

Mixture Regression-Cum-Ratio Estimator Using Multi-Auxiliary Variables and Attributes in Single-Phase Sampling

Teresio Mutembei, John Kung'u, Christopher Ouma

Department of Statistics and Actuarial Science, Kenyatta University, Nairobi, Kenya
Email: mutembeiteresio@gmail.com, johnkungu08@yahoo.com

Received 25 May 2014; revised 30 June 2014; accepted 13 July 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 a class of mixture regression-cum-ratio estimator for estimating population mean by using information on multiple auxiliary variables and attributes simultaneously in single-phase sampling and analyzed the properties of the estimator. An empirical was carried out to compare the performance of the proposed estimator with the existing estimators of finite population mean using simulated population. It was found that the mixture regression-cum-ratio estimator was more efficient than ratio and regression estimators using one auxiliary variable and attribute, ratio and regression estimators using multiple auxiliary variables and attributes and regression-cum-ratio estimators using multiple auxiliary variables and attributes in single-phase sampling for finite population.

Keywords

Regression-Cum-Ratio Estimator, Multiple Auxiliary Variables and Attributes, Single-Phase Sampling

1. Introduction

The work of Neyman [1] may be referred to as the initial works where auxiliary information has been used. Watson [2] used the regression estimator of leaf area on leaf weight to estimate the average area of the leaves on a plant. Cochran [3] used auxiliary information in single-phase sampling to develop the ratio estimator for estimation of population mean. In the ratio estimator, the study variable and the auxiliary variable had a high positive correlation and the regression line was passing through the origin. Hansen and Hurwitz [4] also suggested the use of auxiliary information in selecting the sample with varying probabilities.

Olkin [5] was the first person to use information on more than one supplementary character, which is positively correlated with the variable under study, using a linear combination of ratio estimator based on each auxiliary variable. Shukla [6] proposed that regression estimator using multiple auxiliary was more efficient than regression estimator using single auxiliary variable. Raj [7] suggested a method of using multi-auxiliary information in sample survey. Singh [8] proposed a ratio-cum-product estimator and its multi-variable expression which were more efficient than ratio, product and mean per unit estimators.

Jhajj, Sharma and Grover [9] proposed a family of estimators using information on auxiliary attribute. They used known information of population proportion possessing an attribute that is highly correlated with study variable Y . The attribute is normally used when the auxiliary variable is not available e.g. an amount of milk produced and a particular breed of cow or an amount of yield of wheat and a particular variety of wheat. 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 bi-serial correlation coefficient. The estimator performed better than the usual sample mean and Naik and Gupta [11] estimator. Rajesh, Pankaj, Nirmala and Florentins [10] also used the auxiliary attribute in regression, product and ratio type exponential estimator following the work of Bahl and Tuteja [12].

Hanif, Haq and Shahbaz [13] [14] proposed a general family of estimators using multiple auxiliary attribute in single and double phase sampling. The estimator had a smaller MSE compared to that of Jhajj, Sharma and Grover [9]. They also extended their work to ratio estimator which was generalization of Naik and Gupta [11] estimator in single and double phase sampling with full information, partial information and no information.

The concept of double sampling was first proposed by Neyman [1] in sampling human populations when the mean of auxiliary variable was unknown. It was later extended to multiphase by Robson [15]. In most surveys the auxiliary information is always available and every form of auxiliary information should be used in developing sampling strategies. Samiuddin and Hanif [16] introduced the following approach 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.

Ahmad [17] generalized multivariate ratio and regression estimators for multi-phase sampling. Zahoor, Muhammad and Munir [18] suggested a generalized regression-cum-ratio estimator for two-phase sampling using multiple auxiliary variables in full, partial and no information case. Kung'u and Odongo [19] and [20] proposed ratio-cum-product estimators using multiple auxiliary attributes in single phase sampling and two-phase sampling using multiple auxiliary attributes in full, partial and no information case. Moeen, Shahbaz and Hanif [21] proposed a class of mixture ratio and regression estimators for single-phase sampling for estimating population mean by using information on auxiliary variables and attributes simultaneously.

In this paper, we will incorporate both multiple auxiliary variables and attributes in regression-cum-ratio estimator to form mixture regression-cum-ratio estimator in single-phase sampling and also incorporate Arora and Bansi [22] approach in writing down the mean squared error.

2. Preliminaries

2.1. Notation and Assumption

The following notation will be used in this project. Consider a population of N units. Let Y be the study variable for which we want to estimate the population mean and X_1, X_2, \dots, X_k are k auxiliary variables and $\tau_1, \tau_2, \dots, \tau_t$ are t auxiliary attributes. For single-phase sampling design let n be sample sizes for first phase while x_j and r_j denote the j^{th} auxiliary variables and auxiliary attribute, and y denote the variable of interest from first phase. Let

$$\theta = \left(\frac{1}{n} - \frac{1}{N} \right) \text{ and } y = Y + e_y, x_j = X_j + e_{x_j} \text{ and } p_j = P_j + e_{\tau_j} \quad j = (1, 2, \dots, p) \quad (1.0)$$

where e_y , e_{x_j} and e_{τ_j} are sampling error and are very small. We assume that

$$E(\bar{e}_y) = E(\bar{e}_{x_j}) = E(\bar{e}_{\tau_j}) = 0 \quad (1.1)$$

In defining the attributes we assume complete dichotomy so that;

$$\tau_{ij} = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ unit of population possess } j^{\text{th}} \text{ auxiliary attribute} \\ 0, & \text{otherwise} \end{cases} \quad (1.2)$$

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 $P_j = \frac{A_j}{N}$ and $p_j = \frac{a_j}{n}$ 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.

The coefficient of variations are given by $C_y^2 = \frac{S_y^2}{\bar{Y}^2}$, $C_{x_1}^2 = \frac{S_{x_1}^2}{X_1^2}$, $C_{\tau_j}^2 = \frac{S_{\tau_j}^2}{P_j^2}$ while $\rho_{yx_j} = \frac{S_{yx_j}}{S_y S_{x_j}}$ is the correlation coefficient between study variable and auxiliary variables and $\rho_{pb_j} = \frac{S_{y\tau_j}}{S_y S_{\tau_j}}$ is the bi-serial correlation coefficient between study variable and auxiliary variables. Then for simple random sampling without replacement for both first and second phases we write by using phase wise operation of expectations as:

$$\begin{aligned} E(\bar{e}_y)^2 &= \theta \bar{Y}^2 C_y^2 & E(\bar{e}_y \bar{e}_{x_j})^2 &= \theta \bar{Y} \bar{X}_j C_y C_{x_j} \rho_{yx_j} \\ E(\bar{e}_{x_i} \bar{e}_{x_j}) &= \theta_1 \bar{X}_i \bar{X}_j C_{x_j} C_{x_i} \rho_{x_i x_j} \quad (i \neq j) & E(\bar{e}_{\tau_j})^2 &= \theta_1 P_j^2 C_{\tau_j}^2 \\ E(\bar{e}_{y_2} \bar{e}_{\tau_j})^2 &= \theta \bar{Y} P_j C_{y_2} C_{\tau_j} \rho_{pb_j} & E(\bar{e}_{\tau_i} \bar{e}_{\tau_j}) &= \theta P_i P_j C_{\tau_i} C_{\tau_j} \rho_{\tau_i \tau_j} \quad (i \neq j) \end{aligned} \quad (1.3)$$

$$A^{-1} = \frac{1}{|A|} (C^T)_{ij} = \frac{Adj(A)}{|A|} \quad (1.4)$$

$$\frac{|R|_{y x_p}}{|R|_{x_p}} = (1 - \rho_{y x_p}^2) \quad \text{Arora and Lai [22]} \quad (1.5)$$

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

$|R|_{y x_p}$ Determinant of population correlation matrix of variables $y, x_1, x_2, \dots, x_{p-1}$ and x_p .

$|R_{y x_i}|_{y x_p}$ Determinant of i^{th} minor of $|R|_{y x_p}$ corresponding to the i^{th} element of $\rho_{y x_i}$.

$\rho_{y x_r}^2$ Denotes the multiple coefficient of determination of y on x_1, x_2, \dots, x_{r-1} and x_r .

$\rho_{y x_p}^2$ Denotes the multiple coefficient of determination of y on $y, x_1, x_2, \dots, x_{p-1}$ and x_p .

$|R|_{x_r}$ Determinant of population correlation matrix of variables x_1, x_2, \dots, x_{r-1} and x_r .

$|R|_{x_p}$ Determinant of population correlation matrix of variables x_1, x_2, \dots, x_{p-1} and x_p .

$|R|_{y_i x_r}$ Determinant of the correlation matrix of $y_i, x_1, x_2, \dots, x_{r-1}$ and x_r .

$|R|_{y_i x_p}$ Determinant of the correlation matrix of $y_i, x_1, x_2, \dots, x_{p-1}$ and x_p .

$|R|_{y_i y_j x_r}$ Determinant of the minor corresponding to $\rho_{y_i y_j}$ of the correlation matrix of $y_i, y_j, x_1, x_2, \dots, x_{r-1}$

and $x_r(i \neq j)$.

$|R|_{y_i y_j x_p}$ Determinant of the minor corresponding to $\rho_{y_i y_j}$ of the correlation matrix of $y_i, y_j, x_1, x_2, \dots, x_{p-1}$

and $x_p(i \neq j)$.

2.2. Mean per Unit in Single-Phase Sampling

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

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j \tag{1.6}$$

While its variance is given,

$$V(\bar{y}) = \theta \bar{Y}^2 C_y^2 \tag{1.7}$$

2.3. Ratio and Regression Estimator Using Auxiliary Variable

Let $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ be the unbiased estimator of population means \bar{Y} and \bar{X} respectively.

Then the classical ratio estimator by Cochran [3] and regression estimator by Watson [2] are defined respectively by,

$$t_{R1} = \bar{y} \left(\frac{\bar{X}}{\bar{x}} \right)^{\alpha_1} \quad (\bar{x} \neq 0) \tag{1.8}$$

$$t_{Re1} = \bar{y} + \alpha_2 (\bar{X} - \bar{x}) \tag{1.9}$$

where \bar{X} , the population mean of the auxiliary variable X is known where $\alpha_1 = \frac{C_y^2 \rho_{yx}}{C_x}$ and $\alpha_2 = \frac{\bar{Y} C_y \rho_{yx}}{C_{x1} \bar{X}_1}$

are optimum values of ratio and regression estimator respectively.

The minimum MSE of t_{R1} and t_{Re1} up to the first order of approximation are,

$$MSE(t_{R1}) = \theta \bar{Y}^2 (C_y^2 + \alpha_1^2 C_x^2 - 2\alpha_1 C_x C_y \rho_{yx}) \tag{1.9}$$

$$MSE(t_{Re1}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2) \tag{1.10}$$

2.4. Ratio and Regression Estimator Using Multiple Auxiliary Variables

In case of multiple auxiliary variables, the ratio and regression estimators Ahmad [17] are given by,

$$t_{R2} = \bar{y} \prod_{j=1}^k \left(\frac{\bar{X}_j}{\bar{x}_j} \right)^{\alpha_j} \tag{1.11}$$

$$t_{Re2} = \bar{y} + \sum_{j=1}^k \beta_j (\bar{X}_j - \bar{x}_j) \tag{1.12}$$

where $\alpha_j = \frac{C_y}{C_{xj}} (-1)^{j+1} \frac{|R_{yx_j}|_{y.x_j}}{|R|_{(x,\tau)_t}}$ are the optimum values $\beta_j = \frac{\bar{Y} C_y}{\bar{X}_j C_{xj}} (-1)^{j+1} \frac{|R_{yx_j}|_{y.x_j}}{|R|_{(x,\tau)_t}}$. The minimum mean

squared error of t_{R2} and t_{Re2} up to the first order of approximation are,

$$MSE(t_{R2}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y\bar{x}_k}^2) \tag{1.13}$$

$$MSE(t_{Re2}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y\bar{x}_k}^2) \tag{1.14}$$

2.5. Regression-Cum-Ratio Estimator Using Multiple Auxiliary Variables

The regression-cum-ratio estimator by Zahoor [18] using multiple auxiliary variables is given by,

$$t_{RRe1} = \left(\bar{y} + \beta_j \sum_{j=1}^r (\bar{X}_j - \bar{x}_j) \right) \prod_{j=r+1}^{r+s=k} \left(\frac{\bar{X}_j}{\bar{x}_j} \right)^{\alpha_j} \tag{1.15}$$

The $\alpha_j = (-1)^{j+1} \frac{C_y}{C_{x_j}} \frac{|R_{yx_j}|_{y:\xi_q}}{|R|_{\xi_q}}$ and $\beta_j = \frac{\bar{Y}C_y}{\bar{X}_j C_{x_j}} (-1)^{j+1} \frac{|R_{yx_j}|_{y:\xi_q}}{|R|_{\xi_q}}$ are the optimum values. The minimum

MSE of t_{RRe1} up to the first order of approximation are,

$$\text{MSE}(t_{RRe1}) = \theta \bar{Y}^2 C_y^2 \left(1 - \rho_{y:(\xi_r, \xi_s)}^2\right) \quad (1.16)$$

2.6. Ratio and Regression Estimator Using Auxiliary Attribute

In order to have an estimate of the population mean \bar{Y} the study variable y , assuming the knowledge of the population proportion P , Naik and Gupta [11] defined ratio and regression estimators of population when the prior information of population proportion of units, possessing the same attribute is variable. Using (1.8) and (1.9) Naik and Gupta [11] proposed following estimators:

$$t_{R11} = \bar{y} \left(\frac{P}{p}\right)^{\alpha_1} \quad (1.17)$$

$$t_{Re11} = \bar{y} + \beta_1 (P_1 - p_1) \quad (1.18)$$

The minimum MSE of t_R and t_{Re} up to the first order of approximation are

$$\text{MSE}(t_{R11}) = f_1 \bar{Y}^2 \left(C_y^2 + \alpha_1^2 C_R^2 - 2\alpha_1 C_R C_y \rho_{pb}\right) \quad (1.19)$$

$$\text{MSE}(t_{Re11}) = \bar{Y}^2 C_y^2 + \beta_1^2 P_1^2 C_{r_1}^2 - 2\beta_1 C_{r_1} P_1 \bar{Y} C_y \rho_{pb_1} \quad (1.20)$$

where $\alpha_1 = \frac{C_y \rho_{pb}}{C_r}$ and $\beta_1 = \frac{\bar{Y} C_y \rho_{pb_1}}{C_{r_1} P_1}$ are optimum values of ratio and regression estimator respectively.

2.7. Ratio and Regression Estimator Using Multiple Auxiliary Attributes.

The ratio and regression estimators by Hanif, Haq and Shahbaz [14] for single-phase sampling using information on multiple auxiliary attributes are given by,

$$t_{R22} = \bar{y} \left(\frac{P_1}{p_1}\right)^{\alpha_1} \left(\frac{P_2}{p_2}\right)^{\alpha_2} \cdots \left(\frac{P_q}{p_q}\right)^{\alpha_q} \quad (1.21)$$

$$t_{Re22} = \bar{y} + \beta_1 (P_1 - p_1) + \beta_2 (P_2 - p_2) + \cdots + \beta_k (P_k - p_k) \quad (1.22)$$

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

$$\text{MSE}(t_{R22}) = \theta \bar{Y}^2 \left(C_y^2 + \sum_{j=1}^q C_{r_j}^2 - 2 \sum_{j=1}^q C_y C_{r_j} \rho_{pb_j} + 2 \sum_{j \neq \psi=1}^q C_y C_{r_j} \rho_{j\psi} \right) \quad (1.23)$$

$$\text{MSE}(t_{Re22}) = \theta_1 \left(\bar{Y}^2 C_y^2 + \sum_{j=1}^k \alpha_j^2 C_j^2 P_j^2 - 2 \sum_{j=1}^k \alpha_j C_{r_j} C_y P_j \bar{Y} \rho_{pb_j} + 2 \sum_{j \neq \varphi=1}^k \alpha_j \alpha_\varphi C_{r_j} C_{r_\varphi} P_j P_\varphi \theta_{j\varphi} \right) \quad (1.24)$$

2.8. Regression-Cum-Ratio Estimator Using Multiple Auxiliary Attributes

The regression-cum-ratio estimator using multiple auxiliary attributes is given by,

$$t_{RRe2} = \left(\bar{y} + \beta_j \sum_{j=1}^r (P_j - p_j) \right) \prod_{j=r+1}^{r+s=k} \left(\frac{P_j}{p_j} \right)^{\alpha_j} \quad (1.25)$$

The $\alpha_j = (-1)^{j+1} \frac{C_y}{C_{x_j}} \frac{|R_{yx_j}|_{y:\xi_k}}{|R|_{\xi_k}}$ and $\beta_j = \frac{\bar{Y}C_y}{\bar{X}_j C_{x_j}} (-1)^{j+1} \frac{|R_{yx_j}|_{y:\xi_k}}{|R|_{\xi_k}}$ are the optimum values to the first order

of approximation. The minimum MSE of t_{RRe2} up to the first order of approximation are,

$$MSE(t_{RRe2}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y(x_r, \bar{x}_s)}^2) \tag{1.26}$$

2.9. Mixture Ratio and Regression Using Multiple Auxiliary Variables and Attributes

The mixture ratio estimator based on multiple auxiliary variables and attributes by Moeen, Shahbaz and HanIf [21] is given by:

$$t_{MR(3)} = \bar{y} \prod_{j=1}^k \left(\frac{\bar{X}_j}{\bar{x}_j} \right)^{\beta_j} \prod_{j=k+1}^q \left(\frac{P_j}{p_j} \right)^{\alpha_j} \tag{1.27}$$

$$t_{MRe(3)} = \bar{y} + \sum_{j=1}^k \beta_j (\bar{X}_j - \bar{x}_j) + \sum_{j=1}^q \alpha_j (P_j - p_j) \tag{1.28}$$

The minimum MSE of t_{RM} and t_{ReM} up to the first order of approximation are

$$MSE(t_{RM(3)}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y(x_k, \bar{x}_s)}^2) \tag{1.29}$$

$$MSE(t_{MRe(4)}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y(x_k, \bar{x}_s)}^2) \tag{1.30}$$

In general these estimators have a bias of order n^{-1} . Since the standard error of the estimates is of order $\frac{1}{\sqrt{n}}$,

the quantity bias/s.e is of order $\frac{1}{\sqrt{n}}$ and becomes negligible as n becomes large. In practice, this quantity is usually unimportant in samples of moderate and large sizes.

In this paper, we have combined mixture ratio and mixture regression estimator to form mixture regression-cum-ratio estimator under single-phase sampling and studied the properties of the proposed estimator.

3. Methodology

3.1. Mixture Regression-Cum-Ratio Estimator Using Multi-Auxiliary Variables and Attributes in Single-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 mixture regression-cum-ratio estimator technique for single-phase sampling, a generalized estimator for estimating population mean of study variable Y with the use of multi auxiliary variables and attributes is suggested as:

$$t_{MRR} = (\bar{y} + \alpha_1 (\bar{X}_1 - \bar{x}_1) + \alpha_2 (\bar{X}_2 - \bar{x}_2) + \dots + \alpha_k (\bar{X}_k - \bar{x}_k) + \beta_{k+1} (P_{k+1} - p_{k+1}) + \beta_{k+2} (P_{k+2} - p_{k+2}) + \dots + \beta_q (P_q - p_q)) \left(\frac{\bar{X}_{q+1}}{\bar{x}_{q+1}} \right)^{\lambda_{q+1}} \left(\frac{\bar{X}_{q+2}}{\bar{x}_{q+2}} \right)^{\lambda_{q+2}} \dots \left(\frac{\bar{X}_r}{\bar{x}_r} \right)^{\lambda_r} \left(\frac{P_{r+1}}{p_{r+1}} \right)^{\gamma_{r+1}} \left(\frac{P_{r+2}}{p_{r+2}} \right)^{\gamma_{r+2}} \dots \left(\frac{P_t}{p_t} \right)^{\gamma_t} \tag{2.0}$$

Using (1.0) in (2.0), we get,

$$t_{MRR} = \left(\bar{y} - \sum_{j=1}^k \alpha_j \bar{e}_{x_j} - \sum_{j=k+1}^q \beta_j \bar{e}_{r_j} \right) \prod_{j=q+1}^r \left(1 + \frac{\bar{e}_{x_j}}{\bar{X}_j} \right)^{-\lambda_j} \prod_{j=r+1}^t \left(1 + \frac{\bar{e}_{r_j}}{P_j} \right)^{-\gamma_j} \tag{2.1}$$

Ignoring the second and higher terms for each expansion of product and after simplification, we write,

$$t_{MRR} = \bar{e}_y + \bar{Y} - \sum_{j=1}^k \alpha_j \bar{e}_{x_j} - \sum_{j=k+1}^q \beta_j \bar{e}_{r_j} - \sum_{j=1}^r \lambda_j \bar{Y} \frac{\bar{e}_{x_j}}{\bar{X}_j} - \sum_{j=r+1}^t \gamma_j \bar{Y} \frac{\bar{e}_{r_j}}{P_j} \tag{2.2}$$

The mean squared error of t_{MRR} is given by,

$$MSE(t_{MRR}) = E(t_{MRR} - \bar{Y})^2 = E \left(\bar{e}_y - \sum_{j=1}^k \alpha_j \bar{e}_{x_j} - \sum_{j=k+1}^q \beta_j \bar{e}_{r_j} - \sum_{j=1}^r \lambda_j \bar{Y} \frac{\bar{e}_{x_j}}{\bar{X}_j} - \sum_{j=r+1}^t \gamma_j \bar{Y} \frac{\bar{e}_{r_j}}{P_j} \right)^2 \tag{2.3}$$

We differentiate the Equation (2.3) partially with respect to $\alpha_j (j = 1, 2, \dots, k)$, $\beta_j (j = k + 1, k + 2, \dots, q)$, $\lambda_j (j = q + 1, q + 2, \dots, r)$ and $\gamma_j (j = r + 1, r + 2, \dots, t)$ and equate to zero. The optimum value age given by,

$$\alpha_j = \frac{\bar{Y}C_y}{\bar{X}_j C_{x_j}} (-1)^{j+1} \frac{|R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \quad j = 1, 2, \dots, k \quad (2.4)$$

$$\beta_j = \frac{\bar{Y}C_y}{P_j C_{\tau_j}} (-1)^{j+1} \frac{|R_{y\tau_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \quad j = k + 1, k + 2, \dots, q \quad (2.5)$$

$$\lambda_j = \frac{C_y}{C_{x_j}} (-1)^{j+1} \frac{|R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \quad j = q + 1, q + 2, \dots, r \quad (2.6)$$

$$\gamma_j = \frac{C_y}{C_{x_j}} (-1)^{j+1} \frac{|R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \quad j = r + 1, r + 2, \dots, t \quad (2.7)$$

Using normal equation that is used to find the optimum values given (3.8) we can write,

$$\text{MSE}(t_{\text{MRR}}) = E \left(\bar{e}_y \left(\bar{e}_y - \sum_{j=1}^k \alpha_j \bar{e}_{x_j} - \sum_{j=k+1}^q \beta_j \bar{e}_{\tau_j} - \sum_{j=1}^r \lambda_j \bar{Y} \frac{\bar{e}_{x_j}}{\bar{X}_j} - \sum_{j=r+1}^t \gamma_j \bar{Y} \frac{\bar{e}_{\tau_j}}{P_j} \right) \right) \quad (2.8)$$

Or

$$\text{MSE}(t_{\text{MRR}}) = E(\bar{e}_y^2) - \sum_{j=1}^k \alpha_j E(\bar{e}_y \bar{e}_{x_j}) - \sum_{j=k+1}^q \beta_j E(\bar{e}_y \bar{e}_{\tau_j}) - \sum_{j=1}^r \lambda_j \bar{Y} \frac{E(\bar{e}_y \bar{e}_{x_j})}{\bar{X}_j} - \sum_{j=r+1}^t \gamma_j \bar{Y} \frac{E(\bar{e}_y \bar{e}_{\tau_j})}{P_j} \quad (2.9)$$

Taking expectation of (3.49), we get,

$$\begin{aligned} \text{MSE}(t_{\text{MRR}}) = \theta_1 \left(\bar{Y}^2 C_y^2 - \sum_{j=1}^k \alpha_j \bar{Y} \bar{X}_j C_y C_{x_j} \rho_{yx_j} - \sum_{j=k+1}^q \beta_j \bar{Y} P_j C_y C_{\tau_j} \rho_{Pb_j} \right. \\ \left. - \sum_{j=q+1}^r \lambda_j \bar{Y}^2 C_y C_{x_j} \rho_{yx_j} - \sum_{j=r+1}^t \gamma_j \bar{Y}^2 C_y C_{\tau_j} \rho_{Pb_j} \right) \end{aligned} \quad (2.10)$$

Substituting the optimum (2.4) to (2.7) in (2.10) and after simplification we get,

$$\begin{aligned} \text{MSE}(t_{\text{MRR}}) = \theta \bar{Y}^2 C_y^2 \left(1 - \sum_{j=1}^k \frac{(-1)^{j+1} |R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{yx_j} - \sum_{j=k+1}^q \frac{(-1)^{j+1} |R_{y\tau_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{Pb_j} \right. \\ \left. - \sum_{j=q+1}^r \frac{(-1)^{j+1} |R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{yx_j} - \sum_{j=r+1}^t \frac{(-1)^{j+1} |R_{y\tau_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{Pb_j} \right) \end{aligned} \quad (2.11)$$

Or

$$\begin{aligned} \text{MSE}(t_{\text{MRR}}) = \bar{Y}^2 C_y^2 \left(1 + \sum_{j=1}^k \frac{(-1)^j |R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{yx_j} + \sum_{j=k+1}^q \frac{(-1)^j |R_{y\tau_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{Pb_j} \right. \\ \left. + \sum_{j=q+1}^r \frac{(-1)^j |R_{yx_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{yx_j} + \sum_{j=r+1}^t \frac{(-1)^j |R_{y\tau_j}|_{y:(x,\bar{\epsilon})_t}}{|R|_{(x,\tau)_t}} \rho_{Pb_j} \right) \end{aligned} \quad (2.12)$$

Or

$$\begin{aligned} \text{MSE}(t_{\text{MRR}}) \approx & \frac{\theta_1 \bar{Y}^2 C_y^2}{|R|_{(x,\tau)_t}} \left(|R|_{(x,\tau)_t} + \sum_{j=1}^k (-1)^j |R_{yx_j}|_{y,(x,\tau)_t} \rho_{yx_j} + \sum_{j=k+1}^q (-1)^j |R_{y\tau_j}|_{y,(x,\tau)_t} \rho_{Pb_j} \right. \\ & \left. + \sum_{j=q+1}^r (-1)^j |R_{yx_j}|_{y,(x,\tau)_t} \rho_{yx_j} + \sum_{j=r+1}^t (-1)^j |R_{yx_j}|_{y,(x,\tau)_t} \rho_{Pb_j} \right) \end{aligned} \quad (2.13)$$

$$\begin{aligned} \text{MSE}(t_{\text{MRR}}) = & \frac{\theta_1 \bar{Y}^2 C_y^2}{|R|_{(x,\tau)_t}} \left(|R|_{(x,\tau)_t} + (-1)^1 |R_{yx_1}|_{y,(x,\tau)_t} \rho_{yx_1} + (-1)^2 |R_{yx_2}|_{y,(x,\tau)_t} \rho_{yx_2} + \dots + (-1)^k |R_{yx_k}|_{y,(x,\tau)_t} \rho_{yx_k} \right. \\ & + (-1)^{k+1} |R_{y\tau_{k+1}}|_{y,(x,\tau)_t} \rho_{Pb_{k+1}} + (-1)^{k+2} |R_{y\tau_{k+2}}|_{y,(x,\tau)_t} \rho_{Pb_{k+2}} + \dots + (-1)^q |R_{y\tau_q}|_{y,(x,\tau)_t} \rho_{Pb_q} \\ & + (-1)^{q+q} |R_{yx_{q+1}}|_{y,(x,\tau)_t} \rho_{yx_{q+1}} + (-1)^{q+2} |R_{yx_{q+2}}|_{y,(x,\tau)_t} \rho_{yx_{q+2}} \dots + (-1)^r |R_{yx_r}|_{y,(x,\tau)_t} \rho_{yx_r} \\ & \left. (-1)^{r+1} |R_{yx_{r+1}}|_{y,(x,\tau)_t} \rho_{Pb_{r+1}} + (-1)^{r+2} |R_{yx_{r+2}}|_{y,(x,\tau)_t} \rho_{Pb_{r+2}} + \dots + (-1)^t |R_{yx_t}|_{y,(x,\tau)_t} \rho_{Pb_t} \right) \end{aligned} \quad (2.14)$$

Using (1.8) in (2.14), we get,

$$\text{MSE}(t_{\text{MRR}}) = \theta \bar{Y}^2 C_y^2 \frac{|R|_{y,(x,\tau)_t}}{|R|_{(x,\tau)_t}} \quad (2.15)$$

Using (1.5) in (2.15), we get,

$$\text{MSE}(t_{\text{MRR}}) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y,(x,\tau)_t}) \quad (2.16)$$

3.2. Bias and Consistency of Mixture Regression-Cum-Ratio Estimator

These mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes in single-phase sampling are biased. However, these biases are negligible for moderate and large samples.

It's easily shown that the mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes is a consistent estimator since it is a linear combination of consistent estimators it follows that it is also consistent.

4. Simulation, Result and Discussion

In this section, we carried out data analysis to compare the performance of mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes with already existing estimator namely mean per unit, ratio and regression estimators using one auxiliary variable and attribute, ratio and regression estimators using two auxiliary variables and attributes and regression-cum-ratio estimators using four auxiliary variables and attributes in single-phase sampling for finite population. In the simulated population, the study variable is normally distributed while auxiliary variables and attributes are also normally distributed and strongly positively correlated with the study variable.

Study variable $N = 350$, $n = 50$ mean = 50 standard deviation = 10

For ratio estimator the auxiliary variable and attributes are positively correlated with the study variable and the line passes through the origin.

$N = 350$, $n = 50$ $\rho_{yx_1} = 0.7514$ $\rho_{yx_2} = 0.7213$ $\rho_{y\tau_1} = 0.6713$ $\rho_{y\tau_2} = 0.5781$

For regression estimator the auxiliary variable and attributes are positively correlated with the study variable and the line passes does not pass in the neighborhood of the origin.

$N = 350$, $n = 50$ $\rho_{yx_1} = 0.7622$ $\rho_{yx_2} = 0.5849$ $\rho_{y\tau_1} = 0.6836$ $\rho_{y\tau_2} = 0.5543$

All the results were obtained after carrying out several random sample and taking the average.

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 percent relative efficiency is calculated using the following formulae.

$$\text{eff}(\hat{Y}) = \frac{\text{Var}(\hat{y})}{\text{MSE}(\hat{Y})} * 100 \quad (3.0)$$

The **Table 1** shows percent relative efficiency of mean per unit, ratio and regression estimators using one auxiliary variable and attribute, ratio and regression estimators using two auxiliary variables and attributes and regression-cum-ratio estimators using four auxiliary variables and attributes and mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes with respect to mean per unit estimator for single-phase sampling. It is observed that our proposed mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes using multiple auxiliary variables and attributes is the most efficient of the twelve estimators since it has the highest percent relative efficiency.

5. Conclusion

According to **Table 1**, the proposed mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes using multiple auxiliary variables and attributes has the highest percent relative efficiency compared to mean per unit, ratio and regression estimators using one auxiliary variable and attribute, ratio and regression estimators using two auxiliary variables and attributes and regression-cum-ratio estimators using four auxiliary variables and attributes in single-phase sampling for finite population. This means that the mixture

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

Estimators	Percent relative efficiency with respect to mean per unit
\bar{y}	100
t_{R1}	190
t_{Re1}	129
t_{R11}	206
t_{Re11}	166
t_{R2}	374
t_{Re2}	151
t_{R22}	245
t_{Re22}	234
t_{RRe1}	419
t_{RRe2}	240
$t_{MR(3)}$	490
$t_{MR(3)}$	354
t_{MRR} (proposed)	630

regression-cum-ratio estimator using multiple auxiliary variables and attributes using multiple auxiliary variables and attributes is the most efficient estimator compared to the estimators that utilize auxiliary variables and attributes. The proposed mixture regression-cum-ratio estimator using multiple auxiliary variables and attributes in single-phase sampling is recommended to estimate the finite population mean as it outperforms all the other namely mean per unit, ratio and regression estimators using one auxiliary variable and attribute, ratio and regression estimators using two auxiliary variables and attributes and regression-cum-ratio estimators using four auxiliary variables and attributes in single-phase sampling.

References

- [1] 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>
- [2] Watson, D.J. (1937) The Estimation of Leaf Areas. *Journal of the Agricultural Science*, **27**, 474-504. <http://dx.doi.org/10.1017/S002185960005173X>
- [3] Cochran, W.G. (1940) The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *Journal of the Agricultural Science*, **30**, 262-275. <http://dx.doi.org/10.1017/S0021859600048012>
- [4] Hansen, M.H. and Hurwitz, W.N. (1943) On the Theory of Sampling from Finite Populations. *Annals of Mathematical Statistics*, **14**, 333-362. <http://dx.doi.org/10.1214/aoms/1177731356>
- [5] Olikin, I. (1958) Multivariate Ratio Estimation for Finite Population. *Biometrika*, **45**, 154-165. <http://dx.doi.org/10.1093/biomet/45.1-2.154>
- [6] Shukla, G.K. (1965) Multivariate Regression Estimate. *Journal of the Indian Statistical Association*, **3**, 202-211.
- [7] 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>
- [8] Singh, M.P. (1967) Ratio-Cum-Product Method of Estimation. *Metrika*, **12**, 34-42. <http://dx.doi.org/10.1007/BF02613481>
- [9] 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.
- [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] 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.
- [12] 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>
- [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] Hanif, M., Haq, I. and Shahbaz, M.Q. (2010) Ratio Estimators Using Multiple Auxiliary Attributes. *World Applied Sciences Journal*, **8**, 133-136.
- [15] 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>
- [16] Samiuddin, 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.
- [17] Ahmad, Z. (2008) Generalized Multivariate Ratio and Regression Estimators for Multi-Phase Sampling. Ph.D. Thesis, National College of Business Administration and Economics, Lahore.
- [18] Zahoor, A., Muhammad, H. and Munir, A. (2009) Generalized Regression-Cum-Ratio Estimators for Two-Phase Sampling Using Multiple Auxiliary Variables. *Pakistan Journal of Statistics*, **25**, 93.
- [19] Kung'u, J. and Odongo, L. (2014) Ratio-Cum-Product Estimator Using Multiple Auxiliary Attributes in Single Phase Sampling. *Open Journal of Statistics*, **4**, 239-245. <http://dx.doi.org/10.4236/ojs.2014.44023>
- [20] Kung'u, J. and Odongo, L. (2014) Ratio-Cum-Product Estimator Using Multiple Auxiliary Attributes in Two-Phase Sampling. *Open Journal of Statistics*, **4**, 246-257. <http://dx.doi.org/10.4236/ojs.2014.44024>
- [21] Moeen, M., Shahbaz, Q. and HanIf, M. (2012) Mixture Ratio and Regression Estimators Using Multi-Auxiliary Variable and Attributes in Single Phase Sampling. *World Applied Sciences Journal*, **18**, 1518-1526.
- [22] Arora, S. and Bansil, L. (1989) *New Mathematical Statistics*. Satya Prakashan, New Delhi.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

