

## A CLASS OF RATIO-TYPE ESTIMATORS UNDER TWO-PHASE SAMPLING IN THE PRESENCE OF TWO AUXILIARY VARIABLES

P. A. PATEL

*Department of Statistics, Sardar Patel University, Gujarat, India*  
*Email: patelpraful\_a@yahoo.co.in*

F. H. SHAH\*

*Department of Statistics, Sardar Patel University, Gujarat, India*  
*Email: shah\_fagun@yahoo.co.in*

### SUMMARY

This paper deals with the estimation of population mean under two-phase sampling. Utilizing information on two-auxiliary variables, a class of estimators for estimating the finite population mean is proposed, and its properties, up to the first order of approximation, are studied. Various estimators are suggested as special cases of this class. The performance of the suggested estimators is compared with some contemporary estimators of population mean through numerical illustrations carried over existing datasets of some natural populations. Also, a small scale Monte Carlo simulation is carried out for the empirical comparison.

*Keywords and phrases:* Auxiliary variables, bias, mean square error (MSE), optimum estimation, empirical study, simulation, two-phase sampling.

*AMS Classification:* 62D05; 65C05

## 1 Introduction

In many sample surveys, either full or partial information on auxiliary variables is available prior to the survey or may be collected by allocating a part of the survey budget. Such information can be used either at design stage or at estimation stage or at both stages to improve the precision of estimation of the population parameters of the study variable  $y$ . The problem of estimation of the population mean  $\bar{Y}$  or total  $Y$  of  $y$ , incorporating auxiliary information at estimation stage, has been considered by many authors. For instance, the usual ratio and regression estimators of the population mean of  $y$  under simple random sampling and without replacement (SRSWOR) sampling utilize auxiliary information at estimation stage. Many authors, including Upadhyaya and Singh (1999), Singh and

---

\* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

Taylor (2003), Kadilar and Cingi (2005), Khoshnevisan et al. (2007) among others, have extended these estimators inserting some known parameters of the auxiliary variable  $x$  such as mean ( $\bar{X}$ ), standard deviation ( $\sigma_x$ ), coefficient of variation ( $C_x$ ), skewness ( $\beta_1(x)$ ), kurtosis ( $\beta_2(x)$ ), and correlation coefficient ( $\rho_{yx}$ ).

Use of auxiliary information at estimation stage under two-phase sampling is common in survey sampling. Under partial information of the auxiliary variable  $x$ , for the ordinary ratio and regression estimators of  $\bar{Y}$ , the mean  $\bar{X}$  is estimated by the mean  $\bar{x}'$  of the preliminary sample. Many often, extra auxiliary variable  $z$  closely related to  $x$  but compared to  $x$  remotely relate to  $y$  is available. For example,  $x$  is current value of a characteristic and  $z$  is some previous value of the same characteristic (i.e.  $\rho_{yz} < \rho_{xz}$ ). For such sampling scheme, various estimators have been proposed in the literature. For example, Chand (1975) suggested a chain ratio estimator under the assumption that the relationship between the main auxiliary variable  $x$  and addition auxiliary variable  $z$  can be expressed using a straight line passing through the origin. Kiregyera (1980, 1984), noting that Chand's estimator is not preferable when the regression of  $x$  on  $z$  is linear but does not pass through the origin, suggested the ratio-to-regression, ratio-in-regression and regression-in-regression estimators. Mukerjee et al. (1987) proposed a series of estimators of the form regression-in-regression following Kiregyera (1980) technique. Srivastava et al. (1990) suggested a class of estimators of population mean using multiple auxiliary variables. Sahoo et al. (1993) developed an alternative approach for estimation in double sampling when the population mean of the main auxiliary variable is known. Tracy et al. (1996) introduced a product estimator using two auxiliary variables in the simple random sampling setting. Singh and Espejo (2007) constructed a composite estimator using the ratio and product estimators. Gupta and Shabbir (2007) developed a ratio-cum-ratio estimator using transformed auxiliary variables. Many ratio-cum-ratio type estimators using the know parameters of the additional auxiliary variable have been proposed in the literature. Using these estimators Singh et al. (2011) proposed a group of composite estimators. Shukla et al. (2012) discussed estimation of population mean using two auxiliary sources in sample surveys. Choudhury and Singh (2012) developed a chain ratio-product type estimators for estimating population mean using two auxiliary variables under double sampling scheme, when the information on another additional auxiliary characteristic is available along with the main auxiliary characteristic.

In this paper we seek to estimate the population mean  $\bar{Y}$  of the study variable  $y$ , incorporating auxiliary information on  $x$  and  $z$  variables at the estimation stage using a two-phase sampling scheme, and suggest a class of estimators which include many of the known estimators described above. This paper is organized as follows. In Section 2 we review available estimators and and propose a new class of estimators. Bias and MSE of the estimators are obtained up to the first order of approximation in Section 3. Results from numerical examples and a simulation study are presented in Section 4. Conclusions are given in Section 5.

## 2 The Proposed Class of Estimators

Consider a finite population  $U$  of  $N$  units with characteristics  $(y, x, z)$ , population means  $(\bar{Y}, \bar{X}, \bar{Z})$ , population coefficients of variation  $(C_y, C_x, C_z)$  and correlation coefficients  $(\rho_{yx}, \rho_{yz}, \rho_{xz})$ . We are interested in estimation of  $\bar{Y}$  under two-phase sampling in which a preliminary random sample  $s'$  of size  $n'$  is drawn from the population using simple random sampling without replacement (srswor) and a subsample  $s$  of size  $n$  is drawn using the same sampling design. Assume that the information on  $z_i, i \in U$ ,  $x_i, i \in s'$  and  $y_i, i \in s$ , is available.

Denote  $\bar{u}'$  ( $\bar{u}$ ) is the first-phase (second-phase) sample mean of  $u$  variable,  $u = y, x, z$ . Under two-phase sampling the ratio estimator (Cochran, 1977) is given by

$$\widehat{Y}_{Rd} = \frac{\bar{y}}{\bar{x}} \bar{x}', \quad (2.1)$$

Up to the first order of approximation the bias and MSE of  $\widehat{Y}_{Rd}$  are respectively given by

$$\text{Bias}(\widehat{Y}_{Rd}) = \bar{Y} f_3 (C_x^2 - \rho_{yx} C_y C_x) \quad (2.2)$$

$$\text{MSE}(\widehat{Y}_{Rd}) = \bar{Y}^2 [f_3 (C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x) + f_2 C_y^2], \quad (2.3)$$

where

$$f_1 = n^{-1} - N^{-1}, \quad f_2 = n'^{-1} - N^{-1}, \quad \text{and} \quad f_3 = f_1 - f_2.$$

Chand (1975), making ratio adjustment to  $\bar{x}'$ , suggested a chain ratio estimator

$$t_1 = \frac{\bar{y}}{\bar{x}} \cdot \frac{\bar{x}'}{\bar{z}'} \cdot \bar{Z} \quad (2.4)$$

with approximate bias and MSE are

$$\text{Bias}(t_1) = \text{Bias}(\widehat{Y}_{Rd}) - f_2 \bar{Y} (C_z^2 - \rho_{yz} C_y C_z), \quad \text{and} \quad (2.5)$$

$$\text{MSE}(t_1) = \text{MSE}(\widehat{Y}_{Rd}) + f_2 \bar{Y}^2 \{C_z^2 - 2\rho_{yz} C_y C_z\} \quad (2.6)$$

respectively. A general class of estimators (Singh et al., 2011)

$$t = \frac{\bar{y}}{\bar{x}} \bar{x}' \left[ \frac{a\bar{Z} + b}{a\bar{z}' + b} \right] = \widehat{Y}_{Rd} R_{ab}$$

includes many estimators, where  $a \neq 0$ ,  $b$  are either real numbers or the functions of the known parameters of the auxiliary variable  $z$  such as  $\bar{Z}$ ,  $\sigma_z$ ,  $\beta_1(z)$  and  $\beta_2(z)$ . In Table 1, we review a few of them for comparative discussion.

Using  $t_2 - t_7$  from Table 1, Singh et al. (2011) proposed composite estimators

$$t_i^* = \alpha t_1 + (1 - \alpha) t_i, \quad i = 2, \dots, 7, \quad (2.7)$$

Table 1: Special cases of the class of estimators  $t$  given above

Estimator		a	b
$t_2 = \widehat{Y}_{Rd} R_{1C_z}$	(Singh and Upadhyaya, 1995)	1	$C_z$
$t_3 = \widehat{Y}_{Rd} R_{\beta_2(z)C_z}$	(Upadhyaya and Singh, 2011)	$\beta_2(z)$	$C_z$
$t_4 = \widehat{Y}_{Rd} R_{C_z \beta_2(z)}$	(Upadhyaya and Singh, 2011)	$C_z$	$\beta_2(z)$
$t_5 = \widehat{Y}_{Rd} R_{1\sigma_z}$	(Singh, 2001)	1	$\sigma_z$
$t_6 = \widehat{Y}_{Rd} R_{\beta_1(z)\sigma_z}$	(Singh, 2001)	$\beta_1(z)$	$\sigma_z$
$t_7 = \widehat{Y}_{Rd} R_{\beta_2(z)C_z}$		$\beta_2(z)$	$C_z$

where  $\alpha \in [0, 1]$  is to be determined so that the estimators have minimum MSEs in the class indexed by  $\alpha$ . The MSE of  $t_i^*$  is obtained as

$$\text{MSE}(t_i^*) = \bar{Y}^2 [f_1 C_y^2 + f_3 C_x^2 + (\alpha + \alpha\theta)^2 f_2 C_z^2].$$

From this, one can verify that  $\text{MSE}(t_i^*)$  achieves minimum when  $\alpha$  is given by

$$\alpha_{opt} = (\rho_{yz} C_y C_z^{-1} - \theta) / (1 - \theta), \quad (2.8)$$

where

$$\theta = \frac{a\bar{Z}}{a\bar{Z} + b}, \quad (2.9)$$

and consequently

$$\min\{\text{MSE}(t_i^*)\} = \bar{Y}^2 [f_1 C_y^2 + f_3 (C_x^2 - 2\rho_{yx} C_y C_x) - f_2 \rho_{yz}^2 C_z^2]. \quad (2.10)$$

Motivated from Khoshnevisan et al. (2007), we propose a class of estimators for estimating the mean  $\bar{Y}$  under two-phase sampling that makes use of partial auxiliary information on one variable and complete auxiliary information on the other variable at estimation stage. The proposed class is defined by

$$\widehat{Y}^* = \frac{\bar{y}}{\bar{x}} \bar{x}' \left[ \frac{a\bar{Z} + b}{\lambda(a\bar{z}' + b) + (1 - \lambda)(a\bar{Z} + b)} \right], \quad (2.11)$$

where  $a$  and  $b$  are as defined above,  $\lambda \in [0, 1]$  is to be determined so that  $\widehat{Y}^*$  has minimum MSE. This class of estimators includes many known estimators as special cases. In particular,  $t$  (for  $\lambda = 1$ ) and, consequently,  $t_1$  to  $t_7$  are members of this class for various choice of  $a$  and  $b$ .

### 3 Approximate Bias and MSE

In order to obtain the approximate bias and MSE of  $\widehat{Y}^*$ , let us use the approximate formulae for bias and MSE of any continuous twice-differentiable function  $g(\cdot)$  of  $\hat{\theta}$  (expanded around

$\theta = E(\hat{\theta}))$ ). For more details see Stuart and Ord (1987), Equations (3.1) and (3.2), which are

$$B(g(\hat{\theta})) = \frac{1}{2} \sum_i \sum_j \left[ \frac{\partial^2 g(\hat{\theta})}{\partial \hat{\theta}_i \partial \hat{\theta}_j} \right]_{\hat{\theta}=\theta} E(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) + O(n^{-3}) \quad (3.1)$$

$$V(g(\hat{\theta})) = \sum_i \left[ \frac{\partial g(\hat{\theta})}{\partial \hat{\theta}_i} \right]_{\hat{\theta}=\theta}^2 V(\hat{\theta}_i) + \sum_i \sum_{j \neq i} \left[ \frac{\partial g(\hat{\theta})}{\partial \hat{\theta}_i} \times \frac{\partial g(\hat{\theta})}{\partial \hat{\theta}_j} \right]_{\hat{\theta}=\theta} \text{Cov}(\hat{\theta}_i, \hat{\theta}_j) + O(n^{-3}). \quad (3.2)$$

Write

$$\widehat{Y}^* = g(\bar{y}, \bar{x}, \bar{x}', \bar{z}') = g(\hat{\theta}) \quad \text{and} \quad \bar{Y} = g(\bar{Y}, \bar{X}, \bar{X}, \bar{Z}) = g(\theta), \quad (3.3)$$

where

$$\hat{\theta}_1 = \bar{y}, \hat{\theta}_2 = \bar{x}, \hat{\theta}_3 = \bar{x}', \hat{\theta}_4 = \bar{z}' \quad \text{and} \quad \theta_1 = \bar{Y}, \theta_2 = \bar{X}, \theta_3 = \bar{X}, \theta_4 = \bar{Z}.$$

Noting

$$\begin{aligned} \frac{\partial \widehat{Y}^*}{\partial \bar{y}} &= \left[ \bar{x}' \left\{ \frac{a\bar{Z} + b}{\lambda(a\bar{z}' + b) + (1-\lambda)(a\bar{Z} + b)} \right\} \right]_{\hat{\theta}=\theta} = 1 \\ \frac{\partial \widehat{Y}^*}{\partial \bar{x}} &= \left[ -\frac{\bar{y}\bar{x}'}{\bar{x}^2} \left\{ \frac{a\bar{Z} + b}{\lambda(a\bar{z}' + b) + (1-\lambda)(a\bar{Z} + b)} \right\} \right]_{\hat{\theta}=\theta} = -\frac{\bar{Y}}{\bar{X}} \\ \frac{\partial \widehat{Y}^*}{\partial \bar{x}'} &= \left[ \frac{\bar{y}}{\bar{x}} \left\{ \frac{a\bar{Z} + b}{\lambda(a\bar{z}' + b) + (1-\lambda)(a\bar{Z} + b)} \right\} \right]_{\hat{\theta}=\theta} = \frac{\bar{Y}}{\bar{X}} \\ \frac{\partial \widehat{Y}^*}{\partial \bar{z}'} &= \left[ \frac{\bar{y}}{\bar{x}} \bar{x}' \left\{ \frac{-a\lambda(a\bar{Z} + b)}{[\lambda(a\bar{z}' + b) + (1-\lambda)(a\bar{Z} + b)]^2} \right\} \right]_{\hat{\theta}=\theta} = -\bar{Y} \frac{a\lambda}{a\bar{Z} + b} \end{aligned}$$

and using (3.1) and (3.2) the approximate bias and MSE of  $\widehat{Y}^*$  can be expressed as:

$$\begin{aligned} B(\widehat{Y}^*) &\approx -\frac{1}{\bar{X}} \text{Cov}(\bar{y}, \bar{x}) + \frac{1}{\bar{X}} \text{Cov}(\bar{y}, \bar{x}') - \frac{a\lambda(a\bar{Z} + b)}{[a\bar{Z} + b]^2} \text{Cov}(\bar{y}, \bar{z}') + \frac{\bar{Y}}{\bar{X}^2} V(\bar{x}) \\ &\quad - \frac{\bar{Y}}{\bar{X}^2} \text{Cov}(\bar{x}, \bar{x}') + \frac{\bar{Y}}{\bar{X}^2} \frac{a\lambda(a\bar{Z} + b)}{[a\bar{Z} + b]^2} \text{Cov}(\bar{x}, \bar{z}') - \frac{\bar{Y}}{\bar{X}^2} \frac{a\lambda}{(a\bar{Z} + b)} \text{Cov}(\bar{x}', \bar{z}') \\ &\quad + \bar{Y} \frac{(a\lambda)^2 (a\bar{Z} + b)}{[a\bar{Z} + b]^3} V(\bar{z}') \end{aligned} \quad (3.4)$$

and

$$\begin{aligned} MSE(\widehat{Y}^*) &\approx V(\bar{y}) + \frac{\bar{Y}^2}{\bar{X}^2} V(\bar{x}) + \frac{\bar{Y}^2}{\bar{X}^2} V(\bar{x}') + \bar{Y}^2 \left[ \frac{a\lambda}{a\bar{Z} + b} \right]^2 V(\bar{z}') - 2 \frac{\bar{Y}}{\bar{X}} \text{Cov}(\bar{y}, \bar{x}) \\ &\quad + 2 \frac{\bar{Y}}{\bar{X}} \text{Cov}(\bar{y}, \bar{x}') - 2 \bar{Y} \frac{a\lambda}{a\bar{Z} + b} \text{Cov}(\bar{y}, \bar{z}') - 2 \frac{\bar{Y}^2}{\bar{X}^2} \text{Cov}(\bar{x}, \bar{x}') \\ &\quad + 2 \frac{\bar{Y}^2}{\bar{X}} \frac{a\lambda}{a\bar{Z} + b} \text{Cov}(\bar{x}, \bar{z}') - 2 \frac{\bar{Y}^2}{\bar{X}} \frac{a\lambda}{a\bar{Z} + b} \text{Cov}(\bar{x}', \bar{z}') \end{aligned} \quad (3.5)$$

Inserting expressions for variances and covariances under two-phase srswor sampling in equations 3.4 and 3.5 the approximate bias and MSE of  $\widehat{Y}^*$  are found respectively as

$$B(\widehat{Y}^*) \approx \bar{Y} [f_3 (\rho_{yx} C_y C_x - C_x^2) - f_2 \{(\lambda\theta)^2 C_z^2 + \lambda\theta \rho_{yz} C_y C_z\}] \quad (3.6)$$

$$MSE(\widehat{Y}^*) \approx \bar{Y}^2 [f_3 (C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x) + f_2 \{C_y^2 + (\lambda\theta)^2 C_z^2 - 2\lambda\theta \rho_{yz} C_y C_z\}] \quad (3.7)$$

Differentiating (3.7) with respect to  $(\lambda\theta)$  and setting the derivative equals to zero we obtain the optimum value of  $(\lambda\theta)$  as

$$(\lambda\theta)_{opt} = \rho_{yz} \frac{C_y}{C_z} \quad \text{or} \quad \lambda_{opt} = \rho_{yz} \frac{C_y}{C_z} \cdot \frac{1}{\theta} \quad (3.8)$$

Consequently, inserting (3.8) in (3.6) and (3.7), we obtain the bias and MSE of the optimal  $\widehat{Y}^*$  as

$$\text{Min } B(\widehat{Y}^*) \approx \bar{Y} f_3 \{C_x^2 - \rho_{yx} c_y c_x\} + f_2 (1 - \rho_{yz}^2) C_y^2, \quad \text{and} \quad (3.9)$$

$$\text{Min } MSE(\widehat{Y}^*) \approx \bar{Y}^2 [f_3 (C_y^2 + C_x^2 - 2\rho_{yx} c_y c_x) + f_2 (1 - \rho_{yz}^2) C_y^2], \quad (3.10)$$

respectively.

*Remark 1.* The optimum values of  $a$ ,  $b$  and  $\lambda$  are not separately obtainable.

*Remark 2.* The estimator given at (2.11) is obtained by ratio adjustment whereas (2.7) by composition of estimators  $t_1 - t_7$ . However, surprisingly both have same minimum MSE, see, equations (2.10) and (3.10).

For the choices  $a = 1$ ,  $b = C_z$  corresponding to  $t_2$  given in Table 1 we have  $\theta = \bar{Z}/(\bar{Z} + C_z)$  and therefore the estimate  $\lambda_{opt}$  is given by

$$\hat{\lambda}_{opt} = r_{yz} \frac{c_y}{C_z} \cdot \frac{\bar{Z}}{\bar{Z} + C_z}, \quad (3.11)$$

where  $r_{yz}$  and  $c_y$  are, respectively, subsample correlation coefficient between  $y$  and  $z$  and coefficient of variation of  $y$ . Thus from (2.11) we suggest the estimator

$$\widehat{Y}_2^* = \frac{(\bar{y}/\bar{x})\bar{x}'(\bar{Z} + C_z)}{\lambda_{opt}(\bar{z}' + C_z) + (1 - \hat{\lambda}_{opt})(\bar{Z} + C_z)} = \frac{\widehat{Y}_{Rd}(\bar{Z} + C_z)}{\hat{\lambda}_{opt}(\bar{z}' - \bar{Z}) + (\bar{Z} + c_z)}. \quad (3.12)$$

We also suggest the estimators  $\widehat{Y}_3^*, \dots, \widehat{Y}_7^*$  (for the choices of  $a$  and  $b$  corresponding to  $t_3, \dots, t_7$ , given in Table 1) in Table 2.

## 4 Comparison of Estimators

This section deals with the analytical and empirical comparison of the estimators discussed above.

Table 2: Special cases of the class of estimators  $\widehat{Y}^*$

Estimator	$\lambda_{opt}$	$a$	$b$
$\widehat{Y}_3^* = \widehat{Y}_{Rd}$ $\left[ \frac{\beta_2(z)\bar{Z} + C_z}{\lambda_{opt}\beta_2(z)(\bar{z}' - \bar{Z}) + (\beta_2(z)\bar{Z} + C_z)} \right]$	$r_{yz} \frac{c_y}{C_z} \cdot \frac{\beta_2(z)\bar{Z}}{\beta_2(z)\bar{Z} + C_z}$	$\beta_2(z)$	$C_z$
$\widehat{Y}_4^* = \widehat{Y}_{Rd}$ $\left[ \frac{C_z\bar{Z} + \beta_2(z)}{\lambda_{opt}C_z(\bar{z}' - \bar{Z}) + (C_z\bar{Z} + \beta_2(z))} \right]$	$r_{yz} \frac{c_y}{C_z} \cdot \frac{C_z\bar{Z}}{C_z\bar{Z} + \beta_2(z)}$	$C_z$	$\beta_2(z)$
$\widehat{Y}_5^* = \widehat{Y}_{Rd}$ $\left[ \frac{\bar{Z} + \sigma_z}{\lambda_{opt}(\bar{z}' - \bar{Z}) + (\bar{Z} + \sigma_z)} \right]$	$r_{yz} \frac{c_y}{C_z} \cdot \frac{\bar{Z}}{\bar{Z} + \sigma_z}$	1	$\sigma_z$
$\widehat{Y}_6^* = \widehat{Y}_{Rd}$ $\left[ \frac{\beta_1(z)\bar{Z} + \sigma_z}{\lambda_{opt}\beta_1(z)(\bar{z}' - \bar{Z}) + (\beta_1(z)\bar{Z} + \sigma_z)} \right]$	$r_{yz} \frac{c_y}{C_z} \cdot \frac{\beta_1(z)\bar{Z}}{\beta_1(z)\bar{Z} + \sigma_z}$	$\beta_1(z)$	$\sigma_z$
$\widehat{Y}_7^* = \widehat{Y}_{Rd}$ $\left[ \frac{\beta_2(z)\bar{Z} + \sigma_z}{\lambda_{opt}\beta_2(z)(\bar{z}' - \bar{Z}) + (\beta_2(z)\bar{Z} + \sigma_z)} \right]$	$r_{yz} \frac{c_y}{C_z} \cdot \frac{\beta_2(z)\bar{Z}}{\beta_2(z)\bar{Z} + \sigma_z}$	$\beta_2(z)$	$\sigma_z$

### 4.1 Efficiency Comparisons under Optimality Condition

In this section analytical comparison of  $\widehat{Y}_{Rd}, t_1, \dots, t_7, t_i^*$  and  $\widehat{Y}$  is presented. Observe from (2.3) and (3.10) that

$$MSE(\widehat{Y}_{Rd}) - \text{Min } MSE(\widehat{Y}^*) \geq f_2(\bar{Y}\rho_{yz}C_y)^2, \tag{4.1}$$

and from (5.1) and (2.6) that

$$MSE(t_1) - \text{Min } MSE(\widehat{Y}^*) = f_2\bar{Y}^2(C_z - \rho_{yz}C_y)^2. \tag{4.2}$$

Also, from Equation (3.10) of Singh et al. (2011) and Remark 2 of this paper, we have

$$MSE(t_i) - \text{Min } MSE(\widehat{Y}^*) \geq 0 \quad (i = 2, \dots, 7). \tag{4.3}$$

Combining (4.1) to (4.3), we conclude that the suggested estimator  $\widehat{Y}^*$  is more efficient than  $(\widehat{Y}_{Rd})$  and  $t_1, \dots, t_7$ .

### 4.2 Numerical Illustrations

The various estimators discussed in previous sections were examined by implementing them in four different datasets described in Figure 1.

#### Data set I

(Murthy, 1967) (The observations are replicated 6 times)

- $y$  : Cultivated Area
- $x$  : Fertilizers
- $z$  : Holding size in acres

$$\begin{aligned}
N &= 216, & n' &= 80, & n &= 30, & \bar{Y} &= 2.365, & \bar{X} &= 97.111, & \bar{Z} &= 17.5853 \\
\sigma_z &= 24.99, & \rho_{yz} &= 0.7066, & \rho_{yx} &= 0.3681, & \rho_{xz} &= 0.2549 \\
C_y &= 0.5330, & C_x &= 1.2977, & C_z &= 1.4248, & \beta_1(z) &= 3.6006, & \beta_2(z) &= 17.4666
\end{aligned}$$

### Data set II

(Fisher, 1936) (The observations are replicated 4 times)

$y$  : Sepal width  
 $x$  : Petal length  
 $z$  : Petal width

$$\begin{aligned}
N &= 200, & n' &= 80, & n &= 30, & \bar{Y} &= 3.122, & \bar{X} &= 2.882, & \bar{Z} &= 0.776 \\
\sigma_z &= 0.5701, & \rho_{yx} &= -0.7447, & \rho_{yz} &= -0.6625, & \rho_{xz} &= 0.9345 \\
C_y &= 0.1511, & C_x &= 0.5117, & C_z &= 0.7365, & \beta_1(z) &= 0.1658, & \beta_2(z) &= 1.2865
\end{aligned}$$

### Data set III

(Jobson, 1992) (The observations are replicated 2 times)

$y$  : Highway Rate  
 $x$  : Weight  
 $z$  : Engine size

$$\begin{aligned}
N &= 194, & n' &= 80, & n &= 30, & \bar{Y} &= 68.37, & \bar{X} &= 2973.71, & \bar{Z} &= 27.60 \\
\sigma_z &= 12.1268, & \rho_{yx} &= 0.7790, & \rho_{yz} &= 0.7464, & \rho_{xz} &= 0.8862 \\
C_y &= 0.1869, & C_x &= 0.1761, & C_z &= 0.4395, & \beta_1(z) &= 0.9441, & \beta_2(z) &= 2.5386
\end{aligned}$$

To compare performance of the estimators the relative efficiency (in percentage) of an arbitrary estimator  $\widehat{Y}$  as compared to the Chand estimator  $t_1$  is calculated as

$$RE(\widehat{Y}) = V(t_1) / V(\widehat{Y}) \times 100\%.$$

The REs of various estimators are presented in following table.

From Table 3 it is clear that the use of parameters of auxiliary variable  $z$  in addition to its mean makes the estimators more efficient than the estimators which do not utilize such information. Moreover, the estimator  $t^*$  proposed by Singh et al. (2011) and our proposed estimator  $\widehat{Y}^*$  (see Remark 2) are more efficient than the estimators included in the study.



Table 3: Relative Efficiency in Percentage

Estimator	Relative Efficiency		
	Data set I	Data set II	Data set III
$t_1$	100	100	100
$t_2$	106	135	103
$t_3$	100	130	101
$t_4$	133	148	137
$t_5$	145	130	174
$t_6$	123	155	177
$t_7$	106	126	131
$t^* = \widehat{Y}^*$	153	162	258

### 4.3 Empirical Comparison using a Monte Carlo Simulation

The estimators  $t_1 - t_7, t_2^* - t_7^*$  (given in (2.7)) with the estimates of  $\alpha_{opt}$  (given in (2.8)) and the estimators  $\widehat{Y}_{(2)}^* - \widehat{Y}_{(7)}^*$  were compared empirically on 3 populations given above. For comparison of the estimators, a preliminary sample  $s'$  of size  $n' = 80$  was drawn using srswor and a second-phase sample  $s$  of size  $n = 30$  was drawn using srswor from each of the populations and these estimators were computed. This procedure was repeated  $M = 5000$  times.

For each estimator  $\widehat{Y}$  its relative percentage bias was calculated as

$$RB(\widehat{Y}) = 100 * (\overline{\widehat{Y}} - \bar{Y})/\bar{Y},$$

and the relative efficiency (in percentage) as

$$RE(\widehat{Y}) = MSE_{sim}(t_1)/MSE_{sim}(\widehat{Y}) \times 100,$$

where

$$\overline{\widehat{Y}} = \sum_{j=1}^M \widehat{Y}_j/M \quad \text{and} \quad MSE_{sim}(\widehat{Y}) = \sum_{j=1}^M (\widehat{Y}_j - Y)^2 / (M - 1).$$

The simulation study results (see Table 4 given in Appendix) shows that for Population 1,  $t_i^*$  have lower absolute RBs among all the estimators whereas the reverse is observed for Populations 1 and 2. For Population 2 all the estimators have underestimated the true population mean. However, the absolute values of RBs of all the estimators are within reasonable range. Generally it has been observed that a composite estimator which constructed using any two estimators with mixing parameter  $\alpha \in (0, 1)$  performs better than the estimators used for composition for a very narrow range of the optimal value of  $\alpha$ . The estimators  $t_i^*$ ,

suggested by Singh et al. (2011), are composite estimators and obviously are better than  $t_1, \dots, t_7$ . The same observation has been supported by the simulation. However,  $t_i^*$  are comparable to each other. In other words, there is no significant impact of the parameters of the  $z$ -variable.

Analytically, it has been shown in (4.3) that the suggested estimator  $\widehat{Y}^*$  with optimal value of  $\lambda$  is better than  $t_1, \dots, t_7$ . The same has been reflected in the simulation study. The relative efficiency of the suggested estimators  $\widehat{Y}_{(1)}^*$  to  $\widehat{Y}_{(7)}^*$  depends on the value of  $C_z$ . The efficiency of the suggested estimators increase with decreasing values of  $C_z$ .

The estimators  $t_i^*$  and  $\widehat{Y}^*$  are constructed using different methodologies. However, they have the same minimum Bias and MSE (see Remark 2). However,  $t_i^*$  are comparable to each other. In other words, there is no significant impact of the parameters of  $z$ -variable which is not the case for our estimator. Finally, the REs of  $t_i^*$  and  $\widehat{Y}^*$  based on the true values of the parameters of  $z$ -variable and on the simulated values of these parameters coincide for each population.

## 5 Conclusions

Often, in two-phase sampling schemes, partial auxiliary information on one variable and full auxiliary information on other variable that is positively correlated with main auxiliary variable and remotely related to the study variable is available. In such situations many researchers proposed various estimators of the population mean. In this paper, we have discussed some of those estimators and proposed a new class of estimators. Empirical and simulated outputs have shown that the estimators that are using known parameters of extra auxiliary variable have performed better than the ratio and Chand (1975) estimators. Moreover, the optimal composite estimators given in Singh et al. (2011) and the suggested estimator, obtained using ratio-adjustment to the ordinary two-phase sampling, have performed very well compared to ratio and Chand (1975) estimators. However, all composite estimators seems to have almost same efficiencies whereas the performance of the suggested estimators vary with the values of  $C_z$ . In particular, our estimators have significant difference in efficiency for smaller values of  $C_z$ . Future studies will investigate the extension of our proposed estimators in various ways, e.g., using regression method of estimation, exponential method of estimation, and investigating them in the presence of non-response.

## Acknowledgment

The authors are thankful to the editor and referees for their constructive suggestions.

## A Appendix

Table 4: Relative Bias (%) and Relative Efficiency (%)

Estimator	Relative Bias (%)			Relative Efficiency (%)		
	Population			Population		
	1	2	3	1	2	3
$t_1$	2.70	-1.12	0.13	100	100	100
$t_2$	2.42	-1.44	0.12	107	137	103
$t_3$	2.68	-1.41	0.13	100	132	101
$t_4$	1.46	-1.52	0.06	140	149	139
$t_5$	1.13	-1.41	0.01	152	132	179
$t_6$	1.77	-1.55	0.00	128	156	183
$t_7$	2.42	-1.38	0.07	108	127	133
$t_2^*$	-0.20	-1.89	-0.18	157	161	267
$t_3^*$	-0.32	-1.92	-0.18	157	160	267
$t_4^*$	0.36	-1.79	-0.17	158	161	268
$t_5^*$	0.64	-1.92	-0.15	158	160	269
$t_6^4$	0.15	-1.73	-0.15	158	162	269
$t_7^*$	-0.20	-1.95	-0.17	158	160	268
$\widehat{Y}_{(2)}^*$	1.03	-1.59	-0.11	158	162	270
$\widehat{Y}_{(3)}^*$	1.03	-1.59	-0.11	158	162	270
$\widehat{Y}_{(4)}^*$	0.86	-1.58	-0.13	157	162	263
$\widehat{Y}_{(5)}^*$	1.03	-1.59	-0.11	158	162	270
$\widehat{Y}_{(6)}^*$	0.90	-1.58	-0.14	158	162	249
$\widehat{Y}_{(7)}^*$	1.03	-1.59	-0.11	158	162	270

## References

- Chand, L. (1975), "Some ratio-type estimators based on two or more auxiliary variables," Unpublished PhD thesis, Iowa State University, Ames, Iowa (USA).
- Choudhury, S. and Singh, B. K. (2012), "A class of chain ratio-product type estimators with two auxiliary variables under double sampling scheme," *Journal of the Korean Statistical Society*, 41, 247-256.

- Cochran, W. G. (1977), *Sampling techniques*, New York: John Wiley & Sons, 3rd ed.
- Fisher, R. A. (1936), "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, 7, 179–188.
- Gupta, S. and Shabbir, J. (2007), "On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables," *Journal of statistical planning and inference*, 137, 1606–1611.
- Jobson, J. (1992), *Applied multivariate data analysis*, vol. II, New York: Springer-Verlag.
- Kadilar, C. and Cingi, H. (2005), "A new estimator using two auxiliary variables," *Applied Mathematics and Computation*, 162, 901–908.
- Khoshnevisan, M., Singh, R., Chauhan, P., and Sawan, N. (2007), "A general family of estimators for estimating population mean using known value of some population parameter(s)," *Far east Journal of Theoretical Statistics*, 22, 181–191.
- Kiregyera, B. (1980), "A chain ratio-type estimator in finite population double sampling using two auxiliary variables," *Metrika*, 27, 217–223.
- (1984), "Regression-type estimators using two auxiliary variables and the model of double sampling from finite populations," *Metrika*, 31, 215–226.
- Mukerjee, R., Rao, T., and Vijayan, K. (1987), "Regression type estimators using multiple auxiliary information," *Australian Journal of Statistics*, 29, 244–254.
- Murthy, M. N. (1967), *Sampling Theory and Methods*, Calcutta, India: Statistical Publishing Society.
- Sahoo, J., Sahoo, L., and Mohanty, S. (1993), "A regression approach to estimation in two-phase sampling using two auxiliary variables," *Current Science*, 73–75.
- Shukla, D., Pathak, S., and Thakur, N. S. (2012), "Estimation of population mean using two auxiliary sources in sample surveys," *Statistics in Transition new series*, 13, 21–36.
- Singh, G. (2001), "On the use of transformed auxiliary variable in the estimation of population mean in two-phase sampling," *Statistics in Transition*, 5, 405–416.
- Singh, G. and Upadhyaya, L. (1995), "A class of modified chain type estimators using two auxiliary variables in two-phase sampling," *Metron*, 53, 117–125.
- Singh, H. and Tailor, R. (2003), "Use of known correlation coefficient in estimating the finite population mean," *Statistics in transition*, 6, 555–560.
- Singh, H. P. and Espejo, M. R. (2007), "Double sampling ratio-product estimator of a finite population mean in sample surveys," *Journal of Applied Statistics*, 34, 71–85.

- Singh, R., Chauhan, P., Sawan, N., and Smarandache, F. (2011), "Improvement in Estimating Population Mean using Two Auxiliary Variables in Two-Phase Sampling," *Italian J. of Pure and Applied Mathematics*, N-28, 135–142.
- Srivastava, S. R., Khare, B., and Srivastava, S. (1990), "A generalized chain ratio estimator for mean of finite population," *J Indian Soc Agric Stat*, 42, 108–117.
- Stuart, A. and Ord, J. K. (1987), *Kendall's Advanced Theory of Statistics*, vol. 1, New York: Oxford University Press, 4th ed.
- Tracy, D., Singh, H., and Singh, R. (1996), "An alternative to the ratio-cum-product estimator in sample surveys," *Journal of statistical planning and inference*, 53, 375–387.
- Upadhyaya, L. and Singh, G. (2011), "Chain type estimators using transformed auxiliary variable in two-phase sampling," *Advances in Modeling and Analysis*, 38, 1–10.
- Upadhyaya, L. N. and Singh, H. P. (1999), "Use of transformed auxiliary variable in estimating the finite population mean," *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 41, 627–636.

*Received: December 12, 2018*

*Accepted: May 23, 2019*