Daffodil International University Journal of Business and Economics, Vol. 1, No. 1, July

# Second Order Biases and Mean Squared Errors of Some Estimators Using Single Auxiliary Variable

**M. I. Hossain***
**Masud Ibn Rahman****
**Muhammad Tareq*****

**Abstract :**The ratio, product, chain ratio, chain product, chain regression and ratio-cum product estimators have been considered by Chand (1975), Kiregyera (1980, 1984), Upadhyaya et. al. (1990, 1992), Srivastava et. al. (1990), Prasad et. al. (1992), Sahoo and Sahoo (1993), Singh (1993). Most of them discussed these estimators along with their first order bias and mean square error. In this paper, we have tried to find out the second order biases and mean square errors of these estimators based on simple random sampling. Finally, we have compared the performance of these estimators with some numerical illustration.s.

## 1. Introduction

Let $U = (U_1, U_2, ......, U_i, ......., U_N)$ denote a finite population of $N$ distinct and identifiable units. For estimating the population mean $\overline{Y}$ of a study variable $Y$, let us consider $X$ a an auxiliary variable that is correlated with study variable Y, taking the corresponding values of the units. Further let $Y_i$ be the unknown real variable of $Y$ and $X_i$ be the known variable value of $X$ associated with $U$. Let a sample of size $n$ be drawn from this population using simple random sampling without replacement (SRSWOR) and $y_i$ $(i = 1, 2, ..., n)$ are the values of the study variable and auxiliary variable respectively for the $i$-th units of the sample.

## 2. Some Estimators in Simple Random Sampling

For estimating the population mean of Y, the conventional ratio estimator of the population mean $\overline{Y}$ of Y is given by

$$t_{1s} = \bar{y}\frac{\overline{X}}{\bar{x}} \tag{1.1}$$

Where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ( the notation 's' is used to represent simple random

---

[*] Scientific Officer, Statistics Section, Agril. Eco. Div., BARI, Joydebpur, Gazipur, Dhaka.
[**] Assistant Professor, Faculty of Business and Economics, Daffodil International University, Dhaka.
[**] Lecturer, Department of Management, Islamic University, Khustia.

sampling).

The classical product type estimator is given by

$$t_{2s} = \bar{y}\frac{\bar{x}}{\bar{X}} \quad (1.2)$$

We have from the estimator of Srivastava (1967),

$$t_{3s} = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha} \quad (1.3)$$

Where $\alpha$ is a constant suitably chosen by minimizing MSE of $t_{3s}$. For $\alpha = 1$, $t_{3s}$ is the same as conventional ratio estimator whereas for $\alpha = -1$, it becomes conventional product type estimator.

Again for estimating the population mean of Y, we have from Walsh (1970)'s estimator as

$$t_{4s} = \bar{y}[\bar{X}\{\theta\bar{x} + (1-\theta\bar{X})^{-1}] \quad (1.4)$$

Where $\theta$ is the constant suitably chosen by minimizing mean square error of the estimator $t_{4s}$.

## 3. First Order Biases and Mean Squared Errors

The first order biases and mean squared error (MSE) of the estimators $t_{1s}$ and $t_{2s}$ are given respectively as

$$Bias(t_{1s}) = \bar{Y}\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_x^2 - \rho C_x C_y\right) \quad (2.1)$$

$$Bias(t_{2s}) = \bar{Y}\left(\frac{1}{n} - \frac{1}{N}\right)\rho C_x C_y \quad (2.2)$$

and $$MSE(t_{1s}) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_y^2 + C_x^2 - 2\rho C_x C_y\right)\right] \quad (2.3)$$

$$MSE(t_{2s}) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_y^2 + C_x^2 + 2\rho C_x C_y\right)\right] \quad (2.4)$$

Where $\rho$ is the correlation coefficient between Y and X , $C_y = \frac{S_y}{\bar{Y}}$ and $C_x = \frac{S_x}{\bar{X}}$ are the coefficient of variations of the study variable and the auxiliary variable respectively.
First order biases and mean squared error (MSE) of the estimator $t_{3s}$ are given respectively

$$Bias(t_{3s}) = \frac{\bar{Y}}{2}\left(\frac{1}{n} - \frac{1}{N}\right)\left(\alpha^2 C_x^2 - 2\alpha\rho C_x C_y\right) \text{ and} \tag{2.5}$$

$$MSE(t_{3s}) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_y^2 + \alpha^2 C_x^2 - 2\alpha\rho C_x C_y\right)\right] \tag{2.6}$$

By minimizing $MSE(t_{3s})$, the optimum value of $\alpha$ is obtained as $\alpha_0 = \rho\frac{C_y}{C_x}$ and the expression for the bias and MSE of $t_{3s}$, for the optimum value of $\alpha$ is given respectively by

$$Bias(t_{3s})_{opt} = \frac{\bar{Y}}{2}\left(\frac{1}{n} - \frac{1}{N}\right)\rho^2 C_y \tag{2.7}$$

$$MSE(t_{3s})_{opt} = \left[\left(\frac{1}{n} - \frac{1}{N}\right)S_y^2(1 - \rho^2)\right] \tag{2.8}$$

The expression for the bias and MSE of $t_{4s}$ to the first order of approximation are given respectively as

$$Bias(t_{4s}) = \bar{Y}\left(\frac{1}{n} - \frac{1}{N}\right)\left(\theta^2 C_x^2 - \rho\theta C_x C_y\right) \tag{2.9}$$

and

$$MSE(t_{4s}) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_y^2 + \theta^2 C_x^2 - 2\theta\rho C_x C_y\right)\right] \tag{2.10}$$

By minimizing $MSE(t_{4s})$, the optimum value of $\theta$ is obtained as $\theta_0 = \rho\frac{C_y}{C_x}$ and the expression for the bias and MSE of $t_{4s}$, for the optimum value of $\theta$ are given respectively by

$$Bias(t_{4s})_{opt} = 0 \text{ and} \tag{2.11}$$

$$MSE(t_{4s})_{opt} = \left[\left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 (1-\rho^2)\right] \tag{2.12}$$

We observed that for the optimum cases the biases of the estimators are the difference and the MSE of $t_{3s}$ and $t_{4s}$ are same. It is observed that the mean squared errors of the estimators $t_{3s}$ and $t_{4s}$ are always less than $t_{1s}$ and $t_{2s}$. But the estimators $t_{3s}$ and $t_{4s}$ have the same variance. To find the most efficient estimator among $t_{3s}$ and $t_{4s}$, we have tried to find their second order biases and mean squared errors.

**4. Second Order Biases and Mean Squared Errors**

Now we have discussed the bias and MSE of the given estimators up to the second order of approximation under simple random sampling without replacement (SRSWOR).

Let us define, $\delta_{\bar{y}} = \dfrac{\bar{y} - \bar{Y}}{\bar{Y}}$ and $\delta_{\bar{x}} = \dfrac{\bar{x} - \bar{X}}{\bar{X}}$ then $E(\delta_{\bar{x}}) = E(\delta_{\bar{y}}) = 0$

Before obtaining the bias and MSE of the estimator up to the second order approximation we first discuss the following lemmas.

<u>**Lemma 3.1**</u>

For SRSWOR at both the phases, we have

(i) $$V(\delta\bar{y}) = E\{(\delta\bar{y})^2\} = \frac{N-n}{N-1}\frac{1}{n}C_{02} = L_1 C_{02}$$

(ii) $$V(\delta\bar{x}) = E\{(\delta\bar{x})^2\} = \frac{N-n}{N-1}\frac{1}{n}C_{20} = L_1 C_{20}$$

(iii) $$COV(\delta\bar{x}\,\delta\bar{y}) = E(\delta\bar{x}\,\delta\bar{y}) = \frac{N-n}{N-1}\frac{1}{n}C_{11} = L_1 C_{11}$$

Where $L_1 = \dfrac{(N-n)}{(N-1)}\dfrac{1}{n}$

**Lemma 3.2**

(i) $$E(\delta\overline{x}^2\delta\overline{y}) = \frac{(N-n)(N-2n)}{(N-1)(N-2)}\frac{1}{n^2}C_{21} = L_2C_{21}$$

(ii) $$E(\delta\overline{x}^3) = \frac{(N-n)(N-2n)}{(N-1)(N-2)}\frac{1}{n^2}C_{30} = L_2C_{30}$$

Where $L_2 = \frac{(N-n)(N-2n)}{(N-1)(N-2)}\frac{1}{n^2}$

**Lemma 3.3**

(i) $$E(\delta\overline{x}^3\delta\overline{y}) = L_3C_{31} + 3L_4C_{20}C_{11}$$

(ii) $$E(\delta\overline{x}^4) = L_3C_{40} + 3L_4C_{20}^2$$

(iii) $$E(\delta\overline{x}^2\delta\overline{y}^2) = L_3C_{22} + L_4(C_{20}C_{02}C_{11}^2)$$

Where $L_3 = \frac{(N-n)(N^2+N-6nN+6n^2)}{(N-1)(N-2)(N-3)}\frac{1}{n^3}$

and $L_4 = \frac{N(N-n)(N-n-1)(n-1)}{(N-1)(N-2)(N-3)}\frac{1}{n^3}$

Proof of these lemma are straight forward by using SRSWOR.
We re-write $t_{1s}$ as

$$t_{1s} = \overline{y}\frac{\overline{X}}{\overline{x}} = \overline{Y}(1+\delta\overline{y})(1+\delta\overline{x})^{-1}$$

$$\Rightarrow (t_{1s} - \overline{Y}) = \overline{Y}\left[\delta\overline{y} - \delta\overline{x} - \delta\overline{x}\delta\overline{y} + \delta\overline{x}^2 + \delta\overline{x}^2\delta\overline{y} + \delta\overline{x}^3\delta\overline{y} - \delta\overline{x}^3 + \delta\overline{x}^4 + ......\right]$$

Taking expectation up to second order approximation, we have

$$E(t_{1d} - \overline{Y}) = \overline{Y}\left[E(\delta\overline{x}^2) - E(\delta\overline{x}\delta\overline{y}) + E(\delta\overline{x}^2\delta\overline{y}) + E(\delta\overline{x}^3\delta\overline{y}) - E(\delta\overline{x}^3) + E(\delta\overline{x}^4)\right]$$

Using the lemmas of (4.1), (4.2) and (4.3) we get the following expression for the bias of $t_{1s}$ up to the second order of approximation.

$$Bias_2(t_{1s}) = \bar{Y}\left[L_1C_{20} - L_1C_{11} + L_2C_{21} - L_2C_{30} - 3L_4C_{20}C_{11} - L_3C_{31} + L_3C_{40} + 3L_4C_{20}^2\right]$$
$$= \bar{Y}\left[L_1(C_{20} - C_{11}) - L_2(C_{21} - C_{30}) + L_3(C_{40} - C_{31}) + 3L_4(C_{20}^2 - C_{20}C_{11})\right] \quad (3.1)$$

Now we obtain MSE of $t_{1s}$ and have from squaring of (4.1) and taking expectation and using lemmas we get

$$E(t_{1s} - \bar{Y})^2 = \bar{Y}^2E\left[\delta\bar{y}^2 + \delta\bar{x}^2 - 2\delta\bar{x}\delta\bar{y} + 4\delta\bar{x}^2\delta\bar{y} - 2\delta\bar{x}\delta\bar{y}^2 - 6\delta\bar{x}^3\delta\bar{y} + 3\delta\bar{x}^2\delta\bar{y}^2 - 2\delta\bar{x}^3 + 3\delta\bar{x}^4 + .....\right]$$

Using the lemmas of (4.1), (4.2) and (4.3) we get the following expression for the bias of $t_{1s}$ up to the second order of approximation.

$$MSE_2(t_{1s}) = \bar{Y}^2\begin{bmatrix} L_1C_{02} + L_1C_{20} - 2L_1C_{21} + 4L_2C_{21} - 2L_2C_{12} - 2L_2C_{30} - 6L_3C_{31} \\ + 3L_3C_{22} + 3L_4C_{20}C_{02} + 3L_4 2C_{11}^2 + 3L_3C_{40} + 9L_4C_{40} \end{bmatrix}$$

$$= \bar{Y}^2\begin{bmatrix} L_1(C_{02} + C_{20} - C_{11}) + 2L_2(2C_{21} - C_{12} - C_{30}) + \\ + 3L_3(C_{40} + C_{22} - 2C_{31}) + 3L_4(3C_{40} + C_{20}C_{02} + 2C_{11}^2 \end{bmatrix} \quad (3.2)$$

Again re-write for estimator $t_{2s}$

$$t_{2s} = \bar{y}\frac{\bar{x}}{\bar{X}} = \bar{Y}(1 + \delta\bar{y})(1 + \delta\bar{x})$$

$$\Rightarrow (t_{2s} - \bar{Y}) = \bar{y}\frac{\bar{x}}{\bar{X}} = (\delta\bar{y} + \delta\bar{x} + \delta\bar{x}\delta\bar{y})$$

Taking expectation up to second order approximation, we have

$$E(t_{2s} - \bar{Y}) = E(\delta\bar{y} + \delta\bar{x} + \delta\bar{x}\delta\bar{y})$$

Hence, the second order bias and mean squared error up to second order approximation are given by

$$Bias_2(t_{2s}) = \bar{Y}\,L_1C_{11} \quad (3.3)$$

$$MSE_2(t_{2s}) = \bar{Y}^2\left[L_1C_{02} + L_1C_{20} + L_3C_{22} + 2L_1C_{11} + 2L_2C_{21} + 2L_2C_{12} + L_4C_{20}C_{02} + L_4 2C_{11}^2)\right]$$
$$= \bar{Y}^2\left[L_1(C_{02} + C_{20} + C_{11}) + 2L_2(2C_{21} + C_{12} + L_3C_{22} + L_4(C_{20}C_{02} + 2C_{11}^2)\right] \quad (3.4)$$

For estimator $t_{3s}$

$$t_{3s} = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha} = \bar{Y}(1+\delta\bar{y})(1+\delta\bar{x})^{-\alpha}$$

The second order bias and MSE of $t_{3s}$

$$Bias_2(t_{3s}) = \frac{\bar{Y}}{2}\left[\begin{array}{l} L_1\alpha^2 C_{20} - 2\alpha L_1 C_{11} + \alpha^2 L_2 C_{21} + \dfrac{\alpha^3}{3} L_2 C_{30} - \alpha^3 L_4 C_{20} C_{11} \\ + \dfrac{\alpha^4}{12} L_3 C_{40} + L_3 C_{40} + \dfrac{\alpha^4}{4} L_4 C_{20}^2 \end{array}\right]$$

$$= \frac{\bar{Y}}{2}\left[\begin{array}{l} L_1(\alpha^2 C_{20} - 2\alpha C_{11}) + L_2(\alpha^2 C_{21} - \dfrac{\alpha^3}{3} C_{30}) - L_3(\dfrac{\alpha^4}{12} C_{40} - \dfrac{\alpha^3}{3} C_{31}) \\ + 3L_4(\dfrac{\alpha^4}{12} C_{20}^2 - \dfrac{\alpha^3}{3} C_{20} C_{11}) \end{array}\right] \quad (3.5)$$

and

$$MSE_2(t_{3d}) = \bar{Y}^2\left[\begin{array}{l} L_1(C_{02} + \alpha^2 C_{20} - 2\alpha C_{11}) + L_2(3\alpha C_{21} - 2\alpha C_{12} - \alpha^3 C_{30}) + L_3(\dfrac{7\alpha^4}{4} C_{40} + 2\alpha^2 C_{22} - \dfrac{7\alpha^3}{3} C_{31}) \\ + L_4(\dfrac{21\alpha^4}{4} C_{20}^2 - \dfrac{21\alpha^3}{3} C_{20} C_{11} + 2\alpha^2 C_{20} C_{02} + 4\alpha^2 C_{11}^2) \end{array}\right]$$

… (3.6)

The optimum value of $\alpha$ is obtained by minimizing $MSE_2(t_{3s})$. Theoretically the determination for the optimum value of $\alpha$ is very critical, the solution for the determination of this optimum value is obtained by numerical techniques.

Similarly for the estimator $t_{4s}$

$$t_{4s} = \bar{y}[\bar{X}\{\theta\bar{x} + (1-\theta\bar{X}\}^{-1}]$$

The bias and MSE of second order of this estimator are given by

$$Bias_2(t_{4s}) = \frac{\bar{Y}}{2}\left[\begin{array}{l} L_1(\theta^2 C_{20} - \theta C_{11}) + L_2(\theta^2 C_{21} - \theta^3 C_{30}) + L_3(\theta^4 C_{40} - \theta^3 C_{31}) \\ + 3L_4(\theta^4 C_{20}^2 - \theta^3 C_{20} C_{11}) \end{array}\right]$$

(3.7)

$$MSE_2(t_{4s}) = \bar{Y}^2 \begin{bmatrix} L_1(C_{02} - 2\theta C_{11}) + 2L_2(2\theta C_{21} - \theta C_{12} - \theta^3 C_{30}) + 3L_3(\theta^4 C_{40} - \theta^2 C_{22} - 2\theta^3 C_{31}) \\ + 3L_4(3\theta^4 C_{20}^2 + \theta^2 C_{20}C_{11} - 6\theta^3 C_{20}C_{11} + 2\theta^2 C_{11}^2) \end{bmatrix}$$

(3.8)

The optimum value of $\theta$ is obtained by minimizing $MSE_2(t_{4s})$. Theoretically the determination for the optimum value of $\theta$ is very critical, the solution for the determination of this optimum value is obtained by numerical techniques.

## 5. Numerical Illustration

For the two natural population data, we shall calculate the bias and the mean square error of the estimator and compare Bias and MSE for the first and second order of approximation.

**Data Set-1**

The data for the empirical analysis are taken from 1981, Utter Pradesh District Census Handbook, Aligar. The population consits of 340 villages under koil police station, with Y=Number of agricultural labour in 1981 and X=Area of the villages (in acre) in 1981. The following values are obtained

$\bar{Y} = 73.76765$, $\bar{X} = 2419.04$, $N = 340, n' = 120, n = 70$, $C_{02} = 0.7614,$

$C_{20} = 0.5557,\ C_{11} = 0.2667,\ C_{03} = 2.6942, C_{12} = 0.0747, C_{21} = 0.1589,$
$C_{30} = 0.7877, C_{13} = 0.1321,$

$C_{31} = 0.8851, C_{04} = 17.4275,\ C_{22} = 0.8424, C_{40} = 1.3051$

**Data Set-2**

The data for the empirical analysis are taken from 1981, Utter Pradesh District Census Handbook, Aligar. The population consists of 340 villages under koil police station, with Y=Number of cultivators in the villages in 1981 and X=Area of the villages (in acre) in 1981. The following values are obtained

$\overline{Y} = 141.1294$ , $\overline{X} = 2419.04$ , $N = 340, n' = 120, n = 70$ , $C_{02} = 0.7614,$

$C_{20} = 0.5944,\ C_{11} = 0.2667,\ C_{03} = 2.6942, C_{12} = 0.4720, C_{21} = 0.4897,$
$C_{30} = 0.7877, C_{13} = 1.3923,$

$C_{31} = 1.5586, C_{04} = 17.4275,\ C_{22} = 1.2681, C_{40} = 2.8457$

**Table-4.1 The values of the bias and MSE of the estimators $t_{1s}, t_{2s}, t_{3s}$ and $t_{4s}$ for the data set-1.**

| Estimator | Bias | | MSE | |
|---|---|---|---|---|
| | First order | Second order | First order | Second order |
| $t_{1s}$ | 0.15092 | 0.15072 | 30.19263 | 30.50956 |
| $t_{2s}$ | .013928 | 0.13517 | 71.28859 | 72.53269 |
| $t_{3s}$ | -0.03342 | -0.03344 | 24.40271 | 24.44216 |
| $t_{4s}$ | 0.00000 | -0.00011 | 24.40271 | 24.43308 |

**Table-4.2 The values of the bias and MSE of the estimators $t_{1s}, t_{2s}, t_{3s}$ and $t_{4s}$ for the data set-2.**

| Estimator | Bias | | MSE | |
|---|---|---|---|---|
| | First order | Second order | First order | Second order |
| $t_{1s}$ | 0.21692 | 0.21671 | 84.77982 | 85.69287 |
| $t_{2s}$ | 0.37697 | 0.37789 | 97.58332 | 98.65323 |
| $t_{3s}$ | -0.11964 | -0.11953 | 73.59757 | 74.04931 |
| $t_{4s}$ | 0.00000 | -0.00011 | 73.59757 | 73.81961 |

## 6. Conclusion

Table-4.1 and Table 4.2 present the first order approximation of the estimators $t_{1s,} t_{2s}, t_{3s}$ and $t_{4s}$ for two sets of data. The estimators $t_{2s}$ is a product estimator and it is considered in case of negative correlation. So the bias and mean squared errors are more. For the classical ratio estimator $t_{1s}$, it is observed that the biases and the mean squared errors are increased for second order. The estimator $t_{3s}$ and $t_{4s}$ have the same mean squared error for the first order but the mean squared error of $t_{4s}$ is less than $t_{3s}$ for the second order. So the second order mean squared error of $t_{3s}$ makes a difference from $t_{4s}$ for both data sets. Finally it is clear that the estimator $t_{4s}$ is better than other estimators.

## References


1. Chand, L. (1975). Some ratio-type estimators based on two or more auxiliary variables. Ph.D. Thesis submitted to Iowa State University, *Ames, Iowa.*
2. Kiregyera, B. (1980). A chain ratio-type estimator in finite population under double sampling using two auxiliary variables. Metrika, 27, 217-223.
3. Kiregyera, B. (1984). Regression-type estimators using two-variables and the model of double sampling from finite populations. Metrica, 31, 215-226.
4. Prasad, B., Singh, R.S. and Singh, H. P. (1992). Modified chain ratio estimator for finite population mean using two auxilary variables in double sampling. Statistical Series, 249,1-11.

   Sahoo, J. and Sahoo, L. N. (1993). A class of estimator in two phase sampling using two auxiliary variables. Jour. Ind. Stat. Assoc., 31, 107-114.
5. Singh, H. P. (1993). A chain ratio-cum-difference estimator using two auxiliary varieties in double sampling. Jour. Ravi. Univ., B, 6, 79-83.
6. Srivastava, S.K. (1967. An estimator using auxiliary information in sample surveys. Cal. Stat. Ass. Bull. 15:127-134.
7. Srivastava, Rani. S., Khare, B. B. and Srivastava, S. K., (1990). A generalized chain ratio estimator for mean of finite population. Jour. Ind. Soc. Ag. Stat. 42, 108-117.
8. Upadhyay, L.N., Kushwaha, K.S. and Singh, H.P. (1990). A modified chain-type estimator in two-phase sampling using multi-auxiliary information. *Metron*, **XLVIII**, 381-393.
9. Upadhyay, L.N., Dubey, S.P. and Singh, H.P. (1992). A class of ratio-in-regression estimators using two auxiliary variables in double sampling. *Jour. Sci. Res.*, **42**, 127-134.
10. Walsh, J.B. (1970). Generalization of ratio estimator for population total. Sankhya. Ser. A. 32: 99-106.