, 1) = 1$ holds the certain regularity conditions is better than the low-level calibration estimates, Srivastava and Jhaji [25] and Deng and Wu [23].

5. Class of Calibration Estimators

This section is presenting the class of estimators. A class of estimators is a collection of various estimators under certain regularity conditions that give the same variance. We assume $r = X_a / \sum_{i=1}^n d_i x_i$ and $s = V(X_{HT}) / \widehat{V}(\widehat{X}_{HT})$. The variance of a class of estimators of the sub-population total is

$$V_{SS}(T_{YG,2,a}) = \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij,a} (d_i e_{i,a} - d_j e_{j,a})^2 \right) F(r, s), \quad (33)$$

where the function $F(r, s)$ is of r and s . For example, $F(1, 1) = 1$ possesses the following regularity conditions:

- (1) Function $F(r, s)$ exists for all the values of (r, s) which contain the points $(X_a / \sum_{i=1}^n d_i x_i, V(X_{HT}) / \widehat{V}(\widehat{X}_{HT}))$ inbound subset of two dimensional real spaces.
- (2) First and second-order partial derivatives of the function $F(r, s)$ exist and are also continuous and bounded.

Different members of the class of estimators are exists under regularity conditions. However, three members are

$F(r, s) = r^\eta s^\kappa, F(r, s) = 1 + \eta(r - 1) + \kappa(s - 1), F(r, s) = [1 + \eta(r - 1) + \kappa(s - 1)]^{-1}$ where η and κ both unknown parameters. The value of η and κ are depending on the estimated value. The asymptotic variance of the three estimators will be the same as Srivastava and Jhaji [25] and Singh and Singh [29]. Our proposed estimator is better than the Srivastava and Jhaji [25] and Singh and Singh [29], hence also better than the class of estimator under the regularity restrictions 1 and 2.

5.1. Theoretical Comparison. The theoretical comparison is given due to keeping in mind, the efficiency of the high-level and low-level indirect estimated variances. The high-level variance of the ratio estimator for the subpopulation total is

$$V_H(T_{YG,1,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 \left(\frac{X_a}{\widehat{X}} \right)^2 \left(\frac{S_x^2}{s_x^2} \right). \quad (34)$$

The estimated variance of the ratio estimator through the low-level calibration approach for a^{th} subpopulation total is written as follows:

$$V_L(T_{YG,1,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^n e_{i,a}^2 \left(\frac{X_a}{\widehat{X}} \right)^2. \quad (35)$$

Now, from both (37) and (35), the variance of the high-level is lower as compared to the variance of the low-level of ratio estimate for $S_x^2 < s_x^2$

$$V_H(T_{YG,1,a}) < V_L(T_{YG,1,a}). \quad (36)$$

Furthermore, a comparison of the high-level variance estimate of GREG estimator is

$$V_H(T_{YG,2,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^{n_a} e_{i,a}^2 + \widehat{\psi}_{1,a}(X_a - \widehat{X}) + \widehat{\psi}_{2,a}(X_a - \widehat{X})^2 + \widehat{\psi}_{3,a}(S_x^2 - s_x^2). \quad (37)$$

The low-level variance estimate of the GREG estimator is

$$V_L(T_{YG,2,a}) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^{n_a} e_{i,a}^2 + \widehat{\psi}_{1,a}(X_a - \widehat{X}) + \widehat{\psi}_{2,a}(X_a - \widehat{X})^2. \quad (38)$$

With help of (37) and (38), the restriction is that last term $(s_x^2 - S_x^2) > 0$

$$V_H(T_{YG,2,a}) < V_L(T_{YG,2,a}). \quad (39)$$

We can say that the high-level is more efficient than the low-level calibration approach of ratio and regression estimators.

6. Empirical Study

This section presents an empirical comparison based on the simulation. We take a real data set from Sarndal et al. [30] where the population of 1975 is considered as an auxiliary variable and population of 1985 is considered as a study variable. The subpopulation is considered from Sweden's municipalities regions. However, we take only five regions (subpopulation) 1, 2, 3, 7, and 8 out of the existing eight subpopulations 1, 2, 3, 4, 5, 6, 7, and 8. The study subpopulations have units 25, 48, 32, 15, and 29. For the simulation, we select a random sample of approximately 10%, 20%, and 30% units from the population by study and auxiliary characters y and x , respectively. This process is repeated to the finite times and then obtained the estimated error. Equations (34)–(38) are utilized to obtain the variance of the ratio estimator and generalized estimator of a^{th} subpopulation with low-level and high-level calibration.

The variances of the ratio estimator through high-level calibration are lower as compared to the low-level calibration estimate for all subpopulations. The high-level estimated value is lower as compared to low-level calibration with the sizes 15, 30, and 45, (4.5 to 10.43), (1.75 to 9.5), and (2.53 to 7.1) percentage for domains, respectively. It is observed that the low-level calibration estimate variances are decreasing with the sizes increase from 15 to 45. A similar pattern is also observed for the high-level calibration approach.

The estimated value of the variance of the GREG estimator through the high level is smaller than the low-level estimate for all the subpopulations. The high-level estimated value is lower than the low-level GREG with sizes 15, 30, and 45 in terms of percentage are (2.4 to 12.55), (1.78 to 8.8) and (2.32 to 10.41) for subpopulations, respectively. The width of the generalized is higher than the ratio estimator for the sub-populations. The GREG through low-level calibration estimate is decreasing with the variances when their sizes increase from 15 to 45.

The variance of the ratio and regression estimators with low-level and high-level calibration approach on 15, 30, and 45 units are given. Figures 1–3 are showing the real view of the estimated variance of the ratio estimator on low-level and high-level 15, 30, and 45 sizes. However, Figures 4–6 are

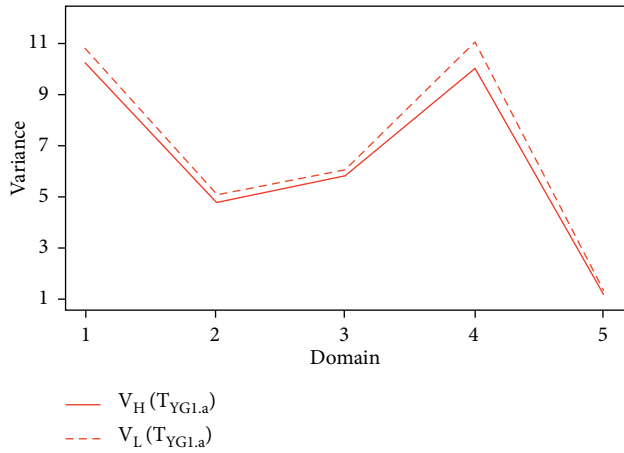


FIGURE 1: The variance of the ratio estimator with low and high levels calibration approach on 15 size. Scale: Y-axis: variance (in thousands), X-axis: domain.

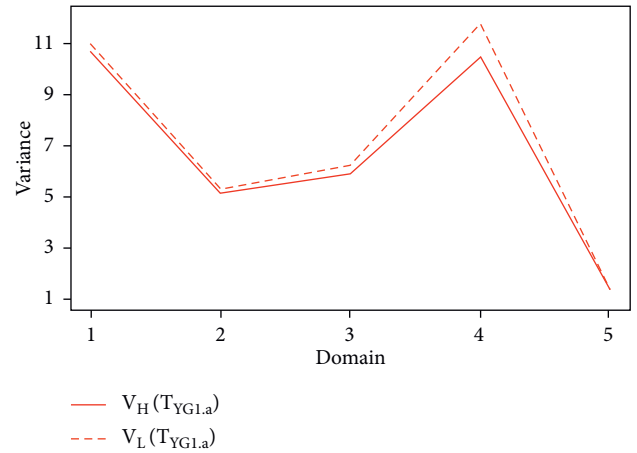


FIGURE 4: The variance of the regression estimator with low and high levels calibration approach on 15 size.

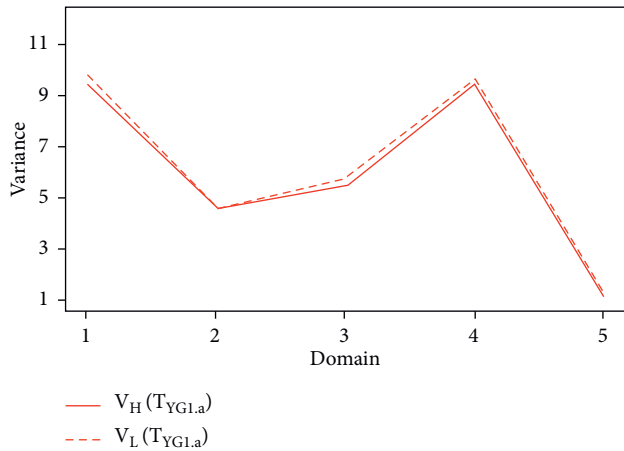


FIGURE 2: The variance of the ratio estimator with low and high levels calibration approach on 30 size.

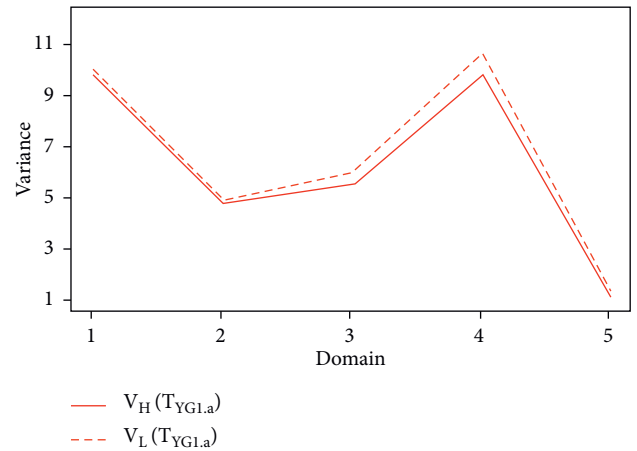


FIGURE 5: The variance of the regression estimator with low and high levels calibration approach on 30 size.

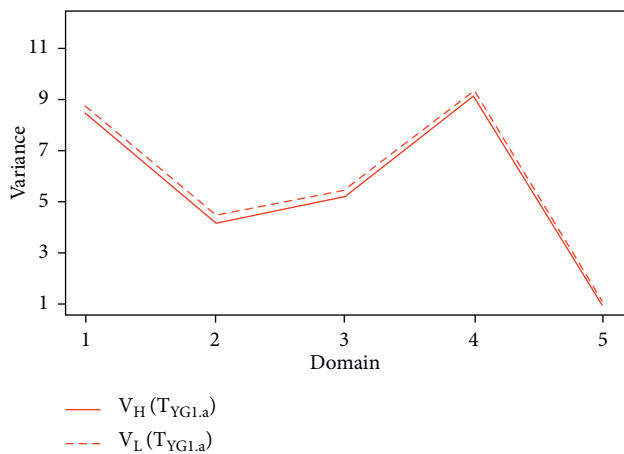


FIGURE 3: The variance of the ratio estimator with low and high levels calibration approach on 45 size.

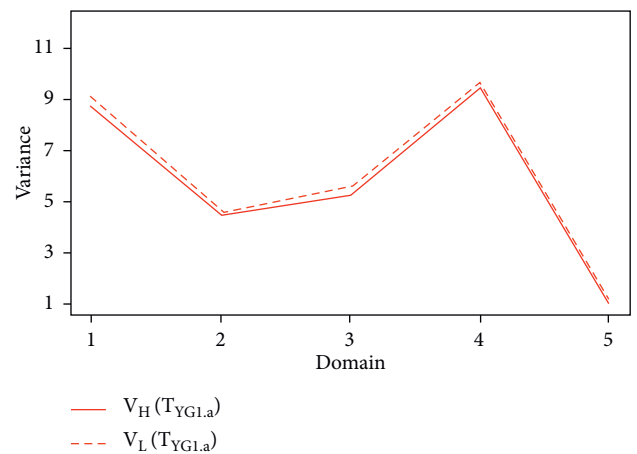


FIGURE 6: The variance of the regression estimator with low and high levels calibration approach on 45 size.

TABLE 1: The value of the variances of high-level of $V_H(T_{YG,1,a})$ and low-level $V_L(T_{YG,1,a})$ of ratio estimate of domains 1, 2, 3, 4, and 5.

Estimator	Size	Subpopulation values				
		1	2	3	4	5
$V_H(T_{YG,1,a})$	15	1027610.1	480190.7	579590.1	1000261.6	118628.6
	30	950956.3	457112.4	548271.6	953223.8	107597.4
	45	845356.7	410954.2	516572.8	911537.8	98217.8
$V_L(T_{YG,1,a})$	15	1079521.2	507254.3	605687.6	1104598.5	128679.5
	30	982150.3	465127.7	578342.3	976731.1	117866.7
	45	873429.6	440132.2	544219.4	934568.4	102283.7

TABLE 2: The variance of the high-level $V_H(T_{YG,2,a})$ and variance of low-level $V_L(T_{YG,2,a})$ of GREG estimate for the domains 1, 2, 3, 4, and 5.

Estimator	Size	Subpopulation				
		1	2	3	4	5
$V_H(T_{YG,2,a})$	15	1073120.5	503254.6	588732.7	1050521.2	126451.3
	30	980427.5	474528.7	552382.3	984327.3	112362.1
	45	875356.7	442734.5	520172.6	947826.2	99879.2
$V_L(T_{YG,2,a})$	15	1099021.4	523459.7	620032.6	1182421.7	130142.6
	30	997831.5	489021.2	593278.4	1070988.6	120345.8
	45	910614.3	456267.6	563267.9	969841.8	110283.7

presenting the real view of the regression estimate variance of high and low-level calibration approach on the different sample sizes (15, 30, and 45).

7. Concluding Results

Tables 1 and 2, and Figures 1–6 show that the high-level calibration estimate of ratio is preferred over than low-level calibration estimate of the subpopulations 1, 2, 3, 4, and 5. The regression estimate through calibration approach ratio is more effective than the generalized estimate. The low-level estimate is poor performance than the high-level estimate in both the estimators. The regression estimate is a higher length interval of the variance for the subpopulations than the ratio estimate. The discussed estimator generalized is a member of the generalized estimate of Srivastava and Jhaji [25]. The ratio estimate of high-level calibration is superior to Deng and Wu [23]. Based on the theoretical and empirical findings, we can conclude that the proposed estimate is more efficient than the regression estimates proposed by Srivastava and Jhaji [25] and Deng and Wu [23] for subpopulations 1, 2, 3, 4, and 5.

7.1. Recommendations and Applications. The following recommendations are given as follows:

- (1) The recommendation points have analysis through theoretical, and empirical for the ratio and generalized estimates for subpopulation total.
- (2) The present estimate is utilized when the domain total of the auxiliary variable is available but the number of units in the subpopulation is small.
- (3) The indirect estimate value also depends on how much the subpopulation value of the auxiliary variable is closed to the estimated sample of the population.

- (4) The subpopulation estimates of the variance of the high-level are a better option than the low-level variance estimates of ratio and regression estimates for the subpopulations. The high-level variance estimate of the ratio estimator can be introduced for the problems related to the health-related problems, environmental issues, and welfare programs like epidemiological issues, estimates for areas that are similar to estimates for those areas which are other parts.

Data Availability

The data is included within the study for finding the results.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] N. J. Purcell and L. Kish, "A biometrics invited paper. Estimation for small domains," *Biometrics*, vol. 35, no. 2, pp. 365–384, 1979.
- [2] J. N. K. Rao, "Some new developments in small area estimation," *Journal of the Iranian Statistical Society*, vol. 2, no. 2, pp. 145–169, 2003.
- [3] S. Singh, "Advanced sampling theory with applications: How Michael "selected" Amy," *Springer Science & Business Media*, vol. 2, 2003.
- [4] A. Rahman, *A Review of Small Area Estimation Problems and Methodological Developments*, NATSEM, University of Canberra, 2008.
- [5] J. N. K. Rao and I. Molina, *Small Area Estimation*, John Wiley & Sons, 2015.
- [6] S. Singh, S. Horn, and F. Yu, "Estimation of variance of the general regression estimator: higher level calibration approach," *Survey Methodology*, vol. 24, no. 1, pp. 41–50, 1998.

- [7] G. C. Tikkiwal, P. K. Rai, and A. Ghiya, "On the performance of generalized regression estimator for small domains," *Communications in Statistics - Simulation and Computation*, vol. 42, no. 4, pp. 891–909, 2013.
- [8] C. E. Särndal, "Efficient estimators with simple variance in unequal probability sampling," *Journal of the American Statistical Association*, vol. 91, no. 435, pp. 1289–1300, 1996.
- [9] A. C. Singh and C. A. Mohl, "Understanding calibration estimators in survey sampling," *Survey Methodology*, vol. 22, no. 2, pp. 107–116, 1996.
- [10] S. Singh, S. Horn, S. Chowdhury, and F. Yu, "Theory & methods: calibration of the estimators of variance," *Australian & New Zealand Journal of Statistics*, vol. 41, no. 2, pp. 199–212, 1999.
- [11] C. Wu and R. R. Sitter, "A model-calibration approach to using complete auxiliary information from survey data," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 185–193, 2001.
- [12] N. Koyuncu and C. Kadilar, "Calibration weighting in stratified random sampling," *Communications in Statistics - Simulation and Computation*, vol. 45, no. 7, pp. 2267–2275, 2016.
- [13] N. Koyuncu, "Calibration estimator of population mean under stratified ranked set sampling design," *Communications in Statistics - Theory and Methods*, vol. 47, no. 23, pp. 5845–5853, 2018.
- [14] B. B. Khare, A. Ashutosh, and P. K. Rai, "A comparative study of a class of direct estimators for domain mean with a direct ratio estimator for domain mean using auxiliary character," *Statistics in Transition New Series*, vol. 22, no. 2, pp. 189–200, 2021.
- [15] D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 663–685, 1952.
- [16] J. C. Deville and C. E. Särndal, "Calibration estimators in survey sampling," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 376–382, 1992.
- [17] E. Slud and T. Maiti, *Mean Squared Error Estimation in Transformed Fay-Herriot Models*, pp. 239–257, Journal of the Royal Statistical Society-Series B, 2006.
- [18] S. Bhusan, A. Kumar, and S. Kumar, "Efficient classes of estimators for population variance using attribute," *International Journal of Mathematics in Operational Research*, vol. 22, no. 1, pp. 74–92, 2021.
- [19] S. Bhusan, A. Kumar, S. Kumar, and S. Singh, "Some modified classes of estimators for population variance using auxiliary attribute," *Pakistan Journal of Statistics*, vol. 38, no. 2, pp. 235–252, 2022.
- [20] C. E. Särndal, B. Swensson, and J. H. Wretman, "The weighted residual technique for estimating the variance of the general regression estimator of the finite population total," *Biometrika*, vol. 76, no. 3, pp. 527–537, 1989.
- [21] T. Zaman and H. Bulut, "A new class of robust ratio estimators for finite population variance," *Scientia Iranica*, vol. 0, no. 0, p. 0, 2022.
- [22] F. Yates and P. M. Grundy, "Selection without replacement from within strata with probability proportional to size," *Journal of the Royal Statistical Society: Series B*, vol. 15, no. 2, pp. 253–261, 1953.
- [23] L. Y. Deng and C. F. J. Wu, "Estimation of variance of the regression estimator," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 568–576, 1987.
- [24] P. Singh and A. K. Srivastava, "Sampling schemes providing unbiased regression estimators," *Biometrika*, vol. 67, no. 1, pp. 205–209, 1980.
- [25] S. K. Srivastava and H. S. Jhajj, "A class of estimators of the population mean in survey sampling using auxiliary information," *Biometrika*, vol. 68, no. 1, pp. 341–343, 1981.
- [26] C. T. Isaki, "Variance estimation using auxiliary information," *Journal of the American Statistical Association*, vol. 78, no. 381, pp. 117–123, 1983.
- [27] C. F. J. Wu, "Variance estimation for the combined ratio and combined regression estimators," *Journal of the Royal Statistical Society: Series B*, vol. 47, no. 1, pp. 147–154, 1985.
- [28] W. A. Fuller, "Sampling with random stratum boundaries," *Journal of the Royal Statistical Society: Series B*, vol. 32, no. 2, pp. 209–226, 1970.
- [29] R. K. Singh and G. Singh, "A class of estimators with estimated optimum values in sample surveys," *Statistics & Probability Letters*, vol. 2, no. 6, pp. 319–321, 1984.
- [30] C. E. Särndal, B. Swensson, and J. Wretman, "Model assisted survey sampling is a universally," *Journal of Official Statistics*, vol. 8, no. 2, pp. 241–246, 1992.