ALTERNATIVE CLASS OF ESTIMATORS OF POPULATION MEAN UNDER DOUBLE SAMPLING SCHEME

[1]Ekpenyong, E. J. (Ph.D) & [2]Enang, E. I (Ph.D)
[1]Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State.
[2]Department of Statistics, University of Calabar, Calabar, Cross River State.
ekpesstat@yahoo.coo & ekkaass@yahoo.com

Abstract
This study proposes a ratio estimator of the population mean under double (two-phase) sampling scheme, in order to tackle the problem of low efficiencies of some existing estimators. The bias and mean square error of the class has been derived. Analytical and numerical results indicate that, the set of optimal estimators of the proposed class of estimators have been found to exhibit greater gains in efficiencies than the existing ones under certain conditions.

Key words: *double sampling, optimal estimator, mean square error, bias.*

Introduction
In double sampling scheme, if information needed to improve on the estimate of the character of interest under study is lacking, and if it is convenient and cheap to do so, then information on the auxiliary variable is collected from a preliminary large sample. While information on the variable of interest, say, Y, is collected from a second sample which is smaller in size than the preliminary sample. The second sample may be a subsample of the preliminary sample or may be an independent sample selected from the entire population. When the second sample is independent of the preliminary sample, information on both the auxiliary and the main characters are obtained from the second sample. The preliminary sample constitutes the first phase sample, and the second sample is the second phase sample (Okafor, 2002).

The subject of ratio and regression estimation in double sampling strategy have also been considered by many authors; some of which have made some extensions and modifications of the existing estimators of population parameters under the sampling scheme using both single and multi-auxiliary characters. Among these authors are Swain (2012b), Pradhan (2005), Handique (2012), Singh (2001), Hidiroglou and Sarndal (1998), Choudhury and Singh (2012), Singh and Choudhury (2012), Malik and Tailor (2013), Dash and Mishra (2011), Singh and Vishwakarma (2007), Yadav, Upadhyaya, Singh, and Chatterjee (2013).
When information on the supplementary variable(X) is unknown, the double sampling estimators for the population mean $(\bar{Y})$ of the character of interest (Y) are used. Sukhatme (1962) proposed the classical ratio estimator of the population mean in double sampling. Singh and Viswakarma (2007) also proposed the two-phase exponential ratio and product estimator, which was a motivation from the Bahl and Tuteja (1991) estimator. Singh and Espejo (2007) identified a family of ratio-cum-product estimator and deduced that members of the class are estimators with optimal performance. The optimal estimators were better than the classical double sampling ratio estimator. Further researches on this area led Singh and Choudhury (2012) to propose a class of product-cum-dual to ratio estimators for estimating the population mean and obtained the bias and mean square errors of members of this class in two different cases of double sampling. The asymptotically optimum estimators (AOE) of the class were identified. They further showed that the asymptotical optimum estimators

were more efficient than other estimators but were as efficient as the regression estimator. Singh *et.al* (2012) suggested a double sampling version of Singh and Tailor [2005(a, b)] estimator along with its properties. The proposed estimator was found to have greater efficiency than the usual unbiased estimator, usual doubling sampling ratio and product estimators. Tailor and Sharma (2013) also proposed the double sampling version of Tailor and Sharma (2009) and studied its properties under two cases – when a subsample from the first phase is drawn and when the subsample is drawn directly from the population, independently of the first phase sample. They discovered that the proposed estimator was of a greater efficiency than the usual unbiased estimator, classical ratio estimator in two-phase sampling, double sampling version of Singh *et.al* (2004) estimator, double sampling version of Sisodia and Dwivedi (1981) and double sampling version of Upadhyaya and Singh (1999) estimators. They finally discovered that taking a subsample independent of the first phase sample was more beneficial when using this type of estimator. However, the proposed estimator was less efficient than the regression estimator in two-phase sampling. In the quest for estimators with greater efficiency, Ozgul and Cingi (2014) proposed a new procedure for estimating the finite population mean using the supplementary variable under double sampling scheme. The bias and Mean Square Error (MSE) of the suggested estimator were derived and efficiency comparison done with other related existing estimators. They deduced that the proposed estimator was always more efficient than classical ratio and regression estimators and Singh and Vishwakarma (2007) exponential estimator in double sampling. In a related development, Khatua and Mishra (2013) proposed a modified exponential method of estimation to estimate finite population mean. They deduced that the new estimator has a greater gain in efficiency than the usual double sample ratio, product, regression estimators and the modified estimators suggested by Singh *et.al* (2007). Singh, Sharma and Tarray (2015) came up with a correction of the Bias and Mean Square Error (MSE) of Khatua and Mishra (2013) estimator and further suggested a new class of exponential ratio and product-type estimators for estimating the finite population mean. They showed that Khatua and Mishra (2013) was a member of this class. They also claimed that their class of estimators at optimal values of parameters was more efficient than other estimators including the regression estimator in two-phase sampling.

Most of the existing estimators have been found to be more efficient when compared with other ratio estimators, however, few comparisons of greater efficiencies over the regression estimators have been made under this scheme.This work therefore seeks to address the above problem by putting forward an alternative class of ratio estimators of population mean under double sampling scheme with improved gain in efficiency under certain conditions.

Review of existing estimators of population mean in two-phase sampling with a single auxiliary variable

Under double sampling scheme, let $\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$ be a finite population of N units. Let *Y* denote the study variable and *X* be the auxiliary variable that will have values on π. To estimate the population mean $\bar{Y}$ of *Y*, consider two cases:

Case I: A large preliminary sample of size n' is selected by simple random sampling without replacement (SRSWOR) from population π of N units. Information on auxiliary variable (X) is obtained from all the n' units and used to estimate, $\bar{X}$. A second subsample of size n units $(n < n')$ is selected from the first phase sample units by simple random sampling without replacement. Information on y and x are obtained from this second phase subsample.

Case II: A second sample of size n, after the first phase sample, is selected from the population, independent of the first phase sample, and information on both the auxiliary and main character are obtained from this sample.
Table 1 presents some related existing estimators under this scheme.

TABLE 1 Some existing and related estimators of population mean in two-phase sampling using a single auxiliary variable and their Mean Square Errors (MSE)

S/N	Estimators	MSE
1.	$\bar{y}$, Sample Mean	$\lambda\bar{Y}^2C_y^2$
2.	$\bar{y}_{dr} = \bar{y}(\frac{\bar{x}'}{\bar{x}})$ Classical Ratio	$\bar{Y}^2\{\lambda C_y^2 + \kappa[C_x^2 - 2\rho C_y C_x]\}$
3.	$\bar{y}_{dlr} = \bar{y} + b(\bar{x}' - \bar{x})$ Regression Estimator	$\bar{Y}^2C_y^2[\lambda - \kappa\rho^2]$
4.	$\bar{y}_{svr} = \bar{y}\exp(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}})$ Singh andVishwakarma (2007)	$\bar{Y}^2\{\lambda C_y^2 + \kappa\left[\frac{C_x^2}{4} - \rho C_y C_x\right]\}$
5.	$\bar{y}_{NH} = \{k_{1N}\bar{y} + k_{2N}(\bar{x}' - \bar{x})\}\exp(\frac{\bar{z}_n' - \bar{z}_n}{\bar{z}_n' + \bar{z}_n})$ Ozgul andCingi (2013)	$\frac{\bar{Y}^2\{Var(\bar{y}_{dlr})[1 - \kappa\theta_N^2 C_x^2] - \frac{\kappa^2\bar{Y}^2\theta_N^2 C_x^2}{4}\}}{\bar{Y}^2 + Var(\bar{y}_{dlr})}$
6.	$\bar{y}_{KM} = \bar{y}[d_1 + d_2(\bar{x}' - \bar{x})]\exp(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}})$ Khatua and Mishra (2013)	$\frac{\bar{Y}^2C_y^2[\lambda - \kappa\rho^2]}{[1 + C_y^2(\lambda - \kappa\rho^2)]}$

where

$$\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i \text{ , population mean of the study variable}$$

$$\bar{X} = N^{-1}\sum_{i=1}^{N} y_i \text{ , population mean of the auxiliary variable}$$

$$\bar{x}' = n'^{-1}\sum_{i=1}^{n} \bar{x}_i' \text{ , sample mean of the auxiliary variable in the first phase sample}$$

$$\bar{x} = n^{-1}\sum_{i=1}^{N} x_i \text{ , sample mean of the auxiliary variable in the second sample}$$

$$\bar{y} = n^{-1}\sum_{i=1}^{N} y_i \text{ , sample mean of the study variable in the second sample}$$

$$\bar{z}_n = a\bar{x} + b, \bar{z}_n' = a\bar{x}' + b$$

$$\theta_N = \frac{a\bar{X}}{2(a\bar{X} + b)}, \qquad \lambda = \frac{1}{n} - \frac{1}{N}, \qquad \lambda' = \frac{1}{n'} - \frac{1}{N}, \kappa = \lambda - \lambda'$$

Proposed estimators for double or two-phase sampling using a single auxiliary variable

The proposed class of estimator of population mean in double sampling strategy is derived from the proposed estimator of the population mean with the least mean squared error in simple random sampling. Therefore, extending the proposed class of estimator to the double sampling procedure we have;

$$\bar{y}_{prd} = \theta_{1d}\bar{y} + \theta_{2d}(\bar{x}' - \bar{x})exp\left[\frac{\delta(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})}\right] \qquad (1)$$

Deriving the Bias and Mean Square Error of this estimator, we express (1) in terms of $e's$ as:

$$\bar{y}_{prd} = \theta_{1d}\bar{Y}(1 + e_y) + \theta_{2d}\bar{X}(e_x' - e_x)exp\left[\frac{\delta'}{2}\frac{(e_x' - e_x)}{\left(1 + \frac{e_x' - e_x}{2}\right)}\right]$$

$$= \theta_{1d}\bar{Y}(1 + e_y) + \theta_{2d}\bar{X}Vexp\left[\frac{\delta'}{2}V\left(1 + \frac{U}{2}\right)^{-1}\right]$$

$$= \theta_{1d}\bar{Y}(1 + e_y) + \theta_{2d}\bar{X}V\left[1 + \frac{\delta'}{2}V\left(1 + \frac{U}{2}\right)^{-1} + \frac{\delta'^2}{8}V^2\left(1 + \frac{U}{2}\right)^{-2} + \cdots\right]$$

To the first degree approximation, we obtain;

$$\bar{y}_{prd} = \bar{Y}\left(\theta_{1d} + \theta_{1d}e_y + \theta_{2d}MV + \frac{\delta'}{2}\theta_{2d}MV^2\right)$$

$$= \bar{Y}\left[\theta_{1d} + \theta_{1d}e_y + \theta_{2d}M(e_x' - e_x) + \frac{\delta'}{2}\theta_{2d}M(e_x' - e_x)^2\right]$$

$$= \bar{Y}\left[\theta_{1d} + \theta_{1d}e_y + \theta_{2d}Me_x' - \theta_{2d}Me_x + \frac{\delta'}{2}\theta_{2d}M\left(e_x'^2 - 2e'e_x + e_x^2\right)\right]$$

$$= \bar{Y}\left[\theta_{1d} + \theta_{1d}e_y + \theta_{2d}Me_x' - \theta_{2d}Me_x + \frac{\delta'}{2}\theta_{2d}Me_x'^2 - \delta'\theta_{2d}Me'e_x + \frac{\delta'}{2}\theta_{2d}Me_x^2\right]$$

$$\bar{y}_{prd} - \bar{Y} = \bar{Y}\left[(\theta_{1d} - 1) + \theta_{1d}e_y + \theta_{2d}Me_x' - \theta_{2d}Me_x + \frac{\delta'}{2}\theta_{2d}Me_x'^2 - \delta'\theta_{2d}Me'e_x + \frac{\delta'}{2}\theta_{2d}Me_x^2\right] (2)$$

Therefore, the bias of the estimator can be obtained from (2) as:

$$B(\bar{y}_{prd}) = E(\bar{y}_{prd} - \bar{Y})$$

$$= E\left\{\bar{Y}\left[(\theta_{1d} - 1) + \theta_{1d}e_y + \theta_{2d}Me_x' - \theta_{2d}Me_x + \frac{\delta'}{2}\theta_{2d}Me_x'^2 - \delta'\theta_{2d}Me_x'e_x + \frac{\delta'}{2}\theta_{2d}Me_x^2\right]\right\} \quad (3)$$

The first order approximation of the mean square error is therefore given as:

$$MSE(\bar{y}_{prd}) = E(\bar{y}_{prd} - \bar{Y})^2$$

$$= E\left\{\bar{Y}^2\left[\begin{array}{l}(\theta_{1d} - 1)^2 + 2(\theta_{1d} - 1)\frac{\delta'}{2}\theta_{2d}Me_x'^2 - 2(\theta_{1d} - 1)\delta'\theta_{2d}Me_x'e_x + (\theta_{1d} - 1)\delta'\theta_{2d}Me_x^2 \\ +\theta_{1d}e_y^2 + 2\theta_{1d}\theta_{2d}Me_ye_x' - 2\theta_{1d}\theta_{2d}Me_ye_x + \theta_{2d}^2M^2e_x'^2 - 2\theta_{1d}^2M^2e_x'e_x + \theta_{2d}^2M^2e_x^2\end{array}\right]\right\}$$

$$= E\left\{\bar{Y}^2\begin{bmatrix}\theta_{1d}^2 - 2\theta_{1d} + 1 + (\theta_{1d}\theta_{2d} - \theta_{2d})\delta' M e_x'^2 - M e_x' e_x + 2\theta_{2d}\delta' M e_x' e_x \\ +\theta_{1d}\theta_{2d}\delta' M e_x^2 - \theta_{2d}\delta' M e_x^2 + \theta_{1d}^2 e_y^2 + 2\theta_{1d}\theta_{2d} M e_y e_x' - 2\theta_{1d}\theta_{2d} M e_y e_x \\ +\theta_{2d}^2 M^2 e_x'^2 - 2\theta_{2d}^2 M^2 e_x' e_x + \theta_{2d}^2 M^2 e_x^2\end{bmatrix}\right\}$$

$$= E\left\{\bar{Y}^2\begin{bmatrix}1 + \theta_{1d}^2\left(1 + e_y^2\right) - 2\theta_{1d} \\ +\theta_{1d}\theta_{2d}\left(\delta' M e_x'^2 - 2\delta' M e_x' e_x + \delta' M e_x^2 + 2M e_y e_x' - 2M e_y e_x\right) \\ -\theta_{2d}\left(\delta' M e_x'^2 - 2\delta' M e_x' e_x + \delta' M e_x^2\right) + \theta_{2d}^2\left(M^2 e_x'^2 - 2M^2 e_x' e_x + M^2 e_x^2\right)\end{bmatrix}\right\}$$

$$= \bar{Y}^2\left\{\begin{matrix}1 + \theta_{1d}^2\left(1 + \lambda C_y^2\right) - 2\theta_{1d} \\ +\theta_{1d}\theta_{2d}\left[\delta' M(\lambda' C_x^2 - 2\lambda' C_x^2 + \lambda C_x^2) + 2M\left(\lambda'\rho C_y C_x - \lambda\rho C_y C_x\right)\right] \\ -\theta_{2d}\delta' M(\lambda' C_x^2 - 2\lambda' C_x^2 + \lambda C_x^2) + \theta_{2d}^2 M^2(\lambda' C_x^2 - 2\lambda' C_x^2 + \lambda C_x^2)\end{matrix}\right\}$$

$$where\ E\left(e_x'^2\right) = \left(\frac{1}{n'} - \frac{1}{N}\right)C_x^2 = \lambda' C_x^2, \qquad \lambda' = \frac{1}{n'} - \frac{1}{N}$$

$$E(e_x' e_x) = \left(\frac{1}{n'} - \frac{1}{N}\right)C_x^2 = \lambda' C_x^2, \qquad E(e_x^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2 = \lambda C_x^2$$

$$E\left(e_y e_x'\right) = \left(\frac{1}{n'} - \frac{1}{N}\right)\rho C_y C_x = \lambda'\rho C_y C_x, \ \ E\left(e_y e_x\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_y C_x = \lambda\rho C_y C_x$$

$$MSE\left(\bar{y}_{prd}\right) = \bar{Y}^2\left\{\begin{matrix}1 + \theta_{1d}^2\left(1 + \lambda C_y^2\right) - 2\theta_{1d} \\ +\theta_{1d}\theta_{2d}\left[\delta' M(\lambda C_x^2 - \lambda' C_x^2) + 2M\left(\lambda'\rho C_y C_x - \lambda\rho C_y C_x\right)\right] \\ -\theta_{2d}\delta' M(\lambda C_x^2 - \lambda' C_x^2) + \theta_{2d}^2 M^2(\lambda C_x^2 - \lambda' C_x^2)\end{matrix}\right\}$$

$$= \bar{Y}^2\left\{\begin{matrix}1 + \theta_{1d}^2\left(1 + \lambda C_y^2\right) - 2\theta_{1d} + \theta_{1d}\theta_{2d}\left[\delta' M C_x^2(\lambda - \lambda') + 2M\rho C_y C_x(\lambda' - \lambda)\right] \\ -\theta_{2d}\delta' M C_x^2(\lambda - \lambda') + \theta_{2d}^2 M^2 C_x^2(\lambda - \lambda')\end{matrix}\right\}$$

$$= \bar{Y}^2\left\{\begin{matrix}1 + \theta_{1d}^2\left(1 + \lambda C_y^2\right) - 2\theta_{1d} - 2\theta_{1d}\theta_{2d} M\kappa C_x^2\left(K - \frac{\delta'}{2}\right) - 2\theta_{2d}\delta' M\frac{\kappa C_x^2}{2} \\ +\theta_{2d}^2 M^2\kappa C_x^2\end{matrix}\right\}$$

$$= \bar{Y}^2\{1 + \theta_{1d}^2 r_1 - 2\theta_{1d} - 2\theta_{1d}\theta_{2d} M r_2' - 2\theta_{2d} M r_3' + \theta_{2d}^2 M^2 r_4'\} \qquad \ldots(4)$$

$$where\ r_1 = \left(1 + \lambda C_y^2\right), \ \ r_2' = \kappa C_x^2\left(K - \frac{\delta'}{2}\right), \ \ r_3' = \delta'\frac{\kappa C_x^2}{2}, \ \ r_4' = \kappa C_x^2, \ \ \kappa = \lambda - \lambda' = \left(\frac{1}{n} - \frac{1}{n'}\right)$$

To obtain the optimal values of θ_{1d}, θ_{2d} and δ' that will minimize (4), we set the following conditions based on our proposition:

(i) $0 < \theta_{1d} \leq 1\ or\ \theta_{1d} \leq 1, \theta_{1d} \geq 0$

(ii) $0 \leq \theta_{2d} \leq \infty\ or\ \ \theta_{2d} \geq 0$

(iii) $\delta' \geq 2$

(iv) $\theta_{1d} \geq 0,\ \ \theta_{2d} \geq 0,\ \ \delta' \geq 0$

Hence, the non-linear programming problem for this case is stated as follows:

Minimize $MSE\left(\bar{y}_{prd}\right)$

Subject to

$$\left.\begin{matrix}-\theta_{1d} \geq -1 \\ \theta_{2d} \geq 0 \\ -\delta' \geq -2\end{matrix}\right\} \qquad \ldots(5)$$

$\theta_{1d}, \theta_{2d}, \delta' \geq 0$

Solving (5) using the Lagrange multiplier method, we use the same procedure as in the case of simple random sampling. Thus, the general objective function with the Lagrange multipliers ζ_i is:

$$G' = \bar{Y}^2 \begin{Bmatrix} 1 + \theta_{1d}^2 r_1 - 2\theta_{1d} - 2\theta_{1d}\theta_{2d}Mr_2' - 2\theta_{2d}Mr_3' + \theta_{2d}^2 M^2 r_4' - \zeta_1(1-\theta_{1d}) \\ -\zeta_2\theta_{2d} - \zeta_3(\delta - 2) \end{Bmatrix}$$

To obtain the values of the unknown variables in the problem, the Kuhn Tucker conditions below are being investigated.

$$\frac{\partial MSE(\bar{y}_{prd})}{\partial \theta_{1d}} = 2\theta_{1d}r_1 - 2 - 2\theta_{2d}Mr_2' + \zeta_1 = 0 \quad \text{... (6)}$$

$$\frac{\partial MSE(\bar{y}_{prd})}{\partial \theta_{2d}} = -2\theta_{1d}Mr_2' - 2Mr_3' + 2\theta_{2d} - \zeta_2 = 0 \quad \text{... (7)}$$

$$\frac{\partial MSE(\bar{y}_{prd})}{\partial \delta'} = \frac{2\theta_{1d}\theta_{2d}MKC_x^2}{2} - \frac{2\theta_{2d}MKC_x^2}{2} - \zeta_3 = 0$$

$$\Longrightarrow \ \theta_{1d}\theta_{2d} - \theta_{2d} - \zeta_3 = 0 \quad \text{... (8)}$$

$$\zeta_1(1-\theta_{1d}) = 0 \quad \text{... (9)}$$

$$\zeta_2\theta_{2d} = 0 \quad \text{... (10)}$$

$$\zeta_3(2-\delta') = 0 \quad \text{... (11)}$$

$$1 - \theta_{1d} \geq 0 \quad \text{... (12)}$$

$$\theta_{2d} \geq 0 \quad \text{... (13)}$$

$$2 - \delta' \geq 0 \quad \text{... (14)}$$

$$\zeta_i \leq 0, i = 1,2,3 \quad (15)$$

Solutions that correspond to the following combinations of $\zeta_i{}'s$ $(i = 1,2,3)$ can be obtained using the procedure earlier followed.

(i) $\zeta_1 = 0$, $\zeta_2 \neq 0$, $\zeta_3 \neq 0$
(ii) $\zeta_1 \neq 0$, $\zeta_2 = 0$, $\zeta_3 \neq 0$
(iii) $\zeta_1 \neq 0$, $\zeta_2 \neq 0$, $\zeta_3 = 0$
(iv) $\zeta_1 = 0$, $\zeta_2 = 0$, $\zeta_3 \neq 0$
(v) $\zeta_1 = 0$, $\zeta_2 \neq 0$, $\zeta_3 = 0$
(vi) $\zeta_1 \neq 0$, $\zeta_2 = 0$, $\zeta_3 = 0$
(vii) $\zeta_1 \neq 0$, $\zeta_2 \neq 0$, $\zeta_3 \neq 0$
(viii) $\zeta_1 = 0$, $\zeta_2 = 0$, $\zeta_3 = 0$

It is now observed that solutions to combinations (iv), (v) and (vii) give the set of feasible optimal solutions required for the minimization problem. Out of these solutions, the best solution which is the solution that gives the least mean square is adopted.

For condition (iii), the solution is

$$\theta_{1d} = 1,\ \theta_{2d} = 0 \text{ from (9) and (10) respectively.}$$

Also, from (6),

$$\zeta_1 = 2 + 2r_1 \geq 0$$

and from (7),

$\zeta_3 = 2M(r_2' + r_3')$, which is either negative or positive.

Since ζ_1 is non-negative, one of the the Kuhn Tucker conditions (15) is not satisfied, then the solution $\theta_{1d} = 1,\ \theta_{2d} = 0,\quad \delta' \geq 0$ is not optimal; though it gives the simple random sample mean as the estimator.

For combination (iv), the solution from (11) is $\delta' = 2$. Putting $\delta' = 2$ in (6) and (7) yields

$$2\theta_{1d}r_1 - 2 - 2\theta_{2d}Mr_{22}' = 0; where\ r_{22}' = \kappa C_x^2(K-1)$$

$$\Longrightarrow \theta_{1d}r_1 - \theta_{2d}Mr_{22}' = 1 \quad \text{.... (16)}$$

and

$$-2\theta_{1d}Mr_{22}' - 2Mr_4' + 2\theta_{2d}M^2r_4' = 0$$

$$\Longrightarrow \theta_{1d}r_{22}' - \theta_{2d}Mr_4' = -r_4' \quad \text{... (17)}$$

Solving (16) and (17) gives

$$\theta_{1d_2} = \frac{r_4' + r_4' r_{22}'}{r_1 r_4' - {r_{22}'}^2} \quad \ldots (18)$$

$$\theta_{2d_2} = \frac{R(r_{22}' + r_1 r_4')}{r_1 r_4' - {r_{22}'}^2} \quad \ldots (19)$$

Substituting (18) and (19) in (4) gives the optimal mean square error for this condition as:

$$MSE(\bar{y}_{prd2}) = \bar{Y}^2 \left[1 - \left(\frac{r_4' + 2r_4' r_{22}' + r_1 {r_4'}^2}{r_1 r_4' - {r_{22}'}^2}\right)\right] \quad \ldots (20)$$

where $r_{22}' = \kappa C_x^2 (K - 1)$

For condition (v), the solution is obtained as:

From (10), $\theta_{2d} = 0$ and from (6), $\theta_{1d} = \frac{1}{r_1}$. Therefore, substituting these results in (4) gives the Mean Square Error of this condition as:

$$MSE(\bar{y}_{prd1}) = \bar{Y}^2 \left(\frac{r_1 - 1}{r_1}\right) = \frac{\lambda \bar{Y}^2 C_y^2}{1 + \lambda C_y^2} \quad (21)$$

Solution for condition (vi) is obtained as follows:

From (9), $\theta_{1d} = 1$ and from (11), if $\lambda_3 = 0$ then $2 - \delta' \neq 0$. Substituting $\theta_{1d} = 1$ in (7), we have:

$$-2Mr_2' - 2Mr_3' + 2\theta_{2d} M^2 r_4' = 0$$

$$\Rightarrow \frac{r_2' + r_3'}{Mr_4'} = \theta_{2d} \quad \ldots (22)$$

Inserting θ_{2d} in (4) provides the mean square error as:

$$MSE(\bar{y}_{prd3}) = \bar{Y}^2 \left[1 + r_1 - 2 - 2r_2' \left(\frac{r_2' + r_3'}{r_4'}\right) - 2r_3' \left(\frac{r_2' + r_3'}{r_4'}\right) + \frac{(r_2' + r_3')^2}{r_4'}\right]$$

$$= \bar{Y}^2 \left[r_1 - 1 - \frac{(r_2' + r_3')^2}{r_4'} \left(\frac{2r_2' + 2r_3' - r_2' - r_3'}{r_2' + r_3'}\right)\right]$$

$$= \bar{Y}^2 \left[r_1 - 1 - \frac{(r_2' + r_3')^2}{r_4'} \left(\frac{r_2' + r_3'}{r_2' + r_3'}\right)\right]$$

$$= \bar{Y}^2 \left[r_1 - 1 - \frac{(r_2' + r_3')^2}{r_4'}\right] \quad \ldots (23)$$

If the expressions for r_1, r_2', r_3' and r_4' are substituted in (23), the resulting expression is the mean square error of $\bar{y}_{prd3}$ expressed in terms of C_y, C_x and ρ as:

$$MSE(\bar{y}_{prd3}) = \bar{Y}^2 C_y^2 [\lambda - \kappa \rho^2] \quad \ldots (24)$$

Remark

1. It is observed here in case (vi) that the optimal value of δ' could not be obtained. Therefore, varying the values of δ' within its range of values produces different estimators of the population mean with the same mean square error.
2. If the values of δ' in case (iv) are varied within the range of δ', different estimators of the population mean in double sampling would be obtained. A summary of these estimators and their mean square errors are presented in Table 2.

TABLE 2. Some members of the proposed family of estimators of population mean in double sampling using single auxiliary variable (Case I)

Estimators	θ_{1d}	θ_{2d}	δ'	MSE
$\bar{y}_{prd_1} = \theta_{1d_1}\bar{y}$	1	0	δ'	$\frac{\lambda\bar{Y}^2C_y^2}{1+\lambda C_y^2}$
$\bar{y}_{prd_2} = \theta_{1d_2}\bar{y} + \theta_{2d_2}(\bar{x}' - \bar{x})exp\left[\frac{2(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})}\right]$	θ_{1d_2}	θ_{2d_2}	2	$\bar{Y}^2\left[1 - \left(\frac{r_4' + 2r_4'r_{22}' + r_1r_4'^2}{r_1r_4' - r_{22}'^2}\right)\right]$
$\bar{y}_{prd_3} = \bar{y} + b(\bar{x}' - \bar{x})exp\left[\frac{(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})}\right]$	1	b	δ'	$\bar{Y}^2C_y^2[\lambda - \kappa\rho^2]$
$\bar{y}_{prd_4} = \theta_{1d_4}\bar{y} + \theta_{2d_4}(\bar{x}' - \bar{x})exp\left[\frac{(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})}\right]$	θ_{1d_4}	θ_{2d_4}	1	$\bar{Y}^2\left[1 - \left(\frac{r_4' + 2r_{24}'r_{34}' + r_1r_{34}'^2}{r_1r_4' - r_{24}'^2}\right)\right]$
$\bar{y}_{prd_5} = \theta_{1d_5}\bar{y} + \theta_{2d_5}(\bar{x}' - \bar{x})$	θ_{1d_5}	θ_{2d_5}	0	$\bar{Y}^2\left[1 - \left(\frac{r_4'}{r_1r_4' - r_{25}'^2}\right)\right]$

$$r_{24}' = \kappa C_x^2\left(K - \frac{1}{2}\right),\quad r_{34}' = \frac{\kappa C_x^2}{2},\qquad r_{25}' = \kappa C_x^2K,\qquad r_{22}' = \kappa C_x^2(K-1), r_4' = \kappa C_x^2$$

Efficiency Comparison

Let the mean square error of the proposed estimator be $MSE(\bar{y}_{pri})$ and the mean square error of estimator to be compared with the proposed estimator be $MSE(.)$, then the percent relative efficiency (PRE) is given as:

$$PRE = \frac{MSE(.)}{MSE(\bar{y}_{pri})} \times 100 \qquad \ldots(25)$$

When (25) is greater than 100, it would be concluded that the proposed estimator is better than the other estimators; otherwise, the other estimators are more efficient. Therefore, the percent relative efficiency is used in comparing the efficiencies of the proposed estimators with other estimators.

Case II: In this case II, the second phase sample is taken independent of the first phase sample.

This means that the second phase sample is obtained from the population instead of the first phase preliminary sample; in which case,

$$\left.\begin{aligned} Cov(\bar{y}, \bar{x}') = 0 \\ Cov(\bar{x}, \bar{x}') = 0 \end{aligned}\right\} \qquad \ldots(26)$$

Therefore, the mean square error of the estimator in (7) would now be:

$$MSE(\bar{y}_{prd}') = E\left\{\bar{Y}^2\left[\begin{matrix} 1 + \theta_{1d}'^2(1 + e_y^2) - 2\theta_{1d}' + \theta_{1d}'\theta_{2d}'(\delta' Me_x'^2 + \delta Me_x^2 - 2Me_ye_x) \\ -\theta_{2d}'(\delta' Me_x'^2 + \delta' Me_x^2) + \theta_{2d}'^2(M^2e_x'^2 + M^2e_x^2) \end{matrix}\right]\right\}$$

$$= \bar{Y}^2\left\{\begin{matrix} 1 + \theta_{1d}'^2(1 + \lambda C_y^2) - 2\theta_{1d}' + \theta_{1d}'\theta_{2d}'[\delta' M(\lambda' C_x^2 + \lambda C_x^2) - 2\lambda M\rho C_yC_x] \\ -\theta_{2d}'\delta' M(\lambda' C_x^2 + \lambda C_x^2) + \theta_{2d}'^2M^2(\lambda' C_x^2 + \lambda C_x^2) \end{matrix}\right\}$$

$$= \bar{Y}^2 \left\{ \begin{array}{c} 1 + \theta_{1d}'^2 (1 + \lambda C_y^2) - 2\theta_{1d}' + 2\theta_{1d}'\theta_{2d}' \left[\frac{\delta' M C_x^2 (\lambda' + \lambda)}{2} - \lambda M \rho C_y C_x \right] \\ -2\theta_{2d}' \frac{\delta' M C_x^2 (\lambda' + \lambda)}{2} + \theta_{2d}'^2 M^2 C_x^2 (\lambda' + \lambda) \end{array} \right\} \quad \ldots (27)$$

$$= \bar{Y}^2 (1 + \theta_{1d}'^2 r_1 - 2\theta_{1d}' + 2\theta_{1d}'\theta_{2d}' M r_2'' - 2\theta_{2d}' M r_3'' + \theta_{2d}'^2 M^2 r_4'') \quad \ldots (28)$$

where $r_1 = (1 + \lambda C_y^2)$, $r_2'' = \frac{\delta' C_x^2 (\lambda' + \lambda)}{2} - \lambda \rho C_y C_x$, $r_3'' = \frac{\delta' C_x^2 (\lambda' + \lambda)}{2}$, $r_4'' = C_x^2 (\lambda' + \lambda)$

The optimal $MSE(\bar{y}_{prd}')$ would be

$$MSE_{opt}(\bar{y}_{prd}') = \bar{Y}^2 \left[1 - \frac{(r_4'' + 2r_3'' r_2'' + r_1 r_3'')}{r_1 r_4'' - r_2''^2} \right] \quad \ldots (29)$$

Following the same procedure as in Case I and varying the values of δ', one can have the corresponding estimators and their MSEs as given in Table 3

TABLE 3 Some members of the proposed family of estimators of population mean in two-phase sampling (Case II)

Estimators	θ_{1d}	θ_{2d}	δ'	MSE
$\bar{y}_{prd_1}' = \theta_{1d_1}' \bar{y}$	1	0	δ'	$\frac{\lambda \bar{Y}^2 C_y^2}{1 + \lambda C_y^2}$
$\bar{y}_{prd_2}' = \theta_{1d_2}' \bar{y} + \theta_{2d_2}' (\bar{x}' - \bar{x}) exp \left[\frac{2(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})} \right]$	θ_{1d_2}	θ_{2d_2}	2	$\bar{Y}^2 \left[1 - \left(\frac{r_4' + 2r_4' r_{22}' + r_1 r_4'^2}{r_1 r_4' - r_{22}'^2} \right) \right]$
$\bar{y}_{prd_3}' = \bar{y} + b(\bar{x}' - \bar{x}) exp \left[\frac{(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})} \right]$	1	b	δ'	$\bar{Y}^2 C_y^2 [\lambda - \kappa \rho^2]$
$\bar{y}_{prd_4}' = \theta_{1d_4}' \bar{y} + \theta_{2d_4}' (\bar{x}' - \bar{x}) exp \left[\frac{(\bar{x}' - \bar{x})}{(\bar{x}' + \bar{x})} \right]$	θ_{1d_4}	θ_{2d_4}	1	$\bar{Y}^2 \left[1 - \left(\frac{r_4' + 2r_{24}' r_{34}' + r_1 r_{34}'^2}{r_1 r_4' - r_{24}'^2} \right) \right]$
$\bar{y}_{prd_5}' = \theta_{1d_5}' \bar{y} + \theta_{2d_5}' (\bar{x}' - \bar{x})$	θ_{1d_5}	θ_{2d_5}	0	$\bar{Y}^2 \left[1 - \left(\frac{r_4'}{r_1 r_4' - r_{25}'^2} \right) \right]$

Numerical Illustrations

This section seeks to make use of numerical data to validate the theoretical results of this work.

TABLE 4 Data sets for numerical illustration in double sampling strategy using a single auxiliary variable

Source	Parameters							
	N	n'	n	ρ	C_y	C_x	$\bar{Y}$	$\bar{X}$
Cingi*et al* (2007)	923	400	200	0.955	1.72	1.86	436.3	114440.5
Murthy (1967)	80	30	10	0.9413	0.35426	0.75067	5182.64	1126.46
Kadilar&Cingi (2006)	104	40	20	0.865	1.866	1.653	13.93	625.37
Murthy (1967)	80	40	20	0.9413	0.354	0.751	51.826	11.265
Handique	34	10	7	0.98	0.75318	0.72	199.44	208.89

(2012)								
Handique (2012)	2500	200	25	0.79	0.95	0.98	4.63	21.09
Das(1988)	278	70	30	0.7213	1.4451	1.6198	39.068	25.111

Table 4 indicates data statistics employed by some authors in their works to validate their theoretical proposition of their estimators. They are also used to compare the efficiency of the proposed estimator of the population mean in two phase sampling with existing ones.

TABLE 5
MSEs and PREs of related existing and proposed exponential ratio-type estimators (Case I) of population mean in two-phase sampling

			Population				
Estimators	I	II	III	IV	V	VI	VII
Sample mean	2205.636	294954.1	54993.75	12.6222	2559.817	0.7661	94.7816
PRE	100	100	100	100	100	100	100
Ratio	944.1017	407516.490	29536.166	16.8864	1631.629	0.3831	72.8886
PRE	233.62	72.38	186.19	81.62	156.89	200.01	130.04
Singh &Vishwakarma (2007)	1163.2680	98974.950	35586.140	5.2863	1874.794	0.3944	64.7653
PRE	189.61	298.01	154.54	163.25	136.54	194.23	146.35
Regression	921.6096	95835.760	29521.360	5.1663	1631.070	0.3435	63.1944
PRE	239.32	307.77	186.28	164.96	156.94	223.02	149.98
Ozgul&Cingi (2013)	914.3010	92247.370	26960.900	5.117	1557.044	0.3343	59.7850
PRE	241.24	319.74	203.98	165.99	164.40	229.19	158.54
Khatua& Mishra (2013)	917.1690	95495.030	27449.340	5.1564	1566.821	0.3381	60.6819
PRE	240.48	308.87	200.35	165.43	163.38	226.59	156.19
$\bar{y}_{prd1}$	2180.3730	291750.300	48214.020	12.5631	2405.040	0.7397	89.2399
PRE	101.16	101.10	114.06	100.47	106.44	103.57	106.21
$\bar{y}_{prd2}$	**892.300**	**84808.020**	**22589.150**	**4.9866**	**1479.035**	**0.3074**	**55.3710**
PRE	**247.19**	**347.79**	**243.45**	**167.86**	**173.07**	**249.26**	**171.18**
$\bar{y}_{prd3}$	921.6096	95835.760	29521.360	5.1663	1631.070	0.3435	63.1944
PRE	239.32	307.77	186.28	164.96	156.94	223.02	149.98
$\bar{y}_{prd4}$	907.4798	92065.380	25403.920	5.0981	1527.886	0.3262	58.4412
PRE	243.05	320.37	216.48	166.38	167.54	234.86	162.18
$\bar{y}_{prd5}$	917.1692	95495.030	27449.340	5.1564	1566.821	0.3381	60.6819
PRE	240.48	308.87	200.35	165.43	163.38	226.59	156.19

Table 5 include existing estimators proposed by some authors and proposed estimators ($\bar{y}_{prdi}, i = 1, 2, \ldots, 5$) in two phase sampling using a single variable under case I, their Mean Square Errors and their Percent Relative Efficiencies.

TABLE 6

MSEs and PREs of proposed exponential ratio-type estimators of population mean in two phase sampling (Case II)

Estimator	Population						
	I	II	III	IV	V	VI	VII
$\bar{y}_{prd1}$	2180.3730	291750.300	48214.020	12.5631	2405.040	0.7397	89.2399
PRE	101.16	101.10	114.06	100.47	106.44	103.57	106.21
$\bar{y}'_{prd2}$	681.318*	66806.77*	14914.68*	3.9091*	760.982*	0.2959*	46.9213*
PRE	323.73*	441.50*	368.72*	322.89*	336.38*	258.91*	202.00*
$\bar{y}'_{prd3}$	921.6096	95835.760	29521.360	5.1663	1631.070	0.3435	63.1944
PRE	239.32	307.77	186.28	164.96	156.94	223.02	149.98
$\bar{y}'_{prd4}$	710.241	78600.580	20370.740	4.129	925.272	0.319	52.724
PRE	310.55	375.26	269.96	305.70	276.66	240.16	179.66
$\bar{y}'_{prd5}$	725.579	83608.440	23671.990	4.228	1017.617	0.333	56.348
PRE	303.98	352.78	232.32	298.54	251.55	230.06	168.19

Table 6 gives the Mean Square Errors (MSEs) and Percent Relative Efficiencies (PREs) of proposed estimators ($\bar{y}'_{prdi}, i = 1, 2, \ldots, 5$)of population mean in two-phase sampling (case II), when one auxiliary information is involved.

Discussion of Results

The mean square errors of the proposed estimator of population mean in two-phase sampling using a single auxiliary variable for both cases- I (when the second phase sample is selected from the first phase sample) and II (when the second phase sample is drawn independent of the first phase sample) are shown in equations (19) and (28) respectively. Optimal estimators with their MSEs obtained from this proposed estimator for case I are shown in Table 2, while those of case II are shown in Table 3. Numerical validation of these analytical results and the comparison of their performances with existing estimators are done using seven (7) populations and presented in Tables 5 and 6. In Table 5, estimator $\bar{y}_{prd2}$ has the least Mean Square Errors and the greatest Percent Relative Efficiencies in all the populations considered. From Table 5, it is also found that $\bar{y}_{prd2}$gives the greatest PRE of 247.19%, 347.79%, 243.45%, 167.86%, 173.07%, 249.26% and 171.18% in all the populations respectively among the existing estimators of Ozgul and Cingi (2013), Khatua and Mishra (2013), Singh and Vishwakarma (2007) and even the regression estimator. This shows that $\bar{y}_{prd2}$ has the greatest gain in efficiency over the existing estimators and the other proposed asymptotic optimum estimators in case I. Similarly, in case II shown in Table 6, it is observed that estimator $\bar{y}'_{prd2}$ has the least Mean Square Errors (MSE) and greatest Percent Relative Efficiencies (PREs) in all the considered populations. This is seconded by $\bar{y}'_{prd4}$. Estimator $\bar{y}'_{prd2}$ has the greatest PRE of 323.73%, 441.5%, 368.72%, 322.89%, 336.38%, 258.91%, and 202.0% in populations I, II, III, IV, V, VI and VII respectively. Authors of the existing estimators considered in this work did not consider their efficiencies in case II, but it has been observed here that $\bar{y}'_{prd2}$, which is the best estimator in case II has consistently possessed greater Percent Relative Efficiencies (PRE) in all the considered populations than any other estimators, both existing and proposed estimators in cases I and II. This efficiency property is at variance with the works of Singh and Vishwakarma (2007), Kalita *et.al* (2013) and Singh and Choudhoury (2012), which showed that their proposed estimators in Cases I and II sometimes perform better, in terms of efficiency, than one another depending on the type of populations considered. Therefore, the proposed estimator of population mean under this two-phase sampling using a single auxiliary variable is highly recommended for case II because of its high gain in efficiency, when the second sample is taken independently of the first phase sample.

Conclusion

Two new estimators of population mean under two-phase sampling strategies using a single auxiliary variable, under two cases, were also proposed in this work with their Biases and Mean Square Errors (MSE'S). Their efficiency comparison with other existing related estimators confirms that the proposed estimators have significant improvements in terms of efficiency over other estimators with $\bar{y}_{prd2}$ as the most efficient in case I, while $\bar{y}'_{prd2}$ has the greatest efficiency in case II for the given data set. These two estimators are better in efficiency than all existing estimators considered in this work with $\bar{y}'_{prd2}$ being the most efficient of all, for the given data set. Other estimators such as $\bar{y}_{prd4}$ and $\bar{y}'_{prd4}$ have also been found to be of improved efficiency than any existing estimators considered in this work.

References

Bahl, S. &Tuteja, R. K., (1991): Ratio and product type exponential estimator. *Journal of Information and Optimization Sciences.* Vol. 12(1), 159 – 164.

Choudhury, S. & Singh, B. K., [2012]: A class of chain ratio-cum-dual to ratio type estimator with two auxiliary characters under double sampling in sample surveys. *Statistics in Transition-New Series*.Vol. 13(3), 519—536.

Dash P. R & Mishra G (2011): An improved class of Estimators in two-phase sampling using two auxiliary variables. *Communications in Statistics-Theory and Methods*, 40: 4347-4352.

Hidiroglou, M. A and Sarndal, C. E (1998): Use of auxiliary Information for two-phase sampling. *Survey Methodology*, 24 (1), 11-20.

Handique, B. K (2012): A class of Regression-cum-Ratio estimators in two-phase sampling for utilizing information from high Resolution Satellite Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1-4, XXII* ISPRS Congress, Melbourne, Australia.

Khatua, P & Mishra, G. (2013): A Modified Generalized Two-phase Exponential Estimator to estimate finite Population Mean. *IOSR Journal of Mathematics (IOSR-JM)*, 8(6), 43-47.

Malik,S. & Tailor, R., (2013): Ratio-type Estimator of population mean double sampling. *International Journal of Advanced Mathematics and Statistics*, 1 (1), 34-39.

Okafor, F. C (2002): Sample Survey Theory with Applications. AFRO-ORBIS Publications, Nsukka, Nigeria.

Ozgul, N &Cingi, H. (2014): A new class of exponential regression-cum-ratio estimator in two phase sampling. *Hacettepe Journal of Mathematics and Statistics*, 43(1), 131-140.

Pradhan, B. K (2005): A chain Regression Estimator in two-phase sampling using multi-auxiliary information. *Bulletin of the Malaysian Mathematical Sciences Society(2)*, 28(1), 81-86.

Singh, G. N.(2001): On the use of transformed auxiliary variable in the estimation of population mean in two-phase sampling. *Statistics in Transition* 5(3), 405-416.

Singh, R., Chauhan, P.,Sawan, N. &Smarandache, F. (2012): Improvement in estimating population mean using two auxiliary variables in two-phase sampling. *Italian Journal of Pure and Applied Mathematics*, 28, 135-142.

Singh, H.P. & Tailor, R., [2005(a)]: Estimation of finite population mean with known coefficient of variation of an auxiliary character. *Statistica, Anno* LXV (3).

Singh, H. P. & Tailor, R., [2005(b)]: Estimation of finite population mean using known correlation coefficient between auxiliary characters. *Statistica, Anno* LXV, 4, 407 – 418.

Singh, H. P., Sharma, N &Tarray, T. A. (2015): An Efficient class of Two-phase exponential ratio and product type estimators for estimating the finite population mean. *Journal of Statistical Applications and Probability Letters*, 2(1), 71-88.

Singh, H. P &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, G. N &Upadhyaya L. N (1995): A class of modified chain-type estimators using auxiliary variable in two-phase sampling. *Metron LIII*, 117-125.

Singh, B. K &Choudhury, S (2012): Double Sampling Product-cum-Dual to Ratio Estimators of finite population mean in sample surveys. *International Journal of Scientific and Engineering Research*, 3(4), 1-7.

Sosidia, B. V. S. &Dwivedi, V. K., (1981): A modified ratio estimator using coefficient of variation of auxiliary variable. *Jour. Ind. Soc. Agri. Stat.*, 33(1), 13 – 18.

Singh, H. P &Vishwakarma, G. K (2007): Modified Exponential and Product Ratio Estimators for finite Population Mean in Double Sampling. *Austrian Journal of Statistics*, 36, (3), 217-225.

Singh, G. N &Majhi, D (2014): Some chain-type exponential estimators of population mean in two-phase sampling. *Statistics in Transition, new series*, 15(2), 221-230.

Singh, H. P, Upadhyaya, L. N & Chandra P. (2004): A general family of estimators for estimating population mean using two auxiliary variables in two phase sampling. *Statistics in Transition,* 6(7), 1055-1077.

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

Swain, A.K.P.C., (2012b): On some classes of modified ratio-type and regression type estimators in two-phase sampling. *Pakistan Journal of Statistics,* 28(4), 445-466.

Tailor, R. & Sharma, B. K. (2013): On the efficiency of ratio-cum-product estimator of finite population mean in double sampling when coefficient of variation and coefficient of kurtosis of auxiliary variable are known. *Journal of Applied Statistical Science*, 20(1), 1-20.

Tailor, R. & Sharma, B. K., (2009): A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis. *Statistics in Transition-New Series*. Vol. 10(1), 15 – 24.

Upadhyaya, L. N. & Singh, H. P., (1999): Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal* 41(5), 627 – 636.