Naïve Family of Estimators for Population Mean Using Auxiliary Parameters in Survey Sampling[i]

Surendra Kumar[1], Brij Mohan Yadav[2], Rekha Srivastava[3], Gautam Gupta[4] and Gagan Kumar[5*]

[1]*Department of Mathematics, Mahamaya Govt. Degree College Mahona, Lucknow, India*

[2]*Department of Mathematics, M P Government Postgraduate College Hardoi, India*

[3]*Department of Sociology, Govt. PG College Musafirkhana, Amethi, India*

[4]*Department of Sociology, Dr. Ambedkar Govt. PG College Unchahar, Raibareli, India*

[5]*Department of Economics, Govt. Degree College Pihani, Hardoi, India*

**Corresponding Author (gkb.gdc@gmail.com)*

Abstract: - In order to estimate the population, mean of a study variable under a straightforward random sampling without replacement framework, this paper presents a naive broad family of ratio estimators that make use of a variety of auxiliary measures. The study looks at particular instances that include auxiliary data like the median, quartile deviation, and coefficient of variation. A first-order approximation is maintained for the introduced estimators' bias and Mean Square Error (MSE). Furthermore, real-world data is used to validate the performance of the introduced estimators, and theoretical conditions are supplied for comparing their efficiency with those of existing estimators. The suggested estimators are more effective than other ratio-based estimators, as shown by numerical analysis, making them suitable for use in a variety of real-world situations.

Keywords: *Study Variable, Auxiliary Parameter, Median, Bias, Mean Square Error.*

1. Introduction

Although it is always better to calculate the parameter in question rather than estimate it, doing so can be costly and time-consuming when the population is large. A labor, money, and time efficient alternative to thorough enumeration is sampling. Since the associated statistics are the best estimate for the parameter under study, the sample mean $\bar{y}$ is most suited for the population mean. The auxiliary variable X is crucial while estimating the parameters of the primary variable Y. If it has a positive connection with Y, ratio type estimators are used for greater population mean $\bar{Y}$ estimates. On the other hand, product type estimators are used to enhance the estimation of when X and Y have a negative correlation. When dealing with large populations or when it is not feasible to collect data from every unit, sampling procedures are commonly employed. These methods provide a practical means of examining the characteristics of a target variable by examining a subset of the population. Sample statistics are typically employed to estimate population parameters; for example, $\bar{y}$ is a frequently used estimator for $\bar{Y}$ because of its objectivity, even when there may be significant volatility. Much work has been done to improve existing methods in order to create estimators that are more effective. Many of these efforts focus on using supplementary data to increase the precision of sample-based population parameter calculations. Many authors from all over the globe used the auxiliary information in different forms for enhanced estimation of $\bar{Y}$ and suggested various efficient estimators of $\bar{Y}$.

Utilized data on X that has a strong positive correlation with Y to create the classical ratio estimator [1]. [2] suggested a better estimator for by adding X data from sample surveys. [3] used the known coefficient of variation (CV) of X to create a modified ratio estimator. [4] investigated techniques for effective estimating $\bar{Y}$ using regression and ratio estimators. While [6] worked on an elevated estimator for by combining the known CV and coefficient of kurtosis of X, [5] suggested a novel estimator for $\overline{Y}$ by using the coefficient of kurtosis. In order to improve $\overline{Y}$ estimate, [7] improved the ratio estimator by employing the known correlation coefficient, and [8] concentrated on creating better ratio estimators. modified ratio estimators that use unconventional auxiliary parameters to estimate $\overline{Y}$ with efficiency [9]. [10] identified several estimators as belonging to their family and presented a general family of ratio estimators for $\overline{Y}$. While [12] enhanced the estimate of $\overline{Y}$ using the known coefficient of skewness of X, [11] created effective estimators for $\overline{Y}$. suggested an effective estimator for using X's known quartile information [13]. [14] used the quartile deviation of X with the coefficient of skewness to create new ratio estimators for $\overline{Y}$. Lastly, a generic family of estimators based on auxiliary parameters was developed by [15].

[16] suggested elevated ratio estimators for $\overline{Y}$ using non-traditional measures of dispersion while [17] created an effective class of ratio estimators for $\overline{Y}$ using known auxiliary parameters. By adding auxiliary factors, [18] and [19] presented enhanced estimators that took non-response cases into consideration. A generalized class of estimators for $\overline{Y}$ the effective estimate of was proposed by [20]. In order to improve the estimation of average peppermint yield using known auxiliary information (X), [22] presented a naive class of estimators. [21] concentrated on the effective estimation of average paddy output using robust measures. [23] used simulation simulations to investigate resilient ratio-type estimators for simple random sampling (SRS).

[24] proposed an advanced class of estimators for $\overline{Y}$ leveraging known auxiliary parameters, while [25] introduced a refined family of estimators utilizing X information for improved estimation of $\overline{Y}$. Numerous other researchers have also contributed to the development of enhanced estimators for $\overline{Y}$ by utilizing known auxiliary parameters. The remainder of this paper is organized into various sections. [24] suggested an advanced class of estimators for leveraging known auxiliary parameters while a revised family of estimators using X information was presented by [25] for better estimation of $\overline{Y}$. The development of improved estimators for using known auxiliary parameters has also been aided by numerous other researchers. The rest of this work is divided into different sections.

The remaining work is broken down into numerous sections, including the comparison of theoretical efficiency, the suggested estimator, the review of estimators, the numerical analysis, and the comparison of the outcomes.

2. Review of Estimators

The most suitable statistic for $\overline{Y}$ is the $\overline{y}$ but with large variance and we search for the estimator with least variance or MSE. This is obtained by using the auxiliary variable having high degree of correlation with the main variable. Table-1 represents different estimators of $\overline{Y}$ using the known auxiliary parameters along with their MSEs.

Table-1: Various estimators and their MSEs

S.No.	Estimator	MSE
1.	$t_R = \overline{y}\left(\frac{\overline{X}}{\overline{x}}\right)$ Cochran [1]	$MSE(t_R) = \lambda \overline{Y}^2[C_y^2 + R^2C_x^2 - 2RC_{yx}]$
2.	$t_1 = \overline{y}\left(\frac{\overline{X}+C_x}{\overline{x}+C_x}\right)$ Sisodia and Dwivedi [3]	$MSE(t_1) = \lambda \overline{Y}^2[C_y^2 + R_1^2C_x^2 - 2R_1C_{yx}]$

3.	$t_2 = \bar{y}\left(\dfrac{\bar{X}+\beta_2}{\bar{x}+\beta_2}\right)$ Singh and Kakran [5]	$MSE(t_2) = \lambda \bar{Y}^2[C_y^2 + R_2^2 C_x^2 - 2R_2 C_{yx}]$
4.	$t_3 = \bar{y}\left(\dfrac{\bar{X}\beta_2 + C_x}{\bar{x}\beta_2 + C_x}\right)$, $t_4 = \bar{y}\left(\dfrac{\bar{X}C_x + \beta_2}{\bar{x}C_x + \beta_2}\right)$ Upadhyaya and Singh [6]	$MSE(t_i) = \lambda \bar{Y}^2[C_y^2 + R_i^2 C_x^2 - 2R_i C_{yx}]$, $i = 2,3$
5.	$t_5 = \bar{y}\left(\dfrac{\bar{X}\dfrac{\beta_1}{R_1} + QD}{\bar{x}\dfrac{\beta_1}{R_1} + QD}\right)$ Neelam et al. [26]	$MSE(t_5) = \lambda \bar{Y}^2[C_y^2 + R_5^2 C_x^2 - 2R_5 C_{yx}]$,

Where,

$$\lambda = \left(\frac{1}{n} - \frac{1}{N}\right), \quad \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2, \quad S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X}), \quad R_1 = \frac{\bar{X}}{\bar{X}+C_x}, \quad R_2 = \frac{\bar{X}}{\bar{X}+\beta_2},$$

$$R_3 = \frac{\bar{X}\beta_2}{\bar{X}\beta_2 + C_x}, \quad R_4 = \frac{\bar{X}C_x}{\bar{X}C_x + \beta_2}, \quad R_5 = \frac{\bar{X}\dfrac{\beta_1}{R_1}}{\bar{X}\dfrac{\beta_1}{R_1} + QD}.$$

3. Proposed estimator

Motivated by [26] and many other authors in the literature, we have introduced a generalized estimator for estimation of $\bar{Y}$ as,

$$t_p = \bar{y}\left(\frac{\bar{X}\dfrac{\beta_1}{R_1} + QD}{\bar{x}\dfrac{\beta_1}{R_1} + QD}\right)^{\delta} \tag{1}$$

Where, δ is the scalar to be obtained such that MSE of t_p is least.

To study the large sampling property of t_p, the following approximations are used,

$\bar{y} = \bar{Y}(1+e_0)$, $\bar{x} = \bar{X}(1+e_1)$, with $E(e_0) = E(e_1) = 0$ and $E(e_0^2) = \lambda C_y^2$, $E(e_1^2) = \lambda C_x^2$, $E(e_0 e_1) = \lambda C_{yx}$

Representing t_p using e_0 and e_1, we obtain,

$$t_p = \bar{Y}(1+e_0)\left(\frac{\bar{X}\frac{\beta_1}{R_1}+QD}{\bar{X}(1+e_1)\frac{\beta_1}{R_1}+QD}\right)^{\delta}$$

$$= \bar{Y}(1+e_0)\left(\frac{\bar{X}\frac{\beta_1}{R_1}+QD}{\bar{X}\frac{\beta_1}{R_1}+QD+\bar{X}\frac{\beta_1}{R_1}e_1}\right)^{\delta}$$

$$= \bar{Y}(1+e_0)\left(1+\frac{\bar{X}\frac{\beta_1}{R_1}e_1}{\bar{X}\frac{\beta_1}{R_1}+QD}\right)^{-\delta}$$

$$= \bar{Y}(1+e_0)(1+R_5 e_1)^{-\delta}$$

$$= \bar{Y}(1+e_0)\left[1-\delta R_5 e_1+\frac{\delta(1+\delta)}{2}R_5^2 e_1^2-\ldots\right]$$

$$t_p = \bar{Y}\left[1+e_0-\delta R_5 e_1-\delta R_5 e_0 e_1+\frac{\delta(1+\delta)}{2}R_5^2 e_1^2-\ldots\right]$$

Subtracting $\bar{Y}$ on both sides of above equation, we have,

$$t_p-\bar{Y} = \bar{Y}\left[e_0-\delta R_5 e_1-\delta R_5 e_0 e_1+\frac{\delta(1+\delta)}{2}R_5^2 e_1^2-\ldots\right] \quad (2)$$

We have bias of t_p as,

$$Bias(t_p) = \lambda\bar{Y}\left[\frac{\delta(1+\delta)}{2}R_5^2 C_x^2-\delta R_5 C_{yx}\right] \quad (3)$$

Squaring on both sides of (2) and for order one, we have,

$$MSE(t_p) = \bar{Y}^2 E[e_0^2+\delta^2 R_5^2 e_1^2-2\delta R_5 e_0 e_1+\ldots]$$

Putting values of different expectation, we get the MSE of t_p, as,

$$MSE(t_p) = \lambda \bar{Y}^2 [C_y^2 + \delta^2 R_5^2 C_x^2 - 2\delta R_5 C_{yx}] \tag{4}$$

Differentiating $MSE(t_p)$ with respect to δ and putting it equal to zero, we get,

$\frac{\partial}{\partial \delta} MSE(t_p) = 0$, which gives,

$$\delta = \frac{C_{yx}}{R_5 C_x^2} = \delta_{opt} \tag{5}$$

Putting the value of δ_{opt} in (4), we get the minimum value of $MSE(t_p)$ as,

$$MSE_{\min}(t_p) = \lambda \bar{Y}^2 \left[C_y^2 - \frac{C_{yx}^2}{C_x^2} \right] \tag{6}$$

4. Efficiency Comparison

A theoretical comparison of efficiencies of different estimators with the proposed estimator has been presented in Table-2 and the efficiency condition has also been obtained.

Table-2: Efficiency conditions of t_p over competing estimators

S.No.	Estimator	MSE
1.	$t_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)$, Cochran [1]	$\left[C_y^2 - \frac{C_{yx}^2}{C_x^2}\right] - [R^2 C_x^2 - 2RC_{yx}] > 0$
2.	$t_1 = \bar{y}\left(\frac{\bar{X} + C_x}{\bar{x} + C_x}\right)$, Sisodia and Dwivedi [3]	$\left[C_y^2 - \frac{C_{yx}^2}{C_x^2}\right] - [R_1^2 C_x^2 - 2R_1 C_{yx}] > 0$
3.	$t_2 = \bar{y}\left(\frac{\bar{X} + \beta_2}{\bar{x} + \beta_2}\right)$, Singh and Kakran [5]	$\left[C_y^2 - \frac{C_{yx}^2}{C_x^2}\right] - [R_2^2 C_x^2 - 2R_2 C_{yx}] > 0$
4.	$t_3 = \bar{y}\left(\frac{\bar{X}\beta_2 + C_x}{\bar{x}\beta_2 + C_x}\right)$, $t_4 = \bar{y}\left(\frac{\bar{X}C_x + \beta_2}{\bar{x}C_x + \beta_2}\right)$ Upadhyaya and Singh [6]	$\left[C_y^2 - \frac{C_{yx}^2}{C_x^2}\right] - [R_i^2 C_x^2 - 2R_i C_{yx}] > 0$, $i = 2,3$

5.	$t_5 = \bar{y}\left(\dfrac{\bar{X}\dfrac{\beta_1}{R_1}+QD}{\bar{x}\dfrac{\beta_1}{R_1}+QD}\right)$, Neelam et al. [26]	$\left[C_y^2 - \dfrac{C_{yx}^2}{C_x^2}\right] - [R_5^2 C_x^2 - 2R_5 C_{yx}] > 0$,

5. Numerical Study

To verify the theoretical results, we have considered the following population, given in [26]. The various parameters of the above population are shown in Table-1.

Table-1: Parameters of the population

$N = 49$	$n = 20$	$\bar{Y} = 116.1633$	$\bar{X} = 98.6735$	$S_y = 98.8286$
$S_x = 102.9709$	$S_{yx} = 7021.7642$	$C_y = 0.8508$	$C_x = 1.0436$	$M_d = 64$
$\beta_2 = 5.9878$	$QD = 78.50$	$\rho = 0.69$	$\lambda = 0.02959$	$C_{yx} = 0.6126$

The MSE and Percentage Relative Efficiency (PRE) of the estimators with respect to t_R are presented in Table-2.

Table-2: MSE and PRE of the estimators

S. No.	Estimator	MSE	PRE
1.	t_R	232.2501	100.00
2.	t_1	228.3407	101.71
3.	t_2	212.0423	109.53
4.	t_3	231.5848	100.29
5.	t_4	212.7862	109.15
6.	t_5	154.8349	150.00
7.	$(t_p)_{opt}$	151.4416	153.36

The graphs and MSE and PRE are presented in Figure-1.

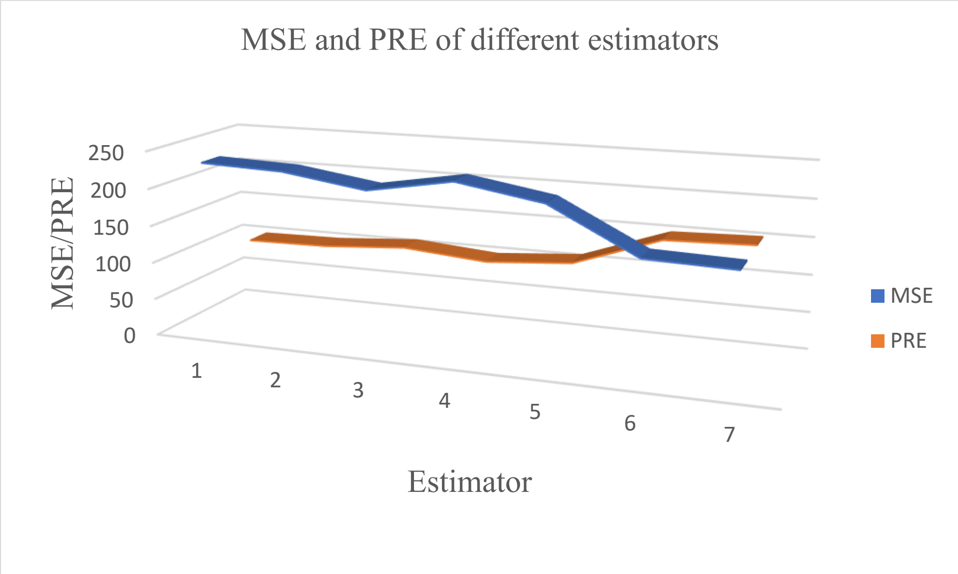

Figure-1: MSE and PRE of different estimators

6. Result and Conclusion

For better estimate in this work, we presented a generalized class of ratio estimators. The first-order approximation of the bias and MSE is obtained. Efficiency requirements are determined by theoretically comparing the performance of the introduced estimators with that of the current estimators. An actual dataset is used in order to validate these theoretical findings. The new estimator's MSE is 151.4416, as Table 2 clearly shows, whilst the competing estimators' MSEs fall between [154.8349, 232.2501]. Furthermore, the PRE of the recommended estimate is 153.36, but the PRE of the competing estimators is within the range [100.00, 150.00]. In Figure-1, these results are also shown as graphs. The suggested estimator is thus the most successful of the compared estimators since it obtains the lowest MSE and the highest PRE, as indicated in Table 2. As a result, the given estimator can be used efficiently for precise estimation in a variety of applications. It should be mentioned that more effective modified estimators may be created in the future and applied in a range of fields. For instance, the average production can be estimated using local information, the factory average production can be estimated using units as auxiliary information, the current year average can be estimated using last year's information, etc.

Refrences

[1] Cochran, W. G. (1940). Sampling Techniques, Third Edition, Wiley Eastern Limited.

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

[3] Sisodia, B. S., & Dwivedi, V. K., (1981). A modified ratio estimator using coefficient of variation of auxiliary variable, Journal of Indian Society AgriculturalStatistics.33, 13–18.

[4] Rao, T. J. (1991). On certain methods of improving ratio and regression estimators, Communication in Statistics Theory and Method, (20), 3325-3340.

[5] Singh, H.P., & Kakran, M.S. (1993). A Modified Ratio Estimator Using Known Coefficient of Kurtosis of an Auxiliary Character. (unpublished).

[6] Upadhyaya, L. N., & Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean, Biometrical Journal, 41 (5), 627–636.

[7] Singh, H. P., & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population means, Statistics in Transition, 6(4), 555–560.

[8] Kadilar, C. & Cingi, H. (2004). Ratio estimators in simple random sampling, Applied mathematics and computation, 151(3), 893-902.

[9] Yadav, S. K., Mishra, S. S., Shukla, A. K., Kumar, S., Singh, R.S. (2016). Use of Non-Conventional Measures of Dispersion for Improved Estimation of Population Mean, American Journal of Operational Research, 6(3), 69-75.

[10] Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., Smarandache, F. A. (2007). general family of estimators for estimating population mean using known value of some population parameter(s), Far East J. Theor. Statist., (22), 181-191.

[11] Koyuncu, N., & Kadilar, C. (2009). Efficient estimators for the population mean, Hacettepe Journal of Mathematics and Statistics, 38(2), 217-225.

[12] Yan, Z., & Tian, B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable, ICICA 2010, Part II, CCIS 106, 103–110.

[13] Subramani, J. & Kumarapandiyan, G. (2012). Modified Ratio Estimators for Population Mean Using Function of Quartiles of Auxiliary Variable, Bonfring International Journal of Industrial Engineering and Management Science, 2(2), 105-110.

[14] Jeelani, M. I., Maqbool, S. & Mir, S.A. (2013). Modified ratio estimators of population mean using linear combination of coefficient of skewness and quartile deviation, International Journal of Modern Mathematical Sciences. 6, 174–183.

[15] Subramani, J. (2013). Generalized Modified Ratio Estimator for Estimation of Finite Population Mean, Journal of Modern Applied Statistical Methods, 12(2), 121-155.

[16] Abid, M., Abbas, N., Sherwani, R.A.K & Nazir, H.Z. (2016). Improved Ratio Estimators for the Population Mean Using Non-Conventional Measures of Dispersion, Pakistan Journal of Statistics and Operations Research,12(2), 353-367.

[17] Subzar, M., Maqbool, S., Raja, T.A. (2017). An improved class of ratio estimators forestimating population mean using auxiliaryinformation in survey sampling, Journal of Reliability and Statistical Studies, 10(2), 65-82

[18] Unal, C., & Kadilar, C. (2019). Improved family of estimators using exponential function for the population mean in the presence of non-response, Communications in Statistics-Theory and Methods. 1-12.

[19] Singh, G.N., & Usman, M. (2019). Efficient combination of various estimators in the presence of non-response, Communications in Statistics-Simulation and Computation, 1-35.

[20] Priam, R. (2019). Visualization of generalized mean estimators using auxiliary information in survey sampling, Communications in Statistics-Theory and Methods, 1-22.

[21] Yadav, S.K., Baghel, S., Saxena, S., & Singh, A.K. (2020). Estimation of Average Paddy Production of Pira Nagar Village at Barabanki District in India, Journal of Reliability and Statistical Studies., 13, 1, 127-148.

[22] Yadav, S.K., Sharma, D.K., & Brown, K. (2021). New Class of Estimators for Enhanced Estimation of Average Production of Peppermint Yield Utilizing Known Auxiliary Variable, International Journal of Mathematics in Operational Research, 20 (2), 281-295.

[23] Zaman, T., Bulut, H. & Yadav, S.K. (2022). Robust ratio-type estimators for finite population mean in simple random sampling: A simulation study, Concurrency and Computation: Practice and Experience, 34(25), e7273. https://doi.org/10.1002/cpe.7273

[24] Ali, M., Yadav, S.K., Gupta, R.K., Kumar, S., & Singh, L. (2023). Enhancement of Class of Estimators for Estimating Population Mean using Known Auxiliary Parameters, International Journal of Agricultural and Statistical Sciences, 19(2), 527-534.

[25] Singh, L., Malik, D., Kumar, M., & Yadav, S.K. (2024). Refined Batch of Estimators for Estimating Population Mean with the help of known Auxiliary Parameters, International Journal of Agricultural and Statistical Sciences, 20(1), 299-307.

[26] Neelam, Sofia, Zaman, Q., Ullah, N., Ihsan, I., & Waseem, D. (2024). A development of new ratio type estimator using the auxiliary information, Power System Technology, 3(3), 2715-2724.

i