Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/090205

A Comparative Study of Ratio Estimators with Regression Estimator

Kuldeep Kumar Tiwari[1], *Sandeep Bhougal*[1,*] *and Sunil Kumar*[2]

[1]School of Mathematics, Shri Mata Vaishno Devi University, Katra 182320, Jammu and Kashmir, India
[2]Department of Statistics, University of Jammu, 180016, Jammu and Kashmir, India

Received: 2 Jan. 2022, Revised: 15 Feb. 2022, Accepted: 18 Feb. 2022
Published online: 1 May 2022

Abstract: Mean is one of the most important measures of central tendency. There is always an attempt to estimate the population mean precisely and accurately. The ratio and regression estimators are the most frequently used estimators for the population mean. It is well known that the regression estimator always performs better than the classical ratio estimator except when the regression line passes through the origin. If the regression line passes through the origin then both the classical ratio and regression estimators are equally efficient. There has been a continuous effort of improving the ratio estimators. The present paper seeks to explore whether the advanced ratio estimators can outweigh the regression estimator in any sense.

Keywords: Ratio estimator, regression estimator, mean squared error, efficiency

1 Introduction

In estimation, there is a lot of history. Bakker [1] tried to collect some historical facts where the sample mean were used to predict the population mean. As the statistics advances, the researchers focus to make estimates more accurate and precise. Supplementary information usually increases the efficiency of the estimate. For example, if we want to estimate the production of cereal then the area of cultivation can be used as auxiliary information. It appears that Cochran [2] first time uses the auxiliary variable to estimate the population mean of the study variable by defining the classical ratio estimator. The ratio estimator usually performs well when there is a positive correlation between the study and auxiliary variables.

Shrivastava [3] generalize the ratio estimator of Cochran [2]. Ratio method of estimation is further improved by Walsh [4], Ray and Sahai [5], Sisodia and Dwivedi [6], Bahl and Tuteja [7] and Upadhayaya and Singh [8]. Kadilar and Cingi [9] advances the ratio method of estimation by proposing a new class of estimators. Some other notable works on various kind of ratio method of estimation are Gupta and Shabbir [10], Yan and Tian [11], Subramani and Kumarapandiyan [12], Abid et al. [13], Singh, Vishwakarma and Gangele [14], Sing and Yadav [15], and Tiwari, Bhougal and Kumar [16]. Here we are going to compare some improved classes of ratio estimators with regression estimator to see their validity.

For y as study variable and x as an auxiliary variable, some notations that used throughout the article are:

- N, n: population and sample size.
- $\bar{Y}$, $\bar{X}$: population means.
- $\bar{y}$, $\bar{x}$: sample means.
- ρ: correlation coefficient between y and x.
- S_y^2, S_x^2, S_{yx} S_y, S_x: population variance, covariance and standard deviation of respective variables.
- s_y^2, s_x^2, s_{yx}: sample variance and covariance.
- $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$: coefficient of variation for y and x.
- $\beta_1(x)$: coefficients of skewness for x.
- $\beta_2(x)$: coefficients of kurtosis for x.

* Corresponding author e-mail: sandeep.bhougal@smvdu.ac.in

• $\lambda = \frac{1-f}{n}$, $f = \frac{n}{N}$, $R = \frac{\bar{Y}}{\bar{X}}$.

2 Materials and Methods

Cochran [2] proposed ratio estimator $t_r = \bar{y}(\frac{\bar{X}}{\bar{x}})$ to estimate the population mean $\bar{Y}$ when there is positive correlation between the study variable y and auxiliary variable x. The mean squared error (MSE) of ratio estimator t_r is

$$MSE(t_r) = \lambda(S_y^2 + R^2 S_x^2 - 2RS_{yx}) \tag{1}$$

Murthy [17] showed that the ratio estimator t_r performs better than usual estimator $t = \bar{y}$ whenever $\frac{1}{2}\frac{C_y}{C_x} < \rho \leq 1$.

The linear regression estimator given by Cochran [18] as $t_{reg} = \bar{y} + b_{yx}(\bar{X} - \bar{x})$; where $b_{yx} = \frac{s_{yx}}{s_x^2}$ is regression coefficient of y on x. The MSE of regression estimator t_{reg} is

$$MSE(t_{reg}) = \lambda S_y^2(1-\rho^2) \tag{2}$$

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of θ, then $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$.

So, if t_r is more efficient than t_{reg} then $MSE(t_r)$ should be less than $MSE(t_{reg})$. From equation (1) and (2) we can find that,

$$MSE(t_r) < MSE(t_{reg}) \text{ if } (RS_x - \rho S_y)^2 < 0 \tag{3}$$

which is not possible i.e. ratio estimator t_r can never be more efficient than regression estimator t_{reg}. It can also be concluded from equation (3) that

$$MSE(t_r) = MSE(t_{reg}) \text{ iff } (RS_x - \rho S_y)^2 = 0$$

That is, t_r is equally efficient to t_{reg} if and only if $R = \rho\frac{S_y}{S_x} = \frac{S_{yx}}{S_x^2}$.

Chakrabarty [19] and Ray & Sahai [5] improved the ratio estimator but their optimal MSEs are the same as regression estimator.

Sisodia & Dwivedi [6] modified the ratio estimator by using coefficient of variation of auxiliary variable by defining $t_{SD} = \bar{y}\left[\frac{\bar{X}+C_x}{\bar{x}+C_x}\right]$. Motivated by Sisodia & Dwivedi [6], Upadhyaya & Singh [8] proposed $t_{US1} = \bar{y}\left[\frac{C_x\bar{X}+\beta_2(x)}{C_x\bar{x}+\beta_2(x)}\right]$ & $t_{US2} = \bar{y}\left[\frac{\beta_2(x)\bar{X}+C_x}{\beta_2(x)\bar{x}+C_x}\right]$. Continuing the work, Singh [20] proposed $t_{S1} = \bar{y}\left[\frac{\bar{X}+S_x}{\bar{x}+S_x}\right]$, $t_{S2} = \bar{y}\left[\frac{\beta_1(x)\bar{X}+S_x}{\beta_1(x)\bar{x}+S_x}\right]$ & $t_{S3} = \bar{y}\left[\frac{\beta_2(x)\bar{X}+S_x}{\beta_2(x)\bar{x}+S_x}\right]$, Singh & Tailor [21] defines $t_{ST} = \bar{y}\left[\frac{\bar{X}+\rho}{\bar{x}+\rho}\right]$ and Singh et al. [22] as $t_{Sr} = \bar{y}\left[\frac{\bar{X}+\beta_1(x)}{\bar{x}+\beta_1(x)}\right]$. Although the estimator t_{SD}, t_{US1}, t_{US2}, t_{S1}, t_{S2}, t_{S3}, t_{ST} and t_{Sr} are more efficient than usual ratio estimator t_r in certain range but none of them matches with regression estimator t_{reg}.

Kadilar & Cingi [9] combined the concept of Sisodia & Dwivedi [6], Upadhyaya & Singh [8] and Singh et al. [22] with regression estimator to construct some new type of ratio estimator as

$$t_{KC_i} = [\bar{y} + b_{yx}(\bar{X} - \bar{x})]A_i;\ i = 1,2,3,4,5.$$

where $A_1 = \frac{\bar{X}}{\bar{x}}$, $A_2 = \frac{\bar{X}+C_x}{\bar{x}+C_x}$, $A_3 = \frac{\bar{X}+\beta_2(x)}{\bar{x}+\beta_2(x)}$, $A_4 = \frac{C_x\bar{X}+\beta_2(x)}{C_x\bar{x}+\beta_2(x)}$, $A_5 = \frac{\beta_2(x)\bar{X}+C_x}{\beta_2(x)\bar{x}+C_x}$.

The MSE of t_{KC_i} is

$$MSE(t_{KC_i}) = \lambda[R_i^2 S_x^2 + S_y^2(1-\rho^2)];\ i = 1,2,3,4,5.$$

where $R_1 = \frac{\bar{Y}}{\bar{X}}$, $R_2 = \frac{\bar{Y}}{\bar{X}+C_x}$, $R_3 = \frac{\bar{Y}}{\bar{X}+\beta_2(x)}$, $R_4 = \frac{C_x\bar{Y}}{C_x\bar{X}+\beta_2(x)}$, $R_5 = \frac{\beta_2(x)\bar{Y}}{\beta_2(x)\bar{X}+C_x}$.

Kadilar & Cingi [9] showed that t_{KC_i} perform better than all existing ratio estimators. Let us take the same population used by Kadilar & Cingi [9] to compare the performance of t_{KC_i} and regression estimator t_{reg}.

The parameters of population are $N = 106$, $n = 20$, $\bar{Y} = 2212.59$, $\bar{X} = 27421.70$, $S_y = 11551.53$, $S_x = 57460.61$, $\rho = 0.86$, $C_x = 2.10$, and $\beta_2(x) = 34.57$.

For this population, the calculated values of MSEs are $MSE(t_{KC_1}) = 2281556$, $MSE(t_{KC_2}) = 2281423$, $MSE(t_{KC_3}) = 2279362$, $MSE(t_{KC_4}) = 2280510$, $MSE(t_{KC_5}) = 2281552$, $MSE(t_r) = 2548180$ and $MSE(t_{reg}) = 1409557$.

It is observed that the MSE of t_{KC_i} are less than that of the usual ratio estimator t_r but the MSE of regression estimator t_{reg} is much less than MSEs of t_{KC_i}.

One can say that t_{KC_i} improves the ratio estimator but results conclude that t_{KC_i} actually impair the regression estimator as t_{KCi} uses regression estimator i.e. the term A_i in $t_{KCi} = [\bar{y} + b_{yx}(\bar{X} - \bar{x})]A_i$ resulted in increment to MSE of regression estimator.

Similarly, Kadilar & Cingi [23] proposed some other ratio estimators as

$$t_{KC2_i} = [\bar{y} + b_{yx}(\bar{X} - \bar{x})]B_i;\ i = 1,2,3,4,5.$$

where $B_1 = \frac{\bar{X}+\rho}{\bar{x}+\rho}$, $B_2 = \frac{C_x\bar{X}+\rho}{C_x\bar{x}+\rho}$, $B_3 = \frac{\rho\bar{X}+C_x}{\rho\bar{x}+C_x}$, $B_4 = \frac{\beta_2(x)\bar{X}+\rho}{\beta_2(x)\bar{x}+\rho}$, $B_5 = \frac{\rho\bar{X}+\beta_2(x)}{\rho\bar{x}+\beta_2(x)}$.

The MSE of t_{KC2_i} is

$$MSE(t_{KC2i}) = \lambda[P_i^2 S_x^2 + S_y^2(1-\rho^2)];\ i = 1,2,3,4,5.$$

where $P_1 = \frac{\bar{Y}}{\bar{X}+\rho}$, $P_2 = \frac{C_x\bar{Y}}{C_x\bar{X}+\rho}$, $P_3 = \frac{\rho\bar{Y}}{\rho\bar{X}+C_x}$, $P_4 = \frac{\beta_2(x)\bar{Y}}{\beta_2(x)\bar{X}+\rho}$, $P_5 = \frac{\rho\bar{Y}}{\rho\bar{X}+\beta_2(x)}$.

In numerical illustration t_{KC2_i} perform almost similar to t_{KC_i}. One can easily verify that for presented data in Kadilar & Cingi [23], the regression estimator t_{reg} is far better than t_{KC2_i}.

In a similar manner, many works have been done. Some of them are Yan & Tian [11], Subramani & Kumarapandiyan [12,24], Abid et al. [13], Abid et al. [25], Abid et al. [26] and Singh & Yadav [15].

3 Results and Discussion

We can generalize all the ratio estimators discussed and cited above in two classes as

$$t_{r_1} = \bar{y}\left[\frac{a\bar{X}+b}{a\bar{x}+b}\right] \tag{4}$$

$$t_{r_2} = [\bar{y} + b_{yx}(\bar{X} - \bar{x})]\left[\frac{a\bar{X}+b}{a\bar{x}+b}\right] \tag{5}$$

where $a(\neq 0)$ and b are either constant or any auxiliary attribute.

Theorem 1. *For any value of a and b, $MSE(t_{r_1}) \geq MSE(t_{reg})$.*

Proof. The MSE of t_{r_1} can be obtained on the lines of Upadhyaya & Singh [8] as

$$MSE(t_{r_1}) = \lambda(S_y^2 + R_{r_1}^2 S_x^2 - 2R_{r_1}S_{yx}) \tag{6}$$

where $R_{r_1} = \frac{a\bar{Y}}{a\bar{X}+b}$.

We can see that all the terms in $MSE(t_{r_1})$ are constant except R_{r_1} as R_{r_1} depends on a and b.

Differentiate $MSE(t_{r_1})$ with respect to R_{r_1} and equate to zero to get the optimum value of R_{r_1} as $R_{r_1}^o = \frac{S_{yx}}{S_x^2}$. Put $R_{r_1}^o$ in equation (6), we get the optimum MSE of t_{r_1}.

$$MSE_{opt}(t_{r_1}) = \lambda S_y^2(1-\rho^2) \tag{7}$$

which is equal to $MSE(t_{reg})$ as given in equation (2).

Again, assume $MSE(t_{r_1}) < MSE(t_{reg})$.

$$\text{i.e.}\qquad \lambda(S_y^2 + R_{r_1}^2 S_x^2 - 2R_{r_1}S_{yx}) < \lambda S_y^2(1-\rho^2) \tag{8}$$

Simplify equation (8), we get $(R_{r_1}S_x - \rho S_y)^2 < 0$, which is not possible. Hence,

$$MSE(t_{r_1}) \geq MSE(t_{reg}) \tag{9}$$

Theorem 2. *For any value of a and b, $MSE(t_{r_2}) \geq MSE(t_{reg})$.*

Proof. The MSE of t_{r_2} can be obtained on the lines of Kadilar & Cingi [9] as

$$MSE(t_{r_2}) = \lambda(S_y^2 + R_{r_2}^2 S_x^2 - 2R_{r_2}S_{yx}) \tag{10}$$

where $R_{r_2} = \frac{a\bar{Y}}{a\bar{X}+b}$.

As in Theorem 1, it can be shown that

$$MSE(t_{r_2}) \geq MSE(t_{reg}) \tag{11}$$

From Theorem 1 and Theorem 2, it is concluded that regression estimator is more efficient than ratio estimators of both classes. It is worth to mention that, ratio estimators of the classes t_{r_1} and t_{r_2} are biased estimators while regression estimator t_{reg} is unbiased. Hence, regression estimator is more accurate and precise than ratio estimators of classes t_{r_1} and t_{r_2}. Regression estimator t_{reg} is one of the simplest estimators of population mean. The results of this comparative study show that the linear regression estimator t_{reg} always works better than both classes of ratio estimators t_{r_1} and t_{r_2}.

Acknowledgement

The authors are grateful to the anonymous referees for a careful checking of the details and for helpful comments that improved this paper.

References

[1] A. Bakker, The early history of average values and implications for education, Journal of Statistics Education, **11(1)**, 1–15, 2003.
[2] W. G. Cochran, 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, 1940.
[3] S. K. Shrivastava, An estimator using auxiliary information in sample surveys, Calcutta Statistical Association Bulletin **16**, 121–132, 1967.
[4] J. Walsh, Generalization of ratio estimator for population total, Sankhya A, **32(1)**, 99–106, 1970.
[5] S. K. Ray and A. Sahai, Efficient families of ratio and product type estimators, Biometrika, **67(1)**, 211–215, 1980.
[6] B. V. S. Sisodia and B. K. Dwivedi, A modified ratio estimator using coefficient of variation of auxiliary variable, Journal of Indian Society Agricultural Statistics, **33**, 13–18, 1981.
[7] S. Bahl and R. Tuteja, Ratio and product-type exponential estimators, Journal of Information and Optimization Sciences, **12(1)**, 159–164, 1991.
[8] L. N. Upadhyaya and H. P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, Biometrical Journal, **41(5)**, 627–636, 1999.
[9] C. Kadilar and H. Cingi, Ratio estimators in simple random sampling, Applied Mathematics and Computation, **151(5)**, 893–902, 2004.
[10] S. Gupta and J. Shabbir, On improvement in estimating the population mean in simple random sampling, Journal of Applied Statistics, **35(5)**, 559–566, 2008.
[11] Z. Yan and B. Tian. Ratio method to the mean estimation using coefficient of skewness of auxiliary variable, in Information Computing and Applications. R. Zhu, Y. Zhang, B. Liu, and C. Liu, Ed. Springer, Berlin, Heidelberg: 103–110, 2010.
[12] J. Subramani and G. Kumarapandiyan, Estimation of population mean using known correlation coefficient and median, Journal of Statistical Theory and Applications, **13(4)**, 333–343, 2014.
[13] M. Abid, N. Abbas, H. Z. Nazir and Z. Lin, Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters, Revista Colombiana de Estadistica, **39(1)**, 63–79, 2016.
[14] A. Singh, G. K. Vishwakarma and R. K. Gangele, Improved predictive estimators for finite population mean using searls technique, Journal of Statistics and Management Systems, **22(8)**, 1555–1571, 2019.
[15] H. P. Singh and A. Yadav, A new exponential approach for reducing the mean squared errors of the estimators of population mean using conventional and non-conventional location parameters, Journal of Modern Applied Statistical Methods, **18(1)**, Article 26, 2020.
[16] K. K. Tiwari, S. Bhougal, and S. Kumar. A Difference-Cum-Exponential Type Efficient Estimator of Population Mean, in Mathematical Modeling and Computation of Real-Time Problems: An Interdisciplinary Approach, 1st ed., R. Kulshrestha, C. Shekhar, M. Jain, and S. R. Chakravarthy, Ed. CRC Press, Boca Raton, FL: 181–194, 2021.
[17] M. Murthy, Product method of estimation, Sankhya A, **26(1)**, 69–74, 1964.
[18] W. G. Cochran, Sampling techniques, Wiley, New York, 1977.
[19] R. P. Chakrabarty, Some ratio estimators, Journal of the Indian Society of Agricultural Statistics, **31(1)**, 49–57, 1979.
[20] G. N. Singh, On the improvement of product method of estimation in sample surveys, Journal of the Indian Society of Agricultural Statistics, **56(3)**, 267–275, 2003.
[21] H. P. Singh and R. Tailor, Use of known correlation coefficient in estimating the finite population mean, Statistics in Transition, **6**, 555–560, 2003.
[22] H. P. Singh, R. Tailor, R. Tailor, and M. S. Kakran, An improved estimator of population mean using power transformation, Journal of the Indian Society of Agricultural Statistics, **58**, 223–230, 2004.
[23] C. Kadilar and H. Cingi, A new ratio estimator using correlation coefficient, Interstat, **4**, 1–11, 2006.
[24] J. Subramani and G. Kumarapandiyan, Estimation of population mean using known median and coefficient of skweness, American Journal of Mathematics and Statistics, **2(5)**, 101–107, 2012.
[25] M. Abid, N. Abbas and M. Riaz, Improved modified ratio estimators of population mean based on deciles, Chiang Mai Journal of Science, **43(1)**, 1311–1323, 2016.
[26] M. Abid, N. Abbas, R. A. K. Sherwani and H. Z. Nazir, Improved ratio estimators for the population mean using non-conventional measure of dispersion, Pakistan Journal of Statistics and Operations Research, **12(2)**, 353–367, 2016.