Mixed-Optimum Estimators for Estimating Finite Population Mean in the Presence of Outliers using Auxiliary Variable under Simple Random Sampling

ABSTRACT

In sample survey the nature of correlation between the study and auxiliary variables plays a crucial role in improving the accuracy of the estimates. In this study a generalized mixed-optimum estimators that handle the three nature of correlation for different values (-1,0,1) of the scalar was proposed for estimating the finite population mean when there is information on the minimum and maximum values of the auxiliary variable and when both the auxiliary and study variables exhibit extreme values. The expression for the mean squared errors and biases were derived to the first order of approximation. The performance of the proposed estimators, relative to conventional methods, has been rigorously analyzed, revealing notable improvements. Theoretical analysis confirmed that correcting the estimators for mitigating maximum and minimum values enhanced its efficiency, and these findings have been empirically validated through comprehensive numerical analysis.

Keywords: Mixed-Optimum estimator, Auxiliary Variable, Outliers, Mean Squared Error.

1. Introductions

The estimation of the population parameters such as mean, total, proportion and even population ratio of the study variable, with greater precision, minimum cost and time is a persistent issue in sampling practice. The use of an auxiliary information can increase the precision of an estimator when study variable, say, *y* is highly correlated with auxiliary variable, say, *x*. For example, oil production and revenue, exchange rate and inflation rate, agricultural output and subsidies, government expenditure and gross domestic growth, malaria incidence and mosquito net distribution, infant mortality rate and immunization coverage patient age and blood pressure levels, body mass index and cholesterol levels. Cochran (1940), used auxiliary information to develop the ratio estimator for estimating population mean of the study variable using single phase sampling design. Several studies among many exist on the use of auxiliary variables which enhanced the precision of the study variable in sample survey, such as: Samiuddin and Hanif (2007), Singh and Espejo (2007), Hanif *el tal.* (2010), Swain (2012), Shahbaz *et al* (2014), Kanwai *et al.* (2016). Estimating the population mean in the presence of outliers can lead to biased and inaccurate results, as these extreme values can significantly influence the sample mean and inflate or deflate its value. Ratio, regression, difference, product and mixed estimators are also affected by the presence of extreme value in the data set which produce less efficient estimate. Extreme values in financial data can lead to inaccurate predications of stock prices and investment risks, in medical research, ignoring extreme values in patient outcomes can skew disease prevalence estimate and treatment effectiveness, extreme values in survey responses can distort public opinion estimates and policy recommendation, in reliability engineering ignoring extreme failure times can lead to inaccurate estimates of product lifespan and maintenance schedules. There have been few studies on the correction of extreme values, among which includes the following: Khan & Shabbir (2013) suggested some modified ratio, product, and regression type estimators when using minimum and maximum values. Al-Hossain and Khan (2014) worked on the estimation of population mean using maximum and minimum values under simple random sampling by incorporating the knowledge of two auxiliary variables. Khan (2015), presents a ratio estimator for the estimation of finite population mean of the study variable under double sampling scheme when there are unusually low and unusually high values and analyzes their properties. Darez *et al.* (2018), suggested an improved class of ratio type estimators in estimating the finite population mean when information on minimum and maximum values of the auxiliary variable is known. Other authors that worked on the correction of extreme values comprises: Mohanty and Sahoo (1995), Walia *et al.* (2015), Cekim and Cingi (2016) and Olayiwola *et al.*, (2021**).** In their studies the proposed estimators are relative to their usual counterparts.

The prior detection of the presence of extreme value in the data set on both the study and auxiliary variables, to the application of estimators, is very important to obtaining efficient estimate. In this study, a generalized modified ratio-product-cum-exponential estimator that can handle the ratio, product and regression challenges for outliers in order to improve the validity of the population parameters is proposed. These proposed estimators can be used for the estimation of the population mean in these areas among others, disease outbreak, hospital-acquired infection, in environmental studies natural disaster (hurricane, earthquakes) nuclear accidents, bank failures, market crashes, students' academic performance

2. Materials and methods

2.1 Research Design

Consider a finite population $U = (U_1, U_2, ---, U_N)$ of size N units. Let y and x denote the study variable and auxiliary variable respectively. A sample of size n $(n < N)$ assuming a simple random sampling without replacement (SRSWOR) s of size n is drawn from the population U $(s \subset U)$. The purpose is to estimate the population mean of the study variable $\bar{Y} = \frac{1}{N}\sum y_i$. It is further assumed that the population mean $\bar{X} = \frac{1}{N}\sum x_i$ of the auxiliary variable x is known. The minimum says (x_{min}) and maximum say (x_{max}) values of the auxiliary variables are also assumed to be known.

In sample survey, mean per unit estimators for finite population mean is very sensitive to unexpected values, if there exists unexpected very large (say y_{max}) and very small (say y_{min}) units in the population, as a result of this, the population mean will be either underestimated (in case the sample contains $y$min) or overestimated (in case it containsy_{max}). To overcome this situation, the methods of Sandal (1972, as cited in Khan and Shabbir 2013)) is employed mitigate the effect of these extreme values. He suggested the following unbiased estimator:

$$\bar{y}_s = \begin{cases} (\bar{y} + c_1), \\ (\bar{y} - c_1), \\ \bar{y} \end{cases} \quad (1)$$

where,

$(\bar{y} + c_1)$ if sample contains y_{min} but not y_{max}

$(\bar{y} - c_1)$ if sample contains y_{max} but not y_{min},

$\bar{y}$ for all other samples, and c is a constant.

The variance of $\bar{y}_s$ is given by:

$$V(\bar{y}_s) = \theta S_y^2 - \frac{2\theta nc}{N-1}(y_{max} - y_{min} - nc) \quad (2)$$

$$V(\bar{y}_s)_{opt} = V(\bar{y}) - \frac{\theta(\Delta y)^2}{2(N-1)} \quad (3)$$

which is always smaller than $V(\bar{y})$

where:

$V(\bar{y})$ is the variance of mean per unit estimator $\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$

$\theta = \frac{1}{n} - \frac{1}{N}$, $S_y^2 = \frac{1}{1-N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$

2.2 Estimators for Estimating the Population Mean

Some of the estimators for estimating the population mean includes:

The sample mean of the study variable is given as:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (4)$$

which is an unbiased estimator of finite population variance $(S_y^2)$and its variance is

$$\text{Var.}(\bar{y}) = \theta \bar{Y}^2 C_y^2 \qquad (5)$$

$$\text{Bias}(\bar{y}) = 0 \qquad (6)$$

The usual ratio [Cochran (1940)] and product [Robson (1957) and Murthy (1964)] as cited in Singh *et al.*, (2020) estimators of population mean have been defined as:

Ratio estimator:

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}} \qquad (7)$$

$$\text{MSE}(\bar{y}_R) = \theta \bar{Y}^2 (C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x) \qquad (8)$$

$$\text{Bias}(\bar{y}_R) = \theta \bar{Y} (C_x^2 - \rho_{yx} C_y C_x) \qquad (9)$$

Product estimator:

$$t_P = \bar{y}\frac{\bar{x}}{\bar{X}} \qquad (10)$$

$$\text{MSE}(\bar{y}_p) = \theta \bar{Y}^2 (C_y^2 + C_x^2 + 2\rho_{yx} C_y C_x) \qquad (11)$$

$$\text{Bias}(\bar{y}_P) = \theta \bar{Y} \rho_{yx} C_y C_x \qquad (12)$$

Bahl and Tuteja (1991, as cited in Abiodun *et al.*, 2021), exponential ratio and product type estimator

Exponential Ratio:

$$\bar{y}_{ER} = \bar{y}\, exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right] \qquad (13)$$

$$\text{MSE}(\bar{y}_{ER}) = \bar{Y}^2 \theta \left[C_y^2 + \frac{1}{4} C_x^2 - \rho_{yx} C_y C_x\right] \qquad (14)$$

$$\text{Bias}\ (\bar{y}_{ER}) = \theta\bar{Y}\left[\frac{3C_x^2}{8} - \frac{\rho_{yx}C_yC_x}{2}\right] \tag{15}$$

Exponential Product:

$$\bar{y}_{EP} = \bar{y}\ exp\left[\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right] \tag{16}$$

$$\text{MSE}\ (\bar{y}_{EP}) = \theta\bar{Y}^2\left[C_y^2 + \frac{1}{4}C_x^2 + \rho_{yx}C_yC_x\right] \tag{17}$$

$$\text{Bias}\ (\bar{y}_{EP}) = \theta\bar{Y}\left[\frac{\rho_{yx}C_yC_x}{2} - \frac{C_y^2}{8}\right] \tag{18}$$

Cochran (1942) regression estimator

$$\bar{y}_{Reg.} = \bar{y} + b_{yx}(\bar{X} - \bar{x}) \tag{19}$$

$$\text{MSE}\ (\bar{y}_{Reg.}) = \theta\bar{Y}^2C_y^2\left(1-\rho_{yx}^2\right) \tag{20}$$

2.3 Proposed Estimators

Here, the mixed optimum, T_{GC1} and mixed non-optimum T_{GC2} estimators are modified to handle the case of extreme values on both the study and auxiliary variables.

$$T_{GC1} = 2^{-1}\bar{y}_{c_0}\left\{\left(\frac{\bar{X}}{\bar{x}_{c_1}}\right)^{\alpha_2} + \left(\frac{\bar{x}_{c_1}}{\bar{X}}\right)^{1-\alpha_2}\right\}\left\{exp\left[\frac{\bar{X}-\bar{x}_{c_1}}{\bar{X}+\bar{x}_{c_1}}\right]\right\} \tag{21}$$

where α_2 is a constant, $-1 \leq \alpha_2 \leq 1$

$$T_{GC1} = \begin{cases} 2^{-1}(\bar{y}+c_0)\left\{\left(\frac{\bar{X}}{\bar{x}+c_1}\right)^{\alpha_2} + \left(\frac{\bar{x}+c_1}{\bar{X}}\right)^{1-\alpha_2}\right\}\left\{exp\left[\frac{\bar{X}-(\bar{x}+c_1)}{\bar{X}+(\bar{x}+c_1)}\right]\right\} \\ 2^{-1}(\bar{y}-c_0)\left\{\left(\frac{\bar{X}}{\bar{x}-c_1}\right)^{\alpha_2} + \left(\frac{\bar{x}-c_1}{\bar{X}}\right)^{1-\alpha_2}\right\}\left\{exp\left[\frac{\bar{X}-(\bar{x}-c_1)}{\bar{X}+(\bar{x}-c_1)}\right]\right\} \\ 2^{-1}\bar{y}\left\{\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha_2} + \left(\frac{\bar{x}}{\bar{X}}\right)^{1-\alpha_2}\right\}\left\{exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]\right\} \end{cases} \tag{22}$$

$$T_{GC2} = 2^{-1}\bar{y}_{c_0}\left\{\left(\frac{\bar{X}}{\bar{x}_{c_1}}\right) + \left(\frac{\bar{x}_{c_1}}{\bar{X}}\right)\right\}\left\{exp\left[\frac{\bar{X}-\bar{x}_{c_1}}{\bar{X}+\bar{x}_{c_1}}\right]\right\} \tag{23}$$

$$T_{GC2} = \begin{cases} 2^{-1}(\bar{y}+c_0)\left\{\left(\frac{\bar{X}}{\bar{x}+c_1}\right)+\left(\frac{\bar{x}+c_1}{\bar{X}}\right)\right\}\left\{exp\left[\frac{\bar{X}-(\bar{x}+c_1)}{\bar{X}+(\bar{x}+c_1)}\right]\right\} \\ 2^{-1}(\bar{y}-c_0)\left\{\left(\frac{\bar{X}}{\bar{x}-c_1}\right)+\left(\frac{\bar{x}-c_1}{\bar{X}}\right)\right\}\left\{exp\left[\frac{\bar{X}-(\bar{x}-c_1)}{\bar{X}+(\bar{x}-c_1)}\right]\right\} \\ 2^{-1}\bar{y}\left\{\left(\frac{\bar{X}}{\bar{x}}\right)+\left(\frac{\bar{x}}{\bar{X}}\right)\right\}\left\{exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]\right\} \end{cases} \tag{24}$$

To obtain the bias and the MSE for T_{GC1} and T_{GC2}, we define the following relative error terms.

$$\bar{y}_{co} = \bar{Y}(1+\bar{e}_o) \; and \; \bar{x}_{c1} = \bar{X}(1+\bar{e}_1) \tag{25}$$

$$E(\bar{e}_o) = E(\bar{e}_1) = 0, \; E(e_0^2) = \frac{\theta}{\bar{Y}^2}\left(S_y^2 - \frac{2nc_0}{N-1}(y_{max} - y_{min} - nc_0)\right),$$

$$E(e_1^2) = \frac{\theta}{\bar{X}^2}\left(S_x^2 - \frac{2nc_1}{N-1}(x_{max} - x_{min} - nc_1)\right), \text{ and}$$

$$E(e_0e_1) = \frac{\theta}{\bar{Y}\bar{X}}\left(S_{yx}^2 - \frac{n}{N-1}(c_1(y_{max} - y_{min}) + c_0(x_{max} - x_{min}) - 2nc_0c_1)\right) \tag{26}$$

Applying (25) in (26) gave (27)

$$T_{p1} = \left[\bar{Y} + \bar{Y}\bar{e}_0 - \alpha_2\bar{Y}\bar{e}_1 - \alpha_2\bar{Y}\,\bar{e}_0\,\bar{e}_1 + \frac{1}{2}\left(\frac{1}{4} + \alpha_2 + \alpha_2^2\right)\bar{Y}\bar{e}_1^2\right] \tag{27}$$

Subtracting $\bar{Y}$ from both sides of (27) and taking expectation of both sides, we get the bias of the estimator T_{GC1}

$$\text{Bias}(T_{GC1}) = E(T_{GC1} - \bar{Y}) = E\left[\bar{Y}\bar{e}_0 - \alpha_2\bar{Y}\bar{e}_1 - \alpha_2\bar{Y}\;\bar{e}_0\,\bar{e}_1 + \frac{1}{2}\left(\frac{1}{4} + \alpha_2 + \alpha_2^2\right)\bar{Y}\bar{e}_1^2\right] \tag{28}$$

Applying (26) in (25), we obtain the bias of T_{GC1}

$$\text{Bias}(T_{GC1}) = \bar{Y}\left[\frac{1}{2}\left(\frac{1}{4} + \alpha_2 + \alpha_2^2\right)\frac{\theta}{\bar{X}^2}\left\{S_x^2 - \frac{2nC_1}{N-1}(\Delta x - nC_1)\right\} - \alpha_2\frac{\theta}{\bar{Y}\bar{X}}\left\{S_{yx} - \frac{n}{N-1}(C_1\Delta y + C_0\Delta x) - 2nC_0C_1\right\}\right] \tag{29}$$

$$\text{Bias}(T_{GC1}) = \bar{Y}\left[\frac{1}{2}\left(\frac{1}{4} + \alpha_2 + \alpha_2^2\right)\frac{\theta R^2}{\bar{Y}^2}\left\{S_x^2 - \frac{2nC_1}{N-1}(\Delta x - nC_1)\right\} - \alpha_2\frac{\theta R}{\bar{Y}^2}\left\{S_{yx} - \frac{n}{N-1}(C_1\Delta y + C_0\Delta x) - 2nC_0C_1\right\}\right] \tag{30}$$

$$\text{Bias}(T_{GC1}) = \frac{\theta R}{\bar{Y}}\left[\frac{1}{2}R\left(\frac{1}{4}+\alpha_2+\alpha_2^2\right)\left\{S_x^2-\frac{2nC_1}{N-1}(\Delta x-nC_1)\right\}-\alpha_2\left\{S_{yx}-\frac{n}{N-1}(\Delta yC_1+\Delta xC_0)-2nC_0C_1\right\}\right] \quad (31)$$

To obtain the MSE of T_{GC1}, we subtract $\bar{Y}$ from both sides of (27), retaining terms in first degree of $e's$, squaring both sides and then taking expectations of both sides, we have:

$$\text{MSE}(T_{GC1}) = E\left(T_{p1ex}-\bar{Y}\right)^2 = \bar{Y}^2E[\bar{e}_0-\alpha_2\bar{e}_1]^2 \quad (32)$$

$$MSE(T_{GC1}) = \bar{Y}^2E[\bar{e}_0^2+\ \alpha_2^2\bar{e}_1^2-2\alpha_2\bar{e}_0\ \bar{e}_1] \quad (33)$$

Substituting equation (26) in (33) will give:

$$MSE(T_{GC1}) = \bar{Y}^2\left[\frac{\theta}{\bar{Y}^2}\left\{S_y^2-\frac{2nC_0}{N-1}(\Delta y-nC_0)\right\}+\frac{\alpha_2^2\theta}{\bar{X}^2}\left\{S_x^2-\frac{2nC_1}{N-1}(\Delta x-nC_1)\right\}-\frac{2\alpha_2\theta}{\bar{Y}\bar{X}}\left\{S_{yx}-\frac{n}{N-1}(\Delta yC_1+\Delta xC_0)-2nC_0C_1\right\}\right] \quad (34)$$

$$MSE(T_{GC1}) = \theta\left[\left(S_y^2\ +\alpha_2^2R^2S_x^2-2\alpha_2RS_{yx}\right)-\frac{2nC_0\Delta y}{N-1}+\frac{2n\alpha_2RC_1\Delta y}{N-1}+\frac{2n^2C_0^2}{N-1}+\frac{2n\alpha_2RC_0\Delta x}{N-1}-\frac{2n\alpha_2^2R^2C_1\Delta x}{N-1}+\frac{2n^2\alpha_2^2R^2C_1^2}{N-1}-\frac{4n^2\alpha_2RC_0\,C_1}{N-1}\right] \quad (35)$$

$$MSE(T_{GC1}) = \theta\left[\left(S_y^2\ +\alpha_2^2R^2S_x^2-2\alpha_2RS_{yx}\right)-\frac{2n}{N-1}(C_0-\alpha_2RC_1)\{\Delta y-\alpha_2R\Delta x-n(C_0-\alpha_2RC_1)\}\right] \quad (36)$$

To obtain the optimum value of C_0 and C_1 we perform partial differential equation of (36) w.r.t C_0 and C_1 and equate each to zero

$$C_0-\alpha_2RC_1 = \frac{\Delta y-\alpha_2R\Delta x}{2n} \quad (37)$$

Also,

$$C_0-\alpha_2RC_1 = \frac{\Delta y-\alpha_2R\Delta x}{2n} \quad (38)$$

In the attempt to perform partial differentiation on the MSE of T_{GC1} with respect to C_0 and C_1,we have one equation with two unknowns so unique, solution is not possible, therefore,

$$C_{0(opt)} = \frac{y_{max} - y_{min}}{2n} = \frac{\Delta y}{2n} \tag{39}$$

$$C_{1(opt)} = \frac{x_{max} - x_{min}}{2n} = \frac{\Delta x}{2n} \tag{40}$$

Substitute (39) and (40) into (36), and simplify we obtain

$$MSE(T_{GC1})_{opt} = \theta\left[\left(S_y^2 + \alpha_2^2 R^2 S_x^2 - 2\alpha_2 R S_{yx}\right) - \frac{2n}{N-1}\left(\frac{\Delta y}{2n} - \alpha_2 R\frac{\Delta x}{2n}\right)\left\{\Delta y - \alpha_2 R\Delta x - n\left(\frac{\Delta y}{2n} - \alpha_2 R\frac{\Delta x}{2n}\right)\right\}\right] \tag{41}$$

$$MSE(T_{GC1})_{opt} = \theta\left[\left(S_y^2 + \alpha_2^2 R^2 S_x^2 - 2\alpha_2 R S_{yx}\right) - \left\{\frac{(\Delta y - \alpha_2 R\Delta x)^2}{2(N-1)}\right\}\right] \tag{42}$$

2.3.1 Special cases of the proposed estimator T_{GC1}

From (42) we try to investigate different common scenarios for the value of α_2

Case (I): $\alpha_2 = -1$,

$$MSE(\bar{y}_{PC})_{opt} = \left[\theta\bar{Y}^2\{C_y^2 + C_x^2 + 2C_y C_x \rho_{yx}\} - \theta\left\{\frac{(\Delta y + R\Delta x)^2}{2(N-1)}\right\}\right] \tag{43}$$

$$MSE(\bar{y}_{PC})_{opt} = \left[\text{MSE}(\bar{y}_P) - \left\{\frac{\theta(\Delta y + \Delta x)^2}{2R(N-1)}\right\}\right] \tag{44}$$

Therefore, equation (44) gives the MSE $(\bar{y}_{PC})$ of the product corrected estimator $(\bar{y}_{PC})$ in the presence of outliers and can be used when the correlation between the study variable and auxiliary variable is less than zero and the lines passes through the origin

Case (II): $\alpha_2 = 0$

$$V(\bar{y}_s) = \theta\left[\bar{Y}^2 C_y^2 - \left\{\frac{(\Delta y)^2}{2(N-1)}\right\}\right] \quad (45)$$

Also, equation (45) is the sample variance for the correction of the sample mean estimator $(\bar{y}_s)$ when there are outliers in the data and suitable when the correlation between the study variable and auxiliary variable is zero

$$V(\bar{y}_s) = \left[V(\bar{y}) - \left\{\frac{\theta(\Delta y)^2}{2(N-1)}\right\}\right] \quad (46)$$

Case (III): $\alpha_2 = 1$,

$$MSE(\bar{y}_{RC}) = \left[\theta\bar{Y}^2\{C_y^2 + C_x^2 - 2C_y C_x \rho_{yx}\} - \theta\left\{\frac{(\Delta y - R\Delta x)^2}{2(N-1)}\right\}\right] \quad (47)$$

More so, equation (47) is the MSE $(\bar{y}_{RC})$ of the ratio corrected estimator $(\bar{y}_{RC})$ when there are outliers and is applicable when the correlation between the study variable and auxiliary variable is greater than zero and the line passes through the origin.

$$MSE(\bar{y}_{RC}) = \left[\text{MSE}(\bar{y}_R) - \theta\left\{\frac{(\Delta y - R\Delta x)^2}{2(N-1)}\right\}\right] \quad (48)$$

Case (Iv): $\alpha_2 = \frac{C_y \rho_{yx}}{C_x}$

$$MSE(Reg.) = \left[\theta\{\bar{Y}^2 C_y^2 + \bar{Y}^2 C_y^2 \rho_{yx}^2 - 2\bar{Y}^2 C_y^2 \rho_{yx}^2\} - \theta\left\{\frac{(\Delta y - \beta\Delta x)^2}{2(N-1)}\right\}\right] \quad (49)$$

where: $\beta = \frac{s_y \rho_{yx}}{s_x}$

$$MSE(\bar{y}_{Reg.}) = \left[\bar{Y}^2 C_y^2(1 - \rho_{yx}^2) - \theta\left\{\frac{(\Delta y - \beta\Delta x)^2}{2(N-1)}\right\}\right] \quad (50)$$

$$MSE(\bar{y}_{Reg.}) = \left[V(\bar{y}_{LR}) - \theta\left\{\frac{(\Delta y - \beta\Delta x)^2}{2(N-1)}\right\}\right] \quad (51)$$

Lastly, equation (51) is the $MSE\left(\bar{y}_{Reg.}\right)$ of the regression estimator $(\bar{y}_{Reg.})$ and is applicable when the correlation between the variables is linear and negative or positive, and the line does not pass through the origin.

Similarly, for the second proposed estimator, applying (25) in (23) we obtained

$$T_{p2} = \left[\bar{Y} + \bar{Y}\bar{e}_0 - \frac{1}{2}\bar{Y}\bar{e}_1 - \frac{1}{2}\bar{Y}\bar{e}_0\,\bar{e}_1 + \frac{7}{8}\bar{Y}\,\bar{e}_1^2\right] \quad (52)$$

Subtracting $\bar{Y}$ from both sides of (52) and taking expectation of both sides, we get the bias of the estimator T_{GC2}

$$\text{Bias}(T_{GC2}) = E(t_{p2} - Y) = E\left[\bar{Y}\bar{e}_0 - \tfrac{1}{2}\bar{Y}\bar{e}_1 - \tfrac{1}{2}\bar{Y}\bar{e}_0\,\bar{e}_1 + \tfrac{7}{8}\bar{Y}\,\bar{e}_1^2\right] \quad (53)$$

Applying (26) in (53), we obtained the bias of T_{GC1}

$$\text{Bias}(T_{GC2}) = \bar{Y}\left[\tfrac{7\theta}{8\bar{X}^2}\left\{S_x^2 - \tfrac{2nC_1}{N-1}(\Delta x - nC_1)\right\} - \tfrac{\theta}{2\bar{Y}\bar{X}}\left\{S_{yx} - \tfrac{n}{N-1}(\Delta y C_1 + \Delta x C_0) - 2nC_0C_1\right\}\right] \quad (54)$$

$$\text{Bias}(T_{GC2}) = \tfrac{\theta}{\bar{X}}\left[\tfrac{7\bar{Y}}{8\bar{X}}\bar{X}^2C_x^2 - \tfrac{1}{2}\bar{Y}C_x\bar{X}\,C_y\rho_{yx} + \tfrac{n}{2(N-1)}\left\{\tfrac{nC_0\Delta x}{2(N-1)} - \tfrac{14n\bar{Y}C_1\Delta x}{8\bar{X}(N-1)} - \tfrac{2n^2C_0C_1}{2(N-1)} + \tfrac{14n^2\bar{Y}C_1^2}{8\bar{X}(N-1)} + \tfrac{n\Delta y C_1}{2(N-1)}\right\}\right] \quad (55)$$

$$\text{Bias}(T_{GC2}) = \left[\bar{Y}\left(\tfrac{7}{8}\theta C_x^2 - \tfrac{1}{2}C_yC_x\,\rho_{yx}\right) + \tfrac{n\theta}{2\bar{X}(N-1)}\left\{\left(C_0 - \tfrac{7\bar{Y}C_1}{2\bar{X}}\right)\Delta x - \left(2C_0 + \tfrac{7\bar{Y}nC_1}{2\bar{X}}\right)nC_1 + C_1\Delta y\right\}\right] \quad (56)$$

To obtain the MSE of T_{GC2}, we subtract $\bar{Y}$ from both sides of (52),retaining terms in first degree of $e's$, squaring both sides and then taking expectations of both sides, we have:

$$T_{GC2} = \left[\bar{Y} + \bar{Y}\bar{e}_0 - \frac{1}{2}\bar{Y}\,\bar{e}_1 - \frac{1}{2}\bar{Y}\bar{e}_0\,\bar{e}_1 + \frac{7}{8}\bar{Y}\,\bar{e}_1^2\right] \quad (57)$$

$$\text{MSE}(T_{GC2}) = E\left(t_{p2} - \bar{Y}\right)^2 = E\left[\bar{Y}\bar{e}_0 - \tfrac{1}{2}\bar{Y}\bar{e}_1\right]^2 \quad (58)$$

$$MSE(T_{GC2}) = \bar{Y}^2E\left[\bar{e}_0^2 + \frac{1}{4}\bar{e}_1^2 - \frac{1}{2}(2\bar{e}_0\bar{e}_1)\right] \quad (59)$$

Let $\frac{1}{2} = a$

$$MSE(T_{GC2}) = \bar{Y}^2E[\bar{e}_0^2 + a^2\bar{e}_1^2 - 2a\bar{e}_0\bar{e}_1] \quad (60)$$

$$MSE(T_{GC2}) = \bar{Y}^2 E\left[\frac{\theta}{\bar{Y}^2}\left\{S_y^2 - \frac{2nC_0}{N-1}(\Delta y - nC_0)\right\} + \frac{a^2\theta}{\bar{X}^2}\left\{S_x^2 - \frac{2nC_1}{N-1}(\Delta x - nC_1)\right\} - \frac{2a\theta}{\bar{Y}\bar{X}}\left\{S_{yx} - \frac{n}{N-1}(\Delta y C_1 + \Delta x C_0) - 2nC_0C_1\right\})\right] \quad (61)$$

$$MSE(T_{GC2}) = \theta\left[\left(S_y^2 + a^2R^2S_x^2 - 2aRS_{yx}\right) - \frac{2nC_0\Delta y}{N-1} + \frac{2naRC_1\Delta y}{N-1} + \frac{2naRC_0\Delta x}{N-1} - \frac{2na^2R^2C_1\Delta x}{N-1} + \frac{2n^2C_0^2}{N-1} + \frac{2n^2a^2R^2C_1^2}{N-1} - \frac{4n^2aRC_0\,C_1}{N-1}\right] \quad (62)$$

$$MSE(T_{GC2}) = \theta\left[\left(S_y^2 + a^2R^2S_x^2 - 2aRS_{yx}\right) - \frac{2n}{N-1}(C_0 - aRC_1)\{\Delta y - aR\Delta x - n(C_0 - aRC_1)\}\right] \quad (63)$$

but

$$C_{0(opt)} = \frac{y_{max} - y_{min}}{2n} = \frac{\Delta y}{2n} \quad (64)$$

$$C_{1(opt)} = \frac{x_{max} - x_{min}}{2n} = \frac{\Delta x}{2n} \quad (65)$$

Substitute (64) and (65) into (63) and simplify, we obtained

$$MSE\left(T_{p2ex}\right) = \theta\left[\left\{S_y^2 + a^2R^2S_x^2 - 2aRS_{yx}\right\} - \frac{2n}{N-1}\left(\frac{\Delta y}{2n} - aR\frac{\Delta x}{2n}\right)\left\{\Delta y - aR\Delta x - n\left(\frac{\Delta y}{2n} - \frac{aR\Delta x}{2n}\right)\right\}\right] \quad (66)$$

$$MSE(T_{GC2}) = \theta\left[\left\{S_y^2 + a^2R^2S_x^2 - 2aRS_{yx}\right\} - \left\{\frac{(\Delta y - aR\Delta x)^2}{2(N-1)}\right\}\right] \quad (67)$$

But $a = \frac{1}{2}$

$$MSE(T_{GC2}) = \theta\left[\bar{Y}^2\left\{C_y^2 + \frac{1}{4}C_x^2 - C_yC_x\rho_{yx}\right\} - \left\{\frac{(2\Delta y - R\Delta x)^2}{8(N-1)}\right\}\right] \quad (68)$$

$$MSE(T_{GC2}) = \left[MSE(T_{ER}) - \theta\left\{\frac{(2\Delta y - R\Delta x)^2}{8(N-1)}\right\}\right] \quad (69)$$

The MSE of (69) has a correction factor of the ratio exponential estimator and it can be used for the correction of the exponential ratio estimator in when there are outliers in the data.

3 Results and discussion

3.2 Comparison of estimators

Comparison with the generalized proposed (mixed optimum) estimator

(a) Comparison of proposed estimator T_{RC} type Estimator

(i) Mean Per Unit Estimator [(5) and (48)]

$V(\bar{y}) - M(T_{RC}) > 0$ or if

$$\rho_{yx} > \frac{1}{2}\left[\frac{RS_x}{S_y} - \frac{(\Delta y - R\Delta x)^2}{2RS_yS_x(N-1)}\right] \quad (70)$$

The corrected ratio estimator of (48) is more efficient than mean per unit estimator given in (5)

(ii) Usual ratio estimator [(8) and (48)]

$M(\bar{y}_R) - M(T_{RC}) > 0$ or if

$$\frac{(\Delta y - R\Delta x)^2}{2RS_yS_x(N-1)} > 0 \quad (71)$$

The corrected ratio estimator of (48) is more efficient than its contemporary ratio estimator of (8) .

(iii) Usual exponential ratio estimator [(16) and (48)]

$M(\bar{y}_{ER}) - M(T_{RC}) > 0$ or if

$$\rho_{yx} > -\frac{1}{2}\left[\frac{5RS_x}{2S_y} + \frac{(\Delta y - R\Delta x)^2}{RS_yS_x(N-1)}\right] \quad (72)$$

Estimator (48) is more efficient than estimator(16)

(b) Comparison of proposed estimator T_{PC} type Estimator

Also, we compare the efficiency of the corrected product estimator with its usual counterpart.

(i) Mean Per Unit Estimator [(5) and (44)

$V(\bar{y}) - M(T_{PC}) > 0$ or if

$$\rho_{yx} > \frac{1}{2}\left[\frac{RS_x}{S_y} - \frac{(\Delta y + R\Delta x)^2}{2RS_yS_x(N-1)}\right] \qquad (73)$$

The corrected product estimator of (44) is more efficient than the mean per unit estimator.

(ii) Usual product estimator [(11) and (44)]

$M(\bar{y}_P) - M(T_{PC}) > 0$ or if

$$\frac{(\Delta y + R\Delta x)^2}{2RS_yS_x(N-1)} > 0 \qquad (74)$$

Corrected product estimator of (44) is more efficient than it product counterpart of (11)

(iii) Usual exponential product estimator [(16) and (44)]

$M(\bar{y}_{EP}) - M(T_{PC}) > 0$ or if

$$\rho_{yx} > -\frac{1}{2}\left[\frac{5RS_x}{2S_y} + \frac{(\Delta y + R\Delta x)^2}{RS_yS_x(N-1)}\right] \qquad (75)$$

The corrected exponential product estimator of(44) is more efficient than the usual product estimator of (16)

(c) Comparison of proposed estimator T_{Reg} type Estimator **[**(5) and (51)]

(i) Mean Per Unit Estimator

$V(\bar{y}) - V(\bar{y}_{lrc}) > 0$ or if

$$\rho_{yx}^2 > -\frac{(\Delta y - \beta\Delta x)^2}{2S_{yx}^2(N-1)} \qquad (76)$$

From (74) it is obvious that the regression estimator of (51) is more efficient than the mean per unit estimator of (5)

(iv) Usual regression estimator [(14) and (51)]

$V(\bar{y}_{lr}) - V(\bar{y}_{lrc}) > 0$ or if

$$\frac{(\Delta y - \beta\Delta x)^2}{2RS_yS_x(N-1)} > 0 \qquad (77)$$

It is showed that the corrected regression estimator of (51) is more efficient than the usual regression estimator of (14)

Comparison with the mixed non-optimum estimator

(a) comparison of proposed estimator T_{GC2}

(i) Mean Per Unit Estimator [(5) and (69)]

$V(\bar{y}) - M(T_{GC2}) > 0$ or if

$$\rho_{yx} > \frac{1}{4}\left[\frac{RS_x}{S_y} - \frac{(2\Delta y - R\Delta x)^2}{2RS_yS_x(N-1)}\right] \quad (78)$$

It is clear from (76) that corrected exponential ratio estimator is more efficient than the exponential ratio estimator.

(ii) Usual exponential ratio [(16) and (69)]

$M(\bar{y}_{ER}) - M(T_{GC2}) > 0$ or if

$$\frac{(2\Delta y - R\Delta x)^2}{8R(N-1)} > 0 \quad (79)$$

The corrected exponential ratio estimator of (69) is more efficient than the usual exponential ratio estimator of (16).

3.3 Numerical illustration

The Percent Relative Efficiency (PRE) is a statistical tool that will be used to measure the efficiency of estimators. Thus,

$$\text{PRE} = \frac{Var\ (\bar{y})}{Var\ (\bar{y}_*) \text{ or } MSE(\bar{y}_*)}\ X\ 100$$

for $*$=1,2,3,4,5,6,7,8,9,10 and $\bar{y}_1 = \bar{y}$, $\bar{y}_2 = \bar{y}_R$,$\bar{y}_3 = \bar{y}_P$, $\bar{y}_4 = \bar{y}_{ER}$,$\bar{y}_5 = \bar{y}_{EP}$, $\bar{y}_6 = \bar{y}_{lr,}$

$\bar{y}_7 = \bar{y}_{RC'}$ $\bar{y}_8 = \bar{y}_{PC}$, $\bar{y}_9 = \bar{y}_{lrC}$, $\bar{y}_{10} = \bar{y}_{ERC}$

Table 1 shows the data that is used to assess the efficiency of the estimators discuss above.

Table 1: Data sets for the Empirical Study

Parameter	Population I	Population II	Population III	Population IV
N	27	34	49	36
n	12	12	12	12
$\bar{Y}$	11.25185	199.441	127.7959	14.77778
$\bar{X}$	10.41111	208.882	103.1429	2.798333
y_{max}	14.8	634	634	33
y_{min}	7.9	6	46	6
x_{max}	14.5	564	507	3.82
x_{min}	6.5	5	2	1.81
S_y	2.025586	150.215	123.1212	6.1788
S_x	2.220586	150.506	104.4051	0.5919
S_{yx}	4.454	22158.05	12619.78	−2.477
ρ_{yx}	0.990	0.980	0.98	−0.6772

Source: Khan and Shabbir (2013)

Table 2 below is the population description of the data of table 1, it states the study variable and the auxiliary variable

Table 2: Population description

Variable	Population I	Population II	Population III	Population IV
y	Milk yield in kg after new food	Area under wheat crop in 1964	Population size in 1930 (in 1000)	Weekly time (hours) spent in nonacademic activities
x	Yield in kg Before new yield	Area under wheat crop in 1963.	Population size in 1920 (in 1000).	Overall grade point average (4.0 bases).

Source: Khan and Shabbir (2013)

The efficiency of the proposed estimators and its contemporary is showed in table 3 below:

Table 3: Percentage Relative efficiency of different estimators with respect to mean per unit estimator

	POPULATION I		POPULATION II		POPULATION III		POPULATION IV	
ESTIMATOR	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}$	0.189956	**100.00**	1216.716	**100.00**	953.872	**100.00**	2.12098	**100.00**
$\bar{y}_R$	0.010895	**1743.56**	48.6438	**2501.28**	39.0462	**2442.93**	4.11721	**51.5149**
$\bar{y}_P$	0.902317	**21.0521**	4611.82	**26.3825**	3974.67	**23.9988**	1.21035	**175.236**
$\bar{y}_{ER}$	0.033763	**562.615**	354.300	**343.414**	233.212	**409.014**	2.98339	**71.0928**
$\bar{y}_{EP}$	0.479474	**39.6176**	2635.89	**46.1596**	2201.02	**43.3376**	1.52996	**138.629**
$\bar{y}_{lr}$	0.003780	**5025.13**	48.1819	**2525.25**	37.7733	**2525.25**	1.14830	**184.706**
$\bar{y}_{RC}$	0.008180	**2322.06**	41.3840	**2940.07**	38.1144	**2502.66**	3.90413	**54.3265**
$\bar{y}_{PC}$	0.899603	**21.1156**	4604.56	**26.4241**	3973.74	**24.0044**	0.997275	**212.677**
$\bar{y}_{lrC}$	0.003686	**5152.93**	42.7899	**2843.46**	37.7607	**2526.09**	1.01845	**208.255**
$\bar{y}_{ERC}$	0.027851	**682.052**	247.751	**491.105**	183.589	**519.569**	2.60992	**81.2659**

The analysis reveals a significant improvement in the precision of estimators that incorporate a correction factor for outliers, compared to their standard counterparts that do not account for such adjustments.

From table 3, in the first population, the estimators that uses an outlier correction factor $\bar{y}_{RC}$, $\bar{y}_{Pc}$, $\bar{y}_{lrC}$, $\bar{y}_{ERC}$ demonstrated it efficiency over the conventional estimators $\bar{y}_{R}$,$\bar{y}_{P}$, $\bar{y}_{lr}$,$\bar{y}_{ER}$ that do not incorporate such adjustments respectively. Also, in population I, II and III estimators $\bar{y}_{RC}$ and $\bar{y}_{ERC}$ performed better than $\bar{y}_{PC}$, while in population IV $\bar{y}_{PC}$ established it superiority over others estimators. But it is obvious that estimator $\bar{y}_{lrC}$ adapted well in all the four populations considered in this study.

Conclusions and Recommendations

In this study a generalized estimator under maximum and minimum values using auxiliary variable was developed that yielded the mean squared error for the correction of the conventional estimators when there are outliers in both the study and auxiliary variable. From the analysis, it can be observed that the ratio and exponential ratio corrected estimator performed better when there is positive correlation between the study variable and the auxiliary variable, while the product exhibited superior performance in the case of negative correlation. The regression estimator, is more flexible and performed well for both positive and negative correlations. Consequently, careful selection of the constant is essential, which depends on the nature of the data. Specifically, when the correlation is negative and the regression line passes through the origin α_2 equal -1; when there is no correlation α_2 equals 0 ; when the correlation is positive and the line passes through the origin α_2 equals 1 and when correlation is either positive or negative but the line does not pass through the origin α_2 is given by $\alpha_2 = \frac{C_y\rho_{yx}}{C_x}$., which is known as the regression coefficient. Therefore, the generalized estimator which accounts for the correction of extreme values demonstrated greater efficiently compared to the conventional ratio, product, sample mean and regression estimators for the different values of the constant.

Thus, the proposed estimators offer significant advantages over traditional methods for practical applications. Their versatility and ability to encompass various known estimators through appropriate adjustment of the constant make them adaptable to different data conditions, enhancing their overall utility.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

References

Abiodun, Y. M.,Ahmed, A., Ishaq, O. O., & Beki, D. O. (2021). An efficient exponential type estimator for estimating finite population mean under simple random sampling. *Anale. Seria Informatică*, *XIX*(fasc.), 46-51

Al-Hossain, Y., & Khan, M. (2014). Efficiency of ratio, product and regression estimators under maximum and minimum values using two auxiliary variables. *Journal of Applied Mathematics, 2014,* Article 1–6.

Cekim, H. O., & Cingi, H. (2016). Some estimator types for population mean using linear transformation with the help of the minimum and maximum values of the auxiliary variable. *Hacettepe Journal of Mathematics and Statistics, 46*(4). https://doi.org/10.15672/hjms.201510114186

Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science, 30*(2), 262–275

Cochran, W. G. (1942). Sampling theory when the sampling units are of unequal sizes. *Journal of the American Statistical Association, 37*(218), 199–212.

Darez, U., Shabbir, J., & Khan, H. (2018). Estimation of finite population mean by using minimum and maximum values in stratified random sampling. *Journal of Modern Applied Statistical Methods, 17*(1), Article eP2730. https://doi.org/10.22237/jmasm/1525133040

Hanif, M., Hamad, N., & Shahbaz, M. Q. (2010). Some new regression type estimators in two phase sampling. *World Applied Sciences Journal, 8*(7), 799–803.

Kanwai, M. E., Asiribo, O. E., & Isah, A. (2016). Use of auxiliary variables and asymptotically optimum estimators in double sampling. *International Journal of Statistics and Probability, 5*(3), 55–62.

Khan, M. (2015). Improvement in estimating the finite population mean under maximum and minimum values in double sampling scheme. *Journal of Statistics Applications & Probability Letters, 2*(2),115–121. http://www.naturalspublishing.com/ContIss.asp?IssID=262

Khan, M., & Shabbir, J. (2013). Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values. *The Scientific World Journal, 2013,* Article 431868. https://doi.org/10.1155/2013/431868

Mohanty, S., & Sahoo, J. (1995). A note on improving the ratio method of estimation through linear transformation using certain known population parameters. *Sankhyā: The Indian Journal of Statistics, Series B, 57*(1), 93–102. https://www.jstor.org/stable/25052879

Olayiwola, O. M., Apantaku, F. S., Imarhia, F. O., Yusuf, K. M., Ogunsola, I. A., & Olawoore, S. A. (2021). A modified estimator for population parameters in the presence of outliers. *Uniport Journal of Engineering and Scientific Research, 6*(1), 85–89.

Samiuddin, M., & Hanif, M. (2007). Estimation of population mean in single phase and two phase sampling with or without additional information. *Pakistan Journal of Statistics*, *23*(2), 99–118.

Shahbaz, S., Zubair, M., & Shafique, N. (2014). A new regression estimator based on two auxiliary variables. *Research Journal of Applied Sciences, Engineering and Technology*, *8*(2), 251–252.

Singh, H. P., & Espejo, M. R. (2007). Double sampling ratio-product estimator of a finite population mean in sample surveys. *Journal of Applied Statistics*, *34*(1), 71–85.

Singh, R., Mishra, P., Ahmed, A., & Khare, S. (2020). Exponential type estimator for estimating finite population mean. *International Journal of Computational Theoretical Statistics*, *7*(1),37- 41. http://dx.doi.org/10.12785/ijcts/070104

Swain, A. K. P. C. (2012). On some classes of modified ratio type and regression type estimators in two phase sampling. *Pakistan Journal of Statistics*, *28*(4), 445–466.

Walia, G. S., Kaur, H., & Sharma, M. (2015). Ratio type estimator of population mean through efficient linear transformation. *American Journal of Mathematics and Statistics*, *5*(3), 144–149. http://article.sapub.org/10.5923.j.ajms.20150503.06.htmlss