Family of Estimators for Estimating Population Median using Auxiliary Information in Survey Sampling

Prayas Sharma*, Anupam Lata, Subhash Kumar Yadav and Muhammad Noor-ul-Amin

Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, India
COMSATS University Islamabad-Lahore Campus, Pakistan
E-mail: prayassharma02@gmail.com; anupam.lata010@gmail.com; drskystats@gmail.com; nooramin.stats@gmail.com
**Corresponding Author*

Received 07 June 2025; Accepted 08 August 2025

Abstract

In this article, we propose a new family of estimators for estimating the unknown population median of a study variable by utilizing auxiliary information under simple random sampling. The choice of the median, as opposed to the mean, is particularly advantageous in the presence of outliers or skewed distributions, where the mean may be unduly influenced. We derive the expressions for the bias and mean square error (MSE) of the proposed class of estimators up to the first order of approximation. Furthermore, we examine several notable subclasses within the proposed family and calculate their respective MSEs. To assess the efficiency and robustness of the proposed estimators, an empirical study is conducted using real-world

Journal of Reliability and Statistical Studies, Vol. 18, Issue 2 (2025), 343–370.
doi: 10.13052/jrss0974-8024.1824

data and benchmarked against existing estimators from the literature. The results of this empirical analysis demonstrate that the proposed estimators achieve lower MSEs, underscoring their practical relevance and effectiveness in survey sampling applications.

Keywords: Auxiliary variable, bias, mean square error, median, simple random sampling, study variable.

1 Introduction

The median is often preferred over the mean when dealing with non-normal distributions such as income, production and consumption, where outliers are commonly present. In such cases, the mean becomes less reliable due to its sensitivity to extreme values. To enhance the precision of estimators, auxiliary information can be effectively utilized in statistical analysis. Several methods – such as ratio, product, and regression techniques – have been developed to incorporate auxiliary information for estimating population parameters. Extensive research has been conducted on estimating parameters like the population mean and median using such information. [4] was among the first to introduce the concept of median estimation in survey sampling. Later, [8] was the first to address the problem of estimating the median using auxiliary information in survey sampling. Since then, a wide variety of estimators have been proposed under different sampling schemes to estimate the population median. [23] proposed an estimator for population median estimation and demonstrated its superiority over existing methods. [20] introduced a generalized estimator using auxiliary information and derived its bias and mean square error (MSE), showing that their estimator performed better in comparison. [16] suggested an estimator for population median estimation under two sampling (SRSWOR and stratified sampling). The estimator they proposed demonstrated superior performance compared to the available ones in the numerical analysis. In a similar way, [17] had given an estimator of difference type for evaluating population median by using sampling scheme (SRSWOR and Two-phase Sampling). [1] had put forward a new estimator and showed its potential through empirical and stimulation study. [22] proposed a ratio estimator based on the known quartiles of the auxiliary variable, and demonstrated its improved performance through both simulation and empirical analyses. [23] proposed an estimator aimed at estimating the population median, wherein the Asymptotic Optimum Estimator (AOE) was examined for each estimator class. [19] introduced a novel estimator for estimating the population mean within a stratified sampling framework and examined several of its subset forms. Empirical evidence substantiated the estimator's efficiency. Furthermore, when the auxiliary variable x is negatively correlated with the study variable y, the product estimator is generally more appropriate than the ratio estimator. [10] had given product method of estimation for estimating population mean. In a similar way [12] and [13] had given alternative estimator and showed that their estimator had minimum MSE. [14] developed a more effective estimator than those of [13] and [25], suitable for cases where the correlation between y and x changes. Important references related to our work include [5, 7, 18, 21, 27] and [3].

The objective of this research is to construct more efficient estimators for median estimation under the influence of extreme observations, making use of auxiliary data. The uniqueness of this study lies in the following points –

1. The flexibility of the proposed estimator allows for the formulation of multiple estimator forms.
2. Rigorous empirical analysis has been undertaken to examine its behavior in scenarios involving significant skewness.

2 Terminologies

Let the population consist of N distinguishable units. Let y_i be the study variable, and x_i the primary auxiliary variable. Suppose z_i is a transformed auxiliary variable, with the relation $z_i = f(x_i)$, where f is a known function. A sample of size n is selected from the target population using the SRSWOR scheme, ensuring that each unit has an equal chance of selection without replacement. Let $\hat{M}_y$ be the sample medians corresponding to the population median M_y with density function $f_y(M_y)$ and $(\hat{M}_x, \hat{M}_z)$ be the sample median corresponding to the population medians (M_x, M_z) with density functions $\{f_x(M_x), f_z(M_z)\}$. Let $\rho_{yx} = \rho_{(\hat{M}_y,\hat{M}_x)} = 4P_{11}(y, x) - 1$, $\rho_{yz} = \rho_{(\hat{M}_y,\hat{M}_z)} = 4P_{11}(y, z) - 1$ and $\rho_{xz} = \rho_{(\hat{M}_x,\hat{M}_z)} = 4P_{11}(x, z) - 1$, be the population correlation coefficients between sample medians represented by their respective subscripts, where $P_{11}(y, x) = P(y \leq M_y \cap x \leq M_x), P_{11}(y, z) = P(y \leq M_y \cap z \leq M_z)$, and $P_{11}(x, z) = P(x \leq M_x \cap z \leq M_z)$. To determine the attributes of estimators, we define the relative error terms as follows: Let $e_0 = (\hat{M}_y - M_y){M_y}^{-1}$, $e_1 = (\hat{M}_x - M_x){M_x}^{-1}$ such that $E(e_i) = 0(i = 0, 1)$, $E(e_0^2) = \lambda C^2_{M_y}$, $E(e_1^2) = \lambda C^2_{M_x}$, $E(e_0e_1) = \lambda C_{M_{yx}}$ where $C_{M_{yx}} = \rho_{yx} C_{M_y} C_{M_x}$, $C_{M_y} = 1/[M_y f_y(M_y)]$, $C_{M_x} = 1/[M_x f_x(M_x)]$, $C_{M_z} = 1/[M_z f_z(M_z)]$, and $\lambda = 0.25(\frac{1}{n} - \frac{1}{N})$.

3 Existing Estimators

In this section, we discuss some prominent estimators commonly employed for the estimation of the population median –

The estimator provided by [4] is given by:

$$\hat{M}_0 = \hat{M}_y. \tag{1}$$

$\hat{M}_0$'s variance is given by:

$$V(\hat{M}_0) = \lambda M_y^2 C_{M_y}^2. \tag{2}$$

The estimation methods outlined in this section are based predominantly on the use of a single auxiliary variable. The ratio estimator given by [8] as follows:

$$\hat{M}_R = \hat{M}_y \left(\frac{M_x}{\hat{M}_x} \right). \tag{3}$$

In survey sampling, the efficiency of an estimator often depends on the nature of the relationship between the study variable y and the auxiliary variable x. When y and x are linearly related, strongly correlated, and the regression line passes through the origin, ratio estimators perform optimally. Conversely, if the regression line does not passes through the origin, then it would be good to use the regression estimator. Up to the first order of approximation, the bias and MSE of $\hat{M}_R$ are given by:

$$Bias(\hat{M}_R) \cong \lambda\{M_y C_{M_x}^2 - M_y C_{M_{yx}}\},$$

$$MSE(\hat{M}_R) \cong \lambda\{M_y^2 C_{M_y}^2 + M_y^2 C_{M_x}^2 - 2M_y^2 C_{M_{yx}}\}. \tag{4}$$

The ratio estimator $\hat{M}_R$ exhibits greater efficiency compared to $\hat{M}_0$ when ρ_{yx} exceeds $\frac{1}{2}\frac{C_{M_x}}{C_{M_y}}$.

[2] presented an estimator of the exponential ratio type, which is defined by

$$\hat{M}_E = \hat{M}_y \exp\left(\frac{M_x - \hat{M}_x}{M_x + \hat{M}_x} \right). \tag{5}$$

This approach is suitable when the regression line of y on x is linear and passes through the origin. The bias and MSE, up to the first-order approximation, are given by:

$$Bias(\hat{M}_E) \cong M_y \lambda \left\{ \frac{3}{8} C_{M_x}^2 - \frac{1}{2} C_{M_{yx}} \right\}$$

and

$$MSE(\hat{M}_E) \cong M_y^2 \lambda \left\{ C_{M_y}^2 + \frac{1}{4} C_{M_x}^2 - C_{M_{yx}} \right\}. \tag{6}$$

The estimator $\hat{M}_E$ based on the exponential ratio method outperforms estimators $\hat{M}_0$ and $\hat{M}_R$ under the condition that $\rho_{yx} > \frac{1}{4}\frac{C_{M_x}}{C_{M_y}}$ and $\rho_{yx} < \frac{3}{4}\frac{C_{M_x}}{C_{M_y}}$.

The idea of difference estimator was first introduced by [6] and is given by:

$$\hat{M}_{D_0} = \hat{M}_y + d_0(M_x - \hat{M}_x), \tag{7}$$

where d_0 is constant.

The minimum MSE of $(\hat{M}_{D_0})_{min}$ is as follows:

$$MSE(\hat{M}_{D_0})_{min} \cong C_{M_y}^2 \lambda (M_y^2 - M_y^2 \rho_{yx}^2). \tag{8}$$

The value of d_0 that minimizes the mean squared error is given by $d_0(opt) = \frac{M_y \rho_{yx} C_{M_y}}{M_x C_{M_x}}$.

Since ρ_{yx}^2 is always greater than 0, the minimum MSE of the estimator for $\hat{M}_{D_0}$ is consistently smaller than the MSE corresponding to $\hat{M}_0$, $\hat{M}_R$, $\hat{M}_E$.

[5, 11] and [15] had given a few more estimators, which are as follows:

$$\hat{M}_{D_1} = d_1 \hat{M}_y + d_2(M_x - \hat{M}_x), \tag{9}$$

here d_1 and d_2 are constant.

The minimum MSE of $\hat{M}_{D_1}$, at optimum value of their constant is given by:

$$MSE(\hat{M}_{D_1})_{min} \cong M_y^2 \left\{ 1 - \frac{B_0}{A_0 B_0 - C_0^2 + B_0} \right\}, \tag{10}$$

The optimum values of d_1 and d_2 are as follows:

$$d_{1(opt)} = \frac{B_0}{A_0 B_0 - C_0^2 + B_0}, \quad d_{2(opt)} = \frac{M_y}{M_x} \frac{C_0}{A_0 B_0 - C_0^2 + B_0},$$

where $A_0 = \lambda C_{M_y}^2$, $B_0 = \lambda C_{M_x}^2$, $C_0 = \lambda C_{M_{yx}}$.

$$\hat{M}_{D_2} = \{d_3 \hat{M}_y + (d_4 M_x - d_4 \hat{M}_x)\}(M_x \hat{M}_x^{-1}), \tag{11}$$

here, d_3 and d_4 are constant.

The minimum MSE of $\hat{M}_{D_2}$, at optimum value of their constant is given by:

$$MSE(\hat{M}_{D_2})_{min} \cong M_y^2 \left\{ 1 - \frac{\begin{array}{c} A_1B_1^2 + B_1C_1^2 - 2B_1C_1D_1 \\ +2B_1C_1 + B_1^2 - 2B_1D_1 + B_1 \end{array}}{A_1B_1 - D_1^2 + B_1} \right\}, \quad (12)$$

The optimum values of d_3 and d_4 are as follows:

$$d_{3(opt)} = \frac{B_1(C_1 - D_1 + 1)}{A_1B_1 - D_1^2 + B_1}, \quad d_{4(opt)} = \frac{M_y}{M_x} \frac{(A_1B_1 - C_1D_1 + B_1 - D_1)}{(A_1B_1 - D_1^2 + B_1)},$$

where, $A_1 = \lambda(C_{M_y}^2 + 3C_{M_x}^2 - 4C_{M_{yx}})$, $B_1 = \lambda C_{M_x}^2$, $C_1 = \lambda(C_{M_x}^2 - C_{M_{yx}})$, $D_1 = \lambda(2C_{M_x}^2 - C_{M_{yx}})$.

$$\hat{M}_{D_3} = \{d_5\hat{M}_y + (d_6M_x - d_6\hat{M}_x)\} \exp\left(\frac{M_x - \hat{M}_x}{M_x + \hat{M}_x}\right), \quad (13)$$

here, d_5 and d_6 are constant.

The minimum MSE of $\hat{M}_{D_3}$, at optimum value of their constant is given by:

$$MSE(\hat{M}_{D_3})_{min} \cong M_y^2 \left\{ 1 - \frac{\begin{array}{c} A_2D_2^2 + B_2C_2^2 - 2C_2D_2E_2 \\ +2B_2C_2 + D_2^2 - 2D_2E_2 + B_2 \end{array}}{A_2B_2 - E_2^2 + B_2} \right\}, \quad (14)$$

The optimum values of d_5 and d_6 are as follows:

$$d_{5(opt)} = \frac{(B_2C_2 - D_2E_2 + B_2)}{(A_2B_2 - E_2^2 + B_2)},$$

$$d_{6(opt)} = \frac{M_y}{M_x} \frac{(A_2D_2 - C_2E_2 + D_2 - E_2)}{(A_2B_2 - E_2^2 + B_2)},$$

where, $A_2 = \lambda(C_{M_y}^2 + C_{M_x}^2 - 2C_{M_{yx}})$, $B_2 = \lambda C_{M_x}^2$, $C_2 = \lambda(\frac{3C_{M_x}^2}{8} - \frac{1}{2}C_{M_{yx}})$, $D_2 = \lambda C_{M_x}^2/2$, $E_2 = \lambda(C_{M_x}^2 - C_{M_{yx}})$.

$$\hat{M}_{D_4} = \{d_7\hat{M}_y + (d_8M_x - d_8\hat{M}_x)\} \exp(M_x\hat{M}_x^{-1} - 1) \quad (15)$$

here, d_7 and d_8 are constant.

The minimum MSE of $\hat{M}_{D_4}$, at optimum value of their constant is given by:

$$MSE(\hat{M}_{D_4})_{min} \cong M_y^2 \left\{ 1 - \frac{\begin{array}{c} A_3B_3^2 + B_3C_3^2 - 2B_3C_3D_3 \\ +B_3^2 + 2B_3C_3 - 2B_3D_3 - B_3 \end{array}}{A_3B_3 - D_3^2 + B_3} \right\} \quad (16)$$

The optimum values of d_7 and d_8 are as follows:

$$d_{7(opt)} = \frac{B_3(C_3 - D_3 + 1)}{A_3B_3 - D_3^2 + B_3}, \quad d_{8(opt)} = \frac{M_y}{M_x} \frac{(A_3B_3 - C_3D_3 + B_3 - D_3)}{(A_3B_3 - D_3^2 + B_3)},$$

where, $A_3 = \lambda(C_{M_y}^2 + 4C_{M_x}^2 - 4C_{M_{yx}})$, $B_3 = \lambda C_{M_x}^2$, $C_3 = \lambda(\frac{3C_{M_x}^2}{2} - C_{M_{yx}})$, $D_3 = \lambda(2C_{M_x}^2 - C_{M_{yx}})$.

[16] introduced the estimator for population median is as follows:

$$\begin{aligned} \hat{M}_{pp}^G &= [m_1\hat{M}_y + (m_2M_x - m_2\hat{M}_x)] \\ &\times \left[\left(\frac{aM_x + b}{a\hat{M}_x + b} \right)^{\alpha_1} \exp\left\{ \frac{\alpha_2(aM_x - a\hat{M}_x)}{\left\{ a(\gamma - 1)M_x + a\hat{M}_x \right\} + 2b} \right\} \right] \end{aligned} \quad (17)$$

Here, unknown population parameters are a and b and scalar quantities are α_1, α_2 and γ, which can take values such as 0 and 1, where m_1 and m_2 are constants.

$$Bias(\hat{M}_{pp}^G) \cong (m_1 - 1)M_y + m_2M_y\lambda \left\{ \frac{3}{2}C_{M_x}^2 - C_{M_y} \right\} + m_2M_x\lambda C_{M_x}^2$$

At optimum value, the MSE of $\hat{M}_{pp}^G$ is given by:

$$\begin{aligned} MSE(\hat{M}_{pp}^G)_{min} &\cong \frac{M_y^2}{1 + \lambda C_{M_y}^2(1 - \rho_{yx}^2)} \\ &\times \left[\lambda C_{M_y}^2(1 - \rho_{yx}^2) - \frac{\lambda^2 C_{M_x}^4}{4} - \lambda^2 C_{M_y}^2 C_{M_x}^2 (1 - \rho_{yx}^2) \right] \end{aligned} \quad (18)$$

The m_1 and m_2's optimum values are stipulated as:

$$m_{1(opt)} = \frac{1 - 0.5\lambda C^2_{M_x}}{1 + \lambda C^2_{M_y}(1 - \rho^2_{yx})},$$

$$m_{2(opt)} = \frac{M_y}{M_x}\left[1 + m_{1(opt)}\left\{\frac{\rho_{yx}C_{M_y}}{C_{M_x}} - 2\right\}\right]$$

[9] stipulated median estimator is given by:

$$\hat{M}_{sm} = \hat{M}_y\left\{m_{12}\left(\frac{M_x}{\hat{M}_x}\right) + m_{13}\left(\frac{\hat{M}_x}{M_x}\right)\right\}$$
$$\times\left[\alpha\exp\left(\frac{M_x - \hat{M}_x}{M_x + \hat{M}_x}\right) + (1-\alpha)\exp\left(\frac{\hat{M}_x - M_x}{M_x + \hat{M}_x}\right)\right], \quad (19)$$

where m_{12} and m_{13} are constant.

$$\begin{aligned}\text{Bias}(\hat{M}_{sm}) = M_y\Bigg[&(m_{12} + m_{13} - 1)\\ &+ m_{12}\lambda\left\{\left(\frac{3}{8} + \frac{3\alpha}{2}\right)C^2_{M_x} - \left(\alpha + \frac{1}{2}\right)C_{M_{yx}}\right\}\\ &+ m_{13}\lambda\left\{\left(\frac{3}{2} - \alpha\right)C_{M_{yx}} + \left(\frac{3}{8} - \frac{\alpha}{2}\right)C^2_{M_x}\right\}\Bigg]\end{aligned} \quad (20)$$

At optimum value, the MSE of $\hat{M}_{sm}$ is evaluated as:

$$MSE(\hat{M}_{sm})_{min} = M^2_y\left[1 - \frac{A_1A^2_5 - 2A_3A_4A_5 + A_2A^2_4}{A_1A_2 - A^2_3}\right] \quad (21)$$

m_{12} and m_{13}'s optimum value are given by:

$$m_{12(opt)} = \left[\frac{A_2A_4 - A_3A_5}{A_1A_2 - A^2_3}\right], \quad m_{13(opt)} = \left[\frac{A_1A_5 - A_3A_4}{A_1A_2 - A^2_3}\right],$$

where

$$A_1 = 1 + \lambda\{C^2_{M_y} + (\alpha^2 + 4\alpha + 1)C^2_{M_x} - 2(2\alpha + 1)C_{M_{yx}}\},$$

$$A_2 = 1 + \lambda\{C^2_{M_y} + (\alpha^2 - 4\alpha + 3)C^2_{M_x} + 2(3 - 2\alpha)C_{M_{yx}}\},$$

$$A_3 = 1 + \lambda\{C^2_{M_y} + 2(1-2\alpha)C_{M_{yx}} + \alpha^2 C^2_{M_x}\},$$
$$A_4 = 1 + \lambda\left\{\left(\frac{3}{8} + \frac{3\alpha}{2}\right)C^2_{M_x} - (\alpha + 0.5)C_{M_{yx}}\right\}$$

and

$$A_5 = 1 + \lambda\left\{\left(\frac{3}{2} - \alpha\right)C_{M_{yx}} + \left(\frac{3}{8} - \frac{\alpha}{2}\right)C^2_{M_x}\right\}.$$

The estimator proposed by [1] is given by:

$$\hat{M}^*_y = \left[\alpha_1 \hat{M}_y \left\{\frac{1}{2}\left(\frac{M_x}{\hat{M}_x} + \frac{\hat{M}_x}{M_x}\right)\right\}\right.$$
$$\left. + (\alpha_2 M_x - \alpha_2 \hat{M}_x)\right] \exp\left(\frac{M_x - \hat{M}_x}{M_x + \hat{M}_x}\right),$$
$$Bias(\hat{M}^*_y) = \left[\alpha_1\left(M_y + \frac{7\lambda M_y C^2_{M_x}}{8} - 0.5\lambda M_y C_{M_{yx}}\right)\right.$$
$$\left. + \alpha_2 M_x \left(\frac{\lambda C^2_{M_x}}{2}\right) - M_y\right] \quad (22)$$

where α_1 andα_2 are constant. At optimum value, the MSE of $\hat{M}^*_y$ is stipulated as:

$$MSE(\hat{M}^*_y)_{min}$$
$$= M^2_y - \frac{M^2_y[64 + \lambda C^2_{M_x}(64 + \lambda\{25C^2_{M_x} - 16C^2_{M_y}(\rho^2_{yx} - 1)\})]}{64[1 + \lambda C^2_{M_x} + \lambda C^2_{M_y}\{1 - \rho^2_{yx}\}]} \quad (23)$$

α_1 and α_2's optimum values are given by:

$$\alpha_{1(opt)} = \frac{8 + 3\lambda C^2_{M_x}}{8(1 + \lambda C^2_{M_x} + C^2_{M_y}(\lambda - \lambda\rho^2_{yx}))},$$
$$\alpha_{2(opt)} = \frac{\begin{array}{c} M_y(\lambda C^3_{M_x} + 8C_{M_y}\rho_{yx} + 3\lambda C^2_{M_x} C_{M_y}\rho_{yx} \\ -4C_{M_x}(1 + \lambda C^2_{M_y}(\rho^2_{yx} - 1))) \end{array}}{8C_{M_x} M_x(1 + \lambda C^2_{M_x} + C^2_{M_y}(\lambda - \lambda\rho^2_{yx}))}.$$

4 The Suggested Family of Estimators

In simple random sampling, we proposed a family of estimators for the population median of the study variable Y as:

$$\hat{M}_{t_m} = \psi\left[\hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\alpha\hat{M}_z + (1-\alpha)M_z}\right) + J_2(M_z - \hat{M}_z)\right] \times \left(\frac{M_z}{\alpha\hat{M}_z + (1-\alpha)\,M_z}\right)^g \tag{24}$$

where $M_z = aM_x + b$ and $\hat{M}_z = a\hat{M}_x + b$, Let $a \neq 0$, and b be either a real number or a function of known parameters of the auxiliary variable X, such as the median Q_2, quartile deviation, standard deviation S_x, coefficient of variation C_x, skewness $\beta_{1(x)}$, kurtosis $\beta_{2(x)}$, or coefficient of correlation ρ_{yx}. Here, $(J_1, J_2, \psi, a, b,$ and $g)$ are constants, while J_1 & J_2 are determined so as to minimize the mean squared error (MSE) of the estimator $\hat{M}_{t_m}$.

Now expressing the generalized class of estimators (24) in terms of e's, we get

$$\hat{M}_{t_m} = \psi\{M_y(1+e_0) + J_1M_y(1+e_0)(1+\alpha\beta e_1)^{-1} - J_2aM_xe_1\}\{1+\alpha\beta e_1\}^{-g} \tag{25}$$

where $\beta = \frac{aM_x}{aM_x+b}$, expanding the Equation (25), and neglecting terms of e's having power greater than two and subtracting M_y on both sides we obtain –

$$\begin{aligned}\hat{M}_{t_m} - M_y = M_y[&\psi\{1 - \alpha\beta ge_1 + 0.5g(g+1)\alpha^2\beta^2e_1^2 + e_0 - \alpha\beta ge_1e_0\}\\ &+ \psi J_1\{1-(g+1)\alpha\beta e_1 + (0.5g(g+1)+g+1)\alpha^2\beta^2e_1^2\\ &+ e_0 - (g+1)\alpha\beta e_1e_0\} - 1 - \psi J_2aM_x\{e_1 - \alpha\beta ge_1^2\}],\end{aligned} \tag{26}$$

By taking the expectation of both sides of Equation (25), the first-order approximation of the bias of $\hat{M}_{t_m}$ is derived as:

$$\begin{aligned}\text{Bias}(\hat{M}_{t_m}) = M_y[&\psi\{1 + 0.5g(g+1)\alpha^2\beta^2\lambda C_{M_x}^2 - \alpha\beta g\lambda C_{M_{yx}}\}\\ &+ \psi J_1\{1 + 0.5g(g+1)\alpha^2\beta^2\lambda C_{M_x}^2 + (g+1)\alpha^2\beta^2\lambda C_{M_x}^2\\ &- (g+1)\alpha\beta\lambda C_{M_{yx}}\} - 1 + \psi J_2aM_x\alpha\beta g\lambda C_{M_x}^2].\end{aligned} \tag{27}$$

To derive the MSE of $\hat{M}_{t_m}$, we square both sides of Equation (26) and take expectation, neglecting higher-order terms of e's (i.e., terms with powers greater than two).

The MSE of $\hat{M}_{t_m}$ to the first order of approximation is given by:

$$MSE(\hat{M}_{t_m}) = \psi^2\{J_1^2 A - 2J_1J_2B + 2J_1C + J_2^2D - 2J_2E + F\} - 2\psi\{J_1G + J_2H + I\} + M_y^2 \tag{28}$$

where,

$$A = M_y^2 - 4\alpha(g+1)\Lambda k_3 + k_2 + \{(g+1)^2\Lambda^2\alpha^2k_1 + 2(g+1)\Lambda^2\alpha^2k_1 + g(g+1)\Lambda^2\alpha^2k_1\},$$

$$B = ak_3 - (2ga\alpha\Lambda k_1 + a\alpha\Lambda k_1),$$

$$C = M_y^2 - 2\alpha\Lambda k_3(2g+1) + k_2 + \alpha^2\Lambda^2k_1(3g+1)(g+1),$$

$$D = a^2k_1,$$

$$E = ak_3 - 2a\alpha g\Lambda k_1,$$

$$F = M_y^2 + k_2 + (g^2\alpha^2\Lambda^2k_1 + g(g+1)\alpha^2\Lambda^2k_1) - 4\alpha g\Lambda k_3,$$

$$G = M_y^2 - \alpha(g+1)\Lambda k_3 + \alpha^2\Lambda^2k_1\{(g+1) + 0.5g(g+1)\},$$

$$H = a\alpha g\Lambda k_1,$$

$$I = M_y^2 - \alpha g\Lambda k_3 + 0.5g(g+1)\alpha^2\Lambda^2k_1.$$

Now differentiating MSE of $\hat{M}_{t_m}$ with respect to J_1 & J_2 we obtain the values of $(J_1)_{opt}$ and $(J_2)_{opt}$ as follows:

$$(J_1)_{opt} = \frac{(DG + BH) - \psi(CD - BE)}{\psi(AD - B^2)}$$

and

$$(J_2)_{opt} = \frac{(BG + AH) - \psi(BC - AE)}{\psi(AD - B^2)}$$

The resulting MSE $(\hat{M}_{t_m})$ at $(J_1)_{opt}$ & $(J_2)_{opt}$ is stipulated by:

$$MSE(\hat{M}_{t_m}) = \frac{\{\psi^2L_1 - 2\psi L_2 + L_3 + (AD - B^2)^2M_y^2\}}{(AD - B^2)^2} \tag{29}$$

Now we differentiate Equation (29) with respect to ψ and we obtain the value of $(\psi)_{opt}$ as follows:

$$(\psi)_{opt} = \frac{L_2}{L_1}$$

Now we obtain minimum MSE of $\hat{M}_{t_m}$ by substituting value of ψ in Equation (29)

$$MSE(\hat{M}_{t_m})_{\min} = \frac{\{L_3 - \frac{L_2^2}{L_1} + (AD - B^2)^2 M_y^2\}}{(AD - B^2)^2} \tag{30}$$

where,

$$\begin{aligned}
L_1 &= (CD - BE)^2 A - 2(CD - BE)(BC - AE)B \\
&\quad - 2(CD - BE)(AD - B^2)C + (BC - AE)^2 D \\
&\quad + 2(BC - AE)(AD - B^2)E + (AD - B^2)^2 F, \\
L_2 &= (DG + BH)(CD - BE)A - (CD - BE)(BG + AH)B \\
&\quad - (DG + HB)(BC - AE)B - (DG + BH)(AD - B^2)C \\
&\quad + (BG + AH)(BC - AE)D + (BG + AH)(AD - B^2)E \\
&\quad - (CD - BE)(AD - B^2)G - (BC - AE)(AD - B^2)H \\
&\quad + (AD - B^2)^2 I \\
L_3 &= (DG + BH)^2 A - 2(DG + BH)(BG + AH)B + (BG + AH)^2 D \\
&\quad - 2(DG + BH)(AD - B^2)G - 2(BG + AH)(AD - B^2)H
\end{aligned}$$

Now we obtain subsets from the generalized estimator $\hat{M}_{t_m}$ by using suitable values of (J_1, J_2, ψ, a, b, and g) are as follows:

Tables 1, 2, 3 and 4 display several members of the proposed estimator family, obtained by assigning different values to the constants. From estimator $\hat{M}_{t_{m_1}}$ to estimator $\hat{M}_{t_{m_{36}}}$, the value of ψ is taken as 1 whereas the values of J_1,J_2,α and g keep changing and taking values 0 or 1 in Tables 1 and 2. In Table 4, J_1 and J_2 are constant whereas α and g keep changing and taking values 0 or 1. Table 1 consists of some generated estimator that are existing in our literature like $\hat{M}_{t_{m_1}}$, $\hat{M}_{t_{m_3}}$ and $\hat{M}_{t_{m_4}}$ are same and given by [4]. Estimator $\hat{M}_{t_{m_2}}$ was given by [8]. In Table 2, estimator $\hat{M}_{t_{m_{17}}}$, $\hat{M}_{t_{m_{19}}}$, and $\hat{M}_{t_{m_{20}}}$ were given by [6].

Table 1 Some members generated from the generalized estimator $\hat{M}_{t_m}$

Estimator	J_1	J_2	ψ	α	g	MSE
$\hat{M}_{t_{m_1}} = \hat{M}_y$ [4]	0	0	1	0	1	$MSE(\hat{M}_{t_{m_1}}) = k_2$
$\hat{M}_{t_{m_2}} = \hat{M}_y \left(\frac{M_x}{\hat{M}_x}\right)$ [8]	0	0	1	1	1	$MSE\left(\hat{M}_{t_{m_2}}\right) = k_2 + \Lambda^2 k_1 - 2\Lambda k_3$
$\hat{M}_{t_{m_3}} = \hat{M}_y$ [4]	0	0	1	1	0	$MSE(\hat{M}_{t_{m_3}}) = k_2$
$\hat{M}_{t_{m_4}} = \hat{M}_y$ [4]	0	0	1	0	0	$MSE(\hat{M}_{t_{m_4}}) = k_2$
$\hat{M}_{t_{m_5}} = 2\hat{M}_y + (M_z - \hat{M}_z)$	1	1	1	0	1	$MSE(\hat{M}_{t_{m_5}}) = 4k_2 + M_y^2 + a^2 k_1 - 4ak_3$
$\hat{M}_{t_{m_6}} = \left[\hat{M}_y + \hat{M}_y \left(\frac{M_x}{\hat{M}_x}\right) + \left(M_z - \hat{M}_z\right)\right] \left(\frac{M_x}{\hat{M}_x}\right)$	1	1	1	1	1	$MSE(\hat{M}_{t_{m_6}}) = 4k_2 + M_y^2 + a^2 k_1 - 4ak_3 + 21\Lambda^2 k_1 + 8a\Lambda k_1 - 18\Lambda k_3$
$\hat{M}_{t_{m_7}} = \left[\hat{M}_y + \hat{M}_y \left(\frac{M_z}{\hat{M}_z}\right) + \left(M_z - \hat{M}_z\right)\right]$	1	1	1	1	0	$MSE(\hat{M}_{t_{m_7}}) = 4k_2 + M_y^2 + a^2 k_1 - 4ak_3 + 3\Lambda^2 k_1 + 2a\Lambda k_1 - 6\Lambda k_3$
$\hat{M}_{t_{m_8}} = 2\hat{M}_y + (M_z - \hat{M}_z)$	1	1	1	0	0	$MSE\left(\hat{M}_{t_{m_8}}\right) = 4k_2 + M_y^2 + a^2 k_1 - 4ak_3$
$\hat{M}_{t_{m_9}} = 2\hat{M}_y$	1	0	1	0	1	$MSE(\hat{M}_{t_{m_9}}) = 4k_2 + M_y^2$
$\hat{M}_{t_{m_{10}}} = \left[\hat{M}_y + \hat{M}_y \left(\frac{M_z}{\hat{M}_z}\right)\right] \left(\frac{M_z}{\hat{M}_z}\right)$	1	0	1	1	1	$MSE(\hat{M}_{t_{m_{10}}}) = 4k_2 + M_y^2 + 21\Lambda^2 k_1 - 18\Lambda k_3$
$\hat{M}_{t_{m_{11}}} = \hat{M}_y + \hat{M}_y \left(\frac{M_z}{\hat{M}_z}\right)$	1	0	1	1	0	$MSE\left(\hat{M}_{t_{m_{11}}}\right) = 4k_2 + M_y^2 + 3\Lambda^2 k_1 - 6\Lambda k_3$
$\hat{M}_{t_{m_{12}}} = 2\hat{M}_y$	1	0	1	0	0	$MSE(\hat{M}_{t_{m_{12}}}) = 4k_2 + M_y^2$
$\hat{M}_{t_{m_{13}}} = \hat{M}_y + (M_z - \hat{M}_z)$	0	1	1	0	1	$MSE\left(\hat{M}_{t_{m_{13}}}\right) = k_2 + a^2 k_1 - 2ak_3$
$\hat{M}_{t_{m_{14}}} = [\hat{M}_y + (M_z - \hat{M}_z)] \left(\frac{M_z}{\hat{M}_z}\right)$	0	1	1	1	1	$MSE(\hat{M}_{t_{m_{14}}}) = k_2 + a^2 k_1 - 2\Lambda k_3 + \Lambda^2 k_1 + 2a\Lambda k_1 - 2ak_3$
$\hat{M}_{t_{m_{15}}} = \hat{M}_y + (M_z - \hat{M}_z)$	0	1	1	1	0	$MSE(\hat{M}_{t_{m_{15}}}) = k_2 + a^2 k_1 - 2ak_3$
$\hat{M}_{t_{m_{16}}} = \hat{M}_y + (M_z - \hat{M}_z)$	0	1	1	0	0	$MSE(\hat{M}_{t_{m_{16}}}) = k_2 + a^2 k_1 - 2ak_3$

Table 2 Some members generated from the generalized estimator $\hat{M}_{t_m}$

Estimator	J_1	J_2	ψ	α	g	MSE
$\hat{M}_{t_{m_{17}}} = \hat{M}_y + J_2(M_z - \hat{M}_z)$ [6]	0	J_2	1	0	1	$MSE(\hat{M}_{t_{m_{17}}}) = k_2 + J_2^2 a^2 k_1 - 2aJ_2k_3$
$\hat{M}_{t_{m_{18}}} = \left[\hat{M}_y + J_2(M_z - \hat{M}_z)\right]\left(\frac{M_z}{\hat{M}_z}\right)$	0	J_2	1	1	1	$MSE(\hat{M}_{t_{m_{18}}}) = k_2 + J_2^2 a^2 k_1 - 2aJ_2k_3 + \Lambda^2 k_1 - 2\Lambda k_3 + 2J_2a\Lambda k_1$
$\hat{M}_{t_{m_{19}}} = \hat{M}_y + J_2(M_z - \hat{M}_z)$ [6]	0	J_2	1	1	0	$MSE(\hat{M}_{t_{m_{19}}}) = k_2 + J_2^2 a^2 k_1 - 2aJ_2k_3$
$\hat{M}_{t_{m_{20}}} = \hat{M}_y + J_2(M_z - \hat{M}_z)$ [6]	0	J_2	1	0	0	$MSE(\hat{M}_{t_{m_{20}}}) = k_2 + J_2^2 a^2 k_1 - 2aJ_2k_3$
$\hat{M}_{t_{m_{21}}} = \hat{M}_y + J_1\hat{M}_y$	J_1	0	1	0	1	$MSE(\hat{M}_{t_{m_{21}}}) = k_2 + J_1^2k_2 + J_1^2M_y^2 + 2J_1k_2$
$\hat{M}_{t_{m_{22}}} = \left[\hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right)\right]\left(\frac{M_z}{\hat{M}_z}\right)$	J_1	0	1	1	1	$MSE(\hat{M}_{t_{m_{22}}}) = k_2 + J_1^2k_2 + J_1^2M_y^2 + 2J_1k_2 - 8\Lambda J_1^2k_3 + 10J_1^2\Lambda^2k_1 - 8\Lambda J_1k_3 - 2\Lambda k_3 + 10J_1\Lambda^2k_1 + \Lambda^2k_1$
$\hat{M}_{t_{m_{23}}} = \hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right)$	J_1	0	1	1	0	$MSE(\hat{M}_{t_{m_{23}}}) = k_2 + J_1^2k_2 + J_1^2M_y^2 + 2J_1k_2 - 4\Lambda J_1^2k_3 + 3J_1^2\Lambda^2k_1 - 2\Lambda J_1k_3$
$\hat{M}_{t_{m_{24}}} = \hat{M}_y + J_1\hat{M}_y$	J_1	0	1	0	0	$MSE(\hat{M}_{t_{m_{24}}}) = k_2 + J_1^2k_2 + J_1^2M_y^2 + 2J_1k_2$
$\hat{M}_{t_{m_{25}}} = 2\hat{M}_y + J_2(M_z - \hat{M}_z)$	1	J_2	1	0	1	$MSE(\hat{M}_{t_{m_{25}}}) = M_y^2 + 4k_2 + J_2^2a^2k_1 - 4aJ_2k_3$
$\hat{M}_{t_{m_{26}}} = \left[\hat{M}_y + \hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right) + J_2(M_z - \hat{M}_z)\right]\left(\frac{M_z}{\hat{M}_z}\right)$	1	J_2	1	1	1	$MSE(\hat{M}_{t_{m_{26}}}) = M_y^2 + 4k_2 + J_2^2a^2k_1 - 4aJ_2k_3 - 18\Lambda k_3 + 21\Lambda^2k_1 + 8J_2a\Lambda k_1$
$\hat{M}_{t_{m_{27}}} = \left[\hat{M}_y + \hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right) + J_2(M_z - \hat{M}_z)\right]$	1	J_2	1	1	0	$MSE(\hat{M}_{t_{m_{27}}}) = M_y^2 + 4k_2 + J_2^2a^2k_1 - 4aJ_2k_3 - 6\Lambda k_3 + 3\Lambda^2k_1 + 2J_2a\Lambda k_1$
$\hat{M}_{t_{m_{28}}} = 2\hat{M}_y + J_2(M_z - \hat{M}_z)$	1	J_2	1	0	0	$MSE(\hat{M}_{t_{m_{28}}}) = M_y^2 + 4k_2 + J_2^2a^2k_1 - 4aJ_2k_3$

Table 3 Some members generated from the generalized estimator $\hat{M}_{t_m}$

Estimator	J_1	J_2	ψ	α	g	
$\hat{M}_{t_{m_{29}}} = \hat{M}_y + J_1\hat{M}_y + \left(M_z - \hat{M}_z\right)$	J_1	1	1	0	1	$MSE(\hat{M}_{t_{m_{29}}}) = k_2 + J_1^2 k_2 + J_1^2 M_y^2 + 2J_1 k_2 - 2aJ_1 k_3 + a^2 k_1 - 2ak_3$
$\hat{M}_{t_{m_{30}}} = \left[\hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right) + \left(M_z - \hat{M}_z\right)\right]\left(\frac{M_z}{\hat{M}_z}\right)$	J_1	1	1	1	1	$MSE(\hat{M}_{t_{m_{30}}}) = J_1^2 M_y^2 - 8\Lambda J_1^2 k_3 + J_1^2 k_2 + a^2 k_1 + 10\Lambda^2 J_1^2 k_1 + 6J_1 a\Lambda k_1 - 2J_1 ak_3 - 5\Lambda^2 k_1 + k_2 - 8J_1\Lambda k_3 + 16J_1\Lambda^2 k_1 + 2J_1 k_2 + 2a\Lambda k_1 - 2ak_3 - 2\Lambda k_3$
$\hat{M}_{t_{m_{31}}} = \hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right) + \left(M_z - \hat{M}_z\right)$	J_1	1	1	1	0	$MSE(\hat{M}_{t_{m_{31}}}) = k_2 + J_1^2 k_2 + J_1^2 M_y^2 + 2J_1 k_2 - 2aJ_1 k_3 + a^2 k_1 - 2ak_3 - 4\Lambda J_1^2 k_3 + 3J_1^2\Lambda^2 k_1 + 2J_1 a\Lambda k_1 - 2J_1\Lambda k_3$
$\hat{M}_{t_{m_{32}}} = \left[\hat{M}_y + J_1\hat{M}_y + \left(M_z - \hat{M}_z\right)\right]$	J_1	1	1	0	0	$MSE(\hat{M}_{t_{m_{32}}}) = k_2 + J_1^2 k_2 + J_1^2 M_y^2 + 2J_1 k_2 - 2aJ_1 k_3 + a^2 k_1 - 2ak_3$
$\hat{M}_{t_{m_{33}}} = \left[(1 + J_1)\hat{M}_y + J_2\left(M_z - \hat{M}_z\right)\right]$	J_1	J_2	1	0	0	$MSE(\hat{M}_{t_{m_{33}}}) = k_2 + J_1^2 k_2 + J_1^2 M_y^2 + 2J_1 k_2 - 2aJ_2 k_3 + a^2 J_2^2 k_1 - 2aJ_1 J_2 k_3$

Table 4 Some members generated from the generalized estimator $\hat{M}_{t_m}$

Estimator	J_1	J_2	ψ	α	g	
$\hat{M}_{t_{m_{34}}} = \left[\hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right) + J_2\left(M_z - \hat{M}_z\right)\right]$	J_1	J_2	1	1	0	$MSE(\hat{M}_{t_{m_{34}}}) = J_1^2M_y^2 - 4J_1^2\Lambda k_3 + J_1^2k_2 + 3J_1^2\Lambda^2k_1 - 2J_1J_2ak_3 + 2J_1J_2a\Lambda k_1 - 2J_1\Lambda k_3 + 2J_1k_2 + J_2^2a^2k_1 - 2J_2ak_3 + k_2$
$\hat{M}_{t_{m_{35}}} = \left[(1+J_1)\hat{M}_y + J_2\left(M_z - \hat{M}_z\right)\right]$	J_1	J_2	1	0	1	$MSE(\hat{M}_{t_{m_{35}}}) = k_2 + J_1^2k_2 + J_1^2M_y^2 + 2J_1k_2 - 2aJ_2k_3 + a^2J_2^2k_1 - 2aJ_1J_2k_3$
$\hat{M}_{t_{m_{36}}} = \left[\hat{M}_y + J_1\hat{M}_y\left(\frac{M_z}{\hat{M}_z}\right) + J_2\left(M_z - \hat{M}_z\right)\right]\left(\frac{M_z}{\hat{M}_z}\right)$	J_1	J_2	1	1	1	$MSE(\hat{M}_{t_{m_{36}}}) = J_1^2M_y^2 - 8J_1^2\Lambda k_3 + J_1^2k_2 + 10J_1^2\Lambda^2k_1 - 2J_1J_2ak_3 + 6J_1J_2a\Lambda k_1 - 8J_1\Lambda k_3 + 2J_1k_2 + 10J_1\Lambda^2k_1 + J_2^2a^2k_1 - 2J_2ak_3 + k_2 + 2J_2a\Lambda k_1 + \Lambda^2k_1 - 2\Lambda k_3$

5 Efficiency Comparisons

This section provides a comparative analysis of the Mean Squared Error (MSE) of the proposed estimator relative to the existing estimators discussed in the paper. From Equations (2) and (18) we get,

$$\begin{aligned} Var(\hat{M}_0) - MSE(\hat{M}^G_{pp})_{min} &= 4\lambda k_2(1-\rho^2_{yx})(C^2_{M_y}+C^2_{M_x}) \\ &\quad + 4k_2\rho^2_{yx} + \lambda^2 M^2_y C^4_{M_x} \geq 0 \end{aligned} \tag{31}$$

From Equations (4) and (18) we get,

$$\begin{aligned} MSE(\hat{M}_R) - MSE(\hat{M}^G_{pp})_{\min} &= \lambda M^2_y(C^2_{M_y}+C^2_{M_x}-2C_{M_{yx}}) \\ &\quad - \frac{M^2_y}{1+\lambda C^2_{M_y}(1-\rho^2_{yx})}\left[\lambda C^2_{M_y}(1-\rho^2_{yx})\right. \\ &\quad \left. - \frac{1}{4}\lambda^2 C^4_{M_x} - \lambda^2 C^2_{M_y}C^2_{M_x}(1-\rho^2_{yx})\right] \geq 0 \end{aligned}$$

from above condition we write it as:

$$\begin{aligned} &\lambda M^2_y\left(C^2_{M_x}+\frac{\lambda C^4_{M_x}}{4}\right)+k_2\rho^2_{yx} \\ &\quad > 2\lambda M^2_y C_{M_{yx}} - \lambda k_2(1-\rho^2_{yx})(2C^2_{M_x}+C^2_{M_y}-2C_{M_{yx}}) \end{aligned} \tag{32}$$

From Equations (6) and (18) we get,

$$\begin{aligned} MSE(\hat{M}_E) - MSE(\hat{M}^G_{pp})_{\min} &= M^2_y\lambda\left\{C^2_{M_y}+\frac{1}{4}C^2_{M_x}-C_{M_{yx}}\right\} \\ &\quad - \frac{M^2_y}{1+\lambda C^2_{M_y}(1-\rho^2_{yx})}\left[\lambda C^2_{M_y}\left(1-\rho^2_{yx}\right)\right. \\ &\quad \left. - \frac{1}{4}\lambda^2 C^4_{M_x} - \lambda^2 C^2_{M_y}C^2_{M_x}(1-\rho^2_{yx})\right] \geq 0 \end{aligned}$$

from above condition we write it as:

$$\begin{aligned} &\frac{1}{4}\lambda M^2_y(C^2_{M_x}+\lambda C^4_{M_x})+k_2\rho^2_{yx} \\ &\quad > \lambda M^2_y C_{M_{yx}} - \lambda k_2(1-\rho^2_{yx})\left(\frac{5C^2_{M_x}}{4}+C^2_{M_y}-C_{M_{yx}}\right) \end{aligned} \tag{33}$$

From Equations (8) and (18) we get,

$$MSE(\hat{M}_{D_0})_{min} - MSE(\hat{M}_{pp}^{G})_{\text{min}}$$

$$= M_y^2 C_{M_y}^2 \lambda(1-\rho_{yx}^2) - \frac{M_y^2}{1+\lambda C_{M_y}^2(1-\rho_{yx}^2)}$$

$$\times \left[\lambda C_{M_y}^2(1-\rho_{yx}^2) - \frac{1}{4}\lambda^2 C_{M_x}^4 - \lambda^2 C_{M_y}^2 C_{M_x}^2 (1-\rho_{yx}^2)\right] \geq 0$$

from above condition we write it as:

$$\lambda k_2(1-\rho_{yx}^2)[C_{M_y}^2(1-\rho_{yx}^2) + C_{M_x}^2] + \frac{M_y^2 \lambda^2 C_{M_x}^4}{4} \geq 0 \qquad (34)$$

From Equations (18) and (23) we get,

$$MSE(\hat{M}_{pp}^{G})_{\text{min}} - MSE(M_y^*)_{\text{min}}$$

$$= \frac{M_y^2}{1+\lambda C_{M_y}^2(1-\rho_{yx}^2)}\left[\lambda C_{M_y}^2(1-\rho_{yx}^2) - \frac{1}{4}\lambda^2 C_{M_x}^4\right.$$

$$\left. - \lambda^2 C_{M_y}^2 C_{M_x}^2(1-\rho_{yx}^2)\right] - M_y^2$$

$$+ \frac{M_y^2[64+\lambda C_{M_x}^2(64+\lambda(25C_{M_x}^2 - 16C_{M_y}^2(\rho_{yx}^2-1)))]}{64[1+\lambda C_{M_x}^2 + \lambda C_{M_y}^2(1-\rho_{yx}^2)]} \geq 0.$$

from above condition we write it as:

$$64\left[k_2(1-\rho_{yx}^2)(1-\lambda C_{M_x}^2) - \frac{\lambda^2 M_y^2 C_{M_x}^4}{4}\right][\lambda C_{M_x}^2 + k_4] - 64\lambda M_y^2 C_{M_x}^2$$

$$+ \lambda M_y^2 C_{M_x}^2 k_4[64 + \lambda\{25C_{M_x}^2 + 16C_{M_y}^2(1-\rho_{yx}^2)\}] \geq 0. \qquad (35)$$

where $k_4 = 1 + \lambda C_{M_y}^2(1-\rho_{yx}^2)$.

From Equations (23) and (28) we get,

$$MSE(M_y^*)_{\text{min}} - MSE(\hat{M}_{t_m})$$

$$= M_y^2 - \frac{M_y^2[64+\lambda C_{M_x}^2(64+\lambda\{25C_{M_x}^2 - 16C_{M_y}^2(\rho_{yx}^2-1)\})]}{64[1+\lambda C_{M_x}^2 + \lambda C_{M_y}^2\{1-\rho_{yx}^2\}]}$$

$$- \psi^2\{J_1^2 A - 2J_1J_2B + 2J_1C + J_2^2 D - 2J_2E + F\}$$
$$+ 2\psi\{J_1G + J_2H + I\} - M_y^2 \geq 0$$

from above condition we write it as:

$$\begin{aligned}
&64[2\psi(J_1G + J_2H + I) \\
&\quad - \psi^2(J_1^2 A - 2J_1J_2B + 2J_1C + J_2^2 D - 2J_2E + F)](k_4 + \lambda C_{M_x}^2) \\
&\quad - M_y^2[64 + \lambda C_{M_x}^2(64 + \lambda(25C_{M_x}^2 - 16C_{M_y}^2(\rho_{yx}^2 - 1)))] \geq 0
\end{aligned} \tag{36}$$

Under the above conditions, the proposed class of estimators outperforms the existing class of estimators. To validate the practical applicability of these conditions, a computational study has been carried out.

6 Empirical Study

To assess the efficiency of the proposed family of estimators, four real-world population datasets have been considered. A comprehensive summary of these datasets is provided in the corresponding Table 5.
For Population 1: Source [24]
Let y=No. of fish caught by fishermen in the year 1995 and x=No. of fish caught by fishermen in the year 1964.
For Population 2: Source [24]
Let y=No. of fish caught by fishermen in the year 1995 and x=No. of fish caught by fishermen in the year 1993.
For Population 3 Source [26]

Table 5 Summary of population parameters for empirical study

	Population 1	Population 2	Population 3	Population 4
N	69	69	144	51
n	17	17	10	11
λ	0.01108	0.01108	0.02327	0.01782
M_y	2068	2068	2023	25.80
M_x	2011	2307	64659	25.60
$f_y(M_y)$	0.00014	0.00014	0.00024	0.0728
$f_x(M_x)$	0.00014	0.00013	0.00001	0.1080
ρ_{yx}	0.1505	0.3166	0.86110	0.9956
R	0.97243	1.11557	31.96193	0.99224
C_{M_y}	3.45399	3.45399	2.05965	0.53242
C_{M_x}	3.55189	3.33433	1.54658	0.36168
$C_{M_{yx}}$	1.84636	3.6462	2.74295	0.19172

Let y=No. of members of faculty and x=No. of students in four different colleges of 36 districts in Punjab.
For Population 4 Source [26]
Let y=Prices of oil in current week of 2017 and x=Prices of oil in previous week of 2017.

In Table 5, for each of the four population datasets, all relevant parameters have been clearly specified. Here, the study and auxiliary variables have been selected to be highly correlated, ensuring that the proposed family of estimators performs effectively. Populations 3 and 4 exhibit a high ρ_{yx} value, indicating strong correlation between the study and auxiliary variables. We now explain the population parameters presented in Table 5. Here, N denotes the size of the respective population, and n represents the sample size drawn from the population using Simple Random Sampling Without Replacement (SRSWOR), $\lambda = 0.25(\frac{1}{n} - \frac{1}{N})$ = Finite population correction factor, M_y = Population median of the study variable of their respective population, M_x = Population median of the auxiliary variable of their respective population. Here as $N \rightarrow \infty$, $n \rightarrow \infty$ then $nN^{-1} \rightarrow f$ (correction factor) and we are assuming as $N \rightarrow \infty$ the distribution of (X,Y) approaches to continuous distribution with marginal densities $f_y(M_y)$ of Y and $f_x(M_x)$ of X. This assumption is necessary for super population model framework for treating values of Y and X as a realization of N independent observations of the population from a continuous distribution. In addition to this we assumed that $f_y(M_y)$ and $f_x(M_x)$ are positive. $\rho_{yx} = \rho_{(\hat{M}_y,\hat{M}_x)}$ = Coefficient of correlation between sample median of auxiliary variable X and study variable Y of their respective population, $R = \frac{M_x}{M_y}$, $C_{M_y} = [M_y f_y(M_y)]^{-1}$ = Coefficient of variation of the median (study variable) of their respective population, $C_{M_x} = [M_x f_x(M_x)]^{-1}$ = Coefficient of variation of the median(auxiliary variable) of their respective population, $C_{M_{yx}} = \rho_{yx} C_{M_y} C_{M_x}$.

In Tables 6 and 7, the Mean Squared Error (MSE) values of the estimators generated from the proposed family have been recorded for Populations 1, 2, 3, and 4. These values were obtained by substituting the population parameters from 5 into the MSE expressions of the estimators given in Tables 1, 2, 3 and 4.

7 Discussion

- To mitigate the influence of outliers, we have utilized the median rather than the mean, as the mean is known to be sensitive to extreme observations in the population.

Table 6 Variances/MSEs/Minimum MSEs of different estimators

Estimators	Population 1	Population 2	Population 3	Population 4
$Var(\hat{M}_0)$	565443.60	565443.60	403886.96	3.36
$MSE(\hat{M}_R)$	988372.80	746752.60	109312.95	0.36
$MSE(\hat{M}_E)$	627420.20	524362.10	199667.99	1.47
$MSE(\hat{M}_0)_{min}$	552636.13	508766.01	104407.52	0.02
$MSE(\hat{M}_{D_1})_{min}$	489395.24	454675.78	101810.20	0.02
$MSE(\hat{M}_{D_2})_{min}$	480458.30	447982.60	101661.14	0.02
$MSE(\hat{M}_{D_3})_{min}$	471131.80	439763.40	100200.79	0.02
$MSE(\hat{M}_{pp}^G)_{min}$	402459.30	384146.80	93055.80	0.02
$MSE(M_y^*)_{min}$	394518.94	376541.31	90648.43	0.02
$MSE(\hat{M}_{t_m})_{min}$	25864.52	35256.30	65459.24	0.27
$MSE(\hat{M}_{t_1})_{min}$	565443.56	565443.56	403886.95	3.36
$MSE(\hat{M}_{t_2})_{min}$	987359.05	746139.48	109315.01	0.47
$MSE(\hat{M}_{t_3})_{min}$	565443.56	565443.56	403886.95	3.36
$MSE(\hat{M}_{t_4})_{min}$	565443.56	565443.56	403886.95	3.36
$MSE(\hat{M}_{t_5})_{min}$	6509560.65	6316763.13	47174020.10	674.96
$MSE(\hat{M}_{t_6})_{min}$	19791542.30	16604742.20	76369028.30	670.73
$MSE(\hat{M}_{t_7})_{min}$	8356199.96	7446185.01	53568701.98	667.73
$MSE(\hat{M}_{t_8})_{min}$	6509560.65	631673.13	47174020.10	674.96
$MSE(\hat{M}_{t_9})_{min}$	6538398.27	6538398.27	5708076.84	679.09
$MSE(\hat{M}_{t_{10}})_{min}$	17496808.51	14477040.81	5789494.31	669.15
$MSE(\hat{M}_{t_{11}})_{min}$	7804144.74	7080486.03	4824361.01	670.43
$MSE(\hat{M}_{t_{12}})_{min}$	6538398.00	6538398.00	5708077.00	679.09
$MSE(\hat{M}_{t_{13}})_{min}$	621705.20	536598.58	50216719.70	1.48
$MSE(\hat{M}_{t_{14}})_{min}$	1624513.54	1304628.63	57200545.50	0.03
$MSE(\hat{M}_{t_{15}})_{min}$	621705.20	536598.58	50216719.71	1.48
$MSE(\hat{M}_{t_{16}})_{min}$	621705.20	536598.58	50216719.71	1.48
$MSE(\hat{M}_{t_{17}})_{min}$	552803.84	509406.65	104592.86	0.02
$MSE(\hat{M}_{t_{18}})_{min}$	1089935.86	879691.81	120502.73	0.02

- Analysis of Tables 6 and 7 leads us to the conclusion that estimators that rely on auxiliary information are more effective. The Table 6 clearly indicates that estimators $\hat{M}_{D_1}$,$\hat{M}_{D_2}$,$\hat{M}_{D_3}$,$\hat{M}_{pp}^G$, and M_y^* exhibit higher efficiency compared to estimator $\hat{M}_0$.
- The members of the class of estimators $\hat{M}_{t_i}$ for i=1,3,4 obtained from the family $\hat{M}_{t_m}$, are equally efficient as $\hat{M}_0$, while $\hat{M}_{t_2}$ is equally efficient as $\hat{M}_R$. Furthermore, the estimators $\hat{M}_{t_i}$ for i $=$ 17,19,20,23,31 exhibit equal efficiency among themselves, whereas the estimators $\hat{M}_{t_i}$ for

Table 7 Variances/MSEs/Minimum MSEs of different estimators

Estimators	Population 1	Population 2	Population 3	Population 4
$MSE(\hat{M}_{t_{19}})_{min}$	556518.39	510803.00	104410.65	0.02
$MSE(\hat{M}_{t_{20}})_{min}$	552803.84	509406.65	104592.86	0.02
$MSE(\hat{M}_{t_{21}})_{min}$	499439.16	499942.45	386574.59	3.36
$\boldsymbol{MSE(\hat{M}_{t_{22}})_{min}}$	**22585.47**	**33042.82**	**62870.99**	**0.47**
$MSE(\hat{M}_{t_{23}})_{min}$	531203.92	541840.35	399312.41	3.36
$MSE(\hat{M}_{t_{24}})_{min}$	499439.16	499942.45	386574.59	3.36
$MSE(\hat{M}_{t_{25}})_{min}$	6503074.91	6381057.77	4824724.21	669.09
$MSE(\hat{M}_{t_{26}})_{min}$	17924182.87	15082738.31	6148968.81	670.92
$MSE(\hat{M}_{t_{27}})_{min}$	8026616.93	7301804.55	4448459.64	664.65
$MSE(\hat{M}_{t_{28}})_{min}$	6503075.00	6381058.00	4824724.00	669.09
$MSE(\hat{M}_{t_{29}})_{min}$	565439.12	491594.25	50407053.60	1.48
$MSE(\hat{M}_{t_{30}})_{min}$	4707854.50	4049840.40	53656773.20	8.02
$MSE(\hat{M}_{t_{31}})_{min}$	540612.56	476446.52	50239156.42	1.48
$MSE(\hat{M}_{t_{32}})_{min}$	565439.12	491594.24	50407053.58	1.48
$MSE(\hat{M}_{t_{33}})_{min}$	489395.24	454675.78	101810.16	0.02
$MSE(\hat{M}_{t_{34}})_{min}$	482783.90	447402.51	83830.11	0.02
$MSE(\hat{M}_{t_{35}})_{min}$	489395.24	454675.78	101810.16	0.02
$\boldsymbol{MSE(\hat{M}_{t_{36}})_{min}}$	**28013.04**	**18201.03**	**24093.33**	**0.01**

i = 21,24,33,34,35 are more efficient than the usual unbiased estimator $\hat{M}_y$.

- In the estimation of the population median M_y, the sample median is known to exhibit bias, particularly in small sample sizes, often resulting in skewed estimates. To mitigate this bias and enhance the efficiency of the estimator, we incorporate a set of constants $(J_1, J_2, \psi, a, b$, and $g)$. These constants are specifically chosen to adjust the estimator such that it more accurately approximates the true population median, while simultaneously reducing variance and improving overall precision. The estimator that incorporates all these constants proves to be the most efficient, achieving the minimum mean squared error (MSE) under optimal conditions. This performance is followed by estimators that use unit values for the constants $\psi = 1$, $\alpha = 1$, $g = 1$, which exhibit relatively higher MSEs.
- The strong correlation between the study variable and the auxiliary variable contributes significantly to the improved performance of the proposed estimator. Furthermore, for all the estimators considered, the mean squared error (MSE) consistently decreases as the sample size increases.

- The estimators presented in the tables that incorporate auxiliary information – specifically the population median M_x and the correlation coefficient ρ_{yx} – demonstrate the lowest mean squared errors (MSEs). As shown in Tables 6 and 7, the estimators $\hat{M}_{22}$ and $\hat{M}_{36}$ exhibit significantly lower MSEs compared to the conventional estimators $\hat{M}_y$, $\hat{M}_R$, and $\hat{M}_E$.
- Among all the estimators that are present in the Tables 6 and 7, $\hat{M}_{36}$ is the best of all for the particular values of constant. Hence we say (ψ=1,$\alpha = 1$ and $g = 1$) are the suitable values for obtaining minimum MSEs among the different subsets of these constant.
- As a potential direction for future work, the proposed approach could be extended and evaluated within the context of stratified sampling.

8 Conclusion

We have proposed a family of estimators for estimating the population median of a study variable using auxiliary information under simple random sampling without replacement (SRSWOR). The proposed estimators are designed to perform effectively even in the presence of skewed distributions or datasets containing outliers. Furthermore, it is observed that our proposed family includes several well-known estimators as special cases – for instance, the conventional unbiased estimator of the population median based on the study variable [4], and the auxiliary variable-based median estimator proposed by [8]. In addition to encompassing existing estimators, several new estimators have also been derived from the proposed family. For these estimators, we have computed the bias and mean squared error (MSE) up to the first order of approximation. The proposed family of estimators is advantageous in that the statistical properties – such as bias and MSE – of individual estimators within the class can be easily obtained from the general form. We ultimately come to the conclusion that there is always a chance to generate estimators from the suggested estimators $\hat{M}_{t_i}$ (i=1,2,3…36) and $\hat{M}_{t_m}$ that are superior to the current estimators. The suggested family of estimators is tested against existing estimators using four distinct datasets of size $n_1 = 69$, $n_2 = 69$, $n_3 = 144$, and $n_4 = 51$. Using normal distributions, the population density functions for X and Y are also computed. In the empirical study, it has been shown that our estimator $\hat{M}_{36}$ is better than the existing estimator as it has minimum MSE among all the existing estimator and confirming higher accuracy in real-world survey.

References

[1] Ahmad, S., Masood, S., Alomair, A. M., and Alomair, M. A. (2024). Improved modified estimator for estimation of median using auxiliary information under simple random sampling. *Scientific Reports*, 14(1):16504.

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

[3] Baig, A., Masood, S., and Ahmed Tarray, T. (2020). Improved class of difference-type estimators for population median in survey sampling. *Communications in Statistics-Theory and Methods*, 49(23):5778–5793.

[4] Gross, S. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, pages 181–184.

[5] Gupta, S., Shabbir, J., and Ahmad, S. (2008). Estimation of median in two-phase sampling using two auxiliary variables. *Communications in Statistics-Theory and Methods*, 37(11):1815–1822.

[6] Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). Sample survey methods and theory. vol. i. methods and applications.

[7] Hussain, M. A., Javed, M., Zohaib, M., Shongwe, S. C., Awais, M., Zaagan, A. A., and Irfan, M. (2024). Estimation of population median using bivariate auxiliary information in simple random sampling. *Heliyon*, 10(7).

[8] Kuk, A. Y. and Mak, T. (1989). Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):261–269.

[9] Muneer, S., Khalil, A., Shabbir, J., and Narjis, G. (2022). Efficient estimation of population median using supplementary variable. *Scientia Iranica*, 29(1):265–274.

[10] Murthy, M. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 69–74.

[11] Rao, T. (1991). On certain methods of improving ratio and regression estimators. *Communications in Statistics-Theory and Methods*, 20(10):3325–3340.

[12] Reddy, V. N. (1973). On ratio and product methods of estimation. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 307–316.

[13] Reddy, V. N. (1974). On a transformed ratio method of estimation. *Sankhyā C*, 36:59–70.

[14] Sahai, A. and Ray, S. (1980). An efficient estimator using auxiliary information. *Metrika*, 27(1):271–275.

[15] Shabbir, J. and Gupta, S. (2015). A note on generalized exponential type estimator for population variance in survey sampling. *Revista Colombiana de EstadÃstica*, 38(2):385–397.

[16] Shabbir, J. and Gupta, S. (2017). A generalized class of difference type estimators for population median in survey sampling. *Hacettepe Journal of Mathematics and Statistics*, 46(5):1015–1028.

[17] Shabbir, J., Gupta, S., and Narjis, G. (2022). On improved class of difference type estimators for population median in survey sampling. *Communications in Statistics-Theory and Methods*, 51(10):3334–3354.

[18] Shahzad, U., Al-Noor, N. H., Hanif, M., and Sajjad, I. (2021). An exponential family of median-based estimators for mean estimation with simple random sampling scheme. *Communications in Statistics-Theory and Methods*, 50(20):4890–4899.

[19] Sharma, P. and Singh, R. (2013). Efficient estimator of population mean in stratified random sampling using auxiliary attribute. *World Applied Sciences Journal*, 27(12):1786–1791.

[20] Sharma, P. and Singh, R. (2015). Generalized class of estimators for population median using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 44(2):443–453.

[21] Singh, G. N., Pandey, A. K., and Singh, C. (2022). Generalized estimation strategy for mean estimation on current occasion in two-occasion rotation patterns. *Communications in Statistics-Simulation and Computation*, 51(4):1661–1684.

[22] Singh, H. P., Sidhu, S. S., and Singh, S. (2006). Median estimation with known interquartile range of auxiliary variable. *International Journal of Applied Mathematics and Statistics*, 4:68–80.

[23] Singh, H. P. and Solanki, R. S. (2013). Some classes of estimators for the population median using auxiliary information. *Communications in Statistics-Theory and Methods*, 42(23):4222–4238.

[24] Singh, S. (2003). *Advanced Sampling Theory With Applications: How Michael "Selected" Amy*, volume 2. Springer Science & Business Media.

[25] Srivastava, S. K. (1967). An estimator using auxiliary information in sample surveys. *Calcutta Statistical Association Bulletin*, 16(2–3): 121–132.

[26] Statistics, B. O. (2009). Punjab development statistics.
[27] Subzar, M., Lone, S. A., Ekpenyong, E. J., Salam, A., Aslam, M., Raja, T., and Almutlak, S. A. (2023). Efficient class of ratio cum median estimators for estimating the population median. *PLOS One*, 18(2):e0274690.

Biographies

Prayas Sharma is currently working as Assistant Professor in the Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow. Dr. Sharma holds a Bachelor's degree in Computer Science & Statistics, Masters and Doctorate degree in Statistics from Banaras Hindu University, Varanasi, India. Dr. Sharma has good knowledge of Statistics, Artificial Intelligence and Machine Learning, Business Analytics & Research Methodology along with strong computational & programming skills.

He has more than 11 years of academic experience, both in the domain of teaching and research. His research interest includes Survey Sampling, Estimation Procedures using Auxiliary Information and Measurement Errors, Predictive Modelling, Business Analytics and Operations Research. Dr. Sharma has published more than 65 research papers in reputed National & International journals along with one book and two chapters in book internationally published. He has more than **800 citations with H-Index 17 & I index of 23**. Dr. Sharma has a keen interest in reading, writing and publishing, he is serving 7 reputed journals as editor/associate editor and more than 30 journals as reviewer and **reviewed more than 150 research papers**.

Anupam Lata is research scholar in the Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow. She has completed Masters in Statistics and pursuing the research in the area of sampling theory.

Subash Kumar Yadav is a faculty in the Department of Statistics at Babasaheb Bhimrao Ambedkar University Lucknow, U.P., India. He earned his M.Sc. and Ph.D. degrees in Statistics from Lucknow University and qualified the National Eligibility Test. Dr. Yadav has published more than 120 papers in SCOPUS/WoS indexed national and international journals of repute and two books from an international publisher. He is a referee for 20 reputed international journals. He has presented papers in more than 20 national and international conferences and also delivered more than 70 invited talks in several conferences and chaired sessions in different national and international conferences. He has been awarded best paper award twice and awarded four timed the Research and Academic Excellence award by his institution.

Muhammad Noor ul Amin is an Associate Professor of Statistics at COMSATS University Islamabad, Lahore Campus. He specializes in statistical quality control, data science, and advanced applied statistics. His research interests include adaptive and robust control charts, ranked set sampling, and machine learning applications in forecasting and health data.