Mathematics

Research article

Efficient estimation combining exponential and ln functions under two phase sampling

Yasir Hassan[1,*], Muhammad Ismail[1], Will Murray[2] and Muhammad Qaiser Shahbaz[3]

[1] Department of Statistics, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan
[2] Department of Mathematics and Statistics, California State University, Long Beach, USA
[3] Department of Mathematics and Statistics, King Abdulaziz University, Jeddah, Saudi Arabia

* Correspondence: Email: yasir.uol@hotmail.com; Tel: +923334427552.

Abstract: In this study, we propose the combination of exponential and ln ratio type estimator to estimate the mean of Y (Study Variable) by incorporating two auxiliary variables in two phase sampling scheme. Under simple random sampling without replacement, the illustration for mean square error and mathematical comparisons are presented. Several approaches are available in literature to estimate the study variable by using information on the variable of interest. The performance of our proposed estimator is compared with other ratio type estimators theoretically and empirically. It is observed that ratio and exponential ratio estimators considered by various researchers and usual unbiased estimator is less efficient than our proposed estimator. An efficiency comparison is also given using five data sets and simulation studies for checking the merits of our proposed estimator and outcomes are sound and moderately illuminating in comparison to different estimators.

1. Introduction

The use of survey sampling is useful in solving real world problems in the field of environmental changes, engineering, management and biological sciences. For instance, environmental sampling is a key tool to verify sources of pollution and adequacy of hygiene process,

refining the frequency and intensity of cleaning and sanitation, identifying problem areas, food safety validation programs, and to give an early cautioning of issues that may require remedial action. Sampling can be conducted to characterize background radiological levels, determine the concentration of radionuclides and make recommendations on environmental surveillance for agricultural products. Over all, it gives the confirmation that product being made will be made under clean conditions.

In case of forming travelling patterns of a certain city residents it is hard to approach every person of the city and ask the information on their travel pattern. A sample data is gathered and based on that information the developers try to understand the travel behavior of residents. Likewise, when choosing the trademark quality of cement, of a specific blend, data resulting from the tests on a sample is used. Similarly, a sample of borehole information is used to find quality soil. The sample is representative of population in all the cases and while doing so the representative ability of the sample data must be ensured statistically.

Velasco-Muñoz, et al. [1] reviewed 25 years of researches on sustainable water use in farming. A bibliometric examination was established to sample 2084 published articles from 1993 to 2017. Results showed that sustainable water used for agriculture attained exponential growth and has turned into a worldwide issue. Serbu, et al. [2] collected water samples from four different locations beside the Cibin River over a time of a year and applied Multiple-Criteria Decision Analysis strategies for the surveying the effect of toxins on the earth.

Ziemer, et al. [3] discussed that how data is collected and in what way the sample of electrical engineering departments is evolved that is used as database. De Mello, et al. [4] proposed theoretical framework that contributes to strengthen representativeness of their outcomes, though some key problems concerning survey study are still open and deserve attention from Software Engineering community. De Mello, et al. [5] described the heterogeneity and members who repeatedly contributed to increase the strength of survey's results. Consequently, De Mello and Travassos [5] believed that sharing of this experience, questionnaire and idea can be helpful for scientists interested on executing surveys on broader scale in Software Engineering.

The information retrieved from auxiliary variables has resulted in wide ranging gain in performance over estimators which don't take such information. When the auxiliary variable X is available in advance or observed easily and high correlation exists with study variable Y, the use of auxiliary information is effective to estimate population mean. In these situations ratio, regression and product estimators are good examples to use. Estimation method of regression was used by Watson [6] to estimate the mean area of leaves on plant. Cochran [7] proposed ratio method in case of strong positive correlation existence between Y and X (study and auxiliary variables). Murthy [8] revisited the idea given by Robson [9] that product method of estimation is appropriate if strong negative correlation exits between auxiliary and study variables. Srivastava [10] proposed the general ratio estimator using a single auxiliary variable and given that population information on this variable is not available. Exponential smoothing is one of the forecast methods to recognize substantial changes in data by incorporating the most recent information. Given that there are numerous different approaches to make forecasts, exponential smoothing is significantly easy to learn, forecasts accurately, and its application to recent observations gives this technique an edge over others.

Bahl, et al. [11] proposed exponential product estimator when study variable and auxiliary variable are negatively correlated and exponential ratio estimator, in case of positive correlation,

exists between Y and X. Numerous authors including Srivastava [10], Hidiroglou [12], Samiuddin, et al. [13], Singh, et al. [14], Singh, et al. [15], Hanif, et al. [16], Hanif, et al. [17], Singh, et al. [18], Singh, et al. [19],Noor-ul-Amin, et al. [20], Tailor, et al. [21] and Shabbir, et al. [22] have also proposed improved ratio and product for estimating the population average of study variable. Al-Marshadi, et al. [23] suggested estimator for estimating population variance by using multi auxiliary variables.

Generally, ratio and product estimators are less efficient than the linear regression estimator when it passes through the origin. It is observed, in most of the situations, that the regression line does not pass through the origin. Considering this fact, Vishwakarma, et al. [24] proposed the ratio product type estimator to improve the performance of Singh and Espejo [14] estimator using two auxiliary variables under two phase sampling.

Vishwakarma, et al. [25] proposed generalized class of estimators by using the information of multi-auxiliary variables in two-phase sampling scheme and claimed that their generalized class of estimators performed well over the Dash, et al. [26] proposed class of estimators.

Mishra, et al. [27] used log function and proposed a ln-product type estimator for estimating the mean value of study variable Y. Akhlaq, et al. [28] proposed an exponential estimator, which is most efficient than the previous estimators for estimating the process variability by using auxiliary information.

To estimate the population mean, considering the N population units $D = \left(D_1, D_2, D_3 .... D_N \right)$, Let y_i and $\left(x_i, z_i \right)$ be the values of study variable $\left(Y \right)$ and auxiliary variables $\left(X, Y \right)$ respectively. Population averages are $\bar{Y} = N^{-1} \sum_{i=1}^{N} Y_i$ and $\bar{X} = N^{-1} \sum_{i=1}^{N} X_i$ where $\bar{Y}$ is the population mean of study variable and $\bar{X}$ is population mean of auxiliary variable.

Auxiliary information in two-phase sampling proves to be effective in estimating the population mean. The first phase estimates the population variable X (auxiliary) in such a manner that when information of the auxiliary variable is not available, it then evaluates the second phase. The first phase sample selection is known as the primary sample, which is comprised of $n_1 \left(n_1 < N \right)$ units. We select a sample from 'N' units by applying simple random sampling without replacement (SRSWR) technique in the first phase. In some situations, another auxiliary variable $\left(Z \right)$ is helpful for obtaining the information of the first auxiliary variable X and both the variables are observed in the first phase sample.

The essential information is estimated in first phase for the auxiliary variable X before moving into the second phase. With the assistance of the auxiliary variable X, the population mean $\bar{Y}$ of Y is estimated in second phase. To serve this purpose, another sample of SRSWR is sorted out from a primary sample of n_2 units $\left(n_2 < n_1 \right)$ Singh, et al. [29]. The sample size n_1 from the population 'N' is a primary sample. The sample n_2 from n_1 is a subsample of our study.

The classical ratio estimator of population mean $\bar{Y}$ under double sampling scheme is given by Cochran [30] as:

$$\bar{Y}_r = \left(\bar{y}_2 \left(\bar{x}_2 \right)^{-1} \bar{x}_1 \right); \bar{x}_2 \neq 0 \tag{1.1}$$

where $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$ and $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$

Upadhyaya, et al. [31] used available information of coefficient of variation $\left(C_z\right)$ and coefficient of kurtosis $\left(\beta_{2(z)}\right)$ from auxiliary variable and proposed estimator for estimating study variable $\bar{Y}$ as:

$$\bar{Y}_{US} = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{C_z\bar{Z} + \beta_{2(Z)}}{C_z\bar{z}' + \beta_{2(Z)}}\right) \tag{1.2}$$

The MSE of [31] estimator up to the order $o\left(n\right)^{-1}$ as:

$$MSE\left(\bar{Y}_{US}\right) = \bar{Y}^2\left[f_1 C_y^2 + f_2\left(\theta_2^2 C_z^2 - 2\theta_2 C_y C_z \rho_{yz}\right) + f_3\left(C_x^2 - 2C_y C_x \rho_{xy}\right)\right] \tag{1.3}$$

A modified version of Bahl and Tuteja [11] exponential ratio estimator given by Singh and Vishwakarma [15] as:

$$\bar{Y}_{SV} = \bar{y}_2 \exp\left[\left(\bar{x}_1 - \bar{x}_2\right)\left(\bar{x}_1 + \bar{x}_2\right)^{-1}\right] \tag{1.4}$$

where $\bar{y}_2$ is the sample mean of study variable in second phase.

Mean square error of $\bar{Y}_{sv}$ is given as:

$$MSE\left(\bar{Y}_{SV}\right) = \bar{Y}^2 C_y^2\left[f_2 + \frac{C_x}{4C_y}\left(f_2 - f_1\right)\left(\frac{C_x}{C_y} - \rho_{xy}\right)\right]. \tag{1.5}$$

Exponential ratio estimator suggested by Singh, et al. [32] given as:

$$\bar{Y}_S = w_{0d}\bar{y}_2 + w_{1d}\bar{y}_{rsd} + w_{2d}\bar{y}_{rsed} \tag{1.6}$$

where $\bar{y}_2, \bar{y}_{rsd}$ and $\bar{y}_{rsed} \in w_r$, while w_r represents the set of all possible ratio-type estimator for estimating study variable $\bar{Y}_s$.

$$\sum_{i=0}^{2} w_{id} = 1, \; t_{rsd} = \bar{y}_2 \exp\left[\left(a\bar{x}_{(1)} + b\right)\left(a\bar{x}_{(2)} + b\right)^{-1}\right],$$

$$t_{rsed} = \bar{y}_2 \exp\left[\frac{\left(a\bar{x}_{(1)} + b\right) - \left(a\bar{x}_{(2)} + b\right)}{\left(a\bar{x}_{(1)} + b\right) + \left(a\bar{x}_{(2)} + b\right)}\right]$$

Minimum MSE of $\bar{Y}_S$ is given as:

$$MSE\left(\bar{Y}_S\right)_{min} = \bar{Y}^2 C_y^2\left(f_2 - \left(f_2 - f_1\right)\rho_{xy}^2\right) \tag{1.7}$$

Noor-ul-Amin and Hanif [20] suggested exponential estimator using ratio with product technique in double sampling as:

$$\bar{Y}_{NH} = \bar{y}_2 \exp\left[\left(\bar{Z} - z_2\right)\left(\bar{Z} + z_2\right)^{-1} - \left(\bar{x}_1 - \bar{x}_2\right)\left(\bar{x}_1 + \bar{x}_2\right)^{-1}\right]. \tag{1.8}$$

The expression for min (MSE) of the $\bar{Y}_{NH}$ is

$$MSE\left(\overline{Y}_{NH}\right) = \overline{Y}^2\left[\frac{f_2}{4}\left(4C_y^2 + C_z^2 - 4C_y C_z \rho_{yz}\right) + \frac{f_1}{4}\left(C_x^2 + 4C_y C_x \rho_{xy} - 2C_x C_z \rho_{xz}\right)\right]. \quad (1.9)$$

Sanaullah, et al. [33] suggested modification in Noor-ul-Amin and Hanif [20] estimator using two auxiliary variables as:

$$\overline{Y}_{SA} = \overline{y}_2 \exp\left[\alpha\left(\overline{Z} - z_2\right)\left(\overline{Z} + z_2\right)^{-1} - \left(1 - \alpha\right)\left(\overline{x}_1 - \overline{x}_2\right)\left(\overline{x}_1 + \overline{x}_2\right)^{-1}\right]. \quad (1.10)$$

Minimum mean square error of $\overline{Y}_{SA}$ is

$$MSE\left(\overline{Y}_{SA}\right) = \left[f_2\overline{Y}^2\left[C_y^2 + \hat{\alpha}^2 C_z^2 - 2C_y C_z \rho_{yz}\right] + f_1\overline{Y}^2\left(1 - \hat{\alpha}\right)\left[\begin{array}{c}\left(\dfrac{1-\hat{\alpha}}{4}\right)C_x^2 + \\ C_y C_x \rho_{yx} - \hat{\alpha} C_x C_z \rho_{xz}\end{array}\right]\right]. \quad (1.11)$$

Kadilar, et al. [34] incorporates two auxiliary variables in classical regression and proposed a new linear regression estimator as:

$$\overline{Y}_{kc} = \overline{y} + b_{xy}\left(\overline{X} - \overline{x}\right) + b_{zy}\left(\overline{Z} - \overline{z}\right) \quad (1.12)$$

The mean square error of $\overline{Y}_{kc}$ is given as:

$$MSE\left(\overline{Y}_{kc}\right) = f_1\overline{Y}^2 C_y^2\left(1 - \rho_{xy}^2 - \rho_{zy}^2 + 2\rho_{xy}\rho_{zy}\rho_{xz}\right) \quad (1.13)$$

where regression coefficients are $b_{xy} = \dfrac{s_{xy}}{s_x^2}$ and $b_{zy} = \dfrac{s_{zy}}{s_z^2}$ respectively.

Where

$$f_1 = \left(n_1^{-1} - N^{-1}\right), \; f_2 = \left(n_2^{-1} - N^{-1}\right), \; \left(f_2 - f_1\right) = f_3 = \left(n_2^{-1} - n_1^{-1}\right)$$

The coefficient of variation X is $C_x = s_x\left(\overline{X}\right)^{-1}$, Y is $C_y = s_y\left(\overline{Y}\right)^{-1}$ and Z is $C_z = s_z\left(\overline{Z}\right)^{-1}$ while correlation coefficient of $\left(Y, X\right)$ is $\rho_{yx} = s_{xy}\left(s_x s_y\right)^{-1}$, is $\left(Y, Z\right)$ $\rho_{yz} = s_{yz}\left(s_y s_z\right)^{-1}$ and $\left(X, Z\right)$ is $r_{xz} = s_{xz}\left(s_x s_z\right)^{-1}$.

The estimators given above are used for estimating mean of study variable in different conditions. The main purpose of this article is to suggest an improved exponential estimator of ratio type and explores its properties. A ln function and two auxiliary variables are used with in proposed estimator and discussed in following section. Third section is comprised of mathematical illustration of suggested estimator and it's comparison with some existing estimators $\overline{Y}, \overline{Y}_S, \overline{Y}_{SV}, \overline{Y}_{NH}$ and $\overline{Y}_{SA}$. In section four empirical study is carried out while conclusion is being discussed in last section.

2. Materials and method

Motivated by Cekim, et al. [35], who proposed ratio estimator using ln function for estimation of population variance, We propose a new estimator by combining the exponential and ln ratio terms using the information of two auxiliary variables Z and X in two-phase sampling for estimating

population mean $\overline{Y}$. Let n_1 and $n_2 < n_1$ be the sample sizes of first and second phase respectively, while an unbiased estimator of $\overline{Z}$ be $\overline{z}_1^* = \left(N\overline{Z} - n_1\overline{z}_{(1)}\right)\big/\left(N - n_1\right)$, as suggested by Bandyopadhyay [36] and Srivenkataramana [37]. The proposed estimator is given as:

$$\overline{Y}_g^{(1)^{**}} = \overline{y}_2\left[\exp\left\{\theta_1\left(\frac{\overline{Z} - \overline{z}_1^*}{\overline{Z} + \overline{z}_1^*}\right)\right\} + \theta_2\ln\left(\frac{\overline{x}_1}{\overline{x}_2}\right)\right]. \tag{2.1}$$

where θ_1, θ_2 are optimization constants and are used to minimize the mean square error while $\overline{z}^*$ introduces as transformed variable. The "ln" function has been introduced to control the variability of ratio for first and second phase variables while exponential function on ratio of transformed variables also helps to reduce the estimator's mean square errors. The properties of the propose estimator can be studied by considering

$$\overline{y} = \left(\overline{Y} + e_y\right), \overline{x} = \left(\overline{X} + e_x\right) and \ \overline{z} = \left(\overline{Z} + e_z\right)$$

Putting $\overline{y} = \left(\overline{Y} + e_y\right), \overline{x} = \left(\overline{X} + e_x\right) and \ \overline{z} = \left(\overline{Z} + e_z\right)$ and $\overline{z}_1^*$ in Eq (2.1), we get

$$\overline{Y}_g^{(1)^{**}} = \overline{y}_2\left[\exp\left\{\theta_1\left(\frac{\overline{Z} - \left(N\overline{Z} - n_1\overline{z}_{(1)}\right)\big/\left(N - n_1\right)}{\overline{Z} + \left(N\overline{Z} - n_1\overline{z}_{(1)}\right)\big/\left(N - n_1\right)}\right)\right\} + \theta_2\ln\left(\frac{\overline{x}_1}{\overline{x}_2}\right)\right]$$

$$\overline{Y}_g^{(1)^{**}} = \left(\overline{Y} + e_y\right)\left[\exp\left\{\theta_1\left(\frac{\overline{Z} - \left(N\overline{Z} - n_1\left(\overline{Z} + e_z\right)\right)\big/\left(N - n_1\right)}{\overline{Z} + \left(N\overline{Z} - n_1\left(\overline{Z} + e_z\right)\right)\big/\left(N - n_1\right)}\right)\right\} + \theta_2\ln\left(\frac{\left(\overline{X} + e_{x(1)}\right)}{\left(\overline{X} + e_{x(2)}\right)}\right)\right]$$

after simplification we have

$$\overline{Y}_g^{(1)^{**}} = \left(\overline{Y} + e_y\right)\left[1 + \frac{n_1\theta_1}{2\left(N - n_1\right)}\frac{e_z}{\overline{Z}} + \theta_2\frac{\left(e_{x(1)} - e_{x(2)}\right)}{\overline{X}}\right]$$

For obtaining the MSE of the suggested estimator $\overline{Y}_g^{(1)^{**}}$, we write

where

$$E\left(e_y\right) = E\left(e_z\right) = E\left(e_{x_1}\right) = E\left(e_{x_2}\right) = 0$$
$$E\left(e_y^2\right) = f_2\overline{Y}^2C_y^2, \ E\left(e_z^2\right) = f_1\overline{Z}^2C_z^2,$$
$$E\left(e_{x_1} - e_{x_2}\right)^2 = \left(f_2 - f_1\right)\overline{X}^2C_x^2,$$
$$E\left(e_ye_z\right) = f_1\overline{YZ}C_yC_z\rho_{yz},$$
$$E\left(e_y\left(e_{x_1} - e_{x_2}\right)\right) = \left(f_1 - f_2\right)\overline{XY}C_xC_Y\rho_{xy}$$

Expanding the right side of (2.1) in term of e's and on further simplification to the first degree of approximation, we have

$$\left(\overline{Y}_g^{(1)**}-\overline{Y}\right)\cong\left[e_{y_2}+\left(\frac{n_1}{N-n_1}\right)\overline{Y}\theta_1\left(\frac{e_{z_1}}{2\overline{Z}}\right)+\frac{\overline{Y}}{\overline{X}}\theta_2\left(e_{x_1}-e_{x_2}\right)\right] \tag{2.2}$$

taking square on both sides and apply expectation

$$E\left(\overline{Y}_g^{(1)**}-\overline{Y}\right)^2\cong E\left[e_{y_2}+\left(\frac{n_1}{N-n_1}\right)\overline{Y}\theta_1\left(\frac{e_{z_1}}{2\overline{Z}}\right)+\frac{\overline{Y}}{\overline{X}}\theta_2\left(e_{x_1}-e_{x_2}\right)\right]^2$$

where θ_1^* and θ_2^* are optimum values of θ_1 and θ_2 respectively and are determined using differential calculus, we get

$$\theta_1=\left(-2\left(\frac{N-n_1}{n_1}\right)\frac{C_y}{C_z}\rho_{yz}\right)=\theta_1^*(say),\ \theta_2=\left(\frac{C_y}{C_x}\rho_{xy}\right)=\theta_2^*(say)\Bigg\} \tag{2.3}$$

using optimum values of θ_1^* and θ_2^*, the minimum MSE of proposed estimator is obtained as:

$$\min MSE\left(\hat{\overline{Y}}_g^{(1)**}\right)\cong\overline{Y}^2C_y^2\left[f_2\left(1-\rho_{xy}^2\right)+f_1\left(\rho_{xy}^2-\rho_{yz}^2\right)\right]. \tag{2.4}$$

2.1. Comparing mathematically $\overline{Y}_g^{(1)**}$ with some existed estimators

In this section, we compared our proposed estimator $\overline{Y}_g^{(1)**}$ with some of existed estimators. The comparison has been made in terms of the mean square errors and we have obtained some conditions under which our proposed estimator has smaller minimum MSE as compared to some already existing estimators. These comparisons are:

Case I

$$MSE\left(\hat{\overline{Y}}_{sv}\right)\geq MSE\left(\hat{\overline{Y}}_g^{(1)**}\right)$$

$\hat{\overline{Y}}_g^{(1)**}$ is more efficent than $\hat{\overline{Y}}_{sv}$

iff

$$f_1\left(\frac{K^2}{4}+\rho_{xy}\left(1-K\right)\right)-f_2\left(\frac{K^2}{4}+\rho_{xy}\left(1-K\right)+\rho_{yz}^2\right)\geq 0 \tag{2.5}$$

where $K=\dfrac{C_x}{4C_y}$

Case II

$$MSE\left(\hat{\overline{Y}}_s\right)\geq MSE\left(\hat{\overline{Y}}_g^{(1)**}\right)\Rightarrow f_1\left(\rho_{yz}\right)^2\geq 0$$

Case III

$$MSE\left(\hat{\bar{Y}}_{NH}\right) \ge MSE\left(\hat{\bar{Y}}_g^{(1)^{**}}\right)$$

iff

$$\left[f_1\left(\frac{C_x^2}{4} + C_x\left(C_y\rho_{xy} - \frac{1}{2}C_z\rho_{xz}\right) - C_y^2\left(\rho_{xy}^2 - \rho_{yz}^2\right)\right) + \\ f_2\left(\frac{C_z^2}{4} - C_y\left(C_z\rho_{yz} - C_y\rho_{xy}^2\right)\right) \ge 0 \right] \tag{2.6}$$

Case IV

$$MSE\left(\hat{\bar{Y}}_{SA}\right) \ge MSE\left(\hat{\bar{Y}}_g^{(1)^{**}}\right)$$

$$\left[f_1(1-\hat{\alpha})\left(C_x\left(\frac{(1-\hat{\alpha})}{4}C_x + C_y\rho_{xy} - \hat{\alpha}C_z\rho_{xz}\right) - C_y^2\left(\rho_{xy}^2 - \rho_{yz}^2\right)\right) + \\ f_2\left(\hat{\alpha}C_z^2 - 2C_yC_z\rho_{yz} + C_y^2\rho_{xy}^2\right) \right] \ge 0 \tag{2.7}$$

If the above given expressions are not met then estimators $\bar{Y}, \bar{Y}_S, \bar{Y}_{SV}, \bar{Y}_{NH}$ and $\bar{Y}_{SA}$ are more efficient than the suggested estimator $\bar{Y}_g^{(1)^{**}}$.

3. Results

In this section some real populations available in literature have been selected for empirical study to obtain mean square error and relative efficiency of our proposed estimator $\hat{\bar{Y}}_g^{(1)^{**}}$. For checking the performance of suggested estimator the following real data sets have been used.

- Sugar Cane Disease "coal of sugar-cane" (This is a disease that is common in sugar-cane plantations in certain areas of Brazil)
- Soil Compositions of Physical and Chemical Characteristics
- Appliances energy prediction Data Set
- Ozone (The data is monthly ozone averages on a very coarse 24 by 24 grid covering Central America, from Jan 1995 to Dec 2000)
- Combined Cycle Power Plant Data Set (The voltage output of engines was measured at various combinations of blade speed and sensor extension)

The comparison of our suggested estimator has been made with the conventional unbiased estimators $\bar{Y}, \bar{Y}_S, \bar{Y}_{SV}, \bar{Y}_{NH}$ and $\bar{Y}_{SA}$. The depiction of variables for each population is given below:

Table 1. Description of variables.

Pop.	Y	X	Z	Populations Sources
1	The number of pieces of the stems, out of 50, planted in each plot.	The number of diseased shoots.	The total number of shoots in each plot.	The data were kindly supplied by Dr. C.G.B. Demetrio of Escola Superior de Agricultura, Universidade de Sao Paolo, Brazil
2	Conductivity	Magnesium in me/100gm	Bulk density in gm	Khattree, R., and Naik, D. N. (2000) *Multivariate Data Reduction and Discrimination with SAS Software.* SAS Institute.
3	Random variable non-dimensional	Pressure (from Chievres weather station), in mm Hg	Humidity outside (from Chievres weather station)	University of Mons (UMONS).
4	Ozone	Radiation	Temperature	Bruntz, S. M., W. S. Cleveland, B. Kleiner, and J. L. Warner. (1974). The Dependence of Ambient Ozone on Solar Radiation, Wind, Temperature, and Mixing Height. In *Symposium on Atmospheric Diffusion and Air Pollution*, pages 125–128. American Meterological Society, Boston.
5	Net hourly electrical energy output	Ambient Pressure (AP) in the range 992.89-1033.30 milibar	Relative Humidity (RH) in the range 25.56% to 100.16%	Faculty of Engineering, Namık Kemal University, Turkey

The following Table 2 depicts the values of means, CV and correlations coefficient, which are essential to estimate the mean square error.

Table 2. Means, CV and coefficient of correlations for populations**.**

Pop.	N	n_1	n_2	$\overline{X}$	$\overline{Y}$	$\overline{Z}$	C_y	C_x	C_z	ρ_{xy}	ρ_{yz}	ρ_{xz}
1	180	79	33	11.94	20.26	118.14	1.21	0.54	0.37	0.08	0.34	0.40
2	48	21	15	8.4646	6.5885	1.3158	0.5989	0.1599	0.1653	0.5083	0.7626	0.4901
3	19735	9000	5000	79.7	24.9	42.9	0.580	0.186	0.121	0.020	0.004	0.4873
4	111	37	17	184.80	42.099	77.792	0.7868	0.4910	0.1219	0.3483	0.6985	0.2940
5	9571	1000	700	1013.2	454.3	73.3	0.037	0.005	0.199	0.514	0.389	0.099

The percent relative efficiency formula for calculating efficiencies is given as,

$$PRE\left(.,\overline{y}\right) = \frac{MSE\left(\overline{y}\right)}{MSE\left(.\right)} \times 100 \quad where\left(.\right) = \overline{Y}, \overline{Y}_S, \overline{Y}_{SV}, \overline{Y}_{NH}, \overline{Y}_{SA} \ and \ \overline{Y}_g^{(1)^{**}}$$

Table 3 shows relative efficiencies of already developed and our proposed estimator.

Table 3. Relative efficiencies.

S.#	Estimators	Pop1	Pop2	Pop3	Pop4	Pop5
1	$\hat{\bar{Y}}$	100.000	100.000	100.000	100.000	100.000
2	$\hat{\bar{Y}}_{SV}$	*	100.6738	*	*	169.3055
3	$\hat{\bar{Y}}_{S}$	100.4436	112.0270	100.025	108.3955	184.5887
4	$\hat{\bar{Y}}_{NH}$	107.1621	122.0960	*	107.1865	*
5	$\hat{\bar{Y}}_{SA}$	113.6877	147.9260	102.0920	143.2637	150.0283
6	$\hat{Y}_{kc}$	465.0547	301.3585	247.1565	322.2064	238.6583
7	$\hat{\bar{Y}}_{g}^{(1)**}$	502.5515	355.1199	247.3330	390.9441	279.0334

* Percent Relative Efficiency is less than 100.

It is observed in Table 3 that most of the proposed estimators give more efficient results than the classical ratio estimator. The relative efficacies table clearly indicated that our proposed estimator $\hat{\bar{Y}}_{g}^{(1)**}$ is more efficient than $\bar{Y}, \bar{Y}_{S}, \bar{Y}_{SV}, \bar{Y}_{NH}$ and $\bar{Y}_{SA}$.

3.1. Simulation results

The simulation study has been steered by generating random populations from a bivariate normal distribution. For this simulation study, a random population of size 50000 was generated, for auxiliary variables X and Z, from standard bivariate normal distribution. Using these auxiliary variables, the study variable Y was generated by using $Y_i = X_i + Z_i + e_i$, where e_i is $N(0,1)$. From this population two phase samples were generated by using three different first phase and second phase samples as for simulation I 20% of 50000 and then 50% of 10000 $(n_1 = 10000, n_2 = 5000)$, for simulation II 10% of 50000 and then 40% of 5000 $(n_1 = 5000, n_2 = 2000)$ and for simulation III 25% of 50000 and then 30% of 12500 $(n_1 = 12500, n_2 = 3750)$ respectively. For each of the sample various estimators were computed. The procedure was repeated for 50000 times and using 50000 values of each estimator, the MSE of each estimator was calculated and results are given in Table 4 below. From the results given in Table 4, we can see that the simulated mean square error of our proposed estimator is approximately half of the mean square of other estimators used in the study for $(n_1 = 5000, n_2 = 2000)$ and $(n_1 = 12500, n_2 = 3750)$. The simulated mean square error of our proposed estimator is approximately 40% of the mean square error of other estimators for $(n_1 = 10000, n_2 = 5000)$. From this simulation study, we can say that our proposed estimator, $\bar{Y}_{g}^{(1)}$, will have smaller mean square error while estimation of population mean as compared with the other available estimators of population mean.

Table 4. Simulation populations.

Estimators	Simulation I $n_1 = 10000$ $n_2 = 5000$	Simulation II $n_1 = 5000$ $n_2 = 2000$	Simulation III $n_1 = 12500$ $n_2 = 3750$
$MSE\left(\overline{Y}_g^{(1)**}\right)$	0.007342740	0.01482553	0.008580045
$MSE\left(\overline{Y}_{sv}\right)$	0.007763214	0.01609473	0.008786188
$MSE\left(\overline{Y}_{s}\right)$	0.014134104	0.03650642	0.019750622
$MSE\left(\overline{Y}_{NH}\right)$	0.020980778	0.06862683	0.039805696
$MSE\left(\overline{Y}_{SA}\right)$	0.021681391	0.07091925	0.041749114

In Table 4, the mean square error of our proposed estimator and already developed estimators have been compared. It is observed from the Table 4 that the MSE of our proposed estimator is minimum in all simulated populations, which proves it to be an efficient estimator since the estimator with least mean square error is considered to be most effective.

4. Discussion

In this study, an exponential ratio type estimator utilizing two auxiliary variables under two-phase sampling is suggested. A theoretical comparison of derived estimator is carried out by establishing the conditions under which proposed estimator, $\hat{\overline{Y}}_g^{(1)**}$, will be more efficient than the exponential ratio type estimator of Singh and Vishwakarma [29] $\left(\hat{\overline{Y}}_{SV}\right)$, Singh, et al. [32] $\left(\hat{\overline{Y}}_{S}\right)$, exponential difference of ratio product estimator of Noor-ul-Amin and Hanif [20] $\left(\hat{\overline{Y}}_{NH}\right)$, Sanaullah, et al. [33] $\left(\hat{\overline{Y}}_{SA}\right)$ and classical regression estimator with two auxiliary variables proposed by Kadilar and Cingi [34] $\left(\hat{Y}_{kc}\right)$ in two phase sampling. We considered real population information as well as simulated data to inspect the performance of proposed estimator under two-phase sampling scheme. In Table 3, the proposed estimator is compared with some of the existing estimators on the basis of their relative efficiencies while Table 4 indicates that the performance of the proposed exponential estimator is better than the Singh and Vishwakarma [29], Singh, et al. [32], Noor-ul-Amin and Hanif [20] and Sanaullah, et al. [33] for all the simulated populations based on their mean square errors. The estimator with higher relative efficiency and least mean square error is considered to be more efficient. It can be clearly observed from Table 3 and Table 4 that the relative efficiency of our proposed estimator is higher for all real populations and its mean square error is minimum in all simulation studies. This proves our proposed estimator to be an efficient estimator.

5. Conclusion

An exponential type ratio estimator in two-phase sampling has been proposed, which obtains mean square error by involving two auxiliary variables. The main purpose of this study is to compare the efficiency of our proposed estimator with some of the existing exponential estimators. MSE equation of our proposed estimator has been derived theoretically and its efficiency has been checked through different simulations and datasets of different populations from different fields. The percent relative efficacies values obtained from table-3 are **(502.5515), (355.1199), (247.3183), (390.9441) and (279.0334),** which are highest for all the populations proving our proposed estimator to be most efficient estimator in this study. Sanaullah, et al. [33] "Ratio Estimator" appears as the second most efficient estimator in term of relative efficiencies for all populations except population 5. Subsequently, looking on the predominance nature of our proposed estimator, we suggest its use for its practical applications especially in the field of environment, engineering and biological sciences.

Acknowledgement

Conflicts of interest

The authors have no conflict of interest.

References

1. J. Velasco-Muñoz, J. Aznar-Sánchez, L. Belmonte-Ureña, et al., *Sustainable water use in agriculture: A review of worldwide research,* Sustainability, **10** (2018), 1084.
2. R. Serbu, B. Marza, S. Borza, *A spatial Analytic Hierarchy Process for identification of water pollution with GIS software in an eco-economy environment,* Sustainability, **8** (2016), 1208.
3. R. Ziemer and W. Tranter, *A survey and analysis of electrical engineering curricula in communications and signal processing.* IEEE Commun. Mag., **18** (1980), 5–12.
4. R. M. De Mello, P. C. Da Silva, G. H. Travassos, *Investigating probabilistic sampling approaches for large-scale surveys in software engineering,* Journal of Software Engineering Research and Development, **3** (2015), 8.
5. R. M. De Mello, G. H. Travassos, *Surveys in software engineering: Identifying representative samples*, Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, 2016.
6. D. J. Watson, *The estimation of leaf area in field crops,* J. Agr. Sci, **27** (1937), 474–483.
7. W. G. Cochran, *The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce,* J. Agr. Sci., **30** (1940), 262–275.
8. M. N. Murthy, *Product method of estimation,* Sankhya, **26** (1964), 294–307.
9. D. S. Robson, *Applications of multivariate polykays to the theory of unbiased ratio-type estimation,* J. Am. Stat. Assoc., **52** (1957), 511–522.

10. S. K. Srivastava, *A generalized estimator for the mean of a finite population using multi-auxiliary information,* J. Am. Stat. Assoc., **66** (1971), 404–407.

11. S. Bahl, R. K. Tuteja, *Ratio and product type exponential estimators,* Journal of information and optimization sciences, **12** (1991), 159–164.

12. M. A. Hidiroglou, *Double sampling,* Survey methodology, **27** (2001), 143–154.

13. M. Samiuddin, M. Hanif, *Estimation of population mean in single and two phase sampling with or without additional information,* Pakistan journal of statistics-all series-, **23** (2007), 99.

14. H. P. Singh, M. R. Espejo, *Double sampling ratio-product estimator of a finite population mean in sample surveys,* J. Appl. Stat., **34** (2007), 71–85.

15. H. P. Singh, G. K. Vishwakarma, *Modified exponential ratio and product estimators for finite population mean in double sampling,* Austrian journal of statistics, **36** (2007), 217–225.

16. M. Hanif, N. Hamad, M. Q. Shahbaz, *A modified regression type estimator in survey sampling,* World Applied Sciences Journal, **7** (2009), 1559–1561.

17. M. Hanif, N. Hamad, M. Q. Shahbaz, *Some new regression type estimators in two phase sampling,* World Applied Sciences Journal, **8** (2010), 799–803.

18. H. P. Singh, R. Tailor, R. Tailor, *On ratio and product methods with certain known population parameters of auxiliary variable in sample surveys,* SORT-Stat Oper. Res. T., **34** (2010), 157–180.

19. R. Singh, F. Smarandache, *Studies in sampling techniques and time series analysis,* Zip publishing, 2011.

20. M. Noor-ul-Amin, M. Hanif, *SOME EXPONENTIAL ESTIMATORS IN SURVEY SAMPLING,* Pakistan Journal of Statistics, **28** (2012).

21. R. Tailor, R. Tailor, R. Parmar, et al., *Dual to Ratio-cum-Product estimator using known parameters of auxiliary variables,* Journal of Reliability and Statistical Studies, **5** (2012), 65–71.

22. J. Shabbir, S. Gupta, Z. Hussain, *Improved estimation of finite population median under two-phase sampling when using two auxiliary variables,* Scientia Iranica, **22** (2015), 1271–1277.

23. A. H. Al-Marshadi, A. H. Alharby, M. Q. Shahbaz, *On some new estimators of population variance in single and two-phase sampling,* Maejo Int. J. Sci. Tech., **12** (2018), 272–281.

24. G. K. Vishwakarma, M. Kumar, *An improved class of chain ratio-product type estimators in two-phase sampling using two auxiliary variables,* Journal of probability and Statistics, 2014.

25. G. K. Vishwakarma, M. Kumar, *An efficient class of estimators for the mean of a finite population in two-phase sampling using multi-auxiliary variates,* Commun. Math. Stat., **3** (2015), 477–489.

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

27. P. Mishra, N. K. Adichwal, R. Singh, *A New Log-Product-Type Estimator Using Auxiliary Information,* Journal of Scientific Research, BHU, Varanasi, **61** (2017), 179–183.

28. T. Akhlaq, M. Ismail, M. Q. Shahbaz, *On Efficient Estimation of Process Variability,* Symmetry, **11** (2019), 554.

29. H. P. Singh, G. K. Vishwakarma, *Modified exponential ratio and product estimators for finite population mean in double sampling,* Austrian journal of statistics, **36** (2007), 217–225.

30. W. G. Cochran, *Sampling techniques,* 3rd Edition, 1977.

31. L. Upadhyaya, G. Singh, *Chain-type estimators using transformed auxiliary variable in two-phase sampling,* ADVANCES IN MODELLING AND ANALYSIS-A-, **38** (2001), 1a–10a.

32. R. Singh, M. Kumar, F. Smarandache, *ALMOST UNBIASED ESTIMATOR FOR ESTIMATING POPULATION MEAN USING KNOWN VALUE OF SOME POPULATION PARAMETER(S)*, Pakistan Journal of Statistics and Operation Research, **4** (2008), 63–76.

33. A. Sanaullah, H. Khan, H. A. Ali, et al., *Improved exponential ratio-type estimators in survey sampling,* Journal of Reliability and Statistical Studies, **5** (2012), 119–132.

34. C. Kadilar, H. Cingi, *A new estimator using two auxiliary variables,* Appl. Math. Comput., **162** (2005), 901–908.

35. H. O. Cekim, C. Kadilar, *In-type estimators for the population variance in stratified random sampling,* Communications in Statistics-Simulation and Computation, **49** (2020), 1665–1677.

36. S. Bandyopadhyay, *Improved ratio and product estimators,* Sankhya, **42** (1980), 45–49.

37. T. Srivenkataramana, *A dual to ratio estimator in sample surveys,* Biometrika, **67** (1980), 199–204.

Appendix

Nomenclature section

Symbols	Description
$\overline{y}, \overline{x}, \overline{z}$	Sample means corresponding to the population means $\overline{Y}, \overline{X}$ and $\overline{Z}$ respectively
N	Complete population
n_1, n_2	Samples sizes Phase one and phase two respectively
$\beta_{2_{(z)}}$	Kurtosis
C_x, C_y, C_z	Coefficient of variation of x, y and z
σ	Standard deviation
$\rho_{xy}, \rho_{yz}, \rho_{xz}$	Correlation coefficients between of $(Y, X),\ (Y, Z)$ and (X, Z) respectively

Simulation Code:

```
tyasir1<-c()
sv01<-c()
nh12<-c()
su12<-c()
us01_2<-c()
singh1_1<-c()
kadilar1_1<-c()

tyasir11<-c()
sv02<-c()
nh121<-c()
su121<-c()
us01_22<-c()
singh1_11<-c()
kadilar1_11<-c()

tyasir111<-c()
sv03<-c()
nh1211<-c()
su1211<-c()
us01_222<-c()
singh1_111<-c()
kadilar1_111<-c()

for(m in 1:50000)                    {#loop1
```

```
x<-(rweibull(50000,0.4,2))
z<-(rweibull(50000,0.5,1))
ei<-(rnorm(50000,0,1))
Y<-x+z+ei
n1<-0.2*50000
n2<-0.5*n1
m1<-0.1*50000
m2<-0.4*m1
p1<-0.25*50000
p2<-0.3*p1
i<-sample(c(1:50000),n1)
i<-sort(i)
x1<-x[i]
z1<-z[i]
j<-sample(i,n2)
x2<-x[j]
z2<-z[j]
y<-Y[j]
k<-sample(c(1:50000),m1)
k<-sort(k)
x11<-x[k]
z11<-z[k]
l<-sample(k,m2)
x22<-x[l]
z22<-z[l]
y1<-Y[l]

t<-sample(c(1:50000),p1)
t<-sort(t)
x111<-x[t]
z111<-z[t]
v<-sample(t,p2)
x222<-x[v]
z222<-z[v]
y11<-Y[v]

xbar<-mean(x)
zbar<-mean(z)
x1bar<-mean(x1)
x2bar<-mean(x2)
z1bar<-mean(z1)
z2bar<-mean(z2)
x11bar<-mean(x11)
x22bar<-mean(x22)
```

```
z11bar<-mean(z11)
z22bar<-mean(z22)

x111bar<-mean(x111)
x222bar<-mean(x222)
z111bar<-mean(z111)
z222bar<-mean(z222)

z1st<-(m*z1bar-n1*z2bar)/(m-n1)
z11st<-(m*z11bar-m1*z22bar)/(m-m1)
z111st<-(m*z111bar-p1*z222bar)/(m-p1)

ybar<-mean(y)
y1bar<-mean(y1)
y11bar<-mean(y11)
Cy<-(sd(Y)/mean(Y))
Cx<-(sd(x)/mean(x))
Cz<-(sd(z)/mean(z))
Rxy<-cor(x,Y)
Rzy<-cor(z,Y)
Rxz<-cor(x,z)

m1a<-Cy*Cz*Rzy
m2a<-((Cx)^2)/2
m3a<-Cx*Cy*Rxy
m4a<-Cx*Cz*Rxz
m5a<-(Cz)^2

A1<-((1/n1)-(1/m))
A2<-((1/n2)-(1/m))

B1<-((1/m1)-(1/m))
B2<-((1/m2)-(1/m))

C1<-((1/p1)-(1/m))
C2<-((1/p2)-(1/m))

alpha1<-((2*A2*m1a)+(A1*(m2a-m3a-m4a))/(2*A2*m5a)+A1*(m2a-2*m4a))
alpha2<-((2*B2*m1a)+(B1*(m2a-m3a-m4a))/(2*B2*m5a)+B1*(m2a-2*m4a))
alpha3<-((2*C2*m1a)+(C1*(m2a-m3a-m4a))/(2*C2*m5a)+C1*(m2a-2*m4a))

theta1<-2*(1-m/n1)*(Cy/Cz)*Rzy
theta2<-(Cy/Cx)*Rxy
```

```
theta11<-2*(1-m/m1)*(Cy/Cz)*Rzy
theta22<-(Cy/Cx)*Rxy

theta111<-2*(1-m/p1)*(Cy/Cz)*Rzy
theta222<-(Cy/Cx)*Rxy

sv1<-ybar*exp((x1bar-x2bar)/(x1bar+x2bar))# sigh and Vishwakarma (2007)
nh1<-ybar*exp(((zbar-z2bar)/(zbar+z2bar))-((x1bar-x2bar)/(x1bar+x2bar))) #Noor-ul-Amin & Hanif,
 2012
su1<-ybar*exp((alpha1*(zbar-z2bar)/(zbar+z2bar))-((1-alpha1)*(x1bar-x2bar)/(x1bar+x2bar)))
 #Sanaullah, Khan, Ali, and Singh (2012)
si1<-ybar*(x1bar/x2bar)*(zbar+sd(z))/(z1bar+sd(z))      # singh 2001 1

sv11<-y1bar*exp((x11bar-x22bar)/(x11bar+x22bar))# sigh and Vishwakarma (2007)
nh11<-y1bar*exp(((zbar-z22bar)/(zbar+z22bar))-((x11bar-x22bar)/(x11bar+x22bar)))
 #Noor-ul-Amin & Hanif, 2012
su11<-y1bar*exp((alpha2*(zbar-z22bar)/(zbar+z22bar))-((1-alpha2)*(x11bar-x22bar)/(x11bar+x22b
 ar))) #Sanaullah, Khan, Ali, and Singh (2012)
si11<-y1bar*(x11bar/x22bar)*(zbar+sd(z))/(z11bar+sd(z))      # singh 2001 1

sv111<-y11bar*exp((x111bar-x222bar)/(x111bar+x222bar))# sigh and Vishwakarma (2007)
nh111<-y11bar*exp(((zbar-z222bar)/(zbar+z222bar))-((x111bar-x222bar)/(x111bar+x222bar)))
 #Noor-ul-Amin & Hanif, 2012
su111<-y11bar*exp((alpha3*(zbar-z222bar)/(zbar+z222bar))-((1-alpha3)*(x111bar-x222bar)/(x111b
 ar+x222bar))) #Sanaullah and Singh (2012)
si111<-y11bar*(x111bar/x222bar)*(zbar+sd(z))/(z111bar+sd(z))      # singh 2001 1

tyasir<-ybar*(exp(theta1*((zbar-z1st)/(zbar+z1st)))+theta2*log(x1bar/x2bar))# my estimator
tyasira<-y1bar*(exp(theta11*((zbar-z11st)/(zbar+z11st)))+theta22*log(x11bar/x22bar))#          my
estimator
tyasirb<-y11bar*(exp(theta111*((zbar-z111st)/(zbar+z111st)))+theta222*log(x111bar/x222bar))#
my estimator
tyasir1<-c(tyasir1,tyasir)
tyasir11<-c(tyasir11,tyasira)
tyasir111<-c(tyasir111,tyasirb)

sv01<-c(sv1,sv01)
nh12<-c(nh1,nh12)
su12<-c(su1,su12)
singh1_1<-c(si1,singh1_1)

sv02<-c(sv11,sv02)
nh121<-c(nh11,nh121)
su121<-c(su11,su121)
```

```
singh1_11<-c(si11,singh1_11)

sv03<-c(sv111,sv03)
nh1211<-c(nh111,nh1211)
su1211<-c(su111,su1211)
singh1_111<-c(si111,singh1_111)
                                    }#loop1
simulation1<-c(var(sv01),var(singh1_1),var(nh12),var(su12),var(tyasir1))
simulation2<-c(var(sv02),var(singh1_11),var(nh121),var(su121),var(tyasir11))
simulation3<-c(var(sv03),var(singh1_111),var(nh1211),var(su1211),var(tyasir111))
simulation1
simulation2
simulation3
```