

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

John Kung'u, Leo Odongo

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

Received 21 March 2014; revised 21 April 2014; accepted 28 April 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we have proposed three classes of ratio-cum-product estimators for estimating population mean of study variable for two-phase sampling using multi-auxiliary attributes for full information, partial information and no information cases. The expressions for mean square errors are derived. An empirical study is given to compare the performance of the estimator with the existing estimator that utilizes auxiliary attribute or multiple auxiliary attributes. The ratio-cum-product estimator in two-phase sampling for full information case has been found to be more efficient than existing estimators and also ratio-cum-product estimator in two-phase sampling for both partial and no information case. Finally, ratio-cum-product estimator in two-phase sampling for partial information case has been found to be more efficient than ratio-cum-product estimator in two-phase sampling for no information case.

Keywords

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

1. Introduction

The use of supplementary information is a widely discussed issue 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 variance 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 characteristic, which is po-

sitively 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 variable. 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 which was more efficient than the previous ratio-cum product estimator.

Jhajj, Sharma and Grover [8] 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 are normally used when the auxiliary variables are not available, e.g. amount of milk produced, a particular breed of cow or amount of yield of wheat and a particular variety of wheat. Bahland Tuteja [9] proposed ratio and product type exponential estimators using auxiliary attribute. Rajesh, Pankaj, Nirmala and Florentins [10] 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 kurtosis and point biserial correlation coefficient. The estimator performed better than the usual sample mean and Naikand Gupta [8] estimator. Rajesh, Pankaj, Nirmala and Florentins [10] also used the auxiliary attributes in ratio-product type exponential estimator following the work of Bahland Tuteja [9], the estimator was more efficient compared to mean per unit, ratio and product type exponential estimators as well as Naik and Gupta [11] estimator.

The concept of double sampling was first proposed by Neyman [12] in sampling human populations when the mean of auxiliary variable(s) was unknown. It was later extended to multiphase by Robson [13]. It is advantageous when the gain in precision is substantial as compared to the increase in the cost due to collection of information on the auxiliary variable for large samples. In most survey, the auxiliary information is always available and every form of auxiliary information should be used in developing sampling strategies. Samiuddin and Hanif [14] introduced the following approach of using auxiliary variable.

- 1) Full information case: Information for all auxiliary variables is available
- 2) No information case: Information for all auxiliary variables is not available.
- 3) Partial information case: Information for some auxiliary variable is available for all population units.

We have used these strategies to develop ratio-cum-product estimators using multiple auxiliary attributes for full information, partial and no information cases.

Hanif, Haq and Shahbaz [15] proposed a general family of estimators using multiple auxiliary attribute in single and double phase sampling. The estimators had a smaller MSE compared to that of Jhajj, Sharma and Grover [8]. They also extended their work to ratio, product and regression estimators which were generalization of Naik and Gupta [11] 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 [16], and then extended to ratio and product estimators. In this paper, we have incorporated the multiple auxiliary attributes in ratio-cum-product estimator in two-phase sampling as proposed by Singh [7] and used strategies introduced by Samiuddin and Hanif [14] and also incorporate Arora and Bansi [17] approach in writing down the mean squared error.

2. Preliminaries and Notations

2.1. Notations

Consider a population of N units. Let Y be the variable for which we want to estimate the population mean and P_1, P_2, \dots, P_q are q auxiliary attributes. For two-phase sampling design, let n_1 and n_2 ($n_2 < n_1$) be sample sizes for first and second phase respectively. In defining the attributes we assume complete dichotomy so that

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

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 pos-

sessing attribute τ_j . Let $P_j = \frac{A_j}{N}$ and $p_{j(2)} = \frac{a_j}{n}$ be the corresponding proportion of units possessing a specific attributes τ_j and \bar{y} is the mean of the main variable at second phase. Let $P_{j(1)}$ and $P_{j(2)}$ denote the j^{th} auxiliary attribute form first and second phase samples respectively and y_2 denote the variable of interest from second phase. Let P_j and C_{P_j} denote the population means and coefficient of variation of j^{th} auxiliary attribute respectively and ρ_{Pb_j} denotes the population bi-serial correlation coefficient of Y and P_j . Let $P_{j(1)}$ be proportion of units possessing attribute r_j in first phase sample of size n_1 while $P_{j(2)}$ be proportion of units possessing attribute r_j in second phase sample of size n_2 . Further, let

$$\theta_1 = \left(\frac{1}{n_1} - \frac{1}{N} \right) \quad \theta_2 = \left(\frac{1}{n_2} - \frac{1}{N} \right) \quad (\theta_2 < \theta_1)$$

$$\bar{e}_{y_2} = \bar{y}_2 - \bar{Y}, \quad \bar{e}_{\tau_{j(1)}} = p_{j(1)} - P_j, \quad \bar{e}_{\tau_{j(2)}} = p_{j(2)} - P_j \quad j = (1, 2, \dots, q) \tag{1.0}$$

where e_{y_2} , $e_{\tau_{j(1)}}$ and $e_{\tau_{j(2)}}$ are sampling error which are assumed to be very small. We let

$$v_j = \frac{p_j}{P_j}, \quad v_j - 1 = \frac{p_j - P_j}{P_j} = \frac{1}{P_j} \bar{e}_{r_j}, \quad v_j = 1 + \frac{1}{P_j} \bar{e}_{r_j} \tag{1.1}$$

while $v_{jd} = \frac{p_{j(2)}}{p_{j(1)}}$, $v_{jd} - 1 = \frac{p_{j(2)} - p_{j(1)}}{p_{j(1)}}$

$$v_{jd} - 1 = \frac{\bar{e}_{r_{j(2)}} - \bar{e}_{r_{j(1)}}}{P_j + \bar{e}_{r_{j(1)}}} = \frac{(\bar{e}_{r_{j(2)}} - \bar{e}_{r_{j(1)}})}{P_j} \left(1 + \frac{\bar{e}_{r_{j(1)}}}{P_j} \right)^{-1} \tag{1.2}$$

Here we shall take $(v_{jd} - 1)$ to, term of order $\frac{1}{n}$ as

$$v_{jd} \approx 1 - \frac{(\bar{e}_{r_{1j}} - \bar{e}_{r_{2j}})}{P_j} \quad \text{for } j = (1, 2, \dots, q) \tag{1.3}$$

Let $C_y^2 = \frac{S_{y_2}^2}{\bar{Y}^2}$ and $C_{\tau_j}^2 = \frac{S_{\tau_j}^2}{P_j}$ 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 $\rho_{Pb_j} = \frac{S_{y\tau_j}}{S_y S_{\tau_j}}$. Then for simple random sampling without replacement for both first and second phases we write

by using phase wise operation of expectations as:

$$E(\bar{e}_{y_2})^2 = \theta_2 \bar{Y}^2 C_{y_2}^2, \quad E(\bar{e}_{\tau_{i(2)}}^2) = \theta_2 P_i^2 C_{\tau_j}^2, \quad E(\bar{e}_{y_2} \bar{e}_{\tau_{j(2)}}) = \theta_2 \bar{Y} P_j C_{y_2} C_{\tau_j} \rho_{Pb_j}$$

$$E(\bar{e}_{y(2)} (\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}})) = (\theta_1 - \theta_2) \bar{Y} P_j C_y C_{\tau_i} \rho_{y\tau_i}, \quad E(\bar{e}_{\tau_{i(2)}} \bar{e}_{\tau_{i(2)}}) = \theta_2 P_i P_j C_{\tau_i} C_{\tau_j} \rho_{\tau_i \tau_j} \quad (i \neq j) \tag{1.4}$$

$$E(\bar{e}_{\tau_{j(2)}} (\bar{e}_{\tau_{j(1)}} - \bar{e}_{\tau_{j(2)}})) = (\theta_1 - \theta_2) \bar{X}_j^2 C_{\tau_j}^2, \quad E(\bar{e}_{\tau_{j(2)}} (\bar{e}_{\tau_{j(1)}} - \bar{e}_{\tau_{j(2)}})) = (\theta_1 - \theta_2) P_i P_j C_{\tau_j} C_{\tau_i} \rho_{\tau_i \tau_j}; (i \neq j)$$

$$E(\bar{e}_{\tau_{j(1)}} - \bar{e}_{\tau_{j(2)}})^2 = (\theta_2 - \theta_1) P_j^2 C_{\tau_j}^2, \quad E((\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}) (\bar{e}_{\tau_{j(1)}} - \bar{e}_{\tau_{j(2)}})) = (\theta_2 - \theta_1) P_i P_j C_{\tau_j} C_{\tau_i} \rho_{\tau_j \tau_i}; (i \neq j)$$

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

$$A^{-1} = \frac{1}{|A|} (C^T)_{ij} = \frac{Adj(A)}{|A|} \tag{1.5}$$

$$\frac{|R|_{y\tau_p}}{|R|_{\tau_p}} = (1 - \rho_{y:\tau_p}^2) \quad \text{Arora and Lai [1]} \quad (1.6)$$

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

$$\begin{aligned} |R|_{y\tau_p} & \text{ Determinant of population correlation matrix of attributes } y, \tau_1, \tau_2, \dots, \tau_{q-1} \text{ and } \tau_q. \\ |R|_{y\tau_j} |_{y\tau_p} & \text{ Determinant of } j^{\text{th}} \text{ minor of } |R|_{y\tau_q} \text{ corresponding to the } j^{\text{th}} \text{ element of } \rho_{pb_j}. \\ \rho_{pb_r}^2 & \text{ Denotes the multiple coefficient of determination of } y \text{ on } \tau_1, \tau_2, \dots, \tau_{r-1} \text{ and } \tau_r. \\ \rho_{pb_r}^2 & \text{ Denotes the multiple coefficient of determination of } y \text{ on } y, \tau_1, \tau_2, \dots, \tau_{q-1} \text{ and } \tau_q. \\ |R|_{\tau_r} & \text{ Determinant of population correlation matrix of attributes } \tau_1, \tau_2, \dots, \tau_{r-1} \text{ and } \tau_r. \\ |R|_{y_i\tau_r} & \text{ Determinant of the correlation matrix of } y_i, \tau_1, \tau_2, \dots, \tau_{r-1} \text{ and } \tau_r. \\ |R|_{y_i\tau_q} & \text{ Determinant of the correlation matrix of } y_i, \tau_1, \tau_2, \dots, \tau_{q-1} \text{ and } \tau_q. \\ |R|_{y_i:y_j\tau_r} & \text{ Determinant of the minor corresponding to } \rho_{pb_{ij}} \text{ of the correlation matrix of } y_i, y_j, \tau_1, \tau_2, \dots, \tau_{r-1} \text{ and } \tau_r \text{ (} i \neq j \text{)}. \\ |R|_{y_i:y_j\tau_q} & \text{ Determinant of the minor corresponding to } \rho_{pb_{ij}} \text{ of the correlation matrix } y_i, y_j, \tau_1, \tau_2, \dots, \tau_{q-1} \text{ and } \tau_q \text{ (} i \neq j \text{)}. \end{aligned} \quad (1.7)$$

2.2. Mean per Unit in Two-Phase Sampling

The sample mean \bar{y}_2 using simple random sampling without replacement is given by,

$$\bar{y}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i \quad (1.8)$$

While the variance of \bar{y}_2 is given by,

$$V(\bar{y}_2) = \bar{Y}^2 \theta_2 C_y^2 \quad (1.9)$$

2.3. Ratio and Product Estimator in Two-Phase Sampling Using One Auxiliary Attributes

In order to have an estimate of the population mean \bar{Y} of the study variable y , assuming the knowledge of the population proportion P , Naik and Gupta [8] 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 [8] proposed the following estimators:

$$t_{R(2)} = \bar{y}_2 \left(\frac{P}{p_2} \right)^{\alpha_1} \quad (1.10)$$

$$t_{P(2)} = \bar{y}_2 \left(\frac{p_2}{P} \right)^{\alpha_2} \quad (1.11)$$

The MSE of $t_{R(2)}$ and $t_{P(2)}$ up to the first order of approximation are given respectively by,

$$MSE(t_{P(2)}) = \theta_2 \bar{Y}^2 (C_y^2 + \alpha_1^2 C_P^2 - 2\alpha_1 C_P C_y \rho_{pb}) \quad (1.12)$$

$$MSE(t_{R(2)}) = \theta_2 \bar{Y}^2 (C_y^2 + \alpha_2^2 C_P^2 + 2\alpha_2 C_P C_y \rho_{pb}) \quad (1.13)$$

The optimum value are $\alpha_1 = \frac{C_{y_2} \rho_{pb}}{C_r}$ and $\alpha_2 = -\frac{C_{y_2} \rho_{pb}}{C_r}$ for ratio and product estimator respectively.

2.3. Ratio and Product Estimator Using Multiple Auxiliary Attributes in Two-Phase Sampling

The ratio and product estimators by Hanif, Haq and Shahbaz [5] for single phase sampling using information on multiple auxiliary attributes are given respectively by,

$$t_{R2(2)} = \bar{y}_2 \left(\frac{P_1}{P_{(2)1}} \right)^{\beta_1} \left(\frac{P_2}{P_{(2)2}} \right)^{\beta_2} \dots \left(\frac{P_q}{P_{(2)q}} \right)^{\beta_q} \tag{1.13}$$

$$t_{P2(2)} = \bar{y}_2 \left(\frac{P_{(2)1}}{P_1} \right)^{\alpha_1} \left(\frac{P_{(2)2}}{P_2} \right)^{\alpha_2} \dots \left(\frac{P_{(2)k}}{P_k} \right)^{\alpha_j} \tag{1.14}$$

The MSE of the $t_{R2(2)}$ and $t_{P2(2)}$ up to the first order of approximation are given respectively by,

$$MSE(t_{R2(2)}) = \theta_2 \bar{Y}^2 C_y^2 (1 - \rho_{pb\tau_q}^2) \tag{1.15}$$

$$MSE(t_{P2(2)}) = \theta_2 \bar{Y}^2 C_y^2 (1 - \rho_{pb\tau_q}^2) \tag{1.16}$$

3. Methodology

3.1. Ratio-Cum-Product Estimator Using Multiple Auxiliary Attributes for Full Information Case in Two-Phase Sampling

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

$$t_{RP(3.1)} = \bar{y}_2 \left(\frac{P_1}{P_{(2)1}} \right)^{\alpha_1} \left(\frac{P_2}{P_{(2)2}} \right)^{\alpha_2} \dots \left(\frac{P_k}{P_{(2)k}} \right)^{\alpha_k} \left(\frac{P_{(2)k+1}}{P_{k+1}} \right)^{\beta_{k+1}} \left(\frac{P_{(2)k+2}}{P_{k+2}} \right)^{\beta_{k+2}} \dots \left(\frac{P_{(2)q}}{P_q} \right)^{\beta_q} \tag{3.1}$$

Using (1.0), (1.1) in (3.1) and ignoring the second and higher terms for each expansion of product and after simplification, we write,

$$t_{RP(3.1)} = \bar{y}_2 - \sum_{j=1}^k \alpha_j \frac{\bar{Y}}{P_j} \bar{e}_{(2)j} + \sum_{j=k+1}^{h+k=q} \beta_j \frac{\bar{Y}}{P_j} \bar{e}_{(2)j} \tag{3.2}$$

The mean squared error of ratio-cum-product estimator is:

$$MSE(t_{RP(3.1)}) = E_2 \left(\bar{e}_{y_2} - \sum_{j=1}^k \alpha_j \frac{\bar{Y}}{P_j} \bar{e}_{(2)j} + \sum_{j=k+1}^{h+k=q} \beta_j \frac{\bar{Y}}{P_j} \bar{e}_{(2)j} \right)^2 \tag{3.3}$$

We differentiate Equation (3.3) partially with respect to α_i ($i=1,2,\dots,k$) and β_i ($i=k+1,k+2,\dots,p$) then equate to zero, using (1.5), (1.7) and (1.4), we get

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

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

Using normal equations that are used to find the optimum values of α_j and β_j (3.3) can be written in simplified form as:

$$MSE(t_{RP(3.1)}) = E \left(\bar{e}_{y_2} \left(\bar{e}_{y_2} - \sum_{j=1}^k \alpha_j \bar{Y} \frac{\bar{e}_{\tau(2)j}}{P_j} + \sum_{j=k+1}^{h+k=q} \beta_j \bar{Y} \frac{\bar{e}_{\tau(2)j}}{P_j} \right) \right) \tag{3.6}$$

Using (1.4) in (3.6), we get,

$$MSE(t_{RP(3.1)}) = \theta_2 \bar{Y}^2 \left(C_y^2 - \sum_{j=1}^k \alpha_j P_j C_{y_2} C_{\tau_j} \rho_{Pb_j} + \sum_{j=1}^{k+h=q} \beta_j P_j C_{y_2} C_{\tau_j} \rho_{Pb_j} \right) \tag{3.7}$$

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

$$MSE(t_{RP(3.1)}) = \theta_2 \bar{Y}^2 \left(C_y^2 + \sum_{j=1}^k (-1)^j \frac{C_{y_2}}{P_j C_{\tau_j}} \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} P_j C_y C_{\tau_j} \rho_{Pb_j} + \sum_{j=k+1}^{k+h=q} (-1)^j \frac{C_{y_2}}{P_j C_{\tau_j}} \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} P_j C_y C_{\tau_j} \rho_{Pb_j} \right) \tag{3.8}$$

Or

$$MSE(t_{RP(3.1)}) = \frac{\theta_2 \bar{Y}^2 C_y^2}{|R|_{\tau_q}} \left(|R|_{\tau_q} + \sum_{j=1}^q (-1)^j |R_{y\tau_j}|_{y\tau_q} \rho_{Pb_j} \right) \tag{3.9}$$

Or

$$MSE(t_{RP(3.1)}) = \theta_2 \bar{Y}^2 C_y^2 \frac{|R|_{y\tau_q}}{|R|_{\tau_q}} \tag{3.11}$$

Using (1.6) in (3.11), we get,

$$MSE(t_{RP(3.1)}) = \theta_2 \bar{Y}^2 C_y^2 \left(1 - \rho_{y.Pb\tau_q}^2 \right) \tag{3.12}$$

3.2. Ratio-Cum-Product Estimator Using Multiple Auxiliary Attributes for Partial Information Case in Two-Phase Sampling

In this section, we proposed a ratio-cum-product estimator using multiple auxiliary attributes for partial information case in two-phase sampling using k auxiliary attributes with “ s ” known and “ $k - s$ ” unknown attributes which are positively correlated with study variable Y and $g - k$ auxiliary attributes with “ $g - t$ ” known and “ $g - k + t$ ” unknown attributes which are negatively correlated with study variable (Y). The proposed ratio-cum-product estimator for partial information case is as follows,

$$t_{RP(3.2)} = \bar{y}_2 \left(\frac{P_{1(1)}}{P_{1(2)}} \right)^{\alpha_1} \left(\frac{P_1}{P_{1(2)}} \right)^{\beta_1} \left(\frac{P_{2(1)}}{P_{2(2)}} \right)^{\alpha_2} \left(\frac{P_2}{P_{2(2)}} \right)^{\beta_2} \dots \left(\frac{P_{s(1)}}{P_{r(2)}} \right)^{\alpha_r} \left(\frac{P_s}{P_{r(2)}} \right)^{\beta_r} \cdot \left(\frac{P_{s+1(1)}}{P_{r+1(2)}} \right)^{\alpha_{r+1}} \left(\frac{P_{s+2(1)}}{P_{r+2(2)}} \right)^{\alpha_{r+2}} \dots \left(\frac{P_{k(1)}}{P_{k(2)}} \right)^{\alpha_k} \left(\frac{P_{k+1(2)}}{P_{k+1(1)}} \right)^{\gamma_{k+1}} \left(\frac{P_{k+1(2)}}{P_{k+1}} \right)^{\lambda_{k+1}} \left(\frac{P_{k+2(2)}}{P_{k+2(1)}} \right)^{\gamma_{k+2}} \cdot \left(\frac{P_{k+2(2)}}{P_{k+2}} \right)^{\lambda_{k+2}} \dots \left(\frac{P_{h(2)}}{P_{h(1)}} \right)^{\gamma_h} \left(\frac{P_{h(2)}}{P_h} \right)^{\lambda_h} \left(\frac{P_{h+1(2)}}{P_{h+1(1)}} \right)^{\gamma_{h+1}} \left(\frac{P_{h+2(2)}}{P_{h+2(1)}} \right)^{\gamma_{h+2}} \dots \left(\frac{P_{g(2)}}{P_{g(1)}} \right)^{\gamma_g} \tag{3.13}$$

Using (1.0), (1.1) in (3.13) and ignoring the second and higher terms for each expansion of product and after simplification, we write,

$$t_{RP(3.2)} = \bar{y}_2 + \sum_{j=1}^r \frac{\bar{Y}}{P_j} (\bar{e}_{rj(1)} - \bar{e}_{rj(2)}) - \sum_{j=1}^r \frac{\bar{Y}}{P_j} \bar{e}_{r(2)j} + \sum_{j=r+1}^k \frac{\bar{Y}}{P_j} (\bar{e}_{rj(1)} - \bar{e}_{rj(2)}) - \sum_{j=k+1}^{t+k=h} \frac{\bar{Y}}{P_j} (\bar{e}_{rj(1)} - \bar{e}_{rj(2)}) + \sum_{i=k+1}^{t+k=h} \frac{\bar{Y}}{P_j} \bar{e}_{r(2)j} - \sum_{i=k+h+1}^g \frac{\bar{Y}}{P_j} (\bar{e}_{rj(1)} - \bar{e}_{rj(2)}) \tag{3.14}$$

Mean squared error of $t_{RP(3.2)}$ is given by

$$MSE(t_{RP(3.2)}) = E \left(\bar{e}_{y_2} + \sum_{j=1}^r \frac{\bar{Y}}{P_j} (\bar{e}_{r_j(1)} - \bar{e}_{r_j(2)}) - \sum_{j=1}^r \frac{\bar{Y}}{P_j} \bar{e}_{r(2)j} + \sum_{j=r+1}^k \frac{\bar{Y}}{P_j} (\bar{e}_{r_j(1)} - \bar{e}_{r_j(2)}) \right. \\ \left. - \sum_{j=k+1}^{r+k=h} \frac{\bar{Y}}{P_j} (\bar{e}_{r_j(1)} - \bar{e}_{r_j(2)}) + \sum_{j=k+1}^{r+k=h} \frac{\bar{Y}}{P_j} \bar{e}_{r(2)j} - \sum_{j=h+1}^g \frac{\bar{Y}}{P_j} (\bar{e}_{r_j(1)} - \bar{e}_{r_j(2)}) \right)^2 \quad (3.15)$$

We differentiate Equation (3.15) with respect to

$$\alpha_i (i = 1, 2, \dots, r), \quad \beta_i (i = 1, 2, \dots, r), \quad \alpha_i (i = r+1, r+2, \dots, k), \\ \gamma_i (i = k+1, k+2, \dots, h), \quad \lambda_i (i = k+1, k+2, \dots, h), \quad \gamma_i (i = h+1, h+2, \dots, p)$$

and equate to zero and use (1.4), (1.6) and (1.7). The optimum value is as follows,

$$\alpha_j = (-1)^{j+1} \frac{C_y}{C_{\tau_j}} \left(\frac{|R_{y\tau_i}|_{\tau_g}}{|R|_{\tau_g}} - \frac{|R_{y\tau_i}|_{\tau_m}}{|R|_{\tau_m}} \right) \quad (3.16)$$

$$\beta_i = \frac{C_y}{|R|_{\tau_m}} \frac{1}{C_{\tau_j}} \left((-1)^{j+1} |R_{y\tau_j}|_{\tau_m} \right) \quad (3.17)$$

$$\alpha_j = (-1)^{j+1} \frac{C_{y_2}}{C_{\tau_i}} \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \quad (3.18)$$

$$\gamma_j = -(-1)^{j+1} \frac{C_{y_2}}{C_{\tau_j}} \left(\frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} - \frac{|R_{yx_j}|_{\tau_m}}{|R|_{\tau_m}} \right) \quad (3.19)$$

$$\lambda_j = -\frac{C_y}{|R|_{\tau_m}} \frac{1}{C_{\tau_j}} \left((-1)^{j+1} |R_{y\tau_j}|_{\tau_m} \right) \quad (3.20)$$

$$\gamma_j = -(-1)^{j+1} \frac{C_y}{|R|_{\tau_g}} \frac{1}{C_{\tau_j}} |R_{y\tau_j}|_{\tau_g} \quad (3.21)$$

Using normal equations that are used to find the optimum values of $\alpha_i, \beta_i, \alpha_i, \gamma_i, \lambda_i$ and γ_i (3.15) can be written in simplified form as

$$MSE(t_{RP(3.2)}) \\ = E_1 E_{2/1} \left(\bar{e}_{y_2} + \sum_{j=1}^r \bar{Y} \alpha_j \frac{(\bar{e}_{x(1)j} - \bar{e}_{x(2)j})}{P_j} - \sum_{j=1}^r \bar{Y} \beta_j \bar{e}_{x(2)j} + \sum_{j=r+1}^{s+r=k} \bar{Y} \alpha_j \frac{(\bar{e}_{x(1)j} - \bar{e}_{x(2)j})}{P_j} \right. \\ \left. - \sum_{j=k+1}^{k+t=h} \bar{Y} \gamma_j \frac{(\bar{e}_{x(1)j} - \bar{e}_{x(2)j})}{P_j} + \sum_{j=h+1}^{k+t=h} \bar{Y} \lambda_j \frac{\bar{e}_{x(2)j}}{P_j} - \gamma_j \frac{(\bar{e}_{x(1)j} - \bar{e}_{x(2)j})}{P_j} \right) \quad (3.22)$$

Substituting (1.4), (3.14) to (3.19) in (3.20), we get

$$\begin{aligned}
 MSE(t_{RP(3.2)}) = & \bar{Y}^2 C_y^2 \left(\theta_2 + (\theta_1 - \theta_2) \sum_{j=1}^r (-1)^{j+1} \left(\frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} - \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \right) \rho_{Pb_j} - \theta_2 \sum_{j=1}^r (-1)^{j+1} \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \rho_{Pb_j} \right. \\
 & + (\theta_1 - \theta_2) \sum_{j=r+1}^{r+s=k} (-1)^{j+1} \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \rho_{Pb_j} + (\theta_1 - \theta_2) \sum_{j=k+1}^{t+k=h} (-1)^{j+1} \left(\frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} - \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \right) \rho_{Pb_j} \\
 & \left. - \theta_2 \sum_{j=k+1}^{t+k=h} (-1)^{j+1} \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \rho_{Pb_j} + (\theta_1 - \theta_2) \sum_{j=h+1}^p (-1)^{j+1} \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \gamma_j \rho_{Pb_j} \right) \quad (3.23)
 \end{aligned}$$

Or

$$\begin{aligned}
 = & \bar{Y}^2 C_y^2 (\theta_2 - \theta_1) \left(1 + \left(\sum_{j=1}^r (-1)^j \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \rho_{Pb_j} + \sum_{j=r+1}^{r+s=k} (-1)^j \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \rho_{Pb_j} \right. \right. \\
 & \left. \left. + \sum_{j=k+1}^{t+k=h} (-1)^j \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \rho_{Pb_j} + \sum_{j=h+1}^p (-1)^j \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \gamma_j \rho_{Pb_j} \right) \right) \quad (3.24) \\
 & + \theta_1 \left(1 + \left(\sum_{j=1}^r (-1)^j \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \rho_{Pb_j} + \sum_{j=k+1}^{t+k=h} (-1)^j \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \rho_{Pb_j} \right) \right)
 \end{aligned}$$

Or

$$\begin{aligned}
 = & \bar{Y}^2 C_y^2 \left(\frac{(\theta_2 - \theta_1)}{|R|_{\tau_g}} \left(|R|_{\tau_g} + \sum_{j=1}^p (-1)^j \frac{|R_{y\tau_j}|_{\tau_g}}{|R|_{\tau_g}} \rho_{Pb_j} \right) \right) \quad (3.25) \\
 & + \frac{\theta_1}{|R|_{\tau_m}} \left(|R|_{\tau_g} + \left(\sum_{j=1}^r (-1)^j \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \rho_{Pb_j} + \sum_{j=k+1}^{t+k=h} (-1)^j \frac{|R_{y\tau_j}|_{\tau_m}}{|R|_{\tau_m}} \rho_{Pb_j} \right) \right)
 \end{aligned}$$

Or

$$MSE(t_{RP(3.2)}) = \bar{Y}^2 C_y^2 \left((\theta_2 - \theta_1) \frac{|R|_{y\tau_g}}{|R|_{\tau_g}} + \theta_1 \frac{|R|_{y\tau_m}}{|R|_{\tau_m}} \right) \quad (3.26)$$

Using (1.6) in (3.26), we get,

$$MSE(t_{RP(3.2)}) = \bar{Y}^2 C_y^2 \left((\theta_2 - \theta_1) (1 - \rho_{Pb\tau_g}^2) + \theta_1 (1 - \rho_{Pb\tau_m}^2) \right) \quad (3.27)$$

Or

$$MSE(t_{RP(3.2)}) = \bar{Y}^2 C_y^2 \left(\theta_2 (1 - \rho_{Pb\tau_g}^2) + \theta_1 (\rho_{Pb\tau_g}^2 - \rho_{Pb\tau_m}^2) \right) \quad (3.28)$$

3.3. Ratio-Cum-Product Estimator in Two-Phase Sampling (No Information Case)

If we estimate a study variable when information on all auxiliary variables is unavailable from population, it is

utilized in the form of their means. By taking the advantage of ratio-cum-product technique for two-phase sampling, a generalized estimator for estimating population mean of study variable Y with the use of multi auxiliary variables are suggested as:

$$t_{RP(3.3)} = \bar{y}_2 \left(\frac{P_{(1)1}}{P_{(2)1}} \right)^{\alpha_1} \left(\frac{P_{(1)2}}{P_{(2)2}} \right)^{\alpha_2} \dots \left(\frac{P_{(1)k}}{P_{(2)k}} \right)^{\alpha_s} \left(\frac{P_{(2)k+1}}{P_{(1)k+1}} \right)^{\gamma_{k+1}} \left(\frac{P_{(2)k+2}}{P_{(1)k+2}} \right)^{\gamma_{k+2}} \dots \left(\frac{P_{(2)q}}{P_{(1)q}} \right)^{\gamma_q} \tag{3.29}$$

Using (1.0), (1.1) in (3.29) and ignoring the second and higher terms for each expansion of product and after simplification, we write,

$$t_{RP(3.3)} = \bar{y}_2 + \sum_{j=1}^k \bar{Y} \alpha_j \frac{\bar{e}_{x_{j(1)}} - \bar{e}_{x_{j(2)}}}{P_j} - \sum_{j=1}^{k+h=q} \bar{Y} \gamma_j \frac{\bar{e}_{x_{j(1)}} - \bar{e}_{x_{j(2)}}}{P_j} \tag{3.30}$$

Mean squared error of $t_{RP(3.3)}$ estimator is given by

$$MSE(t_{RP(3.3)}) = E_2 \left(\bar{e}_{y_2} + \sum_{j=1}^k \bar{Y} \alpha_j \frac{\bar{e}_{x_{j(1)}} - \bar{e}_{x_{j(2)}}}{P_j} - \sum_{j=1}^{k+h=q} \bar{Y} \gamma_j \frac{\bar{e}_{x_{j(1)}} - \bar{e}_{x_{j(2)}}}{P_j} \right)^2 \tag{3.31}$$

We differentiate Equation (3.31) partially with respect to α_i ($i=1,2,\dots,k$) and β_i ($i=k+1,k+2,\dots,p$) then equate to zero, using (1.5), (1.7) and (1.4), we get:

$$\alpha_j = (-1)^{j+1} \frac{C_y}{C_{\tau_j}} \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} \tag{3.32}$$

$$\beta_j = -(-1)^{j+1} \frac{C_y}{C_{\tau_j}} \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} \tag{3.33}$$

Using normal equations that are used to find the optimum values of α_j and β_j (3.31) can be written in simplified form as:

$$MSE(t_{RP(3.3)}) = E_2 \left(\bar{e}_{y_2} \left(\bar{e}_{y_2} + \sum_{j=1}^k \bar{Y} \alpha_j \frac{\bar{e}_{x_{j(1)}} - \bar{e}_{x_{j(2)}}}{P_j} - \sum_{j=1}^{k+h=q} \bar{Y} \gamma_j \frac{\bar{e}_{x_{j(1)}} - \bar{e}_{x_{j(2)}}}{P_j} \right) \right) \tag{3.34}$$

Using (1.4) in (3.34), we get,

$$= \bar{Y}^2 C_y^2 \left(\theta_2 + (\theta_1 - \theta_2) \sum_{j=1}^k \alpha_j P_j C_y C_{\tau_j} \rho_{Pb_j} - (\theta_1 - \theta_2) \sum_{j=k+1}^{k+h=q} \beta_j P_j C_y C_{\tau_j} \rho_{Pb_j} \right) \tag{3.35}$$

Substituting equation (3.32) and (3.33) in (3.35), we get

$$MSE(t_{RP(3.3)}) = \theta_2 \bar{Y}^2 C_y^2 \left(\theta_2 + (\theta_2 - \theta_1) \sum_{j=1}^k (-1)^j \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} \rho_{Pb_j} + (\theta_2 - \theta_1) \sum_{j=1}^{k+h=q} (-1)^j \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} \rho_{Pb_j} \right) \tag{3.36}$$

$$MSE(t_{RP(3.3)}) = \bar{Y}^2 C_y^2 \left(\frac{(\theta_2 - \theta_1)}{|R|_{\tau_q}} \left(|R|_{\tau_q} + \sum_{j=1}^p (-1)^j \frac{|R_{y\tau_j}|_{\tau_q}}{|R|_{\tau_q}} \rho_{Pb_j} + \theta_1 \right) \right) \tag{3.37}$$

Or

$$= \bar{Y}^2 C_y^2 \left((\theta_2 - \theta_1) \frac{|R|_{y\tau_q}}{|R|_{\tau_q}} + \theta_1 \right) \tag{3.38}$$

Using (1.6) in (3.38), we get,

$$MSE(t_{RP(3.3)}) = \bar{Y}^2 C_y^2 (\theta_2 (1 - \rho_{pb\tau_q}^2) + \theta_1 \rho_{pb\tau_q}^2) \tag{3.39}$$

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

These ratio-cum-product estimators using multiple auxiliary attributes in two-phase sampling are biased. However, these biases are negligible for moderate and large samples.

It is easily shown that the ratio-cum-product estimators are consistent estimators using multiple auxiliary variables since they are linear combinations of consistent estimators it follows that they are also consistent.

4. Simulation, Result and Discussion

In this section, we carried out some data simulation experiments to compare the performance of ratio-cum-product estimator in two-phase sampling using multiple auxiliary attributes with existing estimators of finite population that uses one or multiple auxiliary attributes namely mean per unit, ratio and product estimator using one auxiliary attributes and ratio and product estimators using two auxiliary attributes. The simulated data for the empirical study include a study variable and auxiliary attributes that are normally distributed with the following variables

$N = 300, n = 45, \text{Mean} = 45, \text{standard deviation} = 5$

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

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

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

Table 1 shows the percent relative efficiency of existing and proposed estimator with respect to mean per unit estimator for two-phase sampling. It is observed that ratio and product estimators using one auxiliary attribute are more efficient than mean per unit in the two-phase sampling. Again, ratio and product estimator using multiple auxiliary attributes are more efficient than mean per unit and ratio and product estimator using one auxiliary attribute in the two-phase sampling. Finally, Ratio-cum-product estimator in the two-phase sampling for full information case using multiple auxiliary attributes is the most efficient of the five estimators since it has the highest percent relative efficiency.

Table 2 shows percent relative efficiency of ratio-cum-product estimators with respect to mean per unit estimator in two-phase sampling. It is observed that the ratio-cum-product estimators are more efficient than mean per unit in the second phase sampling.

Finally, **Table 3** compares the efficiency of full information case and partial case to no information case and full to partial information case. It is observed that the full information case and partial information case are more efficient than no information case because they have higher Percent Relative Efficiency than no information case. In addition, the full information case is more efficient than the partial information case because it has a higher Percent Relative Efficiency than partial information case.

5. Conclusion

Ratio-cum-product estimator using multiple auxiliary attributes in full information case in two-phase sampling is recommended to estimate population mean as it outperforms other estimator in two-phase sampling. If some auxiliary attributes are known, the ratio-cum-product estimator using multiple auxiliary attributes in partial infor-

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

Estimator	Relative percent efficiency with respect to mean per unit
\bar{y}_2	100
$t_{R(2)}$	142
$t_{P(2)}$	116
$t_{R2(2)}$	193
$t_{P2(2)}$	179
$t_{RP(3,1)}$	210

Table 2. Relative efficiency of existing and proposed estimators with respect to mean per unit estimator for two-phase sampling.

Estimator	Relative efficiency of existing and proposed estimators with respect to mean per unit estimator for two-phase sampling
\bar{y}_2	100
$t_{RP(3,3)}$	135
$t_{RP(3,2)}$	162
$t_{RP(3,1)}$	211

Table 3. Comparisons of full, partial and no information cases for proposed ratio-cum-product estimator using multiple auxiliary variables.

Population	Percent relative efficiency of full and partial to no information		Percent relative efficiency of full to partial in formation case		
Estimator	$t_{RP(3,2)}$	$t_{RP(3,1)}$	$t_{RP(3,0)}$	$t_{RP(3,1)}$	$t_{RP(3,0)}$
Relative percent efficiency	100	130	168	100	135

mation case should be used but if all the auxiliary attributes are unknown, and ratio-cum-product estimator using multiple auxiliary attributes in no information case should be used to estimate finite population mean. This is clear from **Table 3**.

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

- 159-163. <http://dx.doi.org/10.1080/02522667.1991.10699058>
- [10] Rajesh, S., Pankaj, C., Nirmala, S. and Florentins, S. (2007) Ratio-Product Type Exponential Estimator for Estimating Finite Population Mean Using Information on Auxiliary Attributes. Renaissance High Press, USA.
- [11] Jhaji, 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.
- [12] Neyman, J. (1938) Contribution to the Theory of Sampling Human Populations. *Journal of the American Statistical Association*, **33**, 101-116. <http://dx.doi.org/10.1080/01621459.1938.10503378>
- [13] Robson, D.S. (1952) Multiple Sampling of Attributes. *Journal of the American Statistical Association*, **47**, 203-215. <http://dx.doi.org/10.1080/01621459.1952.10501164>
- [14] 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.
- [15] 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.
- [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 Bansi, Lal. (1989) New Mathematical Statistics. Satya Prakashan, New Delhi.