

**GENERALIZED REGRESSION-CUM-RATIO ESTIMATORS FOR  
TWO-PHASE SAMPLING USING MULTI-AUXILIARY VARIABLES**

**Zahoor Ahmad<sup>1</sup>, Muhammad Hanif<sup>2</sup> and Munir Ahmad<sup>3</sup>**

<sup>1</sup> University of Gujrat, Gujrat, Pakistan. Email: zahoor\_ahmed\_stat@yahoo.com

<sup>2</sup> Lahore University of Management Sciences, Lahore, Pakistan.  
Email: hanif@lums.edu.pk

<sup>3</sup> National College of Business Administration and Economics, Lahore, Pakistan.  
Email: drmunir@brain.net.pk

**ABSTRACT**

In this paper we suggest three classes of regression-cum-ratio estimators for estimating population mean of variable of interest for two-phase sampling using multi-auxiliary variables for full information, partial information and no information cases. The expressions for mean square errors are derived. Theoretical comparison is given. Special cases of estimators are also identified.

**KEY WORDS**

Regression-cum-ratio estimator; two-phase sampling; auxiliary variable

**1. INTRODUCTION**

The use of auxiliary information is a widely discussed topic in sampling theory to obtain improved designs and precise estimates of some population parameters like mean or variance. It is well known that when the auxiliary information is utilized at the estimation stage. The ratio, product and regression methods are employed in many such situations.

The estimation of the population mean is an unrelenting issue in sampling theory and several efforts have been made to improve the precision of the estimates in the presence of multi-auxiliary variables. A variety of estimators have been proposed following different ideas linking together ratio, product or regression estimators.

Olkin (1958) was the first author to deal with the problem of estimating the mean of a survey variable when auxiliary variables are made available. He suggested the use of information on more than one auxiliary variable, positively correlated with the study variable analogously to Olkin; Singh (1967a) gave a multivariate expression of Murthy's (1964) product estimator, while Raj (1965) suggested a method for using multi-auxiliary variables through a linear combination of single difference estimators. Moreover, Singh (1967b) considered the extension of the ratio-cum-product estimators to multi-auxiliary variables Shukla (1965) suggested a multiple regression estimator while Rao and Mudholkar (1967) proposed a multivariate estimator based on a weighted sum of single ratio and product estimators.

John (1969) suggested two multivariate generalizations of ratio and product estimators which actually reduce to the Olkin's (1958) and Singh's (1967a) estimators. Srivastava (1971) proposed a general ratio-type estimator which generates a large class of estimators including most of the estimators up to that time proposed.

Robinson (1994) proposed a regression estimator ignoring some of the assumptions usually adopted in the literature (see, e.g., Srivastava (1971)), Tracy et al. (1996) and Perri (2004) proposing an alternative to Singh's (1965, 1967b) ratio-cum-product estimators, when two auxiliary variables are available. Ceccon and Diana (1996) provided a multivariate extension of the Naik and Gupta (1991) univariate class of estimators. Agarwal et al. (1997), moving from Raj (1965), illustrated a new approach to form a multivariate difference estimator which does not require the knowledge of any population parameters. Abu-Dayyeh et al. (2003) introduced two estimators which are definitely members of the class proposed by Srivastava (1971), while Kadilar and Cingi (2004, 2005) analyzed combinations of regression type estimators in the case of two auxiliary variables. In the same situation, Perri (2005) proposed some new estimators obtained from Singh's (1965, 1967b) estimators. Pradhan (2005) suggested a chain regression estimator for two-phase sampling using three auxiliary variables when the population mean of one auxiliary variable is unknown and other auxiliary population means are known.

In practical surveys, the problem is to estimate population means of variables of interest. For example, in a typical socio-economic survey conducted in rural areas in Indo-Pak subcontinent, the multiple variables of interests may be size of household, monthly income and expenditure of the household, number of unemployed persons, number of illiterates, number of persons engaged in agriculture, amount of land owned, leased and leased out, number of cattle owned etc. In some situations the auxiliary information may be available through the past census data or conveniently collected. For example in a village land survey, the information on the variables such as area of the village, cultivable area, grazing grounds etc. may be easily obtained through the past census data and may be used to estimate the means of variables of interest.

If we have information on multi-auxiliary variables practically sometimes either information for all these auxiliary is available from population or available for some variables or not available for all auxiliary variables. By considering these practical situations, we suggest general classes of regression-cum-ratio estimators for estimating the population mean of study variable for two-phase sampling using multi-auxiliary variables by considering the following three cases (see Samiuddin and Hanif (2007)).

1. Estimators when information on all auxiliary variables is known for population (Full Information Case).
2. Estimators when information on some auxiliary variables is known for population (Partial Information Case).
3. Estimators when information on all auxiliary variables is unknown for population (No Information Case).

Before suggesting the estimators we provide two-phase sampling scheme and some useful notations and results in the following section.

## 2. TWO-PHASE SAMPLING USING MULTI-AUXILIARY VARIABLES

Consider a population of  $N$  units. Let  $Y$  be the variable for which we want to estimate the population mean and  $X_1, X_2, \dots, X_q$  are  $q$  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.  $x_{(1)i}$  and  $x_{(2)i}$  denote the  $i^{\text{th}}$  auxiliary variables from first and second phase samples respectively and  $y_2$  denote the variable of interest from second phase.  $\bar{X}_i$  and  $C_{x_i}$  denote the population means and coefficient of variation of  $i^{\text{th}}$  auxiliary variables respectively and  $\rho_{yx_i}$  denotes the population correlation coefficient of  $Y$  and  $X_i$ . Further let  $\theta_1 = \frac{1}{n_1} - \frac{1}{N}$ ,  $\theta_2 = \frac{1}{n_2} - \frac{1}{N}$ ,  $y_{(2)} = Y + e_{y_{(2)}}$ ,  $x_{(1)i} = X_i + e_{x_{(1)i}}$  and  $x_{(2)i} = X_i + e_{x_{(2)i}}$ ; ( $i = 1, 2, \dots, k$ ) where  $e_{y_{(2)}}$ ,  $e_{x_{(1)i}}$  and  $e_{x_{(2)i}}$  are sampling errors and are of very small quantities. We assume that  $E_2(e_{y_{(2)}}) = E_1(e_{x_{(1)i}}) = E_2(e_{x_{(2)i}}) = 0$ . Then for simple random sampling without replacement for both first and second phases we write by using phase wise operation of expectations as:

$$\begin{aligned} E_2(e_{y_2})^2 &= \left(1 - \frac{n_2}{N}\right) S_Y^2, \quad E_2(\bar{e}_{y_2})^2 = \left(1 - \frac{n_2}{N}\right) \frac{S_Y^2}{n_2} = \theta_2 \bar{Y}^2 C_Y^2, \\ E_2(e_{y_2} e_{x_{(2)i}}) &= \left(1 - \frac{n_2}{N}\right) S_{YX_i} = \theta_2 \bar{Y} \bar{X}_i C_Y C_{x_i} \rho_{yx_i}, \\ E_2(\bar{e}_{y_2} \bar{e}_{x_{(2)i}}) &= \left(1 - \frac{n_2}{N}\right) \frac{S_{YX_i}}{n_2} = \theta_2 \bar{Y} \bar{X}_i C_Y C_{x_i} \rho_{yx_i}, \\ E_1 E_{2|1} \left[ e_{y_2} (e_{x_{(1)i}} - e_{x_{(2)i}}) \right] &= E_1 E_{2|1} (e_{y_2} e_{x_{(1)i}}) - E_2 (e_{y_2} e_{x_{(2)i}}) \\ &= \left(1 - \frac{n_1}{N}\right) S_{YX_i} - \left(1 - \frac{n_2}{N}\right) S_{YX_i} = \frac{1}{N} (n_2 - n_1) S_{YX_i}, \\ E_1 E_{2|1} \left[ \bar{e}_{y_2} (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right] &= \left(1 - \frac{n_1}{N}\right) \frac{S_{YX_i}}{n_1} - \left(1 - \frac{n_2}{N}\right) \frac{S_{YX_i}}{n_2} = (\theta_1 - \theta_2) \bar{Y} \bar{X}_i C_Y C_{x_i} \rho_{yx_i}. \end{aligned}$$

Similarly

$$\begin{aligned} E_1 E_{2|1} \left[ \bar{e}_{x_{(2)i}} (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right] &= (\theta_1 - \theta_2) \sigma_{x_i}^2 = (\theta_1 - \theta_2) \bar{X}_i^2 C_{x_i}^2, \\ E_1 E_{2|1} \left[ \bar{e}_{x_{(1)i}} (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right] &= 0, \\ E_1 E_{2|1} \left( \bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}} \right)^2 &= (\theta_2 - \theta_1) \sigma_{x_i}^2 = (\theta_2 - \theta_1) \bar{X}_i^2 C_{x_i}^2, \end{aligned}$$

$$E_1 E_2 | \left[ \left( \bar{e}_{x(1)i} - \bar{e}_{x(2)i} \right) \left( \bar{e}_{x(1)j} - \bar{e}_{x(2)j} \right) \right] = (\theta_2 - \theta_1) \sigma_{x_i x_j}$$

$$= (\theta_2 - \theta_1) \bar{X}_i \bar{X}_j C_{x_i} C_{x_j} \rho_{x_i x_j}; (i \neq j),$$

and

$$E_1 E_2 | \left[ \left( \bar{e}_{x(2)i} \right) \left( \bar{e}_{x(1)j} - \bar{e}_{x(2)j} \right) \right] = (\theta_1 - \theta_2) \sigma_{x_i x_j} = (\theta_1 - \theta_2) \bar{X}_i \bar{X}_j C_{x_i} C_{x_j} \rho_{x_i x_j}, (i \neq j).$$

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

- $|R|_{y x_q}$  Determinant of population correlation matrix of variables  $y, x_1, x_2, \dots, x_{q-1}$  and  $x_q$ .
- $|R|_{y x_i}$  Determinant of  $i^{th}$  minor of  $|R|_{y x_q}$  corresponding to the  $i^{th}$  element of  $\rho_{y x_i}$ .
- $\rho_{y x_s}^2$  Denotes the multiple coefficient of determination of  $y$  on  $x_1, x_2, \dots, x_{r-1}$  and  $x_r$ .
- $\rho_{y x_q}^2$  Denotes the multiple coefficient of determination of  $y$  on  $x_1, x_2, \dots, x_{q-1}$  and  $x_q$ .
- $|R|_{x_s}$  Determinant of population correlation matrix of variables  $x_1, x_2, \dots, x_{r-1}$  and  $x_r$ .
- $|R|_{x_q}$  Determinant of population correlation matrix of variables  $x_1, x_2, \dots, x_{q-1}$  and  $x_q$ .
- $|R|_{y_i x_s}$  Determinant of the correlation matrix of  $y_i, x_1, x_2, \dots, x_{r-1}$  and  $x_r$ .
- $|R|_{y_i x_q}$  Determinant of the correlation matrix of  $y_i, x_1, x_2, \dots, x_{q-1}$  and  $x_q$ .
- $|R|_{y_i y_j x_s}$  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  $x_r$ , for  $(i \neq j)$ .
- $|R|_{y_i y_j x_q}$  Determinant of the minor corresponding to  $\rho_{y_i y_j}$  of the correlation matrix of  $y_i, y_j, x_1, x_2, \dots, x_{q-1}$  and  $x_q$ , for  $(i \neq j)$ .

### 2.1 Result: 1

The following result will help us in deriving the mean square errors of suggested estimators

$$\frac{|R|_{y x_q}}{|R|_{x_q}} = (1 - \rho_{y x_q}^2). \quad [\text{Arora and Lal (1989)}]$$

## 2.2 Regressions-Cum-Ratio Estimator (Full Information Case)

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 Regression-cum-Ratio technique for two-phase sampling, a generalized estimator for estimating population mean of study variable  $\bar{Y}$  with the use of multi-auxiliary variables are suggested as:

$$\begin{aligned} t_1 &= \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i (\bar{X}_i - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\gamma_i} \\ &= \left[ \bar{Y} + \bar{e}_{y_2} - \sum_{i=1}^r \alpha_i \bar{e}_{x_{(2)i}} \right] \prod_{i=r+1}^{r+s=q} \left( 1 + \frac{\bar{e}_{x_{(2)i}}}{\bar{X}_i} \right)^{-\gamma_i}. \end{aligned}$$

Ignoring second and higher terms for each expansion of product and after simplification, we write  $t_1 = \left[ \bar{Y} + \bar{e}_{y_2} - \sum_{i=1}^r \alpha_i \bar{e}_{x_{(2)i}} - \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i \bar{e}_{x_{(2)i}} \right]$ .

The mean square error is

$$MSE(t_1) = E_2 \left[ \bar{e}_{y_2} - \sum_{i=1}^r \alpha_i \bar{e}_{x_{(2)i}} - \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i \bar{e}_{x_{(2)i}} \right]^2. \quad (3.2.1)$$

The optimum values of  $\alpha_i$  and  $\gamma_i$  for which the mean square error of estimator  $t_1$  is minimum to term of  $o(1/n)$  are

$$\alpha_i = (-1)^{i+1} \frac{\bar{Y}}{\bar{X}_i} \frac{C_y}{C_{x_i}} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} = (-1)^{i+1} \beta_{yx_i \cdot x_q} \quad (i = 1, 2, \dots, r \text{ and } r + s = q)$$

and

$$\gamma_i = (-1)^{i+1} \frac{C_y}{C_{x_i}} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} = (-1)^{i+1} \frac{\bar{X}_i}{\bar{Y}} \beta_{yx_i \cdot x_q} \quad (i = r+1, r+2, \dots, r+s \text{ and } r + s = q).$$

The unknown constants are related to the partial regression coefficients of study variable and auxiliary variables. If these partial regression coefficients are not known, these will be estimated from second phase because the estimator  $t_1$  utilizes the information on  $q$  auxiliary variables and study variable from second phase sample.

Using normal equations that are used to find the optimum values of  $\alpha_i$  and  $\gamma_i$ , (3.2.1) can be written in simplified form as:

$$\begin{aligned}
MSE(t_1) &= E_2 \left[ \bar{e}_{y_2} \left( \bar{e}_{y_2} - \sum_{i=1}^r \alpha_i \bar{e}_{x_{(2)i}} - \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i \bar{e}_{x_{(2)i}} \right) \right] \\
&= E_2 (\bar{e}_{y_2}^2) - \sum_{i=1}^r \alpha_i E_2 (\bar{e}_{y_2} \bar{e}_{x_{(2)i}}) - \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i E_2 (\bar{e}_{y_2} \bar{e}_{x_{(2)i}}) \\
&= \theta_2 \bar{Y}^2 C_y^2 - \sum_{i=1}^r \alpha_i \theta_2 \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i} - \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i \theta_2 \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i}.
\end{aligned}$$

Using the values of  $\alpha_i$  and  $\gamma_i$  and after simplification, we get:

$$\begin{aligned}
MSE(t_1) &= \bar{Y}^2 C_y^2 \left[ \theta_2 - \theta_2 \sum_{i=1}^r (-1)^{i+1} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \rho_{yx_i} - \theta_2 \sum_{i=r+1}^{r+s=q} (-1)^{i+1} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \rho_{yx_i} \right] \\
&= \bar{Y}^2 C_y^2 \left[ \theta_2 + \theta_2 \sum_{i=1}^q (-1)^i \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \rho_{yx_i} \right] \\
&= \frac{\theta_2 \bar{Y}^2 C_y^2}{|R|_{x_q}} \left[ |R|_{x_q} - \rho_{yx_1} |R_{yx_1}|_{yx_q} + \rho_{yx_2} |R_{yx_2}|_{yx_q} \right. \\
&\quad \left. - \rho_{yx_3} |R_{yx_3}|_{yx_q} + \dots + (-1)^q \rho_{yx_q} |R_{yx_q}|_{yx_q} \right]
\end{aligned}$$

or

$$MSE(t_1) = \theta_2 \bar{Y}^2 C_y^2 \frac{|R|_{yx_q}}{|R|_{x_q}}.$$

Using Result 1, we get:

$$MSE(t_1) = \theta_2 \bar{Y}^2 C_y^2 (1 - \rho_{y.x_q}^2).$$

### 2.3 Regressions-Cum-Ratio Estimator (Partial Information Case)

In this case suppose we have no information on all  $q$  auxiliary variables but only for  $r$  auxiliary variables from population. Considering Regression-Cum-Ratio technique of estimating technique, the population mean of study variable  $\bar{Y}$  can be estimated for two-phase sampling using multi-auxiliary variables as:

$$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i \left( \bar{x}_{(1)i} - \bar{x}_{(2)i} \right) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i}$$

$$= \left[ \bar{Y} + \bar{e}_{y_2} + \sum_{i=1}^r \alpha_i'' (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right] \prod_{i=r+1}^{r+s=q} \left( 1 + \frac{\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}}{\bar{X}_i} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( 1 + \frac{\bar{e}_{x_{(2)i}}}{\bar{X}_i} \right)^{-\delta_i''}.$$

Ignoring second and higher terms for each expansion of products and after simplification we write

$$t_2 = \left[ \bar{Y} + \bar{e}_{y_2} + \sum_{i=1}^r \alpha_i'' (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) + \sum_{i=r+1}^{r+s} \gamma_i'' \frac{\bar{Y}}{\bar{X}_i} (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) - \sum_{i=r+1}^{r+s} \delta_i'' \frac{\bar{Y}}{\bar{X}_i} \bar{e}_{x_{(2)i}} \right].$$

The mean square error of  $t_2$  is written as

$$\begin{aligned} MSE(t_2) = E_1 E_2 / 1 \left[ \bar{e}_{y_2} + \sum_{i=1}^r \alpha_i'' (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right. \\ \left. + \sum_{i=r+1}^{r+s} \gamma_i'' \frac{\bar{Y}}{\bar{X}_i} (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) - \sum_{i=r+1}^{r+s} \delta_i'' \frac{\bar{Y}}{\bar{X}_i} \bar{e}_{x_{(2)i}} \right]^2 \end{aligned} \quad (3.3.1)$$

The optimum values of  $\alpha_i''$ ,  $\gamma_i''$  and  $\delta_i''$  for which the mean square error of  $t_2$  is minimum to term of  $o(1/n)$  are:

$$\alpha_i'' = (-1)^{i+1} \frac{\bar{Y}}{\bar{X}_i} \frac{C_y}{C_{x_i}} \frac{|R_{yx_i}|_{y_{x_q}}}{|R|_{x_q}} = (-1)^{i+1} \beta_{yx_i \cdot x_q}, \quad (i = 1, 2, \dots, r).$$

$$\begin{aligned} \gamma_i'' &= (-1)^{i+1} \frac{C_y}{C_{x_i}} \left[ \frac{|R_{yx_i}|_{y_{x_q}}}{|R|_{x_q}} - \frac{|R_{yx_i}|_{y_{x_s}}}{|R|_{x_s}} \right] \\ &= (-1)^{i+1} \frac{\bar{X}_i}{\bar{Y}} (\beta_{yx_i \cdot x_q} - \beta_{yx_i \cdot x_s}), \quad (i = r+1, r+2, \dots, r+s) \end{aligned}$$

and

$$\delta_i'' = (-1)^{i+1} \frac{C_y}{C_{x_i}} \frac{|R_{yx_i}|_{y_{x_s}}}{|R|_{x_s}} = (-1)^{i+1} \beta_{yx_i \cdot x_s}, \quad i = r+1, r+2, \dots, r+s.$$

The optimum values are related to the partial regression coefficients of variable of interest and auxiliary variables. Usually these partial regression coefficients are unknown then these can be estimated from sample data. The estimator  $t_2$  utilizes the information on  $q$  auxiliary variables from both first and second phase. Keenly observing the estimators  $t_2$  the optimum values of unknown constants  $\alpha_i''$  and  $\gamma_i''$  will be estimated from first phase sample and  $\delta_i''$  from the second phase. Using normal equations that are

used to find the optimum values of  $\alpha_i''$ ,  $\gamma_i''$  and  $\delta_i''$ , (3.3.1) can be written in a simplified form as:

$$\begin{aligned}
MSE(t_2) &= E_1 E_{2/1} \left[ \bar{e}_{y_2} \left( \bar{e}_{y_2} + \sum_{i=1}^r \alpha_i'' \left( \bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}} \right) \right. \right. \\
&\quad \left. \left. + \sum_{i=r+1}^{r+s} \gamma_i'' \frac{\bar{Y}}{\bar{X}_i} \left( \bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}} \right) - \sum_{i=r+1}^{r+s} \delta_i'' \frac{\bar{Y}}{\bar{X}_i} \bar{e}_{x_{(2)i}} \right) \right] \\
&= E_1 E_{2/1} \left( \bar{e}_{y_2}^2 \right) + \sum_{i=1}^r \alpha_i'' E_1 E_{2/1} \left\{ \bar{e}_{y_2} \left( \bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}} \right) \right\} \\
&\quad + \sum_{i=r+1}^{r+s} \gamma_i'' \frac{\bar{Y}}{\bar{X}_i} E_1 E_{2/1} \left\{ \bar{e}_{y_2} \left( \bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}} \right) \right\} - \sum_{i=r+1}^{r+s} \delta_i'' \frac{\bar{Y}}{\bar{X}_i} E_1 E_{2/1} \left( \bar{e}_{y_2} \bar{e}_{x_{(2)i}} \right) \\
&= \theta_2 \bar{Y}^2 C_y^2 + (\theta_1 - \theta_2) \sum_{i=1}^r \alpha_i'' \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i} \\
&\quad + (\theta_1 - \theta_2) \sum_{i=r+1}^{r+s} \gamma_i'' \frac{\bar{Y}}{\bar{X}_i} \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i} - \theta_2 \sum_{i=r+1}^{r+s} \delta_i'' \frac{\bar{Y}}{\bar{X}_i} \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i}.
\end{aligned}$$

Using the values of  $\alpha_i''$ ,  $\gamma_i''$  and  $\delta_i''$  and after simplification we get:

$$\begin{aligned}
MSE(t_2) &= \bar{Y}^2 C_y^2 \left[ \theta_2 - (\theta_1 - \theta_2) \sum_{i=1}^r (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \right. \\
&\quad \left. - (\theta_1 - \theta_2) \sum_{i=r+1}^{r+s} (-1)^i \rho_{yx_i} \left\{ \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} - \frac{|R_{yx_i}|_{yx_s}}{|R|_{x_s}} \right\} + \theta_2 \sum_{i=r+1}^{r+s} (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_s}}{|R|_{x_s}} \right] \\
&= \bar{Y}^2 C_y^2 \left[ (\theta_2 - \theta_1) + (\theta_2 - \theta_1) \sum_{i=1}^r (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \right. \\
&\quad \left. + (\theta_2 - \theta_1) \sum_{i=r+1}^{r+s} (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} + \theta_1 + \theta_1 \sum_{i=r+1}^{r+s} (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_s}}{|R|_{x_s}} \right]
\end{aligned}$$

or

$$= \bar{Y}^2 C_y^2 \left[ (\theta_2 - \theta_1) \left( 1 + \sum_{i=1}^q (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \right) + \theta_1 \left( 1 + \sum_{i=r+1}^{r+s} (-1)^i \rho_{yx_i} \frac{|R_{yx_i}|_{yx_s}}{|R|_{x_s}} \right) \right],$$

or

$$MSE(t_2) = \bar{Y}^2 C_y^2 \left[ \theta_2 \frac{|R|_{y x_q}}{|R|_{x_q}} + \theta_1 \left( \frac{|R|_{y x_s}}{|R|_{x_s}} - \frac{|R|_{y x_q}}{|R|_{x_q}} \right) \right].$$

Using Result 1 we get:

$$MSE(t_2) = \bar{Y}^2 C_y^2 \left[ \theta_2 (1 - \rho_{y \cdot x_q}^2) + \theta_1 (\rho_{y \cdot x_q}^2 - \rho_{y \cdot x_s}^2) \right].$$

#### 2.4 Regressions-Cum-Ratio Estimator (No Information Case)

We consider the following regression-cum-ratio estimator for two-phase sampling using multi-auxiliary variables for estimating the population mean  $\bar{Y}$  when information on all auxiliary variables is not available from a population as:

$$\begin{aligned} t_3 &= \left[ \bar{y}_2 + \sum_{i=1}^r \alpha'_i (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma'_i} \\ &= \left[ \bar{Y} + \bar{e}_{y_2} + \sum_{i=1}^r \alpha'_i (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right] \prod_{i=r+1}^{r+s=q} \left( 1 + \frac{\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}}{\bar{X}_i} \right)^{\gamma'_i}. \end{aligned}$$

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

$$t_3 = \left[ \bar{Y} + \bar{e}_{y_2} + \sum_{i=1}^r \alpha'_i (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) + \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma'_i (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right].$$

The mean square error is

$$MSE(t_3) = E_2 \left[ \bar{e}_{y_2} + \sum_{i=1}^r \alpha'_i (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) + \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma'_i (\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}}) \right]^2. \quad (3.4.1)$$

The optimum values of  $\alpha'_i$  and  $\gamma'_i$  for which the mean square error of  $t_3$  is minimum to the order  $o(1/n)$  are:

$$\alpha'_i = (-1)^{i+1} \frac{\bar{Y}}{\bar{X}_i} \frac{C_y}{C_{x_i}} \frac{|R_{y x_i}|_{y x_q}}{|R|_{x_q}} = (-1)^{i+1} \beta_{y x_i \cdot x_q} \quad (i = 1, 2, \dots, r)$$

and

$$\gamma'_i = (-1)^{i+1} \frac{C_y}{C_{x_i}} \frac{|R_{y x_i}|_{y x_q}}{|R|_{x_q}} = (-1)^{i+1} \frac{\bar{X}_i}{\bar{Y}} \beta_{y x_i \cdot x_q}, \quad (i = r+1, r+2, \dots, r+s).$$

In this case the optimum values are also related to the partial regression coefficients of study variable and auxiliary variables. Mostly these partial regression coefficients are

unknown but these can be estimated from sample data. The estimator  $t_3$  utilizes the information on  $q$  auxiliary variables from both first and second phase. Analyzing the estimators  $t_3$  the optimum values of unknown constants  $\alpha_i$  and  $\gamma_i$  will be estimated from first phase sample in the form of sample regression coefficients.

(3.4.1) is also written in a simplified form as:

$$\begin{aligned} MSE(t_3) &= E_2(\bar{e}_{y_2}^2) + \sum_{i=1}^r \alpha_i' E_2\left(\bar{e}_{y_2}(\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}})\right) - \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i' E_2\left(\bar{e}_{y_2}(\bar{e}_{x_{(1)i}} - \bar{e}_{x_{(2)i}})\right) \\ &= \theta_2 \bar{Y}^2 C_y^2 + (\theta_1 - \theta_2) \sum_{i=1}^r \alpha_i' \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i} + (\theta_1 - \theta_2) \sum_{i=r+1}^{r+s=q} \frac{\bar{Y}}{\bar{X}_i} \gamma_i' \bar{Y} \bar{X}_i C_y C_{x_i} \rho_{yx_i}. \end{aligned}$$

Using the values of  $\alpha_i$  and  $\gamma_i$  and after simplification we get:

$$\begin{aligned} MSE(t_3) &= \bar{Y}^2 C_y^2 \left[ \theta_2 + (\theta_1 - \theta_2) \sum_{i=1}^r (-1)^{i+1} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \rho_{yx_i} \right. \\ &\quad \left. + (\theta_1 - \theta_2) \sum_{i=r+1}^{r+s=q} (-1)^{i+1} \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \rho_{yx_i} \right] \\ &= \bar{Y}^2 C_y^2 \left[ \theta_2 + (\theta_2 - \theta_1) \sum_{i=1}^q (-1)^i \frac{|R_{yx_i}|_{yx_q}}{|R|_{x_q}} \rho_{yx_i} \right] \\ &= \bar{Y}^2 C_y^2 \left[ (\theta_2 - \theta_1) \frac{|R|_{yx_q}}{|R|_{x_q}} + \theta_1 \right]. \end{aligned}$$

Using Result1, we get:

$$MSE(t_3) = \bar{Y}^2 C_y^2 \left[ (\theta_2 - \theta_1) (1 - \rho_{y \cdot x_q}^2) + \theta_1 \right] = \bar{Y}^2 C_y^2 \left[ \theta_2 (1 - \rho_{y \cdot x_q}^2) + \theta_1 \rho_{y \cdot x_q}^2 \right].$$

### 3. THEORETICAL COMPARISON OF NEW ESTIMATORS

The MSE criterion is most common for comparing various estimators [Lee and Peddada (1987) and Cox and Hinkley (1974)]. We suggest three estimators in this paper in which first one is for full information case, second one is for partial information case and last one is for no information case. The estimator for full information case is more efficient than the estimator for partial information case and the estimator for partial information case is more efficient than for the no information case. It can be checked by comparing their MSE's as:

$$MSE(t_1) - MSE(t_2) = -\theta_1 \bar{Y}^2 C_y^2 (\rho_{y.x_q}^2 - \rho_{y.x_s}^2) < 0; \text{ as } q > s$$

and  $MSE(t_2) - MSE(t_3) = -\theta_1 \bar{Y}^2 C_y^2 \rho_{y.x_s}^2 < 0.$

**4. SPECIAL CASES**

We summarize the results for the three cases as follows:

Proposed Estimators	Unknown Constants		Special Cases
<b>I. Full Information Case</b>	$\alpha_i$	$\gamma_i$	
$t_1 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i (\bar{X}_i - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\gamma_i}$	0	$\gamma_i$	A class of ratio estimators with s auxiliary variables for Full Information Case
$t_1 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i (\bar{X}_i - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\gamma_i}$	$\alpha_i$	0	A class of regression estimators with r auxiliary variables for Full Information Case
<b>II. Partial Information Case</b>	$\alpha_i''$	$\gamma_i''$	$\delta_i''$
$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i'' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''}$	0	$\gamma_i''$	$\delta_i''$
$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i'' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''}$	$\alpha_i''$	0	$\delta_i''$
$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i'' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''}$	$\alpha_i''$	$\gamma_i''$	0
$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i'' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''}$	$\alpha_i''$	0	0

Proposed Estimators	Unknown Constants			Special Cases
$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i'' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''}$	0	$\gamma_i''$	0	A class of ratio estimators with s auxiliary variables for No Information Case
$t_2 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i'' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i''} \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{X}_i}{\bar{x}_{(2)i}} \right)^{\delta_i''}$	0	0	$\alpha_i''$	A class of ratio estimators with s auxiliary variables for Partial Information Case
<b>III. No Information Case</b>				
$t_3 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i'}$	0	$\gamma_i'$		A class of ratio estimators with s auxiliary variables for No Information Case
$t_3 = \left[ \bar{y}_2 + \sum_{i=1}^r \alpha_i' (\bar{x}_{(1)i} - \bar{x}_{(2)i}) \right] \prod_{i=r+1}^{r+s=q} \left( \frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\gamma_i'}$	$\alpha_i'$		0	A class of regression estimators with r auxiliary variables for No Information Case

Obviously the classes of special cases are not efficient than suggested classes of estimators.

### 5. NUMERICAL ILLUSTRATION

Description of populations is given in Table 1 and mean square errors of suggested estimators are given in Table 2.

**Table 1:**  
**Description of Population**

<b>Source</b>	"Measurement of four characters of: Flucus Religiosament" by Pradhan (2000)						
<b>y</b>	Length of petiole						
$x_1$	Length of lamina (blade) of the leaf						
$x_2$	Width of the leaf at its widest paint						
$x_3$	Width of leaf half way along the blade						
	N	$\rho_{yx_1}$	$\rho_{yx_2}$	$\rho_{yx_3}$	$\rho_{x_1x_2}$	$\rho_{x_1x_3}$	$\rho_{x_2x_3}$
	160	0.5423	0.6166	0.2704	0.8568	0.7424	0.8027

**Table 2:**  
**Mean Square Errors of Estimators**

Estimator	Auxiliary Variables for which Information is Known for the Population	Auxiliary Variables for which Information is unknown for the Population	Relative Efficiency when $N = 160, n_1 = 50, n_2 = 20$
$t_3$ (NIC)	-	$x_1, x_2, x_3$	100
$t_2$	$t_{21}$ $x_2, x_3$	$x_1$	129.34
PIC	$t_{22}$ $x_1, x_3$	$x_2$	123.46
	$t_{23}$ $x_1, x_2$	$x_3$	156.65
	$t_{24}$ $x_1,$	$x_2, x_3$	168.07
	$t_{25}$ $x_2$	$x_1, x_3$	149.43
	$t_{26}$ $x_3$	$x_1, x_2$	164.12
$t_1$ (FIC)	$x_1, x_2, x_3$	-	135.73

In Table 2 we provide MSE's of eight estimators, first estimator is for full information case, last estimator is for no information case and other six estimators are for partial information case with all possible combinations of auxiliary variables with known and unknown information from population. In theoretical comparisons  $t_1$  is more efficient than  $t_2$  and  $t_2$  is more efficient than  $t_3$ . But in empirical comparison, we see that a special case of partial information i.e.  $t_{24}$  performs better than all others. It means that population characteristics like mean, coefficient of variation, variances, sample sizes of both phases and especially correlation coefficients of study variable with auxiliary variables and correlation coefficients within auxiliary variables count a lot for suggesting an estimator for use in real life situations. This can be adequately judged by considering at least ten different types of natural populations and MSE's should be calculated for study variable in the presence of at least five auxiliary variables.

#### REFERENCES

1. Abu-Dayyeh, W.A., Ahmed, M.S., Ahmed, R.A. and Muttlak, H.A. (2003). Some estimators of finite population mean using auxiliary information. *App. Math. and Compu.*, 139, 287-298.
2. Agarwal, S.K., Sharma, U.K. and Kashyap, S. (1997). A new approach to use multivariate auxiliary information in sample surveys. *J. Statist. Plann. and Infer.*, 60, 261-267.
3. Arora, S. and Lal, Bansi (1989). *New Mathematical Statistics*. Satya Prakashan, New Delhi.
4. Ceccon, C. and Diana, G. (1996). Classi di stimatori di tipo rapporto in presenza di pi' u variabili ausiliarie, *Proceedings of Italian Statistical Society*, XXXVIII Scientific Meeting, Rimini, 9-13 April 1996, 119-126.
5. Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Champman and Hall, London.

6. John, S. (1969). On multivariate ratio and product estimators. *Biometrika*, 56, 533-536.
7. Kadilar, C. and Cingi, H. (2004). Estimator of a population mean using two auxiliary variables in simple random sampling. *Int. Math. J.*, 5, 357-367.
8. Kadilar, C. and Cingi, H. (2005). A new estimator using two auxiliary variables, *App. Math. and Compu.*, 162, 901-908.
9. Lee, Carl M.S. and Peddada, S.D. (1987). Comparisons of some common ratio estimators using Pitman measure of nearness. *Comm. Statist. Theo Math.*, 16(7), 2017-2027.
10. Murthy, M.N. (1964). Product method of estimation. *Sankhya*, 26, 294-307.
11. Naik, V.D. and Gupta, P.C. (1991). A general class of estimators for estimating population means using auxiliary information. *Metrika*, 38, 11-17.
12. Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika*, 45, 154-165.
13. Perri, P.F. (2004). Alcune considerazioni sull'efficienza degli stimatori rapporto-cum-prodotto. *Statistica and Applicazioni*, 2(2), 59-75.
14. Perri, P.F. (2005). Combining two auxiliary variables in ratio-cum-product type estimators. *Proceedings of Italian Statistical Society, Intermediate Meeting on Statistics and Environment*, Messina, 21-23 September 2005, 193-196.
15. Pradhan, B.K. (2000). *Some problems of estimation in multi-phase sampling*. Ph.D. Dissertation (unpublished), Utkal University, Orissa, India.
16. Pradhan, B.K. (2005). A Chain Regression Estimator in Two Phase Sampling Using Multi-auxiliary Information. *Bull. Malays. Math. Sci. Soc.*, (2) 28(1), 81-86.
17. Raj, D. (1965). On a method of using multi-auxiliary information in sample surveys. *J. Amer. Statist. Assoc.*, 60, 309, 270-277.
18. Rao, P.S.R.S. and Mudholkar, G.S. (1967). Generalized multivariate estimators for the mean of a finite population. *J. Amer. Statist. Assoc.*, 62, 1009-1012.
19. Robinson, P.M. (1994). A class of estimators for the mean of a finite population using auxiliary information. *Sankhya*, B, 56, 389-399.
20. Samiuddin, M. and Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. *Pak. J. Statist.*, 23(2), 99-118.
21. Shukla, G.K. (1965). Multivariate regression estimate. *J. Indian Statist. Assoc.*, 3, 202-211.
22. Singh, M.P. (1965). On the estimation of ratio and product of population parameters. *Sankhya*, Ser. C, 27, 321-328.
23. Singh, M.P. (1967a). Multivariate product method of estimation for finite populations. *J. Ind. Soc. of Agricul. Statist.*, 31, 375-378.
24. Singh, M.P. (1967b). Ratio cum product method of estimation. *Metrika*, 12, 34-43.
25. Srivastava, S.K. (1971). A generalized estimator for the mean of a finite population using multi-auxiliary information. *J. Amer. Statist. Assoc.*, 66, 404-407.
26. Tracy, D.S., Singh, H.P. and Singh, R. (1996). An alternative to the ratio-cum-product estimator in sample surveys. *J. Statist. Plann. and Infer.*, 53, 375-387.