A Generalized Two Phase Sampling Estimator of Ratio of Population Means Using Auxiliary Information

Tarunpreet Kaur Ahuja[1], Peeyush Misra[1,*] and O. K. Belwal[2]

[1]*Department of Statistics, D.A.V.(P.G.) College, Dehradun-248001, Uttarakhand, India*
[2]*Department of Statistics, HNB Garhwal University, Srinagar-246174, Garhwal, Uttarakhand, India*
E-mail: tarunpreetkaur23@gmail.com; dr.pmisra.dav@gmail.com; okbelwal@rediffmail.com
[*]*Corresponding Author*

Received 02 July 2020; Accepted 20 December 2020;
Publication 08 March 2021

Abstract

This paper addresses the problem of estimating ratio of two population means by using quantitative auxiliary knowledge in the form of first and second moments. Through this paper, an improved generalized two phase sampling estimator has been proposed. The relative bias and mean squared error of the suggested estimator has been derived and studied. Also, a comparative study with the conventional estimators has been included to establish its superiority. Besides theoretical comparisons, a subset of optimum estimators having the same minimum mean squared error (MSE) is also explored. An empirical study is also carried out to support theoretical results.

Keywords: Auxiliary Character, Two-phase Sampling, Taylor's Series, Bias, Mean Squared Error and Efficiency.

Journal of Reliability and Statistical Studies, Vol. 14, Issue 1 (2021), 1–16.
doi: 10.13052/jrss0974-8024.1411

1 Introduction

Sampling theory deals with optimum combination of sampling and estimation procedures so that inferences about the population parameters are made with minimum error. The challenging issue of estimating ratio of two population means assumes great importance in the enormous literature of sampling theory. Since past sixty decades, numerous authors have highlighted the concept of incorporating auxiliary information at estimation stage resulting in enhancement of the efficiency and precision of estimators. In almost every field of scientific study like agriculture, forestry, economics surveys, management, biomedical sciences and the likes, estimation of population ratio assumes significant importance. Input-output ratio in an industrial survey, outlay on employees to the entire expenses, proportion of liquid to total assets, profitability rate, crop production rate, literacy rate are some illustrations of estimating ratio of two population parameters. Many a times, a health analyst may be interested in estimating growth index by measuring the ratio of weight to height using chest and skull circumference as auxiliary variables. Many elite survey statisticians have made meritorious efforts to estimate population ratio. For greater knowledge one may see Murthy (1967), Cochran (1977), Sukhatme et al. (1984), Singh and Chaudhary (1997) and Mukhopadhyay (2012).

A wide literature depicts the contribution of several authors who addressed this problem and estimated population ratio by using supplementary knowledge in whatsoever form available. To acquire knowledge on the various historical developments in this context, distinguished works of Singh (1965, 1967, 1969), Shah and Shah (1978), Tripathi (1980), Singh (1982), Singh (1998), Biradar and Singh (1997–98), Upadhyaya et al. (2000), Singh and Rani (2005, 2006), Singh and Naqvi (2015) and Kumar and Srivastava (2018) can be revisited. Although, ample of similar estimators by renowned statisticians are registered in sampling literature, yet there always remains potential and possibilities for improvements. An earnest effort in this regard is made in the subsequent sections of this manuscript.

2 Proposed Estimator

Subsidiary information on one or more auxiliary variable may be known beforehand through census reports, pilot survey, and historical data. Many a times a sampler may encounter a practical situation wherein parametric information associated with the auxiliary variables is not known apriori. This subsidiary information may sometimes be completely or partially lacking. To overcome such difficult situations, Neyman (1938) proposed double or two phase sampling technique. This sampling technique is highly recommended as it happens to be more flexible, robust and considerably cost effective (economical) procedure to develop reliable estimates of unknown population characteristics.

Let (y_1, y_2) be the variables under the reference of study highly correlated with the auxiliary variable x. Using SRSWOR design in either phases, a double sampling or two-phase sampling technique is described as:

(i) At the first phase, we select a preliminary large sample $(x'_1, x'_2, \ldots, x'_{n'})$ whose size is n/ from a population U having N distinct units. The first phase sample is taken on only ancillary variable$X$ and its sample mean is denoted by$\bar{x}'$.

(ii) At the second phase, we select a small sub sample $\{(y_{11}, y_{21}, x_1), (y_{12}, y_{22}, x_2), \ldots, (y_{1n}, y_{2n}, x_n)\}$ of size n from the large first phase sample. The second phase sample is observed on both the study variable Y_1, Y_2 and the auxiliary variable X and their respective means is represented by $\bar{y}_1, \bar{y}_2$ and $\bar{x}$.

Let us denote

$$\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x'_i, \quad \bar{y}_1 = \frac{1}{n}\sum_{i=1}^{n} y_{1i}, \quad \bar{y}_2 = \frac{1}{n}\sum_{i=1}^{n} y_{2i}, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Let population parameters $\overline{Y}_1, \overline{Y}_2$ and $\bar{X}$ denotes population mean and $S^2_{Y_1}, S^2_{Y_2}$ and S^2_X denotes population variance of main variable under study and correlated ancillary character.

We have

$$\overline{Y}_1 = \frac{1}{N}\sum_{i=1}^{N} Y_{1i}, \quad \overline{Y}_2 = \frac{1}{N}\sum_{i=1}^{N} Y_{2i}, \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

$$S^2_{Y_1} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{1i} - \overline{Y_1})^2, \quad S^2_{Y_2} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{2i} - \overline{Y_2})^2,$$

$$S^2_X = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

$$S_{y_1x} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{1i} - \bar{Y}_1)(X_i - \bar{X}),$$

$$S_{y_2x} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{2i} - \bar{Y}_2)(X_i - \bar{X})$$

$$S_{y_1y_2} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2) \quad \text{and}$$

$$\mu_{rst} = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - \bar{Y}_1)^r(Y_{2i} - \bar{Y}_2)^s(X_i - \bar{X})^t; \quad r, s, t = 0, 1, 2, 3, 4.$$

$$\sigma^2_{Y_1} = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - \overline{Y_1})^2, \quad \sigma^2_{Y_2} = \frac{1}{N}\sum_{i=1}^{N}(Y_{2i} - \overline{Y_2})^2,$$

$$\sigma^2_X = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2, \quad \rho = \frac{S_{Y_1Y_2}}{S_{Y_1}S_{Y_2}}, \quad \rho_1 = \frac{S_{Y_1X}}{S_{Y_1}S_X},$$

$$\rho_2 = \frac{S_{Y_2X}}{S_{Y_2}S_X}, \quad C^2_{Y_1} = \frac{S^2_{Y_1}}{\bar{Y}^2_1}, \quad C^2_{Y_2} = \frac{S^2_{Y_2}}{\bar{Y}^2_2}, \quad C^2_X = \frac{S^2_X}{\bar{X}^2}.$$

It is indeed essential to emphasize that auxiliary information in terms of moments about zero, that is $\bar{x}, \bar{x}'$, $\bar{\theta}_x$ and $\bar{\theta}'_x$ has been utilized to define a generalized class of double sampling estimator represented as $\hat{R}_g$ for efficient estimation of population ratio as

$$\hat{R}_g = g\left(\bar{y}_1, \bar{y}_2, \frac{\bar{x}}{\bar{x}'}, \frac{\bar{\theta}_x}{\bar{\theta}'_x}\right) = g(\bar{y}_1, \bar{y}_2, u_1, u_2) \tag{1}$$

where $R = \frac{\bar{Y}_1}{\bar{Y}_2}$, $\hat{R} = \frac{\bar{y}_1}{\bar{y}_2}$, $\frac{\bar{x}}{\bar{x}'} = u_1$, $\frac{\bar{\theta}_x}{\bar{\theta}'_x} = u_2$, $\bar{\theta}_x = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, $\bar{\theta}'_x = \frac{1}{n'}\sum_{i=1}^{n} x_i'^2$ and $g(\bar{y}_1, \bar{y}_2, u_1, u_2)$ satisfies the validity conditions of Taylor's series expansion is a bounded function of $t = (\bar{Y}_1, \bar{Y}_2, u_1, u_2)$ such that

(i) At the point $T = (\bar{Y}_1, \bar{Y}_2, 1, 1)$ we have

$$g(t = T) = R = \frac{\bar{Y}_1}{\bar{Y}_2} \tag{2}$$

(ii) The first order partial derivatives are

$$g_1 = \left(\frac{\partial g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_1}\right)_T = \frac{1}{\bar{Y}_2},$$
$$g_2 = \left(\frac{\partial g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_2}\right)_T = -\frac{\bar{Y}_1}{\bar{Y}_2^2}$$
$$g_3 = \left(\frac{\partial g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial u_1}\right)_T \quad \text{and} \quad g_4 = \left(\frac{\partial g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial u_2}\right)_T \tag{3}$$

(iii) Also, the second order partial derivatives are

$$g_{11} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_1^2}\right)_T = 0, \quad g_{22} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_2^2}\right)_T$$
$$= \frac{2\bar{Y}_1}{\bar{Y}_2^3}, \quad g_{12} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_1 \partial \bar{y}_2}\right)_T = -\frac{1}{\bar{Y}_2^2},$$
$$g_{33} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial u_1^2}\right)_T, \quad g_{44} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial u_2^2}\right)_T,$$
$$g_{13} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_1 \partial u_1}\right)_T, \quad g_{14} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_1 \partial u_2}\right)_T,$$
$$g_{23} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_2 \partial u_1}\right)_T, \quad g_{24} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial \bar{y}_2 \partial u_2}\right)_T,$$
$$g_{34} = \left(\frac{\partial^2 g(\bar{y}_1, \bar{y}_2, u_1, u_2)}{\partial u_1 \partial u_2}\right)_T. \tag{4}$$

3 The Expression for Bias and Mean Squared Error

For analyzing distinct properties relating to suggested estimator, we define

$$\bar{y}_1 = \overline{Y_1}(1 + e_1) \quad \bar{y}_2 = \overline{Y_2}(1 + e_2)$$
$$\bar{x} = \overline{X}(1 + e_3) \quad \bar{x}' = \overline{X}(1 + e_3')$$
$$\bar{\theta}_x = \bar{\theta}_X(1 + e_4) \quad \bar{\theta}_x' = \bar{\theta}_X(1 + e_4') \tag{5}$$

Since population under consideration is large enough relative to sample, for simplicity we ignore finite population correction terms. So that

$$E(e_i) = 0 \quad (i = 1, 2, 3, 4) \quad \text{and} \quad E(e'_j) = 0 \quad (j = 3, 4) \tag{6}$$

$$E(e_1^2) = \frac{1}{n} C_{Y_1}^2 \quad E(e_2^2) = \frac{1}{n} C_{Y_2}^2$$

$$E(e_3^2) = \frac{1}{n} C_X^2 \quad E(e_3'^2) = \frac{1}{n'} C_X^2$$

$$E(e_4^2) = \frac{1}{n\bar{\theta}_X^2}(\mu_{004} + 4\bar{X}\mu_{003} + 4\bar{X}^2\mu_{002} - \mu_{002}^2)$$

$$E(e_4'^2) = \frac{1}{n'\bar{\theta}_X^2}(\mu_{004} + 4\bar{X}\mu_{003} + 4\bar{X}^2\mu_{002} - \mu_{002}^2)$$

$$E(e_1 e_2) = \frac{1}{n}\rho C_{Y_1} C_{Y_2} = \frac{1}{n\bar{Y}_1\bar{Y}_2}\mu_{110}$$

$$E(e_1 e_3) = \frac{1}{n}\rho_1 C_{Y_1} C_X = \frac{1}{n\bar{Y}_1\bar{X}}\mu_{101}$$

$$E(e_1 e'_3) = \frac{1}{n'}\rho_1 C_{Y_1} C_X, \quad E(e_2 e_3) = \frac{1}{n}\rho_2 C_{Y_2} C_X = \frac{1}{n\bar{Y}_2\bar{X}}\mu_{011}$$

$$E(e_2 e'_3) = \frac{1}{n'}\rho_2 C_{Y_2} C_X, \quad E(e_1 e_4) = \frac{1}{n\bar{Y}_1\bar{\theta}_X}(\mu_{102} + 2\bar{X}\mu_{101})$$

$$E(e_1 e'_4) = \frac{1}{n'\bar{Y}_1\bar{\theta}_X}(\mu_{102} + 2\bar{X}\mu_{101}),$$

$$E(e_2 e_4) = \frac{1}{n\bar{Y}_2\bar{\theta}_X}(\mu_{012} + 2\bar{X}\mu_{011})$$

$$E(e_2 e'_4) = \frac{1}{n'\bar{Y}_2\bar{\theta}_X}(\mu_{012} + 2\bar{X}\mu_{011}), \quad E(e_3 e'_3) = \frac{1}{n'} C_X^2 = \frac{1}{n'\bar{X}^2}\mu_{002}$$

$$E(e_3 e_4) = \frac{1}{n\bar{X}\bar{\theta}_X}(\mu_{003} + 2\bar{X}\mu_{002}),$$

$$E(e_3 e'_4) = \frac{1}{n'\bar{X}\bar{\theta}_X}(\mu_{003} + 2\bar{X}\mu_{002})$$

$$E(e'_3 e_4) = \frac{1}{n'\bar{X}\bar{\theta}_X}(\mu_{003} + 2\bar{X}\mu_{002}),$$

$$E(e'_3 e'_4) = \frac{1}{n'\bar{X}\bar{\theta}_X}(\mu_{003} + 2\bar{X}\mu_{002})$$

$$E(e_4 e'_4) = \frac{1}{n'\bar{\theta}_X^2}(\mu_{004} + 4\bar{X}\mu_{003} + 4\bar{X}^2\mu_{002} - \mu_{002}^2) \tag{7}$$

For further simplifications, we now use Taylor's series to expand $g(\bar{y}_1, \bar{y}_2, u_1, u_2)$ about the point $(\bar{Y}_1, \bar{Y}_2, 1, 1)$, we have

$$\begin{aligned}
\hat{R}_g = {} & g(\bar{Y}_1, \bar{Y}_2, 1, 1) + (\bar{y}_1 - \bar{Y}_1)g_1 + (\bar{y}_2 - \bar{Y}_2)g_2 + (u_1 - 1)g_3 \\
& + (u_2 - 1)g_4 + \frac{1}{2!}\{(\bar{y}_1 - \bar{Y}_1)^2 g_{11} + (\bar{y}_2 - \bar{Y}_2)^2 g_{22} + (u_1 - 1)^2 g_{33} \\
& + (u_2 - 1)^2 g_{44} + 2(\bar{y}_1 - \bar{Y}_1)(\bar{y}_2 - \bar{Y}_2)g_{12} + 2(\bar{y}_1 - \bar{Y}_1)(u_1 - 1)g_{13} \\
& + 2(\bar{y}_1 - \bar{Y}_1)(u_2 - 1)g_{14} + 2(\bar{y}_2 - \bar{Y}_2)(u_1 - 1)g_{23} + 2(\bar{y}_2 \\
& - \bar{Y}_2)(u_2 - 1)g_{24} + 2(u_1 - 1)(u_2 - 1)g_{34}\} \\
& + \frac{1}{3!}\left\{(\bar{y}_1 - \bar{Y}_1)\frac{\partial}{\partial \bar{y}_1} + (\bar{y}_2 - \bar{Y}_2)\frac{\partial}{\partial \bar{y}_2} + (u_1 - 1)\frac{\partial}{\partial u_1}\right. \\
& \left. + (u_2 - 1)\frac{\partial}{\partial u_2}\right\}^3 g(\bar{y}_1*, \bar{y}_2*, u_1*, u_2^*)
\end{aligned} \tag{8}$$

where $\bar{y}_1* = \bar{Y}_1 + h(\bar{y}_1 - \bar{Y}_1)$, $\bar{y}_2* = \bar{Y}_2 + h(\bar{y}_2 - \bar{Y}_2)$, $u_1^* = 1 + h(u_1 - 1)$, $u_2^* = 1 + h(u_2 - 1)$ for $0 < h < 1$.

Retaining only second order terms and rewriting above equation (8) in terms of e_i's, the result obtained to the approximation of order one is

$$\begin{aligned}
\hat{R}_g - R = {} & (\mathrm{Re}_1 - \mathrm{Re}_2 + e_3 g_3 - e'_3 g_3 + e_4 g_4 - e'_4 g_4) - e_3 e'_3 g_3 \\
& + e'^2_3 g_3 - e_4 e'_4 g_4 + e'^2_4 g_4 + \frac{1}{2!}\{\bar{Y}_1^2 e_1^2 g_{11} + \bar{Y}_2^2 e_2^2 g_{22} \\
& + 2\bar{Y}_1\bar{Y}_2 e_1 e_2 g_{12} + (e_3^2 + e'^2_3 - 2e_3 e'_3)g_{33} \\
& + (e_4^2 + e'^2_4 - 2e_4 e'_4)g_{44} + 2\bar{Y}_1(e_1 e_3 - e_1 e'_3)g_{13} \\
& + 2\bar{Y}_1(e_1 e_4 - e_1 e'_4)g_{14} + 2\bar{Y}_2(e_2 e_3 - e_2 e'_3)g_{23} \\
& + 2\bar{Y}_2(e_2 e_4 - e_2 e'_4)g_{24} + 2(e_3 e_4 - e_3 e'_4 - e'_3 e_4 + e'_3 e'_4)g_{34}\}
\end{aligned} \tag{9}$$

Further we take expectation on both the sides of Equation (9), the bias of the formulated generalized double sampling estimator $\hat{R}_g$ up to terms of order $O(1/n)$ is expressed as

$$Bias(\hat{R}_g) = \frac{1}{2}\left\{\frac{\bar{Y}_1^2}{n}C_{Y_1}^2 g_{11} + \frac{\bar{Y}_2^2}{n}C_{Y_2}^2 g_{22} + 2\frac{\bar{Y}_1\bar{Y}_2}{n}\rho C_{Y_1}C_{Y_2}g_{12}\right\}$$
$$+ 2\bar{Y}_1\rho_1 C_{Y_1}C_X g_{13} + \frac{1}{2}\left(\frac{1}{n} - \frac{1}{n'}\right)$$
$$\times\left\{C_X^2 g_{33} + \frac{1}{\bar{\theta}_X^2}(\mu_{004} + 4\bar{X}\mu_{003} + 4\bar{X}^2\mu_{002} - \mu_{002}^2)g_{44}\right.$$
$$+ \frac{2}{\bar{\theta}_X}(\mu_{102} + 2\bar{X}\mu_{101})g_{14} + 2\bar{Y}_2\rho_2 C_{Y_2}C_X g_{23}$$
$$\left. + \frac{2}{\bar{\theta}_X}(\mu_{012} + 2\bar{X}\mu_{011})g_{24} + \frac{2}{\bar{X}\bar{\theta}_X}(\mu_{003} + 2\bar{X}\mu_{002})g_{34}\right\} \tag{10}$$

We now square Equation (9), the obtained $MSE(\hat{R}_g)$ after taking expectation, is given by

$$MSE(\hat{R}_g) = E\{(\hat{R}_g) - R\}^2$$
$$= R^2E(e_1^2) + R^2E(e_2^2) - 2R^2E(e_1e_2)$$
$$+ g_3^2\{E(e_3^2) + E(e_3'^2) - 2E(e_3e_3')\} + g_4^2\{E(e_4^2) + E(e_4'^2)$$
$$- 2E(e_4e_4')\} + 2Rg_3\{E(e_1e_3) - E(e_1e_3')\} - 2Rg_3\{E(e_2e_3)$$
$$- E(e_2e_3')\} + 2Rg_4\{E(e_1e_4) - E(e_1e_4')\} - 2Rg_4\{E(e_2e_4)$$
$$- E(e_2e_4')\} + 2g_3g_4\{E(e_3e_4) - E(e_3e_4')$$
$$- E(e_3'e_4) + E(e_3'e_4')\}$$

The above expression is simplified further by substituting expected values given in Equations (6) and (7) as

$$MSE(\hat{R}_g) = \frac{\mathrm{R}^2}{n}(C_{Y_1}^2 + C_{Y_2}^2 - 2\rho C_{Y_1}C_{Y_2}) + \left(\frac{1}{n} - \frac{1}{n'}\right)$$
$$\left[C_X^2 g_3^2 + g_4^2\frac{1}{\bar{\theta}_X^2}(\mu_{004} + 4\bar{X}\mu_{003} + 4\bar{X}^2\mu_{002} - \mu_{002}^2)\right.$$
$$+ 2Rg_3C_X(\rho_1 C_{Y_1} - \rho_2 C_{Y_2})$$

$$+ 2Rg_4\frac{1}{\bar{\theta}_X}\left\{\frac{(\mu_{102} + 2\bar{X}\mu_{101})}{\bar{Y}_1} - \frac{(\mu_{012} + 2\bar{X}\mu_{011})}{\bar{Y}_2}\right\}$$
$$\left. + 2g_3g_4\frac{1}{\bar{X}\bar{\theta}_X}(\mu_{003} + 2\bar{X}\mu_{002})\right] \tag{11}$$

The value of $MSE(\hat{R}_g)$ given in Equation (11) depends on the values of g_3 and g_4, Hence we differentiate the above Equation (11) with respect to g_3 and g_4. The obtained optimum result of g_3 and g_4 for which $MSE(\hat{R}_g)$ in Equation (11) attains the minimum value are

$$g_3 = -\frac{R}{CC_X^3}\left\{C^2C_X^2 + \frac{\delta_2(\delta_2 - \delta_1)}{\Delta}\right\} \tag{12}$$

$$g_4 = \frac{R\bar{X}\bar{\theta}_X(\delta_2 - \delta_1)}{\Delta} \tag{13}$$

where

$$\Delta = C_X^2\bar{X}^2(\mu_{004} + 4\bar{X}\mu_{003} + 4\bar{X}^2\mu_{002} - \mu_{002}^2) - (\mu_{003} + 2\bar{X}\mu_{002})^2 \geq 0$$

$$\delta_1 = C_X^2\bar{X}\left\{\frac{(\mu_{102} + 2\bar{X}\mu_{101})}{\bar{Y}_1} - \frac{(\mu_{012} + 2\bar{X}\mu_{011})}{\bar{Y}_2}\right\}$$

$$\delta_2 = (\rho_1C_{Y_1} - \rho_2C_{Y_2})C_X(\mu_{003} + 2\bar{X}\mu_{002})$$

$$C = (\rho_1C_{Y_1} - \rho_2C_{Y_2})$$

Additionally, the resultant value for minimum mean squared error of $\hat{R}_g$ represented by $MSE(\hat{R}_g)_{\min}$ can be acquired by substituting results obtained in Equations (12) and (13) in Equation (11) as

$$MSE(\hat{R}_g)_{\min} = \frac{R^2}{n}(C_{Y_1}^2 + C_{Y_2}^2 - 2\rho C_{Y_1}C_{Y_2})$$
$$- \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{R^2C^2 + \frac{R^2\left(\delta_2 - \delta_1\right)^2}{\Delta C_X^2}\right\}$$

Or

$$MSE(\hat{R}_g)_{\min} = MSE(\hat{R}) - \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{R^2C^2 + \frac{R^2(\delta_2 - \delta_1)^2}{\Delta C_X^2}\right\} \tag{14}$$

4 Efficiency Comparison

(i) The usual estimator of ratio of two population means and its particular cases with their respective mean squared error are given as follows

Table 1 Some particular estimators with relative MSE

Estimators	MSE
$\hat{R}=\frac{\bar{y}_1}{\bar{y}_2}$	$\frac{\mathrm{R}^2}{n}(C^2_{Y_1}+C^2_{Y_2}-2\rho C_{Y_1}C_{Y_2})$
$\hat{R}_1=\frac{\bar{y}_1}{\bar{y}_2}+k(\bar{x}-\bar{X})=\hat{R}+k(\bar{x}-\bar{X})$	$MSE(\hat{\mathrm{R}})-\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2D^2$
$\hat{R}_2=\frac{\bar{y}_1+k(\bar{x}-\bar{X})}{\bar{y}_2}$	$MSE(\hat{\mathrm{R}})-\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2D^2$
$\hat{R}_3=\frac{\bar{y}_1}{\bar{y}_2}.\frac{\bar{x}}{\bar{X}}$	$MSE(\hat{\mathrm{R}})+\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2(1+2D)$
$\hat{R}_4=\frac{\bar{y}_1}{\bar{y}_2}.\frac{\bar{X}}{\bar{x}}$	$MSE(\hat{\mathrm{R}})+\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2(1-2D)$
$\hat{R}_5=\frac{\bar{y}_1+k(\bar{x}-\bar{X})}{\bar{y}_2+(\bar{x}-\bar{X})}$	$MSE(\hat{\mathrm{R}})-\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2D^2$
$\hat{R}_6=\frac{\bar{y}_1+k(\bar{x}-\bar{X})}{\bar{y}_2}\left(\frac{\bar{x}}{\bar{X}}\right)$	$MSE(\hat{\mathrm{R}})-\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2D^2$

(15)

where

$$D=\left(\frac{\rho_1C_{Y_1}-\rho_2C_{Y_2}}{C_X}\right)$$

(ii) The generalized estimator of ratio of two population means by Singh and Naqvi (2015) and its respective mean squared error is given as $\hat{R}_7=g(\bar{y}_1,\bar{y}_2,\bar{x})$

$$MSE(\hat{R}_7)_{\min}=MSE(\hat{R})-\left(\frac{1}{\mathrm{n}}-\frac{1}{N}\right)C^2_XR^2D^2 \tag{16}$$

The proposed generalized estimator for ratio of two populations mean $\hat{R}_g$ has minimum mean squared error as

$$MSE(\hat{R}_g)_{\min}=MSE(\hat{R})-\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{R^2C^2+\frac{R^2\left(\delta_2-\delta_1\right)^2}{\Delta C^2_X}\right\} \tag{17}$$

From *MSE* given in Equations (15)–(17) and the section given below, it can be clearly concluded that the suggested class of generalized double sampling estimator utilizing known values of first and second moment about zero has lesser MSE as compared to the usual estimator where in no such information is used. Therefore, for obtaining precise results, the use of proposed estimator under practical situation is recommended.

5 An Empirical Study

To support theoretical results, a numerical illustration has been carried out using the data set given on page 177, Singh and Chaudhary (2009). The summary of the population data set is as follows.

$$\bar{Y}_1 = 856.4118, \quad \bar{Y}_2 = 208.8824, \quad \bar{X} = 199.4412, \quad C_{Y_1} = 0.8372,$$
$$C_{Y_2} = 0.7205, \quad C_X = 0.7532, \quad \rho_{Y_1Y_2} = 0.2090, \quad \rho_{Y_1X} = 0.2105,$$
$$\rho_{Y_2X} = 0.9801, \quad n = 12, \quad n' = 34$$

Table 2 MSE and PRE comparison of traditional estimators with $\hat{R}_g$

Estimators	MSE	PRE
$\hat{R}$	1.555166	159
$\hat{R}_1$	1.103932	113
$\hat{R}_2$	1.103932	113
$\hat{R}_3$	1.184017	121
$\hat{R}_4$	3.749341	384
$\hat{R}_5$	1.103932	113
$\hat{R}_6$	1.103932	113
$\hat{R}_7$	1.103932	113
$\hat{R}_g$	0.974356	100

6 Conclusions and Discussion

(i) The minimum MSE for the estimator represented by $\hat{R}_g$ is

$$MSE(\hat{R}_g)_{\min} = MSE(\hat{R}) - \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{R^2C^2 + \frac{R^2(\delta_2 - \delta_1)^2}{\Delta C_X^2}\right\} \tag{18}$$

Any estimator belonging to the suggested generalized class of estimators represented by $\hat{R}_g$ cannot have mean squared error smaller than the expression (18).

(ii) There exists a subset of estimators satisfying Equations (12) and (13) in the class $\hat{R}_g$ such that every member of this subset attains the similar minimum mean squared error (MSE) as obtained in Equation (14). For example, the estimators

$$\hat{R}_{P_1} = \frac{1}{\overline{y}_2}\left\{\overline{y}_1\left(\frac{\overline{x}}{\overline{x}'}\right)\left(\frac{\bar{\theta}_x}{\bar{\theta}'_x}\right)\right\} \tag{19}$$

$$\hat{R}_{P_2} = \frac{1}{\overline{y}_2}\left\{\overline{y}_1\left(\frac{\overline{x}}{\overline{x}'}-1\right)\left(\frac{\bar{\theta}_x}{\bar{\theta}'_x}-1\right)\right\} \tag{20}$$

$$\hat{R}_{P_3} = \frac{1}{\overline{y}_2}\left\{\overline{y}_1\left(\frac{\overline{x}}{\overline{x}'}\right)^{k_1}\left(\frac{\bar{\theta}_x}{\bar{\theta}'_x}\right)^{k_2}\right\} \tag{21}$$

$$\hat{R}_{P_4} = \frac{1}{\overline{y}_2}\left\{\overline{y}_1 + k_1\left(\frac{\overline{x}}{\overline{x}'}-1\right) + k_2\left(\frac{\bar{\theta}_x}{\bar{\theta}'_x}-1\right)\right\} \tag{22}$$

are some particular members of the proposed generalized class and also attains the similar minimum mean squared error as given in Equation (14).

(iii) The MSE $(\hat{R}_g)$ of the formulated estimator $\hat{R}_g$ is minimized for the optimum values given in Equations (12) and (13), the obtained optimum values of g_3 and g_4 are

$$g_3 = -\frac{R}{CC_X^3}\left\{C^2C_X^2 + \frac{\delta_2(\delta_2-\delta_1)}{\Delta}\right\} \tag{23}$$

$$g_4 = \frac{R\bar{X}\bar{\theta}_X(\delta_2-\delta_1)}{\Delta} \tag{24}$$

Under many practical situations, the values of some unknown parameters involved in optimum values may not be known a priori. Hence to overcome such situations it is suggested to use unbiased estimators of unknown parameters of the optimum values.

(iv) From theoretical and empirical efficiency comparison it can be reasonably concluded that the suggested estimator will yield valid and accurate results and is also relatively efficacious than the traditional estimators. Therefore, the application of proposed estimator for practical situations is substantially advisable.

Acknowledgement

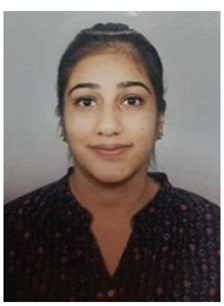
The authors are very much indebted to the editor-in-chief Prof. (Dr.) Vinod Kumar and learned referees for their valuable suggestions leading to improvement of the quality in the contents of the paper.

References

[1] Biradar, R.S. and Singh, H.P. (1997–98). A class of estimators for population parameters using supplementary information, Aligarh J. Statist., 17 & 18, pp. 54–71.
[2] Cochran, W.G. (1977). Sampling Techniques, 3rd edition, John Wiley and Sons, New York.
[3] Kumar, K. and Srivastava, U. (2018). Estimation of ratio and product of two population means using exponential type estimators in sample surveys, International Journal of Mathematics and Statistics, 19–3, pp. 102–109.
[4] Mukhopadhyay, P. (2012). Theory and Methods of Survey and Sampling, 2nd edition, PHI Learning Private Limited, New Delhi, India.
[5] Murthy, M. (1967). Sampling Theory and Methods, 1st edition, Calcutta Statistical Publishing Society, Kolkata, India.
[6] Neyman, J. (1938). Contribution to the theory of sampling human populations, Journal of the American Statistical Association, 33(201), pp. 101–116.
[7] Shah S.M. and Shah D.N (1978). Ratio cum product estimator for estimation ratio (product) of two population parameters, Sankhya C, 40, pp. 156–166.
[8] Singh, D. and Chaudhary, F.S. (1997). Theory and Analysis of Sampling Survey Designs, New Age International Publishers, New Delhi, India.
[9] Singh, D. and Chaudhary, F.S. (2009). Sampling Techniques (2nd ed.), New Age International Publishers, New Delhi, India.
[10] Singh, G.N. and Rani, R. (2005, 2006). Some linear transformations on auxiliary variable for estimating the ratio of two population means in sample surveys, Model Assisted. Statistics and Applications, 1(1), IOS Press, 1–5.
[11] Singh, H.P. (1998). On the estimation of ratio and product of two finite populations means, Proc. Nat. Acad. Sci. India, Sec. A, 58, pp. 399–402.
[12] Singh, M.P. (1965). On the estimation of ratio and product of population parameters, Sankhya B, 27, pp. 321–328.

[13] Singh, M.P. (1967). Ratio cum product method of estimation, Metrika, 12, pp. 34–43.
[14] Singh, M.P. (1969). Comparison of some ratio cum product estimators, Sankhya B, 31, pp. 375–378.
[15] Singh, R. K. and Naqvi, N. (2015). A generalized class of estimator of the ratio of two population means using auxiliary information, Sri Lankan Journal of Applied Statistics, 16–3, pp. 179–193.
[16] Singh, R.K. (1982). On estimating ratio and product of population parameters, Calcutta Statist. Assoc. Bull., 20, pp. 39–49.
[17] Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok, C. (1984). Sampling Theory of Surveys with Applications, 3rd Edition, Ames, Iowa (USA) and Indian Society of Agricultural Statistics, New Delhi, India.
[18] Tripathi, T.P. (1980). A general class of estimators of population ratio, Sankhya C, 42, pp. 63–75.
[19] Upadhyaya, L.N., Singh, G.N. and Singh, H.P. (2000). Use of transformed auxiliary variable in the estimation of population ratio in sample survey, Statistics in Transition, 4, pp. 1019–1027.

Biographies



Tarunpreet Kaur Ahuja is a Research Scholar in Department of Statistics, D.A.V. (P.G) College, Dehradun, which is affiliated to Hemvati Nandan Bahuguna Garhwal University (A Central University), Srinagar, Uttarakhand. She has completed her B.Sc. (Physics, Mathematics & Statistics) degree in 2014 and M.Sc. in Statistics in 2016 from D.A.V. (P.G) College, Dehradun, Uttarakhand. She worked as an intern for Data Analytics team in Pitney Bowes, a commerce company in 2016. Her research interest is in the area of developing sampling theory and she has also published four research papers in different national and international journals of repute.

Peeyush Misra is working as an Associate Professor in Department of Statistics, D.A.V. (P.G) College, Dehradun, which is affiliated to Hemvati Nandan Bahuguna Garhwal University (A Central University), Srinagar, Uttarakhand. He has completed his Graduation, Post Graduation and PhD degree in Statistics from the University of Lucknow, Lucknow, Uttar Pradesh. He is contributing to the scientific field of Statistics mainly in the area of sampling theory. He has published more than sixty research papers in various national and international journals of repute. He has also successfully completed one research project sponsored by UGC.

O. K. Belwal is currently working as a Professor and Head of Department of Statistics, Hemvati Nandan Bahuguna Garhwal University (A Central University), Srinagar, Uttarakhand. He has contributed to the various fields of Statistics through many research publications in different national and international journals of repute. He has also successfully completed some research projects in the field of Statistics. He has also successfully organized some national seminars/conferences of repute. He has also supervised many research scholars.