AIMS Mathematics

AIMS Mathematics, 11(7): 19983–20006.
DOI: 10.3934/math.2026811
Received: 07 May 2026
Revised: 11 June 2026
Accepted: 16 June 2026
Published: 08 July 2026

https://www.aimspress.com/journal/Math

Research article

Optimized generalized estimator for valuation of population mean under probability proportional to size sampling using auxiliary variables

Abdulaziz S. Alghamdi[1] **and Sohaib Ahmad**[2,*]

[1] Department of Mathematics, College of Science &Arts, King Abdulaziz University, P.O. Box 344, Rabigh 21911, Saudi Arabia
[2] Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan

* **Correspondence:** Email: sohaib_ahmad@awkum.edu.pk.

Abstract: The key issue in survey sampling is the efficient estimation of population mean under probability proportional to size sampling (PPS), especially when population units vary considerably in size. In this article, an optimization of generalized estimator of population mean under PPS sampling was proposed by including two auxiliary variables. The proposed class of estimators was designed in such a way that it was flexible in using known population parameters of auxiliary variables, thereby enhancing the efficiency of the estimators. The bias and the mean squared error (MSE) were derived up to the first-order approximation for all the considered estimators. Theoretical comparisons showed that in realistic conditions, the proposed estimator performed better than existing estimators. To validate the theoretical results, an empirical study using real and simulated data was carried out, demonstrating significant improvements in efficiency. The numerical findings showed that the incorporation of auxiliary information in the PPS framework can greatly increase the precision of population mean estimation. The proposed methodology offers a useful contribution to the field of survey sampling and can be effectively implemented, when the population size varies from one individual to another.

Keywords: PPS sampling; auxiliary variables; population mean, bias, mean squared error; efficiency
Mathematics Subject Classification: 62D05, 62L05, 62L10, 62L12, 62F12, 62F25, 62A86, 62G07

1. Introduction

The estimation of population parameters is a common goal in survey sampling, where the population mean is often a primary measure of interest in the process of decision-making in various disciplines, including economics, agriculture, public health, and social sciences. In many practical applications, it is not only impossible but also excessively costly to fully enumerate a population, which means that sample surveys are an indispensable component of statistical inference. The efficiency of estimators derived from such surveys largely depends on the sampling design and the incorporation of auxiliary information. One of the unequal probability sampling methods, PPS sampling, has been of significant interest due to its practical applicability in cases where sampling units differ significantly in size. PPS sampling is characterized by a known measure of size, the probability of which is proportional to the measure of size. This method is especially appropriate in large-scale surveys, where units vary in size, and such differences may be ignored when making estimates, which can be inefficient.

When auxiliary information is used appropriately, it can significantly increase the accuracy of estimators. Ratio, product, and regression estimators are well-known examples that exploit the correlation between the study variable and auxiliary variables. However, over the years, many extensions and generalizations of these estimators have been proposed to suit sampling schemes and data structures. In PPS sampling, the size measure often becomes an auxiliary variable or is closely related to it. The advances in survey sampling have centered on the construction of generalized classes of estimators, which contain several estimators as special cases. These classes are flexible and can adopt to population properties and be optimized using criteria, usually the minimization of MSE. This flexibility is further enhanced by the use of several auxiliary variables, which enable the estimator to capture more complex relationships in the data.

El-Saeed et al. [1] proposed a population mean estimator under PPS sampling. The authors focused on enhancing resistance to outliers and showed that their estimator is more robust to outliers than classical estimators in skewed populations. The study, however was limited by its application to a specific field (radiation data), which may limit generalizability. To enhance efficiency, Alghamdi et al. [2] designed a new estimator that uses auxiliary information under the PPS sampling. Their method exhibits a greater accuracy in comparison to the conventional estimators. Nevertheless, the approach is based on existing estimator designs with variations, so the theoretical part is not particular novel. Azeem et al. [3] presented a better estimator of the population mean in PPS sampling, which was again confirmed using radiation data sets. The research stresses efficiency improvements with perfected weighting plans. One weakness is the lack of broader studies of simulations, which undermines arguments about universal applicability. Singh et al. [4] introduced a better estimator of the total population in PPS sampling. The work offers good theoretical support and derivations. But it is somewhat outdated, as it entails mainly totals rather than means and does not include modern methods of auxiliary-based optimization. Ghorbal [5] stressed the importance of sound estimation methods in the PPS sampling, especially in engineering and radiation data. This enhances the literature by addressing the sensitivity of outliers, but as with other literature, it is not theoretically innovative, and it uses applied case studies to support the research. The study by Hussein et al. [6] became the first to use combination of strong regression and PPS sampling to estimate the population mean. It is a strength, as it is a hybrid that will be more resistant to influential observations. The cost of this increased complexity in the model and the assumptions can restrict practical application.

Shah et al. [7] suggested a new class of estimators of the population distribution function (CDF) in PPS sampling. This also adds theoretical development and computational analysis, thereby making it more comprehensive than mean-based studies. Nevertheless, the work is computationally intensive and might not be as easily accessible to practitioners. Ahmad et al. [8]. emphasized the significance of auxiliary information in enhancing efficiency. However, the estimator family is cumulative, based on preexisting forms as opposed to new concepts. Azeem [9] presented an optimal method for randomized response in PPS sampling, for sensitive survey data. It expands the uses of PPS to more than the estimation of means. Nevertheless, it is more specialized, which restricts direct applicability to estimating parameters of general populations. Mustafa et al. [10] suggested new estimation algorithms using auxiliary information, which have been tested on real and simulated data. One of the strengths is the incorporation of simulation. However, the research is not well-theoretically justified, and the originality of the studies is mainly in adjusting the parameters.

Wang et al. [11] discussed estimation in the context of PPS sampling of data with extreme values using double sampling methods. This plays a crucial role in handling non-normal and contaminated data. Nevertheless, the double sampling design also adds to the cost and complexity of operations. Ahmad et al. [12], developed an improved estimator based on two auxiliary variables, which is more efficient when used in PPS sampling. The application of many auxiliary variables is a strength, which introduces dependency on the availability and accuracy of auxiliary data, which can be a practical limitation. Khan et al. [13] suggested an estimator of predictive approach under PPS sampling. It incorporates the prediction theory to improve the accuracy of estimations. Nevertheless, the assumptions of the predictive model are not always applicable, which reduces its performance on real-world data.

Researchers have developed ideas to improve the reliability analysis of structures by combining machine learning and advanced optimization methods. An improved soft Monte Carlo simulation has been used along with Support Vector Regression (SVR) to enhance the efficiency and accuracy of reliability estimation, as done by Yang et al. [14]. In another study, Yang et al. [15] proposed the framework of ASVR-GPSO, which combines the active support vector regression and global-best partial swarm optimization to obtain more accurate and efficient reliability assessments. Moreover, Yang et al. [16] presented an adaptive Kriging-assisted method with an improved sparrow search algorithm and first-order reliability method in an augmented Lagrangian framework, which is a powerful and efficient solution to complex nonlinear reliability problems. The studies show, as a whole, that surrogate modeling and intelligent optimization techniques are increasingly becoming of paramount importance in enhancing the accuracy and computational efficiency of structural reliability analysis [14–16].

Elkalzah et al. [17] proposed a generalized type of estimators of population mean, which is flexible and more efficient. The advantage is in the fact that it is generalized; however, such models are characterized by over-parameterization and the impossibility of optimum parameter choice. Alomani et al. [18] proposed a better estimator of population mean to be used for data on radiation. It shows efficiency improvements in comparison with classical estimators. Nonetheless, like most of the research, it is application-oriented and not much developed in terms of theory.

However, better estimators are needed that can effectively integrate the benefits of PPS sampling and auxiliary information. Specifically, much attention is given to the development of optimized generalized estimators that can achieve lower bias and MSE in realistic conditions. These estimators are particularly useful in applications where high precision in needed but resources are scarce.

These factors inspire our work and propose an optimized generalized estimator of the population mean under PPS sampling with the use of auxiliary variables. The proposed estimator is designed to effectively leverage known population parameters of auxiliary variables, in a flexible manner, thereby achieving greater efficiency than traditional methods. The formulation is as general as possible to enable some existing estimators to arise as special cases, thus providing a single framework for analyzing them. Theoretical properties of the proposed estimator are explored by obtaining the expressions of the bias and mean squared error of the proposed estimator up to the first order of approximation. Optimality conditions are set by minimizing the MSE for the parameters in question. The results are useful for making a rigorous comparison with existing estimators and for providing information on the circumstances in which the proposed estimator performs better. An empirical study is done to determine the practical utility of the proposed estimator using real data sets. The results show that the proposed estimator is consistently better than conventional estimators in a variety of situations, especially when the auxiliary variables are strongly correlated with the study variable. This underscores the importance of appropriately using auxiliary information to improve the effectiveness of estimators in PPS sampling.

1.1. Research contribution

The theoretical, methodological and practical contributions that this study has to the field of survey sampling and specifically, within the context of PPS sampling entail auxiliary information.

First, we present a general class of estimators that have been optimized for estimating the population mean. The proposed class is more comprehensive than conventional estimators that adopt fixed functional forms, as the class integrates auxiliary variables more effectively. This generalization unifies a few estimators as special cases, and thus offers a more general and flexible estimation scheme.

Second, we select the best parameter values for the generalized estimator by minimizing MSE. Such optimization is a crucial improvement of estimator efficiency over the traditional PPS estimators. The theoretical derivations impose conditions under which the proposed estimator is superior to existing estimators, providing a clear analytical benefit.

Third, the suggested estimator is effective at leveraging auxiliary variables (mean, variance, or other known parameters) to enhance accuracy. Although some of the past investigations use auxiliary information, we go a step further by optimally combining auxiliary inputs within a single structure, resulting in better robustness and efficiency.

Fourth, the research provides a thorough theoretical comparison with existing estimators for PPS sampling. Efficiency conditions are established to demonstrate the superiority of the proposed estimator in that it attains lower bias and reduced MSE, particularly when dealing with heterogeneous populations.

Fifth, the practical applicability of the proposed estimator is evidenced by real-life datasets and/or simulation studies, which ensure that the theoretical gains are reflected in the actual performance gains. This enhances the empirical applicability of the study.

Last, the research can be useful by addressing a major gap in the literature. As most researchers focus on incremental changes or single auxiliary variables, we develop a systematic, optimized, and generalized estimation framework. This not only enhances the accuracy of estimation but also serves as a basis for future extensions to the PPS sampling methodologies.

1.2. Aims and objectives

Our primary goal of this paper is to develop an effective estimation process of the population mean under PPS sampling in the presence of auxiliary information. The targeted goals are as follows:
(1) To obtain a generalized class of estimators of the population mean with PPS sampling by employing two auxiliary variables.
(2) To obtain the statistical characteristics of the proposed estimator, such as the bias and MSE, up to the first-order approximation.
(3) To find the best values of the parameters used in the proposed estimator with the least mean squared error.
(4) To set up some theoretical requirements based on which the proposed estimator could perform better than the existing estimators like Hansen-Hurwitz and ratio-type estimators.
(5) To compare the efficiency of the proposed estimator with traditional estimators using analytical and numerical measures.
(6) To verify the action of the proposed estimator with real-life data and to explore the effect of auxiliary variables, and their relation to the research variable on the effectiveness of the estimator.

This article is organized as follows: In Section 2, we introduce the notation and methodology used in the study. In Section 3, we examine the estimators to be compared. A new generalized class of estimators is proposed in Section 4, and their theoretical properties are derived. In Section 5, we present a detailed numerical study using actual data, and Section 6 conducts a simulation study to evaluate the performance of the proposed estimators across a wide range of sampling scenarios. The implications and the results of the numerical and simulation studies are discussed in Section 7. In Section 8, we summarize our findings, and in Section 9, we discuss the study's limitations and future research directions.

2. Materials and methods

Consider a population $\omega = \{\omega_1, \omega_2, ..., \omega_N\}$ comprising of N identifiable units. Here, we utilize one study and two auxiliary variables. We use both auxiliary variables at the probability assignment stage in PPS sampling, as opposed to two auxiliary variables at the estimation stage in the literature. In particular, the correlation structure of the study variable with both auxiliary variables is investigated first. Since $\rho_{vx} = 0.90$ and $\rho_{vz} = 0.85$, the variable x_i with the higher correlation is preferred as the primary size measurement, as it is x_i.

However, when z_i has a low correlation, then z_i serves as an alternative size measure, and selection probabilities are assigned as:

$$P_i = \frac{z_i}{\sum_{i=1}^{n} z_i},$$

where

$$\sum_{i=1}^{n} P_i .$$

This ensures every unit receives a well-defined positive probability of selection proportional to z_i. Moreover, P_i is further used as

$$u_i = \frac{y_i}{NP_i}, \quad v_i = \frac{x_i}{NP_i}, \quad \bar{u} = \frac{\sum_{i=1}^{n} u_i}{n} = \bar{y}_{pps}, \quad \bar{v} = \frac{\sum_{i=1}^{n} v_i}{n} = \bar{x}_{pps} .$$

This modification is unique in three ways:
(1) The choice of size measure is not arbitrary, but rather correlation-driven.
(2) The choice of size measure is flexible with regard to two auxiliary variables available, depending on the level of correlation.
(3) The modification is a true generalization of classical single-auxiliary PPS estimators, which are recovered when only one auxiliary variable is available.

$$E\left(e_0^2\right) = \frac{c_u^2}{n} = \mathcal{V}_{20}, \quad E\left(e_1^2\right) = \frac{c_v^2}{n} = \mathcal{V}_{02}, \quad E\left(e_0 e_1\right) = \mathcal{Y}\rho_{uv} C_u C_v = \mathcal{V}_{22},$$

$$\rho_{uv} = \frac{\sum_{i=1}^{N} P_i \left(u_i - \bar{Y}\right)\left(v_i - \bar{X}\right)}{S_u S_v},$$

$$S_u^2 = \sum_{i=1}^{N} P_i \left(u_i - \bar{Y}\right)^2, \quad S_v^2 = \sum_{i=1}^{N} P_i \left(v_i - \bar{X}\right)^2,$$

$$S_u = \sqrt{\sum_{i=1}^{N} P_i \left(u_i - \bar{Y}\right)^2}, \quad S_v = \sqrt{\sum_{i=1}^{N} P_i \left(v_i - \bar{X}\right)^2},$$

$$S_{uv} = \sum_{i=1}^{N} P_i \left(u_i - \bar{Y}\right)\left(v_i - \bar{X}\right),$$

where

$$\mathcal{Y} = \left(\frac{1}{n}\right).$$

3. Existing estimators

To check the performance of estimators, here we adopt some well-known estimators for estimation of the population mean, which are given along with the following properties:
(i) The traditional estimator for the mean estimation under PPS sampling is given by:

$$\hat{\bar{Y}}_{1PPS} = u , \tag{1}$$

$$Var\left(\hat{\bar{Y}}_{1PPS}\right) = \bar{Y}^2 \mathcal{V}_{20} . \tag{2}$$

(ii) The ratio estimator for population mean under PPS sampling is given by:

$$\hat{\bar{Y}}_{2PPS} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right), \tag{3}$$

$$Bias(\hat{\bar{Y}}_{2PPS}) \cong \bar{Y}\left(\mathcal{V}_{20} - \mathcal{V}_{22}\right), \tag{4}$$

$$MSE(\hat{\bar{Y}}_{2PPS}) \cong \bar{Y}^2\left(\mathcal{V}_{02} + \mathcal{V}_{02} + 2\mathcal{V}_{22}\right). \tag{5}$$

(iii) The product estimator for the population mean under PPS sampling is given by:

$$\hat{\bar{Y}}_{3PPS} = u\left(\frac{\bar{v}}{\bar{X}}\right), \tag{6}$$

$$Bias(\hat{\bar{Y}}_{3PPS}) \cong \bar{Y}\mathcal{Y}\mathcal{V}_{22}, \tag{7}$$

$$MSE(\hat{\bar{Y}}_{3PPS}) \cong \bar{Y}^2\mathcal{Y}\left(\mathcal{V}_{20} + \mathcal{V}_{02} - 2\mathcal{V}_{22}\right). \tag{8}$$

(iv) The regression estimator for the population mean under PPS sampling, is given by:

$$\hat{\bar{Y}}_{4PPS} = u + b\left(\bar{X} - \bar{v}\right), \tag{9}$$

where

$$b = \frac{S_{uv}}{S_u S_v},$$

$$Var(\hat{\bar{Y}}_{4PPS}) = MSE(\hat{\bar{Y}}_{4PPS}) \cong \bar{Y}^2\mathcal{V}_{20}\left(1 - \rho_{uv}^2\right). \tag{10}$$

(v) The exponential ratio and product type estimators are given by:

$$\hat{\bar{Y}}_{5PPS} = \bar{u}\exp\left(\frac{\bar{X} - \bar{v}}{\bar{X} + \bar{v}}\right), \tag{11}$$

$$\hat{\bar{Y}}_{6PPS} = \bar{u}\exp\left(\frac{\bar{v} - X}{\bar{v} + X}\right), \tag{12}$$

$$Bias\left(\hat{\bar{Y}}_{5PPS}\right) \cong \bar{Y}\left(\frac{3}{8}\mathcal{V}_{02} - \frac{1}{2}\mathcal{V}_{22}\right), \tag{13}$$

$$MSE\left(\hat{\bar{Y}}_{5PPS}\right) \cong \bar{Y}^2\left(\mathcal{V}_{20} + \frac{1}{4}\mathcal{V}_{02} - \mathcal{V}_{22}\right), \tag{14}$$

$$Bias\left(\hat{\bar{Y}}_{6PPS}\right) \cong \bar{Y}\left(\mathcal{V}_{11} - \frac{1}{4}\mathcal{V}_{02}\right), \tag{15}$$

$$MSE\left(\hat{\bar{Y}}_{6PPS}\right) \cong \bar{Y}^2\left(\mathcal{V}_{20} + \frac{1}{4}\mathcal{V}_{02} + \mathcal{V}_{11}\right). \tag{16}$$

(vi) Kadilar and Cingi [19] suggested the following estimators under PPS sampling, which are given by:

$$\hat{\bar{Y}}_{7PPS} = \bar{u}\left(\frac{\bar{X} - C_v}{\bar{v} - C_v}\right), \tag{17}$$

$$\hat{\bar{Y}}_{8PPS} = \bar{u}\left(\frac{\bar{X} - \beta_{2(v)}}{\bar{v} - \beta_{2(v)}}\right), \tag{18}$$

$$\hat{\bar{Y}}_{9PPS} = \bar{u}\left(\frac{\bar{X}\beta_{2(F_x)} - C_v}{\bar{v}\beta_{2(v)} - C_v}\right), \tag{19}$$

$$\hat{\bar{Y}}_{10PPS} = \bar{u}\left(\frac{\bar{X}C_v - \beta_{2(v)}}{\bar{v}C_v - \beta_{2(v)}}\right). \tag{20}$$

The MSEs of $\hat{\bar{Y}}_{Ki} (i = 1,2,3,4)$ are given by:

$$MSE\left(\hat{\bar{Y}}_{Ki}\right) \cong P_y^2\left(\mathcal{V}_{20} + R_i^2\mathcal{V}_{02} - 2R_i\mathcal{V}_{22}\right), \quad \left(i = 1,\ 2,\ 3,\ 4\right), \tag{21}$$

$$R_1 = \frac{\bar{X}}{\bar{X} - C_v}, \quad R_2 = \frac{\bar{X}}{\bar{X} - \beta_{2(v)}}, \quad R_3 = \frac{\bar{X}\beta_{2(v)}}{\bar{X}\beta_{2(v)} - C_v}, \quad R_3 = \frac{\bar{X}C_v}{\bar{X}C_v - \beta_{2(v)}},$$

(vii) The authors discussed the exponential estimator by Upadhyaya [20], which are given by:

$$\hat{\bar{Y}}_{11PPS} = \bar{u}\left(\frac{\bar{X} + \beta_{2(P_x)}}{\bar{v} + \beta_{2(P_x)}}\right). \tag{22}$$

The MSE of $\hat{\bar{Y}}_{11PPS}$ is given by:

$$MSE\left(\hat{\bar{Y}}_{11PPS}\right) \cong \bar{Y}^2\left(\mathcal{V}_{20} + R^2\mathcal{V}_{02} - 2R\mathcal{V}_{22}\right), \quad \left(i = 1,\ 2,\ 3,\ 4\right), \tag{23}$$

$$R = \frac{\bar{X}}{\bar{X} + \beta_{2(v)}}.$$

(viii) A log type estimator developed by Mishra et al. [21] using study and auxiliary information:

$$\hat{\bar{Y}}_{12PPS} = \bar{u} + wlog\left(\frac{\bar{v}}{\bar{X}}\right), \tag{24}$$

w is constant. The MSE of $\hat{\bar{Y}}_{12PPS}$ is given by:

$$MSE\left(\hat{\bar{Y}}_{12PPS}\right) \cong \bar{Y}^2\left(A - \frac{C^2}{B}\right), \tag{25}$$

where

$$A = \mathcal{V}_{20},\ B = \mathcal{V}_{02},\ C = \mathcal{V}_{22}.$$

4. Proposed class of estimators

The need for efficient and economical methods of data estimation has risen in people's minds with modern survey sampling due to the increasing complexity and size of real-world data. There are many practical populations of economic units, agricultural holdings, and business establishments whose sizes are very heterogeneous. PPS sampling is a logical and effective sampling method in such cases because larger and more influential units are assigned higher selection probabilities. However, even with its benefits, the effectiveness of estimators with PPS sampling is not always optimal, especially when auxiliary information is not properly utilized. In practice, auxiliary variables are readily accessible and highly related to the study variable. By overlooking such precious information, one may find themselves losing much of the accuracy in the estimation. Even though auxiliary information is used by classical estimators, e.g., ratio and regression estimators, its use under PPS sampling frameworks is usually restricted or inflexible. In addition, most estimators are specific to particular conditions and do not scale up to different population characteristics or the presence of multiple auxiliary variables.

The other critical issue is that, in practice, traditional estimators, including those often used under PPS sampling, might not reach minimal MSE. This limitation underscores the importance of adopting more flexible, optimized estimation methods that can dynamically adapt to the data structure and the strength of relationships among variables. The lack of such optimized generalized estimators results in a gap in the literature. More than that, as the availability of auxiliary data from administrative records and large-scale surveys continues to increase, the prospect of improving estimation efficiency by developing methodology more effectively is growing. However, bringing this information into the PPS framework while being able to track it analytically is not easy.

Given these difficulties, we aim to design an optimized generalized estimator of the population mean under PPS sampling to effectively utilize auxiliary variables. Our aim is to improve accuracy, minimize error, and offer a versatile framework that can be adjusted to various sampling scenarios. By identifying the drawbacks of current techniques and using auxiliary information more efficiently, we aim to contribute to the further development of survey sampling theory and its implementation. The suggested estimator is given by:

$$\hat{\hat{Y}}_{GenPPS}=\left[\hat{U}_{GG}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{a\bar{X}+b}{a\bar{v}+b}\right)^{\gamma_{11}}\exp\left(\frac{a\left(\bar{X}-\bar{v}\right)}{a\left(X+\bar{v}\right)+2b}\right), \tag{26}$$

where

$$\hat{U}_{GG}=\frac{\bar{u}}{2}\left[\exp\left(\frac{a\left(\bar{X}-\bar{v}\right)}{a\left(\bar{X}+\bar{v}\right)+2b}\right)+\exp\left(\frac{a\left(\bar{v}-\bar{X}\right)}{a\left(\bar{X}+\bar{v}\right)+2b}\right)\right],$$

where a and b assign different parameters like kurtosis, coefficient of variation, mean, and correlation. The values of α and β are optimized. To γ_{11}, we assign real numbers.

$$E\left(e_0^2\right)=\frac{c_u^2}{n}=\mathcal{V}_{20},\quad E\left(e_1^2\right)=\frac{c_v^2}{n}=\mathcal{V}_{02},\quad E\left(e_0e_1\right)=\mathcal{Y}\rho_{uv}C_uC_v=\mathcal{V}_{22},$$

$$\hat{\hat{Y}}_{GenPPS}=\left[\frac{\bar{Y}(1+e_0)}{2}\left[\exp\left(\frac{a\left(\bar{X}-\bar{X}(1+e_1)\right)}{a\left(\bar{X}+\bar{X}(1+e_1)\right)+2b}\right)+\exp\left(\frac{a\left(\bar{X}(1+e_1)-\bar{X}\right)}{a\left(\bar{X}+\bar{X}(1+e_1)\right)+2b}\right)\right]+\alpha\left(\bar{X}-\bar{X}(1+e_1)\right)+\beta\bar{Y}(1+e_0)\right] \\ \left(\frac{a\bar{X}+b}{a\bar{X}(1+e_1)+b}\right)^{\gamma_{11}}\exp\left(\frac{a\left(\bar{X}-\bar{X}(1+e_1)\right)}{a\left(\bar{X}+\bar{X}(1+e_1)\right)+2b}\right), \tag{27}$$

$$\hat{\hat{Y}}_{GenPPS}=\left[\frac{\bar{Y}+\bar{Y}e_0}{2}\left[\exp\left(\frac{\left(-a\bar{X}e_1\right)}{a\left(2\bar{X}+\bar{X}e_1\right)+2b}\right)+\exp\left(\frac{\left(a\bar{X}e_1\right)}{a\left(2\bar{X}+\bar{X}e_1\right)+2b}\right)\right]+\left(-a\bar{X}e_1\right)+\beta\bar{Y}(1+e_0)\right] \\ \left(\frac{a\bar{X}+b}{a\bar{X}+a\bar{X}e_1+b}\right)^{\gamma_{11}}\exp\left(\frac{\left(-a\bar{X}e_1\right)}{a\left(2\bar{X}+\bar{X}e_1\right)+2b}\right). \tag{28}$$

Subtracting $\bar{Y}$ from both sides of (28) and simplify, we obtain the following expression:

$$\hat{\hat{Y}}_{GenPPS}-\bar{Y}=\Big[-\varnothing e_1+\frac{3}{8}\varnothing^2e_1^2-\gamma_{11}\varnothing e_1+\gamma_{11}\varnothing^2e_1^2+\gamma_{11}\frac{(\gamma_{11}+1)}{2}\varnothing^2e_1^2 \\ +e_0-\varnothing e_0e_1-\gamma_{11}\varnothing e_0e_1+\frac{\varnothing^2}{8}e_1^2-\alpha Re_1-\alpha\frac{R}{2}\varnothing e_1^2+\alpha\gamma_{11}R\varnothing e_1^2 \\ +\beta+\beta e_0-\beta\frac{\varnothing}{2}e_1+\frac{3}{8}\varnothing^2e_1^2\beta-\beta\gamma_{11}\varnothing e_1+\beta\frac{\gamma_{11}}{2}\varnothing^2e_1^2+\beta\gamma_{11}\frac{(\gamma_{11}+1)}{2}\varnothing^2e_1^2-\beta\frac{\varnothing}{2}e_0e_1-\beta\gamma_{11}\varnothing e_0e_1\Big], \tag{29}$$

where

$$\varnothing=\frac{a\bar{X}}{a\bar{X}+b},\quad R=\frac{\bar{X}}{\bar{Y}}.$$

Squaring both sides and ignoring the higher order, we obtain:

$$\left(\hat{\hat{Y}}_{GenPPS}-\bar{Y}\right)^2=\bar{Y}^2\Big[\beta^2+e_1^2\Big(\varnothing^2+\gamma_{11}^2\varnothing^2+\alpha^2R^2+\frac{\beta^2\varnothing^2}{4}+\beta^2\gamma_{11}^2\varnothing^2+2\varnothing^2\gamma_{11}+2\varnothing R\alpha+\varnothing^2\beta+2\varnothing^2\beta\gamma_{11} \\ +2\gamma_{11}\varnothing R\alpha+\gamma_{11}\beta\varnothing^2+2\gamma_{11}^2\beta\varnothing^2+\alpha R\beta\varnothing+2\alpha R\gamma_{11}\beta\varnothing+\beta^2\varnothing^2\gamma_{11}\Big) \\ +e_0^2\left(1+\beta^2+2\beta\right)+2\beta e_0e_1\left(-\varnothing-\gamma_{11}\varnothing-\beta\varnothing-\beta\gamma_{11}\varnothing\right) \\ +2\beta e_1^2\Big(\frac{3}{8}\varnothing^2+\gamma_{11}\varnothing+\frac{\gamma_{11}(\gamma_{11}+1)}{2}\varnothing^2+\frac{\varnothing^2}{8}-\frac{\alpha}{2}\varnothing R+\alpha\gamma_{11}\varnothing R+\frac{3}{8}\varnothing^2\beta+\frac{\beta}{2}\gamma_{11}\varnothing^2+\beta\frac{\gamma_{11}(\gamma_{11}+1)}{2}\varnothing^2\Big) \\ +2e_0e_1\Big(-\varnothing-\varnothing\beta-\gamma_{11}\varnothing-\gamma_{11}\varnothing\beta-R\gamma_{11}-R\alpha\beta-\frac{\varnothing}{2}\beta^2-\frac{\varnothing}{2}\beta-\gamma_{11}\varnothing\beta-\gamma_{11}\varnothing\beta^2\Big)\Big]. \tag{30}$$

Taking expectations on both sides, we obtain the MSE expression:

$$E\left(\hat{\hat{Y}}_{GenPPS}-\bar{Y}\right)^2=\bar{Y}^2\Big[\beta^2+E(e_1^2)\Big(\varnothing^2+\gamma_{11}^2\varnothing^2+\alpha^2R^2+\frac{\beta^2\varnothing^2}{4}+\beta^2\gamma_{11}^2\varnothing^2+2\varnothing^2\gamma_{11}+2\varnothing R\alpha+\varnothing^2\beta+2\varnothing^2\beta\gamma_{11} \\ +2\gamma_{11}\varnothing R\alpha+\gamma_{11}\beta\varnothing^2+2\gamma_{11}^2\beta\varnothing^2+\alpha R\beta\varnothing+2\alpha R\gamma_{11}\beta\varnothing+\beta^2\varnothing^2\gamma_{11}\Big) \\ +E(e_0^2)\left(1+\beta^2+2\beta\right)+2\beta E(e_0e_1)\left(-\varnothing-\gamma_{11}\varnothing-\beta\varnothing-\beta\gamma_{11}\varnothing\right) \\ +2\beta E(e_1^2)\Big(\frac{3}{8}\varnothing^2+\gamma_{11}\varnothing+\frac{\gamma_{11}(\gamma_{11}+1)}{2}\varnothing^2+\frac{\varnothing^2}{8}-\frac{\alpha}{2}\varnothing R+\alpha\gamma_{11}\varnothing R+\frac{3}{8}\varnothing^2\beta+\frac{\beta}{2}\gamma_{11}\varnothing^2+\beta\frac{\gamma_{11}(\gamma_{11}+1)}{2}\varnothing^2 \\ +2E(e_0e_1)\Big(-\varnothing-\varnothing\beta-\gamma_{11}\varnothing-\gamma_{11}\varnothing\beta-R\gamma_{11}-R\alpha\beta-\frac{\varnothing}{2}\beta^2-\frac{\varnothing}{2}\beta-\gamma_{11}\varnothing\beta-\gamma_{11}\varnothing\beta^2\Big)\Big)\Big]. \tag{31}$$

$$
\begin{aligned}
&MSE\left(\hat{\bar{Y}}_{GenPPS}\right)=P_y^2\Big[\beta^2+\tilde{V}_{02}\Big(\varnothing^2+\gamma_{11}^2\varnothing^2+\alpha^2R^2+\frac{\beta^2\varnothing^2}{4}+\beta^2\gamma_{11}^2\varnothing^2+2\varnothing^2\gamma_{11}+2\varnothing R\alpha+\varnothing^2\beta+2\varnothing^2\beta\gamma_{11}\\
&+2\gamma_{11}\varnothing R\alpha+\gamma_{11}\beta\varnothing^2+2\gamma_{11}^2\beta\varnothing^2+\alpha R\beta\varnothing+2\alpha R\gamma_{11}\beta\varnothing+\beta^2\varnothing^2\gamma_{11}\Big)\\
&+\tilde{V}_{20}\big(1+\beta^2+2\beta\big)+2\beta E(e_0e_1)\big(-\varnothing-\gamma_{11}\varnothing-\beta\varnothing-\beta\gamma_{11}\varnothing\big)\\
&+2\beta\tilde{V}_{02}\Big(\frac{3}{8}\varnothing^2+\gamma_{11}\varnothing+\frac{\gamma_{11}(\gamma_{11}+1)}{2}\varnothing^2+\frac{\varnothing^2}{8}-\frac{\alpha}{2}\varnothing R+\alpha\gamma_{11}\varnothing R+\frac{3}{8}\varnothing^2\beta+\frac{\beta}{2}\gamma_{11}\varnothing^2+\beta\frac{\gamma_{11}(\gamma_{11}+1)}{2}\varnothing^2\\
&+2\tilde{V}_{22}\Big(-\varnothing-\varnothing\beta-\gamma_{11}\varnothing-\gamma_{11}\varnothing\beta-R\gamma_{11}-R\alpha\beta-\frac{\varnothing}{2}\beta^2-\frac{\varnothing}{2}\beta-\gamma_{11}\varnothing\beta-\gamma_{11}\varnothing\beta^2\Big)\Big)\Big].
\end{aligned}
\tag{32}
$$

After simplifying (32), we obtain:

$$
MSE\left(\hat{\bar{Y}}_{GenPPS}\right)=P_y^2\left[\alpha B_{11}+\beta B_{22}+\alpha^2B_{33}+\beta^2B_{44}+\alpha\beta B_{55}+B_{66}\right]. \tag{33}
$$

Differentiating (33) w.r.t α and β, we obtain the optimum values:

$$
\frac{\delta MSE\left(\hat{\bar{Y}}_{GenPPS}\right)}{\delta\alpha}=0,\quad \frac{\delta MSE\left(\hat{\bar{Y}}_{GenPPS}\right)}{\delta\beta}=0,
$$

$$
\alpha_{opt}=\frac{2B_{11}B_{44}-B_{22}B_{55}}{B_{55}^2-4B_{33}B_{44}},\quad \beta_{opt}=\frac{2B_{22}B_{33}-B_{11}B_{55}}{B_{55}^2-4B_{33}B_{44}}.
$$

Putting the optimum values of (33), we obtain the MSE of $\hat{\bar{Y}}_{GenPPS}$:

$$
MSE\left(\hat{\bar{Y}}_{GenPPS}\right)_{min}=\bar{Y}^2\left[B_{66}+\frac{B_{11}^2B_{44}+B_{22}^2B_{33}-B_{11}B_{22}B_{55}}{B_{55}^2-4B_{33}B_{44}}\right], \tag{34}
$$

where

$$
\begin{aligned}
&B_{11}=2\varnothing RC_v^2+2\gamma_{11}\varnothing RC_v^2-2R\rho_{uv}C_uC_v,\\
&B_{22}=\varnothing^2C_v^2+6\gamma_{11}\varnothing^2C_v^2+3\gamma_{11}\varnothing^2C_v^2+2C_u^2-5\varnothing\rho_{uv}C_uC_v-6\varnothing\rho_{uv}C_uC_v\gamma_{11}+\varnothing^2C_v^2,\\
&B_{33}=R^2C_v^2,\\
&B_{44}=1+C_v^2\frac{\varnothing^2}{4}+2C_v^2\gamma_{11}^2\varnothing^2+3C_v^2\gamma_{11}\varnothing^2+C_u^2-3\rho_{uv}C_uC_v\varnothing-4\varnothing\rho_{uv}C_uC_v\gamma_{11}\varnothing+\frac{3}{4}\varnothing^2C_v^2,\\
&B_{55}=4C_v^2R\gamma_{11}\varnothing-2R\rho_{uv}C_uC_v,\\
&B_{66}=\varnothing^2C_v^2+\gamma_{11}^2\varnothing^2C_v^2+2\varnothing^2C_v^2\gamma_{11}+C_u^2-2\rho_{uv}C_uC_v\varnothing-2\rho_{uv}C_uC_v\varnothing\gamma_{11}.
\end{aligned}
$$

5. Numerical analysis

In this section, we use three real-life datasets to assess the efficiency of all considered estimators. We obtain the MSE and PREs of all estimators. The summary statistics and results are available in detail:

Population-I: [Source: Singh [22]]
Y = Number of fish caught in 1992
X = Number of fish caught in 1992
Z = Number of fish caught in 1992

$$N = 69,\ n = 25,\ \lambda = 0.04,\ \bar{Y} = 4514.899,\ \bar{X} = 4954.435,\ \bar{Z} = 4591.072,$$
$$\bar{u} = 12970.7,\ \bar{v} = 14126.22,\ S_u^2 = 2420387,\ S_v^2 = 2007238,\ S_u = 1555.759,$$
$$S_v = 1416.77,\ C_u = 0.3445834,\ C_v = 0.28596,\ \rho_{uv} = 0.6435256.$$

Population-II: [Source: Kadilar and Cingi [23]]
Y = Apple production amount in 1999
X = Apple production amount in 1998
Z = The number of apple trees in 1999

$$N = 923,\ n = 250, \lambda = 0.004,\ \bar{Y} = 436.4345,\ \bar{X} = 333.1647,\ \bar{Z} = 11440.5,$$
$$\bar{u} = 1916.642,\ \bar{v} = 1898.964,\ S_u^2 = 14932.01,\ S_v^2 = 19503.36,\ S_u = 122.1966,$$
$$S_v = 139.6544,\ C_u = 0.2799885,\ C_v = 0.4191754,\ \rho_{uv} = 0.6960153.$$

Population-III: [Source: Murthy [24]]
Y = Production of factories
X = Stationary investment
Z = Number of employees

$$N = 34,\ n = 8, \lambda = 0.125,\ \bar{Y} = 199.4412,\ \bar{X} = 208.8824,\ \bar{Z} = 747.5882,$$
$$\bar{u} = 817.6256,\ \bar{v} = 864.0967,\ S_u^2 = 3825.127,\ S_v^2 = 3689.709,\ S_u = 61.84761,$$
$$S_v = 60.74297,\ C_u = 0.3101045,\ C_v = 0.2907999,\ \rho_{uv} = 0.9018024.$$

6. Simulation study

Our aim of this simulation study is to compare the performance of different estimators of the population mean under the PPS sampling scheme. The two common efficiency measures, MSE and percent relative efficiency (PRE), are used for the comparisons. We also aim to determine whether the suggested estimators provide greater efficiency than the traditional and existing estimators.

To ensure a robust assessment, artificial populations are generated that follow three distinct statistical distributions: Normal, Log-normal, and Gamma. In the case of normal distribution, the study variable Y and auxiliary variable X are generated with the known means and variances, typically Y∼N (50, 10^2) and X∼N (30, 5^2). In the case of the log-normal and gamma distributions, correlated variables are obtained by transforming underlying normal variables using suitable inverse distribution functions. This enables us to sample symmetric and skewed population structures, thus testing the estimators in realistic and diverse conditions. Three levels of correlation are taken into account, that is, low (ρ = 0.3), moderate (ρ = 0.6), and high (ρ = 0.9).

7. Results and discussion

Family member obtained from the generalized class of estimators are given in Table 1. The results of Tables 2–9 and Figures 1 and 2 favor the proposed class of generalized PPS estimators compared to existing estimators considered in this paper and under the three real populations. Moreover, the simulation settings used are statistically significant and consistent. The MSE for all the estimators considered for three real populations are given in Table 2. The MSEs of usual existing estimators are very large, especially in Population-I and Population-III. This poor performance is because traditional estimators are built a fixed functional form, thus not being able to completely capture the population heterogeneity. Under real data conditions, therefore, the variance of these estimators are still relatively large.

Table 1. Family members obtained from generalized estimators.

a	b	Members of generalized estimators
1	0	$\hat{\hat{Y}}_{Gen1PPS}=\left[\hat{U}_{GG1}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}}{\bar{v}}\right)^{\gamma_{11}}\exp\left(\frac{\left(\bar{X}-\bar{v}\right)}{\left(\bar{X}+\bar{v}\right)}\right)$, $\hat{U}_{GG1}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\left(\bar{X}-\bar{v}\right)}{\left(\bar{X}+\bar{v}\right)}\right)+\exp\left(\frac{\left(\bar{v}-\bar{X}\right)}{\left(\bar{X}+\bar{v}\right)}\right)\right]$.
1	$\beta_{2(P_x)}$	$\hat{\hat{Y}}_{Gen2PPS}=\left[\hat{U}_{GG2}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}+\beta_{2(v)}}{\bar{v}+\beta_{2(v)}}\right)^{\gamma_{11}}\exp\left(\frac{\left(\bar{X}-\bar{v}\right)}{\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)$, $\hat{U}_{GG2}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\left(\bar{X}-\bar{v}\right)}{\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)+\exp\left(\frac{\left(\bar{v}-\bar{X}\right)}{\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)\right]$.
1	C_v	$\hat{\hat{Y}}_{Gen3PPS}=\left[\hat{U}_{GG3}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}+C_v}{\bar{v}+C_v}\right)^{\gamma_{11}}\exp\left(\frac{\left(\bar{X}-\bar{v}\right)}{\left(\bar{X}+\bar{v}\right)+2C_v}\right)$, $\hat{U}_{GG3}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\left(\bar{X}-\bar{v}\right)}{\left(\bar{X}+\bar{v}\right)+2C_v}\right)+\exp\left(\frac{\left(\bar{v}-\bar{X}\right)}{\left(\bar{X}+\bar{v}\right)+2C_v}\right)\right]$.
C_v	$\beta_{2(v)}$	$\hat{\hat{Y}}_{Gen4PPS}=\left[\hat{U}_{GG4}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}+\beta_{2(v)}}{\bar{v}+\beta_{2(v)}}\right)^{\gamma_{11}}\exp\left(\frac{C_v\left(\bar{X}-\bar{v}\right)}{C_v\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)$, $\hat{U}_{GG4}=\frac{\bar{u}}{2}\left[\exp\left(\frac{C_v\left(\bar{X}-\bar{v}\right)}{C_v\left(\bar{X}+\bar{v}\right)+2C_v}\right)+\exp\left(\frac{C_v\left(\bar{v}-\bar{X}\right)}{C_v\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)\right]$.
$\beta_{2(v)}$	C_v	$\hat{\hat{Y}}_{Gen5PPS}=\left[\hat{U}_{GG5}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\beta_{2(v)}\bar{X}+C_v}{\beta_{2(v)}\bar{v}+C_v}\right)^{\gamma_{11}}\exp\left(\frac{\beta_{2(v)}\left(\bar{X}-\bar{v}\right)}{\beta_{2(v)}\left(\bar{X}+\bar{v}\right)+2C_v}\right)$, $\hat{U}_{GG5}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\beta_{2(v)}\left(\bar{X}-\bar{v}\right)}{\beta_{2(v)}\left(\bar{X}+\bar{v}\right)+2C_v}\right)+\exp\left(\frac{\beta_{2(v)}\left(\bar{v}-\bar{X}\right)}{\beta_{2(v)}\left(\bar{X}+\bar{v}\right)+2C_v}\right)\right]$.
$\bar{X}$	ρ_{uv}	$\hat{\hat{Y}}_{Gen6PPS}=\left[\hat{U}_{GG6}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}\bar{X}+\rho_{uv}}{\bar{X}\bar{v}+\rho_{uv}}\right)^{\gamma_{11}}\exp\left(\frac{\bar{X}\left(\bar{X}-\bar{v}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2\rho_{uv}}\right)$, $\hat{U}_{GG6}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\bar{X}\left(\bar{X}-\bar{v}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2\rho_{\rho_{uv}}}\right)+\exp\left(\frac{\bar{X}\left(\bar{v}-\bar{X}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2\rho_{uv}}\right)\right]$.
$\bar{X}$	S_v^2	$\hat{\hat{Y}}_{Gen7PPS}=\left[\hat{U}_{GG7}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}\bar{X}+S_v^2}{\bar{X}\bar{v}+S_v^2}\right)^{\gamma_{11}}\exp\left(\frac{\bar{X}\left(\left(\bar{X}-\bar{v}\right)\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2S_v^2}\right)$, $\hat{U}_{GG}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\bar{X}\left(\bar{X}-\bar{v}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2S_v^2}\right)+\exp\left(\frac{\bar{X}\left(\bar{v}-\bar{X}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2S_v^2}\right)\right]$.

Continued on next page

a	b	Members of generalized estimators
$\beta_{2(v)}$	ρ_{uv}	$\hat{\bar{Y}}_{Gen8PPS}=\left[\hat{U}_{GG8}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\beta_{2(v)}\bar{X}+\rho_{uv}}{\beta_{2(v)}\bar{v}+\rho_{uv}}\right)^{\gamma_{11}}\exp\left(\frac{\beta_{2(v)}\left(\bar{X}-\bar{v}\right)}{\beta_{2(v)}\left(\bar{X}+\bar{v}\right)+2\rho_{uv}}\right),$ $\hat{U}_{GG8}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\beta_{2(v)}\left(\bar{X}-\bar{v}\right)}{\beta_{2(v)}\left(\bar{X}+\bar{v}\right)+2\rho_{uv}}\right)+\exp\left(\frac{\beta_{2(v)}\left(\bar{v}-\bar{X}\right)}{\beta_{2(v)}\left(\bar{X}+\bar{v}\right)+2\rho_{uv}}\right)\right].$
$\bar{X}$	$\beta_{2(v)}$	$\hat{\bar{Y}}_{Gen9PPS}=\left[\hat{U}_{GG9}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\bar{X}\bar{X}+\beta_{2(v)}}{\bar{X}\bar{v}+\beta_{2(v)}}\right)^{\gamma_{11}}\exp\left(\frac{\bar{X}\left(\bar{X}-\bar{v}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right),$ $\hat{U}_{GG9}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\bar{X}\left(\bar{X}-\bar{v}\right)}{\bar{X}\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)+\exp\left(\frac{\bar{X}\left(\bar{v}-\bar{X}\right)}{X\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)\right].$
ρ_{uv}	$\beta_{2(v)}$	$\hat{\bar{Y}}_{Gen10PPS}=\left[\hat{U}_{GG10}+\alpha\left(\bar{X}-\bar{v}\right)+\beta\bar{u}\right]\left(\frac{\rho_{uv}\bar{X}+\beta_{2(v)}}{\rho_{uv}\bar{v}+\beta_{2(v)}}\right)^{\gamma_{11}}\exp\left(\frac{\rho_{uv}\left(\bar{X}-\bar{v}\right)}{\rho_{uv}\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right),$ $\hat{U}_{GG10}=\frac{\bar{u}}{2}\left[\exp\left(\frac{\rho_{uv}\left(\bar{X}-\bar{v}\right)}{\rho_{uv}\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)+\exp\left(\frac{\rho_{uv}\left(\bar{v}-\bar{X}\right)}{\rho_{uv}\left(\bar{X}+\bar{v}\right)+2\beta_{2(v)}}\right)\right].$

Table 2. MSE using real data sets.

Estimators	Population-I	Population-II	Population-III
$\hat{\bar{Y}}_{1PPS}$	60083.65	59.72805	478.1409
$\hat{\bar{Y}}_{2PPS}$	266898.4	69.12482	89.91179
$\hat{\bar{Y}}_{3PPS}$	56721.75	318.0752	1707.297
$\hat{\bar{Y}}_{4PPS}$	61780.68	30.79356	89.29399
$\hat{\bar{Y}}_{5PPS}$	165188	30.95845	178.9105
$\hat{\bar{Y}}_{6PPS}$	60081.57	155.4336	987.603
$\hat{\bar{Y}}_{7PPS}$	60144.39	69.00462	89.86493
$\hat{\bar{Y}}_{8PPS}$	60083.83	75.85645	90.45162
$\hat{\bar{Y}}_{9PPS}$	60299.55	69.13736	89.92731
$\hat{\bar{Y}}_{10PPS}$	60023.7	87.32378	92.60136
$\hat{\bar{Y}}_{11PPS}$	56721.75	63.4286	89.54806
$\hat{\bar{Y}}_{12PPS}$	49348.89	30.79356	89.29399
$\hat{\bar{Y}}_{Gen1PPS}$	49392.68	8.91973	86.54828
$\hat{\bar{Y}}_{Gen2PPS}$	49350.15	12.37483	86.95833
$\hat{\bar{Y}}_{Gen3PPS}$	49500.16	9.03113	86.59173
$\hat{\bar{Y}}_{Gen4PPS}$	49349.01	16.12661	87.75296
$\hat{\bar{Y}}_{Gen5PPS}$	49348.89	8.927503	86.56327

Continued on next page

Estimators	Population-I	Population-II	Population-III
$\hat{\bar{Y}}_{Gen6PPS}$	50824.6	8.920287	86.54893
$\hat{\bar{Y}}_{Gen7PPS}$	49350	19.50357	88.35189
$\hat{\bar{Y}}_{Gen8PPS}$	49348.89	8.980176	86.69028
$\hat{\bar{Y}}_{Gen9PPS}$	49416.76	8.931227	86.55038
$\hat{\bar{Y}}_{Gen10PPS}$	60083.65	13.6729	86.99975

The proposed generalized estimators, however, perform much better in terms of MSE values for all three populations. In Population-II, the MSE of the proposed estimator is about 8.9, while the MSE value of classical estimators are typically in the range 30–300. This dramatic decrease can be statistically explained by the inclusion of two auxiliary variables x_i and z_i. We use both auxiliary variables at the probability assignment stage in PPS sampling, as opposed to two auxiliary variables at the estimation stage in the literature. In particular, the correlation structure of the study variable with both auxiliary variables is investigated first. Since $\rho_{yx} = 0.90$ and $\rho_{yz} = 0.85$, variable x_i is higher, and the correlation is preferred as the primary size measurement, as it is x_i. However, when z_i has a low correlation, z_i serves as an alternative size measure, and selection probabilities are assigned as:

$$P_i = \frac{z_i}{\sum_{i=1}^{n} z_i},$$

where

$$\sum_{i=1}^{n} P_i = 1.$$

This ensures every unit receives a well-defined positive probability of selection proportional to z_i. Their simultaneous use enables the estimator to better depict the true population structure, giving a significant decrease in estimation error. The MSE values for the estimators proposed in Population-III are in the range of 86–88, in contrast to some of the traditional estimators, which have MSE values greater than 1700. This difference makes the proposed methodology very practical. It is logical because in populations that have a significant difference in size between units, a correlation-based size measure (from the same population) is much more effective in assigning selection probabilities than a single auxiliary measure, resulting in a more balanced and representative sample, as is most advantageous in this type of population for PPS sampling.

In Table 3, the PRE of all the estimators compared to the usual estimator is displayed. If PRE is above 100, it means that the model has performed better than the baseline estimator. The findings show that when the correlation between the study and auxiliary variables is high, the classical estimators of ratio and regression yield better PRE values than the usual estimator, but the PRE values obtained by the proposed generalized estimators are significantly larger in all three populations.

Table 3. PREs of all considered estimators using real data sets.

Estimators	Population-I	Population-II	Population-III
$\hat{\hat{P}}_{usual}$	100	100	100
$\hat{\hat{P}}_{Ratio}$	161.1345	86.40609	531.7888
$\hat{\hat{P}}_{Product}$	36.27428	18.77797	28.00573
$\hat{\hat{P}}_{Reg}$	170.6849	193.9628	535.4681
$\hat{\hat{P}}_{ER}$	156.7083	192.9297	267.2515
$\hat{\hat{P}}_{EP}$	58.60925	38.42673	48.41428
$\hat{\hat{P}}_{K1}$	161.14	86.5566	532.0661
$\hat{\hat{P}}_{K2}$	160.9717	78.73826	528.615
$\hat{\hat{P}}_{K3}$	161.134	86.39041	531.6971
$\hat{\hat{P}}_{K4}$	160.5575	68.39838	516.3432
$\hat{\hat{P}}_{Upad}$	161.2954	94.16581	533.9489
$\hat{\hat{P}}_{MMPS}$	170.6849	193.9628	535.4681
$\bar{P}_{\mathrm{Pr}oP1}$	196.1857	669.6173	552.4556
$\bar{P}_{\mathrm{Pr}oP2}$	196.0118	482.6575	549.8505
$\bar{P}_{\mathrm{Pr}oP3}$	196.1807	661.3574	552.1784
$\bar{P}_{\mathrm{Pr}oP4}$	195.5862	370.3695	544.8715
$\bar{P}_{\mathrm{Pr}oP5}$	196.1852	669.0343	552.36
$\bar{P}_{\mathrm{Pr}oP6}$	196.1857	669.5755	552.4515
$\bar{P}_{\mathrm{Pr}oP7}$	190.4894	306.2417	541.1779
$\bar{P}_{\mathrm{Pr}oP8}$	196.1813	665.1101	551.5507
$\bar{P}_{\mathrm{Pr}oP9}$	196.1857	668.7553	552.4423
$\bar{P}_{\mathrm{Pr}oP10}$	195.9163	436.8354	549.5888

The deep statistical rationale behind such a high PRE is that the proposed estimators are not based on a specific functional form for the problem. Instead, their generalized and flexible structure makes them more efficient at adaptive preference, obtaining more information from additional variables. This versatility is the central point of theoretical benefit of the proposed methodology, which can be seen in the PRE values obtained consistently in all the populations. When Tables 2 and 3 are combined, a clear and consistent relationship is: MSE values across the estimators are low when PRE values are high and vice versa. The pattern is very strong in all three populations, thus establishing the generalizability and reliability of the proposed estimators for different real data situations.

The results from the simulation study in Tables 4–9 further support and confirm the results of the real data study. In all simulation scenarios, the proposed generalized class of estimators performs the best in terms of MSE and PRE among the other competing estimators. The efficiency gain of the proposed estimator is greater as the sample size grows, which is reasonable given that the more information that is available in the sample, the more efficient the estimator will be. Moreover, the advantages of the proposed estimator are shown to be robust as the simulation results indicate that the strength of the proposed estimator is not significantly affected by different population situations, which suggests its applicability in real-world situations across survey scenarios. The numerical results of the tables are graphically illustrated in Figures 1 and 2. From the line graphs, it is evident that the proposed estimators have consistently lower MSE and higher PRE at all the populations and sample sizes when compared with the traditional estimators. The smooth and stable trend lines for the proposed estimators indicate their statistical stability and robustness, whereas the irregular and higher valued trend lines for classical estimators reveal the fact that they are limited in their ability to efficiently make use of the available auxiliary information.

Table 4. MSE of considered estimators using normal distribution.

Estimator	$\rho = 0.3, N = 500$	$\rho = 0.6, N = 500$	$\rho = 0.9, N = 500$
$\hat{\bar{Y}}_{1PPS}$	25.5809	19.9978	12.0256
$\hat{\bar{Y}}_{2PPS}$	2.1978	1.1544	0.6840
$\hat{\bar{Y}}_{3PPS}$	95.4070	77.6537	46.2446
$\hat{\bar{Y}}_{4PPS}$	2.1975	1.1503	0.6838
$\hat{\bar{Y}}_{5PPS}$	8.0840	5.7245	3.4951
$\hat{\bar{Y}}_{6PPS}$	54.6886	43.9742	26.2754
$\hat{\bar{Y}}_{7PPS}$	2.2281	1.1543	0.6923
$\hat{\bar{Y}}_{8PPS}$	2.4391	1.4515	0.8288
$\hat{\bar{Y}}_{9PPS}$	2.1989	1.1619	0.6864
$\hat{\bar{Y}}_{10PPS}$	2.4522	1.5415	0.8441
$\hat{\bar{Y}}_{11PPS}$	2.3880	1.2616	0.7661
$\hat{\bar{Y}}_{12PPS}$	2.1975	1.1503	0.6838
$\hat{\bar{Y}}_{GenPPS}$	0.7845	0.8339	0.3903

Table 5. MSE of considered estimators using normal distribution.

Estimator	$\rho = 0.3$, $N = 500$	$\rho = 0.6$, $N = 500$	$\rho = 0.9$ $N = 500$
$\hat{\bar{Y}}_{1PPS}$	100.0000	100.0000	100.0000
$\hat{\bar{Y}}_{2PPS}$	1163.9400	1732.3600	1758.1900
$\hat{\bar{Y}}_{3PPS}$	26.8100	25.7500	26.0000
$\hat{\bar{Y}}_{4PPS}$	1164.0900	1738.5000	1758.7200
$\hat{\bar{Y}}_{5PPS}$	316.4400	349.3400	344.0700
$\hat{\bar{Y}}_{6PPS}$	46.7800	45.4800	45.7700
$\hat{\bar{Y}}_{7PPS}$	1148.1100	1732.4300	1736.9800
$\hat{\bar{Y}}_{8PPS}$	1048.7900	1377.7000	1450.9100
$\hat{\bar{Y}}_{9PPS}$	1163.3300	1721.0900	1751.9400
$\hat{\bar{Y}}_{10PPS}$	1043.1900	1297.3000	1424.6700
$\hat{\bar{Y}}_{11PPS}$	1071.2100	1585.1500	1569.7800
$\hat{\bar{Y}}_{12PPS}$	1164.0900	1738.5000	1758.7200
$\hat{\bar{Y}}_{GenPPS}$	3260.7100	2398.2400	3081.4000

Table 6. MSE of considered estimators using lognormal distribution.

Estimator	$\rho = 0.3$, $N = 500$	$\rho = 0.6$, $N = 500$	$\rho = 0.9$, $N = 500$
$\hat{\bar{Y}}_{1PPS}$	19.9953	17.9698	8.7916
$\hat{\bar{Y}}_{2PPS}$	8.6232	4.9097	2.2293
$\hat{\bar{Y}}_{3PPS}$	69.1597	66.2565	29.6093
$\hat{\bar{Y}}_{4PPS}$	7.8742	4.6155	2.2181
$\hat{\bar{Y}}_{5PPS}$	9.5852	7.0365	3.7285
$\hat{\bar{Y}}_{6PPS}$	39.8534	37.7098	17.4185
$\hat{\bar{Y}}_{7PPS}$	8.2938	4.7230	2.2187
$\hat{\bar{Y}}_{8PPS}$	16.1534	33.8979	3.3391
$\hat{\bar{Y}}_{9PPS}$	8.6793	4.9312	2.2343
$\hat{\bar{Y}}_{10PPS}$	14.2994	22.5784	3.3349
$\hat{\bar{Y}}_{11PPS}$	7.9063	5.4692	2.4207
$\hat{\bar{Y}}_{12PPS}$	7.8742	4.6155	2.2181
$\hat{\bar{Y}}_{GenPPS}$	6.7221	4.3837	1.9055

Table 7. PRE of all considered estimators using lognormal distribution.

Estimator	$\rho = 0.3, N = 500$	$\rho = 0.6, N = 500$	$\rho = 0.9, N = 500$
$\hat{\bar{Y}}_{1PPS}$	100.0000	100.0000	100.0000
$\hat{\bar{Y}}_{2PPS}$	231.8800	366.0000	394.3600
$\hat{\bar{Y}}_{3PPS}$	28.9100	27.1200	29.6900
$\hat{\bar{Y}}_{4PPS}$	253.9300	389.3400	396.3500
$\hat{\bar{Y}}_{5PPS}$	208.6100	255.3800	235.7900
$\hat{\bar{Y}}_{6PPS}$	50.1700	47.6500	50.4700
$\hat{\bar{Y}}_{7PPS}$	241.0900	380.4800	396.2600
$\hat{\bar{Y}}_{8PPS}$	123.7800	53.0100	263.2900
$\hat{\bar{Y}}_{9PPS}$	230.3800	364.4100	393.4800
$\hat{\bar{Y}}_{10PPS}$	139.8300	79.5900	263.6200
$\hat{\bar{Y}}_{11PPS}$	252.9000	328.5700	363.1800
$\hat{\bar{Y}}_{12PPS}$	253.9300	389.3400	396.3500
$\hat{\bar{Y}}_{GenPPS}$	699.9700	409.9300	461.3800

Table 8. MSE of all considered estimators using gamma distribution.

Estimator	$\rho = 0.3, N = 500$	$\rho = 0.6, N = 500$	$\rho = 0.9, N = 500$
$\hat{\bar{Y}}_{1PPS}$	25.4255	14.3767	12.5472
$\hat{\bar{Y}}_{2PPS}$	14.4989	6.3573	3.9901
$\hat{\bar{Y}}_{3PPS}$	105.7509	58.0635	44.8518
$\hat{\bar{Y}}_{4PPS}$	10.4271	5.0070	3.7585
$\hat{\bar{Y}}_{5PPS}$	11.2873	5.9086	5.3002
$\hat{\bar{Y}}_{6PPS}$	56.9134	31.7617	25.7311
$\hat{\bar{Y}}_{7PPS}$	13.4070	5.9743	3.8531
$\hat{\bar{Y}}_{8PPS}$	24.8260	9.5640	5.4106
$\hat{\bar{Y}}_{9PPS}$	14.7662	6.4753	4.0298
$\hat{\bar{Y}}_{10PPS}$	20.2198	9.0871	4.9698
$\hat{\bar{Y}}_{11PPS}$	11.2546	5.2531	3.7630
$\hat{\bar{Y}}_{12PPS}$	10.4271	5.0070	3.7585
$\hat{\bar{Y}}_{GenPPS}$	8.4900	4.6991	3.6815

Table 9. PRE of all considered estimators using gamma distribution.

Estimator	$\rho = 0.3$, $N = 500$	$\rho = 0.6$, $N = 500$	$\rho = 0.9$, $N = 500$
$\hat{\bar{Y}}_{1PPS}$	100.0000	100.0000	100.0000
$\hat{\bar{Y}}_{2PPS}$	175.3600	226.1500	314.4600
$\hat{\bar{Y}}_{3PPS}$	24.0400	24.7600	27.9700
$\hat{\bar{Y}}_{4PPS}$	243.8400	287.1300	333.8300
$\hat{\bar{Y}}_{5PPS}$	225.2600	243.3200	236.7300
$\hat{\bar{Y}}_{6PPS}$	44.6700	45.2600	48.7600
$\hat{\bar{Y}}_{7PPS}$	189.6400	240.6400	325.6400
$\hat{\bar{Y}}_{8PPS}$	102.4100	150.3200	231.9000
$\hat{\bar{Y}}_{9PPS}$	172.1900	222.0200	311.3600
$\hat{\bar{Y}}_{10PPS}$	125.7500	158.2100	252.4700
$\hat{\bar{Y}}_{11PPS}$	225.9100	273.6800	333.4400
$\hat{\bar{Y}}_{12PPS}$	243.8400	287.1300	333.8300
$\hat{\bar{Y}}_{GenPPS}$	299.4700	305.9500	340.8200

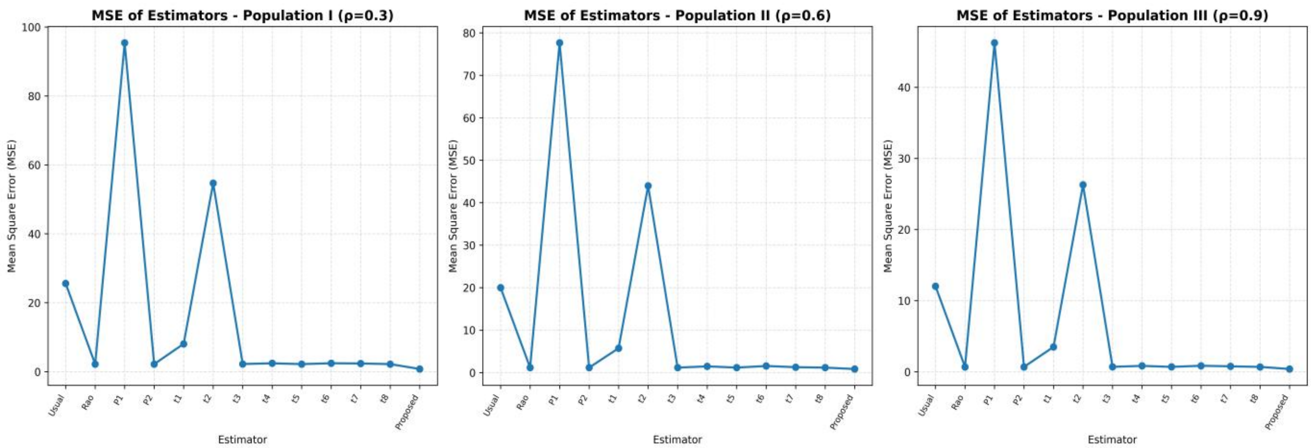

Figure 1. Sowing MSEs of all estimators with the help of line graph.

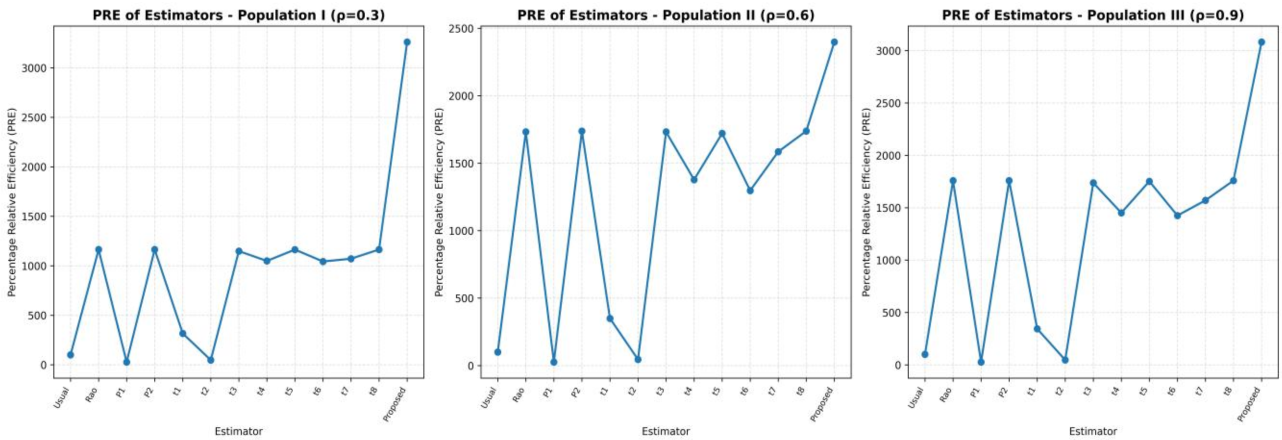

Figure 2. Sowing PREs of all estimators with the help of line graph.

The numerical results obtained from the real data sets and the simulation study indicate that the proposed generalized estimators can contribute much in the PPS sampling literature and the contribution is practically meaningful. The proposed estimators achieve higher precision, lower bias and higher efficiency than all of the existing ones, especially in the populations in which units are significantly different in size a case in which PPS sampling is most suitable and beneficial.

8. Conclusions

In this research, we proposed a more optimal generalized class of estimators for the estimation of the population mean under PPS sampling by appropriately using auxiliary variables. Our main aim was to increase the efficiency of the estimation process by using the available auxiliary information within an integrated framework. The proposed estimator was constructed as a generalized class, with a number of existing estimators being special cases, thereby providing flexibility for various population structures. To the first-order approximation, the statistical properties of the estimator, such as bias and MSE, were obtained. The best conditions for minimizing the MSE were determined, yielding an efficient estimator.

Theoretical comparisons showed that the proposed estimator is better than conventional estimators, under realistic conditions, especially when the study variable is strongly correlated with auxiliary variables. These findings were further supported by empirical results, from a real and a simulated study, demonstrating consistent gains in efficiency. The paper underscores the importance of making effective use of auxiliary information to improve the accuracy of estimators in PPS sampling. The combination of optimization with the generalized estimator offers an efficient and effective approach to working with complex survey data.

To summarize, the proposed methodology will be an important contribution to the field of survey sample estimators, as it will help overcome the limitations of existing estimators and offer a more effective alternative. It is especially applicable in large scale surveys in which the size measures and auxiliary information is easily accessible. Researchers can build on this work by examining higher-order approximations, alternative sampling designs, or applying the proposed estimator to other population parameters.

9. Future extension of the study

Although the presented optimized generalized estimator shows better efficiency in estimating the population mean under PPS sampling, several predictions remain open for further investigation and to improve the methodology.

This research is founded on the first-order approximations of bias and MSE. As such: First, researchers could extend this by deriving higher-order approximations to obtain more accurate analysis results, especially when small sample sizes are involved or the population is heavily skewed. Second, the proposed estimator can be further extended by including additional types of auxiliary information, such as multiple auxiliary variables with complex relationships, non-linear transformations, or auxiliary attributes. The consideration of incorporating such information could result in an additional increase in the efficiency of the estimation.

Third, the estimator performance can be tested with other sampling designs, including stratified PPS sampling, systematic PPS sampling, or multi-stage sampling frameworks. This would make it more applicable to more complex survey environments typically found in practice. Fourth, researchers can consider a robust version of the suggested estimator to address the presence of outliers, measurement errors, or missing data. Such strong estimators would be developed, thereby enhancing reliability in real-world data settings. Fifth, modern computational and optimization methods, including machine learning-based methods, could be used to optimize parameter selection further and enhance the performance of the estimators.

Last, the proposed methodology can be extended to estimate other population parameters, e.g., population totals, variances, or proportions, and thus the usefulness of the proposed methodology in survey sampling theory and applications can be further extended.

Author contributions

Abdulaziz S. Alghamdi: Conceptualization, software, validation, investigation, resources, writing—original draft preparation, supervision, project administration, and funding acquisition; Sohaib Ahmad: Conceptualization, methodology, formal analysis, data curation, writing—original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (IPP: 469-662-2026). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Conflict of interest

The authors declare no conflicts of interest.

References

1. A. R. El-Saeed, S. Ahmad, B. Aloraini, Robust estimator for estimation of population mean under PPS sampling: application to radiation data, *J. Radiat. Res. Appl. Sci.*, **18** (2025), 101384. https://doi.org/10.1016/j.jrras.2025.101384
2. S. M. Alghamdi, S. Ahmad, S. M. Almarzouki, B. Aloraini, M. M. Badr, M. A. Abdelkawy, Constructing a new estimator for estimating population mean utilizing auxiliary information in probability proportional to size sampling, *Alex. Eng. J.*, **110** (2025), 506–511. https://doi.org/10.1016/j.aej.2024.10.029
3. M. Azeem, S. Iftikhar, M. Ijaz, N. Salahuddin, M. Ilyas, An improved estimator of population mean under PPS sampling with application to radiation data sets, *J. Radiat. Res. Appl. Sci.*, **18** (2025), 101543. https://doi.org/10.1016/j.jrras.2025.101543
4. H. P. Singh, A. C. Mishra, S. K. Pal, Improved estimator of population total in PPS sampling, *Commun. Stat. Theory Methods*, **47** (2018), 912–934. https://doi.org/10.1080/03610926.2017.1313987
5. A. B. Ghorbal, Robust estimation of mean under probability proportional to size sampling: An application to radiation science and engineering data, *J. Radiat. Res. Appl. Sci.*, **19** (2026), 102404. https://doi.org/10.1016/j.jrras.2026.102404
6. M. Hussein Mohamud, F. A. Mohamud, Estimation of the mean using robust regression and probability proportional to size sampling, *Stat. Theory Relat. Fields*, **9** (2025), 213–222. https://doi.org/10.1080/24754269.2025.2516339
7. S. Shah, E. Mahmoudi, H. Iftikhar, P. C. Rodrigues, R. I. Gonzales Medina, J. L. López-Gonzales, A novel family of CDF estimators under PPS sampling: computational, theoretical, and applied perspectives, *Axioms*, **14** (2025), 796. https://doi.org/10.3390/axioms14110796
8. S. Ahmad, J. Shabbir, E. Zahid, M. Aamir, Improved family of estimators for the population mean using supplementary variables under PPS sampling, *Sci. Prog.*, **106** (2023), 00368504231180085. https://doi.org/10.1177/00368504231180085
9. M. Azeem, An optimal quantitative randomized response technique under PPS sampling design, *Commun. Stat. Theory Methods*, **55** (2025), 3534–3546. https://doi.org/10.1080/03610926.2025.2580525
10. M. Mustafa, S. Ahmad, E. Zahid, J. Shabbir, S. Masood, Novel methods for estimation of population mean using auxiliary information under PPS sampling: application with real and simulated data sets, *Kurd. Stud.*, **12** (2024), 1553–1562. https://doi.org/10.53555/ks.v12i5.3508
11. J. Wang, S. Ahmad, M. Arslan, S. A. Lone, A. H. Abd Ellah, M. A. Aldahlan, et al., Estimation of finite population mean using double sampling under probability proportional to size sampling in the presence of extreme values, *Heliyon*, **9** (2023), e21418. https://doi.org/10.1016/j.heliyon.2023.e21418

12. S. Ahmad, E. Zahid, J. Shabbir, M. Aamir, R. Onyango, Enhanced estimation of the population mean using two auxiliary variables under probability proportional to size sampling, *Math. Probl. Eng.*, **2023** (2023), 5564360. https://doi.org/10.1155/2023/5564360
13. S. Khan, M. Farooq, S. Ahmad, S. Khan, Improved estimator for the estimation of population mean using a predictive approach under PPS sampling, *VFAST Trans. Math.*, **12** (2024), 1–16.
14. S. Yang, D. Meng, H. Yang, C. Luo, X. Su, Enhanced soft Monte Carlo simulation coupled with support vector regression for structural reliability analysis, *Proc. Inst. Civ. Eng. Transp.*, **178** (2025), 459–474. https://doi.org/10.1680/jtran.24.00128
15. S. Yang, D. Meng, B. Keshtegar, A. M. P. De Jesus, M. Alfouneh, S. P. Zhu, ASVR-GPSO: A novel hybrid active support vector regression and global-best partial swarm optimization for structural reliability analysis, *Struct. Multidiscip. Optim.*, **68** (2025), 196. https://doi.org/10.1007/s00158-025-04134-4
16. S. Yang, D. Meng, H. Yang, B. Keshtegar, A. M. P. De Jesus, S. Zhu, Adaptive Kriging-assisted enhanced sparrow search with augmented-Lagrangian first-order reliability method for highly efficient structural reliability analysis, *Reliab. Eng. Syst. Saf.*, **267** (2026), 111916. https://doi.org/10.1016/j.ress.2025.111916
17. B. Elkalzah, M. El-Morshedy, H. S. Shahen, M. Elgarhy, A new generalized class of estimators for estimation of population mean under PPS sampling: application with radiation science, *J. Radiat. Res. Appl. Sci.*, **18** (2025), 101738. https://doi.org/10.1016/j.jrras.2025.101738
18. G. Alomani, M. Kayid, S. Ahmad, Application to radiation data sets by suggesting an improved mean estimator under probability proportional to size sampling, *J. Radiat. Res. Appl. Sci.*, **18** (2025), 101354. https://doi.org/10.1016/j.jrras.2025.101354
19. C. Kadilar, H. Cingi, An improvement in estimating the population mean by using the correlation coefficient, *Hacet. J. Math. Stat.*, **35** (2006), 103–109.
20. L. N. Upadhyaya, H. P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, *Biom. J.*, **41** (1999), 627–636. Available from: https://api.semanticscholar.org/CorpusID:120685207.
21. P. Mishra, A. Sharma, N. K. Adichwal, S. Rai, R. Singh, Log-product-type estimator for estimation of population variance using auxiliary information, *Thail. Stat.*, **22** (2024), 610–617. Available from: https://ph02.tci-thaijo.org/index.php/thaistat/article/view/254771.
22. S. Singh, *Advanced sampling theory with applications: how Michael "selected" Amy*, Dordrecht: Springer Netherlands, 2003.
23. C. Kadilar, H. Cingi, Ratio estimators in stratified random sampling, *Biom. J.*, **45** (2003), 218–25. https://doi.org/10.1002/bimj.200390007
24. M. N. Murthy, *Sampling theory and methods*, Calcutta: Statistical Publishing Society, 1967.