Open Journal of Statistics, 2014, 4, 355-366
Published Online August 2014 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2014.45035

Mixture Regression Estimators Using Multi-Auxiliary Variables and Attributes in Two-Phase Sampling

John Kung'u, Grace Chumba, Leo Odongo

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

Received 23 May 2014; revised 26 June 2014; accepted 10 July 2014

Abstract

In this paper, we have developed estimators of finite population mean using Mixture Regression estimators using multi-auxiliary variables and attributes in two-phase sampling and investigated its finite sample properties in full, partial and no information cases. An empirical study using natural data is given to compare the performance of the proposed estimators with the existing estimators that utilizes either auxiliary variables or attributes or both for finite population mean. The Mixture Regression estimators in full information case using multiple auxiliary variables and attributes are more efficient than mean per unit, Regression estimator using one auxiliary variable or attribute, Regression estimator using multiple auxiliary variable or attributes and Mixture Regression estimators in both partial and no information case in two-phase sampling. A Mixture Regression estimator in partial information case is more efficient than Mixture Regression estimators in no information case.

Keywords

Regression Estimator, Multiple Auxiliary Variables, Multiple Auxiliary Attributes, Two-Phase Sampling, Bi-Serial Correlation Coefficient

1. Introduction

The history of using auxiliary information in survey sampling is as old as the history of survey sampling. The work of Neyman [1] may be referred to as the initial work where auxiliary information has been used to estimate population parameters. Hansen and Hurwitz [2] also suggested the use of auxiliary information in selecting the sample with varying probabilities. The concept of ratio estimation was introduced in sample survey by Cochran

How to cite this paper: Kung'u, J., Chumba G. and Odongo, L. (2014) Mixture Regression Estimators Using Multi-Auxiliary Variables and Attributes in Two-Phase Sampling. *Open Journal of Statistics*, **4**, 355-366.
http://dx.doi.org/10.4236/ojs.2014.45035

[3]; it is preferred when the study variable is highly positively correlated with the auxiliary variable. Watson [4] used the regression estimator of leaf area on leaf weight to estimate the average area of the leaves on a plant. Olkin [5] was the first person using 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. Raj [6] suggested a method of using multi-auxiliary information in sample survey.

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 [7]. Abdul, Zahoor and Hanif [8] also developed a generalized multivariate regression estimator for multi-phase sampling using multi-auxiliary variables. Zahoor, Abdul, and Muhhamad [9] suggested a generalized regression-cum-ratio estimator for two-phase sampling using multiple auxiliary variables. 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 variate for large samples. It was proved that optimum estimator in the proposed class of estimators was approximately equally efficient with the usual biased linear regression estimator. Samiuddin and Hanif [10] introduced ratio and regression estimation procedures for estimating population mean in two-phase sampling for different three situations depending upon the availability of information on two auxiliary variables for population. They considered three situations, first when information on both auxiliary variables was not available, second when information on one auxiliary variable was available and third, when information was available on both auxiliary variables.

Jhajj, Sharma and Grover [11] proposed a 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 optimum estimate of the proposed family of mean was less biased and more efficient than mean per unit estimator. 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. The estimator performed better than the usual sample mean and Naik and Gupta [12] estimator. Rajesh Pankaj, Nirmala and Florentins [13] used the auxiliary attribute in regression-ratio type exponential estimator following the work of Bahl and Tuteja [14]; the estimator was more efficient compared to mean per unit, ratio and product type exponential estimator as well as Naik and Gupta [12] estimator.

Hanif, Haq and Shahbaz [15] 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 [11]. They also extended their work to ratio and regression estimator which was generalization of Naik and Gupta [12] estimator in single and double phase sampling with full information, partial information and no information. Moeen, Shahbaz and HanIf [16] 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. Kung'u and Odongo [17] and [18] proposed ratio-cum-product estimators using multiple auxiliary attributes in single and two-phase sampling.

In our paper, we will extend the mixture regression estimator proposed by Moeen, Shahbaz and HanIf [16] to two-phase sampling under full, partial and no information case strategies introduced by Samiuddin and Hanif [10] and also incorporate Arora and Bansi [19] approach in writing down the mean squared error.

2. Preliminaries

2.1. Notation and Assumption

Consider a population of *N* units. Let *Y* be the variable for which we want to estimate the population mean and $X_1, X_2, \cdots, X_p$ are *p* auxiliary variables. For two-phase sampling design let n_1 and n_2 $(n_2 < n_1)$ are sample sizes for first and second phase respectively. $x_{i(1)}$ and $x_{i(2)}$ denote the i^{th} auxiliary variables form first and second phase samples respectively and y_2 denote the variable of interest from second phase. X_i and C_{x_i} denote the population means and coefficient of variation of i^{th} auxiliary variables respectively and ρ_{yx_i} denotes the population correlation coefficient of *Y* and X_i.

Further, let

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

$$y_{(2)} = Y + e_{y_{(2)}},\ x_{i(1)} = X_i + e_{x_{i(1)}} \text{ and } x_{i(2)} = X_i + e_{x_{i(2)}} \quad \left(i = (1, 2, \cdots, p)\right) \tag{1.0}$$

where e_{y_2}, $e_{x_{i(1)}}$ and $e_{x_{i(2)}}$ are sampling error and are very small. We assume that

$$E\left(\bar{e}_{y_2}\right)=E\left(\bar{e}_{x_{i(1)}}\right)=E\left(\bar{e}_{x_{i(2)}}\right)=0 \tag{1.1}$$

Consider a sample of size *n* drawn by simple random sampling without replacement from a population of size *N*. Let y_j and denotes the observations on variable *y* and *r* respectively for the j^{th} unit where $j=1,2,\cdots,n$.

In defining the attributes we assume complete dichotomy so that;

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

Let $A_i=\sum_{j=1}^{N}\tau_{ji}$ and $a_i=\sum_{j=1}^{n}\tau_{ji}$ be the total number of units in the population and sample respectively possessing attribute τ_i. Let $P_i=\frac{A_i}{N}$ and $p_{i(2)}=\frac{a_i}{n}$ be the corresponding proportion of units possessing a specific attributes τ_i and $\bar{y}$ is the mean of the main variable at second phase. Let $p_{i(1)}$ and $p_{i(2)}$ denote the i^{th} auxiliary attribute form first and second phase samples respectively and y_2 denote the variable of interest from second phase. The mean of main variable of interest at second phase will be denoted by $\bar{y}_2$. Also let us define

$$\bar{e}_{y_2}=\bar{y}_2-\bar{Y},\ \ \bar{e}_{\tau_{i(1)}}=p_{i(1)}-P_i,\ \ \bar{e}_{\tau_{i(1)}}=p_{i(1)}-P_i \tag{1.3}$$

The coefficient of variation and correlation coefficient are given by

$$C_y^2=\frac{S_{y_2}^2}{\bar{Y}^2},\quad C_y^2=\frac{S_{y_2}^2}{\bar{Y}^2},\quad C_{\tau_1}^2=\frac{S_{x_1}^2}{\bar{P}_1^2},\quad \rho_{yx}=\frac{S_{yx}}{S_yS_x},\quad \rho_{yz}=\frac{S_{yz}}{S_yS_z},\quad \rho_{Pb_i}=\frac{S_{y\tau}}{S_yS_\tau},\ \text{and}\ \rho_{zx}=\frac{S_{zx}}{S_zS_x}$$

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\left(\bar{e}_{y_2}\right)^2=\theta_2\bar{Y}^2C_{y_2}^2,\quad E\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)^2=\left(\theta_2-\theta_1\right)P_i^2C_{\tau_i}^2\\
&E\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)^2=\left(\theta_2-\theta_1\right)\bar{X}_i^2C_{x_i}^2,\quad E\left(\bar{e}_{y_2}\bar{e}_{x_{i(2)}}\right)^2=\theta_2\overline{YX}_iC_{y_2}C_{x_i}\rho_{yx_i}\\
&E\left(\bar{e}_{y_2}\bar{e}_{\tau_{i(2)}}\right)^2=\theta_2\bar{Y}P_iC_{y_2}C_{\tau_i}\rho_{Pb_i},\quad E\left(\bar{e}_{y_{(2)}}\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)\right)=\left(\theta_1-\theta_2\right)\overline{YX}_jC_yC_{x_i}\rho_{yx_i}\\
&E\left(\bar{e}_{x_{i(2)}}\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)\right)=\left(\theta_1-\theta_2\right)\bar{X}_i^2C_{x_i}^2,\quad E\left(\bar{e}_{\tau_{i(2)}}\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)\right)=\left(\theta_1-\theta_2\right)P_i^2C_{\tau_i}^2\\
&E\left(\bar{e}_{x_{i(2)}}\bar{e}_{x_{i(2)}}\right)=\theta_2\bar{X}_i\bar{X}_jC_{x_i}C_{x_j}\rho_{x_ix_j};(i\neq j),\quad E\left(\bar{e}_{\tau_{i(2)}}\bar{e}_{\tau_{i(2)}}\right)=\theta_2P_iP_jC_{\tau_i}C_{\tau_j}\rho_{\tau_i\tau_j};(i\neq j)\\
&E\left(\bar{e}_{y_{(2)}}\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)\right)=\left(\theta_1-\theta_2\right)\overline{YX}_jC_yC_{x_i}\rho_{yx_i}\\
&E\left(\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)\left(\bar{e}_{\tau_{j(1)}}-\bar{e}_{\tau_{j(2)}}\right)\right)=\left(\theta_1-\theta_2\right)P_iP_jC_{\tau_j}C_{\tau_i}\rho_{\tau_j\tau_i};(i\neq j)\\
&E\left(\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)\left(\bar{e}_{x_{j(1)}}-\bar{e}_{x_{j(2)}}\right)\right)=\left(\theta_1-\theta_2\right)\bar{X}_i\bar{X}_jC_{x_j}C_{x_i}\rho_{x_jx_i};(i\neq j)\\
&E\left(\bar{e}_{x_{i(2)}}\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)\right)=\left(i\neq j\right)\bar{X}_i\bar{X}_jC_{x_j}C_{x_i}\rho_{x_jx_i};(i\neq j)\\
&E\left(\bar{e}_{\tau_{i(2)}}\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)\right)=\left(\theta_1-\theta_2\right)P_iP_jC_{\tau_j}C_{\tau_i}\rho_{\tau_j\tau_i};(i\neq j)
\end{aligned} \tag{1.4}$$

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

$$\frac{|R|_{y\underline{x}_q}}{|R|_{\underline{x}_q}}=\left(1-\rho_{y\cdot\underline{x}_q}^2\right)\ \ \text{Arora and Lai [19]} \tag{1.6}$$

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

$|R|_{y\underset{\sim}{x}_p}$ Determinant of population correlation matrix of variables $y,\ x_1,\ x_2,\cdots,\ x_{q-1}$ and x_q.

$\left|R_{yx_i}\right|_{y\underset{\sim}{x}_q}$ Determinant of i^{th} minor of $|R|_{y\underset{\sim}{x}_p}$ corresponding to the i^{th} element of ρ_{yx_i}.

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

$\rho^2_{y\underset{\sim}{x}_q}$ Denotes the multiple coefficient of determination of y on $y,\ x_1,\ x_2,\cdots,\ x_{q-1}$ and x_q.

$|R|_{\underset{\sim}{x}_r}$ Determinant of population correlation matrix of variables $x_1,\ x_2,\cdots,\ x_{r-1}$ and x_r.

$|R|_{\underset{\sim}{x}_p}$ Determinant of population correlation matrix of variables $x_1,\ x_2,\cdots,\ x_{p-1}$ and x_p.

$|R|_{y_i\underset{\sim}{x}_r}$ Determinant of the correlation matrix of $y_i,\ x_1,\ x_2,\cdots,\ x_{r-1}$ and x_r.

$|R|_{y_i\cdot\underset{\sim}{x}_p}$ Determinant of the correlation matrix of $y_i,\ x_1,\ x_2,\cdots,\ x_{p-1}$ and x_p.

$|R|_{y_i\cdot y_j\cdot\underset{\sim}{x}_r}$ Determinant of the minor corresponding to $\rho_{y_iy_j}$ of the correlation matrix of $y_i, y_j, x_1, x_2,\cdots, x_{r-1}$ and $x_{r(i\neq j)}$.

$|R|_{y_i\cdot y_j\cdot\underset{\sim}{x}_p}$ Determinant of the minor corresponding to $\rho_{y_iy_j}$ of the correlation matrix of

$$y_i, y_j, x_1, x_2,\cdots, x_{p-1} \text{ and } x_{q(i\neq j)} \tag{1.7}$$

2.2. Mean per Unit in Two-Phase Sampling

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

$$\bar{y}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} y_i \tag{2.0}$$

While its variance is given,

$$\operatorname{Var}(\bar{y}_2) = \theta_2 \bar{Y}^2 C_y^2 \tag{2.1}$$

2.3. Regression Estimators Using One and Multiple Auxiliary Variables and Attributes

Let $\bar{y}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j$ and $\bar{x}_{1(2)} = \frac{1}{n_2}\sum_{j=1}^{n_2} x_j$ be the unbiased estimator of sample means of $\bar{Y}$ and $\bar{X}$ respectively in two phase sampling. The simple regression estimator for known X suggested by Watson [4] is,

$$t_{\text{REX}} = \bar{y}_2 + \beta\left(\bar{X}_1 - \bar{x}_{1(2)}\right) \tag{2.2}$$

Its mean squared error is given by,

$$\operatorname{MSE}(t_{\text{REX}}) = \theta_2 \bar{Y}^2 C_y^2 \left(1 - \rho_{yx}^2\right) \tag{2.3}$$

In case of multiple auxiliary variables, regression estimator is given by,

$$t_{\text{MREX}} = \bar{y}_2 + \sum_{i=1}^{k} \alpha_i \left(\bar{X}_i - \bar{x}_{1(2)}\right) \tag{2.4}$$

Its mean squared error is given by,

$$\operatorname{MSE}(t_{\text{MREX}}) = \theta_2 \bar{Y}^2 C_y^2 \left(1 - \rho_{y\underset{\sim}{x}_k}^2\right) \tag{2.5}$$

Naik and Gupta [12] defined Regression estimator of population when the prior information of population proportion of units, possessing the same attribute is variable as,

$$t_{\text{REP}} = \bar{y}_2 + \alpha_1\left(P_1 - p_{1(2)}\right) \tag{2.6}$$

Its mean squared error is given by,

$$\text{MSE}\left(t_{\text{REP}}\right) = \theta_2 \bar{Y}^2 C_y^2\left(1 - \rho_{y\tau_1}^2\right) \tag{2.7}$$

$\alpha_1 = \dfrac{\bar{Y} C_y \rho_{\rho b_1}}{C_{r_1} P_1}$ are optimum for Regression estimator. $\rho_{Pb_1} = \dfrac{S_{yr}}{S_y S_r}$ is the bi-serial correlation coefficient.

In case of multiple auxiliary variables, regression estimator is given by,

$$t_{\text{MREP}} = \bar{y}_2 + \sum_{i=1}^{k} \alpha_i\left(P_i - p_{i(2)}\right) \tag{2.8}$$

Its mean squared error is given by,

$$\text{MSE}\left(t_{\text{MREX}}\right) = \theta_2 \bar{Y}^2 C_y^2\left(1 - \rho_{y\tau_k}^2\right) \tag{2.9}$$

The mixture ratio estimator based on multiple auxiliary variables and attributes by Moeen, Shahbaz and HanIf [16] is given by:

$$t_{\text{REXP}} = \bar{y} + \sum_{i=1}^{k} \alpha_i\left(\bar{X}_i - \bar{x}_i\right) + \sum_{i=k+1}^{k} \beta_i\left(P_i - p_i\right) \tag{3.0}$$

It is normally known that the above estimators are biased but the bias being of the order n^{-1}, can be assumed negligible in large samples. It is assumed that the sample of size n is large enough so that the biases of these estimators are negligible.

Our project will extend the mixture regression estimator proposed by Moeen, Shahbaz and Hanif [16] to two-phase sampling under full, partial and no information case strategies introduced by Samiuddin and Hanif [10].

3. Methodology

3.1. Proposed Mixture Regression Estimator in Two-Phase Sampling (Full Information Case)

If we estimate a study variable when information on all auxiliary variables and attributes is available from population, it is utilized in the form of their means. By taking the advantage of Mixture Regression estimator technique for two-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:

$$\begin{aligned} t_{\text{MR}(3.0)} = \Big(&\bar{y}_2 + \alpha_1\left(\bar{X}_1 - \bar{x}_{1(2)}\right) + \alpha_2\left(\bar{X}_2 - \bar{x}_{2(2)}\right) + \cdots + \alpha_k\left(\bar{X}_k - \bar{x}_{k(2)}\right) \\ &+ \beta_{k+1}\left(P_{k+1} - p_{k+1(2)}\right) + \beta_{k+2}\left(P_{k+2} - p_{k+2(2)}\right) + \cdots + \beta_q\left(P_q - p_{q(2)}\right)\Big) \end{aligned} \tag{3.0}$$

Substituting Equation (1.0) and (1.3) in (3.0), we get,

$$t_{\text{MR}(3.0)} = \bar{e}_{y_2} + \bar{Y} - \sum_{i=1}^{k} \alpha_i \bar{e}_{x_{i(2)}} - \sum_{i=k+1}^{q} \beta_i \bar{e}_{\tau_{i(2)}} \tag{3.2}$$

The mean squared error of $t_{\text{MR}(3.0)}$ is given by

$$\text{MSE}\left(t_{\text{MR}(3.0)}\right) = E\left(t_{\text{MR}(3.0)} - \bar{Y}\right)^2 = E_2\left(\bar{e}_{y_2} - \sum_{i=1}^{k} \alpha_i \bar{e}_{x_{(2)i}} - \sum_{i=k+1}^{h+k=m} \beta_i \bar{e}_{\tau_{(2)i}}\right)^2 \tag{3.3}$$

We differentiate the Equation (3.3) partially with respect to $\alpha_i\left(i = 1, 2, \cdots, k\right)$ and $\beta_i\left(i = k+1, k+2, \cdots, m\right)$ then equate to zero, using (1.4), (1.5), (1.6) and (1.7), we get

$$\alpha_i = (-1)^{i+1} \frac{\bar{Y} C_y}{\bar{X}_i C_{x_i}} \frac{\left|R_{yx_i}\right|_{\underline{x}_k}}{\left|R\right|_{\underline{x}_p}} \tag{3.4}$$

$$\beta_i = (-1)^{i+1} \frac{\bar{Y} C_y}{P_i C_{\tau_i}} \frac{\left|R_{y\tau_i}\right|_{\underset{\sim}{\tau}h}}{|R|_{\underset{\sim}{\tau}h}} \tag{3.5}$$

Using normal equation that is used to find the optimum values given (3.3), we can write (3.3) as,

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = E_2\left(\bar{e}_{y_2}\left(\bar{e}_{y_2} - \sum_{i=1}^{k} \alpha_i \bar{e}_{x_{(2)i}} - \sum_{i=k+1}^{h+k=p} \beta_i \bar{e}_{\tau_{(2)i}}\right)\right) \tag{3.6}$$

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = E_2\left(\bar{e}_{y_2}^2\right) - \sum_{i=1}^{k} \alpha_i E_2\left(\bar{e}_{y_2} \bar{e}_{x_{(2)i}}\right) - \sum_{i=k+1}^{k+h=p} \beta_i E_2\left(\bar{e}_{y_2} \bar{e}_{\tau_{(2)i}}\right) \tag{3.7}$$

Taking expectation in (3.7) and substituting (1.4), we get,

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = \theta_2\left(\bar{Y}^2 C_y^2 + \sum_{i=1}^{k} (-1)^i \frac{\bar{Y} C_y}{\bar{\bar{X}}_i C_{x_I}} \frac{\left|R_{yx_i}\right|_{\underline{x}_k}}{|R|_{\underline{x}_k}} \bar{X}_i \bar{Y} C_y C_{x_i} \rho_{yx_i} + \sum_{i=1}^{k+h=p} (-1)^i \frac{\bar{Y} C_y}{P_i C_{\tau_i}} \frac{\left|R_{y\tau_i}\right|_{\underset{\sim}{\tau}h}}{|R|_{\underset{\sim}{\tau}h}} P_i \bar{Y} C_y C_{\tau_i} \rho_{Pb_i}\right) \tag{3.8}$$

Or

$$= \theta_2 \bar{Y}^2 C_y^2\left(1 + \left(\sum_{i=1}^{k} (-1)^i \frac{\left|R_{yx_i}\right|_{\underline{x}_k}}{|R|_{\underline{x}_k}} \rho_{yx_i} + \sum_{i=k+1}^{k+h=q} (-1)^i \frac{\left|R_{y\tau_i}\right|_{\underset{\sim}{\tau}h}}{|R|_{\underset{\sim}{\tau}h}} \rho_{Pb_i}\right)\right) \tag{3.9}$$

Or

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = \theta_2 \bar{Y}^2 C_y^2\left(1 + \left(\sum_{i=1}^{q} (-1)^i \frac{\left|R_{yx_i}\right|_{(\underline{x},\underset{\sim}{\tau})_q}}{|R|_{(\underline{x},\underset{\sim}{\tau})_q}} \rho_{y(x_i,\tau_i)}\right)\right) \tag{3.10}$$

Or

$$\begin{aligned} &\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) \\ &= \frac{\theta_2 \bar{Y}^2 C_y^2}{|R|_{(\underline{x},\underset{\sim}{\tau})_q}}\left(|R|_{\underline{x}_p} - \rho_{y_2,(x_1,\tau_1)}\left|R_{yx_1}\right|_{y,x_1,\tau_1} + \rho_{y_2,(x_2,\tau_2)}\left|R_{yx_2}\right|_{y,x_1,\tau_1} + \cdots + (-1)^q \left|R_{y,(x_q,\tau_q)}\right|_{y,x_q,\tau_q} \rho_{y_q,(x_2,\tau_2)}\right) \end{aligned} \tag{3.11}$$

Or

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = \frac{\theta_2 \bar{Y}^2 C_y^2}{|R|_{(\underline{x},\underset{\sim}{\tau})_q}}\left(|R|_{(\underline{x},\underset{\sim}{\tau})_q} + \sum_{i=1}^{q} (-1)^i \left|R_{y,x_i,\tau_i}\right|_{y,x_i,\tau_i} \rho_{y,x_i,\tau_i}\right) \tag{3.12}$$

Or

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = \theta_2 \bar{Y}^2 C_y^2 \frac{|R|_{y(\underline{x},\underset{\sim}{\tau})_q}}{|R|_{(\underline{x},\underset{\sim}{\tau})_q}} \tag{3.13}$$

Using (1.6) in (3.13), we get

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.0)}\right) = \theta_2 \bar{Y}^2 C_y^2\left(1 - \rho_{y.(\underline{x},\underset{\sim}{\tau})_q}^2\right) \tag{3.14}$$

3.2. Mixture Regression Estimator in Two-Phase Sampling (Partial Information Case)

In this case suppose, we have no information on all t auxiliary variables and h auxiliary attributes from population. Considering Mixture Regression estimator technique, the population mean of study variable Y can be estimated for two-phase sampling using multi-auxiliary variables and attributes as:

$$\begin{aligned} t_{\mathrm{MR}(3.1)} = &\Big(\bar{y}_2 + \alpha_1\left(\bar{x}_{1(1)} - \bar{x}_{1(2)}\right) + \alpha_2\left(\bar{x}_{2(1)} - \bar{x}_{2(2)}\right) + \cdots + \alpha_t\left(\bar{x}_{t(1)} - \bar{x}_{t(2)}\right) + \delta_1\left(\bar{X}_1 - \bar{x}_{1(2)}\right) + \delta_2\left(\bar{X}_2 - \bar{x}_{2(2)}\right) \\ &+ \cdots + \delta_t\left(\bar{X}_t - \bar{x}_{t(2)}\right)\alpha_1\left(\bar{x}_{t+1(1)} - \bar{x}_{t+1(2)}\right) + \alpha_2\left(\bar{x}_{t+2(1)} - \bar{x}_{t+2(2)}\right) + \cdots + \alpha_k\left(\bar{x}_{k(1)} - \bar{x}_{k(2)}\right) \\ &+ \beta_{k+1}\left(p_{k+1(1)} - p_{k+1(2)}\right) + \beta_{k+2}\left(p_{k+2(1)} - p_{2(2)}\right) + \cdots + \beta_g\left(p_{g(1)} - p_{g(2)}\right)\gamma_{k+1}\left(P_{k+1} - p_{k+1(2)}\right) \\ &+ \gamma_{k+2}\left(P_{k+2} - p_{k+2(2)}\right) + \cdots + \gamma_g\left(P_g - p_{g(2)}\right) + \beta_{g+1}\left(P_{g+1(1)} - p_{g+2(2)}\right) + \beta_{g+2}\left(P_{g+2(1)} - p_{g+2(2)}\right) \\ &+ \cdots + \beta_q\left(P_{q(1)} - p_{q(2)}\right)\Big) \end{aligned} \tag{3.15}$$

Substituting (1.0) and (1.3) in (3.15), we get,

$$\begin{aligned} = &\left(\bar{e}_{y_2} + \bar{Y} + \sum_{i=1}^{t}\alpha_i\left(\bar{e}_{x_{i(1)}} - \bar{e}_{x_{i(2)}}\right) - \sum_{i=1}^{t}\delta_i\bar{e}_{x_{i(2)}} + \sum_{i=t+1}^{k}\alpha_i\left(\bar{e}_{x_{i(1)}} - \bar{e}_{x_{i(2)}}\right) + \sum_{i=k+1}^{g}\beta_i\left(\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}\right)\right. \\ &\left. - \sum_{i=k+1}^{g}\gamma_i\bar{e}_{\tau_{i(2)}} + \sum_{i=g+1}^{q}\beta_i\left(\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}\right)\right) \end{aligned} \tag{3.16}$$

Mean squared error of $t_{\mathrm{MR}(3.1)}$ estimator is given by

$$\begin{aligned} \mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right) = E_1E_{2/1}&\left(\bar{e}_{y_2} + \sum_{i=1}^{t}\alpha_i\left(\bar{e}_{x_{i(1)}} - \bar{e}_{x_{i(2)}}\right) - \sum_{i=1}^{t}\delta_i\bar{e}_{x_{i(2)}} + \sum_{i=k+1}^{k}\alpha_i\left(\bar{e}_{x_{i(1)}} - \bar{e}_{x_{i(2)}}\right)\right. \\ &\left. + \sum_{i=k+1}^{g}\beta_i\left(\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}\right) - \sum_{i=k+1}^{g}\gamma_i\bar{e}_{\tau_{i(2)}} + \sum_{i=k+1}^{q}\beta_i\left(\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}\right)\right)^2 \end{aligned} \tag{3.17}$$

We differentiate the Equation (3.24) with respect to $\alpha_i\left(i = 1, 2, \cdots, r\right)$, $\beta_i\left(i = 1, 2, \cdots, r\right)$, $\alpha_i\left(i = r+1, r+2, \cdots, k\right)$, $\gamma_i\left(i = k+1, k+2, \cdots, h\right)$, $\lambda_i\left(i = k+1, k+2, \cdots, h\right)$, $\gamma_i\left(i = h+1, h+2, \cdots, p\right)$ and equate to zero and use (1.6) and (1.7). The optimum value is as follows,

$$\begin{aligned} &\alpha_i = (-1)^{i+1}\frac{\bar{Y}C_y}{\bar{X}_iC_{x_i}}\left(\frac{\left|R_{yx_i}\right|_{\underline{x}_k}}{|R|_{\underline{x}_k}} - \frac{\left|R_{yx_i}\right|_{\underline{x}_s}}{\left|R_{yx_i}\right|_{\underline{x}_s}}\right) && \delta_i = (-1)^{i+1}\frac{\bar{Y}C_y}{|R|_{\underline{x}_t}\bar{X}_iC_{x_i}}\left|R_{yx_i}\right|_{\underline{x}_t} \\ &i = 1, 2, \cdots, r && i = 1, 2, \cdots, r \\ &\alpha_i = (-1)^{i+1}\frac{\bar{Y}C_y}{\bar{X}_iC_{x_i}}\frac{\left|R_{yx_i}\right|_{\underline{x}_s}}{|R|_{\underline{x}_s}} && \beta_i = (-1)^{i+1}\frac{\bar{Y}C_y}{P_iC_{x_i}}\left(\frac{\left|R_{y\tau_i}\right|_{\underline{\tau}_z}}{|R|_{\underline{\tau}_z}} - \frac{\left|R_{y\tau_i}\right|_{\underline{x}_g}}{|R|_{\underline{x}_g}}\right) \\ &i = t+1, t+2, \cdots, k && i = k+1, k+2, \cdots, h \\ &\gamma_i = -\frac{C_y}{|R|_{\underline{\tau}_g}}\frac{\bar{Y}}{P_iC_{x_i}}\left((-1)^{i+1}\left|R_{y\tau_i}\right|_{\underline{\tau}_t}\right) && \beta_i = (-1)^{i+1}\frac{\bar{Y}C_y}{|R|_{\underline{\tau}_h}P_iC_{\tau_i}}\left|R_{y\tau_i}\right|_{\underline{\tau}_h} \\ &i = k+1, k+2, \cdots, h && i = h+1, h+2, \cdots, q \end{aligned} \tag{3.18}$$

Using normal equation that are used to find the optimum values given (3.17) we can write

$$\begin{aligned} \mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right) = E_1E_{2/1}&\left(\bar{e}_{y_2}\left(\bar{e}_{y_2} + \sum_{i=1}^{t}\alpha_i\left(\bar{e}_{x_{i(1)}} - \bar{e}_{x_{i(2)}}\right) - \sum_{i=1}^{t}\delta_i\bar{e}_{x_{i(2)}} + \sum_{i=k+1}^{k}\alpha_i\left(\bar{e}_{x_{i(1)}} - \bar{e}_{x_{i(2)}}\right)\right.\right. \\ &\left.\left. + \sum_{i=k+1}^{g}\beta_i\left(\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}\right) - \sum_{i=k+1}^{g}\gamma_i\bar{e}_{\tau_{i(2)}} + \sum_{i=k+1}^{q}\beta_i\left(\bar{e}_{\tau_{i(1)}} - \bar{e}_{\tau_{i(2)}}\right)\right)\right) \end{aligned} \tag{3.19}$$

Or

$$\begin{aligned}\mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right) = {} & E_2\left(\bar{e}_{y_2}^2\right)+\sum_{i=1}^{t}\alpha_i E_1E_{2/1}\left(\bar{e}_{y_2}\left(\bar{e}_{x_{(1)i}}-\bar{e}_{x_{(2)i}}\right)\right)-\sum_{i=1}^{t}\delta_i E_2\left(\bar{e}_{y_2}\left(\bar{e}_{x_{(2)i}}\right)\right)+\sum_{i=1}^{t}\delta_i E_1E_{2/1}\left(\bar{e}_{y_2}\left(\bar{e}_{x_{(1)i}}-\bar{e}_{x_{(2)i}}\right)\right) \\ & +\sum_{i=k+1}^{g}\alpha_i E_1E_{2/1}\left(\bar{e}_{y_2}\left(\bar{e}_{x_{(1)i}}-\bar{e}_{x_{(2)i}}\right)\right)-\sum_{i=k+1}^{g}\gamma_i E_2\left(\bar{e}_{y_2}\left(\bar{e}_{x_{(2)i}}\right)\right)+\sum_{i=g+1}^{q}\delta_i E_1E_{2/1}\left(\bar{e}_{y_2}\left(\bar{e}_{x_{(1)i}}-\bar{e}_{x_{(2)i}}\right)\right)\end{aligned} \tag{3.20}$$

Using (1.4) in (3.28) we get,

$$\begin{aligned}\mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right) = {} & \theta_2\bar{Y}^2C_y^2+\left(\theta_1-\theta_2\right)\sum_{i=1}^{t}\alpha_i\bar{X}_i\bar{Y}C_yC_{x_i}\rho_{yx_i}-\sum_{i=1}^{t}\theta_2\delta_i\bar{X}_iC_yC_{x_i}\rho_{yx_i}+\left(\theta_1-\theta_2\right)\sum_{i=1}^{k}\alpha_i\bar{X}_iC_yC_{x_i}\rho_{yx_i} \\ & +\left(\theta_1-\theta_2\right)\sum_{i=k+1}^{g}\beta_iP_i\bar{Y}C_yC_{x_i}\rho_{Pb_i}-\sum_{i=k+1}^{g}\theta_2\gamma_iP_iC_yC_{x_i}\rho_{Pb_i}+\left(\theta_1-\theta_2\right)\sum_{i=g+1}^{q}\beta_iP_iC_yC_{x_i}\rho_{Pb_i}\end{aligned} \tag{3.22}$$

Or

$$\begin{aligned} = \bar{Y}^2C_y^2 & \left(\theta_2+\left(\theta_1-\theta_2\right)\sum_{i=1}^{t}(-1)^{i+1}\left(\frac{\left|R_{yx_i}\right|_{y\underline{x}_k}}{|R|_{\underline{x}_p}}-\frac{\left|R_{yx_i}\right|_{y\underline{x}_m}}{\left|R_{\underline{x}_m}\right|}\right)\rho_{yx_i}-\theta_2\sum_{i=1}^{t}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_t}}{|R|_{\underline{x}_m}}\rho_{yx_i}\right. \\ & +\left(\theta_1-\theta_2\right)\sum_{i=t+1}^{k}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_q}}{|R|_{\underline{x}_q}}\rho_{yx_i}+\left(\theta_1-\theta_2\right)\sum_{i=k+1}^{g}(-1)^{i+1}\left(\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_q}}{|R|_{\underline{x}_q}}-\frac{\left|R_{y\tau_j}\right|_{y\underline{\tau}_m}}{|R|_{\underline{\tau}_m}}\right)\rho_{Pb_i} \\ & \left.-\theta_2\sum_{i=k+1}^{g}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_t}}{|R|_{\underline{\tau}_m}}\rho_{\rho Pb_i}+\left(\theta_1-\theta_2\right)\sum_{i=h+1}^{q}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_p}}{|R|_{\underline{\tau}_p}}\gamma_i\rho_{Pb_i}\right)\end{aligned} \tag{3.23}$$

Or

$$\begin{aligned} = \bar{Y}^2C_y^2 & \left(\theta_2+\left(\theta_1-\theta_2\right)\sum_{i=1}^{t}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_p}}{|R|_{\underline{x}_p}}\rho_{yx_i}-\left(\theta_1-\theta_2\right)\sum_{i=1}^{t}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_m}}{|R|_{\underline{x}_m}}\rho_{yx_i}-\theta_2\sum_{i=1}^{t}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_t}}{|R|_{\underline{x}_p}}\rho_{yx_i}\right. \\ & +\left(\theta_1-\theta_2\right)\sum_{i=t+1}^{k}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_p}}{|R|_{\underline{x}_p}}\rho_{yx_i}+\left(\theta_1-\theta_2\right)\sum_{i=k+1}^{g}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_q}}{|R|_{\underline{x}_p}}\rho_{Pb_i}-\left(\theta_1-\theta_2\right)\sum_{i=k+1}^{g}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_m}}{|R|_{\underline{\tau}_m}}\rho_{Pb_i} \\ & \left.-\theta_2\sum_{i=k+1}^{g}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_q}}{|R|_{\underline{\tau}_q}}\rho_{Pb_i}+\left(\theta_1-\theta_2\right)\sum_{i=g+1}^{q}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{\tau}_q}}{|R|_{\underline{\tau}_q}}\gamma_i\rho_{Pb_i}\right)\end{aligned} \tag{3.24}$$

Or

$$\begin{aligned} & \mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right) \\ & = \bar{Y}^2C_y^2\left(\left(\theta_2-\theta_1\right)\left(1-\left(\sum_{i=1}^{t}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_p}}{|R|_{\underline{x}_p}}\rho_{yx_i}+\sum_{i=r+1}^{k}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_p}}{|R|_{\underline{x}_p}}\rho_{yx_i}+\sum_{i=k+1}^{g}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{x}_p}}{|R|_{\underline{\tau}_p}}\rho_{Pb_i}\right)\right)\right. \\ & \left.\quad+\theta_1\left(1-\left(\sum_{i=1}^{t}(-1)^{i+1}\frac{\left|R_{yx_i}\right|_{y\underline{x}_r}}{|R|_{\underline{x}_m}}\rho_{yx_i}+\sum_{i=k+1}^{g}(-1)^{i+1}\frac{\left|R_{yx_j}\right|_{y\underline{x}_t}}{|R|_{\underline{\tau}_m}}\rho_{Pb_i}\right)\right)\right)\end{aligned} \tag{3.25}$$

Or

$$=\bar{Y}^2C_y^2\left(\frac{(\theta_2-\theta_1)}{|R|_{(\underline{x},\underline{\tau})_q}}\left(|R|_{(\underline{x},\underline{\tau})_q}+\sum_{i=1}^{q}(-1)^i\frac{|R_{y,x_i\tau_i}|_{y,\underline{\tau},\underline{x}_p}}{|R|_{(\underline{x},\underline{\tau})_q}}\rho_{y,x_i\tau_i}\right)\right)+\frac{\theta_1}{|R|_{(\underline{x},\underline{\tau})_m}}\left(1+\left(\sum_{i=1}^{t}(-1)^i\frac{|R_{y,x_i,\tau_j}|_{y(\underline{x},\underline{\tau})_m}}{|R|_{(\underline{x},\underline{\tau})_m}}\rho_{y,x_i\tau_i}\right)\right) \quad (3.26)$$

Or

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right)=\bar{Y}^2C_y^2\left((\theta_2-\theta_1)\frac{|R|_{y.(\underline{x},\underline{\tau})_q}}{|R|_{(\underline{x},\underline{\tau})_q}}+\theta_1\frac{|R|_{y.(\underline{x},\underline{\tau})_m}}{|R|_{(\underline{x},\underline{\tau})_m}}\right) \quad (3.27)$$

Using (1.6) in (3.27), we get

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.1)}\right)=\bar{Y}^2C_y^2\left(\theta_2\left(1-\rho^2_{y.(\underline{x},\underline{\tau})_q}\right)+\theta_1\left(\rho^2_{y.(\underline{x},\underline{\tau})_q}-\rho^2_{y.(\underline{x},\underline{\tau})_m}\right)\right) \quad (3.28)$$

3.3. Mixture Regression 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 Mixture Regression estimator technique for two-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:

$$\begin{aligned}t_{\mathrm{MR}(3.2)}=&\left(\bar{y}_2+\alpha_1\left(\bar{x}_{1(1)}-\bar{x}_{1(2)}\right)+\alpha_2\left(\bar{x}_{2(1)}-\bar{x}_{2(2)}\right)+\cdots+\alpha_k\left(\bar{x}_{k(1)}-\bar{x}_{k(2)}\right)\right.\\&\left.+\beta_{k+1}\left(p_{k+1(1)}-p_{k+1(2)}\right)+\beta_{k+2}\left(p_{k+2(2)}-p_{k+2(2)}\right)+\cdots+\beta_q\left(p_{q(2)}-p_{q(2)}\right)\right)\end{aligned} \quad (3.29)$$

Substituting equation (1.0) and (1.1) in (3.29), we get,

$$t_{\mathrm{MR}(3.2)}=\bar{e}_{y(2)}+\bar{Y}+\sum_{i=1}^{k}\alpha_i\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)+\sum_{i=k+1}^{q}\beta_i\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right) \quad (3.30)$$

The mean squared error of $t_{RP(3.2)}$ is given by,

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=E\left(t_{\mathrm{MR}(3.2)}-\bar{Y}\right)^2=E_2\left(\bar{e}_{y_2}+\sum_{i=1}^{k}\alpha_i\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)+\sum_{i=k+1}^{h+k=q}\beta_i\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)\right)^2 \quad (3.31)$$

We differentiate the Equation (3.32) partially with respect to $\alpha_i\left(i=1,2,\cdots,k\right)$ and $\beta_i\left(i=k+1,k+2,\cdots,q\right)$ then equate to zero, using (1.4), (1.5), (1.6) and (1.7), we get

$$\alpha_i=(-1)^{i+1}\frac{\bar{Y}C_y}{\bar{X}_iC_{x_i}}\frac{|R_{yx_i}|_{\underline{x}_P}}{|R|_{\underline{x}_p}} \quad (3.32)$$

$$\beta_i=(-1)^{i+1}\frac{\bar{Y}C_y}{P_iC_{\tau_i}}\frac{|R_{y\tau_i}|_{\underline{\tau}_h}}{|R|_{\underline{\tau}_h}} \quad (3.33)$$

Using normal equation that is used to find the optimum values given (3.31) we can write,

$$\mathrm{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=E_2\left(\bar{e}_{y_2}\left(\bar{e}_{y_2}+\sum_{i=1}^{k}\alpha_i\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)+\sum_{i=k+1}^{h+k=q}\beta_i\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)\right)\right) \quad (3.34)$$

$$\mathrm{MSE}\left(t_{\mathrm{MRP}(3.2)}\right)=E_2\left(\bar{e}^2_{y_2}\right)+\sum_{i=1}^{k}\alpha_iE_2\left(\bar{e}_{y_2}\left(\bar{e}_{x_{i(1)}}-\bar{e}_{x_{i(2)}}\right)\right)+\sum_{i=k+1}^{k+h=q}\beta_iE_2\left(\bar{e}_{y_2}\left(\bar{e}_{\tau_{i(1)}}-\bar{e}_{\tau_{i(2)}}\right)\right) \quad (3.35)$$

Taking expectation and substituting (3.32) and (3.33) and, we get,

$$\operatorname{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=\theta_2\bar{Y}^2C_y^2+\left(\theta_2-\theta_1\right)\sum_{i=1}^{k}(-1)^i\frac{\bar{Y}C_y}{\bar{X}_iC_{x_i}}\frac{\left|R_{yx_i}\right|_{y\underline{x}_k}}{|R|_{\underline{x}_k}}\bar{X}_i\bar{Y}C_yC_{x_i}\rho_{yx_i}$$
$$+\left(\theta_2-\theta_1\right)\sum_{i=k+1}^{k+h=q}(-1)^i\frac{\bar{Y}C_y}{P_iC_{\tau_i}}\frac{\left|R_{y\tau_i}\right|_{y\underline{\tau}_h}}{|R|_{\underline{\tau}_h}}P_i\bar{Y}C_yC_{\tau_i}\rho_{y\tau_i} \tag{3.36}$$

Or

$$\operatorname{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=\bar{Y}^2C_y^2\left(\theta_2+\left(\theta_2-\theta_1\right)\sum_{i=1}^{k}(-1)^i\frac{\left|R_{yx_i}\right|_{y\underline{x}_k}}{|R|_{\underline{x}_k}}\rho_{yx_i}+\left(\theta_2-\theta_1\right)\sum_{i=k+1}^{k+h=q}(-1)^i\frac{\left|R_{y\tau_i}\right|_{y\underline{\tau}_h}}{|R|_{\underline{\tau}_h}}\rho_{Pb_i}\right) \tag{3.37}$$

Or

$$\operatorname{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=\bar{Y}^2C_y^2\left(\left(\theta_2-\theta_1\right)\frac{|R|_{y.(\underline{x},\underline{\tau})_q}}{|R|_{(\underline{x},\underline{\tau})_q}}+\theta_1\right) \tag{3.38}$$

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

$$\operatorname{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=\bar{Y}^2C_y^2\left(\left(\theta_2-\theta_1\right)\left(1-\rho^2_{y.(\underline{x},\underline{\tau})_q}\right)+\theta_1\right) \tag{3.39}$$

Simplifying (3.38) we get,

$$\operatorname{MSE}\left(t_{\mathrm{MR}(3.2)}\right)=\bar{Y}^2C_y^2\left(\theta_2\left(1-\rho^2_{y.(\underline{x},\underline{\tau})_q}\right)+\theta_1\rho^2_{y.(\underline{x},\underline{\tau})_q}\right) \tag{3.40}$$

3.4. Bias and Consistency of Mixture Regression Estimators

These mixture regression estimators using multiple auxiliary variables in two phase sampling are biased. However, these biases are negligible for moderate and large samples. It's easily shown that the mixture regression estimators are consistent estimators using multiple auxiliary variables since they are linear combinations of consistent estimators it follows that they are also consistent.

4. Result and Discussion

In this section, we carried out some data analysis using *R* statistical package to compare the performance of mixture regression estimators with already existing estimator in two-phase sampling for finite population that uses one or multiple auxiliary variables or attributes.

In the natural population, the study variable was body fat and auxiliary variables are Thigh circumference and chest circumference while attributes were abdomen and hip circumference.

$$N=252\quad n'=80\quad n=50$$

$$\rho_{yx_1}=0.5596\quad \rho_{yx_1}=0.7026\quad \rho_{Pb_1}=0.5035\quad \rho_{Pb_2}=0.6862$$

Population: The simulated population was a normally distributed with the following parameters $N=600$, $n'=90$, $n=36$ mean = 75 standard deviation = 5

$$\rho_{yx_1}=0.7616\quad \rho_{yx_1}=0.6437\quad \rho_{Pb_1}=0.7020\quad \rho_{Pb_2}=0.6005$$

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.

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

The Table 1 shows percent relative efficiency of proposed and existing estimator with respect to mean per unit estimator for two phase sampling. It is observed that Regression estimators using one auxiliary variables and attributes are more efficient than mean per unit in the two populations. Again, Regression estimators using multiple auxiliary variables and attributes are more efficient than mean per unit and Regression estimators. Finally, Mixture Regression estimators using multiple auxiliary variables and attributes is the most efficient of the five estimators in the two populations since it has the highest percent relative efficiency.

Finally, Table 2 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. Conclusions

The percent relative efficiency is used in sample survey to compare the efficiency of different estimators. The estimator with the highest percent relative efficiency with respect to mean per unit is normally considered to be more efficient compared to the other estimators.

According to Table 1, the proposed Mixture Regression estimators using multiple auxiliary variables and attributes in two-phase sampling has the highest percent relative efficiency compared to mean per unit, Regression estimators using one auxiliary variable and attributes, Regression estimators using multiple auxiliary variables and attributes. This means that the ratio-cum-product estimator in two-phase sampling is the most efficient estimator compared to the estimators that utilize auxiliary variables and attributes.

The Mixture Regression estimators were then extended to two-phase sampling in partial and no information case. In Table 2, we compared the efficiency of full and partial information case to no information case and

Table 1. Relative efficiency of suggested estimator with respect to mean per unit estimator for two phase sampling.

Estimators	Relative efficiency of suggested estimator with respect to mean per unit estimator for two phase sampling	
	Population I	Population II
$\bar{y}_2$	100	100
t_{REX}	135	140
t_{REP}	127	130
t_{MREX}	199	183
t_{MREP}	183	175
$t_{MR(3.0)}$ (proposed)	211	274

Table 2. Comparisons of full, partial and no information cases for proposed mixture regression estimator.

Population	Percent relative efficiency of full and partial to no information			Percent relative efficiency of full to partial in formation case	
Estimators	$t_{MR(3.2)}$	$t_{MR(3.1)}$	$t_{MR(3.0)}$	$t_{MR(3.1)}$	$t_{MR(3.0)}$
1	100	122	158	100	130
2	100	131	163	100	134

found that the two are more efficient than the no information case. We also compared the efficiency of full information case to partial information case and found that the full information case is more efficient than the partial information case.

The proposed Mixture Regression estimator using multiple auxiliary variables and attributes in two-phase sampling is recommended to estimate the finite populations mean for full information case as it outperforms all the other existing estimators for full information using one auxiliary or multiple auxiliary variables and attributes. It also outperforms Mixture Regression estimators using multiple auxiliary variables and attributes in partial and no information cases.

When some auxiliary variables are unknown, the two-phase sampling is recommended**.** If some auxiliary variables are known, the Mixture Regression estimators using multiple auxiliary variables and attributes in partial information case should be used but if all the auxiliary variables and attributes are unknown. Mixture Regression estimators using multiple auxiliary variables in no information case should be used to estimate the finite population mean.

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] 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

[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 Agricultural Science*, **30**, 262-275. http://dx.doi.org/10.1017/S0021859600048012

[4] Watson, D.J. (1937) The Estimation of Leaf Areas. *Journal of Agricultural Science*, **27**, 474. http://dx.doi.org/10.1017/S002185960005173X

[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] 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

[7] 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

[8] Zahoor, A., Muhhamad, H. and Munir, A. (2009) Generalized Multivariate Ratio Estimator Using Multiple Auxiliary Variables for Multi-Phase Sampling. *Pakistan Journal of Statistic*, **26**, 569-583.

[9] Zahoor, A., Muhhamad, H. and Munir, A. (2009) Generalized Regression-Cum-Ratio Estimators for Two Phase Sampling Using Multiple Auxiliary Variables. *Pakistan Journal of Statistics*, **25**, 93-106.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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

[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] 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.

[17] 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

[18] 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

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