Scientific
Research

Ratio-Cum-Product Estimator Using Multiple Auxiliary Attributes in Single Phase Sampling

John Kung'u, Leo Odongo

Department of Mathematics, Kenyatta University, Nairobi, Kenya
Email: johnkungu08@yahoo.com

Abstract

In this paper, we have proposed a class of ratio-cum-product estimator for estimating population mean of study variable for single phase sampling using multi-auxiliary attributes. The expressions for mean square error are derived. An empirical study is given to compare the performance of the estimator with existing estimators. It has been found that the ratio-cum-product estimator using multiple auxiliary attributes is more efficient than mean per unit, product and ratio estimators using one auxiliary attribute, and Product and Ratio estimators using multiple auxiliary attributes in single phase sampling.

Keywords

Ratio-Cum-Product Estimator, Multiple Auxiliary Attributes, Single Phase Sampling, Bi-Serial Correlation Coefficient

1. Introduction

The use of supplementary information is widely in sampling theory. Auxiliary variables are commonly used in sample survey practices to obtain improved designs and to achieve higher precision in the estimates of some population parameters such as the mean or the total of a study variable. The concept of ratio estimation was introduced in sample survey by Cochran [1]. It is preferred when the study variable is highly positively correlated with the auxiliary variable. Murthy [2] proposed the product estimator for negatively correlated study variable(s) and auxiliary variable which was similar to ratio estimator. Olkin [3] was the first author to use information on more than one supplementary characteristics, which are positively correlated with the variable under study, using a linear combination of ratio estimator based on each auxiliary variable. Raj [4] suggested a method of using

multi-auxiliary information in sample survey. Using this idea, Singh [5] proposed a multivariate expression of product estimator where the study variable was negatively correlated with the multi-auxiliary variables. In the same year, Singh [6] proposed a ratio-cum-product estimator and its multi-variable expression. Singh and Tailor [7] proposed a ratio-cum-product estimator for finite population mean in simple random sampling using coefficient of variation and coefficient of kurtosis.

Shabbir and Gupta [8] proposed estimation of the finite population mean with population proportion of an auxiliary variable while Jhajj, Sharma and Grover [9] proposed a general family of estimators using information on auxiliary attribute. They used known information of population proportion possessing an attribute (highly correlated with study variable Y). The attribute is normally used when the auxiliary variables are not available e.g. amount of milk produced and a particular breed of cow or amount of yield of wheat and a particular variety of wheat. Bahland Tuteja [10] proposed ratio and product type exponential estimators using auxiliary attribute. Rajesh, Pankaj, Nirmala and Florentins [11] used the information on auxiliary attribute in ratio estimator in estimating population mean of the variable of interest using known attributes such as coefficient of variation, coefficient of kurtosis and point bi-serial correlation coefficient. Rajesh, S., Pankaj Nirmala and Florentins [11] also used the auxiliary attributes in ratio-product type exponential estimator following the work of Bahland Tuteja [10]. In most of the survey, the auxiliary information is always available and every form of auxiliary information should be used in developing sampling strategies as introduced by Simiuddinand Hanif [12]. We have used those strategies to develop a ratio-cum product estimator using multiple auxiliary attributes in single phase sampling.

Hanif, Haq and Shahbaz [13] proposed a general family of estimators using multiple auxiliary attribute in single and double phase sampling. They also extended their work to ratio, product and regression estimators which were generalization of Naik and Gupta [14] estimator in single and double phase sampling with full information, partial information and no information. The concept of multiple auxiliary attributes was proposed by Hanif, Haq and Shahbaz [15] and [16], and then extended to ratio and product estimators. In this paper, we have incorporated the multiple auxiliary attributes in Ratio-cum-product estimator proposed by Singh [6] and also incorporate Arora and Bansi [17] approach in writing down the mean squared error.

2. Preliminaries

2.1. Notations

Let $\left(y_i, \tau_{i1}, \tau_{i2}, \cdots, \tau_{iq} \right)$ be the ith sample point from a population of size N where y_i is the value of variable of interest and $\tau_{ij} = \left(j = 1, 2, \cdots, q \right)$ is the value of jth auxiliary attribute. In defining the attributes we assume complete dichotomy so that;

$$\tau_{ij} = \begin{cases} 1, & \text{if } i^{th} \text{ unit of population possess } j^{th} \text{ auxiliary attribute} \\ 0, & \text{otherwise} \end{cases}$$

Let $A_j = \sum_{i=1}^{N} \tau_{ij}$ and $a_j = \sum_{i=1}^{n} \tau_{ij}$ be the total number of units in the population and sample respectively possessing attribute τ_j. Let $P_j = \dfrac{A_j}{N}$ and $p_j = \dfrac{a_j}{n}$ be the corresponding proportion of units possessing a specific attributes τ_j and $\overline{y}$ is the mean of the main variable at first phase.

$$\text{Further, let } \quad \theta = \left(\frac{1}{n_1} - \frac{1}{N} \right), \quad \overline{y} = \overline{Y} + \overline{e}_y \quad \text{and} \quad \overline{e}_{\tau_j} = p_j - P_j \quad \left(j = 1, 2, \cdots, q \right) \tag{1.0}$$

where e_y and $\overline{e}_{\tau_j}$ are sampling errors and are of very small quantities.
We assume that $E\left(\overline{e}_y \right) = E\left(\overline{e}_{\tau_j} \right) = 0$.

$$\text{Let } \quad v_j = \frac{\overline{p}_j}{\overline{P}_j}, \quad v_j - 1 = \frac{p_j}{P_j} - 1 = \frac{p_j - P_j}{P_j} = \frac{\overline{e}_{\tau_j}}{P_j} \quad \text{then} \quad v_j = 1 + \frac{\overline{e}_{\tau_j}}{P_j}. \tag{1.1}$$

Then for simple random sampling without replacement for first phase sampling, we write by using phase wise operation of expectations as:

$$E\left(\overline{e}_y^2\right) = \theta \overline{Y}^2 C_y^2, \quad E\left(\overline{e}_{\tau_j}^2\right) = \theta P_j^2 C_{\tau_j}^2 \text{ and } E\left(\overline{e}_y \overline{e}_{\tau_j}\right) = \theta \overline{Y} P C_y C_{\tau_j} \rho_{pb_j} \quad (1.2)$$

where $C_y^2 = \dfrac{S_y^2}{\overline{Y}^2}$ and $C_{\tau_j} = \dfrac{S_{\tau_i}^2}{P_j}$ are the coefficient of variation of study variable and the auxiliary variables respectively.

The bi-serial correlation coefficient between study variable and auxiliary attributes is given by $\rho_{Pb_j} = \dfrac{S_{y\tau_j}}{S_y S_{\tau_j}}$. If the information proportion P_j is known for the entire auxiliary attribute, then, a general an estimator is suggested as $T_1 = \gamma_w\left(\overline{y}, v_1, v_2, \cdots, v_q\right)$ Where $v_j = \dfrac{p_j}{P_j}$ and $v_j > 0$, p_j is the sample proportion of jth attributes. $T_1 = \gamma_\omega\left(\overline{y}, v_1, v_2, \cdots, v_q\right)$ is the parametric function such that $\gamma_\omega\left(\overline{Y}, 1, 1, \cdots, 1\right) = \overline{Y}$ and the point $\left(\overline{y}, v_1, v_2, \cdots, v_q\right)$ are to be bounded set in R_K containing a point $\left(\overline{Y}, 1, 1, \cdots, 1\right)$

If A is a square matrix, its inverse can be written using adjoint matrix as

$$A^{-1} = \frac{1}{|A|}\left(C^T\right)_{ij} = \frac{Adj(A)}{|A|}, \quad (1.3)$$

$$\frac{|R|_{y\tau_p}}{|R|_{\tau_p}} = \left(1 - \rho_{y.\tau_p}^2\right) \text{ Arora and Lai [17]} \quad (1.4)$$

The following notations will be used in deriving the mean square errors of proposed estimator.

$|R|_{y\tau_p}$ Determinant of population correlation matrix of attributes $y, \tau_1, \tau_2, \cdots, \tau_{p-1}$ and τ_q

$\left|R_{y\tau_j}\right|_{y\tau_p}$ Determinant of jth minor of $|R|_{y\tau_q}$ corresponding to the jth element of ρ_{Pb_j}

$\rho_{Pb_r}^2$ Denotes the multiple coefficient of determination of y on $\tau_1, \tau_2, \cdots, \tau_{r-1}$ and τ_r.

$\rho_{Pb_r}^2$ Denotes the multiple coefficient of determination of y on $y, \tau_1, \tau_2, \cdots, \tau_{q-1}$ and τ_q.

$|R|_{\tau_r}$ Determinant of population correlation matrix of attributes $\tau_1, \tau_2, \cdots, \tau_{r-1}$ and τ_r.

$|R|_{\tau_q}$ Determinant of population correlation matrix of attributes $x_1, x_2, \cdots, x_{p-1}$ and x_p

$|R|_{y_i\tau_r}$ Determinant of the correlation matrix of $y_i, \tau_1, \tau_2, \cdots, \tau_{r-1}$ and τ_r.

$|R|_{y_i.\tau_q}$ Determinant of the correlation matrix of $y_i, \tau_1, \tau_2, \cdots, \tau_{q-1}$ and τ_q.

$$(1.5)$$

2.2. Ratio and Product Estimator Using Auxiliary Attribute

The sample mean $\overline{y}$ using simple random sampling without replacement is given by,

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (2.1)$$

while variance of $\overline{y}$ is given by,

$$V\left(\overline{y}\right) = \overline{Y}^2 \theta C_y^2 \quad (2.2)$$

In order to have an estimate of the population mean $\overline{Y}$ the study variable y, assuming the knowledge of the population proportion P, Naik and Gupta [14] defined ratio and product estimators of population mean when the

prior information of population proportion of units, possessing the same attribute is variable. Naik and Gupta [14] proposed the following estimators:

$$t_{R1(1)} = \overline{y}\left(\frac{P}{p}\right)^{\alpha_1} \tag{2.3}$$

$$t_{P1(1)} = \overline{y}\left(\frac{p}{P}\right)^{\alpha_2} \tag{2.4}$$

The MSE of $t_{R1(1)}$ and $t_{P1(1)}$ up to the first order of approximation are

$$MSE\left(t_{R1(1)}\right) = \theta\overline{Y}^2\left(C_y^2 + \alpha_1^2 C_P^2 - 2\alpha_1 C_P C_y \rho_{Pb}\right) \tag{2.5}$$

$$MSE\left(t_{P1(1)}\right) = \theta\overline{Y}^2\left(C_y^2 + \alpha_2^2 C_P^2 + 2\alpha_2 C_P C_y \rho_{Pb}\right) \tag{2.6}$$

The optimums value are $\alpha_1 = \dfrac{C_y \rho_{Pb}}{C_r}$ and $\alpha_2 = -\dfrac{C_y \rho_{Pb}}{C_r}$ for ratio and product estimator respectively.

2.2. Ratio and Product Estimator Using Multiple Auxiliary Attributes

The ratio and product estimators by Hanif, Haq and Shahbaz [16] for single phase sampling using information on multiple auxiliary attributes are given respectively by,

$$t_{R2(1)} = \overline{y}\left(\frac{P_1}{p_1}\right)^{\beta_1}\left(\frac{P_2}{p_2}\right)^{\beta_2}\cdots\left(\frac{P_q}{p_q}\right)^{\beta_q} \tag{2.7}$$

$$t_{P2(1)} = \overline{y}\left(\frac{p_1}{P_1}\right)^{\alpha_1}\left(\frac{p_2}{P_2}\right)^{\alpha_2}\cdots\left(\frac{p_k}{P_k}\right)^{\alpha_j} \tag{2.8}$$

The MSE of the $t_{R2(1)}$ and $t_{P2(1)}$ up to the first order of approximation are given respectively by,

$$MSE\left(t_{R2(1)}\right) = \theta\overline{Y}^2 C_y^2\left(1 - \rho_{Pb\tau_q}^2\right) \tag{2.9}$$

$$MSE\left(t_{P2(1)}\right) = \theta\overline{Y}^2 C_y^2\left(1 - \rho_{Pb\tau_q}^2\right) \tag{2.10}$$

3. Methodology

3.1. The Proposed Ratio-Cum-Product Estimator Using Multiple Auxiliary Attributes in Single Phase Sampling

If we estimate a study variable when information on all auxiliary attributes is available from population, it is utilized in the form of their means. By taking the advantage of ratio-cum-product technique for single phase sampling, a generalized estimator for estimating population mean of study variable *Y* with the use of multi-auxiliary attributes is given by:

$$t_{RP(3.0)} = \overline{y}\left(\frac{P_1}{p_1}\right)^{\alpha_1}\left(\frac{P_2}{p_2}\right)^{\alpha_2}\cdots\left(\frac{P_k}{p_k}\right)^{\alpha_k}\left(\frac{p_{k+1}}{P_{k+1}}\right)^{\beta_{k+1}}\left(\frac{p_{k+2}}{P_{k+2}}\right)^{\beta_{k+2}}\cdots\left(\frac{p_q}{P_q}\right)^{\beta_q} \tag{3.0}$$

Simplifying (3.0) and substituting (1.0) we get,

$$t_{RP(3.0)} = \left(\overline{e}_y + \overline{Y}\right)\prod_{j=1}^{k}\left(\frac{\overline{e}_{r_j} + P_j}{P_j}\right)^{-1}\prod_{j=k+1}^{h+k=q}\left(\frac{\overline{e}_{r_j} + P_j}{P_j}\right) \tag{3.1}$$

Using (1.0) and (1.1) and ignoring the second and higher terms for each expansion of product and after simplification, we write,

$$t_{RP(3.0)} = \left(\overline{e}_y + \overline{Y} \right) - \sum_{j=1}^{k} \alpha_j \frac{\overline{Y}}{P_j} \overline{e}_{r_j} + \sum_{j=k+1}^{h+k=q} \beta_j \frac{\overline{Y}}{P_j} \overline{e}_{r_j} \tag{3.2}$$

The mean squared error of ratio-cum-product estimator is given by,

$$MSE\left(t_{RP(3.0)} \right) = E\left(t_{RP(3.0)} - \overline{Y} \right)^2 = E\left(\overline{e}_y - \sum_{j=1}^{k} \alpha_j \frac{\overline{Y}}{P_j} \overline{e}_{r_j} + \sum_{j=k+1}^{h+k=q} \beta_j \frac{\overline{Y}}{P_j} \overline{e}_{r_j} \right)^2 \tag{3.3}$$

We differentiate Equation (3.3) partially with respect to α_i $\left(i = 1, 2, \cdots, k \right)$ and β_i $\left(i = k+1, k+2, \cdots, q \right)$ then equate to zero, using (1.2), (1.3) and (1.4), we get.

$$\alpha_j = (-1)^{j+1} \frac{C_y}{|R|_{\tau_q} C_{\tau_j} P_j} \left| R_{y\tau_j} \right|_{\tau_q} \tag{3.4}$$

$$\beta_j = -(-1)^{j+1} \frac{C_y}{|R|_{\tau_p} C_{\tau_j} P_j} \left| R_{y\tau_j} \right|_{\tau_q} \tag{3.5}$$

Using normal equations that are used to find the optimum values of α_i and β_i can be written in simplified form as:

$$MSE\left(t_{RP(3.0)} \right) = E\left(\overline{e}_y \left(\overline{e}_y - \sum_{j=1}^{k} \alpha_j \overline{Y} \frac{\overline{e}_{\tau_j}}{P_j} + \sum_{j=k+1}^{h+k=q} \beta_j \overline{Y} \frac{\overline{e}_{\tau_j}}{P_j} \right) \right) \tag{3.6}$$

Using (1.3) in (3.6), we get,

$$MSE\left(t_{RP(3.0)} \right) = \theta \overline{Y}^2 \left(C_y^2 - \sum_{j=1}^{k} \alpha_j P_j C_y C_{\tau_j} \rho_{Pb_j} + \sum_{j=k+1}^{k+h=q} \beta_j P_j C_y C_{\tau_j} \rho_{Pb_j} \right) \tag{3.7}$$

Using the optimum value α_j and β_j in (3.4), (3.5) in (3.7), we get

$$MSE\left(t_{RP(3.0)} \right) = \theta \overline{Y}^2 \left(C_y^2 + \sum_{j=1}^{k} (-1)^j \frac{C_y}{P_j C_{\tau_j}} \frac{\left| R_{y\tau_j} \right|_{\tau_q}}{|R|_{\tau_q}} P_j C_y C_{\tau_j} \rho_{Pb_j} + \sum_{j=k+1}^{k+h=q} (-1)^j \frac{C_y}{P_j C_{\tau_j}} \frac{\left| R_{y\tau_i} \right|_{\tau_q}}{|R|_{\tau_q}} P_j C_y C_{\tau_j} \rho_{Pb_j} \right) \tag{3.8}$$

$$\text{Or} \quad MSE\left(t_{RP(3.0)} \right) = \theta \overline{Y}^2 C_y^2 \left(1 + \left(\sum_{j=1}^{q} (-1)^j \frac{\left| R_{y\tau_i} \right|_{\tau_q}}{|R|_{\tau_q}} \rho_{Pb_j} \right) \right) \tag{3.9}$$

Or

$$MSE\left(t_{RP(3.0)} \right) = \frac{\theta \overline{Y}^2 C_y^2}{|R|_{\tau_q}} \left(|R|_{\tau_p} + \sum_{j=1}^{q} (-1)^j \left| R_{y\tau_i} \right|_{y\tau_q} \rho_{Pb_j} \right) \tag{3.10}$$

Or

$$MSE\left(t_{RP(3.0)} \right) = \theta \overline{Y}^2 C_y^2 \frac{|R|_{y\tau_q}}{|R|_{\tau_q}} \tag{3.11}$$

$$MSE\left(t_{RP(3.0)} \right) = \theta \overline{Y}^2 C_y^2 \left(1 - \rho_{y.Pb\tau_q}^2 \right) \tag{3.12}$$

3.1. Bias and Consistency of Ratio-Cum-Product Estimators

These ratio-cum-product estimators using multiple auxiliary attribute in single phase sampling is biased. However, these biases are negligible for moderate and large samples. It's easily shown that the ratio-cum-product estimator is consistent estimator using multiple auxiliary variables since it is a linear combinations

Table 1. Relative efficiency of existing and proposed estimator with respect to mean per unit estimator for single phase sampling.

Estimators	Percent relative efficiency with respect with the mean per unit
$\bar{y}$	100
$t_{R1(1)}$	161
$t_{P1(1)}$	128
$t_{R2(1)}$	202
$t_{P2(1)}$	189
$t_{RP(3.0)}$ (Proposed)	230

of consistent estimators it follows that it is also consistent.

4. Simulation, Results and Discussion

In this section, we carried out some data simulation experiments to compare the performance of ratio-cum product estimator with already existing estimators of finite population mean that uses one or multiple auxiliary attributes. The data for the empirical study is a normally distributed simulated population with the following variables

$N = 300$, $n = 45$, mean = 45, standard deviation = 5

$$\rho_{Pb_{\tau_1}} = 0.7678 \quad \rho_{Pb_{\tau_2}} = 0.6582 \quad \rho_{Pb_{\tau_{1p}}} = -0.6821 \quad \rho_{\tau_1\tau_2} = 0.2188 \quad \rho_{\tau_1\tau_{1p}} = -0.1533 \quad \rho_{\tau_2\tau_{1p}} = -0.1864$$

In order to evaluate the efficiency gain we could achieve by using the proposed estimators, we calculated the variance of mean per unit and the mean squared error of all estimators we considered. We calculated percent relative efficiency of each estimator in relation to variance of mean per unit. We then compared the percent relative efficiency of each estimator, the estimator with the highest percent relative efficiency is considered to be the most efficient compared to other estimators. The efficiency was calculated using the following formulae.

$$eff\left(\hat{\bar{Y}}\right) = \frac{MSE\left(\hat{\bar{y}}\right)}{MSE\left(\hat{\bar{Y}}\right)} \times 100 \tag{4.0}$$

According to **Table 1**, the proposed ratio-cum-product estimator using multiple auxiliary attributes in single phase sampling has the highest percent relative efficiency compared to mean per unit, ratio and product estimator using one auxiliary attribute and Ratio and Product estimator using multiple auxiliary attributes in the simulated population. This means that the Ratio-cum-product estimator in single phase sampling is the most efficient estimator compared to the estimators that utilize auxiliary attributes. The proposed ratio-cum-product estimatorusing multiple auxiliary attributesin single phase samplingis recommended to estimate the finite populations mean when all auxiliary attributes are known. In our future work, we will extend this work to two phase sampling.

References

[1] Cochran, W.G. (1940) The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *The Journal of Agricultural Science*, **30**, 262-275. http://dx.doi.org/10.1017/S0021859600048012

[2] Murthy, M.N. (1964) Product Method of Estimator. *The Indian Journal of Statistics Series*, **26**, 69-74.

[3] Olkin, I. (1958) Multivariate Ratio Estimation for Finite Population. *Biometrika*, **45**, 154-165. http://dx.doi.org/10.1093/biomet/45.1-2.154

[4] Raj, D. (1965) On a Method of Using Multi-Auxiliary Information in Sample Surveys. *Journals of the American Statistical Association*, **60**,154-165. http://dx.doi.org/10.1080/01621459.1965.10480789

[5] Singh, M.P. (1967) Multivariate Product Method of Estimation for Finite Population. *Journal of the Indian Society of Agriculture Statistics*, **31**, 375-378.

[6] Singh, M.P. (1967) Multivariate Product Method of Estimation for Finite Population. *Journal of the Indian Society of*

Agriculture Statistics, **31**,375-378.

[7] Singh, H.P. and Tailor, R. (2003) Use of Known Correlation Coefficient in Estimating the Finite Population Mean. *Statistics in Translation*, **6**, 553-560.

[8] Shabbir, J. and Gupta, S. (2007) On Estimating the Finite Population Mean with Population Proportion of an Auxiliary Variable. *Pakistan Journal of Statistics*, **23**, 1-9.

[9] Jhajj, H.S., Sharma, M.K. and Grover, L.K. (2006) A Family of Estimator of Population Mean Using Information on Auxiliary Attributes. *Pakistan Journal of Statistics*, **22**, 43-50.

[10] Bahl, S. and Tuteja, R.K. (1991) Ratio and Product Type Estimator. *Information and Optimization Science*, **12**, 159-163.

[11] Rajesh, S., Pankaj, C., Nirmala, S. and Florentins, S. (2007) Auxiliary Information and a Priori Values in Construction of Improved Estimators. Renaissance High Press, USA.

[12] Simiuddin, M. and Hanif, M. (2007) Estimation of Population Mean in Single and Two Phase Sampling with or without Additional Information. *Pakistan Journal of Statistics*, **23**, 99-118.

[13] Hanif, M., Haq, I.U. and Shahbaz, M.Q. (2009) On a New Family of Estimator Using Multiple Auxiliary Attributes. *World Applied Science Journal*, **11**, 1419-1422.

[14] Naik, V.D. and Gupta, P.C. (1996) A Note on Estimation of Mean with Known Population of Auxiliary Character. *Journal of the Indian Society of Agricultural Statistics*, **48**, 151-158.

[15] Hanif, M., Haq, I.U. and Shahbaz, M.Q. (2010) Ratio Estimator Using Multiple Auxiliary Attributes. *World Applied Sciences Journal*, **8**, 133-136.

[16] Haq, I.U. (2009) A Family of Estimators for Two-Phase Sampling Using Multi-Auxiliary Attributes. Ph.D. Thesis, National College of Business Administration & Economics, Lahore.

[17] Arora, S. and Lal, B. (1989) New Mathematical Statistics. Satya Prakashan, New Delhi.