Eur. J. Stat. 4 (2024) 13
doi: 10.28924/ada/stat.4.13

# An Effective Method for Estimating the Population Mean That Utilizes Dual Auxiliary Information

Muhammad Asim Masood[1], Tarushree Bari[2], Rabbia Mukhtar[1], Nasir Ali[1], Abid Hussain[3,*]

[1]*Department of Statistics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan*
*asimmasood2746@gmail.com, rabbia.mukhtarbaig@gmail.com, nasir_stat@uaar.edu.pk*
[2]*Directorate of Online Education, Manipal Academy of Higher Education, Karnataka, India*
*drtarustats@gmail.com*
[3]*Department of Statistics, Govt. College Khayaban-e-Sir Syed, Rawalpindi, Pakistan*
*abid0100@gmail.com*
[*]*Correspondence: abid0100@gmail.com*

Abstract. To improve the effectiveness of population estimators, researchers have recently implemented dual supplementary information. They employed traditional rankings, the empirical cumulative distribution function, and indicator functions as supplementary sources of information in their analysis. An improved family of population mean estimators is introduced in this article, which utilizes the relative ranks of the auxiliary information's configurations to incorporate the relevant information. A first-order approximation is employed to derive the mathematical expressions for the bias and the mean-squared error (MSE) of the proposed family of estimators. The empirical analysis is investigated to demonstrate the practicality of the proposed estimators in real-world scenarios. Additionally, the theoretical conclusions are effectively validated by the Monte Carlo simulation integration. Our results unequivocally indicate that the proposed family of estimators surpasses their current counterparts.

## 1. Introduction

The primary goal of sample survey theory is to determine the values of population parameters that are presently unknown, such as the mean, proportion, and variance of the study variable. In order to generate a dependable estimation of the parameter of interest by analyzing a scrupulously selected sample of individuals, a precise and efficient methodology is necessary. The efficacy of the estimators is enhanced by incorporating the information of auxiliary variables that are associated with the study variable. The integration of supplementary data to enhance various techniques that were previously used to estimate the attributes of a population under study has a long and diverse

Received: 29 Sep 2024.

*Key words and phrases.* Estimation; auxiliary information; ranks; relative-ranks; simulation; comparison.

history in a variety of academic fields. For additional information regarding this subject, please consult to [1–8].

It is evident from a comprehensive examination of the literature that Pierre-Simon Laplace, a prominent figure in the 18th century, played a substantial role in advocating for the use of supplementary information to assist in the estimation of the parameters of study variables. For instance, in order to improve the precision of the fundamental arithmetic of France's total population during the 18th century, he proposed the utilization of vital birth records, which guarantee the well-being of individuals, as a method for determining the heads count of a vast empire without counting the individuals within it. As illustrated in [9], it is crucial to have an understanding of the population-to-annual-births ratio in order to determine this. The fundamental concepts that establish the mathematical validity of employing supplementary information to estimate the properties of the variable(s) under investigation were disclosed by Cochran's seminal work. The concept of estimating the mean of a small group by utilizing the inherent relationship between the main study variable and an additional variable was first introduced by Cochran [10].

In recent years, there have been significant improvements in the development of rank-based estimators, which provide a novel method for incorporating supplementary information. In situations where the dual use of the auxiliary variable provides valuable information about the research variable, Haq et al.[11] developed estimators that are superior to the aforementioned methods. These estimators are based on the rank of a significant auxiliary variable. They have found that these estimators exhibit resilient characteristics when assessed against both theoretical models and real-world scenarios as a result of their research.

This article introduces a novel class of estimators that are designed to estimate the mean of a finite target population within the design of simple random sampling. In order to enhance the performance of the proposed family of estimators, we implemented the relative ranks as an auxiliary variable in conjunction with auxiliary information. This relative rank approach, which considers the distance between data points, was recently proposed by Hussain et al.[12]. In accordance with estimation theory, the performance of the proposed estimators is improved by a strong correlation between the subject and auxiliary variables. By employing the auxiliary variable and its relative ranking, we have achieved a better estimation of the population mean. The theoretical calculation of the bias and MSE of the suggested estimators is performed using a first-order approximation. The MSE of the proposed estimators is compared to those of their recently proposed competitors. The comparative studies indicated that the proposed estimators consistently outperformed all other estimators, as assessed in both theoretical and numerical evaluations.

The current article is summarized as follows. The conventional and contemporary methods for determining the mean of a finite population are the primary focus of Section 2. Section 3 presents a more precise class of estimators for estimating the mean of a finite population. Section 4 analyzes theoretical comparisons between the proposed and existing estimators. The numerical investigation is presented in Section 5 to analyze and evaluate the performance of the estimators under our consideration. Section 6 contains a concise summary of our findings.

## 2. Existing estimators

This section examines numerous estimators of the finite population mean that are frequently employed in the current literature of survey sampling. We focused on the variances and MSEs of the estimators that were obtained from the first degree of approximation in order to offer a succinct explanation. The conventional method employs $\bar{y}$ (where y is our study variable) with a variance as follows to estimate the unbiased mean per unit

$$Var(\bar{y}) = \lambda \bar{Y}^2 C_y^2. \tag{1}$$

The conventional difference estimator $\hat{\bar{Y}}_D$ is

$$\hat{\bar{Y}}_D = \bar{y} + k(\bar{X} - \bar{x}),$$

where $k$ is unknown constant and $x$ is our auxiliary variable. Demonstrating the unbiasedness of $\hat{\bar{Y}}_D$ is a straightforward task. The minimum variance of the estimator $\hat{\bar{Y}}_D$ at the optimal value of $k$, denoted as $k_{\text{opt}}$, which is equal to $\rho_{yx}(S_y/S_x)$, is expressed as

$$Var_{min}(\hat{\bar{Y}}_D) = \lambda \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2). \tag{2}$$

The difference estimator $\hat{\bar{Y}}_D$ was found to be more efficient than the ratio and product estimators when comparing the efficacy of estimators for estimating $\bar{Y}$, see for instance, [13]. Additionally, the author recommended the upgraded difference type estimator as follows

$$\hat{\bar{Y}}_{R,D} = t_1 \bar{y} + t_2(\bar{X} - \bar{x}),$$

where $t_1$ and $t_2$ are constants. The minimum MSE of $\hat{\bar{Y}}_{R,D}$ at the optimum values

$$t_{1(opt)} = \frac{1}{1 + \lambda C_y^2 (1 - \rho_{yx}^2)},$$

and

$$t_{2(opt)} = \frac{\bar{Y} C_y \rho_{yx}}{\bar{X} C_x \left(1 + \lambda C_y^2 (1 - \rho_{yx}^2)\right)},$$

is given by

$$MSE_{min}(\hat{\bar{Y}}_{R,D}) = \frac{\lambda \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2)}{1 + \lambda C_y^2 (1 - \rho_{yx}^2)}.$$

The better efficiency of $\hat{\bar{Y}}_{R,D}$ over $\hat{\bar{Y}}_D$ can easily be shown as

$$MSE_{min}(\hat{\bar{Y}}_{R,D}) = Var_{min}(\hat{\bar{Y}}_D) - \frac{\lambda^2 \bar{Y}^2 C_y^4 (1 - \rho_{yx}^2)^2}{1 + \lambda C_y^2 (1 - \rho_{yx}^2)}. \tag{3}$$

Grover and Kaur[7] has proposed an additional well-known study on a family of exponential estimators that employ auxiliary data in a difference-based formulation in terms of an exponential functional. Their estimators are generally structured as they are

$$\hat{\bar{Y}}_{GK} = \big(\omega_1 \bar{y} + \omega_2(\bar{X} - \bar{x})\big) exp\left(\frac{a(\bar{X} - \bar{x})}{a(\bar{X} - \bar{x}) + 2b}\right),$$

where $\omega_1$ and $\omega_2$ are constants and $\theta = \frac{a\bar{X}}{a\bar{X}+b}$. The minimum MSE of $\hat{\bar{Y}}_{GK}$ as

$$MSE_{min}(\hat{\bar{Y}}_{GK}) = \frac{\lambda \bar{Y}^2\big(64C_y^2(1-\rho_{yx}^2) - \lambda\theta^4 C_x^4 - 16\lambda\theta^2 C_x^2 C_y^2(1-\rho_{yx}^2)\big)}{64\big(1+\lambda C_y^2(1-\rho_{yx}^2)\big)},$$

with the optimum values of $\omega_1$ and $\omega_2$ are given

$$\omega_{1(opt)} = \frac{8-\lambda\theta^2 C_x^2}{8\big(1+\lambda C_y^2(1-\rho_{yx}^2)\big)},$$

and

$$\omega_{2(opt)} = \frac{\bar{Y}\Big(\lambda\theta^3 C_x^3 + 8C_y\rho_{yx} - \lambda\theta^2 C_x^2 C_y \rho_{yx} - 4\theta C_x\big(1-\lambda C_y^2(1-\rho_{yx}^2)\big)\Big)}{8\bar{X}C_x\{1+\lambda C_y^2(1-\rho_{yx}^2)\}}.$$

The authors showed that the estimate $\hat{\bar{Y}}_{GK}$ consistently outperforms the difference estimator $\hat{\bar{Y}}_D$, i.e.

$$MSE_{min}(\hat{\bar{Y}}_{GK}) = Var_{min}(\hat{\bar{Y}}_D) - \frac{\lambda^2\bar{Y}^2\big(\theta^2 C_x^2 + 8C_y^2(1-\rho_{yx}^2)\big)^2}{64\big(1+\lambda C_y^2(1-\rho_{yx}^2)\big)}. \tag{4}$$

In a recent work, Haq et al.[11] proposed a difference-ratio-type family of exponential estimators, which is expressed as

$$\hat{\bar{Y}}_{AH} = \delta\bar{y} + \delta_2(\bar{X} - \bar{x}) + \delta_3(\bar{R}_x - \bar{r}_x)\exp\left(\frac{a(\bar{X} - \bar{x})}{a(\bar{X} + \bar{x}) + 2b}\right),$$

where $\delta_1$, $\delta_2$, and $\delta_3$ are constants. The MSE of their family of estimators

$$\mathrm{MSE}_{\min}(\hat{\bar{Y}}_{AH}) \cong \frac{\lambda\bar{Y}^2\big(64C_y^2(1-Q_{y.xr_x}^2) - \lambda\theta^4 C_x^4 - 16\lambda\theta^2 C_x^2 C_y^2(1-Q_{y.xr_x}^2)\big)}{64\big(1+\lambda C_y^2(1-Q_{y.xr_x}^2)\big)}, \tag{5}$$

with the following optimal values of the constants

$$\delta_{1(\mathrm{opt})} = \frac{8-\lambda\theta^2 C_x^2}{8\big(1+\lambda C_y^2(1-Q_{y.xr_x}^2)\big)},$$

$$\delta_{2(\mathrm{opt})} = \frac{\bar{Y}\begin{pmatrix}\lambda\theta^3 C_x^3(-1+\rho_{xr_x}^2) + (-8C_y + \lambda\theta^2 C_x^2 C_y)(\rho_{yx} - \rho_{xr_x}\rho_{yr_x}) + \\ 4\theta C_x(-1+\rho_{xr_x}^2)\big(-1+\lambda C_y^2(1-Q_{y.xr_x}^2)\big)\end{pmatrix}}{8\bar{X}C_x(-1+\rho_{xr_x}^2)\big(1+\lambda C_y^2(1-Q_{y.xr_x}^2)\big)},$$

and

$$\delta_{3(\mathrm{opt})} = \frac{\bar{Y}(8-\lambda\theta^2 C_x^2)C_y(\rho_{xr_x}\rho_{yx} - \rho_{yr_x})}{8\bar{R}xC_r(-1+\rho x{r_x}^2)\big(1+\lambda C_y^2(1-Q_{y.xr_x}^2)\big)}.$$

The coefficient of multiple determination is denoted by $Q_{y\cdot xrx}^2 = \frac{\rho_{yx}^2 + \rho_{yr_x}^2 - 2\rho_{yx}\rho_{yr_x}\rho_{xr_x}}{1-\rho_{xr_x}^2}$.

## 3. Proposed family of estimators

**Motivation.** The subsequent factors must be taken into account in order to enhance the precision of an estimator. It is imperative to effectively employ supplementary information during both the design and estimation phases. It is important to mention that supplementary data that is required for the survey conducted within the specified context is typically accessible in the case of other groups of researchers. In other words, the values of the study variable can be represented in the estimated relative rank of the auxiliary variable when there is a strong correlation between the research variable and the auxiliary variable. In light of this information, we suggest an enhanced family of estimators for the mean of a finite population. This proposed class of estimators includes the auxiliary variable's relative ranks and supplementary information.

**Notations.** Let $U = \{x_1, x_2, \ldots, x_N\}$ be the set of $N$ individual values of the $X$ variable in a finite population. Refer to the following formula to calculate the relative rankings, denoted as $R_r$, of the auxiliary variable $X$.

- Define $(R_r)_1 = 1$.
- Define
$$x_r = i - 1 + \frac{(N-1)(x_{(i)} - x_{(i-1)})}{x_{(N)} - x_{(1)}}, \ \ i = 2, 3, \ldots, N.$$
- For $i = 2, 3, \ldots, N$, define $(R_r)_i$ to be the $(i-1)^{th}$ smallest value of $\{(x_r)_2, (x_r)_3, \ldots, (x_r)_N\}$, so that
$$(R_r)_1 < (R_r)_2 < \cdots < (R_r)_N.$$

Let $(\bar{r}_r)_x$, $(\bar{R}_r)_x$, and $S^2_{(\bar{r}_r)x}$ represents the corresponding sample mean, population mean, and population variance for relative ranks. We compute $(\bar{r}_r)_x = \frac{1}{n}\sum_{i=1}^{n}(r_r)_{x,i}/n$, $(\bar{R}_r)_x = \frac{1}{N}\sum_{i=1}^{N}(r_r)_{x,i}/N = (N+1)/2$, and $S^2_{(r_r)_x} = \frac{1}{N-1}\sum_{i=1}^{N}\big((r_r)_{x,i} - (\bar{R}_r)_x\big)^2$, where $(r_r)_{x,i}$ denotes the i$th$ value of $(R_r)_x$ in the population $U$. Let $\rho_{y.(r_r)_x} = \frac{S_{y.(r_r)x}}{S_y S_{(r_r)x}}$ be the correlation coefficient between $Z$ and $(R_r)_x$, where $S_{y.(r_r)_x} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})\big((r_r)_{x,i} - (\bar{R}_r)_x\big)/(N-1)$ is the population covariance between $Y$ and $(\bar{R}_r)$. Let $C_{(r_r)} = \frac{S_{(r_r)x}}{(\bar{R}_r)_x}$ be the coefficient of variation of $(\bar{R}_r)$. We take into consideration the following relative error terms as we try to determine the bias and MSE of the suggested estimators: $\epsilon_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $\epsilon_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$, and $\epsilon_2 = \frac{(\bar{r}_r)_x - (\bar{R}_r)_x}{(\bar{R}_r)_x}$, such that $E(\epsilon_i) = 0$ for $i = 0, 1, 2$. It is easy to show that $E(\epsilon_0^2) = \lambda C_y^2$, $E(\epsilon_1^2) = \lambda C_x^2$, $E(\epsilon_2^2) = \lambda C^2_{(r_r)}$, $E(\epsilon_0\epsilon_1) = \lambda \rho_{yx} C_y C_x$, $E(\epsilon_0\epsilon_2) = \lambda \rho_{y.(r_r)} C_y C_{(r_r)}$, and $E(\epsilon_1\epsilon_2) = \lambda \rho_{x.(r_r)} C_x C_{(r_r)}$.

**Layout.** According to [13], an approach to estimating the mean of a finite population that takes into consideration the auxiliary information with $(\bar{R}_r)_x$ is offered by a difference-type estimator as
$$\hat{\bar{Y}}_{P_r} = \eta_1 \bar{y} + \eta_2(\bar{X} - \bar{x}) + \eta_3\big((\bar{R}_r)_x - (\bar{r}_r)_x\big).$$

where $\eta_1$, $\eta_2$, and $\eta_3$ are constants that will be chosen appropriate. We propose a difference-ratio-type class of exponential estimators, $\hat{\hat{Y}}^*_{Pr}$, based on the studies of [7, 11, 14, 15] as in the following form

$$\hat{\hat{Y}}^*_{P_r} = \eta_1\bar{y} + \eta_2(\bar{X}-\bar{x}) + \eta_3\big((\bar{R}_r)_x - (\bar{r}_r)_x\big)\exp\left(\frac{a(\bar{X}-\bar{x})}{a(\bar{X}+\bar{x})+2b}\right),$$

where $a$ and $b$ are as explained earlier. Upon reformulating $\hat{\hat{Y}}^*_{\mathrm{Pr}}$ in terms of relative error terms, we get

$$\hat{\hat{Y}}^*_{Pr} = \big(\eta_1\bar{Y}(1+\epsilon_0) - \eta_2\bar{X}\epsilon_1 - \eta_3(\bar{R}_r)_x\epsilon_2\big)\left(1 - \frac{\theta\epsilon_1}{2} + \frac{3\theta^2\epsilon_1^2}{8} + \cdots\right).$$

We can write the following by expanding the above equation and restricting terms to only two digits in $\epsilon_i$'s

$$\begin{aligned}(\hat{\hat{Y}}^*_{Pr} - \bar{Y}) \cong & -\bar{Y} + \bar{Y}\eta_1 + \bar{Y}\epsilon_0\eta_1 - \frac{1}{2}\bar{Y}\theta\epsilon_1\eta_1 - \bar{X}\epsilon_1\eta_2 - (\bar{R}_r)_x\epsilon_2\eta_3 \\ & - \frac{1}{2}\bar{Y}\theta\epsilon_0\epsilon_1\eta_1 + \frac{3}{8}\bar{Y}\theta^2\epsilon_1^2\eta_1 + \frac{1}{2}\bar{X}\theta\epsilon_1^2\eta_2 + \frac{1}{2}(\bar{R}_r)_x\theta\epsilon_1\epsilon_2\eta_3.\end{aligned}$$

The bias and MSE of the estimated value $\hat{\hat{Y}}^*_{Pr}$, using the first-order approximation, can be expressed as

$$\begin{aligned}\mathrm{Bias}\left(\hat{\hat{Y}}^*_{Pr}\right) \cong & \frac{1}{8}\Big(-8\bar{Y} + 4\lambda\theta C_x\big(\bar{X}C_x\eta_2 + (\bar{R}_r)_x C_{(r_r)}\eta_3\rho_{x.(r_r)_x}\big) \\ & + \bar{Y}\eta_1\big(8 + \lambda\theta C_x\,(3\theta C_x - 4C_y\rho_{y.x})\,\big)\Big),\end{aligned}$$

and

$$\begin{aligned}\mathrm{MSE}(\hat{\hat{Y}}^*_{Pr}) \cong & \bar{Y}^2 + \lambda\bar{X}C_x^2\eta_2(-\bar{Y}\theta + \bar{X}\eta_2) + \lambda(\bar{R}_r)_x^2 C_{(r_r)}^2\eta_3^2 + \lambda(\bar{R}_r)_x C_x C_{(r_r)} \\ & (-\bar{Y}\theta + 2\bar{X}\eta_2)\eta_3\rho_{x.(r_r)_x} + \bar{Y}^2\eta_1^2\big(1 + \lambda\big(C_y^2 + \theta C_x(\theta C_x - 2C_y\rho_{y.x})\big)\big) \\ & + \frac{1}{4}\bar{Y}\eta_1\Big(-8\bar{Y} + \lambda C_x\big(\theta C_x(-3\bar{Y}\theta + 8\bar{X}\eta_2) + 8(\bar{R}_r)_x\theta C_{r_r}\eta_3\rho_{x.(r_r)_x} \\ & + 4C_y(\bar{Y}\theta - 2\bar{X}\eta_2)\rho_{y.x}\big) - 8(\bar{R}_r)_x\lambda C_y C_{(r_r)}\eta_3\rho_{y.(r_r)_x}\Big).\end{aligned} \tag{6}$$

The optimal values of $\eta_1$, $\eta_2$, and $\eta_3$ obtained by minimizing equation (6) are

$$\eta_{1(\mathrm{opt})} = \frac{8 - \lambda\theta^2 C_x^2}{8\big(1 + \lambda C_y^2(1 - K^2_{y.x.(r_r)_x})\big)},$$

$$\eta_{2(\mathrm{opt})} = \frac{\bar{Y}\begin{pmatrix}\lambda\theta^3 C_x^3(-1+\rho^2_{x.(r_r)_x}) + (-8C_y + \lambda\theta^2 C_x^2 C_y)(\rho_{yx} - \rho_{x.(r_r)_x}\rho_{y.(r_r)_x}) \\ + 4\theta C_x(-1+\rho^2_{x.(r_r)_x})\big(-1 + \lambda C_y^2(1 - K^2_{y.x.(r_r)_x})\big)\end{pmatrix}}{8\bar{X}C_x(-1+\rho^2_{x.(r_r)_x})\big(1 + \lambda C_y^2(1 - K^2_{y.x.(r_r)_x})\big)},$$

and

$$\eta_{3(\mathrm{opt})} = \frac{\bar{Y}(8 - \lambda\theta^2 C_x^2)C_y(\rho_{x.(r_r)_x}\rho_{y.x} - \rho_{y.(r_r)_x})}{8(\bar{R}_r)_x C_{(r_r)}(-1+\rho^2_{x.(r_r)_x})\big(1 + \lambda C_y^2(1 - K^2_{y.x.(r_r)_x})\big)}.$$

The coefficient of multiple determination, $K^2_{y.x.(r_r)_x} = \frac{\rho^2_{y.x} + \rho^2_{y.(rr)_x} - 2\rho_{y.x}\rho_{y.(rr)_x}\rho_{x.(rr)_x}}{1 - \rho^2_{x.(rr)_x}}$, represents the extent to which the variable $Y$ can be explained by both $X$ and $(R_r)_x$. By substituting the optimal values of $\eta_1$, $\eta_2$, and $\eta_3$ into equation (6) and simplifying, we obtain the minimum MSE of $\hat{\hat{Y}}^{*}_{Pr}$ as

$$\mathrm{MSE}_{\min}(\hat{\hat{Y}}^{*}_{P_r}) \cong \frac{\lambda\bar{Y}^2\Big(64C_y^2(1-K^2_{y.x.(r_r)_x})-\lambda\theta^4C_x^4-16\lambda\theta^2C_x^2C_y^2(1-K^2_{y.x.(r_r)_x})\Big)}{64\Big(1+\lambda C_y^2(1-K^2_{y.x.(r_r)_x})\Big)}. \tag{7}$$

## 4. Efficiency comparisons

This section presents a comparison between the suggested estimator and the existing estimators that are being considered in this study.

$$\frac{\lambda\bar{Y}^2\Big(\lambda\theta^4C_x^4+16C_y^2\Big(4K^2_{y.x.(r_r)_x}+\lambda\theta^2C_x^2(1-K^2_{y.x.(r_r)_x})\Big)+64\lambda C_y^4(1-K^2_{y.x.(r_r)_x})\Big)}{64\Big(1+\lambda C_y^2(1-K^2_{y.x.(r_r)_x})\Big)}>0.$$

$MSE_{\min}(\hat{\hat{Y}}^{*}_{\mathrm{Pr}}) < \mathrm{Var}_{\min}(\hat{\hat{Y}}_D)$ if

$$\frac{\lambda^2\bar{Y}^2\Big(\theta^2C_x^2+8C_y^2(1-\rho^2_{yx})^2\Big)}{64\Big(1+\lambda C_y^2(1-\rho^2_{yx})\Big)} + \frac{\lambda\bar{Y}^2C_y^2(\rho_{y.(r_r)_x}-\rho_{yx}\rho_{x.(r_r)_x})^2(-8+\lambda\theta^2C_x^2)^2}{64(1-\rho^2_{x.(r_r)_x})\Big(1+\lambda C_y^2(1-\rho^2_{yx})\Big)\Big(1+\lambda C_y^2(1-K^2_{y.x.(r_r)_x})\Big)}>0.$$

$$\frac{\lambda^2\theta^2\bar{Y}^2C_x\Big(\theta^2C_x^2+16C_y^2(1-\rho^2_{yx})\Big)}{64\Big(1+\lambda C_y^2(1-\rho^2_{yx})\Big)} + \frac{\lambda\bar{Y}^2C_y^2(\rho_{y.(r_r)_x}-\rho_{yx}\rho_{x.(r_r)_x})^2(-8+\lambda\theta^2C_x^2)^2}{64(1-\rho^2_{x.(r_r)_x})\Big(1+\lambda C_y^2(1-\rho^2_{yx})\Big)\Big(1+\lambda C_y^2(1-K^2_{y.x.(r_r)_x})\Big)}>0.$$

$$\frac{\lambda\bar{Y}^2C_y^2(\rho_{y.(r_r)_x}-\rho_{yx}\rho_{x.(r_r)_x})^2(-8+\lambda\theta^2C_x^2)^2}{64(1-\rho^2_{x.(r_r)_x})\Big(1+\lambda C_y^2(1-\rho^2_{yx})\Big)\Big(1+\lambda C_y^2(1-K^2_{y.x.(r_r)_x})\Big)}>0.$$

$$\frac{\lambda C_y^2\Bigg(\Big(\bar{Y}(1-\lambda-\lambda C_y^2)+\theta^2C_x^2\Big((1+2\lambda C_y^2)+\lambda C_x^2\Big)\Big)(Q^2-K^2)\Bigg)+\lambda^2C_y^4(4\bar{Y}^2-\theta^2C_x^2)(Q^4-K^4)}{4(1+\lambda C_y^2-\lambda C_y^2Q^2)(1+\lambda C_y^2-\lambda C_y^2K^2)}>0.$$

It is imperative to underscore that the aforementioned conditions are perpetually legitimate. Consequently, the estimators that are recommended outperform all of the other estimators that were evaluated in this scenario.

## 5. Performance evaluations

This section provides a comprehensive account of the empirical and simulation-based investigations that were employed to assess the relative efficacy of the aforementioned procedures. To guarantee inclusivity, we have selected three authentic datasets that are widely acknowledged in the field of survey methodologies and encompass a diverse array of disciplines. Additionally, we have employed three distinct bivariate distributions to generate three datasets. Additionally, the objectives of conducting a fair comparison are accomplished by considering the frequently employed datasets and the methodologies employed by contemporary researchers, such as the highly regarded [11] family.

5.1. **Empirical evaluation.** In this section, we employ numerical calculations to investigate the relationship between the MSE and PRE of the proposed family of estimators. We have chosen three authentic datasets that are well-known in the survey methodology sector and encompass multiple disciplines.

**Dataset 1.** $\left(Source : Singh[16]\right)$

Y: estimated length of sleep (in minutes) for individuals over the age of 50 years.

X: age of individuals in years.

$N = 36$, $n = 5$, $\bar{Y} = 0.1709$, $\bar{X} = 0.1856$, $\bar{R}_X = 18.5$, $\rho_{YX} = 0.8788$, $\rho_{YR} = 0.8448$, $\rho_{XR} = 0.9582$, $C_y = 0.3709$, $C_x = 0.4050$, $C_r = 0.5694$, $\beta_2(x) = 3.3450$.

**Dataset 2.** $\left(Source : Gujarati[17]\right)$

Y: The millions of eggs that were produced in 1990.

X: The cost (cents) per dozen in 1990.

$N = 50$, $n = 5$, $\bar{Y} = 1357.622$, $\bar{X} = 78.29$, $\bar{R}_X = 25.5$, $\rho_{YX} = -0.2888$, $\rho_{YR} = -0.2469$, $\rho_{XR} = 0.9468$, $C_y = 1.2236$, $C_x = 0.2723$, $C_r = 0.5716$, $\beta_2(x) = 4.0255$.

**Dataset 3.** $\left(Source : Murthy[18]\right)$

Y: Output production of factories in a region.

X: No. of workers in factories in a region.

$N = 80$, $n = 10$, $\bar{Y} = 5182.637$, $\bar{X} = 285.125$, $\bar{R}_X = 40.5$, $\rho_{YX} = 0.9150$, $\rho_{YR} = 0.9836$, $\rho_{XR} = 0.8902$, $C_y = 0.3542$, $C_x = 0.0.9485$, $C_r = 0.5736$, $\beta_2(x) = 3.5808$.

Table 1-3 presents the results of the PREs evaluation of the proposed and existing estimators with respect to $\bar{y}$. The superiority of the proposed estimators over all other estimators studied in this study on real populations 1-3 is evident from Table 1-3 respectively.

5.2. **Simulation study.** This section employs a simulation study to investigate the relationship between the MSE and PRE of the proposed estimators. By generating three finite populations from three distinct and well-known bivariate probability models, each with 1000 realizations. The details are provided below.

**Bivariate normal distribution.** Our first simulated-investigation is follow to a bivariate normal distribution with unique means for the study and auxiliary variables as

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_y = 11 \\ \mu_x = 52 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 = 72 & \rho\sigma_y\sigma_x = 142 \\ \rho\sigma_x\sigma_y = 142 & \sigma_x^2 = 337 \end{pmatrix}\right).$$

We employed a uniform sample size of $n = 100$ and thereafter utilized the following equation to facilitate a comparison of efficiency. We compute the PREs for the proposed and contemporary estimators in relation to the variance of the study variable. The findings are presented in Table 4, demonstrating that the proposed technique outperformed in all instances.

**Bivariate uniform distribution.** The population is assumed to be governed by a bivariate uniform distribution with standardized means and variances along $\rho = 0.9$. For the aim of comparing efficiency, we utilized a sample size of $n = 100$ and employed the subsequent expression. We compute the PREs of both the proposed and contemporary estimators in relation to the variance of the research variable. The findings are presented in Table 5, demonstrating that the proposed technique outperformed in all cases.

**Bivariate t-distribution.** The population is assumed to be regulated by a bivariate student's t distribution with distinct means for both study and auxiliary variables as

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim t\left( \begin{pmatrix} \mu_y = 11 \\ \mu_x = 52 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 = 72 & \rho\sigma_y\sigma_x = 142 \\ \rho\sigma_x\sigma_y = 142 & \sigma_x^2 = 337 \end{pmatrix} \right).$$

For the purpose of comparing efficiency, we utilized a sample size of 100, denoted as $n = 100$, and employed the subsequent expression. We compute the PREs for both the proposed and contemporary estimators in relation to the variance of the research variable. The results are presented in Table 6, and they consistently demonstrate that the recommended estimators outperformed in all scenarios.

## 6. Summary

The present study's findings contribute to the existing literature on survey sampling methodology by addressing the challenges associated with estimating the mean within a finite population. By introducing a novel class of estimators that employ auxiliary variables and their respective rankings to offer additional information, the goals are successfully accomplished. This newly suggested family of estimators is evaluated by an analysis of three real-world datasets and a thorough simulation study utilizing three well recognized bivariate probability models. As predicted, the very high degree of coupling of the relative ranks led to the introduction of dual information in a more object-oriented architecture. As said before, the results of the study validated a rational approach to leverage relative ranks in order to enhance the efficiency of the estimation procedure. Moreover, this improvement is readily apparent in connection to all the examined datasets and through simulation experiments.

Table 1. PREs of the estimators where $\hat{\bar{Y}}$ is used as base-line evaluator by using Dataset 1.

| Estimator | | a | b | Families $\hat{\bar{Y}}_{GK}$ | $\hat{\bar{Y}}_{AH}$ | $\hat{\bar{Y}}_{Pr}$ |
|---|---|---|---|---|---|---|
| $\bar{y}$ | 100 | – | – | – | – | – |
| $\hat{\bar{Y}}_D$ | 439.09 | – | – | – | – | – |
| $\hat{\bar{Y}}_{R,D}$ | 441.46 | – | – | – | – | – |
| 1 | – | 1 | $C_X$ | 441.77 | 441.95 | 444.42 |
| 2 | – | 1 | $\beta_2(x)$ | 441.46 | 441.64 | 444.11 |
| 3 | – | $\beta_2(x)$ | $C_X$ | 442.74 | 442.91 | 445.39 |
| 4 | – | $C_X$ | $\beta_2(x)$ | 441.46 | 441.63 | 444.10 |
| 5 | – | 1 | $\rho_{YX}$ | 441.55 | 441.73 | 444.19 |
| 6 | – | $C_X$ | $\rho_{YX}$ | 441.47 | 441.65 | 444.12 |
| 7 | – | $\rho_{YX}$ | $C_X$ | 441.72 | 441.90 | 444.36 |
| 8 | – | $\beta_2(x)$ | $\rho_{YX}$ | 442.02 | 442.20 | 444.67 |
| 9 | – | $\rho_{YX}$ | $\beta_2(x)$ | 441.46 | 441.64 | 444.10 |
| 10 | – | 1 | $N\bar{X}$ | 441.46 | 441.63 | 444.10 |

Table 2. PREs of the estimators where $\hat{\bar{Y}}$ is used as base-line evaluator by using Dataset 2.

| Estimator | | a | b | Families $\hat{\bar{Y}}_{GK}$ | $\hat{\bar{Y}}_{AH}$ | $\hat{\bar{Y}}_{Pr}$ |
|---|---|---|---|---|---|---|
| $\bar{y}$ | 100 | – | – | – | – | – |
| $\hat{\bar{Y}}_D$ | 109.10 | – | – | – | – | – |
| $\hat{\bar{Y}}_{R,D}$ | 136.05 | – | – | – | – | – |
| 1 | – | 1 | $C_X$ | 136.51 | 137.32 | 138.22 |
| 2 | – | 1 | $\beta_2(x)$ | 136.47 | 137.28 | 138.18 |
| 3 | – | $\beta_2(x)$ | $C_X$ | 136.51 | 137.33 | 138.22 |
| 4 | – | $C_X$ | $\beta_2(x)$ | 136.38 | 137.19 | 138.08 |
| 5 | – | 1 | $\rho_{YX}$ | 136.51 | 137.33 | 138.22 |
| 6 | – | $C_X$ | $\rho_{YX}$ | 136.52 | 137.34 | 138.23 |
| 7 | – | $\rho_{YX}$ | $C_X$ | 136.52 | 137.34 | 138.23 |
| 8 | – | $\beta_2(x)$ | $\rho_{YX}$ | 136.51 | 137.33 | 138.22 |
| 9 | – | $\rho_{YX}$ | $\beta_2(x)$ | 136.73 | 137.55 | 138.44 |
| 10 | – | 1 | $N\bar{X}$ | 136.05 | 136.87 | 137.76 |

Table 3. PREs of the estimators where $\hat{\hat{Y}}$ is used as base-line evaluator by using Dataset 3.

| Estimator | | a | b | Families $\hat{\hat{Y}}_{GK}$ | $\hat{\hat{Y}}_{AH}$ | $\hat{\hat{Y}}_{Pr}$ |
|---|---|---|---|---|---|---|
| $\bar{y}$ | 100 | – | – | – | – | – |
| $\hat{\hat{Y}}_D$ | 614.21 | – | – | – | – | – |
| $\hat{\hat{Y}}_{R,D}$ | 615.31 | – | – | – | – | – |
| 1 | – | 1 | $C_X$ | 663.77 | 6307.63 | 6351.57 |
| 2 | – | 1 | $\beta_2(x)$ | 662.14 | 6182.12 | 6224.33 |
| 3 | – | $\beta_2(x)$ | $C_X$ | 664.20 | 6342.04 | 6386.45 |
| 4 | – | $C_X$ | $\beta_2(x)$ | 662.02 | 6173.26 | 6215.35 |
| 5 | – | 1 | $\rho_{YX}$ | 663.79 | 6309.30 | 6353.26 |
| 6 | – | $C_X$ | $\rho_{YX}$ | 663.76 | 6306.83 | 6350.75 |
| 7 | – | $\rho_{YX}$ | $C_X$ | 663.71 | 6303.25 | 6347.13 |
| 8 | – | $\beta_2(x)$ | $\rho_{YX}$ | 664.21 | 6342.51 | 6386.93 |
| 9 | – | $\rho_{YX}$ | $\beta_2(x)$ | 661.94 | 6167.00 | 6209.01 |
| 10 | – | 1 | $N\bar{X}$ | 615.31 | 3993.05 | 4010.96 |

Table 4. PREs of the estimators where $\hat{\hat{Y}}$ is used as base-line evaluator by using bivariate normal distribution.

| Estimator | | a | b | Families $\hat{\hat{Y}}_{GK}$ | $\hat{\hat{Y}}_{AH}$ | $\hat{\hat{Y}}_{Pr}$ |
|---|---|---|---|---|---|---|
| $\bar{y}$ | 100 | – | – | – | – | – |
| $\hat{\hat{Y}}_D$ | 593.2314 | – | – | – | – | – |
| $\hat{\hat{Y}}_{R,D}$ | 593.7678 | – | – | – | – | – |
| 1 | – | 1 | $C_X$ | 593.9451 | 594.5442 | 594.5445 |
| 2 | – | 1 | $\beta_2(x)$ | 593.9274 | 594.5265 | 594.5268 |
| 3 | – | $\beta_2(x)$ | $C_X$ | 593.9468 | 594.5460 | 594.5462 |
| 4 | – | $C_X$ | $\beta_2(x)$ | 593.8983 | 594.4974 | 594.4977 |
| 5 | – | 1 | $\rho_{YX}$ | 593.9411 | 594.5402 | 594.5405 |
| 6 | – | $C_X$ | $\rho_{YX}$ | 593.9300 | 594.5291 | 594.5294 |
| 7 | – | $\rho_{YX}$ | $C_X$ | 593.9448 | 594.5440 | 594.5443 |
| 8 | – | $\beta_2(x)$ | $\rho_{YX}$ | 593.9454 | 594.5446 | 594.5449 |
| 9 | – | $\rho_{YX}$ | $\beta_2(x)$ | 593.9256 | 594.5247 | 594.5250 |
| 10 | – | 1 | $N\bar{X}$ | 593.7678 | 594.3668 | 594.3671 |

Table 5. PREs of the estimators where $\hat{\bar{Y}}$ is used as base-line evaluator by using bivariate uniform distribution.

| Estimator | | a | b | Families | | |
|---|---|---|---|---|---|---|
| | | | | $\hat{\bar{Y}}_{GK}$ | $\hat{\bar{Y}}_{AH}$ | $\hat{\bar{Y}}_{Pr}$ |
| $\bar{y}$ | 100 | – | – | – | – | – |
| $\hat{\bar{Y}}_D$ | 537.7205 | – | – | – | – | – |
| $\hat{\bar{Y}}_{R,D}$ | 538.0210 | – | – | – | – | – |
| 1 | – | 1 | $C_X$ | 538.1143 | 538.6669 | 538.6681 |
| 2 | – | 1 | $\beta_2(x)$ | 538.0403 | 538.5929 | 538.5941 |
| 3 | – | $\beta_2(x)$ | $C_X$ | 538.1899 | 538.7427 | 538.7438 |
| 4 | – | $C_X$ | $\beta_2(x)$ | 538.0287 | 538.5813 | 538.5825 |
| 5 | – | 1 | $\rho_{YX}$ | 538.0746 | 538.6272 | 538.6283 |
| 6 | – | $C_X$ | $\rho_{YX}$ | 538.0452 | 538.5978 | 538.5990 |
| 7 | – | $\rho_{YX}$ | $C_X$ | 538.1037 | 538.6564 | 538.6575 |
| 8 | – | $\beta_2(x)$ | $\rho_{YX}$ | 538.1304 | 538.6830 | 538.6842 |
| 9 | – | $\rho_{YX}$ | $\beta_2(x)$ | 538.0374 | 538.5900 | 538.5911 |
| 10 | – | 1 | $N\bar{X}$ | 538.0210 | 538.5735 | 538.5747 |

Table 6. PREs of the estimators where $\hat{\bar{Y}}$ is used as base-line evaluator by using bivariate Student's t-distribution.

| Estimator | | a | b | Families | | |
|---|---|---|---|---|---|---|
| | | | | $\hat{\bar{Y}}_{GK}$ | $\hat{\bar{Y}}_{AH}$ | $\hat{\bar{Y}}_{Pr}$ |
| $\bar{y}$ | 100 | – | – | – | – | – |
| $\hat{\bar{Y}}_D$ | 593.5117 | – | – | – | – | – |
| $\hat{\bar{Y}}_{R,D}$ | 594.0485 | – | – | – | – | – |
| 1 | – | 1 | $C_X$ | 594.2259 | 594.8235 | 594.8240 |
| 2 | – | 1 | $\beta_2(x)$ | 594.2081 | 594.8057 | 594.8062 |
| 3 | – | $\beta_2(x)$ | $C_X$ | 594.2276 | 594.8252 | 594.8257 |
| 4 | – | $C_X$ | $\beta_2(x)$ | 594.1790 | 594.7766 | 594.7770 |
| 5 | – | 1 | $\rho_{YX}$ | 594.2219 | 594.8195 | 594.8200 |
| 6 | – | $C_X$ | $\rho_{YX}$ | 594.2107 | 594.8084 | 594.8089 |
| 7 | – | $\rho_{YX}$ | $C_X$ | 594.2256 | 594.8233 | 594.8237 |
| 8 | – | $\beta_2(x)$ | $\rho_{YX}$ | 594.2262 | 594.8239 | 594.8243 |
| 9 | – | $\rho_{YX}$ | $\beta_2(x)$ | 594.2063 | 594.8040 | 594.8044 |
| 10 | – | 1 | $N\bar{X}$ | 594.0485 | 594.6459 | 594.6464 |

## References


[1] C. Kadilar, H. Cingi, Ratio estimators in simple random sampling, Appl. Math. Comp. 151 (2004), 893–902. https://doi.org/10.1016/S0096-3003(03)00803-8.

[2] C. Kadilar, H. Cingi, An improvement in estimating the population mean by using the correlation coefficient, Hacettepe J. Math. Stat. 35 (2006), 103–109.

[3] C. Kadilar, H. Cingi, Improvement in estimating the population mean in simple random sampling, Appl. Math. Lett. 19 (2006), 75–79. https://doi.org/10.1016/j.aml.2005.02.039.

[4] S. Gupta, J. Shabbir, On improvement in estimating the population mean in simple random sampling, J. Appl. Stat. 35 (2008), 559–566. https://doi.org/10.1080/02664760701835839.

[5] A. Haq, J. Shabbir, Improved family of ratio estimators in simple and stratified random sampling, Commun. Stat. - Theory Meth. 42 (2013), 782–799. https://doi.org/10.1080/03610926.2011.579377.

[6] H.P. Singh, R.S. Solanki, An efficient class of estimators for the population mean using auxiliary information, Commun. Stat. - Theory Meth. 42 (2013), 145–163. https://doi.org/10.1080/03610926.2011.575519.

[7] L.K. Grover, P. Kaur, A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable, Commun. Stat. - Simul. Comp. 43 (2014), 1552–1574. https://doi.org/10.1080/03610918.2012.736579.

[8] J. Shabbir, A. Haq, S. Gupta, A new difference-cum-exponential type estimator of finite population mean in simple random sampling, Rev. Colomb. Estad. 37 (2014), 199-211.

[9] S. Lohr, Sampling: design and analysis, Duxbury Press, (1999).

[10] W.G. Cochran, The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce, J. Agric. Sci. 30 (1940), 262–275. https://doi.org/10.1017/S0021859600048012.

[11] A. Haq, M. Khan, Z. Hussain, A new estimator of finite population mean based on the dual use of the auxiliary information, Commun. Stat. - Theory Meth. 46 (2017), 4425–4436. https://doi.org/10.1080/03610926.2015.1083112.

[12] A. Hussain, S. Drekic, S.A. Cheema, A relative-rank measure for the rank transformation, Stat. Prob. Lett. 204 (2024), 109932. https://doi.org/10.1016/j.spl.2023.109932.

[13] T.J. Rao, On certail methods of improving ration and regression estimators, Commun. Stat. - Theory Meth. 20 (1991), 3325–3340. https://doi.org/10.1080/03610929108830705.

[14] L.K. Grover, P. Kaur, Ratio type exponential estimators of population mean under linear transformation of auxiliary variable: theory and methods, South Afr. Stat. J. 45 (2011), 205–230. https://hdl.handle.net/10520/EJC99138.

[15] J. Shabbir, S. Gupta, On estimating finite population mean in simple and stratified random sampling, Commun. Stat. - Theory Meth. 40 (2010), 199–212. https://doi.org/10.1080/03610920903411259.

[16] S. Singh, Advanced sampling theory with applications: How Michael "selected" Amy, Springer, (2003).

[17] D.N. Gujarati, Basic econometrics, McGraw-Hill, (2009).

[18] M.N. Murthy, Sampling theory and methods, Statistical Publication Society, Calcutta, (1967).