EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2025, Vol. 18, Issue 2, Article Number 6019
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

On Use of Predictive Technique for Estimation of Population Mean Using Auxiliary Information: Application on Real Data Sets

Sajid Khan[1], Muhammad Farooq[1], Sohaib Ahmad[2,*], Manahil SidAhmed Mustafa[3], Elsiddig Idriss Mohamed[3], Elfarazdag M. M.Hussein[3], Fatima Ibrahim Abdallah Albadwi[4]

[1] *Department of Statistics, University of Peshawar, Peshawar, Pakistan*
[2] *Department of Statistics Abdul Wali Khan Univerity, Mardan, Pakistan*
[3] *Department of Statistics, Faculty of Science, University of Tabuk, Kingdom of Saudi Arabia*
[4] *Department of Management Information Systems, College of Business and Econometrics, Qassim University, Buraydah, Saudi Arabia*

Abstract. In this article, we used model base method (predictive approach) using auxiliary variables under simple random sampling. Based on the predictive approach the unobserved population along with the observed ones is considered. We found the results for bias, mean squared error both theoretically and numerically. The research analyzes sophisticated through predictive models because it successfully models complex non-linear research variable interactions combined with their auxiliary variables. The current work is evaluated through numerous applications which produce better improved estimation performances. The corresponding optimal strategies of the suggested estimators are explored together with their empirical and graphical analogues with various contemporary estimators of population mean. Some actual data sets are used in an empirical research. By means of a comparison between the mean square error of the proposed estimators with the mean square error of existing estimators under which the suggested class of estimator dominates the existing estimators. The favorable empirical results clearly demonstrate the superiority of the proposed estimators over the existing estimators.

2020 Mathematics Subject Classifications: 62D05

Key Words and Phrases: Predictive approach, estimation of mean, bias, means squared error, efficiency

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v18i2.6019

Email addresses: sajidkhan2022@uop.edu.pk (S. Khan), m.farooq@uop.edu.pk (M. Farooq), sohaib_ahmad@awkum.edu.pk (S. Ahmad), msida@ut.edu.sa (M. S. Mustafa), eidriss@ut.edu.sa (E. I. Mohamed), e.hussain@ut.edu.sa (E. M. M. Hussein), F.albadwi@qu.edu.sa (F. I. A. Albadwi)

 Copyright: © 2025 The Author(s). (CC BY-NC 4.0)

1. Introduction

To achieve more efficient estimators, survey sampling is the technique that contains design based and model based method using auxiliary information. Sample survey can be obtain through prediction theory that is a general frame work for statistical inference referring to a character of a finite population. In many cases, the assessment of the study variable is established for every unit of the population before the sample is selected. After selecting and observing the sample units, we get information about the sampled values only; while non-sampled values remain unknown. This ignorance of y-values (study variable) leads to a need of predicting some function of those values mathematically in order to have an estimator or predictor for the full population. For example, it is typical in many scientific disciplines, such as economics and medicine, for an outcome or research variable to be costly to detect while its covariates are relatively cheap. For example, it is very costly to measure consumption poverty since it involves administering a long questionnaire that people must fill out over a long period of time. However, its variate, such as asset holdings, water and lightning resources are less expensive to observe. When destructive testing is too costly and inaccurate, non-destructive testing might be utilized instead in some industries. This is where prediction estimators as a replacement. It must be noted that we use the term "prediction" in the sense of making statistical guess about the non-sampled (un-observed) y-values, not in the sense of forecasting future values. The reactivity of the customary estimators of the population mean of sampling is really better understood with the help of this approach. The predictive approach has been applied by researchers for examining the existing estimators, or developing new estimators. The authors [1–4] used the auxilary variables for estimation of population parameters under differenct sampling schemes. The readers can find some related findings in predictive approach under simple random sampling include [5] proposed for stratified two phase random sampling a generalised exponential chain ratio estimator. [6] proposed a two-phase sample predictive estimate of finite population mean in case of incomplete data. [7] proposed ranked set sampling-based new predictive estimators. [8] proposed both internal and external validation of predictive models by means of a small sample bias and precision simulation analysis. [9] addressed a developing enhanced predictive estimator for finite population mean with auxiliary information. [10] addressed population mean predictive estimate in ranked set sampling. [11] effective finite population mean estimators derived from predictive estimating in simple random sampling. [12–18] discussed an enhanced estimators based on auxiliary information for predictive approach under different sampling scheme. Authors in [19] proposed a better estimator based on predictive technique under PPS sampling for population mean estimate.

In survey sampling, the method which is designed based rely survey scheme like stratified, cluster and randomization to achieve inferences about the unknown population. These techniques provide unbiased estimators. These methods ensure unbiased estimators and avoid constructing assumptions about the distribution of the population by trusting simply on the sample design for inferential processes. Whenever the designed prepared well for the collection of sample, it is more suitable and having more application. Alternatively

the model based utilizes the helping information and having statistical models for the survey scheme. To boost the efficiency of estimate, we utilize predictive approach whenever the sample size is small. By integrating information from numerous bases, model-based approach provides more precise estimates. In many circumstances when consistent data is accessible, then model-based technique may outclass design-based approach; though, it is important to carefully deliberate model assumptions. For good estimation of population mean, predictive approaches are important for making learned predictions about the target variable using more information. By integrating information from causes other than the sample, predictive models are capable to diminish estimation errors. With predictive approaches, properties are assigned competently and interferences are focused where they are most desirable, leading to better-informed policies and actions. For the assessment of population mean the predictive approach based on simple random sampling is indispensable in many domains. Mostly in situation when the sample numbers are limited, the predictive approach boosts the efficiency of estimators. Enhancements in efficiency and a reduction in sample inconsistency lead to more consistent estimations of population parameters. Using this prediction method, researchers are capable to assign possessions appropriately and make knowledgeable conclusions based on estimated demographic influences. Eventually, the social sciences, public health, and economics stand to help from a well sympathetic of population dynamics, better decision-making, and well-organized use of resources when simple random sampling is supplemented by predictive techniques used to assignment population means. The research gap that this work fills is produced by the comprehensive integration of finite population mean estimate into an integrated structure incorporating auxiliary information considerations and a predictive technique under simple random sampling. While some studies have looked at these components alone or in pairs, very few have integrated them all. The new method that our study introduces combines predictive approach estimate with basic random sampling techniques is a significant contribution to the field. In sophisticated sampling circumstances, this complete background provides a more precise technique of measuring population mean. By examining all these fundamentals collectively, our study not only fills extant holes in the literature but also gives a more thorough and practical solution for real-world survey circumstances where these aspects often co-exist. Several fields, such as healthcare, weather prediction, and finance, make use of this research in their estimation of improved population mean. As a result, healthcare planning, weather forecasting, and financial decision-making are all enhanced. Improving estimating approaches and ensuring higher precision, the study takes into account variables like predictive approach, simple random sampling, and population mean in sample surveys.

Most current survey estimation methods including ratio, regression and difference estimators require strong assumptions of linear relationships and normal data distribution to function. The assumptions fail to maintain validity when used to analyze actual datasets which show characteristics of complexity and non-linear behavior and contain substantial noise levels. Research on predictive approaches for parameter estimation shows few applications regarding population mean estimation with auxiliary variables. Many academic works neglect to perform comparative research on traditional estimators and modern

predictive approaches particularly when they are used with genuine datasets. The available literature lacks adequate guides for actual implementation as well as performance standards when sampling different scenarios. Studies fail to demonstrate how machine learning methods can be properly adapted to survey sampling applications that retain interpretability features together with minimal bias effects. A thorough research of population mean estimation through predictive techniques needs immediate attention using diverse empirical datasets for verification. The development of such research leads to improved estimation techniques that better match modern data collection scenarios.

The continuing of the article is structured as: Section 2, includes the methods and materials regarding the considered study. Section 3, includes some relevant existing adopted estimator based on a predictive approach under simple random sampling. The suggested generalized estimator are given in Section 4. Section 5 include numerical study. Applications of the study are given in section 6. Argumentation of the article is given in Section 7. Lastly, Section 8, contain the concluding remarks of the article.

2. 2. Methodology

Let$K = K_1, K_1, \cdots, K_N$ contains of N diverse units. Suppose y be the study variable and x and z be the auxiliary variables. Thus $(y_j, x_j, z_j), j = 1, 2, \ldots N$ denote the j^{th} observations for the main and auxiliary variables. We are concerned in approximating the population mean: $\bar{Y} = \frac{1}{N}\sum_{j=1}^{N} Y_j$ in the occurrence of the auxiliary variables x and z. Further, let S signify the fixed size from the population K. Let s be a member of S and let n_s signify the effective sample size in s and $\bar{s}$ denote the gathering of units of K which are not included in s.

$$\bar{Y} = \frac{1}{n}\sum_{j\epsilon s} y_j \ \ (sample\ mean)$$

$$\bar{Y}_{\bar{s}} = \frac{1}{N-n}\sum_{j\epsilon \bar{s}} y_j \ \ (mean\ of\ non-population\ values\ of\ the\ study\ variable)$$

For any $s\epsilon S$, we have : $\bar{Y} = \frac{n}{N}\bar{Y}_s + \frac{n}{(N-n)}\bar{Y}_s$, $\bar{y} = \frac{1}{n}\sum_{j\epsilon s} y_j$
That is,$\bar{y} = \bar{Y}_s$
Thus the $\bar{Y}_s$ can be replaced by$\bar{y}$ as:

$$t = \frac{n}{N}\bar{y} + \frac{n}{N-n}T$$

where T is the predictor of the population mean $\bar{Y}_{\bar{s}}$ of unobserved units of the population. We introduce the following symbols and notations:
$\bar{X}_{\bar{S}} = \frac{N\bar{X}-n\bar{x}}{N-n}$, Non-sampled population mean of x.
$\bar{Z}_{\bar{S}} = \frac{N\bar{Z}-n\bar{z}}{N-n}$, Non-sampled population mean of z.
$\bar{Y} = \frac{1}{N}\sum_{j=1}^{N} y_j$, $\bar{X} = \frac{1}{N}\sum_{j=1}^{N} x_j$, and $\bar{Z} = \frac{1}{N}\sum_{j=1}^{N} z_j$, donates the population means.

$\bar{y} = \frac{1}{N}\sum_{j=1}^{N} y_j$, Sample mean of y, $\bar{X} = \frac{1}{N}\sum_{j=1}^{N} x_j$, Sample mean of x, and $\bar{z} = \frac{1}{N}\sum_{j=1}^{N} z_j$, Sample mean of z.

$C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$, $C_z = \frac{S_z}{\bar{Z}}$, the coefficient of variation of variation of y, x and z.

$S_x^2 = \sum_{j=1}^{n} \frac{x_j - \bar{X}}{N-1}$, $S_y^2 = \sum_{j=1}^{n} \frac{y_j - \bar{Y}}{N-1}$, $S_z^2 = \sum_{j=1}^{n} \frac{z_j - \bar{Z}}{N-1}$, denote the population variance of the study variable y.

$S_{xy} = \sum_{j=1}^{N} \frac{(x_j - \bar{X})(y_j - \bar{Y})}{N-1}$, $S_{zy} = \sum_{j=1}^{N} \frac{(z_j - \bar{z})(y_j - \bar{Y})}{N-1}$, $S_{xz} = \sum_{j=1}^{N} \frac{(x_j - \bar{X})(z_j - \bar{Z})}{N-1}$, denote the population covariance ,

where

$$C = \rho_{xy}\frac{C_y}{C_x},\ \lambda = \frac{1-f}{n},\ f = \frac{n}{N}$$

Let

$$e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}, e_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}, e_2 = \frac{\bar{z} - \bar{Z}}{\bar{Z}}$$

Such that $E(e_0) = 0 = E(e_1) = E(e_2)$

$E(e_0^2) = \lambda C_y^2$, $E(e_1^2) = \lambda C_x^2$, $E(e_2^2) = \lambda C_z^2$, $E(e_0 e_1) = \lambda \rho_{xy} C_x C_y$, $E(e_0 e_2) = \lambda \rho_{zy} C_z C_y$, $E(e_1 e_2) = \lambda \rho_{xz} C_x C_z$,

3. Adopted existing estimators

In this section, we have deliberated some existing adopted estimators for estimation of mean using predictive approach under simple random sampling.

(i) [20] proposed the predictive estimators of usual mean, ratio and product estimators for predicting the mean $\bar{Y}_{\bar{s}}$ is given by:

$$t_1 = \frac{n\bar{y}}{N} + (\frac{N-n}{N})\bar{Y}_{\bar{s}},\ where\ \bar{Y}_s = \bar{y} \tag{1}$$

The bias and variance of t_1 are given by:

$$Bias(t_1) = 0 \tag{2}$$

$$Var(t_1) \cong \lambda \bar{Y}^2 C_y^2 \tag{3}$$

$$t_2 = \frac{n\bar{y}}{N} + (\frac{N-n}{N})(\bar{y}\frac{\bar{x}_{\bar{s}}}{\bar{x}}) \tag{4}$$

$$t_3 = \frac{n\bar{y}}{N} + (\frac{N-n}{N})(\bar{y}\frac{\bar{x}_{\bar{s}}}{\bar{y}}) \tag{5}$$

The bias and MSE of t_2 and t_3 are given by:

$$Bias(t_2) \cong \lambda \bar{Y} C_x^2 [1 - C] \tag{6}$$

$$Bias(t_3) \cong \lambda \bar{Y} C_x^2 [C + \frac{f}{1-f}] \tag{7}$$

and

$$MSE(t_2) \cong \lambda \bar{Y}^2 [C_y^2 + C_x^2(1-2C)] \tag{8}$$

$$MSE(t_3) \cong \lambda \bar{Y}^2 [C_y^2 + C_x^2(1+2C)] \tag{9}$$

(ii) [21] recommended the following enhanced estimators, which are given by:

$$t_4 = \frac{n\bar{y}}{N} + \frac{N-n}{N}\bar{y} exp(\frac{\bar{X}_{\bar{s}} - \bar{x}}{\bar{X}_{\bar{s}} + \bar{x}}) \tag{10}$$

$$t_5 = \frac{n\bar{y}}{N} + \frac{N-n}{N}\bar{y} exp(\frac{\bar{x} - \bar{X}_{\bar{s}}}{\bar{X}_{\bar{s}} + \bar{x}}) \tag{11}$$

The bias and MSEs of t_3 and t_4 are given by:

$$Bias(t_4) \cong \frac{\lambda}{8} \bar{Y} C_x^2 [3 - 4(C+f)] \tag{12}$$

$$Bias(t_4) \cong \frac{\lambda}{8} \bar{Y} C_x^2 [3 - 4(C+f)] \tag{13}$$

and

$$MSE(t_4) \cong \lambda \bar{Y}^2 [C_y^2 + \frac{C_x^2}{4}(1-4C)] \tag{14}$$

$$MSE(t_5) \cong \lambda \bar{Y}^2 [C_y^2 + \frac{C_x^2}{4}(1+4C)] \tag{15}$$

(iii) [7]recommended efficient exponential estimators, which are given by:

$$t_6 = k_1 [\frac{n\bar{y}}{N} + (\frac{N-n}{N})\bar{y} exp\left\{\frac{\bar{X}_{\bar{s}} - \bar{x}}{\bar{X}_{\bar{s}} + \bar{x}}\right\}] \tag{16}$$

$$t_7 = k_2 [\frac{n\bar{y}}{N} + (\frac{N-n}{N})\bar{y} exp\left\{\frac{\bar{x} - \bar{X}_{\bar{s}}}{\bar{X}_{\bar{s}} + \bar{x}}\right\}] \tag{17}$$

The bias and mean square errors of t_6 and t_7 are given by:

$$Bias(t_6) \cong \bar{Y}[(k_1 - 1) + k_1 \left\{\frac{3}{8} - \frac{1}{2}f\right\} \lambda C_x^2 - \frac{1}{2}\lambda C C_x^2] \tag{18}$$

$$Bias(t_7) \cong \bar{Y}[(k_2 - 1) + k_2 \left\{\frac{1}{8(1-f)}\right\} \lambda C_x^2 - \frac{1}{2}\lambda C C_x^2] \tag{19}$$

and

$$MSE(t_6) \cong \bar{Y}^2(1 - \frac{A_1^2}{B_1^2}) \tag{20}$$

$$MSE(t_7) \cong \bar{Y}^2(1 - \frac{A_2^2}{B_2^2}) \tag{21}$$

where

$$A_1 = 1 + \frac{1}{8}(3-4f)\lambda C_x^2 - \frac{1}{2}\lambda C C_x^2$$

$$A_2 = 1 - \frac{1}{8(1-f)}\lambda C_x^2 + \frac{1}{2}\lambda C C_x^2$$

$$B_1 = 1 + \lambda C_x^2 - 2\lambda C C_x^2 + \frac{1}{8}(5-4f)\lambda C_x^2$$

$$B_2 = 1 + \lambda C_x^2 - 2\lambda C C_x^2 + \frac{f}{4(1-f)}\lambda C_x^2$$

(iv) The author in [22] recommended improved exponential estimators, which are given by:

$$t_8 = \frac{n\bar{y}}{N} + (\frac{N-n}{N})\bar{y}exp\left\{\frac{\sqrt{\bar{\bar{X}}_{\bar{s}}} - \sqrt{\bar{\bar{x}}}}{\sqrt{\bar{\bar{X}}_{\bar{s}}} + \sqrt{\bar{\bar{x}}}}\right\} \tag{22}$$

$$t_9 = \frac{n\bar{y}}{N} + (\frac{N-n}{N})\bar{y}exp\left\{\frac{\sqrt{\bar{\bar{x}}} - \sqrt{\bar{\bar{X}}_{\bar{s}}}}{\sqrt{\bar{\bar{X}}_{\bar{s}}} + \sqrt{\bar{\bar{x}}}}\right\} \tag{23}$$

The bias and MSE are given by:

$$Bias(t_8) \cong \lambda \bar{Y}^2 C_x^2(\frac{(3-4f)}{(32(1-f))} - \frac{1}{4}C) \tag{24}$$

$$Bias(t_9) \cong \lambda \bar{Y}^2 C_x^2(\frac{(1)}{(4C))} - \frac{1-4f}{32(1-f)}) \tag{25}$$

and

$$MSE(t_8) \cong \lambda \bar{Y}^2(C_y^2 + \frac{C_x^2}{16} - \frac{CC_x^2}{2}) \tag{26}$$

$$MSE(t_8) \cong \lambda \bar{Y}^2(C_y^2 + \frac{C_x^2}{16} + \frac{CC_x^2}{2}) \tag{27}$$

4. The suggested estimator

Research surveys and population investigations require the evaluation of population mean as their core objective. The ratio and regression methods serve as traditional tools to enhance the estimate efficiency when working with auxiliary information. Increasing amounts of complex datasets demand new predictive techniques to detect intricate variables interactions due to rising needs in robust forecasting. Research teams now seek to find contemporary predictive modeling techniques which boost the accuracy and reliability of population mean estimation. We suggest a new enhanced generalized estimator for population mean using two auxiliary variables under simple random sampling using predictive approach. This work is considered unique as, to the best of our knowledge, no one has before explored the predictive method of estimating the finite population mean. Such techniques provide improved accuracy and enhanced robustness for prediction through their ability to handle complicated or dimensionally complex auxiliary information. The suggested estimator is given in equation (28):

$$t_p = \frac{n\bar{y}}{N} + \left(\frac{N-n}{N}\right)\left[\bar{y} + k_1\left(\bar{X}_{\bar{s}} - \bar{x}\right) + k_2\left(\bar{Z}_{\bar{z}} - \bar{z}\right)\right] exp\left\{\frac{\gamma(\bar{X}_{\bar{s}} - \bar{x})}{\bar{X}_{\bar{s}+\bar{x}}}\right\} \tag{28}$$

where k_1, and k_2 are unknown constants.

MSE of the above estimator when γ may take the values 1 and -1. We get (29):

$$t_p = \frac{n\bar{y}}{N} + (\frac{N-n}{N})\bar{y}exp\left\{\frac{\gamma(\bar{X}_{\bar{s}} - \bar{x})}{\bar{X}_{\bar{s}+\bar{x}}}\right\} + (\frac{N-n}{N})k_1\left(\bar{X}_{\bar{s}} - \bar{x}\right) exp\left\{\frac{\gamma(\bar{X}_{\bar{s}} - \bar{x})}{\bar{X}_{\bar{s}+\bar{x}}}\right\}$$
$$+(\frac{N-n}{N})k_2\left(\bar{Z}_{\bar{z}} - \bar{z}\right) exp\left\{\frac{\gamma(\bar{X}_{\bar{s}} - \bar{x})}{\bar{X}_{\bar{s}+\bar{x}}}\right\} \tag{29}$$

where

$$\bar{X}_{\bar{S}} = \frac{N\bar{X} - n\bar{x}}{N-n}$$

Simplifying the right-hand side of Equation (29), multiplying and expressing the above equation in terms of e's as follows:

$$t_p = \bar{Y}(1+e_0)\left[(f + (1-f)exp\left\{-\frac{\gamma e_1}{2(1-f)}(1 + \frac{1-2f}{2(1-f)})^{-1}\right\}\right]$$
$$-\frac{k_1(1-f)\bar{X}e_1}{1-f}exp\left\{-\frac{\gamma e_1}{2(1-f)}(1 + \frac{1-2f}{2(1-f)})^{-1}\right\}$$
$$-\frac{k_2(1-f)\bar{Z}e_2}{1-f}exp\left\{-\frac{\gamma e_1}{2(1-f)}(1 + \frac{1-2f}{2(1-f)})^{-1}\right\}$$

After simplifying the above equation, we got equation (30):

$$t_p = \bar{Y}(1+e_0)(f + 1 - f) + (1-f)\left\{1 - \frac{\gamma e_1}{2(1-f)} + \frac{\gamma^2 + 2\gamma - 4\gamma f}{8(1-f)^2}\right\}e_1^2$$

$$-\frac{k_1(1-f)\bar{X}e_1}{1-f}\left\{1-\frac{\gamma e_1}{2(1-f)}\right\}-\frac{k_2(1-f)\bar{Z}e_2}{1-f}\left\{1-\frac{\gamma e_1}{2(1-f)}\right\} \tag{30}$$

Expanding and approximating terms of (30) up to the first degree, we got equation (31):

$$t_p-\bar{Y}\cong\bar{Y}e_0-\frac{1}{2}\bar{Y}\gamma e_1-\bar{X}k_1e_1-\bar{Z}k_2e_2-\frac{1}{2}\bar{Y}\gamma e_0e_1+\frac{\bar{Y}(\gamma^2+2\gamma-4\gamma f)e_1^2}{8(1-f)}$$
$$+\frac{\bar{X}k_1\gamma e_1^2}{2(1-f)}+\frac{\bar{Z}k_1\gamma e_1e_2}{2(1-f)} \tag{31}$$

Taking expectations on both sides equation (31), we develop the bias which is given in equation (32):

$$Bias(t_p)\cong\bar{Y}E(e_0)-\frac{1}{2}\bar{Y}\gamma E(e_1)-\bar{X}k_1E(e_1)-\frac{1}{2}\bar{Y}\gamma E(e_0e_1)$$
$$+\frac{\bar{Y}(\gamma^2+2\gamma-4\gamma f)}{8(1-f)}E(e_1^2)+\frac{\bar{X}k_1\gamma}{2(1-f)}E(e_1^2)+\frac{\bar{Z}k_1\gamma}{2(1-f)}E(e_1e_2)$$

$$Bias(t_p)\cong-\frac{1}{2}\bar{Y}\gamma\rho_{xy}C_xC_y+\frac{\bar{Y}(\gamma^2+2\gamma-4\gamma f)}{8(1-f)}\lambda C_x^2+\frac{\bar{X}k_1\gamma}{2(1-f)}\lambda C_x^2+\frac{\bar{Z}k_1\gamma}{2(1-f)}\gamma\lambda\rho_{xz}C_xC_z \tag{32}$$

Expanding square and taking expectations of equation (31), we got equation (33):

$$\mathrm{MSE}(t)\cong\bar{Y}^2E(e_0^2)-\frac{2\bar{Y}^2\gamma E(e_0e_1)}{2}-2\bar{X}\bar{Y}k_1E(e_0e_1)-2\bar{Z}\bar{Y}k_2E(e_0e_1)+\frac{\bar{Y}^2\gamma^2E(e_1^2)}{4}$$
$$+\frac{2\bar{X}\bar{Y}k_1\gamma E(e_1^2)}{2}+\bar{X}^2k_1^2E(e_1^2)+\bar{Z}k_2^2E(e_2^2)+\frac{2\bar{Z}\bar{Y}k_2\gamma E(e_1e_2)}{2}+2\bar{X}\bar{Z}k_1k_2E(e_1e_2) \tag{33}$$

Substituting the values of expectations in equation (33), we got equation (34):

$$\mathrm{MSE}(t)\cong\bar{Y}^2\lambda C_y^2-\bar{Y}^2\gamma\lambda\rho_{xy}C_xC_y-2\bar{X}\bar{Y}k_1\lambda\rho_{xy}C_xC_y-2\bar{Z}\bar{Y}k_2\lambda\rho_{zy}C_zC_y+\frac{\bar{Y}^2\gamma^2\lambda C_x^2}{4}$$
$$+\bar{X}\bar{Y}k_1\gamma\lambda C_x^2+\bar{X}^2k_1^2\lambda C_x^2+\bar{Z}k_2^2\lambda C_z^2+\bar{Z}\bar{Y}k_2\gamma\lambda\rho_{yz}C_yC_z+2\bar{X}\bar{Z}k_1k_2\lambda\rho_{xz}C_xC_z \tag{34}$$

To find values of k_1 and k_2 , we partially differentiate equation (34) with respect to sk_1 and $k_2$and set to zero i.e

$$\frac{\partial MSE(t_P)}{\partial k_1}=0$$
$$\frac{\partial MSE(t_p)}{\partial k_2}=0$$

We get following two normal equations.

$$-2\bar{X}\bar{Y}\lambda\rho_{xy}C_xC_y + \bar{X}\bar{Y}\gamma\lambda C_x^2 + 2\bar{X}^2k_1\lambda C_x^2 + 2\bar{X}\bar{Z}k_2\lambda\rho_{xz}C_xC_z = 0$$
$$-2\bar{Z}\bar{Y}\lambda\rho_{zy}C_zC_y + 2\bar{Z}k_2\lambda C_z^2 + \bar{Z}\bar{Y}\gamma\lambda\rho_{yz}C_yC_z + 2\bar{X}\bar{Z}k_2\lambda\rho_{xz}C_xC_z = 0$$

Solving these equations, we get the values of k_1 and k_2, i.e.

$$k_{1opt} = -\frac{\bar{Y}(\gamma C_x\rho_{xz}^2 - 2C_y\rho_{xz}\rho_{yz} - \gamma C_x + 2C_y\rho_{xy})}{2XC_x(\rho_{xz}^2 - 1)}$$

and

$$k_{2opt} = \frac{\bar{Y}C_y(\rho_{xy}\rho_{xz} - \rho_{yz})}{\bar{Z}C_z(\rho_{xz}^2 - 1)}$$

Putting values in equation (34), we get bias and MSE of the proposed estimator, which are given by:

$$Bias(t_p)_{opt} \cong \frac{\gamma\lambda C_x\bar{Y}(4f^2C_x - f\gamma C_x + 4fC_y\rho_{xy} - 6fC_x + 2C_x)}{8(1-f)} \tag{35}$$

and

$$MSE(t_p)_{opt} \cong \bar{Y}^2\lambda C_y^2 - \bar{Y}^2\gamma\lambda\rho_{xy}C_xC_y + \frac{1}{4}\bar{Y}^2\gamma^2\lambda C_x^2$$
$$+\frac{\bar{Y}^2(\gamma C_x\rho_{xz}^2 - 2C_y\rho_{xz}\rho_{yz} - \gamma C_x + 2C_y\rho_{xy})\lambda\rho_{xy}C_y}{(\rho_{xz}^2 - 1)}$$
$$-\frac{2\bar{Y}^2\lambda C_y^2\rho_{yz}(\rho_{xy}\rho_{xz} - \rho_{yz})}{(\rho_{xz}^2 - 1)} + \frac{\bar{Y}^2\lambda C_y^2(\rho_{xy}\rho_{xz} - \rho_{yz})}{(\rho_{xz}^2 - 1)}$$
$$-\frac{\bar{Y}^2\gamma\lambda C_x(\gamma C_x\rho_{xz}^2 - 2C_y\rho_{xz}\rho_{yz} - \gamma C_x + 2C_y\rho_{xy})}{2(\rho_{xz}^2 - 1)}$$
$$+\frac{\bar{Y}\lambda(\gamma C_x\rho_{xz}^2 - 2C_y\rho_{xz}\rho_{yz} - \gamma C_x + 2C_y\rho_{xy})^2}{4(\rho_{xz}^2 - 1)}$$
$$+\frac{\bar{Y}^2\gamma\lambda C_xC_y\rho_{xz}(\rho_{xy}\rho_{xz} - \rho_{yz})}{(\rho_{xz}^2 - 1)}$$
$$+\frac{\bar{Y}^2\lambda(\gamma C_x\rho_{xz}^2 - 2C_y\rho_{xz}\rho_{yz} - \gamma C_x + 2C_y\rho_{xy})}{4(\rho_{xz}^2 - 1)}$$
$$-\frac{\bar{Y}^2\lambda C_y\rho_{xz}(\gamma C_x\rho_{xz}^2 - 2C_y\rho_{xz}\rho_{yz} - \gamma C_x + 2C_y\rho_{xy})(\rho_{xy}\rho_{xz} - \rho_{yz})}{(\rho_{xz}^2 - 1)}$$

After simplification the above equation we got the minimum MSE of t_p, which is given in equation (36):

$$MSE(t_p)_{opt} \cong \frac{\bar{Y}^2\lambda C_y^2(1-\rho_{xy}^2-\rho_{xz}^2-\rho_{yz}^2+2\rho_{xy}\rho_{xz}\rho_{yz})}{(1-\rho_{xy}^2)} \tag{36}$$

For different values of k_1, k_2, and γ in equation (28), some members of the family of the proposed estimator can be obtained, which is given in Table 1.

Table 1: Some members of the generalized family of regression cum exponential type estimators.

Estimators	γ	k_1	k_2
$t_{p1} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\bar{Y}_{\bar{s}} = \bar{y}$, *sample mean*	0	0	0
$t_{p2} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + b(\bar{X}_{\bar{s}} - \bar{x})\right\}$, *Srivastava* [8]	0	B	0
$t_{p3} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\bar{y}exp(\frac{\bar{X}_{\bar{s}}-\bar{x}}{\bar{X}_{\bar{s}}+\bar{x}})$, *Singh et al.* [6]	1	0	0
$t_{p4} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\bar{y}exp(\frac{\bar{x}-\bar{X}_{\bar{s}}}{\bar{X}_{\bar{s}}+\bar{x}})$, *Singh et al.* [6]	-1	0	0
$t_{p5} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + b(\bar{X}_{\bar{s}} - \bar{x})\right\}exp(\frac{\bar{X}_{\bar{s}}-\bar{x}}{\bar{X}_{\bar{s}}+\bar{x}})$	1	B	0
$t_{p6} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + b(\bar{X}_{\bar{s}} - \bar{x})\right\}exp(\frac{\bar{x}-\bar{X}_{\bar{s}}}{\bar{X}_{\bar{s}}+\bar{x}})$	-1	b	0
$t_{p7} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + k_1(\bar{X}_{\bar{s}} - \bar{x})\right\} + k_2(\bar{Z}_{\bar{s}} - \bar{z})$	0	k_1	k_2
$t_{p8} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + b(\bar{Z}_{\bar{s}} - \bar{z})\right\}exp(\frac{\bar{X}_{\bar{s}}-\bar{x}}{\bar{X}_{\bar{s}}+\bar{x}})$	1	0	b
$t_{p9} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + b(\bar{Z}_{\bar{s}} - \bar{z})\right\}exp(\frac{\bar{x}-\bar{X}_{\bar{s}}}{\bar{X}_{\bar{s}}+\bar{x}})$	-1	0	0

Case 1: When $\gamma = 1$ in equation (28) the proposed estimator becomes the predictive regression-cum-exponential ratio type estimator, which is given in equation (37):

$$t_{PR} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + k_1(\bar{X}_{\bar{s}} - \bar{x}) + k_2(\bar{Z}_{\bar{s}} - \bar{z})exp(\frac{\bar{X}_{\bar{s}} - \bar{x}}{\bar{X}_{\bar{s}} + \bar{x}})\right\} \tag{37}$$

Optimum bias and MSE of equation (37), after simplifying we got:

$$Bias(t_{PR})_{opt} \cong \frac{\theta C_x^2\bar{Y}(4f^2+4fC-7f+1)}{8(1-f)} \tag{38}$$

and

$$MSE(t_{PR})_{opt} \cong \frac{\bar{Y}^2C_y^2(1-\rho_{xy}^2-\rho_{xz}^2-\rho_{yz}^2+2\rho_{xy}\rho_{xz}\rho_{yz})}{1-\rho_{xz}^2} \tag{39}$$

Case 2: For $\gamma = -1$, the proposed estimator becomes the predictive regression-cum-exponential product type estimator, which is given in equation (40):

$$t_{PR} = \frac{n\bar{y}}{N} + \frac{N-n}{N}\left\{\bar{y} + k_1(\bar{X}_{\bar{s}} - \bar{x}) + k_2(\bar{Z}_{\bar{s}} - \bar{z})exp(\frac{\bar{x} - \bar{X}_{\bar{s}}}{\bar{X}_{\bar{s}} + \bar{x}})\right\} \tag{40}$$

The bias and MSE of the proposed predictive regression-cum-exponential product type estimator are given by:

$$Bias(t_{PP})_{opt} \cong \frac{\theta C_x^2\bar{Y}(4f^2+4fC-5f+3)}{8(1-f)} \tag{41}$$

and

$$MSE(t_{PP})_{opt} \cong \frac{\bar{Y}^2 C_y^2(1-\rho_{xy}^2-\rho_{xz}^2-\rho_{yz}^2+2\rho_{xy}\rho_{xz}\rho_{yz})}{1-\rho_{xz}^2} \tag{42}$$

5. 1. Numerical study

This section contains comparisons of estimators using some actual data sets. To check the relative performances of estimators we compute PRE. The PRE is expressed as:

$$PRE(t_l) = Var(\bar{y})/MSE(t_u)100, \quad u = [1.2, \cdots, p]$$

To check the bias of different estimators, we calculated the following quantities:

$$Q_(t_2) = |\frac{Bias(t_2)}{\lambda \bar{Y} C_x^2}|,\ Q_(t_3) = |\frac{Bias(t_3)}{\lambda \bar{Y} C_x^2}|,\ Q_(t_4) = |\frac{Bias(t_4)}{\lambda \bar{Y} C_x^2}|,\ Q_(t_5) = |\frac{Bias(t_5)}{\lambda \bar{Y} C_x^2}|$$

$$Q_(t_6) = |\frac{Bias(t_6)}{\lambda \bar{Y} C_x^2}|,\ Q_(t_7) = |\frac{Bias(t_7)}{\lambda \bar{Y} C_x^2}|, Q_(t_8) = |\frac{Bias(t_8)}{\lambda \bar{Y} C_x^2}|,\ Q_(t_9) = |\frac{Bias(t_9)}{\lambda \bar{Y} C_x^2}|$$

$$Q_(t_{10}) = |\frac{Bias(t_{10})}{\lambda \bar{Y} C_x^2}|$$

As the term $(\lambda \bar{Y} C_x^2)$ is common in bias terms of all estimators, so we divide the each bias with this common quantity to simplify the calculations. Values of quantities are found numerically and findings are shown in the following Table 2.

6. Application of the study

Applications of the suggested estimator are validated using some real data sets, which is given below:

Data 1: [Source:[2]]

Y = Number of kids, X = number of polio cases, Z = Number of polio cases.

$N = 34, n = 15, \bar{Y} = 4.92, \bar{X} = 2.59, \bar{Z} = 2.91, \rho_{xy} = 0.7326, \rho_{yz} = 0.643, \rho_{xz} = 0.6837, C_x = 1.23187, C_y = 1.01232, C_z = 1.053516$

Data 2: [Source: [3]]

Y = Number of fish caught in 1995, X = Number of fish caught in 1994, Z = Number of fish caught in 1993.

$N = 69, n = 15, \bar{Y} = 4514.899, \bar{X} = 4954.435, \bar{Z} = 4591.072, \rho_{xy=0.9601}, \rho_{yz} = 0.9564, \rho_{xz} = 0.9729, C_x = 1.4247, C_y = 1.3508, C_z = 1.3755.$

Data 3: [Source: [4]]

Y = Tomato supply in tons of Pakistan for in 2003,X = Tomato supply in tons of Pakistan in 2002, Z = Tomato supply in tons of Pakistan in 2001.

$N = 97, n = 30, \bar{Y} = 3135.6168, \bar{X} = 3050.2784, \bar{Z} = 2743.9587, \rho_{xy} = 0.9872, \rho_{yz} = 0.8501, \rho_{xz} = 0.6122, C_x^2 = 5.4812, C_y^2 = 4.8674, C_z^2 = 6.2422$

Table 2: Biases of all considered estimators based on real data sets

Quantities	Data1	Data2	Data3
Q_{t2}	0.3979	0.0034	0.2393
Q_{t3}	0.1466	0.2683	0.1496
Q_{t4}	1.6031	3.6445	2.7324
Q_{t5}	2.7046	0.0211	1.2113
Q_{t6}	0.2471	0.1743	0.56
Q_{t7}	0.1583	0.2795	0.2133
Q_{t8}	0.2319	0.4173	0.2993
Q_{t9}	0.1553	0.8142	0.1335
Q_{tp}	0.5894	0.5356	0.4864
Q_{tPR}	0.0553	0.0814	0.0335
Q_{tPP}	0.5894	0.5356	0.4864

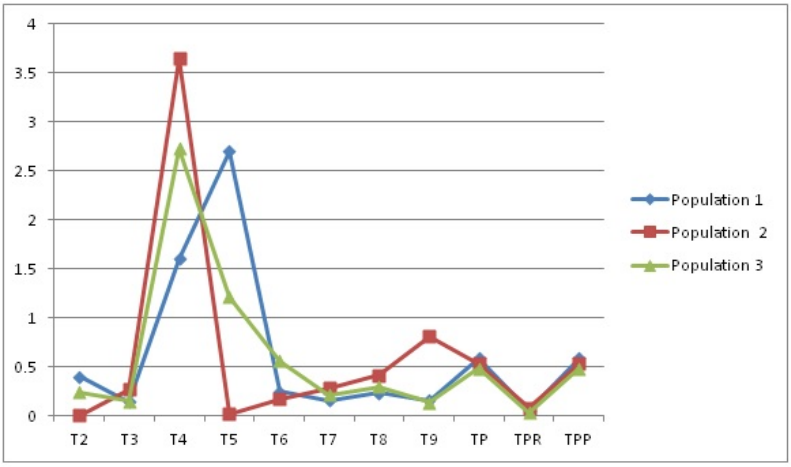

Figure 1: Biases of all considered estimators based on real data sets

7. Discussion

Table 1, included some members of our generalized class of estimators. Table 2, consists the results of bias for all the considered estimators in the article. The results of mean squared error for all the adopted and recommended estimators are given in Table 3. Table 4, include the result of percentage relative efficiency. We observed from Table 2 the following points: The proposed regression-cum exponential ratio-type estimator t_{PR} is less biased than all considered estimators. For population 2, suggested predictive estimator is less biased than other estimator. Also for population 3, the suggested estimator is less biased than all estimators. For all three populations the proposed regression-cum exponential product-type estimator t_{PP} is less biased. The result provide by $t_3, t_5 and t_9$ for all the three population is maximum mean squared error. Also we observed from Table 4, that estimators $t_3, t_5 and t_9$ provide less efficient result as compared to all considered es-

Table 3: MSEs of all considered estimators using real data sets

Estimators	Data 1	Data 2	Data 3
t_1	44.39061	47526246	34042153
t_2	39.34197	3993781	14057273
t_3	219.2408	193161470	130697896
t_4	20.6411	12997169	14465855
t_5	110.5905	107581013	72786167
t_6	22.34624	190873849	7112580
t_7	23.48469	17226857	18578610
t_8	27.20956	27070996	21858040
t_9	72.18426	74362919	51018195
t_p	18.88242	3274143	4963220

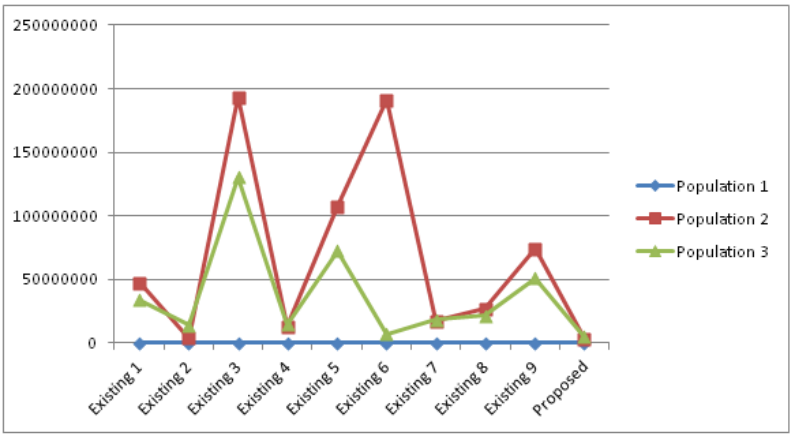

Figure 2: MSEs of all considered estimators using real data sets

timators. The MSEs for both t_{PR} and t_{PP} are the same, so we consider of t_p comparing its MSE with all other considered estimator in Table 3. The result of bias for all estimators are visualize, and are given in Figure 1. The mean squared error and percentage relative efficiency of estimators are also shown in Figures 2 and Figure 3. It is detected from Table 3 that the mean squared error of the proposed estimator is minimum than all considered adopted estimators in all populations. It is detected from Table 4 that the suggested predictive regression-cum exponential estimator performs improved in terms of PRE than all considered estimators. The research value of this study rests in both its programmable applications together with advancements in methodology. Real datasets benefit from predictive techniques as the research creates an alignment between theory development and real-world implementation. Modern predictive tools allow the proper use of auxiliary data to enhance population mean estimation processes. The technique delivers essential consequences for disciplines including public health and agriculture and economics together with the social sciences since their policy implementation usually depends on accurate population estimations. The research adds important knowledge to current academic support

Table 4: PREs of all considered estimators using real data sets

Estimators	Data 1	Data 2	Data 3
t_1	100	100	100
t_2	112.8327	1190.006	242.1676
t_3	20.24743	24.60441	26.0464
t_4	215.0593	365.6661	235.3276
t_5	40.13962	44.17717	46.7701
t_6	198.6491	24.8993	478.6189
t_7	189.0194	275.8846	183.2334
t_8	163.1434	175.5615	155.742
t_9	61.49625	63.91122	66.7251
t_p	235.0896	1451.563	685.8881

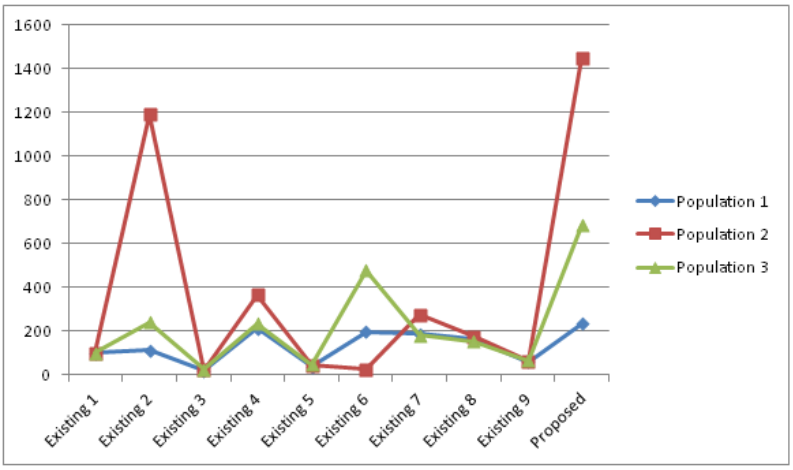

Figure 3: PREs of all considered estimators using real data sets

which promotes combining data science technologies with traditional statistical practices. The research promotes adoption of data-oriented and adaptable procedures for survey analysis. Practitioners and policymakers can learn from these findings to improve their use of existing auxiliary information during situations that require either expensive or time-consuming or difficult data collection processes. This research contributes evidence to the creation of better yet efficient and precise and extendable estimation methods for current data setups.

8. Conclusion

This articles aims to estimate population mean using predictive approach under simple random sampling. The terminologies of the bias and MSE are computed to first order. We found the bias and MSE of the suggested and existing counterparts both theoretically and numerically. To check the efficiency of estimators we utilized three actual data sets. Based

on the numerically result, it is shown that the suggested estimators has minimum MSE and higher PRE. By applying the suggested estimator to future studies that attempt to estimate the means of finite populations will yield drastically better results. Using auxiliary variables and predictive modeling approaches, this estimator improves the efficiency and accuracy of estimates. Reducing sampling variability and improving precision is especially achieved when there is a high correlation among the study variable and auxiliary variables. The current work can be easily extended to estimation population mean using predictive approach under stratified random sampling. Predictive methods should add the combination of linear regression and ratio estimators through non-parametric and ensemble approaches to include random forests along with gradient boosting and neural networks. The models flourish at detecting complicated patterns between study variables and auxiliary variables that deviate from linear structures. The evaluation of performance metrics needs to be established under various sampling approaches (including stratified and systematic together with adaptive methods) to widen practical use for complicated survey methodologies. The technique needs to be analyzed for real-time estimation applying streaming data with dynamic populations to properly serve environments that experience rapid changes such as health surveillance and economic monitoring systems. Testing predictive estimators across different population structures through simulation studies and selecting targets in agriculture and demographic and health sectors to conduct authentic validations will promote their practical implementation between sectors.

Data availability The datasets generated and/or analyzed during the current study are available in the current study are available from the corresponding author on reasonable request.

Conflict of interest

The authors declare no conflict of interest

References

[1] MFA. Crops area production. *The Journal of Agricultural Science*, 2004. Islamabad, Pakistan: Ministry of Food and Agriculture.

[2] Shashi Bahl and R. K. Tuteja. Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12(1):159–164, 1991.

[3] Sarjinder Singh. *Advanced sampling theory with applications: how Michael "selected" Amy*. Springer, 2003. 2 Vols.

[4] W. G. Cochran. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*, 30(2):262–275, 1940.

[5] Aamir Sanaullah, Humera Ameer Ali, Muhammad Noor ul Amin, and Muhammad Hanif. Generalized exponential chain ratio estimators under stratified two-phase random sampling. *Applied Mathematics and Computation*, 226:541–547, 2014.

[6] Lovleen Grover and Anchal Sharma. Predictive estimation of finite population mean in case of missing data under two-phase sampling. *Journal of Statistical Theory and Applications*, 22, 2023.

[7] Shashi Bhushan and Anoop Kumar. Novel predictive estimators using ranked set sampling. *Concurrency and Computation: Practice and Experience*, 35(3):e7435, 2023.

[8] Ewout W. Steyerberg, Sacha E. Bleeker, Henriëtte A. Moll, Diederick E. Grobbee, and Karel G. M. Moons. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, 56(5):441–447, 2003.

[9] Subhash Kumar Yadav and S. S. Mishra. Developing improved predictive estimator for finite population mean using auxiliary information. *Statistika: Statistics and Economy Journal*, 95:76–84, 2015.

[10] Shakeel Ahmed, Javid Shabbir, and Sat Gupta. Predictive estimation of population mean in ranked set sampling. *REVSTAT-Statistical Journal*, 17(4):551–562, 2019.

[11] Rajesh Singh, Hemant Verma, and Prayas Sharma. Efficient estimators of finite population mean using predictive estimation in simple random sampling. *Journal of Statistics*, 23:1–11, 2016.

[12] Shashi Bhushan, Anoop Kumar, Md Tanwir Akhtar, and Showkat Ahmad Lone. Logarithmic type predictive estimators under simple random sampling. *AIMS Mathematics*, 7(7):11992–12010, 2022.

[13] Sohaib Ahmad, Javid Shabbir, Erum Zahid, and Muhammad Aamir. Improved family of estimators for the population mean using supplementary variables under PPS sampling. *Science Progress*, 106(2):00368504231180085, 2023. PMID: 37341780.

[14] Sohaib Ahmad, Sardar Hussain, Aned Al Mutairi, Mustafa Kamal, Masood Ur Rehman, and Manahil SidAhmed Mustafa. Improved estimation of population distribution function using twofold auxiliary information under simple random sampling. *Heliyon*, 10(2):e24115, 2024.

[15] M. K. Pandey, G. N. Singh, Tolga Zaman, Aned Al Mutairi, and Manahil SidAhmed Mustafa. Improved estimation of population variance in stratified successive sampling using calibrated weights under non-response. *Heliyon*, 10(6):e27738, 2024.

[16] Ashok K. Jaiswal, M. K. Pandey, and G. N. Singh. Optimizing population mean estimation using regression and factor type estimators in the presence of non-response. *Franklin Open*, 7:100096, 2024.

[17] Anoop Kumar and Asra Sayyed Siddiqui. Enhanced estimation of population mean using simple random sampling. *Research in Statistics*, 2(1):2335949, 2024.

[18] Abdullah Ali H. Ahmadini, Rajesh Singh, Yashpal Singh Raghav, and Anamika Kumari. Estimation of population mean using ranked set sampling in the presence of measurement errors. *Kuwait Journal of Science*, 51(3):100236, 2024.

[19] Sajid Khan, Muhammad Farooq, Sohaib Ahmad, and Sardar Hussain. Improved estimator for estimation of population mean using predictive approach under PPS sampling. *VFAST Transactions on Mathematics*, 12, 2024.

[20] S. K. Srivastava. Predictive estimation of finite population mean using product estimator. *Metrika*, 30(1):93–99, 1983.

[21] Housila P. Singh, Ramkrishna Solanki, and Alok Singh. Predictive estimation of finite population mean using exponential estimators. *Statistika: Statistics and Economy Journal*, 94:41–53, 2014.

[22] Viplav Kumar Singh and Rajesh Singh. Predictive estimation of finite population mean using generalised family of estimators. *Istatistik: Journal of the Turkish Statistical Association*, 7(2):43–54, 2014.