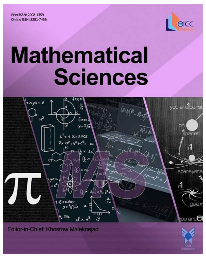


Mathsci. 2025; 19(3): Article 20

Mathematical Sciences

https://doi.org/10.57647/mathsci.2025.1904.20

Improvement of the Population Mean Using Cosine Robust Regression Estimator

Muhammad Ijaz [1,*], Muneeba Slaeem [1], Sadiah M. Aljeddani [2]

[1] *Department of Mathematics and Statistics, The University of Haripur, Haripur, Pakistan*

[2] *Department of Mathematics, Al-Lith University College, Umm Al-Qura University, Al-Lith, Saudi Arabia*

*

Corresponding authors: m.ijaz@uoh.edu.pk

Research Article

Received:
25 September 2025

Revised:
16 November 2025

Accepted:
20 November 2025

Published in Issue:
31 December 2025

Abstract

This paper introduces a novel family of cos-robust regression-type estimators along with special members to estimate the finite population mean under SRSWOR. The new family of estimators is produced by hybridizing the auxiliary information with the cos function. To reduce the impact of outliers, various robust regression techniques, namely Huber's M- estimation, Mallows' GM-estimation, Schweppe's GM-estimation, and SIS GM-estimation, are employed and theoretically compared with the Ordinary Least Squares (OLS) method. Simulated and actual data are used to generate and validate theoretical properties, such as bias and Mean Square Error (MSE). According to the findings, the robust estimators perform better than the conventional OLS approach in terms of MSE and PRE.

Mathematics Subject Classification (2020). 62D05,62G05, 62G35, 62J05

Keywords: Auxiliary information; Comparison; Cos function; MSE; Robust statistics; Variance

Cite this article: Ijaz, M., Slaeem, M., Aljeddani, S. M. Improvement of the Population Mean Using Cosine Robust Regression Estimator. *Mathsci* **19**, 04 (2025). https://doi.org/10.57647/mathsci.2025.1904.20

1. Introduction

As studying the entire population is often impossible in large-scale data analysis, prob- ability sampling techniques such as Simple Random Sampling (SRS) are employed to estimate the population parameter; however, the existence of outliers and variability in data may render traditional estimators under SRS ineffective.

To address this problem, researchers employed supplementary information and advanced statistical methods to enhance precision and mitigate the impact of outliers. Recent studies proposed new modifications to the existing one that incorporate auxiliary information, for example, [1,2] and[6] defined the ratio type estimator, [3] improved the population mean by using the coefficient of variation, [4, 5] explained the robust regression estimator by using the quartile based skewness coefficient and robust statistics.[7–9] presented a new modification by using the coefficient of skewness and coefficient of Kurtosis, [10] improved by using conventional measures of dispersion. For a more detailed study, we refer to see[12–14] for the ratio type estimator and [15–17] for the regression estimator.In the existing literature, this is a common practice among researchers to improve classical estimators by increasing auxiliary variables, increasing the number of parameters, or introducing a logarithmic operator. For example, [18, 19] introduced robust regression estimators using multiple auxiliary information,[20] increased the number of parameters, and [21–23] have used a logarithmic operator to increase the efficiency of the estimator. However, these methods have limitations; for example, increasing the number of parameters and auxiliary information creates computational complexity, and the log operator will distort the efficiency if there are zeros or negative values in the data.

Secondly, these methods do not sufficiently increase efficiency in the presence of outliers. To tackle these issues, the current research presents a new method that combines the cosine function with a robust estimation framework. Due to its bounded and oscillatory nature, the cosine transformation is especially effective in reducing the impact of extreme observations, as it compresses large deviations and stabilizes estimator performance in the presence of outliers.

Our proposed method leverages this property to provide an alternative means of achieving efficiency gains and robustness enhancements, separating our contribution from prior alterations of classical estimators.

The rest of the paper is organized as follows: Section 2 presents the literature review, Section 3 defines the derivation of the proposed estimators along with the MSE, Section 4 explains theoretical conditions, while Sections 5 and 6 explain the applications to real and simulated data. The paper concluded with section 7.

2. Notation

The following notations will be used in the paper.

N– Population size

n– Sample size

Y– Response variable

X– Auxiliary variable

$\bar{X}, \bar{Y}$– Population true means

$\bar{x}, \bar{y}$– Sample means

S_x, S_y– Population standard deviations of Xand Y

C_x, C_y– Population coefficients of variation of Xand Y

ρ_{yx}– Correlation between Xand Yin the population

β_1, β_2– Moment ratios

QD– Quartile deviation

$Q_i = Q_1, Q_2, Q_3$– First, second, and third quartiles

MD– Mean deviation

D_{10}– 10th decile

$a = S_x, \rho$

$d = \frac{Q_1+Q_2}{2}, Q_1, Q_2, Q_3, Q_3 - Q_1, MD, D_{10}$

$\theta = \frac{\bar{X}a}{\bar{X}a+d}$

b_j– Regression coefficient under OLS, Huber, Mallows, Schweppe GM, Sis GM

γ– Scalar constant

$S_k = \frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$

$R = \frac{\bar{Y}S_k}{\bar{X}S_k+Q_i}$

$R_1 = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1+QD}$

3. Existing Estimators

The classical estimator is

$$\bar{Y}_r \frac{\bar{Y}}{\bar{x}} X \quad (3.1)$$

The mathematical equations for MSE and Bias are obtained as

$$Bias(\bar{Y}_r) = \bar{Y}\left(C_x^2 - \rho_{yx}C_yC_x\right) \quad (3.2)$$

$$MSE(\bar{Y}_r) = \bar{Y}^2\left(C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\right) \quad (3.3)$$

Ijaz et al. [1] proposed robust regression-type ratio estimators using auxiliary information. The proposed estimators are as

$$t_s = \frac{\bar{y}}{(\bar{x})\gamma}(\bar{X})\gamma \quad (3.4)$$

$$t_s = \left[\frac{\bar{X}+(\beta_1-\beta_2)}{\bar{x}+(\beta_1-\beta_2)}\right] \quad (3.5)$$

The mathematical expression for the MSE of the estimator t_s is given as

$$MSE(t_s) = \frac{1-f}{n}\bar{Y}^2\left[C_y^2(1-\rho^2)\right] \quad (3.6)$$

$$MSE(t_s) = \frac{1-f}{n}\bar{Y}^2\left[C_y^2 + \sigma^2C_x^2 - 2\rho\theta C_yC_x\right] \quad (3.7)$$

Jeelani etal.[14] proposed the improved regression estimator using the coefficient of skewness and the quartile deviation

$$\bar{Y}_r = \frac{\bar{y}+b_{ols}(\bar{X}-\bar{x})}{(\bar{x}\beta_1+QD)}.(\bar{X}\beta_1 + QD) \quad (3.8)$$

The mathematical expression for $\bar{Y}_r$ was obtained as

$$MSE(\bar{Y}_r) = \frac{1-f}{n}\bar{Y}\left[R_1^2S_x^2 + S_y^2(1-\rho^2)\right] \quad (3.9)$$

Subzar et al. [4] suggested the following proposed ratio-type estimator to estimate the mean of the unknown population:

$$\bar{Y}_r = \frac{\bar{y}+b_j(\bar{X}-\bar{x})}{(\bar{x}S_K+Q_i)}.(\bar{x}S_K + Q_i) \quad (3.10)$$

The MSE of estimator $\bar{Y}_r$ is obtained as

$$MSE(\bar{Y}_r) \cong \frac{1-f}{n}\left[R^2S_x^2 + 2b_jR^2S_x^2 + b_j^2S_x^2 - 2RS_{xy} - 2b_jS_{xy} + S_y^2\right] \quad (3.11)$$

4. Methodology of the proposed estimators

Definition 4.1. The robust type of regression estimator is proposed by integrating the *cos* function with the auxiliary information. The estimator is proposed by developing a doubly robust structure because b_j will internally act as a robust estimator in the first part $\bar{y} + b_j(2\bar{X} - \bar{x})$ of the estimator and entire *Cos* function $cos\left[\frac{\bar{X}-\bar{x}}{2\bar{X}-\bar{x}} + \frac{\bar{x}a+d}{\bar{X}a+d}\right]$ will cap the influence of outliers of auxiliary variable as an

10.57647/mathsci.2025.1904.20

external robustness in the second part. This dual mechanism of robustness will yield a new estimator that will be more efficient for the heavy-tailed distributions and outliers. The proposed estimator is defined as

$$y_{ms} = \left(\bar{y} + b_j(2\bar{X} - \bar{x})\right)cos\left[\frac{\bar{X}-\bar{x}}{2\bar{X}-\bar{x}} + \frac{\bar{x}a+d}{\bar{X}a+d}\right] \quad (4.1)$$

Where

$$b_j = b_{OLS}, b_{Huber}, b_{mallows}, b_{schew}{}'_{sGM}, b_{SIS\ GM}$$

$$a = s_x, \rho \text{ and } d = \frac{Q_1 + Q_2}{2}, Q_1, Q_2, Q_3, Q_3 - Q_1, MD, D_{10}$$

The robust estimations of the regression coefficient b_j have been implemented in R using the packages MASS, robustbase, and rrcov. Specifically, rlm() with psi.huber and psi.bisquare was used for Huber and Mallows estimators, while lmrob() with the method= "SMDM" option was employed for Schweppe-type estimators. The objective functions, along with the tuning constant, are as follows.

Huber's M-Estimator

The objective function is defined by

$$min_{a,b}\sum_{i=1}^{n}\rho H\left(\frac{y_i - a - b_{xi}}{\sigma}\right)$$

where

$$\rho H(u) = \begin{cases} \frac{1}{2}u^2, & |u| \le k \\ k\left(|u| - \frac{1}{2}k\right), & |u| > k \end{cases} \text{ and } \psi_H(u) = \begin{cases} u, & |u| \le k \\ k\ sign(u), & |u| > k \end{cases}$$

tuning constant k = 1.345.

Mallows' GM-Estimator

The Mallows' objective function is given by

$$\rho M(u) = \begin{cases} \frac{c^2}{6}\left[1 - \left(1 - \left(\frac{u}{c}\right)^2\right)^3\right], & |u| < c, \\ \frac{c^2}{6}, & |u| \ge k \end{cases}$$

and

$$\psi_M(u) = \begin{cases} u\left(1 - \left(\frac{u}{c}\right)^2\right)^2, & |u| < c \\ 0, & |u| \ge c \end{cases}$$

tuning constant c = 4.685

Schweppe's GM-Estimator

Schweppe's used the following objective function

$$min_{a,b}\sum_{i=1}^{n}\rho S\left(\frac{y_i - a - b_{xi}}{\sigma_{wi}}\right)$$

with

$$\psi_S(u_i) = w_i\,\psi_M\left(\frac{u_i}{w_i}\right)$$

tuning constant c = 4.685.

$$\psi_{SIS}(u_i) = v_i\,\psi_M\left(\frac{u_i}{v_i}\right)$$

tuning constants are c = 4.685.

SIS GM-Estimator

The SIS GM-estimator utilizes the objective function

$$min_{a,b}\sum_{i=1}^{n}\rho SIS\left(\frac{y_i - a - b_{xi}}{\sigma_{Vi}}\right)$$

With

$$\psi_{SIS}(u_i) = v_i\,\psi_M\left(\frac{u_i}{v_i}\right)$$

tuning constants are c = 4.685 and c_1 = 0.2.

Now, to find the MSE of (4.1) under a simple random sampling without replacement (SRSWOR), we will utilize the following supposition

$$E(e_y) = 0, E(e_x) = 0, E(e_y e_x) = \frac{1-f}{n}\rho C_y C_x, E(e_x^2) = \frac{1-f}{n}C_x^2, E(e_y^2) = \frac{1-f}{n}C_y^2$$

$$\bar{y} = \bar{Y}(1 + e_y), \bar{x} = \bar{X}(1 + e_x)$$

By adding the above error terms in (4.1), the expression will be

$$y_{ms} = \left[\bar{Y}(1 + e_y) + b_j(\bar{X} - \bar{X}e_x)\right]cos\left[\frac{\bar{X}(1+e_x)}{\bar{X}-\bar{X}e_x} + \frac{\bar{X}a+d+a\bar{X}e_x}{\bar{X}a+d}\right] \quad (4.2)$$

$$y_{ms} = \left[\bar{Y}(1 + e_y) + b_j(\bar{X} - \bar{X}e_x)\right]cos\left[\frac{\bar{X}(1+e_x)}{\bar{X}(1-e_x)} + \frac{\bar{X}a+d\left(1+\frac{a}{\bar{X}}e_x\right)}{\bar{X}a+d}\right] \quad (4.3)$$

By simplifying the above expression, we get (4.3)

$$y_{ms} = \left[\bar{Y}(1 + e_y) + b_j(\bar{X} - \bar{X}e_x)\right]cos[(1 + e_x)(1 - e_x)^{-1} + (1 + \theta e_x)] \quad (4.4)$$

where

$$\theta = \frac{\bar{X}a}{\bar{X}a + d}$$

Expanding (4.4) using the inverse and cos series up to the first-order approximation, we have

$y_{ms} = \left[\bar{Y}\left(1+e_y\right) + b_j\bar{X}(1-e_x)\right]$

$\left[1 - 2\theta e_x - \frac{\theta^2 e_x^2}{2}\right]$ (4.5)

$y_{ms} = \left[\bar{Y}\left(1+e_y\right)(1-2\theta e_x)\right] + \left[b_j\bar{X}(1-e_x)\right][1 - 2\theta e_x]$ (4.6)

$y_{ms} = \left[\bar{Y} + \bar{Y}e_y + \bar{Y}2\theta e_x + 2\bar{Y}\theta e_x e_y\right] - b_j\bar{X}[(1-2\theta-1)e_x + 2\theta e_x^2\,]$ (4.7)

Subtracting $\bar{Y}$ from both sides, we have (4.7)
$(y_{ms} - \bar{Y}) = \left[\bar{Y}e_y + \bar{Y}2\theta e_x + 2\bar{Y}\theta e_x e_y\right] - b_j\bar{X}[(1 - 2\theta - 1)e_x + 2\theta e_x^2]$ (4.8)

Squaring both sides of (4.8), we get
$(y_{ms} - \bar{Y})^2 = \left[\left[\bar{Y}e_y + \bar{Y}2\theta e_x + 2\bar{Y}\theta e_x e_y\right] - b_j\bar{X}[(1 - 2\theta - 1)e_x + 2\theta e_x^2]\right]^2$ (4.9)

$$(y_{ms} - \bar{Y})^2 = \bar{Y}^2\left(e_y^2 + 4\theta e_x^2 + 4\theta e_x e_y\right) + b_j^2\bar{X}^{\,2}\frac{(1-(2\theta-1)e_x)^2}{2\bar{Y}b_j\bar{X}(4\theta(2\theta-1)e_x - (4\theta(2\theta-1)e_x^2}$$

Applying expectation on both sides and ignoring higher order terms to get MSE of y_{ms}, we have

$$MSE(y_{ms}) \cong \frac{1-f}{n}\left[\left[\bar{Y}^2\left(C_y^2 + 4\theta C_x^2 + 4\theta\rho C_x C_y\right)\right] + b_j^2\bar{X}^{\,2}\frac{(1-(2\theta-1)C_x)^2}{2\bar{Y}b_j\bar{X}(4\theta(2\theta-1))C_x - 4\theta(2\theta-1)C_x^2}\right] \quad (4.10)$$

Hence, the final expression for MSE is given as

$$MSE(y_{ms}) \cong \frac{1-f}{n}\left[\bar{Y}^2 C_y^2 + AC_x^2 + B\rho C_x C_y + b_j^2\bar{X}^{\,2}\right] \quad (4.11)$$

where

$A = \left[\bar{Y}^2 4\theta^2 - b_j^2\bar{X}^{\,2}(2\theta-1)^2 + 2\bar{Y}b_j\bar{X}(4\theta(2\theta-1)\right]$
$B = \left[\bar{Y}^2 4\theta - 2\bar{Y}b_j\bar{X}(4\theta(2\theta-1)\right]$

MSE measures the average deviation between the predicted values and the true value, while the Percentage relative efficiency (PRE) measures how efficient an estimator is as compared to a benchmark estimator.

The following mathematical expressions will be used to quantify the performance of each estimator.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$$

$$PRE = \left(\frac{MSE\ of\ referance\ estimator}{MSE\ of\ compared\ estimator}\right) \times 100$$

Generally, an estimator with a smaller MSE and a large PRE will be considered a preferable estimator.

10.57647/mathsci.2025.1904.20

5. Efficiency comparison of the OLS method with other robust regression estimators

In this section, a theoretical comparison between the OLS estimator with the robust regression estimator is presented. The robust regression estimator will lead to better performance if the following conditions are satisfied

$$MSE\left(y_{ms(b_{rob})}\right) < MSE\left(y_{ms(b_{ols})}\right) \quad (5.1)$$

That is,

$$\left[\bar{Y}^2C_y^2 + \bar{Y}^2 4\theta^2 - b_{rob}^2\bar{X}^{\,2}(2\theta-1)^2 + 2\bar{Y}b_{rob}\bar{X}(4\theta(2\theta-1)C_x^2\right] + \left[\bar{Y}^2 4\theta - 2\bar{Y}b_j\bar{X}(4\theta(2\theta-1)\right]\rho C_x C_y + b_{rob}^2\bar{X}^{\,2} <$$

$$\left[\bar{Y}^2C_y^2 + \bar{Y}^2 4\theta^2 - b_{ols}^2\bar{X}^{\,2}(2\theta-1)^2 + 2\bar{Y}b_{ols}\bar{X}(4\theta(2\theta-1)C_x^2\right] + [\bar{Y}^2 4\theta - 2\bar{Y}b_{ols}\bar{X}(4\theta(2\theta-1)]\rho C_x C_y + b_{ols}^2\bar{X}^{\,-2}$$

Let's simplify the above term using
Δ := LHS – RHS<0
By combining the like terms and canceling out identical terms on both sides, we determined

$$\Delta = -(b_{rob}^2 - b_{ols}^2)\bar{X}^{\,2}(2\theta-1)^2 + 2\bar{Y}\bar{X}(4\theta(2\theta-1)C_x^2(b_{rob} - b_{ols}) + \rho C_x C_y\left[\bar{Y}^2 4\theta - 2\bar{Y}b_{rob}\bar{X}(4\theta(2\theta-1) - [\bar{Y}^2 4\theta - 2\bar{Y}b_{ols}\bar{X}(4\theta(2\theta-1)]\right] + (b_{rob}^2 - b_{ols}^2)\bar{X}^{\,-2} \quad (5.2)$$

$$\Delta = -(b_{rob}^2 - b_{ols}^2)\bar{X}^{\,2}(2\theta-1)^2 + 2\bar{Y}\bar{X}(4\theta(2\theta-1)C_x^2(b_{rob} - b_{ols}) - 2\bar{Y}\bar{X}(4\theta(2\theta-1)\rho C_x C_y(b_{rob} - b_{ols})\rho C_x C_y + (b_{rob}^2 - b_{ols}^2)\bar{X}^{\,-2} \quad (5.3)$$

Group terms proportional to $(b_{rob}^2 - b_{ols}^2)$ and proportional to $(b_{rob} - b_{ols})$ factor terms

$$\Delta = (b_{rob}^2 - b_{ols}^2)[-\bar{X}^{\,2}(2\theta-1)^2 + \bar{X}^{\,2}] + 2\bar{Y}\bar{X}(4\theta(2\theta-1)\left(C_x^2 - \rho C_x C_y\right)(b_{rob} - b_{ols}) \quad (5.4)$$

$$\Delta = (b_{rob} - b_{ols})\left\{(b_{rob} + b_{ols})[-\bar{X}^{\,2}(2\theta-1)^2 + \bar{X}^{\,-2}] + 2\bar{Y}\bar{X}(4\theta(2\theta-1)\left(C_x^2 - \rho C_x C_y\right)\right\} \quad (5.5)$$

Since $\Delta < 0$, by plugging in and dividing the braces factor, we get

$$\left\{(b_{rob} - b_{ols}) < -\frac{2\bar{Y}\bar{X}(4\theta(2\theta-1)\left(C_x^2 - \rho C_x C_y\right)}{(b_{rob} + b_{ols})[-\bar{X}^{\,2}(2\theta-1)^2 + \bar{X}^{\,-2}]}\right\}$$

After a bit of simplification, we derived

$$(b_{rob} - b_{ols}) < \frac{-2\bar{Y}\bar{X}(4\theta(2\theta-1)\left(C_x^2 - \rho C_x C_y\right)}{[2\bar{X}^{\,2}(2\theta-1)^2 C_x^2]}$$

which simplifies

$$(b_{rob} - b_{ols}) < \left[\frac{-\bar{Y}(4\theta(2\theta-1)(C_x^2-\rho C_x C_y)}{[(2\theta-1)^2 C_x^2]}\right] \quad (5.6)$$

The above conditions will be satisfied if

$\theta(2\theta - 1)(C_x^2 - \rho C_x C_y) > 0$

and this condition will yield

$\theta(2\theta - 1) > 0 \qquad \Rightarrow \theta > \frac{1}{2}$

$C_x^2 - \rho C_x C_y > 0 \Rightarrow \rho < \frac{C_x}{C_y}, \quad C_x > 0, \ C_y > 0$

Fig. 1 shows that the OLS dominates over a very small region of the parameter space where $\rho \to 1$ *and* $\frac{C_x}{C_y} \to 0$, even when θ approaches 1. While robust dominates in wide range of parameters space where$\theta >$ 0.5 and $\frac{C_x}{C_y}$ is not extremely small.

Numerical Illustration

This section presents the significance of the proposed estimator as compared to the classical one by using the Real and simulated data sets. The real data set represents apple production amount (as a study variable) and the number of apple trees (as an auxiliary variable), which was recently cited by Subzar et al. [4]. Table 1, describes the summary statistics. Fig. 2 shows the Box plot and the scatter plot of the real data set, which demonstrates the presence of outliers in the data. Table 2 defines the theoretical conditions that guarantee supremacy of the robust regression estimator as compared to the OLS. Table 3 defines the MSE, while Table 4 and Table 5, respectively, explain the PRE of OLS vs robust regression estimator and Huber estimator vs other robust estimators. The results concluded that the robust estimators are superior to the OLS.

Simulated data

Simulated data is generated from auxiliary variable X using the heavy-tailed Laplace distribution with a mean $m = 50$ and $s = 10$ by using 1000 replications and 123 random seeds under SRSWOR.

The study variable Y is then simulated using the linear regression model, where the error term follows a heavy-tailed behavior. The regression coefficient b_j was computed using four robust methods by their corresponding objective and ψ function. The detail is given in section 4.

The summary statistics of the simulated data are given in Table 6. Table 7 demonstrates theoretical justification, and hence, the performance of the robust estimator will yield better results.

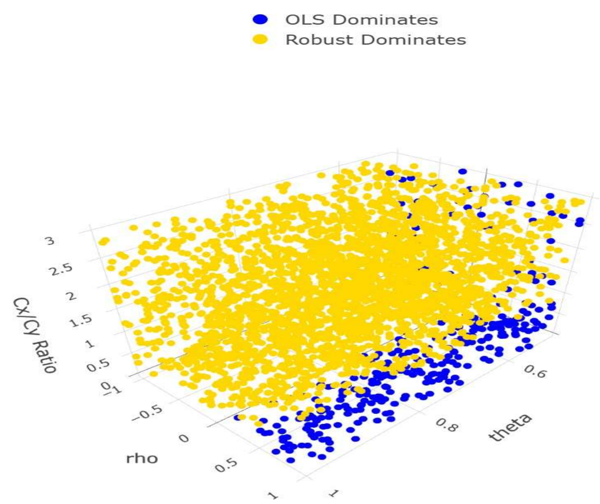
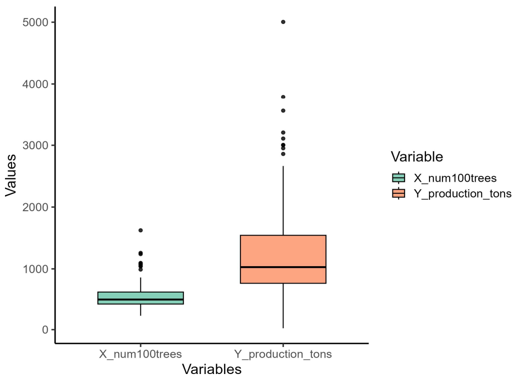


Figure 1. Parameter Space Dominance Plot for Robust and OLS

Table 1. Statistics of real data

$N = 117$	$Q_1 = 300.1$	$b_{huber} = 2.16$
$n = 40$	$Q_2 = 600.6$	$b_{mallows} = 1.01$
$\sigma^2 = 862$	$Q_3 = 701.0$	$b_{schewppe} = 0.96$
$S_Y = 235.5$	$c_{\bar{x}} = 0.9728$	$b_{SIS\ GM} = 0.85$
$\bar{X} = 560.0$	$C_X = 0.7395$	SK = 0.6654
$\bar{Y} = 1263$	$b_{ols} = 3.19$	$\rho = 0.987$



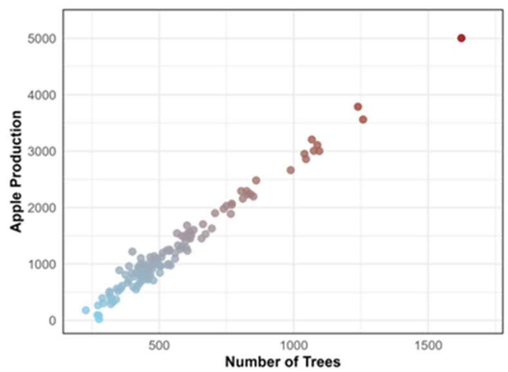


Figure 2. Box plot and scatter plot of the real data set

Table 2. Theoretical Conditions of real data

Estimator	Huber	Mallows	Schweppe	SIS GM
y_{m1}	$-1.03 < 30, 205, 086$	$-2.18 < 30, 205, 086$	$-2.23 < 30, 205, 086$	$-2.34 < 30, 205, 086$
y_{m2}	$-1.03 < 54, 957, 015$	$-2.18 < 54, 957, 015$	$-2.23 < 54, 957, 015$	$-2.34 < 54, 957, 015$
y_{m3}	$-1.03 < 1, 863, 307$	$-2.18 < 1, 863, 307$	$-2.23 < 1, 863, 307$	$-2.34 < 1, 863, 307$
y_{m4}	$-1.03 < 47, 124, 419$	$-2.18 < 4, 7124, 419$	$-2.23 < 47, 124, 419$	$-2.34 < 47, 124, 419$
y_{m5}	$-1.03 < 10, 322, 970$	$-2.18 < 10, 322, 970$	$-2.23 < 10, 322, 970$	$-2.34 < 10, 322, 970$
y_{m6}	$-1.03 < 389, 710, 155$	$-2.18 < 389, 710, 155$	$-2.23 < 389, 710, 155$	$-2.34 < 389, 710, 155$
y_{m7}	$-1.03 < 786, 377, 276$	$-2.18 < 786, 377, 276$	$-2.23 < 786, 377, 276$	$-2.34 < 786, 377, 276$

Table 3. MSE of the OLS and Robust Estimators

Estimator	Combinations	OLS	Huber	Mallows	Schweppe	SISGM
y_{m1}	$a = S_x, d = Q1 + Q2/2$	106015	77051.08	57584.96	57046.68	55952.83
y_{m2}	$a = S_x\ ,\ \ d = Q_1$	93943.65	87017.14	69371.81	68912.24	67991.41
y_{m3}	$a = S_x, d = Q_3 - Q_1$	108718.6	80031.07	60927.16	60405.91	59349.9
y_{m4}	$a = \rho,\ \ d = M.D$	105235.4	76223.4	56680.86	56138.72	55036.2
y_{m5}	$a = \rho,\ \ d = D_{20}$	121515.7	104520.7	96330.11	96232.1	96092.18
y_{m6}	$a = S_x,\ \ d = Q_2$	98418.97	69433.18	49618.04	49056.82	47910.21
y_{m7}	$a = S_x,\ \ d = Q_3$	93943.65	65300.36	45594.41	45031.39	43878.93

Table 4. PRE of the OLS VS Robust Estimators using real data

Estimator	OLS/Mallows GM	OLS/SchweppeGM	OLS/SIS GM
y_{m1}	184.101	185.839	189.472
y_{m2}	135.420	136.323	138.169
y_{m3}	178.440	179.980	183.182
y_{m4}	185.663	187.456	191.211
y_{m5}	126.145	126.273	126.457
y_{m6}	198.353	200.622	205.423
y_{m7}	206.042	208.618	214.097

Table 5. PRE of the Huber vs other Robust Estimators using real data

Estimator $y_{m(i)}$	Combinations	H.M / M.GM	H.M / S.GM	H.M / SIS.GM
y_{m1}	$a = S_x, d = Q1 + Q2/2$	133.804	135.066	137.707
y_{m2}	$a = S_x,\ \ d = Q_1$	125.435	126.272	127.982
y_{m3}	$a = S_x,\ \ d = Q_3\ Q_1$	131.355	132.488	134.846
y_{m4}	$a = \rho,\ \ d = M.D$	134.478	135.776	138.496
y_{m5}	$a = \rho,\ \ d = D_{10}$	108.502	108.613	108.771
y_{m6}	$a = S_x,\ \ d = Q_2$	139.935	141.536	144.923
y_{m7}	$a = S_x,\ \ d = Q_3$	143.220	145.010	148.819

Table 6. Statistics of Simulated data

N = 123	Q_1 = 31.79	b_{huber} = 0.4680
n = 100	Q_2 = 39.54	$b_{mallows}$ = 0.4599
$\bar{X}$ =49.79	Q_3 = 51.37	$b_{schweppe}$ = 0.4720
$\bar{Y}$ =41.79	C_Y = 0.4434	$b_{sis.gm}$ = 0.4610
D_{10} = 40.64	C_X = 0.2795	S_K = 0.6654
MD = 14.46	B_{ols} = 0.4891	ρ = 0.368

Table 7. Theoretical conditions of a simulated data

Estimator	Huber	Mallows	Schweppe	SIS GM
y_{m1}	−0.021 < 3, 516.485	−0.029 < 3, 516.485	−0.017 < 3, 516.485	−0.028 < 3, 516.485
y_{m2}	−0.021 < 3, 140.268	−0.029 < 3, 140.268	−0.017 < 3, 140.268	−0.028 < 3, 140.268
y_{m3}	−0.021 < 1, 688.393	−0.029 < 1, 688.393	−0.017 < 1, 688.393	−0.028 < 1, 688.393
y_{m4}	−0.021 < 2, 513.659	−0.029 < 2, 513.659	−0.017 < 2, 513.659	−0.028 < 2, 513.659
y_{m5}	−0.021 < 5, 332.519	−0.029 < 5, 332.519	−0.017 < 5, 332.519	−0.028 < 5, 332.519
y_{m6}	−0.021 < 3, 848.57	−0.029 < 3, 848.57	−0.017 < 3, 848.57	−0.028 < 3, 848.57
y_{m7}	−0.021 < 4, 616.499	−0.029 < 4, 616.499	−0.017 < 4, 616.499	−0.028 < 4, 616.499

Table 8. MSE of the OLS and Robust estimators using simulated data

Estimator	Combinations	OLS	Huber	Mallows	Schweppe	SISGM
y_{m1}	$a = S_x, d = Q1 + Q2/2$	1.784	1.690	1.699	1.708	1.618
y_{m2}	$a = S_x, \ d = Q_1$	1.844	1.750	1.758	1.768	1.677
y_{m3}	$a = S_x, \ d = Q_3$	Q_1	2.121	2.024	2.033	2.042
y_{m4}	$a = \rho, \ d = M.D$	1.952	1.857	1.866	1.875	1.784
y_{m5}	$a = \rho, \ d = D_{20}$	1.482	1.392	1.400	1.409	1.322
y_{m6}	$a = S_x, \ d = Q_2$	1.732	1.639	1.648	1.657	1.568
y_{m7}	$a = S_x, \ d = Q_3$	1.613	1.522	1.530	1.539	1.451

Table 9. PRE of the OLS vs Robust Estimators Using Simulated Data

Estimator	OLS/Mallows GM	OLS/Schweppe GM	OLS/SIS GM
y_{m1}	105.002	104.441	110.213
y_{m2}	104.864	104.319	109.919
y_{m3}	104.330	103.847	108.790
y_{m4}	104.637	104.118	109.438
y_{m5}	105.863	105.201	112.068
y_{m6}	105.128	104.552	110.482
y_{m7}	105.449	104.835	111.170

Table 10. PRE of the Huber vs other Robust estimators using simulated data

Estimator Yst(i)	Combinations	H.M / M.GM	H.M / S.GM	H.M / SIS.GM
y_{m1}	$a = S_x, d = Q1 + Q2/2$	99.515	98.984	104.454
y_{m2}	$a = S_d, \ d = Q_1$	99.529	99.011	104.326
y_{m3}	$a = S_d, \ d = (Q_1 Q_2)$	99.580	99.119	103.837
y_{m4}	$a = P, \ d = M.D$	99.551	99.057	104.118
y_{m5}	$a = P, \ d = D_w$	99.432	98.810	105.260
y_{m6}	$a = S_d, \ d = Q_2$	99.503	98.958	104.571
y_{m7}	$a = S_d, \ d = Q_3$	99.472	98.894	104.870

Table 8 represents the MSE of the simulated data, while Tables 9 and 10 define the PRE of the OLS vs Robust and Huber vs other robust estimators. The results conclude the dominance of the robust

10.57647/mathsci.2025.1904.20

regression estimators as compared to OLS.

8. Conclusion

This research has introduced a new class of Cos robust-type regression estimator to estimate the population mean under (SRSWOR).

The proposed estimators utilize auxiliary information with a cosine function and robust statistical techniques to address the limitations of traditional estimators, especially in the presence of outliers. The results of the analysis indicate that the proposed estimators outperform the conventional OLS estimator.

The results show that SIS GM estimation turned out to be the most efficient, having the lowest MSE and highest PRE values. It was also established that GM-estimates reduced the bias as compared to OLS, which provides an alternative.

The proposed estimator will become computationally complex as compared to the OLS, but will be highly applicable in an environment where outliers are common.

Acknowledgements

We would like to thank the reviewers comments for further improvement of the manuscript.

Author contributions

All the co-authors have contributed equally to all aspects of the preparation of this submission.

Conflict of interest statement

The authors declare that they have no known compet- ing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

No funds are available for the publication of this work.

Data availability

The data have been taken from the literature and has been used to verify the results of the proposed estimator.

References

[1] J. Cheng, J. Nakagawa, M. Yamamoto, T. Yamazaki, Uniqueness in an inverse problem for a one-dimensional fractional diffusion equation, Inverse problems, **25** (2009) 115002.

[2] M. Ijaz, A. U. Khan, S. M. Asim, et al., "A new ratio type estimator to estimate the population mean using auxiliary information," *Adv. Appl. Statist.*, vol. 63, no. 1, pp. 97–108, 2020.

[3] M. Subzar, S. Maqbool, and T. A. Raja, "A class of efficient ratio type estimators for the estimation of population mean using the auxiliary information in survey sampling," *Int. J. Eng. Dev. Res.*, vol. 5, no. 2, 2017.

[4] J. Subramani and G. Kumarapandiyan, "Estimation of population mean using coefficient of variation and median of an auxiliary variable," Int. J. Probab. Statist., vol. 1, pp. 111–118, 2012.

[5] M. Subzar, A. I. Al-Omari, and A. R. A. Alanzi, "The robust regression methods for estimating of finite population mean based on SRSWOR in case of outliers," CMC: Computers, Materials & Continua, vol. 65, no. 1, pp. 125–138, 2020.

[6] M. Ijaz, S. M. Asim, A. Ullah, and I. Mahariq, "Flexible robust regression-ratio type estimators and its applications," Math. Probl. Eng., vol. 2022, Article ID 8977392, 2022.

[7] C. Kadilar and H. Cingi, "Ratio estimators in simple random sampling," Appl. Math. Comput., vol. 151, no. 3, pp. 893–902, 2004.

[8] Z. Yan and B. Tian, "Ratio method to the mean estimation using coefficient of skew- ness of auxiliary variable," in Inf. Comput. Appl., vol. 106, pp. 103–110, 2010.

[9] J. Subramani and G. Kumarapandiyan, "Estimation of population mean using known median and coefficient of skewness," Amer. J. Math. Statist., vol. 2, pp. 101–107, 2012.

[10] J. Subramani and G. Kumarapandiyan, "Modified ratio estimators using known me- dian and coefficient of kurtosis," Amer. J. Math. Statist., vol. 2, pp. 95–100, 2012.

[11] S. K. Yadav, S. S. Mishra, A. K. Shukla, S. Kumar, and R. S. Singh, "Use of non- conventional measures of dispersion for improved estimation of population mean," Amer. J. Oper. Res., vol. 6, no. 3, pp. 69–75, 2006.

[12] D. S. Robson, "Applications of multivariate polykays to the theory of unbiased ratio- type estimation," J. Amer. Statist. Assoc., vol. 52, no. 280, pp. 511–522, 1957.

[13] M. Kumar, R. Singh, A. K. Singh, and F. Smarandache, "Some ratio type estimators under observational errors," World Appl. Sci. J., vol. 14, pp. 272–276, 2011.

[14] S. Mohanty, "Combination of regression and ratio estimate," J. Indian Statist. Assoc., vol. 5, pp. 16–19, 1967.

[15] M. Jeelani, S. Maqbool, and S. A. Mir, "Modified ratio estimators of population mean using linear combination of coefficient of skewness and quartile deviation," Int.

[16] J. Mod. Math. Sci., vol. 6, pp. 174–183, 2013.

[17] L. N. Upadhyaya and H. P. Singh, "Use of transformed auxiliary variable in estimating the finite population mean," Biometrical J., vol. 41, no. 5, pp. 627–636, 1999.

[18] R. Mukerjee, T. J. Rao, and K. Vijayan, "Estimators using multiple auxiliary in regression type formation," Aust. J. Statist., vol. 29, no. 3, pp. 244–254, 1987.

[19] M. Z. Hasan, M. Sultana, K. Fatema, Md. A. Hossain, and M. M. Hossain, "A new regression type estimator and its application in survey sampling," Open J. Statist., vol. 10, pp. 1010–1019, 2020.

[20] U. Shahzad, "Mean estimation using robust quantile regression with two auxiliary variables," Scientia Iranica, 2023.

[21] N. Ali, I. Ahmad, M. Hanif, and U. Shahzad, "An improved regression type estimator of population mean with two auxiliary variables and its variant using robust regression method," J. Comput. Appl. Math., vol. 382, Art. no. 113072, 2019.

[22] H. Ertaş, S. Toker, and S. Kaçıranlar, "Robust two-parameter ridge M-estimator for linear regression," J. Appl. Stat., vol. 42, no. 7, pp. 1490–1502, 2015.

[23] W. Babayemi, G. I. Onwuka, and A. B. Isah, "A Generalized Class of Log-Type Estimators of Finite Population Mean Based

on Correlation Coefficient," Int. J. Sci. Glob. Sustainability, vol. 9, no. 3, pp. 85–91, 2023.

[24] A. Adejumobi, M. A. Yunusa, Y. A. Erinola, and K. Abubakar, "An Efficient Log- arithmic Ratio Type Estimator of Finite Population Mean under Simple Random Sampling," Int. J. Eng. Appl. Phys., vol. 3, no. 2, pp. 700–705, 2023.

[25] G. R. V. Triveni, F. Danish, and M. Alrasheedi, "Application of Log-Type Estimators for Addressing Non-Response in Survey Sampling Using Real Datasets," Mathematics, vol. 13, no. 7, Art. 1089, 2025.

10.57647/mathsci.2025.1904.20

Appendix 1: Raw Data

Table 11. Raw Data for 117 Villages (*X*: Number of Apple Trees100, *Y*: Apple Production in Tons)

X	Y	X	Y	X	Y	X	Y
657.49	1453.29	482.02	1139.61	707.78	1901.05	1067.56	3206.25
466.78	800.25	466.79	1120.59	1095.24	3002.45	749.88	2030.29
431.78	1103.24	672.44	1531.98	432.64	724.43	432.31	815.50
579.22	1262.70	276.18	22.83	291.76	303.43	418.85	605.91
362.50	615.09	600.63	1282.39	374.72	808.35	320.35	288.71
1039.92	2952.20	468.09	1043.67	530.05	984.29	319.15	424.37
421.31	943.04	541.97	975.15	347.15	519.63	619.31	1513.29
413.65	842.60	457.92	722.84	413.51	550.83	1239.00	3788.34
508.35	1117.56	357.42	565.84	769.95	2069.47	339.71	372.96
569.63	1329.96	272.52	265.04	328.68	328.33	566.18	1543.81
739.51	1977.16	558.91	1094.02	485.70	1085.53	456.49	754.01
314.00	483.30	397.98	838.18	433.03	690.61	860.48	2482.01
609.50	1499.02	288.51	396.09	603.59	1683.25	443.96	809.77
403.54	593.32	695.50	1629.30	849.97	2197.56	810.86	2158.59
383.02	670.89	455.26	922.93	605.76	1542.90	828.03	2233.11
430.41	996.19	474.42	917.15	352.02	888.49	342.31	563.24
766.26	1889.42	988.96	2664.33	493.12	1088.21	839.04	2240.14
615.00	1560.61	407.72	778.41	614.92	1451.33	1074.89	3005.84
502.50	845.39	1088.00	3108.65	227.18	179.27	769.71	2050.28
535.39	1001.49	456.81	914.18	536.68	1232.14	270.36	94.15
469.05	955.71	613.61	1464.75	1045.77	2859.59	424.92	824.58
386.80	714.98	427.23	933.38	804.77	2293.28	605.00	1234.69
423.33	639.32	662.85	1705.94	538.14	1255.22	825.33	2292.92
400.27	1219.71	452.48	953.65	442.93	997.37	315.42	511.58
595.20	1480.41	584.82	1308.98	513.28	1199.35	466.72	818.93
320.06	363.34	438.77	757.80	450.23	877.76	387.57	960.75
478.31	709.18	628.09	1604.08	1257.97	3561.14	559.82	1196.97
583.79	1499.80	492.49	1028.01	275.75	86.77	504.93	964.96
528.07	1241.80	1623.62	5004.38	473.35	979.96	596.82	1402.34
502.79	966.04						

Appendix 2: Objective, ψ function, tuning constants, and Algorithms for Robust Estimation Methods

The algorithms and mathematical formulations for the robust regression estimators used in the simulation investigation are provided in this appendix. Each method uses a weighted least squares approach to iteratively update the regression coefficients while minimizing a robust objective function $\rho(\cdot)$.

Huber's M-Estimator

(1) Initialize $a^{(0)}, b^{(0)}$using ordinary least squares (OLS).

(2) Compute residuals$r_i^{(t)} = y_i - a^{(t)} - b_{xi}^{(t)}$

(3) Compute weights$w_i^{(t)} = \Psi_H\left(\frac{r_i^{(t)}}{\sigma}\right)\left(\frac{r_i^{(t)}}{\sigma}\right)$

(4) Update (*a, b*) by solving the weighted least squares:$min_{a,b}\sum_{i=1}^{n} w_i^{(t)}(y_i - a - b_{xi})^2$

Repeat steps 2–4 until convergence.

Mallows' GM-Estimator

(1) Start with initial OLS estimates ($a^{(0)}$, $b^{(0)}$).

(2) Compute standardized residuals $w_i^{(t)} = r_i^{(t)}/\sigma$

(3) Calculate weights $w_i^{(t)} = \Psi_M(u_i^t)/u_i^t$

(4) Update(a, b) using weighted least squares with weights $w_i^{(t)}$

(5) Iterate until parameter changes are negligible.

10.57647/mathsci.2025.1904.20

Schweppe's GM-Estimator

(1) Compute leverage weights $w_i = \sqrt[1]{h_{ii}}$
(2) Initialize (*a, b*) using OLS.
(3) Compute residuals $u_i^{(t)} = \left(y_i - a^{(t)} - b_{xi}^{(t)}\right)/(\sigma w_i)$
(4) compute$\psi_s(u_i^t) = w_i \psi_M(u_i^t)/w_i$ Update (*a, b*) via weighted least squares using weights.$w_i^{(t)} = \psi_s(u_i^t)/u_i^t$
(5) Repeat until convergence.

SIS GM-Estimator

(1) Compute initial S-estimates $a^{(0)}, b^{(0)}, \sigma^{(0)(0)}$).
(2) Determine leverage-based and residual-based weights $v_i^{(t)}$
(3) Compute standardized residuals $u_i^{(t)} = \left(y_i - a^{(t)} - b_{xi}^{(t)}\right)/(\sigma^t {v_i}^t)$
(4) Calculate $\psi_{sis}(u_i^t) = {v_i}^t \psi_M(u_i^t)/{v_i}^t$
(5) Update parameters using weighted least squares and update σ via robust scale equation.
(6) Iterate until convergence

10.57647/mathsci.2025.1904.20

10.57647/mathsci.2025.1904.20