Journal of Statistical Theory and Applications
Vol. 19(4), December 2020, pp. 534–539
DOI: https://doi.org/10.2991/jsta.d.201228.001; ISSN 1538-7887
https://www.atlantis-press.com/journals/jsta

An Unbiased Estimator of Finite Population Mean Using Auxiliary Information

B. Mahanty[1,*], G. Mishra[2]

[1]P.G. Department of Statistics, Utkal University, Bhubaneswar-751004, India
[2]P.G. Department of Statistics, Utkal University, Bhubaneswar-751004, India

ARTICLE INFO

Article History

Received 20 July 2020
Accepted 25 Dec 2020

Keywords

Simple random sampling
Auxiliary variable
Unbiased estimator
Optimality
Efficiency

ABSTRACT

In this paper, an unbiased estimator is constructed by using a linear combination of an estimator of study variable and mean per unit estimator of an auxiliary variable under simple random sampling without replacement scheme. The efficiency of the estimator under optimality compared with the mean per unit estimator, an almost unbiased ratio estimator, an unbiased product estimator, and a regression estimator both theoretically and with the numerical illustration.

1. INTRODUCTION

In the survey sampling method, it is a common practice to utilize auxiliary information, which is frequently acknowledged to the higher precision of the estimators of population parameters. The classical ratio estimator, product estimator, and regression estimator are good examples in this context. When the correlation between the study variable (y) and auxiliary variable (x) is highly positively correlated, the ratio method of estimation is quite effective. Similarly when there is an existence of a high negative correlation between y and x then the product method of estimation is effectively used.

The property of unbiasedness is one of the important properties of an estimator. So it is desirable to construct unbiased estimators simultaneously keeping in mind its efficiency.

Tin [1] suggested an almost unbiased ratio estimator estimate the population means. Tin showed that this estimator is an almost unbiased and more efficient than the conventional ratio estimator suggested by Cochran [2]. Robson [3] suggested an estimator when there exists a negative correlation between study variable and auxiliary variable and is known as product estimator. Further, Robson constructed an unbiased estimator to estimate the population means by subtracting estimated bias from the product estimator.

In this paper, the unbiased estimator is suggested and its efficiency is compared with the mean per unit estimator, with an almost unbiased ratio estimator, suggested by Tin [1], unbiased product estimator suggested by Robson [3] and regression estimator by Watson [4].

2. UNBIASED ESTIMATORS

Let's consider a finite population $U = (U_1, U_2, U_3, \ldots, U_N)$ size N. Let (y, x) be the study and auxiliary variables respectively. Now we consider (y_i, x_i), i = 1, 2, 3, … , n. denotes a sample of size "n" on the characteristics y and x have drawn from the population U with SRSWOR.

We denote $\overline{Y}$ and $\overline{X}$ are the population means of y and x respectively.

Let $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are the sample mean of y and x respectively.

*Corresponding author. Email: bulu.mahanty@gmail.com

Further, we denote $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2$, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$, and $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)$

Tin [1] suggested an almost unbiased ratio type estimator and it is given by

$$t_{Tin} = \overline{y}\frac{\overline{X}}{\overline{x}}\left[1+\theta\left(\frac{s_{yx}}{\overline{xy}}-\frac{s_x^2}{\overline{x}^2}\right)\right] \tag{1}$$

where, $\theta = \left(\frac{1}{n}-\frac{1}{N}\right)$

$$V\left(t_{Tin}\right) = \theta\overline{Y}^2\left[C_y^2+C_x^2-2C_{yx}\right] \text{ Considering up to } O\left(\frac{1}{n}\right). \tag{2}$$

Robson [3] suggested an unbiased product type estimator and it is given by

$$t_{Rob} = \frac{\overline{yx}}{\overline{X}} - \theta\frac{s_{yx}}{\overline{X}}. \tag{3}$$

$$V\left(t_{Rob}\right) = \theta\overline{Y}^2\left[C_y^2+C_x^2+2C_{yx}\right] \text{ Considering up to } O\left(\frac{1}{n}\right). \tag{4}$$

Further, the regression estimator which is biased but in most cases, it is more precise is given by

$$t_{lr} = \overline{y} + b\left(\overline{X}-\overline{x}\right) \tag{5}$$

$$V\left(t_{lr}\right) = \theta\overline{Y}^2 C_y^2\left(1-\rho^2\right) \tag{6}$$

where, $b = \frac{s_{yx}}{s_x^2}$, the regression coefficient of y on x.

2.1. Proposed Unbiased Estimator

Considering the linear combination of Robson's an unbiased estimator and the mean per unit estimator of auxiliary variable x, we construct an unbiased estimator, given by

$$t_{MU} = \lambda_0\left(\overline{y}\frac{\overline{x}}{\overline{X}} - \theta\frac{s_{xy}}{\overline{X}}\right) + \lambda_1\overline{x} \tag{7}$$

where, λ_0 and λ_1 are two suitable chosen constants such that estimator t_{MU} is unbiased.

Hence, we put an unbiased condition

$$E\left(t_{MU}\right) = E\left[\lambda_0 t_{Rob} + \lambda_1\overline{x}\right] = \overline{Y}$$

$$\Rightarrow \lambda_0 E\left(t_{Rob}\right) + \lambda_1 E(\overline{x}) = \overline{Y}$$

$$\Rightarrow \lambda_0\overline{Y} + \lambda_1\overline{X} = \overline{Y}$$

$$\Rightarrow \left(\lambda_0 - 1\right)\overline{Y} + \lambda_1\overline{X} = 0 \tag{8}$$

Now, we derive the variance of the proposed unbiased estimator by considering the condition. The value of λ_0 and λ_1, which makes the estimator unbiased.

3. VARIANCE OF THE PROPOSED ESTIMATOR t_{MU}

The variance of the estimator t_{MU} is given as

$$\begin{aligned}\mathrm{V}\left(t_{MU}\right) &= \left[\lambda_0^2 V\left(t_{Rob}\right) + \lambda_1^2 V(\overline{x}) + 2\lambda_0\lambda_1 Cov\left(\overline{x}, t_{Rob}\right)\right]\\ &= \lambda' S_2 \lambda \end{aligned} \tag{9}$$

where, $\lambda' = \begin{bmatrix} \lambda_0 & \lambda_1 \end{bmatrix}$, $S_2 = \begin{bmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{bmatrix}$, and $\lambda = \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix}$

Further, $S_{00} = V\left(\overline{y}_{Rob}\right)$, $S_{10} = S_{01} = Cov\left(\overline{x}, t_{Rob}\right)$, and $S_{11} = V(\overline{x})$

$S_{00} = V\left(t_{Rob}\right) = \theta \overline{Y}^2 \left[C_y^2 + C_x^2 + 2C_{yx}\right]$, Considering up to $O\left(\frac{1}{n}\right)$

$S_{10} = S_{01} = Cov\left(t_{Rob}, \overline{x}\right) = \theta \overline{Y}\overline{X}\left[C_{yx} + C_x^2\right]$, Considering up to $O\left(\frac{1}{n}\right)$

$S_{11} = V(\overline{x}) = \theta \overline{X}^2 C_x^2$, Considering up to $O\left(\frac{1}{n}\right)$

The value of λ_0 and λ_1 which minimize the $V(t_{MU})$ given in (9) and subject to the condition of unbiasedness of t_{MU} given in (7).

The optimum value of λ is obtained as

$$\lambda = \overline{Y}\frac{S_2^{-1}Q_2}{Q_2' S_2^{-1} Q_2} \tag{10}$$

where, S_2^{-1} is the inverse of the matrix of S_2, i.e.,

$$S_2^{-1} = \frac{1}{S_{00}S_{11} - S_{10}^2}\begin{bmatrix} S_{11} & -S_{10} \\ -S_{01} & S_{00} \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} \overline{Y} \\ \overline{X} \end{bmatrix}$$

where, Q_2' is the transpose of Q_2 i.e.,

$$Q_2' = \begin{bmatrix} \overline{Y} & \overline{X} \end{bmatrix}$$

Hence, after the expansion

$$\lambda = \frac{\overline{Y}\begin{bmatrix} \overline{Y}S_{11} - \overline{X}S_{01} \\ -\overline{Y}S_{10} + \overline{X}S_{00} \end{bmatrix}}{\overline{Y}^2 S_{11} - 2\overline{X}\overline{Y}S_{10} + \overline{X}^2 S_{00}} \tag{11}$$

From the Equation (11) we get the optimum values of λ_0 and λ_1

$$\begin{aligned}\lambda_0 &= \frac{\overline{Y}\left(\overline{Y}S_{11} - \overline{X}S_{10}\right)}{\overline{Y}^2 S_{11} - 2\overline{X}\overline{Y}S_{10} + \overline{X}^2 S_{00}}\\ &= \frac{R^2 V(\overline{x}) - RCov(\overline{x}, t_{Rob})}{V(t_{Rob}) + R^2 V(\overline{x}) - 2RCov(\overline{x}, t_{Rob})}\end{aligned} \tag{12}$$

$$\begin{aligned}\lambda_1 &= \frac{\overline{Y}\left(\overline{X}S_{00} - \overline{Y}S_{10}\right)}{\overline{Y}^2 S_{11} - 2\overline{X}\overline{Y}S_{10} + \overline{X}^2 S_{00}}\\ &= R.\frac{V\left(t_{Rob}\right) - RCov\left(\overline{x}, t_{Rob}\right)}{V\left(t_{Rob}\right) + R^2 V(\overline{x}) - 2RCov\left(\overline{x}, t_{Rob}\right)}\end{aligned} \tag{13}$$

where, $R = \frac{\overline{Y}}{\overline{X}}$

The optimum variance is obtained by substituting the optimum value of λ_0 and λ_1

$$V\left(t_{MU}\right)_{opt} = \frac{\overline{Y}^2}{Q' S_2^{-1} Q} = \frac{\overline{Y}^2 \left(S_{00} S_{11} - S_{10}^2\right)}{\overline{Y}^2 S_{11} - 2\overline{X}\overline{Y} S_{10} + \overline{X}^2 S_{00}} = \theta \overline{Y}^2 C_x^2 \left(1 - \rho^2\right) \tag{14}$$

4. COMPARISON OF EFFICIENCY

i. **Comparison of t_{MU} under optimality with mean per unit estimator $\bar{y}$.**
The variance of mean per unit estimator is

$$V(\bar{y}) = \theta \overline{Y}^2 C_y^2 \tag{15}$$

Comparing the variance of suggested unbiased estimator t_{MU} under optimum value from the Equation (14) with the variance of Mean per unit estimator from Equation (15) we have,

$$V(\bar{y}) - V\left(t_{MU}\right)_{opt} = \theta \overline{Y}^2 C_y^2 - \theta \overline{Y}^2 C_x^2 \left(1 - \rho^2\right) = \theta \overline{Y}^2 \left[C_y^2 - C_x^2 \left(1 - \rho^2\right)\right] \tag{16}$$

Hence, the proposed estimator t_{MU} under optimality more efficient than the mean per unit estimator if

$$C_y^2 > C_x^2 \left(1 - \rho^2\right). \tag{17}$$

ii. **Comparison of t_{MU} under optimality with an almost unbiased estimator due to Tin t_{Tin}.**
From the above Equations (2) and (14)

$$\text{MSE}\left(t_{Tin}\right) - \text{V}\left(t_{MU}\right) = \theta \overline{Y}^2 \left(C_y^2 + C_x^2 - 2\rho C_y C_x\right) - \theta \overline{Y}^2 C_x^2 \left(1 - \rho^2\right) = \theta \overline{Y}^2 \left(C_y^2 + \rho^2 C_x^2 - 2\rho C_y C_x\right) = \theta \overline{Y}^2 \left(C_y - \rho C_x\right)^2 > 0 \tag{18}$$

Since, $\left(C_y - \rho C_x\right)^2$ always positive the proposed estimator t_{MU} is more efficient than the Tin estimator t_{Tin}.

iii. **Comparison of t_{MU} under optimality with an unbiased product estimator due to Robson t_{Rob}.**
From the above Equations (4) and (14)

$$\text{MSE}\left(t_{Rob}\right) - \text{V}\left(t_{MU}\right) = \theta \overline{Y}^2 \left(C_y^2 + C_x^2 + 2\rho C_y C_x\right) - \theta \overline{Y}^2 C_x^2 \left(1 - \rho^2\right) = \theta \overline{Y}^2 \left(C_y^2 + \rho^2 C_x^2 + 2\rho C_y C_x\right) = \theta \overline{Y}^2 \left(C_y + \rho C_x\right)^2 > 0 \tag{19}$$

Since, $\left(C_y + \rho C_x\right)^2$ always positive the proposed estimator t_{MU} is more efficient than the unbiased product estimator t_{Rob}.

iv. **Comparison of t_{MU} under optimality with regression estimator t_{lr}**
From the above Equations (6) and (14)
Now,

$$\text{MSE}\left(t_{lr}\right) - \text{V}\left(t_{MU}\right) = \theta \overline{Y}^2 C_y^2 \left(1 - \rho^2\right) - \theta \overline{Y}^2 C_x^2 \left(1 - \rho^2\right) = \theta \overline{Y}^2 \left(1 - \rho^2\right) \left[C_y^2 - C_x^2\right] > 0 \tag{20}$$

when, C_y^2 is greater than, C_x^2, t_{MU} is more efficient than t_{lr}

5. NUMERICALLY ILLUSTRATIONS

To study the performance of the estimators, $\bar{y}$ (mean per unit estimator), t_{Tin}, t_{Rob}, t_{lr}, t_{MU} numerically, we consider seven natural populations with positive correlation and seven natural population with negative correlation described in the Tables 1 and 2. The comparison is based on the computation of the variance of different estimators.

Remarks: Table 3, which shows that the variance of the proposed estimator t_{MU} is the least when, C_y^2 is greater than C_x^2, followed by the regression estimator.

Remarks: Table 4, which shows that the variance of the estimator t_{MU} is the least when, C_y^2 is greater than C_x^2, followed by the regression estimator.

Table 1 Description of the population (Correlation coefficient is positive).

Pop.ⁿ No.	Sources	Popⁿ size (N)	Y	x	C_y	C_x	ρ
1	Daniel and Cross [5], p. 474	25	Paired Serum	Dry Blood Spot Specimens	0.5177	0.4358	0.95
2	Gujarati [6], p. 598	10	Income (thousands of dollars)	No. of Families Owning a House	0.3096	0.1811	0.36
3	Gupta [7], p. 451	11	Fathers Height (cm)	Sons Height (cm)	0.0014	0.0015	0.55
4	Armitage and Berry [8], p. 161	17	Birth Weight	Increase in Weight	0.6076	0.2104	0.64
5	Sukhatme and Sukhatme [9], p. 166	20	No. of Banana Bunches	No. of Banana Pits	0.0605	0.0431	0.75
6	Daniel and Cross [5], pp. 455–456	20	Age (years)	Bilirubin Levels (mg/dl)	0.4518	0.0483	0.46
7	Cochran [10], p. 186	21	Number of Family Members	Number of Cars	0.5231	0.1156	0.97

Table 2 Description of the population (Correlation coefficient is negative).

Pop.ⁿ No.	Sources	Popⁿ Size (N)	y	x	C_y	C_x	ρ
1	Singh and Mangat [11], p. 71	10	Number of Tube wells	Samples of Villages	0.7412	0.3981	−0.19
2	Singh and Mangat [11], p. 227	36	Samples Student OGPA	Number of hours per Week Devoted to TV Viewing	0.2852	0.0190	−0.58
3	Singh and Mangat [11], p. 195	37	Collected in respect of OGPA	Said nonacademic Activities	0.1648	0.0443	−0.36
4	Armitage and Berry [8], p. 161	32	Increase in Weight after 70–100 days	Birth Weight,	0.1069	0.0256	−0.68
5	Maddala [12], p. 316	16	Veal (Consumption per capita LB)	Veal (Price per pound)	0.0519	0.0097	−0.68
6	Maddala [12], p. 316	16	Lamb (Consumption per capita LB)	Lamb (Price per pound)	0.011	0.0105	−0.75
7	Gujarati [6], p. 598	10	Hypothetical data on Income 1000$	No. of Families at Income	0.344	0.1423	−0.43

Table 3 MSE of different estimators (sample size n = 4) (Correlation coefficient is positive).

Pop.ⁿ No.	$\bar{y}$	t_{Tin}	t_{Rob}	t_{lr}	t_{MU}
1	137.4757	13.5067	492.9409	13.3607	11.2491
2	18.954	19.5847	40.4995	16.4883	9.6457
3	0.9545	0.8289	2.9216	0.6572	0.6340
4	3.2559	1.9426	6.8246	1.9350	0.6701
5	9842.1	4470.5	4438.3	6232.4	3128.54
6	4.4792	3.5936	6.3234	3.5075	0.3753
7	0.3111	0.0949	0.6648	0.0158	0.0035

Table 4 MSE of different estimators (sample size n = 4) (Correlation coefficient is negative).

Pop.ⁿ No.	$\bar{y}$	t_{Tin}	t_{Rob}	t_{lr}	t_{MU}
1	2800.20	5087.90	3521.06	2698.20	1449.48
2	2.9360	4.0209	2.2425	1.9258	0.1283
3	8.0612	13.2546	7.2073	7.0078	1.8862
4	115.5169	221.0861	65.4413	60.9492	14.6398
5	0.568	0.3396	0.3412	0.7866	0.0568
6	10.657	5.4721	5.4356	8.3346	4.8318
7	18.954	37.359	16.2325	15.3967	6.3700

6. CONCLUSION

In this paper, an almost unbiased estimator t_{MU} perform better than the estimators mean per unit estimator ($\bar{y}$), Tin estimator (t_{Tin}), Robson estimator (t_{Rob}), and regression estimator (t_{lr}) both theoretically and numerically when the coefficient variation of y is greater than the coefficient variation of x.

CONFLICTS OF INTEREST

We the author(s) declare(s) that there is no conflict of interest.

AUTHORS' CONTRIBUTIONS

B. Mahanty developed the theoretical formalism, performed the analytic calculations and performed the numerical illustrations. Both B. Mahanty and G. Mishra. authors contributed to the final version of the manuscript. G.Mishra. supervised the manuscript.

Funding Statement

The author(s) received no financial support to carry out this research in the manuscript.

ACKNOWLEDGMENTS

The authors are indebted to two anonymous referees and the Editor for their productive comments and suggestions, which led to improve the presentation of this manuscript.

REFERENCES

1. M. Tin, J. Am. Stat. Assoc. 60 (1965), 294–307.
2. W.G. Cochran, J. Agric. Sci. 30 (1940), 262–275.
3. D.S. Robson, J. Am. Stat. Assoc. 52 (1957), 511–522.
4. D.J. Watson, J. Agric. Sci. 27 (1937), 474–483.
5. W.W. Daniel, L.C. Cross, Biostatistics: Basic Concepts and Methodology for the Health Sciences, International Student Version, tenth ed., Wiley India Pvt. Ltd., New Delhi, India, 2014.
6. D.N. Gujarati, Basics Econometrics, fourth ed., McGraw-Hill, New Delhi, India, 2003.
7. S.P. Gupta, Elementary Statistical Methods, Sultan and Chand & Sons, New Delhi, India, 2008.
8. P. Armitage, G. Berry, Statistical Methods in Medical Research, fourth ed., Blackwell Publishing Company, Macmillan Noida, India Ltd.
9. P.V. Sukhatme, B.V. Sukhatme, Sampling Theory of Surveys with Applications, first ed., Piyush Publications, Delhi, India, 1997.
10. W.G. Cochran, Sampling Techniques, third ed., John Wiley and Sons, Inc., New York, NY, USA, 1977.
11. R. Singh, N. Singh Mangat, Elements of Survey Sampling, vol. 15, B.V. Kluwer Academic Publishers, Punjab, India, 1996.
12. G.S. Maddala, Introduction to Econometrics, second ed., Macmillan Publishing Company, New York, NY, USA,1992.