International Journal of Advanced Statistics and Probability, 13 (1) (2026) 115-129

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
https://doi.org/10.14419/37wmqj74
Research paper

Development of Robust Exponential Estimators for Population Mean Estimation Under PPS Sampling: Applications In Medical, Agricultural, And Displacement Contexts in The USA

Anchal Yadav [1] *, Anuj Yadav [2]

[1] Department of Statistics, University of Lucknow, Lucknow-226007, Uttar Pradesh, India
[2] Department of Mathematics, National Institute of Technology, Rourkela-769008, Odisha, India
**Corresponding author E-mail: yanchal635@gmail.com*

Received: 26-07-2026, Accepted: 26-07-2026, Published: 10-08-2026

Abstract

The main objective of this research is to explore the effective use of auxiliary variables to enhance the estimation of the population mean of the study variable under consideration. The proposed generalized estimators demonstrate superior performance compared to existing methods, particularly in terms of reduced mean squared error (MSE), which is used as the primary measure of precision. Using first-order approximation techniques, the study examines important statistical properties of the proposed estimators, with a specific focus on their bias and MSE under the Probability Proportional to Size (PPS) sampling framework. PPS sampling assigns higher inclusion probabilities to units with larger size measures, making it highly efficient for heterogeneous populations. A new generalized class of estimators is developed by incorporating auxiliary information to improve the accuracy of population mean estimation under PPS sampling. The methodology is applied to a real-world radiation science dataset, where average ambient radiation levels are recorded across different geographical regions selected based on area size and population distribution. In addition, a simulation study is conducted using artificially generated datasets to validate the performance of the proposed estimators under controlled conditions. Graphical analysis and comparative charts are also pre-pared using both simulated and real datasets to visually demonstrate the superiority of the proposed estimators. The results clearly indicate that the suggested estimators achieve lower MSE compared to existing approaches, making them highly suitable for radiological risk as-assessment and environmental health monitoring applications.

Keywords: *PPS Sampling; Mean Squared Error (MSE); Percentage Relative Efficiency (PRE); Simulation Study; Auxiliary Variable; Efficiency; Visualization.*

1. Introduction

Auxiliary information plays a highly significant role in survey sampling, as it provides an effective means of improving the efficiency and precision of estimators. Such information can be incorporated at multiple stages of the survey process, including during the design phase, the estimation phase, or even both simultaneously for optimal results. When there is a meaningful correlation between the study variable and auxiliary variables, estimation techniques such as ratio, product, and regression estimators become particularly powerful tools. Over time, researchers have gone beyond these classical approaches and developed a wide variety of advanced estimators by combining them, including exponential-type estimators that further exploit inter-variable relationships. In real-world applications, collecting data on the primary variable of interest is often expensive, labor-intensive, and time-consuming. In contrast, auxiliary variables are typically easier, quicker, and more economical to measure. This practical advantage has led to their increasing use among survey statisticians to enhance both the accuracy and reliability of estimators. Consequently, many modern surveys deliberately integrate auxiliary information into their frameworks, making use of its correlation with the study variable to achieve better estimation outcomes at both the design and estimation stages.

A major challenge in survey sampling has always been improving the accuracy of estimators used for population parameter estimation. To address this, numerous methodologies based on auxiliary variables, especially regression, ratio, and product estimators, have been extensively studied and documented. In addition, several estimators have been proposed for situations involving multiple auxiliary variables, with the aforementioned approaches being the most widely used. Researchers have also investigated the effectiveness of different statistical measures, such as variance, kurtosis, and the coefficient of variation, in predicting population characteristics. Ensuring representativeness requires selecting a sufficiently large sample size, particularly when dealing with variability in the population. In homogeneous populations, simple random sampling methods, with or without replacement, are often adequate. However, the application of ratio, product, or regression estimators generally requires prior knowledge of population parameters related to auxiliary variables. By properly calibrating and incorporating such information, many improved estimators have been introduced in the literature. The estimation of population parameters is the fundamental objective of survey sampling and is essential across various fields, including social sciences, economics, health sciences, agriculture, and environmental studies. Among all parameters, the population mean is especially important as it provides a measure of the central tendency of the entire population. Designing an efficient sampling strategy is crucial to maintaining accuracy and consistency, particularly when dealing with heterogeneous populations where units differ in size and importance. Probability Proportional to Size (PPS) sampling is one of the most widely used techniques, in which units are selected based on an auxiliary variable reflecting their size or relevance. This method is particularly advantageous when larger units contribute more significantly to the variable of interest. For example, in environmental and radiation studies, variables such as geographic area and population density are often closely associated with radiation exposure and environmental risk. Increasing the probability of including such influential units enhances estimation efficiency.

Despite these advantages, classical estimators such as the Horvitz-Thompson estimator, when used with PPS sampling, may suffer from high variance, particularly when auxiliary information is not effectively utilized or when the data distribution is highly skewed. To overcome these limitations, researchers have developed numerous modified estimators aimed at reducing bias and mean squared error through the incorporation of additional information. While ratio, product, regression, and exponential estimators have demonstrated strong performance in simple random and stratified sampling frameworks, there remains a noticeable gap in research concerning specialized estimators tailored for PPS sampling. This gap is especially evident in complex real-world applications such as radiation monitoring. The field of radiation science introduces unique challenges due to spatial variability, industrial activities, uneven population distribution, and varying proximity to natural and artificial radiation sources. Therefore, obtaining accurate estimates of average radiation levels is critically important for public health planning, environmental protection, policy formulation, and post-nuclear event assessment. The development of finite population estimation has evolved significantly from classical ratio and regression estimators to highly sophisticated classes of estimators incorporating auxiliary information, multiple sampling schemes, and simulation-based validation. Early foundational work by Murthy (1967) established the theoretical basis of sampling theory, which later researchers expanded into more efficient estimation strategies using auxiliary variables.

Initial improvements in estimation accuracy were introduced through ratio-type and product-type estimators. Chakrabarty (1979) and Sisodia and Dwivedi (1981) contributed modified ratio estimators incorporating auxiliary variable properties such as the coefficient of variation. Bahl and Tuteja (1991) further advanced this field by proposing exponential ratio and product-type estimators, which improved efficiency under simple random sampling (SRS). Subramani and Kumarapandiyan (2012) extended these ideas using quartiles of auxiliary variables, demonstrating that non-traditional auxiliary functions could significantly reduce mean squared error (MSE). Parallel developments in sampling strategies emphasized probability proportional to size (PPS) sampling. Singh et al. (2018), Latpate et al. (2021), Olayiwolla et al. (2019), and Sinha and Khanna (2022) highlighted improvements in PPS-based estimation, particularly in handling variability and measurement errors. Singh, Mishra, and Pal (2018) proposed improved estimators for population totals under PPS sampling, strengthening the applicability of unequal probability designs.

A major shift in modern research focuses on the dual or multiple use of auxiliary variables. Kadilar and Cingi (2007) demonstrated early improvements using correlation coefficients. This idea was significantly expanded in recent works by Ahmad and Shabbir (2018, 2024), Ahmad et al. (2021-2024), Hussain et al. (2020-2022), and Semary et al. (2024), who developed estimators using dual auxiliary information under SRS, stratified random sampling (STRS), and PPS schemes. These studies consistently showed improved efficiency in estimating population means, proportions, variance, and distribution functions. Ahmad et al. (2022, 2023, 2024) made substantial contributions by introducing enhanced estimators using two auxiliary variables, often validated through simulation studies and real datasets. Their work addressed challenges such as non-response, extreme values, and stratification, providing robust and adaptable estimation frameworks. Similarly, Hussain et al. (2020, 2021, 2022) focused on distribution function estimation and non-response problems, demonstrating that dual auxiliary information significantly enhances estimator reliability. Recent studies have further generalized these approaches. Iftikhar et al. (2022), Wang et al. (2023), Gupta et al. (2024), Kumar et al. (2024), and Raja and Maqbool (2024) developed hybrid and exponential-type estimators integrating auxiliary attributes and variables. These models have been validated using simulation techniques, reinforcing their superiority over traditional estimators.

Moreover, research has extended beyond mean estimation to include estimation of variance, proportions, and distribution functions. Semary et al. (2024) and Iqbal et al. (2024) introduced generalized estimator families capable of handling complex sampling designs. Lawson (2022) and Zaman (2018, 2019) contributed to improved estimator classes using prior information and auxiliary attributes.

Overall, the literature reflects a clear progression: from classical ratio estimators to advanced multi-auxiliary, model-based, and simulation-validated estimation techniques under diverse sampling frameworks such as SRS, STRS, and PPS. The consistent conclusion across studies is that auxiliary information, especially when used in dual or multiple forms, significantly improves estimator efficiency and robustness.

Research Gap

Probability proportional to size (PPS) sampling is widely regarded as a highly effective and reliable technique in survey sampling, particularly in situations where the population units exhibit significant variation in size, scale, or relative importance. In such contexts, the Horvitz–Thompson estimator has been widely used for its desirable theoretical property of unbiasedness. However, despite this advantage, the estimator often suffers from the drawback of high variance, especially when the population is highly heterogeneous or when auxiliary information is not efficiently incorporated into the estimation process. This limitation reduces its overall effectiveness in practical applications and creates challenges for practitioners who require both accuracy and stability in their estimates. Over the years, several improved estimators such as ratio, product, regression, and exponential-type estimators have been proposed to address these issues. Nevertheless, these enhanced estimators often face practical constraints when applied directly within the PPS framework, primarily due to the complexity involved in their optimization and implementation.

Furthermore, there exists a notable gap in the literature regarding the development of flexible and generalized classes of estimators specifically designed for PPS sampling that can effectively utilize auxiliary information to minimize the mean squared error (MSE). This gap becomes even more significant in applied scientific domains such as radiation science, where the adoption of advanced statistical methodologies has been relatively limited. In such fields, data often exhibit high variability due to spatial heterogeneity, industrial influences, and varying levels of exposure across regions. At the same time, auxiliary information such as population density, geographical area, and historical radiation levels is often readily available and can be leveraged to improve estimation accuracy.

Despite this favorable setting, limited research has been conducted to design estimators that fully exploit these auxiliary variables under PPS sampling. This highlights the need for developing a more adaptable and robust class of estimators that not only satisfies theoretical requirements such as unbiasedness and minimum variance but also addresses practical considerations in real-world applications. In particular, in radiation physics and environmental monitoring, accurate estimation of population parameters is crucial for public safety assessments, policy formulation, and risk evaluation. Therefore, the present study aims to bridge this research gap by proposing a novel and flexible estimator framework. The proposed approach is designed to effectively incorporate auxiliary information, reduce estimation error, and improve efficiency. Its performance is further validated through both theoretical analysis and empirical evaluation using simulated as well as real-world datasets, thereby ensuring its applicability in practical scenarios.

Motivation

Advanced sampling and estimation techniques are essential in survey sampling to obtain accurate results, especially when dealing with highly heterogeneous populations such as those encountered in radiation science. Radiation exposure levels vary considerably across geographical regions due to differences in land use practices, population distribution, and proximity to radiation sources. To account for such variability, sampling procedures must incorporate population-specific characteristics, for which Probability Proportional to Size (PPS) sampling provides an effective solution. PPS sampling improves representativeness; however, classical Horvitz–Thompson estimators may exhibit high variance unless auxiliary information is properly utilized. This study focuses on enhancing the efficiency of PPS sampling methods, as they may still lead to higher bias and mean squared error (MSE) in estimation. Incorporating auxiliary variables through generalized estimation techniques significantly improves the accuracy and reliability of population mean estimation. The proposed generalized estimator framework is applicable across multiple domains, including transportation, healthcare, manufacturing, retail, and economics, where high prediction accuracy is required.

- The proposed method supports predictive maintenance in industrial environments by systematically analyzing large volumes of manufacturing equipment data collected through advanced sensor systems installed across production lines.
- By utilizing real-time and historical sensor data, the approach can accurately predict potential failures or breakdowns in critical machine components within automobile manufacturing factories, thereby enabling timely maintenance actions that significantly reduce unexpected downtime and repair costs.
- This predictive capability ultimately leads to a more streamlined, efficient, and continuous production process, while simultaneously improving overall operational productivity and minimizing unnecessary maintenance expenditures across industrial systems.
- The estimators contribute significantly to economic forecasting by generating more precise and reliable predictions of key macroeconomic indicators such as Gross Domestic Product (GDP) growth rates, unemployment trends, and inflation levels.
- These improved forecasting capabilities assist both governmental bodies and private sector organizations in making more informed decisions, optimizing resource allocation, and designing more effective economic policies based on accurate data-driven insights.
- In environmental applications, the proposed methodology enhances the estimation and monitoring of pollutant concentrations as well as broader environmental changes occurring over time across different geographic regions.
- It further supports the prediction of ecological and climate-related patterns, enabling researchers to better understand long-term environmental impacts and assess risks associated with climate change and ecosystem disruption.
- Within the transportation sector, the method improves route optimization by analyzing traffic flow patterns and predicting delays in advance, which helps logistics and delivery companies plan more efficient travel routes.
- As a result, transportation systems benefit from reduced fuel consumption, shorter delivery times, and improved operational efficiency in logistics and supply chain management.
- In public health applications, the estimators enhance epidemiological modeling by providing more accurate forecasts of disease transmission patterns, healthcare demand, and potential outbreak scenarios.
- This improved predictive ability supports better healthcare planning, resource distribution, and intervention strategies, ensuring that medical systems are better prepared to respond to future health challenges.
- In the retail sector, the approach improves demand forecasting for various consumer products, allowing businesses to maintain optimal inventory levels and avoid issues such as overstocking or stock shortages.
- By achieving better balance in inventory management, retail organizations can enhance customer satisfaction, reduce wastage, and improve overall supply chain efficiency.
- In the field of medical imaging, the proposed method enhances diagnostic accuracy by integrating multiple imaging modalities such as computed tomography (CT) scans and magnetic resonance imaging (MRI) for more comprehensive analysis.
- This integration facilitates earlier and more accurate detection of diseases, including cancer, by improving image segmentation and enabling more precise classification of abnormal tissues such as tumor regions.
- Overall, this demonstrates the versatility and wide applicability of the proposed approach across multiple domains, highlighting its potential to significantly transform decision-making processes in industry, healthcare, environment, and economic systems.

2. Notations and Terminology

Let us consider a population of $\psi_1, \psi_{2,\dots,}\psi_N$ should be considered. It highlights the characteristics associated with the auxiliary variables. $X_{(i)}; Z_{(i)}$ and the study variable is $Y_{(i)}$. Under PPS sampling, a sample of size n is selected from the population.

$P_i = \frac{z_i}{\sum_{i=1}^{n} z_i}$identity the PPS sample for i^{th} units.

$$u_i = \frac{y_i}{NP_i}, v_i = \frac{x_i}{NP_i},$$

$$\bar{u} = \frac{\sum_{i=1}^{n} u_i}{n} = \bar{y}_{PPS} \text{ and } \bar{v} = \frac{\sum_{i=1}^{n} v_i}{n} = \bar{x}_{PPS},$$

$$\xi_0 = \frac{\bar{u}-\bar{Y}}{\bar{Y}} \text{ or } \bar{u} = \bar{Y}(1+\xi_0)$$

$$\xi_1 = \frac{\bar{v}-\bar{X}}{\bar{X}} \text{ or } \bar{v} = \bar{Y}(1+\xi_1)$$

$E(\xi_i) = 0$, as $i = 0,1$

$E(\xi_0^2) = \lambda C_u^2$, $E(\xi_1^2) = \lambda C_v^2$, $E(\xi_0\xi_1) = \lambda\rho_{uv}C_uC_v$

$C_u^2 = \frac{S_u^2}{\bar{u}}$, $C_v^2 = \frac{S_v^2}{\bar{v}}$

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

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

$S_{uv} = \sum_{i=1}^{N} P_i(u_i - \bar{Y})(v_i - \bar{X})$ and $\rho_{uv} = \frac{\sum_{i=1}^{N} P_i(u_i-\bar{Y})(v_i-\bar{X})}{S_uS_v}$.

3. Existing estimators under PPS sampling

Under the framework of Probability Proportional to Size (PPS) sampling, a variety of estimators have been proposed for the estimation of population mean and total. These estimators differ in structure and efficiency depending on the use of auxiliary information. The following section presents a brief overview of the existing estimators along with their bias and mean squared error (MSE).

1) Under PPS sampling, the conventional estimator and its corresponding variance are defined as follows.

$$\mathrm{Var}(t_{1(pps)}) = \lambda\bar{Y}^2C_u^2 \; t_{1(pps)} = \bar{u} \tag{1}$$

2) Under PPS sampling, the ratio estimator and its corresponding bias and MSE are defined as follows.

$$t_{2(pps)} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right)$$

$$\mathrm{Bias}(t_{2(pps)}) \cong \bar{Y}\lambda[C_u^2 - \rho_{uv}C_uC_v] \tag{2}$$

$$\mathrm{MSE}(t_{2(pps)}) \cong \bar{Y}^2\lambda[C_u^2 + C_v^2 - 2\rho_{uv}C_uC_v] \tag{3}$$

3) Under PPS sampling, the product estimator and its corresponding bias and MSE are defined as follows.

$$t_{3(pps)} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right) \tag{4}$$

$$\mathrm{Bias}(t_{3(pps)}) \cong \bar{Y}\lambda\rho_{uv}C_uC_v \tag{5}$$

$$\mathrm{MSE}(t_{3(pps)}) \cong \bar{Y}^2\lambda[C_u^2 + C_v^2 + 2\rho_{uv}C_uC_v]$$

4) Under PPS sampling, by the Punjab Bureau of Statistics (2021-22), its corresponding bias and MSE are defined as follows.

$$t_{4(pps)} = \bar{u}\exp\left(\frac{\bar{X}-\bar{v}}{\bar{X}+\bar{v}}\right) \tag{6}$$

$$\mathrm{Bias}(t_{4(pps)}) \cong \bar{Y}\lambda\left(\frac{3}{8}C_v^2 - \frac{1}{2}\rho_{uv}C_uC_v\right) \tag{7}$$

$$\mathrm{MSE}(t_{4(pps)}) \cong \bar{Y}^2\lambda\left(C_u^2 + \frac{1}{4}C_v^2 - \rho_{uv}C_uC_v\right).$$

5) Under PPS sampling, by the Punjab Bureau of Statistics (2021-22) its corresponding bias and MSE are defined as follows.

$$t_{5(pps)} = \bar{u}\exp\left(\frac{\bar{v}-\bar{X}}{\bar{v}+\bar{X}}\right)$$

$$\mathrm{Bias}(t_{5(pps)}) \cong \bar{Y}\lambda\left(\rho_{uv}C_uC_v - \frac{1}{4}C_v^2\right) \tag{8}$$

$$\mathrm{MSE}(t_{5(pps)}) \cong \bar{Y}^2\lambda\left(C_u^2 + \frac{1}{4}C_v^2 + \rho_{uv}C_uC_v\right) \tag{9}$$

6) Under PPS sampling, the conventional ratio estimator and its corresponding variance are defined as follows.

$$t_{6(pps)} = (1 - \varrho_1)\bar{u} + \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right)$$

$$\mathrm{Bias}(t_{6(pps)}) \cong \bar{Y}\lambda\left(\frac{\varrho_1}{2}C_v^2 - \varrho_1\rho_{uv}C_uC_v\right) \tag{10}$$

$$\mathrm{MSE}(t_{6(pps)}) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_1^2C_v^2 - 2\varrho_1\rho_{uv}C_uC_v]. \tag{11}$$

7) Under PPS sampling, Sisodia and Dwivedi (1981) suggested an improved ratio-type estimator using the coefficient of variation estimator and its corresponding bias and MSE defined as follows.

$$t_{7(pps)} = \bar{u}\exp\left(\frac{\bar{X}-C_v}{\bar{X}+C_v}\right)$$

$$\text{Bias}\left(t_{7(pps)}\right) \cong \bar{Y}\lambda(\varrho_2^2 C_v^2 - \varrho_2\rho_{uv}C_uC_v) \quad (12)$$

$$\text{MSE}\left(t_{7(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_2^2 C_v^2 - 2\varrho_2\rho_{uv}C_uC_v]. \quad (13)$$

8) Under PPS sampling, Kadilar and Cingi (2007) suggested a ratio estimator by using the coefficient of variation and correlation between the study variable and auxiliary variables, with bias and MSE defined as follows:

$$t_{8(pps)} = \left(\frac{\bar{u}}{\bar{v}C_v+\rho_{uv}}\right)[\bar{X}\,C_v + \rho_{uv}]$$

$$\text{Bias}\left(t_{8(pps)}\right) \cong \bar{Y}\lambda(\varrho_3^2 C_v^2 - \varrho_3\rho_{uv}C_uC_v) \quad (14)$$

$$\text{MSE}\left(t_{8(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_3^2 C_v^2 - 2\varrho_3\rho_{uv}C_uC_v]. \quad (15)$$

$$\varrho_3 = \frac{\bar{X}\,C_v}{\bar{X}\,C_v+\rho_{uv}}$$

9) Under PPS sampling, Subramani and Kumarpandiyan (2012) suggested a ratio estimator by using the third quartile, with bias and MSE defined as follows:

$$t_{9(pps)} = \left(\frac{\bar{u}}{\bar{v}C_v+\rho_{uv}}\right)[\bar{X}\,C_v + \rho_{uv}]$$

$$\text{Bias}\left(t_{9(pps)}\right) \cong \bar{Y}\lambda(\varrho_4^2 C_v^2 - \varrho_4\rho_{uv}C_uC_v) \quad (16)$$

$$\text{MSE}\left(t_{9(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_4^2 C_v^2 - 2\varrho_4\rho_{uv}C_uC_v]. \quad (17)$$

$$\varrho_4 = \frac{\bar{X}}{\bar{X}+Q_{3(v)}}$$

10) Under PPS sampling, Subramani and Kumarpandiyan (2012) suggested a ratio estimator by using the first quartile, with bias and MSE defined as follows:

$$t_{10(pps)} = \bar{u}\left(\frac{\bar{X}+Q_{1(v)}}{\bar{v}+Q_{1(v)}}\right)[\bar{X}\,C_v + \rho_{uv}]$$

$$\text{Bias}\left(t_{10(pps)}\right) \cong \bar{Y}\lambda(\varrho_5^2 C_v^2 - \varrho_5\rho_{uv}C_uC_v) \quad (18)$$

$$\text{MSE}\left(t_{10(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_5^2 C_v^2 - 2\varrho_5\rho_{uv}C_uC_v]. \quad (19)$$

$$\varrho_5 = \frac{\bar{X}}{\bar{X}+Q_{1(v)}}$$

11) Under PPS sampling, Murthy (1967) suggested a ratio estimator by using the median of the auxiliary variables, with bias and MSE defined as follows:

$$t_{11(pps)} = \bar{u}\left(\frac{\bar{X}\,C_v+M_x}{\bar{v}C_v+M_x}\right)$$

$$\text{Bias}\left(t_{11(pps)}\right) \cong \bar{Y}\lambda(\varrho_6^2 C_v^2 - \varrho_6\rho_{uv}C_uC_v) \quad (20)$$

$$\text{MSE}\left(t_{11(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_6^2 C_v^2 - 2\varrho_6\rho_{uv}C_uC_v]. \quad (21)$$

$$\varrho_6 = \frac{\bar{X}\,C_v}{\bar{X}\,C_v+M_x}$$

12) Under PPS sampling, Murthy (1967) suggested a ratio estimator by using the median of the auxiliary variables, with bias and MSE defined as follows

$$t_{12(pps)} = \bar{u}\left(\frac{\bar{X}+Q.D_x}{\bar{v}+Q.D_x}\right)$$

$$\text{Bias}\left(t_{12(pps)}\right) \cong \bar{Y}\lambda(\varrho_7^2 C_v^2 - \varrho_7\rho_{uv}C_uC_v) \quad (22)$$

$$\text{MSE}\left(t_{12(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_7^2 C_v^2 - 2\varrho_7\rho_{uv}C_uC_v]. \quad (23)$$

$Q.D_x = \frac{Q_3-Q_1}{2}$, $\varrho_7 = \frac{\bar{X}}{\bar{X}+Q.D_x}$

13) Under PPS sampling, Murthy (1967) suggested a ratio estimator by using the median of the auxiliary variables, with bias and MSE defined as follows

$$t_{13(pps)} = \bar{u}\left(\frac{\bar{X}+Q.Da_{(x)}}{\bar{v}+Q.Da_{(x)_x}}\right)$$

$$\text{Bias}\left(t_{13(pps)}\right) \cong \bar{Y}\lambda(\varrho_8^2C_v^2 - \varrho_8\rho_{uv}C_uC_v) \tag{24}$$

$$\text{MSE}\left(t_{13(pps)}\right) \cong \bar{Y}^2\lambda[C_u^2 + \varrho_8^2C_v^2 - 2\varrho_8\rho_{uv}C_uC_v]. \tag{25}$$

$Q.Da_{(x)} = \frac{Q_3+Q_1}{2}$, $\varrho_7 = \frac{X}{\bar{X}+Q.Da_{(x)}}$

4. Proposed Estimators under PPS Sampling

Motivated by the work of Abiodun (2021) and Kanwai (2024), this study proposes a new class of estimators under Probability Proportional to Size (PPS) sampling. Their contributions emphasize the importance of incorporating auxiliary information to improve estimation efficiency. Building on this idea, the proposed estimators make effective use of auxiliary variables to enhance the accuracy of estimation. In this study, we have proposed two estimators within the suggested class. Both estimators are designed to utilize size measures along with known auxiliary information to reduce bias and improve precision. The use of auxiliary information helps in strengthening the relationship between the study variable and sampling design. The proposed approach aims to achieve lower mean squared error (MSE) compared to existing estimators. The estimators are also flexible and can be adapted under different sampling situations. Theoretical properties, including bias and MSE, are derived to evaluate their performance. Overall, the proposed estimators are expected to provide improved efficiency over traditional PPS estimators.

Proposed Estimator-1

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{u}\left[\left(\frac{\bar{v}}{\bar{X}}\right)^{\alpha} + \exp\left(\frac{\bar{X}-\bar{v}}{\bar{X}+\bar{v}}\right)\right] \tag{26}$$

Here, α is considered the optimal constant whose value is selected so that the mean squared error is minimized to its lowest attainable level.

Proposed Estimator-2

$$t_{Prop_2}^{\mathfrak{F}} = 2^{-1}\bar{u}\left\{\left(\frac{\bar{X}}{\bar{v}}\right)^{\epsilon} + \left(\frac{\bar{v}}{\bar{X}}\right)^{1-\epsilon}\right\}\left\{\exp\left(\frac{\bar{X}-\bar{v}}{\bar{X}+\bar{v}}\right)\right\} \tag{27}$$

Here, ϵ represents the minimizing constant, chosen in such a way that the mean squared error attains its minimum possible value.

Derivation of Bias and MSE of Proposed Estimator-1

The given equation (26) may be expressed in terms of its corresponding error term in order to better analyze the underlying error structure.

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[\left(\frac{\bar{X}(1+\xi_1)}{\bar{X}}\right)^{\alpha} + \exp\left[\frac{\bar{X}-\bar{X}(1+\xi_1)}{\bar{X}+\bar{X}(1+\xi_1)}\right]\right] \tag{28}$$

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[(1+\xi_1)^{\alpha} + \exp\left[\frac{-\bar{X}\xi_1}{2\bar{X}+\bar{X}\xi_1}\right]\right] \tag{29}$$

Upon expanding the terms in equation (29) and performing the required algebraic manipulations, we arrive at the following result.

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[(1+\xi_1)^{\alpha} + \exp\left\{-\frac{\xi_1}{2}\left(1+\frac{\xi_1}{2}\right)^{-1}\right\}\right] \tag{30}$$

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[\left(1+\alpha\xi_1+\frac{\alpha(\alpha-1)}{2}\right) + \exp\left[-\frac{\xi_1}{2}\left(1-\frac{\xi_1}{2}+\frac{1}{4}\xi_1^2\right)\right]\right] \tag{31}$$

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[\left(1+\alpha\xi_1+\frac{\alpha(\alpha-1)}{2}\right) + \exp\left[\left(-\frac{\xi_1}{2}+\frac{\xi_1^2}{4}\right)\right]\right] \tag{32}$$

By expanding and simplifying the right-hand side of equation (32), we obtain a more reduced and manageable form of the expression.

$$t_{Prop_1}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[\left(1+\alpha\xi_1+\frac{\alpha(\alpha-1)}{2}\right) + \left[1-\frac{\xi_1}{2}+\frac{3}{8}\xi_1^2\right]\right] \tag{33}$$

$$t_{Prop_1}^{\mathfrak{F}} = \bar{Y}(1+\xi_0)\left[1+\left(\frac{\alpha}{2}-\frac{1}{4}\right)\xi_1 + \left(\frac{\alpha(\alpha-1)}{2}+\frac{3}{16}\right)\xi_0\xi_1\right] \tag{34}$$

$$t_{Prop_1}^{\mathfrak{F}} = \bar{Y} + \bar{Y}\left[\left(\frac{\alpha}{2}-\frac{1}{4}\right)\xi_1 + \left(\frac{\alpha(\alpha-1)}{4}+\frac{3}{16}\right)\xi_1^2 + \xi_0 + \left(\frac{\alpha}{2}-\frac{1}{4}\right)\xi_0\xi_1\right] \tag{35}$$

$$\mathrm{Bias}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right) = \mathrm{E}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}} - \bar{Y}\right)$$

When the population mean $(\bar{Y})$ is subtracted from both sides of equation (35) and the expectation is taken, the bias expression is obtained in the following form.

$$\mathrm{E}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}} - \bar{Y}\right) = \bar{Y}\left[\left(\frac{\alpha}{2} - \frac{1}{4}\right)\mathrm{E}(\xi_1) + \left(\frac{\alpha(\alpha-1)}{4} + \frac{3}{16}\right)\mathrm{E}(\xi_1^2) + \left(\frac{\alpha}{2} - \frac{1}{4}\right)\mathrm{E}(\xi_0\xi_1)\right] \tag{36}$$

$$\mathrm{Bias}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right) = \lambda\bar{Y}\left[\left(\frac{\alpha}{2} - \frac{1}{4}\right)C_v^2 + \left(\frac{\alpha}{2} - \frac{1}{4}\right)\rho_{uv}C_uC_v\right] \tag{37}$$

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right) = \mathrm{E}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}} - \bar{Y}\right)^2$$

By considering the first-order error terms in equation (36), squaring them, and then taking their expectation, we obtain the expression for the mean squared error.

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right) = \bar{Y}^2\mathrm{E}\left[\left(\frac{\alpha}{2} - \frac{1}{4}\right)\xi_1 + \xi_0\right]^2 \tag{38}$$

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right) = \bar{Y}^2\lambda\left(C_u^2 + \left(\frac{\alpha}{2} - \frac{1}{4}\right)^2 C_v^2 + 2\left(\frac{\alpha}{2} - \frac{1}{4}\right)\rho_{uv}C_uC_v\right) \tag{39}$$

$$\alpha_{\mathrm{opt}}^* = \frac{1}{2} - \frac{2\rho C_u}{C_v}$$

After substituting the optimal value of g into Eq (39), the corresponding expression for the minimum mean squared error is obtained.

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right)_{\min} = \lambda\bar{Y}^2C_u^2(1 - \rho_{uv}^2) \tag{40}$$

Derivation of Bias and MSE of Proposed Estimator-2

Equation (27) can be rewritten in terms of the corresponding error term to provide a clearer representation of the associated variability.

$$t_{\mathrm{Prop}_2}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)\left[\left(\frac{\bar{X}}{\bar{X}(1+\xi_1)}\right)^{\epsilon} + \left(\frac{\bar{X}(1+\xi_1)}{\bar{X}}\right)^{1-\epsilon}\right]\left[\exp\left\{\frac{\bar{X}-\bar{X}(1+\xi_1)}{\bar{X}-\bar{X}(1+\xi_1)}\right\}\right] \tag{41}$$

$$t_{\mathrm{Prop}_2}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)[(1+\xi_1)^{-\epsilon} + (1+\xi_1)^{1-\epsilon}]\left[\exp\left(\frac{-\bar{X}\xi_1}{2\bar{X}+\bar{X}\xi_1}\right)\right] \tag{42}$$

$$t_{\mathrm{Prop}_2}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)[(1+\xi_1)^{-\epsilon} + (1+\xi_1)^{1-\epsilon}]\left[\exp\left\{-\frac{\xi_1}{2}\left(1+\frac{\xi_1}{2}\right)^{-1}\right\}\right] \tag{43}$$

After expanding the given equation (43) and carrying out the necessary simplifications, we obtain the following result.

$$t_{\mathrm{Prop}_2}^{\mathfrak{F}} = 2^{-1}\bar{Y}(1+\xi_0)[(2+\xi_1 - 2\epsilon\xi_1 + \epsilon^2\xi_1^2)]\left[1 - \frac{\xi_1}{2} + \frac{3}{8}\xi_1^2\right] \tag{44}$$

After carrying out the expansion and necessary simplification of the right-hand side of equation (), we arrive at a simplified version of the expression.

$$t_{\mathrm{Prop}_2}^{\mathfrak{F}} = \left[\bar{Y} + \bar{Y}\xi_0 - \bar{Y}\epsilon\xi_1 - \epsilon\bar{Y}\xi_0\xi_1 + \frac{1}{2}\left(\frac{1}{4} + \epsilon + \epsilon^2\right)\bar{Y}\xi_1^2\right] \tag{45}$$

$$\mathrm{Bias}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \mathrm{E}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}} - \bar{Y}\right)$$

By subtracting the population mean $(\bar{Y})$ from both sides of equation (45) and subsequently taking the expectation, we obtain the resulting expression for the bias.

$$t_{\mathrm{Prop}_2}^{\mathfrak{F}} - \bar{Y} = \left[\bar{Y} + \bar{Y}\mathrm{E}(\xi_0) - \bar{Y}\epsilon\mathrm{E}(\xi_1) - \epsilon\bar{Y}\mathrm{E}(\xi_0\xi_1) + \frac{1}{2}\left(\frac{1}{4} + \epsilon + \epsilon^2\right)\bar{Y}\mathrm{E}(\xi_1^2)\right] \tag{46}$$

$$\mathrm{Bias}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \lambda\bar{Y}\left[\frac{1}{2}\left(\frac{1}{4} + \epsilon + \epsilon^2\right)C_v^2 - \epsilon\rho_{uv}C_uC_v\right] \tag{47}$$

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \mathrm{E}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}} - \bar{Y}\right)^2$$

The mean squared error is obtained from equation (46) by taking the first-order error terms, squaring the expression, and then evaluating its expectation.

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \lambda\bar{Y}^2[C_u^2 + \epsilon^2C_v^2 - \epsilon\rho_{uv}C_uC_v] \tag{48}$$

Special cases of proposed generalized estimators:

To analyze three common situations involving the value of ϵ.

Case 1: $\epsilon = -1$

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \lambda\bar{Y}^2[C_u^2 + \epsilon^2 C_v^2 + \epsilon\rho_{uv}C_uC_v]$$

So, $\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \mathrm{MSE}\left(t_{3(pps)}\right)$
This is equal to the product estimator.
Case 2: $\epsilon = 0$

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \mathrm{Var}\left(t_{1(pps)}\right)$$

It leads to the sample variance associated with the sample mean.
Case 3: $\epsilon = 1$

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \lambda\bar{Y}^2[C_u^2 + \epsilon^2 C_v^2 - \epsilon\rho_{uv}C_uC_v]$$

This results in the mean squared error of the ratio estimator proposed by Murthy (1964), as cited in Single et al. 2020).

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right) = \mathrm{MSE}\left(t_{2(pps)}\right)$$

By taking the partial derivative of equation (e) with respect to ϵ and then equating the resulting expression to zero, the optimum value of ϵ can be determined.

$$\epsilon_{\mathrm{opt}}^{*} = \frac{C_u\rho_{uv}}{C_v}$$

After substituting the obtained value into equation ($\epsilon_{\mathrm{opt}}^{*}$), we derive the expression for the mean squared error as follows.

$$\mathrm{MSE}\left(t_{\mathrm{Prop}_2}^{\mathfrak{F}}\right)_{\min} = \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) \tag{49}$$

5. Efficiency Comparisons

This section presents a comparative analysis of efficiency across the evaluated methods. The goal is to identify performance differences and highlight which approach offers superior resource utilization under given conditions.

- From Eq (1) and Eq (40), we get,

$$\mathrm{Var}\left(t_{1(pps)}\right) - \mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right)_{\min} > 0,$$

Or,

$$\lambda\bar{Y}^2C_u^2 - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$$

- From Eq (3) and Eq (40), we get,

$$\mathrm{MSE}\left(t_{2(pps)}\right) - \mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right)_{\min} > 0,$$

Or

$$,\bar{Y}^2\lambda[C_u^2 + C_v^2 - 2\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$$

- From Eq (5) and Eq (40), we get,

$$\mathrm{MSE}\left(t_{3(pps)}\right) - \mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right)_{\min} > 0,$$

Or,

$$\bar{Y}^2\lambda[C_u^2 + C_v^2 + 2\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$$

- From Eq (7) and Eq (40), we get,

$$\mathrm{MSE}\left(t_{4(pps)}\right) - \mathrm{MSE}\left(t_{\mathrm{Prop}_1}^{\mathfrak{F}}\right)_{\min} > 0,$$

Or,

$$\bar{Y}^2\lambda\left(C_u^2 + \frac{1}{4}C_v^2 - \rho_{uv}C_uC_v\right) - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$$

- From Eq (9) and Eq (40), we get,

$\mathrm{MSE}\left(t_{5(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda\left(C_u^2 + \frac{1}{4}C_v^2 + \rho_{uv}C_uC_v\right) - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (11) and Eq (40), we get,

$\mathrm{MSE}\left(t_{6(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_1^2C_v^2 - 2\varrho_1\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (13) and Eq (40), we get,

$\mathrm{MSE}\left(t_{7(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_2^2C_v^2 - 2\varrho_2\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (15) and Eq (40), we get,

$\mathrm{MSE}\left(t_{8(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_3^2C_v^2 - 2\varrho_3\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (17) and Eq (40), we get,

$\mathrm{MSE}\left(t_{9(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_4^2C_v^2 - 2\varrho_4\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (19) and Eq (40), we get,

$\mathrm{MSE}\left(t_{10(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_5^2C_v^2 - 2\varrho_5\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (21) and Eq (40), we get,

$\mathrm{MSE}\left(t_{11(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_6^2C_v^2 - 2\varrho_6\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (23) and Eq (40), we get,

$\mathrm{MSE}\left(t_{12(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{Prop_1}\right)_{min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_7^2C_v^2 - 2\varrho_7\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

- From Eq (25) and Eq (40), we get,

$\mathrm{MSE}\left(t_{13(pps)}\right) - \mathrm{MSE}\left(t^{\mathfrak{F}}_{\mathrm{Prop}_1}\right)_{\min} > 0,$

Or,

$\bar{Y}^2\lambda[C_u^2 + \varrho_8^2 C_v^2 - 2\varrho_8\rho_{uv}C_uC_v] - \lambda\bar{Y}^2C_u^2(1-\rho_{uv}^2) > 0.$

The study focuses on proposed estimator-1 in the comparison, given that both estimators possess identical MSE expressions and demonstrate improved performance over the existing estimators.

6. Numerical Study

a) Real Data Study

In this section, a comprehensive numerical study is carried out using several real-life data sets to evaluate and compare the performance of the proposed estimators with existing estimators available in the literature. The main objective of this empirical investigation is to examine the practical usefulness and efficiency gains of the proposed methods under realistic conditions. The performance of each estimator is assessed using two widely accepted criteria, namely Percentage Relative Efficiency (PRE) and Mean Squared Error (MSE), which provide a clear measure of both accuracy and precision. A brief description of each data set, along with its essential summary statistics, is presented in Table 1 to facilitate a better understanding of the underlying structure of the data used in the study.
Y= No. of IDPs in USA in1996
X= No. of IDPs in the USA in 1995
Population-2 [Source: Singh (2003)]
Y= No of beds allocated for COVID patients
X= Beds allocated for COVID
Z= Beds used by patients
Population-3 [Source Bahl and Tuteja (1991)]
Y= Agricultural maize in 1964
X= Agricultural maize in 1963

R_{X1} = rank of X_1

Table 1: Statistical Summary of the Three Populations (1–3)

Parameters	Population-1	Population-2	Population-3
N	51	36	34
n	12	8	8
λ	0.08333333	0.125	0.125
$\bar{Y}$	13903.24	660.1389	199.4412
$\bar{X}$	15464.06	215.6389	208.8824
$\bar{u}$	61284.25	3645.545	817.6256
$\bar{v}$	64373.61	1308.779	864.0967
C_u	0.02931779	0.481517	0.07564295
C_v	0.04196459	0.4155077	0.07429192
ρ_{uv}	0.735276	0.1726467	0.9018024
S_u	1796.718	317.8681	61.84761
S_v	2571.768	89.59963	60.74297
C_u^2	0.0008595	0.2318587	0.005721856
C_v^2	0.001761027	0.1726467	0.005519289

Percentage Relative Efficiency (PRE) is a measure used in statistics to compare the efficiency of two estimators, usually by expressing how much better one estimator performs compared to another in terms of variance or mean squared error (MSE).
It is commonly defined as:

$$\mathrm{PRE} = \frac{\mathrm{Var}\left(t_{1(pps)}\right)}{\mathrm{MSE}(t_i)} \times 100$$

$t_i = t_{1(pps)}, t_{2(pps)}, t_{3(pps)}, \ldots t_{13(pps)}$

Table 2: Analysis of MSE and PRE Based on the Application of Real Populations

Estimator	Population1		Population-2		Population-3	
	MSE	PRE	MSE	PRE	MSE	PRE
$t_{1(pps)}$	12530.02	100.00	13845.65	100.00	28.44963	100.00
$t_{2(pps)}$	10751.70	116.51	13069.12	105.94	5.496623	517.50
$t_{3(pps)}$	32217.47	38.90	71326.44	19.41	105.2865	27.03
$t_{4(pps)}$	9389.713	133.46	6357.621	217.87	10.11263	281.20
$t_{5(pps)}$	20672.60	60.62	35506.6	38.98	60.40695	47.10
$t_{6(pps)}$	11174.89	112.19	10560.76	131.11	18.32734	155.20
$t_{7(pps)}$	9607.107	130.47	12671.49	109.26	24.85612	114.47
$t_{8(pps)}$	9383.661	112.19	7774.261	178.08	26.44176	107.58
$t_{9(pps)}$	12661.87	130.47	8063.307	171.70	20.52987	138.50
$t_{10(pps)}$	10934.16	133.56	11507.23	120.30	28.08695	101.29

$t_{11(pps)}$	12176.43	114.61	8597.433	161.07	26.23298	108.47
$t_{12(pps)}$	12130.42	102.90	8334.263	166.16	26.61554	106.93
$t_{13(pps)}$	12216.52	103.28	7821.631	177.03	4.854178	4586.20
$t^{\mathfrak{F}}_{Prop_1}$	9305.647	134.66	6360.265	217.68	5.313036	535.50
$t^{\mathfrak{F}}_{Prop_2}$	9306.447	134.65	6360.244	217.64	5.31326	535.50

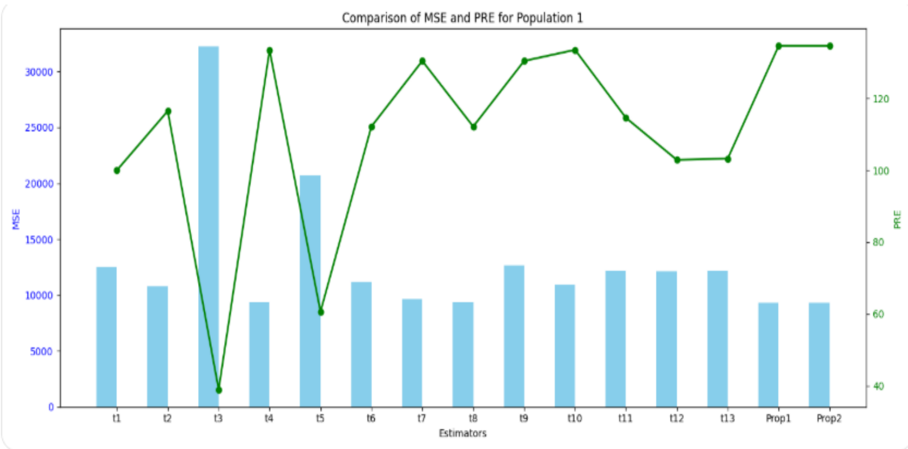

Fig. 1: MSE and PRE of Estimators Using Population 1.

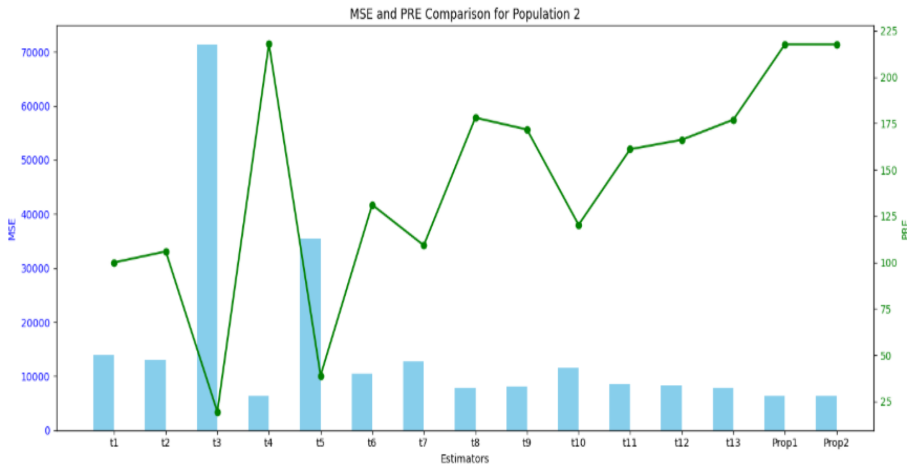

Fig. 2: MSE and PRE of Estimators Using Population 2

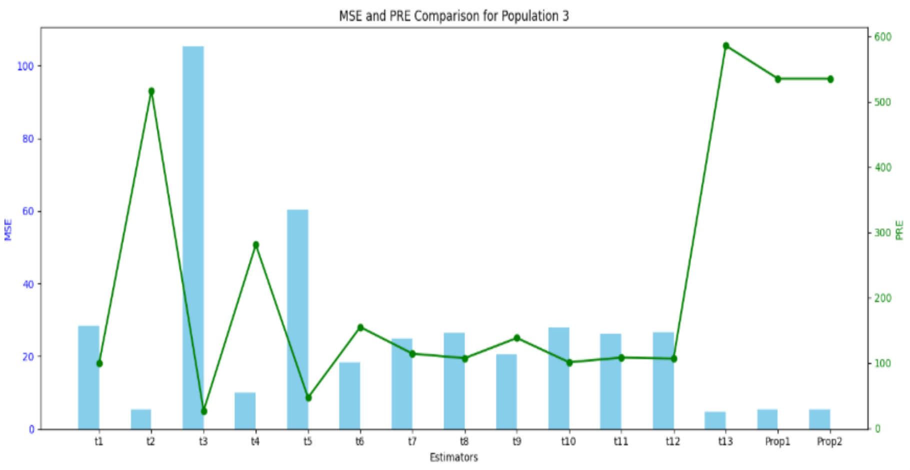

Fig. 3: MSE and PRE of Estimators Using Population 3.

b) Simulation study

This section, taking motivation from Mustafa (2024), presents a simulation study conducted to evaluate the performance of the proposed estimators under controlled conditions. The study assumes that the population follows a multivariate normal distribution, with known mean vectors and covariance matrices. To ensure reliability and accuracy of the results, a total of 5000 iterations are performed. In each iteration, random samples are drawn from the specified multivariate normal population. The estimators are then computed for each sample, and their corresponding Mean Squared Error (MSE) and Percentage Relative Efficiency (PRE) values are evaluated. The results are averaged over all iterations to obtain stable estimates of performance measures. This simulation framework allows for a comprehensive assessment of the estimators under repeated sampling conditions.

The following populations are considered in the simulation study algorithm to examine the performance of the proposed estimators. $\mathrm{MSE}\left(t^{\mathfrak{F}}_{\mathrm{Prop}_1}\right)_{\min}$ and $\mathrm{MSE}\left(t^{\mathfrak{F}}_{\mathrm{Prop}_2}\right)_{\min}$ over the existing estimators t_{ipps} where, $i = 1,2,3, \ldots 13$.

Population-I

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

And

$$\Sigma = \begin{bmatrix} 1000 & 800 & 810 \\ 800 & 850 & 820 \\ 810 & 820 & 840 \end{bmatrix}$$

Population-II

$$\mu_2 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix}$$

And

$$\Sigma = \begin{bmatrix} 1200 & 870 & 820 \\ 870 & 900 & 800 \\ 820 & 800 & 740 \end{bmatrix}$$

Population-III

$$\mu_3 = \begin{bmatrix} 500 \\ 500 \\ 500 \end{bmatrix}$$

And

$$\Sigma = \begin{bmatrix} 400 & 270 & 220 \\ 270 & 500 & 300 \\ 220 & 300 & 200 \end{bmatrix}$$

1) The formula representing the mean squared error (MSE) of the estimator is given as follows.

$\mathrm{MSE}(\bar{y}_i) = \frac{1}{k}\sum_{j=1}^{k}\left(y_{ij} - \bar{y}\right)$; where k =5000

2) The formula representing the percentage relative efficiency (PRE) of the estimator is given as follows.

$\mathrm{PRE}(i) = \frac{\mathrm{var}\left(t_{1(\mathrm{pps})}\right)}{\mathrm{MSE}\left(t_{i(\mathrm{pps})}\right)} \times 100$, where, $\left(i = 1,2,3, \ldots, 13, \mathrm{MSE}\left(t^{\mathfrak{F}}_{\mathrm{Prop}_1}\right)_{\min}, \mathrm{MSE}\left(t^{\mathfrak{F}}_{\mathrm{Prop}_2}\right)_{\min}\right)$

Table 2: Analysis of MSE And PRE Based on the Application of Simulated Populations

	Population1		Population-2		Population-3	
Estimator	**MSE**	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
$t_{1(\mathrm{pps})}$	0.4823719	100	0.7099683	100	0.705276	100
$t_{2(\mathrm{pps})}$	0.10172353	484.0421	0.08037702	884.1764	0.1887632	367.1186
$t_{3(\mathrm{pps})}$	1.756176	27.06687	2.62583	28.09745	1.865729	38.34789
$t_{4(\mathrm{pps})}$	0.1006472	477.7854	0.07984346	887.9876	0.1572187	421.8976
$t_{5(\mathrm{pps})}$	0.18821907	260.7841	0.2467754	285.7051	0.357869	185.0303
$t_{6(\mathrm{pps})}$	1.901455	48.52763	0.2467755	48.29767	1.26788	69.07064
$t_{7(\mathrm{pps})}$	0.3210123	153.3571	1.456889	163.9876	0.4727098	155.0888
$t_{8(\mathrm{pps})}$	0.468445	105.125	0.4342843	108.0765	0.587000	106.9897
$t_{9(\mathrm{pps})}$	0.1019016	482.3087	0.6684345	771.9876	0.2273451	312.6268
$t_{10(\mathrm{pps})}$	0.417834	117.2609	0.086333	118.8051	0.6360145	112.6654
$t_{11(\mathrm{pps})}$	0.2645915	187.087	0.604161	193.8765	0.704722	151.5044
$t_{12(\mathrm{pps})}$	0.4866002	102.1492	0.369586	101.876	0.628896	100.5899
$t_{13(\mathrm{pps})}$	0.4139208	118.2231	0.7002404	118.8976	0.641878	114.8214
$t^{\mathfrak{F}}_{\mathrm{Prop}_1}$	0.4246351	572.7890	0.001435	2738.635	0.0552775	1276.0987
$t^{\mathfrak{F}}_{\mathrm{Prop}_2}$	0.4256352	572.6280	0.001433	2738.435	0.0552772	1276.0986

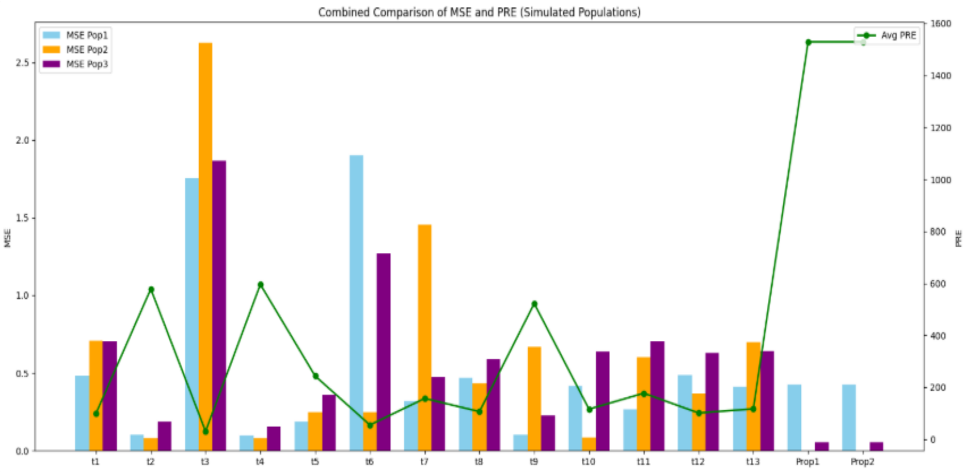

Fig. 4: MSE and PRE Using Simulation Populations.

7. Results and Discussion

- Table 1 presents a comparative analysis of Mean Squared Error (MSE) and Percentage Relative Efficiency (PRE) for various estimators across three populations. For Population 1, the estimator $t_{3(pps)}$ shows the highest MSE (32217.47) and lowest PRE (38.90), indicating poor performance, while the proposed estimators $t^{\mathfrak{F}}_{Prop_1}$ and $t^{\mathfrak{F}}_{Prop_2}$ achieve the lowest MSE (≈9305) and highest PRE (≈134.66), demonstrating superior efficiency. In Population 2, a similar pattern is observed, where $t_{3(pps)}$ again performs worst with the highest MSE (71326.44) and lowest PRE (19.41), whereas the proposed estimators and show the best performance with minimum MSE (≈6360) and maximum PRE (≈217.68). For Population 3, the estimator $t_{13(pps)}$ stands out with the lowest MSE (4.854178) and extremely high PRE (4586.20), followed by the proposed estimators with low MSE (~5.31) and high PRE (~535.50). Overall, the table clearly demonstrates an inverse relationship between MSE and PRE, confirming that estimators with lower error exhibit higher efficiency.
- Figure 1 jointly compares MSE and PRE for various estimators in Population 1, revealing substantial differences in performance. The MSE ranges from a maximum of 32217.47 (for $t_{3(pps)}$ to a minimum of around 9305.647-9306.447 for the proposed estimators, indicating improved accuracy for the latter. Similarly, PRE values vary from a low of 38.90 for $t_{3(pps)}$ respectively. These results clearly demonstrate that the proposed estimators. $t^{\mathfrak{F}}_{Prop_1}$ and $t^{\mathfrak{F}}_{Prop_2}$ with lower MSE achieve higher efficiency, with the proposed estimators outperforming all others.
- Figure 2 presents a dual-axis visualization of MSE and PRE for various estimators under Population 2, where MSE is displayed as bar charts and PRE as a line graph, enabling a joint evaluation of estimation accuracy and efficiency. The estimator $t_{3(pps)}$ shows the highest MSE value of 71326.44, indicating the poorest performance in terms of accuracy, enabling a joint evaluation of estimation accuracy and efficiency $t^{\mathfrak{F}}_{Prop_1}$ and $t^{\mathfrak{F}}_{Prop_2}$ achieve the lowest MSE values, approximately 6360.265 and 6360.244, respectively, indicating strong precision. Similarly, PRE results indicate that $t_{4(pps)}$ have the highest efficiency, whereas $t_{13(pps)}$ records the lowest value of 19.41, reflecting weak performance. In summary, the figure confirms a strong inverse relationship between MSE and PRE and highlights the superior performance of the proposed estimators in Population 2.
- Figure 3 shows a dual-axis comparison of MSE and PRE for estimators under Population 3. The MSE values are displayed as bars, while PRE is illustrated using a line plot, allowing simultaneous assessment of accuracy and efficiency. The estimator $t_{3(pps)}$ has the highest MSE (105.2865) and a low PRE (27.03), indicating poor performance. In contrast $t_{13(pps)}$ attains the lowest MSE (4.854178) and the highest PRE (586.20), making it the best-performing estimator. The proposed estimators $t^{\mathfrak{F}}_{Prop_1}$ an $t^{\mathfrak{F}}_{Prop_2}$ also demonstrate strong performance with low MSE (~5.31) and high PRE (~535.50). Overall, the figure confirms an inverse relationship between MSE and PRE.
- The simulation Table 2 compares estimators using MSE and PRE across three populations. A clear inverse relationship is observed: lower MSE leads to higher PRE. The proposed estimators $t^{\mathfrak{F}}_{Prop_1}$ and $t^{\mathfrak{F}}_{Prop_2}$ consistently show the lowest MSE and highest PRE, especially in Populations 2 and 3, indicating superior performance. In contrast, $t_{3(pps)}$ performs worst with high MSE and low PRE, while a few estimators show moderate efficiency improvements.
- This figure illustrates a combined comparison of Mean Squared Error (MSE) and Percentage Relative Efficiency (PRE) for all estimators across three simulated populations. The MSE values are displayed using grouped bar charts, enabling a clear comparison of estimation accuracy among the populations, while PRE is represented through an average line plot to indicate overall efficiency. It is observed that the proposed estimators consistently produce very low MSE values across all populations, especially in Population 2 and Population 3, where the errors are minimal. In parallel, these estimators attain very high PRE values, demonstrating superior efficiency.

8. Final Thoughts and Suggestions

In conclusion, this study makes a significant contribution to the field of survey sampling by developing a robust and generalized class of exponential estimators for population mean estimation under the Probability Proportional to Size (PPS) sampling framework. The proposed estimators effectively utilize auxiliary information to enhance estimation accuracy, particularly in heterogeneous populations where traditional estimators often perform inadequately. Through rigorous theoretical derivations, the bias and mean squared error (MSE) expressions were obtained and optimized, demonstrating that the proposed estimators achieve minimum MSE under optimal conditions. The empirical validation, based on both real-world datasets and extensive simulation studies, consistently confirms the superiority of the proposed estimators over conventional estimators in terms of lower MSE and higher percentage relative efficiency (PRE). The graphical and numerical comparisons further reinforce these findings, highlighting the stability and reliability of the proposed approach across varying population structures. Moreover, the study emphasizes the practical relevance of the methodology in diverse application domains such as radiation science, healthcare, agriculture, environmental monitoring, and economic forecasting, where accurate estimation is crucial for informed decision-making. Future research should focus on extending the proposed framework to more advanced sampling designs, such as multistage sampling, stratified PPS techniques, and adaptive sampling approaches. Moreover, addressing practical issues like non-response, measurement errors, and missing observations would significantly improve the robustness of the estimators. Incorporating modern computational tools, including machine learning methods and data-driven optimization strategies, may further enhance estimation accuracy, particularly in large-scale and real-time data settings.

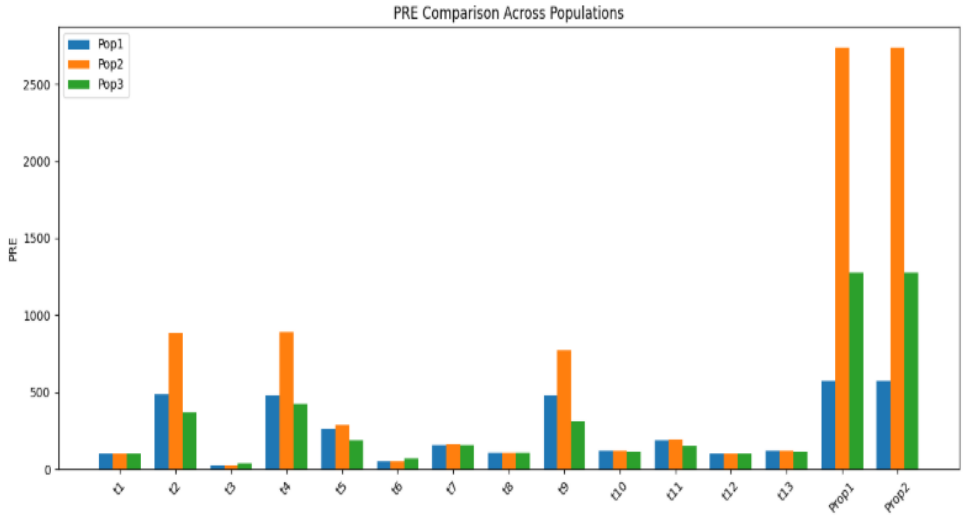

Fig. 5: PRE Comparison across All the Simulated Data for Population-1, 2 & 3.

References

[1] S. Ahmad, N. K. Adichwal, M. Aamir, J. Shabbir, N. Alsadat, M. Elgarhy, and H. Ahmad, “An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling,” *Scientific Reports*, vol. 13, no. 1, p. 21444, 2023. https://doi.org/10.1038/s41598-023-44169-5.

[2] S. Ahmad and J. Shabbir, “Use of extreme values to estimate finite population means under PPS sampling scheme,” *Journal of Reliability and Statistical Studies*, pp. 99–112, 2018.

[3] S. Ahmad and J. Shabbir, “Proportion estimation using enhanced class of estimators under simple random sampling: Application with real data sets and simulation,” *Measurement: Interdisciplinary Research and Perspectives*, Advance online publication, 2024. https://doi.org/10.1080/15366367.2024.2346426.

[4] S. Ahmad, S. Hussain, M. Aamir, F. Khan, M. N. Alshahrani, and M. Alqawba, “Estimation of finite population mean using dual auxiliary variable for non-response using simple random sampling,” *AIMS Mathematics*, vol. 7, no. 3, pp. 4592–4613, 2022. https://doi.org/10.3934/math.2022256.

[5] S. Ahmad, S. Hussain, M. Aamir, U. Yasmeen, J. Shabbir, and Z. Ahmad, “Dual use of auxiliary information for estimating the finite population means under the stratified random sampling scheme,” *Journal of Mathematics*, vol. 2021, Art. no. 3860122, 2021. https://doi.org/10.1155/2021/3860122.

[6] S. Ahmad, S. Hussain, K. Ullah, E. Zahid, M. Aamir, J. Shabbir, and W. Alajlan, “A simulation study: Improved ratio-in-regression type variance estimator based on dual use of auxiliary variable under simple random sampling,” *PLoS ONE*, vol. 17, no. 11, Art. no. e0276540, 2022. https://doi.org/10.1371/journal.pone.0276540.

[7] S. Ahmad, S. Hussain, and S. Ahmad, “Finite population distribution function estimation using auxiliary information under simple random sampling,” *Statistics, Computing and Interdisciplinary Research*, vol. 3, no. 1, pp. 29–38, 2021. https://doi.org/10.52700/scir.v3i1.25.

[8] S. Ahmad, J. Shabbir, M. Aamir, S. Hussain, and K. Ullah, “Improved mean estimation for a finite population with twofold auxiliary variables under simple random sampling,” *International Journal of Applied and Computational Mathematics*, vol. 10, no. 1, Art. no. 31, 2024. https://doi.org/10.1007/s40819-023-01624-1.

[9] S. Ahmad, J. Shabbir, E. Zahid, and M. Aamir, “Improved family of estimators for the population mean using supplementary variables under PPS sampling,” *Science Progress*, vol. 106, no. 2, Art. no. 00368504231180085, 2023. https://doi.org/10.1177/00368504231180085.

[10] S. Ahmad, E. Zahid, J. Shabbir, M. Aamir, and R. Onyango, “Enhanced estimation of the population mean using two auxiliary variables under probability proportional to size sampling,” *Mathematical Problems in Engineering*, vol. 2023, Art. ID 1–??, 2023. https://doi.org/10.1155/2023/5564360.

[11] A. A. Al-Omari, A. A. Jemain, and K. Ibrahim, “New ratio estimators of the mean using simple random sampling and ranked set sampling methods,” *Investigación Operacional*, vol. 30, pp. 97–108, 2009.

[12] S. Bahl and R. Tuteja, “Ratio and product type exponential estimators,” *Journal of Information and Optimization Sciences*, vol. 12, no. 1, pp. 159–164, 1991. https://doi.org/10.1080/02522667.1991.10699058.

[13] R. P. Chakrabarty, “Some ratio-type estimators,” *Journal of the Indian Society of Agricultural Statistics*, vol. 25, pp. 49–57, 1979.

[14] R. Gupta, M. Ali, S. K. Yadav, G. Kumar, and S. Kumar, “Improved estimation of the population mean by integrating additional auxiliary parameters,” *Korean Journal of Physiology and Pharmacology*, vol. 28, no. 1, pp. 205–210, 2024.

[15] S. Hussain, S. Ahmad, S. Akhtar, A. Javed, and U. Yasmeen, “Estimation of finite population distribution function with dual use of auxiliary information under non-response,” *PLoS ONE*, vol. 15, no. 12, Art. no. e0243584, 2020. https://doi.org/10.1371/journal.pone.0243584.

[16] S. Hussain, S. Ahmad, M. Saleem, and S. Akhtar, “Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling,” *PLoS ONE*, vol. 15, no. 9, Art. no. e0239098, 2020. https://doi.org/10.1371/journal.pone.0239098.

[17] S. Hussain, M. Zichuan, S. Hussain, A. Iftikhar, M. Asif, S. Akhtar, and S. Ahmad, “On estimation of distribution function using dual auxiliary information under nonresponse using simple random sampling,” *Journal of Probability and Statistics*, vol. 2020, Art. no. 1693612, 2020. https://doi.org/10.1155/2020/1693612.

[18] A. Iftikhar, S. Hongbo, S. Hussain, S. Khan, B. Alnssyan, J. Shabbir, and M. Aamir, "A new exponential factor-type estimator for population distribution function using dual auxiliary variables under stratified random sampling," *Mathematical Problems in Engineering*, vol. 2022, Art. no. 7673945, 2022. https://doi.org/10.1155/2022/7673945.
[19] K. Iqbal, S. M. Raza, M. M. Butt, H. Ahmad, and S. Askar, "On exploring the generalized mixture estimators under simple random sampling and application in health and finance sector," *AIP Advances*, vol. 14, no. 1, 2024. https://doi.org/10.1063/5.0176945.
[20] C. Kadilar and H. Cingi, "An improvement in estimating the population mean by using the correlation coefficient," *Quality Control and Applied Statistics*, vol. 52, no. 4, pp. 443–446, 2007.
[21] M. E. Kanwai, Y. Yakubu, A. Isah, and A. Usman, "Exponential type estimator for estimating finite population mean with auxiliary variables under simple random sampling," *Science World Journal*, vol. 19, no. 3, pp. 701–708, 2024, https://doi.org/10.4314/swj.v19i3.17.
[22] A. Kumar and A. S. Siddiqui, "Enhanced estimation of population mean using simple random sampling," *Research in Statistics*, vol. 2, no. 1, Art. no. 2335949, 2024. https://doi.org/10.1080/27684520.2024.2335949.
[23] A. Kumar, S. Bhushan, M. S. Mustafa, R. Aldallal, H. M. Aljohani, and F. A. Almulhim, "Novel imputation methods under stratified simple random sampling," *Alexandria Engineering Journal*, vol. 95, pp. 236–246, 2024. https://doi.org/10.1016/j.aej.2024.03.088.
[24] A. Kumar, A. S. Siddiqui, M. S. Mustafa, E. Hussam, H. M. Aljohani, and F. A. Almulhim, "Mean estimation using an efficient class of estimators based on simple random sampling: Simulation and applications," *Alexandria Engineering Journal*, vol. 91, pp. 197–203, 2024. https://doi.org/10.1016/j.aej.2024.02.003.
[25] N. Lawson, "An improved class of population mean estimators by utilizing some prior information in simple random sampling using Searl's approach," *Lobachevskii Journal of Mathematics*, vol. 43, no. 11, pp. 3376–3383, 2022. https://doi.org/10.1134/S1995080222140232.
[26] R. Latpate, J. Kshirsagar, V. K. Gupta, G. Chandra, et al., "Probability proportional to size sampling," in *Advanced Sampling Methods*, pp. 85–98, 2021. https://doi.org/10.1007/978-981-16-0622-9_7.
[27] M. N. Murthy, *Sampling Theory and Methods*. Statistical Publishing Society, 1967.
[28] M. S. A. Mustafa, S. Ahmad, E. Zahid, J. Shabbir, and S. Masood, "Novel methods for estimation of population mean using auxiliary information under PPS sampling: Application with real and simulated data sets," *Kurdish Studies*, vol. 12, no. 5, pp. 1553–1562, 2024, https://doi.org/10.53555/ks.v12i5.3508.
[29] Y. Mojdeed, A. Ahmed, I. O. Olatunji, and B. O. Daud, "Efficient exponential type estimator for estimating finite population mean under simple random sampling," *Anale. Seria Informatică*, vol. 19, no. 1, pp. 46–52, 2021.
[30] O. M. Olayiwolla, F. S. Apantaku, and O. A. Wale-Orojo, "Probability proportional to size (PPS) method to enhance efficiency of estimator in two-stage sampling," *Annals of Computer Science Series*, vol. 17, pp. 311–315, 2019.
[31] Punjab Bureau of Statistics, *Annual Report*, 2021–2022.
[32] T. A. Raja and S. Maqbool, "Ratio type estimators for estimating population mean using some auxiliary information," *International Journal of Agriculture Sciences*, 2024.
[33] H. E. Semary, S. Ahmad, W. A. Hamdi, O. Albalawi, I. Elbatal, C. Chesneau, and S. M. Almarzouki, "Improved novel estimation for estimation of population distribution function using auxiliary information under stratified sampling strategy," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 3, Art. no. 101011, 2024. https://doi.org/10.1016/j.jrras.2024.101011.
[34] H. E. Semary, S. Ahmad, I. Elbatal, C. Chesneau, M. Elgarhy, and E. M. Almetwally, "New comprehensive class of estimators for population proportion using auxiliary attribute: Simulation and an application," *Alexandria Engineering Journal*, 2024. https://doi.org/10.1016/j.aej.2024.04.065.
[35] H. P. Singh, A. C. Mishra, and S. K. Pal, "Improved estimator of population total in PPS sampling," *Communications in Statistics—Theory and Methods*, vol. 47, no. 4, pp. 912–934, 2018. https://doi.org/10.1080/03610926.2017.1313987.
[36] S. Singh, *Advanced Sampling Theory with Applications*, vol. 2. Springer, 2003. https://doi.org/10.1007/978-94-007-0789-4.
[37] R. R. Sinha and B. Khanna, "Estimation of population mean under probability proportional to size sampling with and without measurement errors," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 18, Art. no. e7023, 2022. https://doi.org/10.1002/cpe.7023.
[38] B. V. S. Sisodia and V. K. Dwivedi, "Modified ratio estimator using coefficient of variation of auxiliary variable," *Journal of the Indian Society of Agricultural Statistics*, vol. 33, no. 2, 1981.
[39] J. Subramani and G. Kumarapandiyan, "Modified ratio estimators for population mean using function of quartiles of auxiliary variable," *Bonfring International Journal of Industrial Engineering and Management Science*, vol. 2, no. 2, pp. 19–23, 2012. https://doi.org/10.9756/BIJIEMS.1356.
[40] J. Wang, S. Ahmad, M. Arslan, S. A. Lone, A. H. Abd Ellah, M. A. Aldahlan, and M. Elgarhy, "Estimation of finite population mean using double sampling under PPS sampling in the presence of extreme values," *Heliyon*, vol. 9, no. 11, 2023. https://doi.org/10.1016/j.heliyon.2023.e21418.
[41] T. Zaman, "New family of estimators using two auxiliary attributes," *International Journal of Advanced Research in Engineering & Management*, vol. 4, no. 11, pp. 11–16, 2018.
[42] T. Zaman and E. Toksoy, "Improvement in estimating the population mean using two auxiliary attributes and numerical application in agricultural engineering," *Fresenius Environmental Bulletin*, vol. 28, no. 6, pp. 4584–4590, 2019.

Dedicated to Our Mother