VFAST Transactions on Mathematics
http://vfast.org/journals/index.php/VTM@ 2024 ISSN(e): 2309-0022, ISSN(p): 2411-6343
Volume 12, Number 2, July-December, 2024 pp: 01-16

Improved estimator for estimation of population mean using predictive approach under PPS sampling

Sajid Khan[1*], **Muhammad Farooq**[1], **Sohaib Ahmad**[2,3], **Sardar Hussain**[3]

[1]Department of Statistics, University of Peshawar, Peshawar, Pakistan; [2]Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan ; [3]Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan

Keywords: *Predictive approach, simulation study, PPS sampling, study and auxiliary variables, bias, MSE, PRE.*

Subject Classification: *62D05*

Journal Info:
Submitted: September 05, 2024
Accepted: September 25, 2024
Published: September 29, 2024

Abstract In this study, we apply a predictive approach to the problem of finding new estimators for the estimation of finite population mean using auxiliary variable under probability proportional to size (PPS) sampling. This work is deemed novel due to the fact that, as far as we are aware, no one has before investigated the predictive approach to estimating the finite population mean under probability proportional to size sampling. The expressions for the bias and mean square error (MSE) are derived up to the first order. In order to verify the theoretical results, numerical and simulation investigations are conducted respectively. Based on the numerical result, it is shown that the suggested estimator performs well in terms of minimum MSE and higher percentage relative efficiency (PRE). The conditions under which the suggested estimator is more efficient than the other estimators are described numerically.

***Correspondence Author Email Address:**
sohaib_ahmad@awkum.edu.pk
DOI: 10.21015/vtm.v12i2.1942

1 Introduction

In survey sampling, it is to be noted that generally the use of the auxiliary information either at the design or at the estimation stage or at both stages may increase the efficiency of an estimator. In the literature of survey sampling, auxiliary information may be used in ratio, product, and regression type estimators when there is correlation among the variables. A number of estimators have been developed using unique ideas by combining ratio, product, regression, or exponential type estimators, where each estimator takes advantage of the correlation among the variables. In order to make the estimators more accurate and efficient, survey statisticians progressively use data regarding auxiliary variables. While measuring the main variable could be costly, auxiliary information is usually easy to quantify. Numerous surveys make use of the auxiliary variable to improve the estimators' ability by taking advantage of the relationship between it and the variables being examined, either in the sample design or during the estimating stage.

Traditional methods for estimating population means, such as ratios, regression, and exponentials, have seen extensive application in recent years due to their computational ease and simple structures. In order to better estimate the unknown population mean of the study variable using a number of auxiliary variable-related characteristics, including the population mean, skewness, kurtosis, and coefficients of variation, several scholars have enhanced the standard ratio and exponential estimators. For the ratio, exponential, and regression estimation methods to work, it is necessary to know the auxiliary variable's population characteristics beforehand. Researchers can develop these research findings by examining [1], [2], [6], and [7].

When the population of interest is homogeneous, then simple random sampling provides better results. Alternatively, when the sampling units change considerably in size, then units may be chosen by adopting the probability proportional to size (PPS) sampling. For instance, if we are required to compute the population in a town in a city, we may choose a variable on which we have information as the auxiliary variable, e.g.,...

(a) Size of each town in the area (correlation with auxiliary variable = 0.85, say).

(b) Number of families in each town in the area (correlation with the study variable = 0.96, say).

Using the above evidence, we select the auxiliary variables that have the higher correlation with the study variable. Therefore, we select information on (b) as an auxiliary variable whenever we select a sampling using probability proportional without replacement (PPS) sampling. Similarly, if we carried out a sample survey related to the income of families, the size of families may vary with respect to its sizes.

[8] suggested an enhanced estimator using the auxiliary information under PPS sampling. [9] recommended an improved estimator using two-phase sampling under PPS sampling. [10] discussed optimum estimation for estimation of PPS using multi-auxiliary information. [11] and [12] discussed novel estimators for the estimation of the mean using PPS sampling.

Survey sampling describes various methodologies, including model-based and design-based approaches that use the auxiliary information to build more effective estimators. An encompassing framework for statistical inferences regarding the environment of a finite population can be model-based theory or prediction methods. Deriving statistical inferences regarding the population parameters of the study variable can be achieved by applying model-based theory or predictive estimate theory of survey sampling. In order to draw valid conclusions about aspects of a population, a sample survey is an easy, cheap, and practical technique to collect data. The main purpose of a sample survey is to get an accurate representation of the population. So, to make parameter estimators more efficient, one can use the auxiliary information. Various approaches for estimating population means using the auxiliary data are discussed in the literature, including product, ratio, and regression methods.

Using a prediction approach, [14] attempted to determine the population mean. He combined the unobserved unit mean with the understudied population mean based on his predictions or estimations of the unobserved unit mean. Following the work in [15], [16], and [17], some improved estimators for population mean using a predictive approach in the case of simple and successive sampling were considered. The authors in [18], [19], and [20] suggested product type and some alternative predictive estimators. The authors in [21]–[47] developed modified estimators using auxiliary information under predictive and other related estimators for the estimation of population parameters.

The studies discussed up to this point are either as efficient as, or even less efficient than, the conventional regression estimator. We introduce a new class of estimators for the research variable's population mean estimation in this study, which uses data from auxiliary variables derived from simple random sampling. The primary goal of this work has been outlined by:

1. The goal of this study is to introduce a novel estimator for population mean using auxiliary variables under PPS sampling using a predictive approach.
2. The properties of the recommended estimator are established up to the first order of approximation.
3. The proposed estimator is shown to be useful in a simulation study as well as with actual data.
4. We believe that our proposed estimator could be useful in many contexts and brings something new and substantial to the study of population mean estimation.

The remainder of the article is organized as follows. Section 2 contains the materials and methods related to the article. The existing estimators are given in Section 3. The suggested estimator is given in Section 4. To check the efficiency of our suggested estimator, we present a mathematical study in Section 5. The discussion of the article is given in Section 6. Argumentation is given in Section 7. Lastly, concluding remarks are provided in Section 8.

Let a population $\omega = (\omega_1, \omega_2, \omega_3, \ldots, \omega_N)$ of size N having distinct and recognizable units. Let y_i and (x_i, z_i, q_i) be the characteristics of the study variable (Y) and the auxiliary variables (X, Z, Q) respectively. Let a sample n be chosen from a population by consuming PPSWR. Let

$$P_i = \frac{q_i}{\sum_{i=1}^{N} q_i}$$

represent the probability of choosing an auxiliary variable from a set of known values Q. Let Y, (X, Z, Q) represent the population mean corresponding to the sample means $\bar{y}$, $(\bar{x}, \bar{z}, \bar{q})$ respectively.

We define the following symbols under PPS sampling:

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i, \quad \bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i, \quad \bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i,$$

where

$$u_i = \frac{y_i}{NP_i}, \quad v_i = \frac{x_i}{NP_i}, \quad w_i = \frac{z_i}{NP_i}.$$

The errors are defined as:

$$e_0 = \frac{\bar{u} - \bar{Y}}{\bar{Y}}, \quad e_1 = \frac{\bar{v} - \bar{X}}{\bar{X}}, \quad e_2 = \frac{\bar{w} - \bar{Z}}{\bar{Z}},$$

with the following properties:

$$E(e_i) = 0, \quad i = 0, 1, 2,$$

$$E(e_0^2) = \vartheta C_u^2, \quad E(e_1^2) = \vartheta C_v^2, \quad E(e_2^2) = \vartheta C_w^2,$$

$$E(e_0 e_1) = \vartheta \rho_{uv} C_u C_v, \quad E(e_0 e_2) = \vartheta \rho_{uw} C_u C_w, \quad E(e_1 e_2) = \vartheta \rho_{vw} C_v C_w,$$

where

$$C_u = \frac{\sigma_u}{\bar{Y}}, \quad C_v = \frac{\sigma_v}{\bar{X}}, \quad C_w = \frac{\sigma_w}{\bar{Z}},$$

$$\sigma_u = \sum_{i=1}^{N} P_i (u_i - \bar{Y})^2, \quad \sigma_v = \sum_{i=1}^{N} P_i (v_i - \bar{X})^2, \quad \sigma_w = \sum_{i=1}^{N} P_i (w_i - \bar{Z})^2,$$

$$\sigma_{uv} = \sum_{i=1}^{N} P_i (u_i - \bar{Y})(v_i - \bar{X}), \quad \sigma_{uw} = \sum_{i=1}^{N} P_i (u_i - \bar{Y})(w_i - \bar{Z}), \quad \sigma_{vw} = \sum_{i=1}^{N} P_i (v_i - \bar{X})(w_i - \bar{Z}).$$

Anticipated values that were not included in the sample using a predictive method lead to the following population mean estimate under PPS sampling:

$$\bar{Y} = \frac{\vartheta(s)}{N} \bar{u} + \frac{N - \vartheta(s)}{N} \bar{Y}_s,$$

where

$$\bar{u} = \frac{1}{\vartheta(s)} \sum_{i \in s} u_i, \quad \bar{Y}_s = \frac{1}{N - \vartheta(s)} \sum_{i \in \bar{s}} u_i,$$

are the sample mean and non-sampled unit mean estimators of the population mean $\bar{Y}$. This leads to the following prediction for the overall population mean $\bar{Y}$ when we use PPS with size $\vartheta(s) = n$:

$$\bar{Y} = \frac{n}{N} \bar{u} + \frac{N - n}{N} \bar{Y}_s.$$

The mean estimate for the population is given by:

$$t_{i(\text{pps})} = f \bar{u} + (1 - f) \alpha_{i(\text{pps})},$$

where

$$f = \frac{n}{N}, \quad \vartheta = \frac{1}{1 - f}.$$

Existing Estimators

Several existing estimators under PPS sampling using the predictive approach have been discussed:

1. **Traditional Estimator**:

$$\alpha_{1(\text{pps})} = \bar{u}$$

The variance of $\alpha_{1(\text{pps})}$ is:

$$\text{Var}(\alpha_{1(\text{pps})}) = \vartheta \bar{Y}^2 C_u^2.$$

2. **Usual Ratio Estimator**:

$$\alpha_{2(\text{pps})} = f \bar{u} + (1 - f) \frac{\bar{X}_s}{\bar{v}}$$

The bias and mean squared error (MSE) of $\alpha_{2(\text{pps})}$ are:

$$\text{Bias}(\alpha_{2(\text{pps})}) \approx \vartheta\bar{Y}(C_v^2 - \rho_{uv}C_uC_v),$$

$$\text{MSE}(\alpha_{2(\text{pps})}) \approx \vartheta\bar{Y}^2(C_u^2 + C_v^2 - 2\rho_{uv}C_uC_v).$$

3. **Traditional Exponential Estimator**:

$$\alpha_{3(\text{pps})} = f\bar{u} + (1-f)\bar{u}\exp\left(\frac{\bar{X}_s - \bar{v}}{\bar{X}_s + \bar{v}}\right)$$

The bias and MSE of $\alpha_{3(\text{pps})}$ are:

$$\text{Bias}(\alpha_{3(\text{pps})}) \approx \vartheta\bar{Y}\left(\frac{3-4f}{8(1-f)}C_v^2 - \frac{\rho_{uv}C_uC_v}{2}\right),$$

$$\text{MSE}(\alpha_{3(\text{pps})}) \approx \vartheta\bar{Y}^2\left(C_u^2 + \frac{C_v^2}{4} - \rho_{uv}C_uC_v\right).$$

4. **Difference Estimator**:

$$\alpha_{4(\text{pps})} = f\bar{u} + (1-f)\left[\bar{u} + d\left(\frac{\bar{X}_s - \bar{v}}{\bar{v}}\right)\right],$$

where d is a constant. The MSE of $\alpha_{4(\text{pps})}$ is:

$$\text{MSE}(\alpha_{4(\text{pps})}) = \vartheta\bar{Y}^2C_u^2(1-\rho_{uv}^2).$$

Proposed Estimator

An estimator's efficiency during the design and estimating stages can be enhanced with the proper utilization of an auxiliary variable. Moreover, numerous studies have collected data on auxiliary variables and population unit characteristics. Since they are straightforward to record and cost-effective, the study variable and several auxiliary variables that are linearly connected are recorded. In order to estimate the population mean of the research variable, a mixture estimator is used together with these auxiliary variables, which are often seen as both quantitative and qualitative (attributes) variables. Taking motivation from the difference in difference estimator and [3] exponential ratio type estimator, we suggest a new improved estimator for estimation of population mean using predictive approach under PPSWR, which is given by:

$$\alpha_{5(PPS)} = f\bar{u} + (1-f)\left[\bar{u} + K_1\left(\bar{X}_s - \bar{v}\right) + K_2\left(\bar{Z}_s - \bar{w}\right)\right]\exp\left(\frac{\bar{X}_s - \bar{v}}{\bar{X}_s + \bar{v}}\right)$$

where

$$\bar{X}_s = \frac{1}{N-n}\sum_{i\in s^c}\bar{u}_i = \frac{N\bar{X} - n\bar{v}_s}{N-n}, \quad \bar{Z}_s = \frac{1}{N-1}\sum_{i\in s^c}\bar{w}_i = \frac{N\bar{Z} - n\bar{w}_s}{N-n}$$

$$\alpha_{5(PPS)} = f\bar{u} + (1-f)\left[\bar{u} + K_1\left(\bar{X}_s - \bar{v}\right) + K_2\left(\bar{Z}_s - \bar{w}\right)\right]\exp\left(\frac{\bar{X}_s - \bar{v}}{\bar{X}_s + \bar{v}}\right)$$

$$\alpha_{5(\text{PPS})} = \left[f\overline{Y}\left(1+e_0\right) + (1-f)\left(\overline{Y}\left(1+e_0\right) - \frac{K_1\overline{X}e_1}{1-f} - \frac{K_2\overline{Z}e_2}{1-f}\right)\right]\left[1 - \frac{e_1}{2(1-f)} + \frac{3-4f}{8(1-f)^2}e_1^2\right]$$

$$\alpha_{5(\text{PPS})} - \overline{Y} = \overline{Y}\left[e_0 - \frac{e_1}{2} + \frac{3-4f}{8(1-f)}e_1^2 - \frac{e_0e_1}{2}\right] - K_1\overline{X}\left[e_1 - \frac{e_1^2}{2(1-f)}\right] - K_2\overline{Z}\left[e_2 - \frac{e_1e_2}{2(1-f)}\right]$$

Taking both sides expectation of (10), we get the bias of our proposed estimators, which is given by:

$$\text{Bias}(\alpha_{5(\text{PPS})}) = \vartheta\overline{Y}\left[\frac{3-4f}{8(1-f)}C_v^2 - \frac{\rho_{uv}C_uC_v}{2}\right] + K_1\left(\frac{\overline{X}\vartheta C_v^2}{2(1-f)}\right) + K_2\overline{Z}\vartheta\rho_{uv}C_uC_v$$

Squaring and taking expression on both sides of (10) we got MSE expression, which is given by:

$$\text{MSE}(\alpha_{5(\text{PPS})}) = \vartheta\overline{Y}^2\left[C_u^2 + \frac{C_v^2}{n} - \rho_{uv}C_uC_v\right] - \frac{AD^2 + BC^2 - 2CDE}{AB - E^2}$$

where

$$A = \vartheta C_v^2, \quad B = \vartheta C_w^2, \quad C = \vartheta\left(\rho_{uv}C_uC_v - \frac{C_v^2}{n}\right), \quad D = \vartheta\left(\rho_{uv}C_uC_v - \frac{\rho_{uv}C_uC_v}{2}\right), \quad E = \vartheta\rho_{uv}C_uC_v$$

The values of $K_{i(\text{opt})}$ (i = 1, 2) are given by:

$$K_1(\text{opt}) = \frac{\overline{Y}}{\overline{\overline{X}}}\left(\frac{BC - DE}{AB - E^2}\right), \quad K_2(\text{opt}) = \frac{\overline{Y}}{\overline{\overline{Z}}}\left(\frac{AD - CE}{AB - E^2}\right)$$

Numerical Study

Here, we compare the performance of our suggested estimator to existing estimators by means of a numerical study. We take three real data sets. The PRE of $\alpha_{5(\text{PPS})}$ estimator with respect to $\alpha_{1(\text{PPS})}$ is carried out using the below expressions:

$$\text{PRE} = \frac{\text{Var}(\alpha_{1(\text{PPS})})}{\text{MSE}(\alpha_{i(\text{PPS})})} \times 100, \quad i = (1,2,3,4,5)$$

Population-1: [Source: [5]]

Let

$$Y = \text{Total fertility rate}, \quad X = \text{Crude birth rate}, \quad Z = \text{Crude death rate}, \quad Q = \text{Expectation of life at birth}$$

$$N = 96, \quad n = 20, \quad f = 0.2083333, \quad \vartheta = 0.05, \quad \overline{Y} = 3.501562, \quad \overline{X} = 26.01146, \quad \overline{Z} = 10.87187$$

$$\overline{u} = 19.38656, \quad \overline{v} = 143.022, \quad \overline{w} = 60.60931$$

$$S_u^2 = 6.87443, \quad S_u = 2.621914, \quad S_v^2 = 327.7338, \quad S_v = 18.10342, \quad S_w^2 = 68.62184, \quad S_w = 8.28383$$

$$C_u = 0.7487839, \quad C_v = 0.6959786, \quad C_w = 0.7619505$$

$$C_u^2 = 0.5606773, \quad C_v^2 = 0.4843863, \quad C_w^2 = 0.5805685$$

$$\rho_{uv} = 0.99267, \quad \rho_{uw} = 0.7917016, \quad \rho_{vw} = 0.7924682$$

$$S_{uv} = 47.11768, \quad S_{uw} = 17.19535, \quad S_{vw} = 118.843, \quad C_{uv} = 0.5173176, \quad C_{uw} = 0.4516944, \quad C_{vw} = 0.5173176$$

Population-2: [Source: [5]]

Let

Y = Fish trapped in 1995, X = Fish trapped in 1994, Z = Fish trapped in 1993, Q = Fish trapped in 1992

$$N = 69,\quad n = 12,\quad f = 0.173913,\quad \vartheta = 0.08333333,\quad \overline{Y} = 4514.899,\quad \overline{X} = 4954.435,\quad \overline{Z} = 4591.072$$

$$\overline{u} = 30440.48,\quad \overline{v} = 33470.67,\quad \overline{w} = 34206.84$$

$$S_u^2 = 3361460,\quad S_u = 1833.428,\quad S_v^2 = 3599389,\quad S_v = 1897.205,\quad S_w^2 = 2245140,\quad S_w = 1498.379$$

$$C_u = 0.4060841,\quad C_v = 0.3829308,\quad C_w = 0.326368$$

$$C_u^2 = 0.1649043,\quad C_v^2 = 0.146636,\quad C_w^2 = 0.1065161$$

$$\rho_{uv} = 0.7546307,\quad \rho_{uw} = 0.4251362,\quad \rho_{vw} = 0.5370245$$

$$S_{uv} = 2624900,\quad S_{uw} = 1167922,\quad S_{vw} = 1526617,\quad C_{uv} = 0.1173466,\quad C_{uw} = 0.05634451,\quad C_{vw} = 0.1173466$$

Population-3: [Source: [13]]

Let

Y = Total beds, X = Allocated for Covid, Z = Used by Covid, Q = Patients treated

$$N = 38,\quad n = 12,\quad f = 0.3157895,\quad \vartheta = 0.08333333,\quad \overline{Y} = 9965.342,\quad \overline{X} = 13932.08,\quad \overline{Z} = 21217.45$$

$$\overline{u} = 33893.82,\quad \overline{v} = 40838.68,\quad \overline{w} = 75067.66$$

$$S_u^2 = 69251593,\quad S_u = 8321.754,\quad S_v^2 = 39362574,\quad S_v = 6273.96,\quad S_w^2 = 76580967,\quad S_w = 8751.055$$

$$C_u = 0.8350696,\quad C_v = 0.4503247,\quad C_w = 0.4124462$$

$$C_u^2 = 0.6973412,\quad C_v^2 = 0.2027924,\quad C_w^2 = 0.1701119$$

$$\rho_{uv} = 0.3599156,\quad \rho_{uw} = 0.7039386,\quad \rho_{vw} = 0.5709211$$

$$S_{uv} = 18791322,\quad S_{uw} = 51263714,\quad S_{vw} = 31345718,\quad C_{uv} = 0.1353472,\quad C_{uw} = 0.2424514,\quad C_{vw} = 0.1353472$$

Table 1. Table 1: MSE and PRE using real data sets

Estimators	MSE(Data 1)	PRE(Data 1)	MSE(Data 2)	PRE(Data 2)	MSE(Data 3)	PRE(Data 3)
$\alpha_{(1(PPS))}$	0.3437215	100	280121.6	100	5770966	100
$\alpha_{(2(PPS))}$	0.006393021	5376.512	130539.3	214.5879	5209032	110.7877
$\alpha_{(3(PPS))}$	0.005020491	6846.372	120901.5	231.6941	5023399	114.8817
$\alpha_{(4(PPS))}$	0.1008194	340.928	143058.1	195.8096	5070438	113.8159
$\alpha_{(5(PPS))}$	0.004997	6878.557	120445.9	232.5705	2896195	199.2602

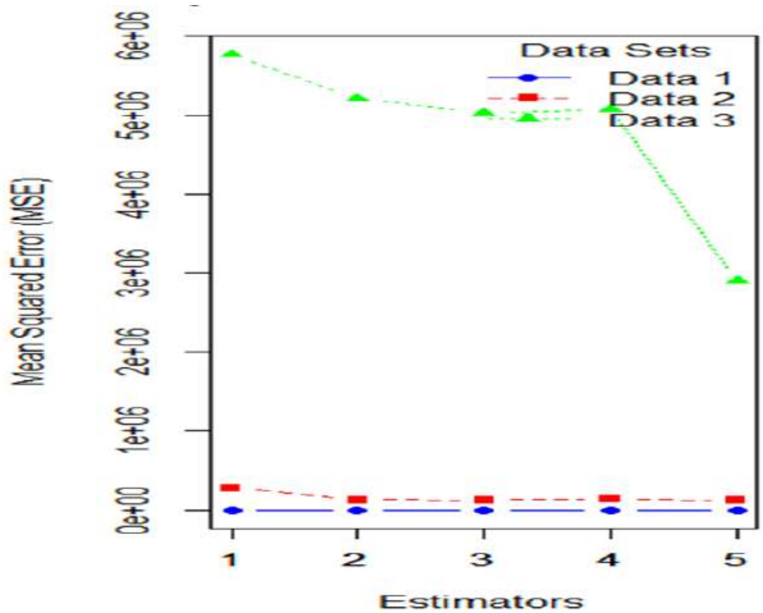

Figure 1. Show MSEs based on real data sets

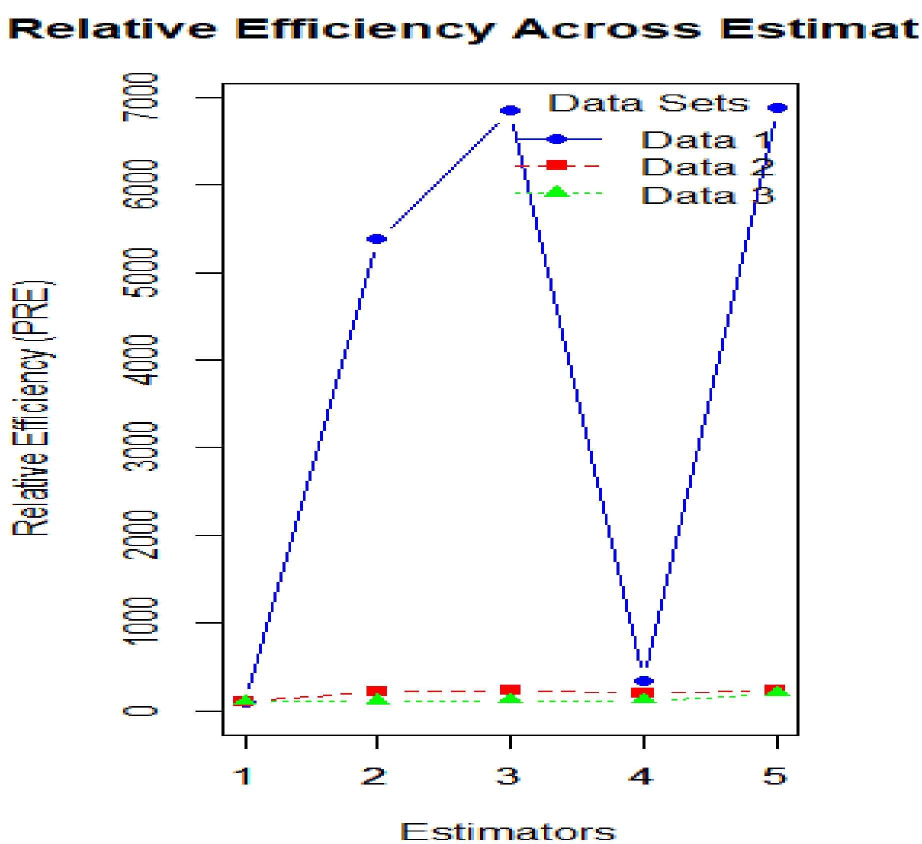

Figure 2. Show PREs based on real data sets

Simulation study

In this part, we show how three simulated data sets with sizes of 5,000 were generated utilizing various covariance matrices followed by a multivariate normal distribution.

Population-I:

$$\mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \\ 500 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 1000 & 800 & 810 \\ 850 & 830 & 815 \\ 840 & 900 & 925 \end{bmatrix}$$

Population-II:

$$\mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \\ 500 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 1600 & 860 & 830 \\ 900 & 920 & 940 \\ 830 & 860 & 880 \end{bmatrix}$$

Population-III:

$$\mu_1 = \begin{bmatrix} 500 \\ 500 \\ 500 \\ 500 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 450 & 370 & 320 \\ 280 & 600 & 200 \\ 230 & 300 & 555 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 420 & 430 & 415 \end{bmatrix}$$

Table 2. Table 2: MSE and PRE using simulation study

Estimators	MSE (Data 1)	PRE (Data 1)	MSE(Data 2)	PRE(Data 2)	MSE(Data 3)	PRE(Data 3)
$\alpha_{(1(PPS))}$	6.036505	100	0.3177546	100	0.1468262	100
$\alpha_{(2(PPS))}$	0.06800332	8876.779	0.005891808	5393.16	0.001745721	8410.631
$\alpha_{(3(PPS))}$	0.05967757	10115.2	0.004631402	6860.873	0.001519015	9665.877
$\alpha_{(4(PPS))}$	1.667502	362.0088	0.09316036	341.0835	0.04077223	360.1131
$\alpha_{(5(PPS))}$	0.05964771	10120.26	0.00460999	6892.74	0.001517999	9672.347

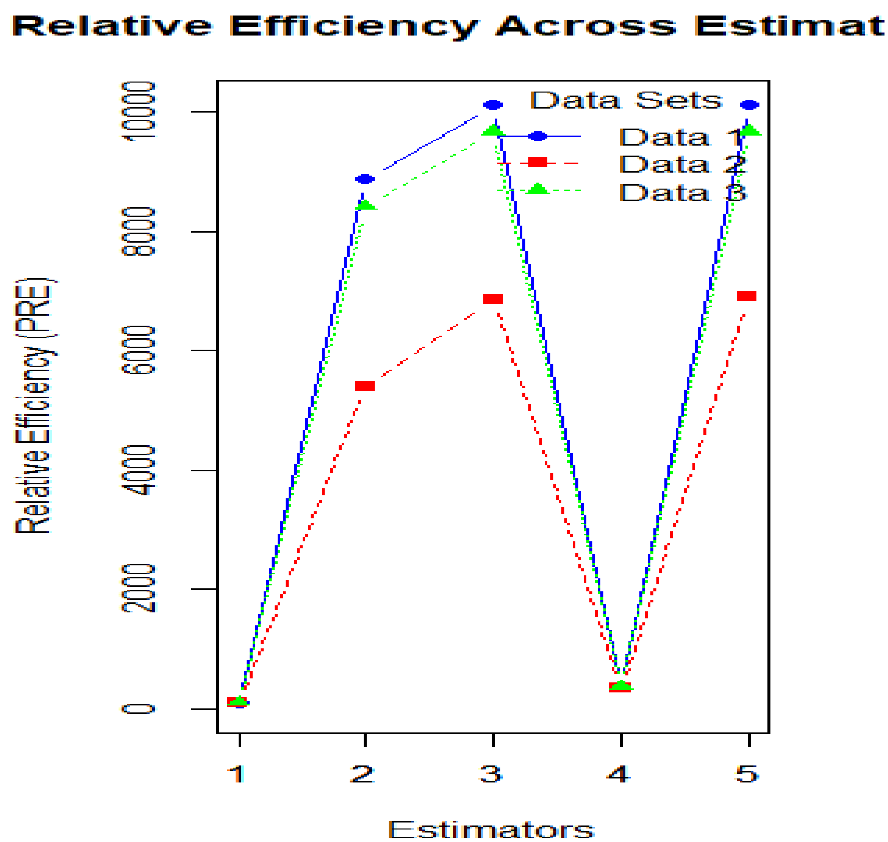

Figure 4. Show the PREs of all estimators based on simulation study

Discussion

The proposed estimators are demonstrated to be practical by administering sample surveys and evaluating the efficiency of each estimator with respect to actual data. This section discusses the study main points, offers an analysis and interpretation of those points, and explains and interprets the results. To evaluate the impact the suggested generalized estimators, we conducted a simulation study using four actual data sets. Tables 1 and Table 2 illustrate the mathematical results of MSE and PRE based on real data sets and a simulation research. From the numerical results of MSE and PRE which are displayed in

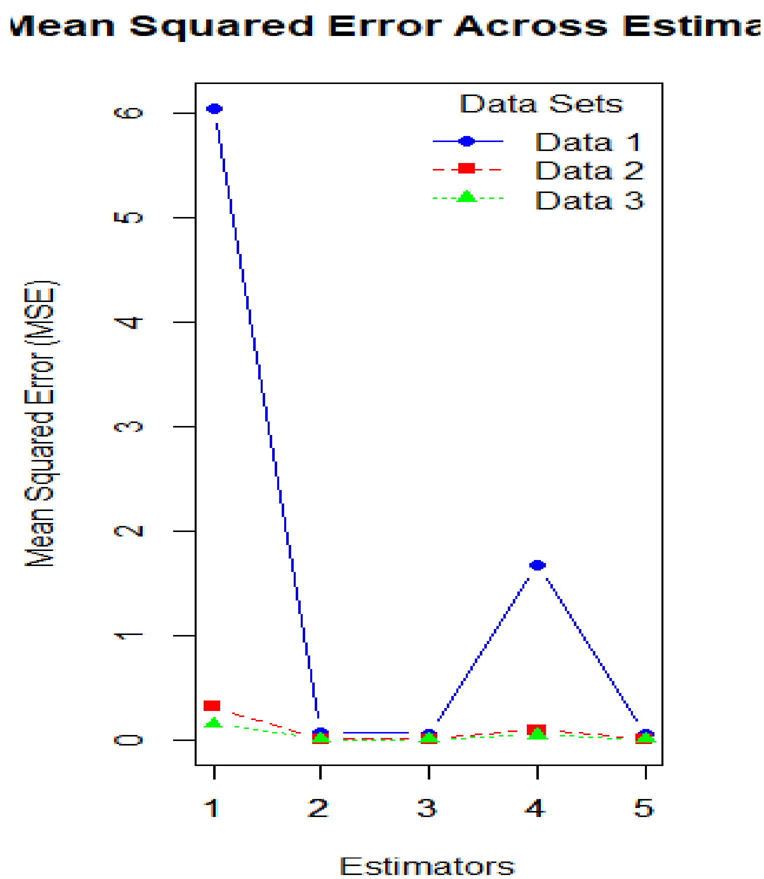

Figure 3. Show the MSEs of all estimators based on simulation study

Table 1, it is noted that the suggested estimator has minimum MSE and higher PRE. The result of MSEs and PREs for the real data sets and a simulation study is also shown with the help of graphs, which is given in Figure 1-4. The results show that the proposed estimators under the probability proportional to size (PPS) prediction method significantly influence Data-I, Data-II, and Data-III. The simulation study may possibly validate the same findings. The productivity of the suggested class of estimators is highlighted as being superior to that of its existing counterparts constructed utilizing the numerical example.

Conclusion

In this article, we suggest improved estimator for estimation of population mean using predictive method under probability proportional to size (PPS) sampling. It is conceivable to derive the numerical computations of the bias and MSEs of all estimators up to the first approximation. Three real data sets are employed to validate the efficiency of an estimator. The result of MSE and PRE is given in Table 1. The efficiency situations are further analyzed with simulation investigation. The results based on simulation study are given in Table 2. With a smaller MSE and a higher PRE than the current estimators, the suggested estimator in these Tables performs better. Based on the numerical results it has been shown that the suggested estimator have minimum MSE and higher PRE. Given its enhanced performance, the suggested estimator is recommended for implementation across diverse disciplines of study, such as to the biological and agricultural sciences, commerce, engineering, economics, fisheries, medical and social sciences. It is possible to extend the current work by estimating the population means using a predictive technique under systematic sampling. Finally, survey statisticians and practitioners may be interesting in the proposed estimators for their potential practical applications due to their positive characteristics under predictive PPS sampling.

Funding

There is no specific funding to conduct this study.

Conflict of interest

The authors declares no conflict of interest.

Data availability

All the data are available within the manuscript.

Author Contributions

Sajid Khan: Writing main manuscript **Mohammad Farooq**: Supervision **Sohaib Ahmad**: Writting and Review the main manuscript, data curation, analysis, and methodology **Sardar Hussain**: conceptualization, software, data analysis and methodology.

Compliance with Ethical Standards

It is declared that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in this study.

Author Information

ORCID:
Sajid Khan: 0009-0009-7880-5769
Muhammad Farooq: 0009-0002-2408-3819
Sohaib Ahmad: 0000-0003-2582-2265
Sardar Khan: 0000-0001-9274-0396

References

[1] Ahmad, S., Shabbir, J., Aamir, M., Hussain, S., & Ullah, K. (2024) 'Improved Mean Estimation for a Finite Population with Twofold Auxiliary Variables Under Simple Random Sampling', International Journal of Applied and Computational Mathematics, 10(1), p. 31.

[2] Ahmad, S., Hussain, S., Al Mutairi, A., Kamal, M., Rehman, M. U., & Mustaafa, M. S. (2024) 'Improved estimation of population distribution function using twofold auxiliary information under simple random sampling', Heliyon, 10(2).

[3] Bahl, S., & Tuteja, R. (1991) 'Ratio and product type exponential estimators', Journal of Information and Optimization Sciences, 12(1), pp. 159-164.

[4] Bhushan, S., & Kumar, A. (2023) 'Evaluating the performance of logarithmic type estimators using auxiliary attribute', Life Cycle Reliability and Safety Engineering, pp. 1-8.

[5] Singh, S. (2003) Advanced Sampling Theory With Applications: How Michael "Selected" Amy. Vol. 2. Springer Science & Business Media.

[6] Singh, H. P., Gupta, A., & Tailor, R. (2023) 'Efficient class of estimators for finite population mean using auxiliary attribute in stratified random sampling', Scientific Reports, 13(1), p. 10253.

[7] Yadav, R., Singh, H. P., Upadhyaya, L. N., & Yadav, S. K. (2024) 'Adroit family of estimators of population mean using known auxiliary parameters', Journal of Computational and Applied Mathematics, 437, p. 115455.

[8] Al-Marzouki, S., Chesneau, C., Akhtar, S., Nasir, J. A., Ahmad, S., Hussain, S., ... & El-Morshedy, M. (2021) 'Estimation of finite population mean under PPS in presence of maximum and minimum values', AIMS Mathematics, 6(5), pp. 5397-5409.

[9] Wang, J., Ahmad, S., Arslan, M., Lone, S. A., Abd Ellah, A. H., Aldahlan, M. A., & Elgarhy, M. (2023) 'Estimation of finite population mean using double sampling under probability proportional to size sampling in the presence of extreme values', Heliyon, 9(11).

[10] Amab, R. (2004) 'Optimum estimation of a finite population total in PPS sampling with replacement for multi-character surveys', Journal of Indian Society of Agricultural Statistics, 58(2), pp. 231-243.

[11] Ahmad, S., Shabbir, J., Zahid, E., Aamir, M., & Alqawba, M. (2023) 'New generalized class of estimators for estimation of finite population mean based on probability proportional to size sampling using two auxiliary variables: A simulation study', Science Progress, 106(4), p. 00368504231208537.

[12] Singh, H. P., Mishra, A. C., & Pal, S. K. (2018) 'Improved estimator of population total in PPS sampling', Communications in Statistics-Theory and Methods, 47(4), pp. 912-934.

[13] Punjab Bureau of Statistics (PBS), Pakistan (2021–2022).

[14] Basu, D. (1971) 'An essay on the logical foundations of survey sampling, Part I', Foundations of Statistical Inferences, V. P. Godambe and D. A. Sprott, pp. 203-204.

[15] Bhushan, S., Jaiswal, P., & Pandey, S. (2020) 'An improved predictive approach for estimation of population mean', International Journal of Research in Applied Science and Engineering Technology, 8(IX), pp. 251-256.

[16] Grover, L. K., & Sharma, A. (2023) 'Predictive estimation of finite population mean in case of missing data under two-phase sampling', Journal of Statistical Theory and Applications, 22(4), pp. 283-308.

[17] Maji, R., Singh, G. N., & Bandyopadhyay, A. (2021) 'On the use of predictive approach in estimation of population mean under successive sampling scheme', Journal of Information and Optimization Sciences, 42(6), pp. 1237-1270.

[18] Ahmad, S., Hussain, S., & Ahmad, S. (2021) 'Finite population distribution function estimation using auxiliary information under simple random sampling', Statistics, Computing and Interdisciplinary Research, 3(1), pp. 29-38.

[19] Ahmad, S., Hussain, S., Aamir, M., Khan, F., Alshahrani, M. N., & Alqawba, M. (2022) 'Estimation of finite population mean using dual auxiliary variable for non-response using simple random sampling', AIMS Mathematics, 7(3), pp. 4592-4613.

[20] Yadav, S. K. (2022) 'Modified Searls predictive estimation of population mean using known auxiliary character', Thailand Statistician, 20(1), pp. 53-65.

[21] Yadav, S. K., & Mishra, S. S. (2015) 'Developing improved predictive estimator for finite population mean using auxiliary information', Statistika: Statistics and Economy Journal, 95(1).

[22] Hussain, S., Zichuan, M., Hussain, S., Iftikhar, A., Asif, M., Akhtar, S., & Ahmad, S. (2020) 'On estimation of distribution function using dual auxiliary information under nonresponse using simple random sampling', Journal of Probability and Statistics, 2020(1), p. 1693612.

[23] Iftikhar, A., Hongbo, S., Hussain, S., Khan, S., Alnssyan, B., Shabbir, J., ... & Aamir, M. (2022) 'A new exponential factor-type estimator for population distribution function using dual auxiliary variables under stratified random sampling', Mathematical Problems in Engineering, 2022(1), p. 7673945.

[24] Semary, H. E., Ahmad, S., Hamdi, W. A., Albalawi, O., Elbatal, I., Chesneau, C., & Almarzouki, S. M. (2024) 'Improved novel estimation for estimation of population distribution function using auxiliary information under stratified sampling strategy', Journal of Radiation Research and Applied Sciences, 17(3), p. 101011.

[25] Ahmad, S., Adichwal, N. K., Aamir, M., Shabbir, J., Alsadat, N., Elgarhy, M., & Ahmad, H. (2023) 'An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling', Scientific Reports, 13(1), p. 21444.

[26] Ahmad, S., Ullah, K., Zahid, E., Shabbir, J., Aamir, M., Alshanbari, H. M., & El-Bagoury, A. A. A. H. (2023) 'A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling', Scientific Reports, 13(1), p. 5415.

[27] Nayak, R., & Sahoo, L. (2012) 'Some alternative predictive estimators of population variance', Revista Colombiana de Estadística, 35(3), pp. 509-521.

[28] Srivastava, S. K. (1983) 'Predictive estimation of finite population mean using product estimator', Metrika, 30(1), pp. 93-99.

[29] Sahoo, L. N., & Sahoo, R. K. (2001) 'Predictive estimation of finite population mean in two phase sampling using two auxiliary variables', Journal of Indian Society of Agricultural Statistics, 54(4), pp. 258-264.

[30] Ahmad, S., & Shabbir, J. (2024) 'Proportion estimation using enhanced class of estimators under simple random sampling: Application with real data sets and simulation', Measurement: Interdisciplinary Research and Perspectives, pp. 1-13.

[31] Ahmad, S., Masood, S., Alomair, A. M., & Alomair, M. A. (2024) 'Improved modified estimator for estimation of median using auxiliary information under simple random sampling', Scientific Reports, 14(1), p. 16504.

[32] Semary, H. E., Ahmad, S., Elbatal, I., Chesneau, C., Elgarhy, M., & Almetwally, E. M. (2024) 'New comprehensive class of estimators for population proportion using auxiliary attribute: Simulation and an application', Alexandria Engineering Journal, 99, pp. 130-136.

[33] Hussain, S., Ahmad, S., Saleem, M., & Akhtar, S. (2020) 'Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling', PLOS ONE, 15(9), p. e0239806.

[34] Ahmad, S., Zahid, E., Shabbir, J., & Aamir, M. (2024) 'New generalized family of estimators for finite population mean based on stratified random sampling', Scientific Reports, 14(1), p. 1720.

[35] Ahmad, S., Hussain, S., Zahid, E., Ullah, K., Shabbir, J., & Aamir, M. (2024) 'New family of estimators for estimating population mean using auxiliary variable in stratified random sampling', Journal of Computational and Applied Mathematics, 437, p. 115455.

[36] Ahmad, S., & Shabbir, J. (2024) 'A new class of estimators for estimation of finite population mean under stratified random sampling using two auxiliary variables', Measurement: Interdisciplinary Research and Perspectives, pp. 1-13.

[37] Yadav, S. K., & Mishra, S. S. (2019) 'Improved estimation of population mean using known median of auxiliary character in two-phase sampling', Journal of Reliability and Statistical Studies, 12(1), pp. 15-27.

[38] Srivastava, S. K. (1983) 'Predictive estimation of finite population mean using product estimator', Metrika, 30(1), pp. 93-99.

[39] Maiti, M. (1989) 'Some alternative predictive estimators of finite population mean using two auxiliary variables', Journal of Statistical Planning and Inference, 23(1), pp. 131-144.

[40] Basu, D. (1971) 'An essay on the logical foundations of survey sampling, Part I', Foundations of Statistical Inferences, V. P. Godambe and D. A. Sprott, pp. 203-204.

[41] Abbas, N., & Mukherjee, S. (2021) 'A new exponential ratio-type estimator for finite population mean using auxiliary information', Journal of Reliability and Statistical Studies, 14(1), pp. 59-75.

[42] Ahmed, S. E. (1996) 'Estimation of population variance in simple random sampling using auxiliary information', Journal of Indian Society of Agricultural Statistics, 48(2), pp. 107-115.

[43] Shabbir, J., & Gupta, S. (2005) 'A new estimator of population variance in stratified sampling', Journal of the Korean Statistical Society, 34(3), pp. 295-301.

[44] Abd El-Salam, A., & Ibrahim, N. (2012) 'Estimation of population variance in stratified random sampling using auxiliary information', Journal of Statistical Planning and Inference, 142(6), pp. 1397-1405.

[45] Singh, H. P., & Tailor, R. (2005) 'Estimation of finite population mean using auxiliary information', Applied Mathematics and Computation, 170(2), pp. 1351-1360

[46] El-Morshedy, M., Hussain, S., Ullah, K., Khalil, A., Shabbir, J., & Mansoor, W. (2022). Finite Population Mean Estimation Under Systematic Sampling Scheme in Presence of Maximum and Minimum Values Using Two Auxiliary Variables. Mathematical Problems in Engineering, 2022(1), 2703178.

[47] Junaid, M., Manzoor, S., Hussain, S., Bakr, M.E., Balogun, O.S. & Rasheed, S., 2024. An optimal estimation approach in non-response under simple random sampling utilizing dual auxiliary variable for finite distribution function. Heliyon, 10(19).

[48] Akhtar, S., Hassan, F., Saqlain, S.R., Ali, A. and Hussain, S., 2023. The prevalence of peripheral neuropathy among the patients with diabetes in Pakistan: a systematic review and meta-analysis. Scientific Reports, 13(1), p.11744.

[49] Khan, S., Farooq, M., Sardar Hussain, S.A., Atif, M. and Ilyas, M., 2024. Comprehensive Class Of Estimators For Estimation Of Population Mean Under Stratified Sampling: Application With Real Data Sets And Simulation Analysis.

[50] Ahmad, S., Hussain, S., Yasmeen, U., Aamir, M., Shabbir, J., El-Morshedy, M., Al-Bossly, A. & Ahmad, Z., 2022. A simulation study: Using dual ancillary variable to estimate population mean under stratified random sampling. Plos one, 17(11), p.e0275875.

[51] Ahmad, S., Hussain, S., Ullah, K., Zahid, E., Aamir, M., Shabbir, J., Ahmad, Z., Alshanbari, H.M. and Alajlan, W., 2022. A simulation study: Improved ratio-in-regression type variance estimator based on dual use of auxiliary variable under simple random sampling. PLoS One, 17(11), p.e0276540.

[52] Ahmad, S., Hussain, S., Shabbir, J., Zahid, E., Aamir, M. and Onyango, R., 2022. Improved estimation of finite population variance using dual supplementary information under stratified random sampling. Mathematical Problems in Engineering, 2022(1), p.3813952.

[53] Hussain, S., Akhtar, S. and El-Morshedy, M., 2022. Modified estimators of finite population distribution function based on dual use of auxiliary information under stratified random sampling. Science Progress, 105(3), p.00368504221128486.