Journal of Statistical Research *https://doi.org/10.3329/jsr.v55i2.58811*
2021, Vol. 55, No. 2, pp. 377-393 *ISSN 0256 - 422 X*

REGRESSION-TYPE ESTIMATION OF A FINITE POPULATION MEAN IN TWO-PHASE SAMPLING USING AUXILIARY VARIABLE AND ATTRIBUTE

P. A. PATEL
Department of Statistics, Sardar Patel University, Vallabh Vidhyanagar 388 120
Email: patelpraful_a@yahoo.co.in

FAGUN SHAH*
Department of Statistics, Sardar Patel University, Vallabh Vidhyanagar 388 120
Email: shah_fagun@yahoo.co.in

SUMMARY

In this paper, by making regression adjustment, a class of estimators of the finite population mean under two-phase sampling is suggested which incorporates auxiliary information on quantitative and qualitative variables. Making approximation up to first order, bias and mean squared error (MSE) are obtained. A few particular cases of the estimators are discussed. The numerical and empirical comparisons of these estimators with ordinary ratio and regression estimators are carried out using a Monte Carlo simulation.

Keywords and phrases: Attribute; auxiliary variables; bias; MSE; optimum estimation; empirical study; simulation; two-phase sampling

1 Introduction

In sample surveys quit often information on one, two or multiple auxiliary variables, which are linearly related to the survey variable, is available prior to surveys. In such situations, it can be incorporated at estimation stage to improve the efficiency of the ordinary estimators of finite population parameters under simple random sampling without replacement (SRSWOR). The problem of estimation of the population total or mean has been extensively studied by many authors under single-phase sampling. Cochran (1940), under the assumption that the study variable and auxiliary variables are highly positively correlated, introduced the ratio estimator where as Watson (1937) introduced the regression estimator. This estimator is preferable when the relation between the study variable and auxiliary variable is linear but not passes through origin. Mohanty (1967), Khare and Srivastava (1981), Sahoo et al. (1993), Samiudin and Hanif (2006, 2007), Kadilar and Cingi (2004, 2005), Kadilar et al. (2007) among others have extended these estimators by incorporating auxiliary information on two or more auxiliary variables. Many researchers, by utilizing some known parameters of the auxiliary variable x such as population mean ($\bar{X}$), standard deviation (σ_x), coefficient

* Corresponding author

of variation (C_x), skewness $(\beta_1(x))$, kurtosis $(\beta_2(x))$, correlation coefficient (ρ_{yx}) etc., used the transformed variables to increase the precision of the ordinary ratio and regression estimators. For details, one may refer Gupta and Shabbir (2007) and other related references cited in this paper. Moreover, Dash and Mishra (2011), Khan (2016), Khan and Khan (2018) have extended the results of one auxiliary variable by using two auxiliary variables. Considering single-phase SRSWOR, Jhajj et al. (2006) proposed a general family of estimators under auxiliary attribute. Singh et al. (2008) suggested many ratio and ratio-product type exponential estimators incorporating known parameters of auxiliary attribute φ such as coefficient of variation C_{φ}, coefficient of kurtosis $\beta_2(\varphi)$, Elfattah et al. (2010) have proposed composite estimators, by using the estimators given in Singh et al. (2008). Khan (2018) developed a class of almost unbiased estimators of the population mean by using auxiliary variable and attribute under systematic sampling. Shabbir and Gupta (2010) proposed an estimator based on auxiliary attributes. Moeen et al. (2012) suggested the mixture regression estimator based on multiple auxiliary variables and attributes. Most of the estimators proposed in single phase, discussed in the above work, have been extended under two-phase sampling. Sukhatme (1962) introduced the ratio estimator using single auxiliary variable under double sampling. Cochan (1977) proposed the ratio and regression estimators under two-phase sampling. Raj (1965), Mohanty (1967), Srivastava (1971), Hidiroglou and Särndal (1998), Fuller (2000), Hidiroglou (2000) and many more researchers have suggested various estimators under two-phase sampling. Kung'u et al. (2014) have extended the estimators suggested by Moeen et al. (2012) under two phase sampling when full, partial and no auxiliary information is available.

Many times addition information on auxiliary variable z, closely related to the main auxiliary variable x but compared to x remotely related to the study variable y (i.e., $\rho_{yz} \leq \rho_{xz}$), for entire population is available. This information can be utilized to estimate the population mean $\bar{X}$ of auxiliary variable using ratio or regression methods of estimation. Chand (1975) suggested a chain ratio estimator whereas Kiregyera (1980, 1984) proposed the ratio-to-regression, ratio-in-regression and regression-in-regression estimators under two-phase sampling. Sisodia and Dwivedi (1981), Upadhyaya and Singh (1999), Singh and Upadhyaya (2001), Singh (2001), Gupta and Shabbir (2007), Singh (2011), Singh et al. (2011) among others have used the transformed auxiliary variables to increase the precision of the ordinary ratio and regression estimators of $\bar{X}$, Motivated from Kiregyera (1980, 1984), we propose a class of estimators for the population mean $\bar{Y}$ incorporating auxiliary information variable x and an attribute φ, highly positively correlated with x, at the estimation stage under a two-phase sampling. We select a preliminary sample s' of size n' using SRSWOR and select a subsample s of size n using SRSWOR.

This article is organized as follows. In Section 2, a class of estimators is suggested and bias and MSE of estimators are obtained by considering approximation up to the first order. An empirical comparison using a Monte Carlo simulation is made in Section 3. Conclusion is given in Section 4.

2 The Proposed Class of Estimators

Consider a finite population U of N identifiable units labeled as $1, \ldots, N$. Let $(y_i, x_i, z_i, \varphi_i)$ be the values of the study variable y, auxiliary variables x and z, and an auxiliary attribute φ for the population unit $i\,(=1,2,\ldots N)$. We wish to estimate the population mean under two-phase sampling by using the auxiliary information available on x and φ, The examples we may consider are (i) $y=$ amount of milk produced, $x=$ amount of feed and $\varphi=$ a particular breed of cow and (ii) $y=$ amount of yield of wheat, $x=$ area under crop and $\varphi=$ a particular variety of wheat. Further, let φ take only two values 0 and 1. $\varphi_i = 1$, if the i^{th} population unit possesses attribute φ and $\varphi_i=0$, otherwise.

We denote the population mean, variance and coefficient of variation of a variable v respectively by

$$\bar{V}=\frac{\sum_{i\in U} v_i}{N},\ S_v^2=S_{vv}=\frac{\sum_{i\in U}\left(v_i-\bar{V}\right)^2}{N},\ \text{ and }\ C_v=\frac{S_v}{\overline{V}}.$$

The population correlation coefficient between two variables v and $w\,(v,w=y,x,z,\varphi)$ is denoted by $\rho_{uv}=S_{vw}/\sqrt{S_vS_w}$, where

$$S_{vw}=\frac{\sum_{i\in U}\left(v_i-\overline{V}\right)\left(w_i-\overline{W}\right)}{(N-1)},\ \ S_v=\sqrt{S_{vv}}.$$

The simple Karle–Pearson correlation coefficient between the variable y and the attribute $\varphi\,(\rho_{y\varphi})$ is called bi-serial correlation. Suppose a sample s of size n is selected using SRSWOR. For discussions to follow we need the following notations

$$f_1=\left(n^{-1}-N^{-1}\right),\ \ \overline{v}=\sum_s\frac{v_i}{n},\ \ s_{vw}=\frac{\sum_s\left(v_i-\overline{v}\right)\left(w_i-\overline{w}\right)}{(n-1)},$$
$$s_v^2=s_{vv},\ \ V\left(\overline{v}\right)=f_1\overline{V}^2C_v^2,\ \ Cov\left(\overline{v},\ \overline{w}\right)=f_1\overline{V}\,\overline{W}\rho_{vw}C_vC_w.$$

Here, $\sum_A(\cdot)$ stands for $\sum_{i\in A}(\cdot)$.

Under SRSWOR, using bi-serial correlation between y and φ, Naik and Gupta (1996) proposed the ratio and regression estimators of $\overline{Y}$ as

$$\hat{\overline{Y}}_{NGR}=\overline{y}\frac{P}{p},\qquad(2.1)$$
$$and\ \ \hat{\overline{Y}}_{NGReg}=\overline{y}+b_{y\varphi}\left(P-p\right),$$

respectively. Here $P=\overline{\varphi}=\sum_U(\varphi_i/N)$ and $p=\sum_s(\varphi_i/n)$ are the proportions of units possessing attribute φ in the population and in a random sample s of size n and $b_{y\varphi}=s_{y\varphi}/s_\varphi^2$ is the sample regression coefficient between the variables y and φ. The approximate MSEs of these estimators are given respectively by

$$\text{MSE}\left(\hat{\overline{Y}}_{NGR}\right)=f_1\overline{Y}^2\left[C_y^2+C_\varphi^2-2\rho_{y\varphi}C_yC_\varphi\right],$$
$$\text{MSE}\left(\hat{\overline{Y}}_{NGReg}\right)=f_1\overline{Y}^2C_y^2\left(1-\rho_{y\varphi}^2\right).$$

It is well-known that if the relationship between y and x is a straight line then under SRSWOR the ordinary ratio and regression estimators are preferable. These estimators are useful only when complete auxiliary information about the population is available. In case only partial auxiliary information is available, two-phase sampling method is adopted. In this sampling method the first-phase sample s' of size n' is selected from U, according to SRSWOR, to obtain a good estimator of $\overline{X}$. Given s', a second-phase sample s of size n is selected from s' acording to SRSWOR. Let $\overline{x}' = (\sum_{s'} x_i)/n'$, $p' = (\sum_{s'} \varphi_i)/n'$, and b'_{yx} denote the regression coefficient of y and x for preliminary sample. The two-phase ordinary ratio and regression estimators (Cochran, 1977, pp 358, 353) are given by

$$\widehat{\widehat{Y}}_{Rd} = \frac{\overline{y}}{\overline{x}}\overline{x}', \tag{2.2}$$

$$\widehat{\widehat{Y}}_{Regd} = \overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right). \tag{2.3}$$

Up to the first order of approximation, the biases and MSEs of these estimators are given by

$$\begin{aligned}
B\left(\widehat{\widehat{Y}}_{Rd}\right) &= \overline{Y} f_3\left(C_x^2 - \rho_{yx}C_yC_x\right),\\
MSE\left(\widehat{\widehat{Y}}_{Rd}\right) &= \overline{Y}^2 f_3\left(C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\right) + f_2\overline{Y}^2C_y^2. \text{ and}\\
B\left(\widehat{\widehat{Y}}_{Regd}\right) &= -\beta_{yx}\frac{N}{N-1}f_3\left(\frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{03}}{\mu_{02}}\right),\\
MSE\left(\widehat{\widehat{Y}}_{Regd}\right) &= \overline{Y}^2\Big[f_2C_y^2 + f_3\left(1-\rho_{yx}^2\right)C_y^2\Big] = \overline{Y}^2C_y^2\left(f_1 - f_3\rho_{yx}^2\right),
\end{aligned}$$

where

$$f_2 = (1/n') - (1/N), f_3 = f_1 - f_2, \beta_{yx} = \frac{S_{yx}}{S_x^2},\ and\ \mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^r\left(x_i - \overline{X}\right)^s.$$

In particularly,

$$\mu_{12} = N^{-1}\sum_{1}^{N}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)^2,\ \mu_{03} = N^{-1}\sum_{1}^{N}\left(x_i - \overline{X}\right)^3$$

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

$$\widehat{\widehat{Y}}_C = \frac{\overline{y}}{\overline{x}}\cdot\frac{\overline{x}'}{\overline{z}'}\cdot\overline{Z}. \tag{2.4}$$

Kiregyera (1980, 1984), noting that this estimator is not preferable when the regression of x on z is linear but not passing through the origin, suggested the ratio-to-regression, ratio-in-regression and regression-in-regression estimators

$$\widehat{\widehat{Y}}_{k1} = \frac{\overline{y}}{\overline{x}}\Big[\overline{x}' + b'_{xz}\left(\overline{Z} - \overline{z}'\right)\Big], \tag{2.5}$$

$$\widehat{\widehat{Y}}_{k2} = \overline{y} + b_{yx}\left(\overline{x}'\overline{Z}/\overline{z}' - \overline{x}\right), \tag{2.6}$$

$$\widehat{\widehat{Y}}_{k3} = \overline{y} + b_{yx}\Big[\overline{x}' + b'_{xz}\left(\overline{Z} - \overline{z}'\right) - \overline{x}\Big]. \tag{2.7}$$

The approximate MSEs of above estimators and many more other estimators are given in Ahmed et al. (2013).

Mohanty (1967) developed a regression-cum-ratio type estimator in the presence of two auxiliary variables as

$$\hat{\hat{Y}}_M = \left[\overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right)\right]\frac{\overline{Z}}{\overline{z}}. \tag{2.8}$$

Mukherjee et al. (1987) proposed a series of estimators of the form regression-in-regression following Kiregyera's (1984) technique as

$$\hat{\hat{Y}}_{MRV1} = \overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right) + b_{yz}\left(\overline{z}' - \overline{z}\right), \tag{2.9}$$

$$\hat{\hat{Y}}_{MRV2} = \overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right) + b_{yz}\left(\overline{Z} - \overline{z}'\right), \tag{2.10}$$

$$\hat{\hat{Y}}_{MRV3} = \overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right) + b_{yx}b_{xz}\left(\overline{Z} - \overline{z}'\right) + b_{yz}(\overline{Z} - \overline{z}). \tag{2.11}$$

Sahoo and Sahoo (1993) suggested modified regression type estimators as

$$\hat{\hat{Y}}_{SS1} = \overline{y}\frac{\overline{x}}{\overline{x}' + b_{xz}\left(\overline{Z} - \overline{z}'\right)}, \tag{2.12}$$

$$\hat{\hat{Y}}_{SS2} = \overline{y} + b_{yx}\left(\overline{x}'\frac{\overline{z}'}{\overline{Z}} - \overline{x}\right), \tag{2.13}$$

$$\hat{\hat{Y}}_{SS3} = \overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right) + b_{yz}\left(\overline{Z} - \overline{z}'\right) = \hat{\hat{Y}}_{MRV2}. \tag{2.14}$$

A generalized ratio estimator was presented by Mishra and Rout (1997) in presence of two auxiliary variables as

$$\hat{\hat{Y}}_{MR} = \overline{y} + d_1\left(\overline{x}^{'} - \overline{x}\right) + d_2\left(\overline{Z} - \overline{z}^{'}\right) + d_3\left(\overline{Z} - \overline{z}\right), \tag{2.15}$$

where d_1, d_2 and d_3 are suitably chosen constants.

Kung'u et al. (2014) have extended the estimator suggested by Moeen et al. (2012), regressing the study variable on multi-auxiliary variables and attributes, under two-phase sampling. The estimators given in (2.9) - (2.11), (2.14), (2.15) can be seen as special cases of the estimators suggested in Kung'u et al. (2014). In particular, for one auxiliary variable and one auxiliary attribute, their estimators are

$$\begin{aligned}
\hat{\hat{Y}}_{KCO1} &= \overline{y} + b_{yx}\left(\overline{X} - \overline{x}'\right) + b_{y\varphi}\left(P - p'\right), \\
\hat{\hat{Y}}_{KCO2} &= \overline{y} + b_{yx}\left(\overline{x}^{'} - \overline{x}\right) + b_{y\varphi}\left(P - p'\right), \\
\hat{\hat{Y}}_{KCO3} &= \overline{y} + b_{yx}\left(\overline{x}^{'} - \overline{x}\right) + b_{y\varphi}\left(p^{'} - p\right).
\end{aligned} \tag{2.16}$$

Remark 1. The estimators given in (2.4) to (2.15) can be modified replacing variable z by attribute φ.

We propose the following class of estimators by regressing the study variable y on the main auxiliary variable x and making ratio-type adjustment to $\overline{x}^{'}$, using an attribute φ

$$\hat{\hat{Y}}^{(1)} = \overline{y} + b_{yx}\left[\overline{x}'\left[\frac{P + a\left(p' - P\right)}{P + b\left(p' - P\right)}\right]^J - \overline{x}\right], \tag{2.17}$$

where $a \neq 0, b$ and J are either real numbers or the functions of the known parameters of the auxiliary attribute such as P , σ_φ, $\beta_1(\varphi)$, $\beta_2(\varphi)$.

Remark 2. This class of estimators is different from the estimators suggested by Kung'u et al. (2014) since the underlying assumptions are different. Moreover, the estimator given in (2.16) can be derived from (2.17), see Equation (2.23).

The rationale for introducing the term $\left[\frac{P+a\left(p'-P\right)}{P+b(p'-P)}\right]^J$ in our proposed estimator is as follows.

(i) This term gives equal and unequal weights to the population and preliminary sample proportions.

(ii) The estimator given in (2.17) gives better results as compared to the ordinary regression type estimator $\overline{y}+b_{yx}\left[\overline{x}'\frac{P}{p'}-\overline{x}\right]$ by capturing the effect of the auxiliary information over a longer range.

(iii) The proposed class of estimators includes the following unknown estimators (with obvious modifications under two-phase sampling of the estimators suggested by Singh (1969), Reddy (1974), Sahai (1979), Srivenkataramana and Tracy (1979)) with suitable choice of a, b, and J.

$$\widehat{\widehat{Y}}_S=\frac{\overline{y}}{\overline{x}}\overline{x}'\frac{P}{p'-(1-w)(p'-P)},\ \widehat{\widehat{Y}}_{Re}=\frac{\overline{y}}{\overline{x}}\overline{x}'\frac{P}{P+b(p'-P)},$$
$$\widehat{\widehat{Y}}_{Sahai}=\frac{\overline{y}}{\overline{x}}\overline{x}'\frac{wP+(1-w)p'}{wp'+(1-w)P},\ \widehat{\widehat{Y}}_{ST}=\frac{\overline{y}}{\overline{x}}\overline{x}'\left[1+a\frac{p'-P}{P}\right].$$

Theorem 1. *The bias and MSE of $\widehat{\widehat{Y}}^{(1)}$ to the terms of order $O\left(n^{-1}\right)$ are given by*

$$B\big(\widehat{\widehat{Y}}^{(1)}\big)=B\big(\widehat{\widehat{Y}}_{Regd}\big)+\Delta\frac{Nf_2}{N-1}\left[\frac{\mu_{011}}{\overline{X}}-\{(J-1)(a-b)-(a+b)\}\frac{\mu_{002}}{P}+\frac{\mu_{111}}{\mu_{110}}-\frac{\mu_{021}}{\mu_{020}}\right],$$

and

$$MSE\big(\widehat{\widehat{Y}}^{(1)}\big)=f_3\left(1-\rho_{xy}^2\right)S_y^2+f_2\left(S_y^2+\Delta^2S_\varphi^2+2\Delta S_{y\varphi}\right), \tag{2.18}$$

where

$$\Delta=\frac{J\beta_{yx}\overline{X}(a-b)}{P}\ \ and\ \ \mu_{rst}=\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\overline{Y}\right)^r\left(x_i-\overline{X}\right)^s\left(\varphi_i-P\right)^t. \tag{2.19}$$

Proof. See Appendix A, where we have used Taylor series expansion to obtain the approximate bias and MSE. For more details about the method see Dash and Mishra (2011). □

It can easily be seen that the optimum value of Δ is

$$\Delta_{opt}=-\frac{S_{y\varphi}}{S_\varphi^2}=-\beta_{y\varphi}. \tag{2.20}$$

Inserting (2.20) in (2.18) we obtain

$$Min\ MSE\left(\widehat{\overline{Y}}^{(1)}\right) = \overline{Y}^2 C_y^2\Big[f_3\left(1-\rho_{xy}^2\right) + f_2\left(1-\rho_{y\varphi}^2\right)\Big],$$

or equivalently

$$Min\ MSE\left(\widehat{\overline{Y}}^{(1)}\right) = MSE\left(\widehat{\overline{Y}}_{Regd}\right) - f_2\overline{Y}^2\rho_{y\varphi}^2 C_y^2. \tag{2.21}$$

In the following corollary, we proposed a subclass of (2.17) that exploits the linear relationship between y and x which is passing through the origin.

Corollary 2.1. The class of estimators

$$\widehat{\overline{Y}}^{(2)} = \frac{\overline{y}}{\overline{x}}\overline{x}'\left[\frac{P + a\left(p^{'} - P\right)}{P + b\left(p' - P\right)}\right]^J$$

has the minimum MSE

$$Min\ MSE\left(\widehat{\overline{Y}}^{(2)}\right) = MSE\left(\widehat{\overline{Y}}_{Rd}\right) - f_2\overline{Y}^2\rho_{y\varphi}^2 C_y^2. \tag{2.22}$$

Remark 3. The optimums values of a, b and J are not separately obtainable.

Remark 4. The optimum value $\Delta_{opt} = -\beta_{y\varphi}$, given in (2.20) is usually unknown and must be estimated using subsample information.

Motivated from (2.17), (2.19) and (2.20) we suggest the following estimators of $\overline{Y}$.

(1) For $b = 0, J = 1$ in (2.17), a takes the optimal value $a = -(\beta_{y\varphi}P)/(\beta_{yx}\overline{X})$. Inserting a by its sample estimate, viz. $\hat{a} = -(b_{y\varphi}P)/(b_{yx}\overline{x}')$, in (2.17) we obtain the estimator of the form

$$\widehat{\overline{Y}}^{(1)}_{prop} = \overline{y} + b_{yx}\left(\overline{x}' - \overline{x}\right) + b_{y\varphi}\left(P - p'\right) = \widehat{\overline{Y}}_{KCO2}. \tag{2.23}$$

(2) For $a = 0, J = 1$, inserting b by its sample estimate, viz. $\hat{b} = (b_{y\varphi}P)/(b_{yx}\overline{x}')$, in (2.17) we find

$$\widehat{\overline{Y}}^{(2)}_{prop} = \overline{y} + b_{yx}\left[\frac{\overline{x}'}{\frac{1-b_{y\varphi}(P-p')}{\overline{x}'}} - \overline{x}\right]. \tag{2.24}$$

(3) Inserting $b = 0, J = 1,$ and $\hat{a} = -(b_{y\varphi}P)/(\overline{x}')$ in (2.17) we have

$$\widehat{\overline{Y}}^{(3)}_{prop} = \overline{y} + b_{yx}\Big[\left(\overline{x}' - \overline{x}\right) + b_{y\varphi}\left(P - p'\right)\Big] = \widehat{\overline{Y}}_{k3}. \tag{2.25}$$

3 Comparison of Estimators

This section deals with the analytical and empirical comparison of the proposed estimators with the bench mark estiamtors. Here, we compare the optimal estimator with the ordinary estimators. One can extend this comparison including the estimators discussed in Section 2.

3.1 Comparison under optimality conditions

We observe from (2.21) and (2.22) that

$$MSE\left(\widehat{\overline{Y}}_{Regd}\right) - Min\ MSE\left(\widehat{\overline{Y}}^{(1)}\right) = \overline{Y}^2 f_2 \rho_{y\varphi}^2 C_y^2 \geq 0,$$

$$MSE\left(\widehat{\overline{Y}}_{Rd}\right) - Min\ MSE\left(\widehat{\overline{Y}}^{(2)}\right) = \overline{Y}^2 \left[f_3\left(C_x - \rho_{yx} C_y\right)^2 + f_2 \rho_{y\varphi}^2 C_y^2\right] \geq 0,$$

$$Min\ MSE\left(\widehat{\overline{Y}}^{(2)}\right) - Min\ MSE\left(\widehat{\overline{Y}}^{(1)}\right) = \overline{Y}^2 f_3\left(C_x - \rho_{yx} C_y\right)^2 \geq 0.$$

Therefore, our proposed estimator $\widehat{\overline{Y}}^{(1)}$ given in (2.17) is at least as good as $\widehat{\overline{Y}}_{Rd}$, $\widehat{\overline{Y}}_{Regd}$ and $\widehat{\overline{Y}}^{(2)}$.

3.2 Numerical illustration

The estimators $\widehat{\overline{Y}}_{NG}$, $\widehat{\overline{Y}}_{Rd}$, $\widehat{\overline{Y}}_{Regd}$, $\widehat{\overline{Y}}^{(1)}$ and $\widehat{\overline{Y}}^{(2)}$, given in (2.1), (2.2), (2.3), (2.23) and (2.24) respectively, were examined with help of three real data sets described below.

Data set I: (Source: The Fuel Consumption Guide 1985 published by Transport Canada. Also see Jobson, 1992) (The observations are replicated 2 times)

y : Highway Rate, x : Engine size

φ : Automat transmission ($\varphi = 1$ if automatic and $\varphi = 0$ if standard)

$N = 194,\ \overline{Y} = 68.37,\ \overline{X} = 27.5979,\ P = 0.5979,$

$\sigma_\varphi = 12.1268,\ \rho_{yx} = 0.7464,\ \rho_{y\varphi} = 0.7464,\ \rho_{x\varphi} = 0.2508,$

$C_y = 0.1869,\ C_x = 0.4394,\ C_\varphi = 0.4395,\ \beta_1(\varphi) = 0.9441,\ \beta_2(\varphi) = 2.5386.$

Data set II : (Source: Fisher R.A. (1936). Also see Anderson, 1958) (The observations are replicated 4 times)

y : Sepal length, x : Petal length

φ : Species ($\varphi = 1$ if unit come from Versicolor family and $\varphi = 0$ otherwise)

$N = 200,\ \overline{Y} = 6.264,\ \overline{X} = 4.936,\ P = 0.5,$

$\sigma_\varphi = 0.5013,\ \rho_{yx} = 0.7957,\ \rho_{y\varphi} = -0.4321,\ \rho_{x\varphi} = -0.7804,$

$C_y = 0.1007,\ C_x = 0.1634,\ C_\varphi = 1.0025,\ \beta_1(\varphi) = 0,\ \beta_2(\varphi) = 1.$

Data set III : (Source: Narula, S. C. and Wellington, J. F. (1977), Technometrics,19. Also see Montgomery et al., 2003) (The observations are replicated 9 times)

y : Sale price of house, x : Living space

φ : Number of baths ($\varphi = 1$ if number of baths is 1 and $\varphi = 0$ otherwise)

$N = 216, \;\; \overline{Y} = 34.612, \;\; \overline{X} = 1.38, \;\; P = 0.708,$

$\sigma_{\varphi} = 0.4556, \;\; \rho_{yx} = 0.7101, \;\; \rho_{y\varphi} = -0.6475, \;\; \rho_{x\varphi} = -0.7582,$

$C_y = 0.1701, \;\; C_x = 0.1959, \;\; C_{\varphi} = 0.6431, \;\; \beta_1(\varphi) = -0.9167, \;\; \beta_2(\varphi) = 1.8403.$

The percentage relative efficiency was calculated by considering $\hat{\hat{\overline{Y}}}_{Rd}$ as a benchmark estimator.

Table 1: Relative Efficiency in Percentage

Estimator	Data set I	Data set II	Data set III
$\hat{\hat{\overline{Y}}}_{Rd}$	100.00	100.00	100.00
$\hat{\hat{\overline{Y}}}_{NG}$	14.95	0.95	3.84
$\hat{\hat{\overline{Y}}}_{Regd}$	423.45	193.96	122.26
$\hat{\hat{\overline{Y}}}^{(1)}$	458.99	213.72	149.34
$\hat{\hat{\overline{Y}}}^{(2)}$	101.86	105.01	117.41

The ordinary two-phase regression estimator $\hat{\hat{\overline{Y}}}_{Regd}$ has performed very well for all the populations under study. The proposed estimator $\hat{\hat{\overline{Y}}}^{(1)}$ has exhibited substantial improvement over $\hat{\hat{\overline{Y}}}_{Regd}$ and the use of additional auxiliary attribute φ makes the estimators more efficient than the other estimators which do not utilize such information.

3.3 Empirical comparison using a Monte Carlo simulation

Here, we include the following estimators (motivated from Singh et al., 2008) in addition to the estimators given in (2.2), (2.3), (2.23), (2.24), (2.25) for empirical comparison.

$$t_1 = \overline{y} + b_{yx}\left(\left\{\overline{x}' + b_{x\varphi}^{'}\left(P - p^{'}\right)\right\}\frac{P}{p'} - \overline{x}\right),$$

$$t_2 = \overline{y} + b_{yx}\left(\left\{\overline{x}' + b_{x\varphi}^{'}\left(P - p^{'}\right)\right\}\frac{P + \beta_2(\varphi)}{p' + \beta_2(\varphi)} - \overline{x}\right),$$

$$t_3 = \overline{y} + b_{yx}\left(\left\{\overline{x}' + b_{x\varphi}^{'}\left(P - p^{'}\right)\right\}\frac{P + C_{\varphi}}{p' + C_{\varphi}} - \overline{x}\right),$$

$$t_4 = \overline{y} + b_{yx}\left(\left\{\overline{x}' + b_{x\varphi}^{'}\left(P - p^{'}\right)\right\}\frac{P\beta_2(\varphi) + C_{\varphi}}{p'\beta_2(\varphi) + C_{\varphi}} - \overline{x}\right).$$

The above estimators are the members of the class

$$\hat{\hat{\overline{Y}}}^{(3)} = \overline{y} + b_{yx}\left(\left\{\overline{x}' + b_{x\varphi}^{'}\left(P - p^{'}\right)\right\}\frac{cP + d}{cp' + d} - \overline{x}\right),$$

where $c \neq 0$ and b are either real numbers or the functions of the known parameters of the auxiliary attribute φ such as σ_φ, C_φ, $\beta_1(\varphi)$, $\beta_2(\varphi)$ etc.

For comparison of the estimators, a preliminary sample s' of size $n' = 50\,(80)$ was drawn using SRSWOR and a second-phase sample s of size $n = 20\,(30)$ was drawn using SRSWOR from each of the above populations and these estimators were computed. This procedure was repeated $M = 5000$ times. The performances of the estimators were measured in terms of relative percentage bias and relative percentage efficiency. Here many known and unknown estimators and the estimators which are members of the three classes $\widehat{\overline{Y}}^{(i)}, i = 1, 2, 3$, given above were included in the simulation; however a few of them performed very well. The performances of such estimators were reported in Tables 2 and 3 (see Appendix B).

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

$$RB\left(\widehat{\overline{Y}}\right) = 100 * \frac{\left(\overline{\widehat{\overline{Y}}} - \overline{Y}\right)}{\overline{Y}},$$

and the relative percentage efficiency as

$$RE\left(\widehat{\overline{Y}}\right) = MSE_{sim}\left(\widehat{\overline{Y}}_{Rd}\right) / MSE_{sim}\left(\widehat{\overline{Y}}\right) \times 100,$$

where

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

Here $\widehat{\overline{Y}}_{Rd}$ was considered as the benchmark estimator.

Table 2: Simulated Relative Efficiency in Percentage for different set of sample sizes

	$n' = 50,\ n = 20$			$n' = 80,\ n = 30$		
Estimator	Pop 1	Pop 2	Pop 3	Pop 1	Pop 2	Pop 3
$\widehat{\overline{Y}}_{Rd}$	100.00	100.00	100.00	100.00	100.00	100.00
$\widehat{\overline{Y}}_{Regd}$	375.75	173.97	111.72	427.64	193.74	121.21
t_1	387.34	83.73	132.73	446.05	103.63	113.45
t_2	375.11	83.75	59.64	431.80	103.83	74.45
t_3	373.71	83.76	95.95	423.99	103.41	56.69
t_4	390.53	83.40	123.37	451.60	103.15	125.30
$\widehat{\overline{Y}}^{(1)}_{prop}$	393.45	189.23	145.10	455.85	206.80	136.17
$\widehat{\overline{Y}}^{(2)}_{prop}$	393.12	188.99	145.20	455.54	206.67	135.94
$\widehat{\overline{Y}}^{(3)}_{prop}$	394.98	186.87	146.82	457.09	206.91	137.49

The scatter plot of Population 1 revealed that a linear model $y_i = \alpha + \beta x_i + \varepsilon_i$ might be appropriate and the relationship between y and x is strong. For Population 2 scatter plot exhibited two different straight lines (because the observations are taken from two different families - Iris Setosa and Iris Versicolor) whereas for Population 3 no systematic pattern was found though the correlation between y and x is moderate to high. Clearly, the populations had not fulfilled the requirements for the suggested estimators.

From Tables 2 and 3, we have the following interesting observations.

(1) All the estimators have very small relative bias and decreased with increased in sample size for Populations 1 and 2.

(2) Among t_1, t_2, t_3 and t_4, t_4 is better performer.

(3) The proposed estimators $\widehat{\overline{Y}}^{(1)}_{prop}$, $\widehat{\overline{Y}}^{(2)}_{prop}$ and $\widehat{\overline{Y}}^{(3)}_{prop}$ have exhibited very good performance over all the estimators. However, they are comparable among each other.

(4) The relative efficiency of the proposed estimators increased with sample size for Populations 1 and 2.

4 Conclusion

Many times addition auxiliary attribute φ is closely positively related to x; but compared to x, it is remotely related to y. We have utilized such variables to develop a class of estimators of the population mean under two-phase sampling and obtained the minimum MSE for the proposed class. Theoretically, we have shown that the class of estimators is at least as good as the ordinary two-phase ratio and regression estimators provided there is positive biserial correlation between x and φ, In addition, to support these theoretical findings a small scale empirical study is carried out. Our proposed estimators have performed better than the other estimators. Finally, our suggested estimators are useful for those survey statisticians who are interesting in regression approach to estimation of the population characteristic using auxiliary variable and attribute in two-phase sampling.

Acknowledgements

The authors are thankful to the referees and Editor for their constructive and helpful comments, which helped to improve the original manuscript.

Appendices

A Proof of Theorem 1

In order to obtain the approximate bias and MSE of $\widehat{\overline{Y}}^{*}$, 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 detail see Stuart and Ord (1987), Equation (10.12) or Wolter (2007) p. 230) as follows

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

$$V(g(\underline{\hat{\theta}})) = \sum_i\left[\frac{\partial g(\underline{\hat{\theta}})}{\partial\hat{\theta}_i}\right]^2_{\underline{\hat{\theta}}=\underline{\theta}} V\left(\hat{\theta}_i\right)+\sum\sum_{i\neq j}\left[\frac{\partial g(\underline{\hat{\theta}})}{\partial\hat{\theta}_i}\cdot\frac{\partial g(\underline{\hat{\theta}})}{\partial\hat{\theta}_j}\right]_{\underline{\hat{\theta}}=\underline{\theta}} \mathrm{Cov}\left(\hat{\theta}_i,\hat{\theta}_j\right)+O\left(n^{-3}\right). \tag{A.2}$$

Write $\widehat{\overline{Y}}^{(1)} = g\left(\overline{y},\overline{x},\overline{x}',p',\mathrm{s}_{\mathrm{x}}^2,\mathrm{s}_{\mathrm{yx}}\right)=g(\underline{\hat{\theta}})$ and $\overline{Y}=g\left(\overline{Y},\overline{X},\overline{X},P,\mathrm{S}_{\mathrm{x}}^2,\mathrm{S}_{\mathrm{yx}}\right)=g(\underline{\theta})$, where

$$\hat{\theta}_1=\overline{y},\quad \hat{\theta}_2=\overline{x},\quad \hat{\theta}_3=\overline{x}',\quad \hat{\theta}_4=p',\quad \hat{\theta}_5=s_x^2,\quad \hat{\theta}_6=s_{yx},$$
$$\theta_1=\overline{Y},\quad \theta_2=\overline{X},\quad \theta_3=\overline{X},\quad \theta_4=P,\quad \theta_5=S_{\mathrm{x}}^2,\quad \theta_6=S_{\mathrm{yx}}.$$

Since

$$\frac{\partial\widehat{\overline{Y}}^{(1)}}{\partial\overline{y}}=1,\ \frac{\partial\widehat{\overline{Y}}^{(1)}}{\partial\overline{x}}=\left[-\frac{s_{xy}}{s_x^2}\right]_{\underline{\hat{\theta}}=\underline{\theta}}=-\frac{S_{xy}}{S_x^2}=-\beta_{yx},\frac{\partial\widehat{\overline{Y}}^{(1)}}{\partial\overline{x}'}=\left[\frac{s_{xy}}{s_x^2}\cdot\left(\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right)\right]_{\underline{\hat{\theta}}=\underline{\theta}}=\beta_{yx},$$

$$\frac{\partial^2\widehat{\overline{Y}}^{(1)}}{\partial p'}=\left[J\frac{s_{xy}}{s_x^2}\overline{x}'\cdot\left(\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right)^{J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\right]_{\underline{\hat{\theta}}=\underline{\theta}}=J\beta_{yx}\overline{X}\frac{a-b}{P},$$

$$\frac{\partial^2\widehat{\overline{Y}}^{(1)}}{\partial s_x^2}=\left[-\frac{s_{xy}}{\left(s_x^2\right)^2}\cdot\left\{\overline{x}'\left(\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right)^J-\overline{x}\right\}\right]_{\underline{\hat{\theta}}=\underline{\theta}}=0,$$

$$\frac{\partial^2\widehat{\overline{Y}}^{(1)}}{\partial s_{yx}}=\left[\frac{1}{s_x^2}\cdot\left\{\overline{x}'\left(\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right)^J-\overline{x}\right\}\right]_{\underline{\hat{\theta}}=\underline{\theta}}=0,$$

we obtain using (A.1) and (A.2) the approximate bias and MSE of $\widehat{\overline{Y}}^{(1)}$ as

$$\begin{aligned}B\left(\widehat{\overline{Y}}^{(1)}\right)&=\beta_{yx}\overline{X}\left[\frac{\mathrm{Cov}\left(\overline{x},s_x^2\right)}{\overline{X}S_x^2}-\frac{\mathrm{Cov}\left(\overline{x},s_{xy}\right)}{\overline{X}S_{xy}}-\frac{\mathrm{Cov}\left(\overline{x}',s_x^2\right)}{\overline{X}S_x^2}+\frac{\mathrm{Cov}\left(\overline{x}',s_{xy}\right)}{\overline{X}S_{xy}}\right]\\
&+J\beta_{yx}\overline{X}(a-b)\left[\frac{\mathrm{Cov}\left(\overline{x},p'\right)}{\overline{X}P}-\{(J-1)(a-b)-(a+b)\}\frac{V\left(p'\right)}{P^2}+\frac{\mathrm{Cov}\left(p',s_{xy}\right)}{PS_{xy}}-\frac{\mathrm{Cov}\left(p',s_x^2\right)}{PS_x^2}\right]\\
&=-\beta_{yx}\frac{Nf_3}{N-1}\left(\frac{\mu_{120}}{\mu_{110}}-\frac{\mu_{030}}{\mu_{020}}\right)+\Delta\frac{Nf_2}{N-1}\left[\frac{\mu_{011}}{\overline{X}}-\{(J-1)(a-b)-(a+b)\}\frac{\mu_{002}}{P}+\frac{\mu_{111}}{\mu_{110}}-\frac{\mu_{021}}{\mu_{020}}\right].\end{aligned}$$

and

$$
\begin{aligned}
&MSE\left(\widehat{\overline{Y}}^{(1)}\right) = V(\overline{y}) + \left(-\frac{s_{xy}}{s_x^2}\right)^2 V(\overline{x}) + \left(\frac{s_{xy}}{s_x^2}\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J\right)^2 V\left(\overline{x}'\right) \\
&+\left(\frac{s_{xy}}{s_x^2}\overline{x}'J\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^{J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\right)^2 V\left(p'\right) \\
&+\left[\frac{1}{s_x^2}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\right]^2 V\left(s_{xy}\right) + \left[-\frac{s_{xy}}{s_x^2}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\right]^2 V\left(s_x^2\right) \\
&-\frac{2s_{xy}}{s_x^2}\operatorname{Cov}(\overline{y},\overline{x}) + \frac{2s_{xy}}{s_x^2}\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J\operatorname{Cov}\left(\overline{y},\overline{x}'\right) + \frac{2s_{xy}}{s_x^2}\overline{x}'\left[\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^{J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\operatorname{Cov}\left(\overline{y},p'\right)\right. \\
&+\frac{2}{s_x^2}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(\overline{y},s_{yx}\right) - \frac{2s_{xy}}{s_x^2}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(\overline{y},s_x^2\right) \\
&-2\left(\frac{s_{xy}}{s_x^2}\right)^2\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J\operatorname{Cov}\left(\overline{x},\overline{x}'\right) - 2\left(\frac{s_{xy}}{s_x^2}\right)^2\overline{x}'J\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^{J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\operatorname{Cov}\left(\overline{x},p'\right) \\
&-\frac{2s_{xy}}{s_x^4}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(\overline{x},s_{yx}\right) + 2\left(\frac{s_{xy}}{s_x^2}\right)^2\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(\overline{x},s_x^2\right) \\
&+2\left(\frac{s_{xy}}{s_x^2}\right)^2\overline{x}'J\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^{2J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\operatorname{Cov}\left(\overline{x}',p'\right) \\
&+\frac{2s_{xy}}{s_x^4}\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(\overline{x}',s_{yx}\right) \\
&-2\left(\frac{S_{xy}}{s_x^2}\right)^2\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(\overline{x}',s_x^2\right) \\
&+\frac{2s_{xy}}{s_x^4}\overline{x}'J\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^{J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(p',s_{yx}\right) \\
&-2\left(\frac{s_{xy}}{s_x^2}\right)^2\overline{x}'J\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^{J-1}\frac{(a-b)P}{\left(P+b\left(p'-P\right)\right)^2}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)\operatorname{Cov}\left(p',s_x^2\right) \\
&-\frac{2s_{xy}}{s_x^4}\left(\overline{x}'\left[\frac{P+a\left(p'-P\right)}{P+b\left(p'-P\right)}\right]^J-\overline{x}\right)^2\operatorname{Cov}\left(s_{yx},s_x^2\right).
\end{aligned}
$$

evaluated at $\hat{\theta}=\theta$.

Evaluating the above expression of MSE at $\hat{\theta}=\theta$ and inserting expressions for variances and covariances under two-phase sampling we obtain

$$
\begin{aligned}
\mathrm{MSE}\left(\widehat{\overline{Y}}^{(1)}\right) &= f_1\left[S_y^2+\beta_{yx}^2S_x^2-2\beta_{yx}S_{yx}\right]+f_2\left[\beta_{yx}^2S_x^2+2\beta_{yx}S_{yx}-2\beta_{yx}^2S_x^2\right] \\
&\quad + f_2\left[\left(J\beta_{yx}\overline{X}\frac{a-b}{P}\right)^2S_\varphi^2+2J\beta_{yx}\overline{X}\frac{a-b}{P}S_{y\varphi}\right].
\end{aligned}
$$

Upon simplification

$$\mathrm{MSE}\left(\widehat{\overline{Y}}^{(1)}\right) = f_3\left(S_y^2 + \beta_{yx}^2 S_x^2 - 2\beta_{yx} S_{yx}\right) + f_2\left(S_y^2 + \Delta^2 S_\varphi^2 + 2\Delta S_{y\varphi}\right),$$

where $\Delta = J\beta_{yx}\overline{X}(a-b)/P$.

Table A1: Relative Bias (in $\%$) for different set of sample sizes

	$n'=50,\ n=20$			$n'=80,\ n=30$		
Estimator	Pop 1	Pop 2	Pop 3	Pop 1	Pop 2	Pop 3
$\widehat{\overline{Y}}_{Rd}$	0.20	-1.20	-0.04	0.11	-0.87	0.13
$\widehat{\overline{Y}}_{Regd}$	-0.14	0.41	0.11	-0.05	-0.33	0.22
t_1	0.33	-0.42	0.12	-0.02	-0.32	0.25
t_2	0.45	-0.41	0.02	-0.01	-0.32	0.35
t_3	0.44	-0.40	0.01	-0.01	-0.32	0.41
t_4	0.41	-0.41	0.02	-0.03	-0.32	0.22
$\widehat{\overline{Y}}^{(1)}_{prop}$	-0.41	-0.42	0.12	-0.06	-0.32	0.17
$\widehat{\overline{Y}}^{(2)}_{prop}$	0.41	-0.54	0.16	-0.04	-0.33	0.19
$\widehat{\overline{Y}}^{(3)}_{prop}$	-0.49	-0.41	0.14	-0.04	-0.34	0.17

References

Abd-Elfattah, A.M., El-Sherpieny, E.A., Mohamed, S.M. and Abdou, O.F. (2010), "Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute," *Applied Mathematics and Computation*, 215, 4194–4202.

Ahmad, Z., Shahbaj, M. Q. and Hanif, M. (2013), *Two Phase Sampling*, Cambridge Scholars Publishing, UK.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

Chand L. (1975). *Some ratio type estimators based on two or more auxiliary variables,* Unpublished Ph. D. thesis, Iowa State University, Ames, Iowa (USA).

Cochran, W. G. (1940). "The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce". The Journal of Agricultural Science, 30, 262-275.

Cochran, W. G. (1977). *Sampling Techniques, 3rd ed.*, John Wiley and Sons, New York

Dash, P. and Mishra, G. (2011), "An Improved Class of Estimators in Two-Phase Sampling Using Two Auxiliary Variables," *Communication in Statistics- Theory and Methods*, 40, 4347–4352.

Fuller, W. A. (2000), "Two-phase sampling," *In Proceedings of the Annual Meeting of the Survey Methods Section of the Statistical Society of Canada*, 23–30.

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.

Hidiroglou, M. A. (2000), "Double sampling," *Survey Methodology*, 27, 143–154.

Hidiroglou, M. A. and Sarndal, C.E. (1998), "Use of auxiliary information for two-phase sampling," *Survey Methodology*, 24, 11–20.

Jhajj, H.S., Sharma M.K., and L.K. Grover (2006), "A family of estimators of population mean using information on auxiliary attribute," *Pak. J. Statist.*, 22, 43–50.

Jobson, J. (1992), *Applied Multivariate Data Analysis*, Vol. II, Springer-Verlag.

Kadilar, C. and Cingi, H. (2004), "Ratio Estimators in Simple Random Sampling," *Applied Mathematics and Computation*, 151, 893–902.

Kadilar and Cingi (2005), "A new ratio estimator using two auxiliary variables," *Applied Mathematics and Computation*, 162, 901–908.

Kadilar, C., Candan, M. and Cingi, H. (2007), "Ratio estimators using robust regression," *Hacettepe Journal of Mathematics and Statistics*, 36, 181–188.

Khan (2016). *A Class of Generalized Estimators of Population Mean Using Auxiliary Variables in Single and Two Phase Sampling with Properties of the Estimators*, Unpublished PhD thesis, Department of Statistics, GC University, Lahore, Pakistan..

Khare, B.B. and Srivastava, S.R. (1981), "A general regression ratio estimator for the population mean using two auxiliary variables," *Aligarh Journal of Statistics*, 1, 43–51.

Khan, Z. (2018), "A class of almost unbiased estimators of population mean using auxiliary variable and auxiliary attribute in systematic sampling," *Journal of Statistical Theory and Practice*, 12, 412–422.

Khan, H. and Khan, M. (2018), "New Exponential-Ratio Type Estimators of Population Mean in Two- Phase Sampling Using No Information Case on Auxiliary Variables," *Journal of Reliability and Statistical Studies*, 10, 95–103.

Kiregyera, B. (1980), "A chain ratio-type estimator in finite population mean in double sampling using two auxiliary variables," *Metrika*, 27, 217–223.

Kiregyera, B. (1984), "Regression-type estimator using two auxiliary variables and model of double sampling from finite populations," *Metrika*, 31, 215–223.

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.

Mishra, G. and Rout, K. (1997). "A regression estimator in two-phase sampling in presence of two auxiliary variables." *Metron*, 52, 177–186.

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.

Mohanty, S. (1967). "Combination of Regression and Ratio Estimate." *Jour. Ind. Statist. Assoc.*, 5, 16-19.

Montgomery, D.C., Perk, E.A. and Vining, G.G. (2003), *Introduction to Linear Regression Analysis*, Wiley India, Third edition.

Mukerjee R., Rao T. J. and Vijayan K. (1987). "Regression type estimators using multiple auxiliary information", *Australian Journal of Statistics*, 29(3), 244—254.

Muzaffar, H. (2016), "A Class of Generalized Estimators of Population Mean Using Auxiliary Variables in Single and Two Phase Sampling with Properties of the Estimators," *Unpublished PhD thesis*, Department of Statistics, GC University, Lahore, Pakistan.

Naik, V.D. and Gupta, P.C. (1991). "A General class of estimators for estimating population mean using auxiliary information", *Metrika*, 38, 11-17.

Raj, D. (1965). "On a Method of Using Multi-Auxiliary Information in Sample Surveys," *Journals of the American Statistical Association,* 60, 154-165.

Riaz, S. and Darda, M.A. (2016), "Some classes of estimators of estimators in the presence of non-response using auxiliary attribute," *SpringerPlus*, 5, 1271.

Sahai, A. (1979), "An efficient variant of the product and ratio estimators," *Statistica Neerlandica*, 33, 27–35.

Sahoo, J., Sahoo, L.N. and Mohanty, S. (1993), "A regression approach to estimation in two phase sampling using two auxiliary variables," *Current Sciences*, 65, 73–75.

Sahoo, J., Sahoo, L. N. (1993). "A class of estimators in two-phase sampling using two auxiliary variables," *J. Ind. Statist. Assoc.* 37:107-114.

Samiuddin, M. and Hanif, M. (2006), "Estimation in two phase sampling with complete and incomplete information," *Proc. 8th Islamic Countries Conference on Statistical Sciences*, 13, 479–495.

Shabbir, J. and Gupta, S. (2007), "On the use of transformed auxiliary variables in estimating population mean," *Journal of Statistical Planning and Inference*, 137, 1606–1611.

Shabbir, J. and Gupta, S. (2010), "Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes," *Hacettepe Journal of Mathematics and Statistics*, 39, 121–129.

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.

Singh G.N. (2011). “On the use of transformed auxiliary variable in estimation of population mean in two-phase sampling”, *Statistics in Transition* , 5(3), 405-416.

Singh, P., Chauhan, N., Sawan, R. and Smarandache, F. (2008), “Ratio estimators in simple random sampling using information on auxiliary attribute,” *Pak. J. Stat. Oper. Res.*, 4, 47–53.

Singh, H.P and Upadhyaya, L.N. (2001), “Chain type estimators using transformed auxiliary variable in two-phase sampling,” *Advances in Modeling and Analysis*, 38, 1–10.

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2011). “Improvement in estimating population mean using two auxiliary variables in two-phase sampling”, *Italian Journal of Pure and Applied Mathematics,* N-28, 135-142.

Sisodia, B.V.S. and Dwivedi, V.K. (1981), “A modified ratio estimator using coefficient of variation of auxiliary variable,” *Jour. Ind. Soc. Agr. Stat.*, 33, 13–18.

Srivastava, S.K. (1970). “A two phase sampling estimator in sample surveys”, *Australian Journal of Statistics,* 12, 23-27.

Stuart, A. and Ord, K. (1987), *Kendall’s Advanced Theory of Statistics*, Volume 1, Distribution Theory, Fourth Edition.

Sukhatme, B. V. (1962). “Some ratio-type estimators in two-phase sampling,” *Journal of the American Statistical Association,* 57, 628-632.

Upadhyaya, L.N. and Singh, H.P. (1999), “Use of transformed auxiliary variable in estimating the finite population mean,” *Biometrical Journal*, 41, 627–636.

Watson, D.J. (1937). “The Estimation of Leaf Areas,” *Journal of Agricultural Science,* 27, 474.

Wolter, K. M. (2007), *Introduction to Variance Estimation* (Second Edition), Springer.

Received: July 9, 2020

Accepted: April 7, 2021