

Use of auxiliary information in estimation of the finite population mean: An exponential type of estimator

^aKalim Ullah, ^bHudson, I. L., ^cCheema, S. A. ^aKhan, A. ^aAbdu-R-Rahman and ^dHussian, Z.

^aDepartment of Statistics, Quaid-i-Azam University Islamabad, Pakistan ^bMathematical Sciences, College of STEM, School of Science, Royal Melbourne Institute of Technology (RMIT), Melbourne, Australia.

^cDepartment of Applied Sciences, National Textile University Faisalabad, Pakistan.

^dDepartment of Social and Allied Sciences, Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Pakistan.

Email: irene.hudson@rmit.edu.au

Abstract: The use of auxiliary information has a long history in statistical theory and estimation procedures. The utility of supplementary knowledge becomes vital when information about the study variable is limited. In this paper, we present a more competent mechanism to utilise auxiliary information in the estimation of the finite population mean. We propose a new exponential type of estimator for the estimation of finite population mean in the scenario where a simple random sampling scheme is adopted. Our proposed procedure is based on the dual use of the supportive information to maximise additional gain and involves the use of the mean of the auxiliary variable along with its rank to increase the extent of relevant information. The comparative performance of the proposed scheme is demonstrated with respect to 10 most used, classic, and some recent procedures in estimation theory literature. These are the classic mean estimator \hat{Y}_{SRS} , the so-called traditional ratio, product, and regression estimators \hat{Y}_R , \hat{Y}_P and \hat{Y}_{reg} , respectively, along with the difference type estimator $\hat{Y}_{R.D.}$. In addition, the more recent estimators investigated are the ratio-product exponential type $\hat{Y}_{S.RP}$, difference exponential type \hat{Y}_{GK} , ratio exponential $\hat{Y}_{BT,R}$, product exponential $\hat{Y}_{BT,P}$ and the ratio-product-exponential \hat{Y}_{SHG} , all used for comparison. Moreover, we consider three data sets of a multi-disciplinary nature, encompassing health surveillance, industrial production and poultry. The choice of data sets is mainly motivated by two reasons; (i) these data sets have been topics of contemporary techniques and, (ii) the considered data sets do offer a wide range of parametric settings, including lower extent of correlation between the study variable with the auxiliary variable and they also vary in sample sizes.

In addition, we consider cases of a higher positive and higher negative degree of linear relationship extant between the study variable and auxiliary variable in these data sets. Along with the opportunity of conducting a fair comparison of our suggested strategy with contemporary techniques, the above approach allows for us to observe various patterns prevalent in the resultant gains of our newly devised scheme. Improvements are quantified by the mean square errors of the competing estimators, which are further transformed into relative percentage efficiencies to attain a comprehensive view of the research effort. Overall, we observe a noticeable amount of decrease in mean square error for our proposed estimator as compared to existing estimators, evident for all the considered data sets. However, there are a few observant patterns in the efficiency gains coinciding with assigned pre-defined parametric settings, in that the extent of the correlation between the auxiliary variable and output variable plays a pivotal role in the performance of estimation procedures. The improvement in the efficiency becomes more obvious as the degree of linear relationship between the output variable with the auxiliary variable strengthens. For example, minimum gain in percentage relative efficiencies (PREs) is observed for the 1st data set, wherein the correlation coefficient, $\rho_{out,aux}$, remains minimal. For the two other data sets the gain remains clearer as the correlation coefficient takes higher values, say, $|\rho_{out,aux}| > 0.85$. We also note the varying performance hierarchy among contemporary estimators with respect to varying features of each population. Our proposed estimator outperforms the existing methods studied here in all cases. The mathematical expressions for the bias and mean squared error of the proposed estimator is derived under the first order of approximation. The theoretical and empirical studies show that the proposed estimator performs uniformly better than the existing estimators in terms of the percentage relative efficiency. We advocate that in future exponential smoothing will be used to quantify changes given updates by auxiliary information and recent observations.

Keywords: Auxiliary information, exponential type estimator, mean estimation, ranks, simple random sampling

1. INTRODUCTION

1.1 Motivation

The use of auxiliary information to enhance the performance of estimation procedures is well documented in the statistical and allied sciences literature, see for example, Zhang and Chambers (2004), Chou et al. (2017) and Bai et al. (2021). In practice auxiliary information is obtainable from various sources, such as census, survey reports and expert opinion, see also Rao et al. (1990) and Biemer and Peytchev (2013). It is customary to incorporate supportive information as correlates of study variable(s) in estimation procedures, see Kreuter et al. (2010). Many researchers have understood the utility of supplementary information in multi-disciplinary investigations. For example, early on Haberman (1984), documented the efficacy of such additional information to improve the projections of probability measures. Further, Holt and Elliot (1991) advocated the use of auxiliary information as a damage control tool in survey-based research to minimize the distortion caused by non-responses. Olson (2013) demonstrated the value of supportive information in the estimation of distribution functions that can suffer from lost non-ignorable missing data. Recently, Haq et al. (2020) argued for the value-add of auxiliary information in the early detection of process shift in industrial engineering. Indeed, information retrieved via auxiliary variables has resulted in a wide-ranging gain in performance over estimators which do not employ such information. When the auxiliary variable X is available in advance or observed readily, and a high correlation exists with the study variable Y , use of auxiliary information is shown to be effective for estimating the population mean. In these situations, ratio, regression, and product estimators are suggested; note that estimation via regression is traditional and well known. Cochran (1940) proposed the ratio method in the case when there is strong positive correlation between Y and X . Later Murthy (1964) re-evaluated the product method of estimation suggesting it appropriate if a strong negative correlation exists between auxiliary and study variables. Srivastava (1971) recommended general ratio estimator using a single auxiliary variable. Bhal and Tuteja (1991) proposed an exponential product estimator for the scenario when study and auxiliary variable are negatively correlated and proposed a simple exponential ratio estimator, in the case when a positive correlation exists between Y and X . There are some similarities between the procedures in this paper and Hussain et al. 2020, however the target Hussain et al., (2020) is estimation of population distribution function. This paper primarily focuses on the development of a new estimation procedure capable of more competent incorporation of supplementary information in estimating the finite population mean. We propose an exponential type of estimator of the mean when random sampling is conducted. Our approach involves the dual use of an auxiliary variable by accommodating not only the mean of the variable but ranks of the auxiliary variable. The applicability of the proposed scheme is explored by studying three leading data sets from the existing literature on estimation procedures. The comparative performance of the newly suggested estimator is also evaluated with respect to ten most used existing measures. We consider various parametric settings for this purpose. The gains of our proposed use of auxiliary information are documented. In subsection 1.2 and 1.3, we present the mathematical preliminaries frequently used in this research along with a brief introduction of contemporary methods to assess their comparative competency, respectively. This paper is divided into three major sections. In section 2, we document the mathematical foundations and developments of our proposed estimator, wherein the expressions for the bias and mean squared error of the proposed estimator are derived under the first order of approximation. Section 3 reports the main findings of the comparative investigations. Section 4 summarises major outcomes of the research and future research.

1.2 Preliminaries

Let $Z = \{Z_1, Z_2, \dots, Z_N\}$ be a finite population of size N , where n represents the size of a sample drawn from the population using simple random sampling without replacement. Let (Y, F) be the measurable space with P the set of all probability measures on (Y, F) . Then Y_i represents study variable for i 'th population unit, where $i = 1, 2, \dots, N$ indicating identically and independently distributed vector of random variables. Also, X_i denotes auxiliary variable from the i 'th data source along with R_i the ranks of the auxiliary variable. The key assumption on the auxiliary variable is the existence of a shared parameter with the study variable. Then for we denote $\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$ as the population mean of the study variable along with $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ as sample mean. Similarly, $\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$ is the population mean of the auxiliary variable, whereas $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ represents the sample mean. Moreover, $\bar{R}_x = \frac{\sum_{i=1}^N R_{x_i}}{N}$ denotes the average rank of the auxiliary with $\bar{r}_x = \frac{\sum_{i=1}^n r_{x_i}}{n}$ as its sample counterpart. The population variances of study variable, auxiliary variable and related ranks are given as, $S_y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$, $S_x^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$ and $S_{r_x}^2 = \frac{\sum_{i=1}^N (R_{x_i} - \bar{R}_x)^2}{N-1}$, respectively, with the sample variants such as $s_y^2 =$

$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$, $s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ and $s_{x_r}^2 = \frac{\sum_{i=1}^n (r_{xi} - \bar{r}_x)^2}{n-1}$. Error terms are defined as, $\epsilon_0 = \frac{(\bar{y} - \bar{Y})}{\bar{Y}}$, $\epsilon_1 = \frac{(\bar{x} - \bar{X})}{\bar{X}}$ and $\epsilon_2 = \frac{(\bar{r}_x - \bar{R}_x)}{\bar{R}_x}$, where we write $E(\epsilon_i) = 0$ for $i = 0, 1$. Furthermore, $E(\epsilon_0^2) = \lambda C_y^2$, $E(\epsilon_1^2) = \lambda C_x^2$, $E(\epsilon_2^2) = \lambda C_{r_x}^2$, $E(\epsilon_0 \epsilon_1) = \lambda C_y C_x \rho_{yx}$, $E(\epsilon_0 \epsilon_2) = \lambda C_y C_{r_x} \rho_{yr_x}$ and $E(\epsilon_1 \epsilon_2) = \lambda C_x C_{r_x} \rho_{xr_x}$, where $\lambda = \left(\frac{1}{n} - \frac{1}{N}\right)$ is known as the finite population correction. Also, C_y , C_x and C_{r_x} represent coefficients of variation such that $C_y = \frac{s_y}{\bar{y}}$, $C_x = \frac{s_x}{\bar{x}}$ and $C_{r_x} = \frac{s_{r_x}}{\bar{r}_x}$, respectively. The shared parameter between study variable and auxiliary variable is launched through correlation coefficients, such that $\rho_{yx} = \frac{s_{yx}}{s_y s_x}$, $\rho_{yr_x} = \frac{s_{yr_x}}{s_y s_{r_x}}$ and $\rho_{xr_x} = \frac{s_{xr_x}}{s_x s_{r_x}}$.

1.3 Existing Methods

This sub-section provides a brief introduction of the most used to date existing methods of estimation of the finite population mean with respect to which we compare the performance of our newly proposed scheme.

(i)- Conventional estimator \hat{Y}_{srs} :

The mean square error (MSE) of conventional unbiased mean per unit estimator is given as;

$$MSE(\hat{Y}_{srs}) = \lambda \bar{Y}^2 C_y^2, \quad (1)$$

where, λ and C_y are documented above.

(ii)- Ratio estimator \hat{Y}_R :

Cochran (1940) proposed the classic ratio estimator as;

$$\hat{Y}_R = \bar{y} \left(\frac{\bar{X}}{\bar{x}} \right), \quad (2)$$

where the bias is given as, $Bias(\hat{Y}_R) = \lambda \bar{Y} (C_x^2 - C_y C_x \rho_{yx})$. The resultant MSE can be written as;

$$MSE(\hat{Y}_R) = \lambda \bar{Y}^2 (C_y^2 + C_x^2 - C_y C_x \rho_{yx}). \quad (3)$$

The boundary conditions of the ratio estimator are given as $\rho_{yx} > \frac{C_x}{2C_y}$.

(iii)- Product estimator \hat{Y}_P : Murthy (1964) proposed the product estimator as;

$$\hat{Y}_P = \bar{y} \left(\frac{\bar{x}}{\bar{X}} \right), \quad (4)$$

with the associated bias of $Bias(\hat{Y}_P) = \lambda \bar{Y} (C_x^2 - C_y C_x \rho_{yx})$. The MSE of the estimator is derived as,

$$MSE(\hat{Y}_P) = \lambda \bar{Y}^2 (C_y^2 + C_x^2 - C_y C_x \rho_{yx}). \quad (5)$$

The boundary conditions of the ratio estimator are given as $\rho_{yx} < -\frac{C_x}{2C_y}$.

(iv)- Regression estimator \hat{Y}_{reg} : Cochran (1977) provided the famous asymptotically unbiased regression estimator such as;

$$\hat{Y}_{reg} = \bar{y} + b(\bar{X} - \bar{x}), \quad (6)$$

with b as traditional slope estimate of regression line. The asymptotic MSE of the estimator is given as,

$$MSE(\hat{Y}_{reg}) = \lambda \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2). \quad (7)$$

The boundary conditions are $\lambda \bar{Y}^2 \rho_{yx}^2 C_y^2 > 0$, $\lambda \bar{Y}^2 (C_x - C_{yx} C_y)^2 > 0$ and $\lambda \bar{Y}^2 (C_x + C_{yx} C_y)^2 > 0$.

(v)- Exponential type ratio estimator $\hat{Y}_{BT,R}$: Bahl and Tuteja (1991) offered exponential type version of ratio estimator such as;

$$\hat{Y}_{BT,R} = \bar{y} \exp \left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right), \quad (8)$$

where, bias is defined as $Bias(\hat{Y}_{BT,R}) = \frac{\lambda\bar{Y}}{8}(3C_x^2 - 4C_y C_x \rho_{yx})$. The expression of MSE is given as,

$$MSE(\hat{Y}_{BT,R}) = \frac{\lambda\bar{Y}^2}{4}(4C_y^2 + C_x^2 - 4C_y C_x \rho_{yx}). \quad (9)$$

(vi)- **Exponential type product estimator $\hat{Y}_{BT,P}$** : Bahl and Tuteja (1991) offered another exponential type estimator as an extension of the product estimator. The expression is given as;

$$\hat{Y}_{BT,P} = \bar{y} \exp\left(\frac{\bar{X} + \bar{x}}{\bar{X} - \bar{x}}\right). \quad (10)$$

The associated bias is $Bias(\hat{Y}_{BT,P}) = \frac{\lambda\bar{Y}}{8}(3C_x^2 + 4C_y C_x \rho_{yx})$. The MSE of the estimator is derived as,

$$MSE(\hat{Y}_{BT,P}) = \frac{\lambda\bar{Y}^2}{4}(4C_y^2 + C_x^2 + 4C_y C_x \rho_{yx}). \quad (11)$$

(vii)- **Difference type estimator $\hat{Y}_{R,D}$** : Rao (1991) suggested a difference type estimator as follows;

$$\hat{Y}_{R,D} = t_1 \bar{y} + t_2 (\bar{X} - \bar{x}), \quad (12)$$

where, t_1 and t_2 are suitably chosen constants having optimal values such as $t_{1(opt)} = \frac{1}{1 + \lambda C_y^2 (1 - \rho_{yx}^2)}$ and

$t_{2(opt)} = \frac{\bar{Y} C_y \rho_{yx}}{\bar{X} C_x [1 + \lambda C_y^2 (1 - \rho_{yx}^2)]}$. The bias of the estimator up to first order is given as $Bias(\hat{Y}_{R,D}) = \bar{Y}(t_1 - 1)$.

The minimum MSE of $\hat{Y}_{R,D}$, up to the first order of approximation, derived as:

$$MSE_{min}(\hat{Y}_{R,D}) = \frac{\lambda\bar{Y}^2 C_y^2 (1 - \rho_{yx}^2)}{1 + \lambda C_y^2 (1 - \rho_{yx}^2)}. \quad (13)$$

(viii)- **Ratio-product exponential type estimator $\hat{Y}_{S,RP}$** : Singh et al. (2008) proposed new estimator as;

$$\hat{Y}_{S,RP} = \bar{y} \left[t_3 \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) + (1 - t_3) \exp\left(\frac{\bar{x} - \bar{X}}{\bar{x} + \bar{X}}\right) \right], \quad (14)$$

where, t_3 is any arbitrary constant with optimal value of $t_3 = \frac{1}{2} + \frac{C_y \rho_{yx}}{C_x}$. Also, the bias of the estimator up to

first order is defined as $Bias(\hat{Y}_{S,RP}) = \frac{\lambda\bar{Y} C_x}{2} ((1 + t_3) C_x - C_y \rho_{yx})$. The minimal MSE expression is;

$$MSE_{min}(\hat{Y}_{S,RP}) = \lambda\bar{Y}^2 C_y^2 (1 - \rho_{yx}^2). \quad (15)$$

(ix)- **Grover and Kaur estimator \hat{Y}_{GK}** : Grover and Kaur (2011) provided this estimator,

$$\hat{Y}_{GK} = (t_4 \bar{y} + t_5 (\bar{X} - \bar{x})) \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right), \quad (16)$$

with $t_{4(opt)} = \frac{-8 + \lambda C_x^2}{8[-1 - \lambda C_y^2 (1 - \rho_{yx}^2)]}$ and $t_{5(opt)} = \frac{\bar{Y}[-8C_y \rho_{yx} + C_x\{4 - \lambda C_x^2 + \lambda C_y C_x \rho_{yx} - 4\lambda C_y^2 (1 - \rho_{yx}^2)\}]}{8\bar{X} C_x [-1 + \lambda C_y^2 (1 - \rho_{yx}^2)]}$ as optimally

chosen constants. Further, the bias is given as, $Bias(\hat{Y}_{GK}) = \bar{Y}(t_4 - 1) + \frac{3}{8}\lambda t_4 C_x^2 + \frac{1}{2}\lambda t_5 \bar{X} C_x^2 - \frac{1}{2}\lambda t_4 \bar{Y} C_y C_x \rho_{yx}$. The resultant expression of minimum MSE is written as;

$$MSE_{min}(\hat{Y}_{GK}) = \frac{\lambda\bar{Y}^2 [\lambda C_x^4 + 16C_y^2 (1 - \rho_{yx}^2) (4 - \lambda C_x^2)]}{64[-1 - \lambda C_y^2 (1 - \rho_{yx}^2)]}. \quad (17)$$

(x)- **Shabbir et al. estimator \hat{Y}_{SHG}** : Shabbir et al. (2014) suggested another estimator as follows,

$$\hat{Y}_{SHG} = \left[\frac{\bar{y}}{2} \left\{ \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) + \exp\left(\frac{\bar{x} - \bar{X}}{\bar{x} + \bar{X}}\right) \right\} + t_6 (\bar{X} - \bar{x}) + t_7 \bar{y} \right] \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right). \quad (18)$$

The constants, $t_{6(opt)} = \frac{\bar{Y}[-8C_y \rho_{yx} + C_x\{2 - \lambda C_x^2 + \lambda C_y C_x \rho_{yx} - 2\lambda C_y^2 (1 - \rho_{yx}^2)\}]}{4\bar{X} C_x [-1 - \lambda C_y^2 (1 - \rho_{yx}^2)]}$ and $t_{7(opt)} = \frac{\lambda\{C_x^2 + 4C_y^2 (1 - \rho_{yx}^2)\}}{4[-1 - \lambda C_y^2 (1 - \rho_{yx}^2)]}$, are

and $Bias(\hat{Y}_{SHG}) = \frac{1}{8} [8t_7 \bar{Y} + \lambda C_x^2 \{4t_6 \bar{X} + \bar{Y}(4 + 3t_7)\} - 4\lambda \bar{Y} C_y C_x \rho_{yx} (1 + t_7)]$, the minimum of MSE is;

$$MSE_{min}(\hat{Y}_{SHG}) = \frac{\lambda\bar{Y}^2 [\lambda C_x^4 + 8C_y^2 (1 - \rho_{yx}^2) (2 - \lambda C_x^2)]}{64[-1 - \lambda C_y^2 (1 - \rho_{yx}^2)]}. \quad (19)$$

2. SUGGESTED SCHEME

We proceed by proposing a new exponential-type estimator of the finite population mean while employing dual use of auxiliary information. The proposed estimator is written in the form as follows,

$$\hat{Y}_{Pr} = \left[\frac{\bar{y}}{2} \left\{ \exp\left(\frac{\bar{x}-\bar{x}}{\bar{x}+\bar{x}}\right) + \exp\left(\frac{\bar{x}-\bar{x}}{\bar{x}+\bar{x}}\right) \right\} + t_8(\bar{X} - \bar{x}) + t_9\bar{y} + t_{10}\left(\frac{\bar{R}-\bar{r}}{\bar{R}}\right) \right] \exp\left(\frac{\bar{x}-\bar{x}}{\bar{x}+\bar{x}}\right), \quad (20)$$

where, t_8 , t_9 and t_{10} are unknown constants chosen to attain minimal mean square error. Equation (20) when rewritten in terms of errors obtains,

$$\hat{Y}_{Pr} = \left\{ \bar{Y}(1 + t_9)(1 + \epsilon_0) - t_8\epsilon_1 - t_{10}\epsilon_2 + \frac{1}{8}\bar{Y}\epsilon_1^2 \right\} \left\{ 1 - \frac{1}{2}\epsilon_1 + \frac{1}{8}\bar{Y}\epsilon_1^2 + \dots \right\}.$$

By keeping errors up to second degree, we attain

$$\begin{aligned} (\hat{Y}_{Pr} - \bar{Y}) &= \left\{ t_9\bar{Y} + t_9\bar{Y}\epsilon_0 + \bar{Y}\epsilon_0 - t_8\epsilon_1 - t_{10}\epsilon_2 - \frac{1}{2}\bar{Y}\epsilon_1 - \frac{1}{2}t_9\bar{Y}\epsilon_1 + \frac{1}{2}\bar{Y}\epsilon_1^2 + \frac{3}{8}t_9\bar{Y}\epsilon_1^2 + \frac{1}{8}t_8\epsilon_1^2 - \right. \\ &\quad \left. \frac{1}{2}\bar{Y}\epsilon_0\epsilon_1 - \frac{1}{2}t_9\epsilon_0\epsilon_1 + \frac{1}{2}t_{10}\epsilon_1\epsilon_2 \right\} \end{aligned} \quad (21)$$

The bias of the estimator is calculated by applying expectation to both sides of the above equation (21) as,

$$\text{Bias}(\hat{Y}_{Pr} - \bar{Y}) = \left\{ t_9\bar{Y} + \frac{1}{2}\lambda\bar{Y}C_x^2 + \frac{3}{8}t_9\lambda\bar{Y}C_x^2 + \frac{1}{8}t_8C_x^2 - \frac{1}{2}\lambda\bar{Y}C_yC_x\rho_{yx} - \frac{1}{2}t_9\lambda C_yC_x\rho_{yx} + \frac{1}{2}t_{10}\lambda C_xC_r\rho_{xr_x} \right\} \quad (22)$$

The MSE of the proposed estimator is then found by squaring and taking expectation of equation (21) as,

$$\begin{aligned} \text{MSE}(\hat{Y}_{Pr}) &= -\lambda\bar{Y}^2C_yC_x\rho_{yx} + \frac{3}{2}t_9\lambda\bar{Y}^2C_x^2 + t_9^2\lambda\bar{Y}^2C_x^2 + t_8\lambda\bar{Y}C_x^2 - 2t_9^2\lambda\bar{Y}^2C_yC_x\rho_{yx} + \lambda\bar{Y}^2C_y^2 + \\ &\quad t_9^2\bar{Y}^2 + t_{10}\lambda\bar{Y}C_yC_r\rho_{yr_x} - 3t_9\lambda\bar{Y}^2C_yC_x\rho_{yx} - 2t_8\lambda\bar{Y}^2C_yC_x\rho_{yx} - 2t_{10}\lambda\bar{Y}^2C_yC_r\rho_{yr_x} - \\ &\quad 2t_9t_{10}\lambda\bar{Y}^2C_yC_r\rho_{yr_x} + \frac{1}{4}\lambda\bar{Y}^2C_x^2 + 2t_9\lambda\bar{Y}^2C_y^2 + t_8^2\lambda C_x^2 + 2t_8t_9\lambda\bar{Y}^2C_x^2 - \\ &\quad 2t_8t_9\lambda\bar{Y}C_yC_x\rho_{yx} + 2t_8t_{10}\lambda C_xC_r\rho_{xr_x} + t_9^2\lambda\bar{Y}^2C_y^2 + t_{10}^2\lambda C_r^2 + 2t_9t_{10}\lambda C_xC_r\rho_{xr_x}. \end{aligned} \quad (23)$$

The optimal values of t_8 , t_9 and t_{10} are obtained by minimising the above expression (23) to give,

$$\begin{aligned} t_{8(opt)} &= \frac{\bar{Y} \left[\lambda C_x^3 \rho_{yx}^2 - \lambda C_y C_x^2 \rho_{yr_x} \rho_{xr_x} - 4 C_y^2 C_x \rho_{yx} \rho_{yr_x} \rho_{xr_x} - \lambda C_x^3 + \lambda C_y C_x^2 \rho_{yx} + 2 \lambda C_y^2 C_x \rho_{yx}^2 + 2 \lambda C_y^2 C_x \rho_{yr_x}^2 + 2 \lambda C_y^2 C_x \rho_{xr_x}^2 \right]}{4 C_x \{-2 \lambda C_y^2 \rho_{yx} \rho_{yr_x} \rho_{xr_x} + \lambda C_y^2 \rho_{yx}^2 + \lambda C_y^2 \rho_{yr_x}^2 - \lambda C_y^2 + \rho_{xr_x}^2 - 1\}}, \\ t_{9(opt)} &= \frac{\lambda \{ C_x^2 \rho_{xr_x}^2 - 8 C_y^2 \rho_{yx} \rho_{yr_x} \rho_{xr_x} - (C_x^2 + 4 C_y^2 \rho_{yx}^2 + 4 C_y^2 \rho_{yr_x}^2 + 4 C_y^2 \rho_{xr_x}^2 - 4 C_y^2) \}}{4 \{-2 \lambda C_y^2 \rho_{yx} \rho_{yr_x} \rho_{xr_x} + \lambda C_y^2 \rho_{yx}^2 + \lambda C_y^2 \rho_{yr_x}^2 - \lambda C_y^2 + \rho_{xr_x}^2 - 1\}}, \text{ and} \\ t_{10(opt)} &= \frac{\bar{Y} C_y \{ \rho_{yx} \rho_{xr_x} - \lambda C_x^2 \rho_{yr_x} - 4 \rho_{yx} \rho_{xr_x} + 4 \rho_{yr_x} \}}{4 \{-2 \lambda C_y^2 \rho_{yx} \rho_{yr_x} \rho_{xr_x} + \lambda C_y^2 \rho_{yx}^2 + \lambda C_y^2 \rho_{yr_x}^2 - \lambda C_y^2 + \rho_{xr_x}^2 - 1\}}. \end{aligned}$$

The minimum MSE is thus derived by reusing optimal values of t 's such as,

$$\text{MSE}_{min}(\hat{Y}_{Pr}) = \frac{\lambda\bar{Y}^2\{16C_y^2(1-\rho_{yxr_x}^2) - \lambda C_x^4 - 8C_y^2C_x^2(1-\rho_{yxr_x}^2)\}}{64\{1+\lambda C_y^2(1-\rho_{yxr_x}^2)\}}, \quad (24)$$

$$\text{where, } \rho_{yxr_x}^2 = \frac{\rho_{yx}^2 + \rho_{yr_x}^2 - 2\rho_{yx}\rho_{yr_x}\rho_{xr_x}}{(1-\rho_{xr_x}^2)}.$$

Below we delineate the mathematical foundations to demonstrate the efficiency of the proposed estimator as compared to the existing methods. These are given in Table 1.

Table 1. The efficiency expressions relative to the existing methods (i) – (x)

(i) \hat{Y}_{srs}	$64C_y^2\{1 + \lambda C_y^2(1 - \rho_{yxr_x}^2)\} - \{16C_y^2(1 - \rho_{yxr_x}^2) - \lambda C_x^4 - 8C_y^2C_x^2(1 - \rho_{yxr_x}^2)\} > 0$
(ii) \hat{Y}_R	$64\{1 + \lambda C_y^2(1 - \rho_{yxr_x}^2)\}(C_y^2 + C_x^2 - C_yC_x\rho_{yx}) - \{16C_y^2(1 - \rho_{yxr_x}^2) - \lambda C_x^4 - 8C_y^2C_x^2(1 - \rho_{yxr_x}^2)\} > 0$
(iii) \hat{Y}_P	$64\{1 + \lambda C_y^2(1 - \rho_{yxr_x}^2)\}(C_y^2 + C_x^2 + C_yC_x\rho_{yx}) - \{16C_y^2(1 - \rho_{yxr_x}^2) - \lambda C_x^4 - 8C_y^2C_x^2(1 - \rho_{yxr_x}^2)\} > 0$
(iv) \hat{Y}_{reg}	$64\{1 + \lambda C_y^2(1 - \rho_{yxr_x}^2)\}C_y^2(1 - \rho_{yx}^2) - \{16C_y^2(1 - \rho_{yxr_x}^2) - \lambda C_x^4 - 8C_y^2C_x^2(1 - \rho_{yxr_x}^2)\} > 0$

(v) $\widehat{Y}_{BT,R}$	$16\{1 + \lambda C_y^2(1 - \rho_{yxr_x^2})\}(4C_y^2 + C_x^2 - 4C_y C_x \rho_{yx}) - \{16C_y^2(1 - \rho_{yxr_x^2}) - \lambda C_x^4 - 8C_y^2 C_x^2(1 - \rho_{yxr_x^2})\} > 0$
(vi) $\widehat{Y}_{BT,P}$	$16\{1 + \lambda C_y^2(1 - \rho_{yxr_x^2})\}(4C_y^2 + C_x^2 + 4C_y C_x \rho_{yx}) - \{16C_y^2(1 - \rho_{yxr_x^2}) - \lambda C_x^4 - 8C_y^2 C_x^2(1 - \rho_{yxr_x^2})\} > 0$
(vii) $\widehat{Y}_{R,D}$	$64\{1 + \lambda C_y^2(1 - \rho_{yxr_x^2})\}C_y^2(1 - \rho_{yx^2}) - (1 + \lambda C_y^2(1 - \rho_{yx^2}))\{16C_y^2(1 - \rho_{yxr_x^2}) - \lambda C_x^4 - 8C_y^2 C_x^2(1 - \rho_{yxr_x^2})\} > 0$
(viii) $\widehat{Y}_{S,RP}$	$64\{1 + \lambda C_y^2(1 - \rho_{yxr_x^2})\}C_y^2(1 - \rho_{yx^2}) - \{16C_y^2(1 - \rho_{yxr_x^2}) - \lambda C_x^4 - 8C_y^2 C_x^2(1 - \rho_{yxr_x^2})\} > 0$
(ix) \widehat{Y}_{GK}	$\{1 + \lambda C_y^2(1 - \rho_{yxr_x^2})\}[\lambda C_x^4 + 16C_y^2(1 - \rho_{yx^2})(4 - \lambda C_x^2)] - \{-1 - \lambda C_y^2(1 - \rho_{yx^2})\}\{16C_y^2(1 - \rho_{yxr_x^2}) - \lambda C_x^4 - 8C_y^2 C_x^2(1 - \rho_{yxr_x^2})\} > 0$
(x) \widehat{Y}_{SHG}	$\{1 + \lambda C_y^2(1 - \rho_{yxr_x^2})\}[\lambda C_x^4 + 8C_y^2(1 - \rho_{yx^2})(2 - \lambda C_x^2)] - \{-1 - \lambda C_y^2(1 - \rho_{yx^2})\}\{16C_y^2(1 - \rho_{yxr_x^2}) - \lambda C_x^4 - 8C_y^2 C_x^2(1 - \rho_{yxr_x^2})\} > 0$

The above results in Table 1 involve demonstrating that for each earlier existing method (i) – (x), that $MSE_{min}(\widehat{Y}_{existing}) - MSE_{min}(\widehat{Y}_{Pr}) > 0$, which is verifiable via using the relevant expressions of optimal MSE’s while comparing with the optimal MSE of our proposed estimator

3. DATA AND RESULTS

This section documents empirical evaluations of the comparative performance of the proposed estimator with respect to the afore-mentioned existing techniques. Summaries of three data sets along with their sources are:

Data 1, source Gujarati (2009): Study variable, Y , the eggs produced (millions) in 1990. Auxiliary variable, X , the price per dozen (cents) in 1991. $N = 50, n = 5, \bar{Y} = 1357.62, \bar{X} = 78.29, \bar{R}_x = 25.50, C_y = 1.2236, C_x = 0.2923, C_{r_x} = 0.5716, \rho_{yx} = -0.3022, \rho_{yr} = -0.2662$ and $\rho_{xr_x} = 0.9574$.

Data 2, source Gujarati (2009): Study variable, Y , sleep (minutes) of people age over 50. Auxiliary variable, X , corresponding age (years). $N = 30, n = 5, \bar{Y} = 384.2, \bar{X} = 67.27, \bar{R}_x = 15.50, C_y = 0.1558, C_x = 0.1373, C_{r_x} = 0.5674, \rho_{yx} = -0.8552, \rho_{yr} = -0.8394$ and $\rho_{xr_x} = 0.9889$.

Data 3, source Murthy (1967): Study variable, Y , factory output. Auxiliary variable, X , number of workers. $N = 80, n = 10, \bar{Y} = 5182.64, \bar{X} = 285.13, \bar{R}_x = 40.50, C_y = 0.3542, C_x = 0.9485, C_{r_x} = 0.5738, \rho_{yx} = 0.9150, \rho_{yr} = 0.9836$ and $\rho_{xr_x} = 0.8902$.

The above data sets include various parametric settings primarily incorporating varying strengths of shared reports the comparative performance of rival estimators using \widehat{Y}_{SRS} as a base technique in the calculation of percentage relative efficiencies (PRE) using optimal MSE values, that is $PRE = \frac{MSE(\widehat{Y}_{SRS})}{MSE(\widehat{Y}_{Other})} \times 100$.

As seen in Table 2, gains of our exponential-type estimator are significant. The efficiency of the devised estimator becomes more pronounced as the degree of correlation between the study and auxiliary variable increases, consistent with traditional theory regarding use of auxiliary information. Table 2 also presents the underlying performance ordering of the estimators under study.

Table 2. Percentage relative efficiency (PRE) and Performance ordering versus Data sets 1,2 and 3.

Data	Estimators										
	\widehat{Y}_{SRS}	\widehat{Y}_R	\widehat{Y}_P	\widehat{Y}_{reg}	$\widehat{Y}_{BT,R}$	$\widehat{Y}_{BT,P}$	$\widehat{Y}_{R,D}$	$\widehat{Y}_{SR,P}$	\widehat{Y}_{GK}	\widehat{Y}_{SHG}	\widehat{Y}_{Pr}
1	100	83.22	109.56	110.05	92.04	106.14	137.01	110.05	137.53	138.07	552.29
2	100	30.45	371.43	372.35	51.34	372.35	226.97	372.75	373.10	373.55	1494.22
3	100	300.57	7.65	614.21	291.95	19.07	615.31	614.21	664.37	827.06	3308.25
Performance hierarchy of estimators with respect to each data set: Performance ordering											
1	$\widehat{Y}_{Pr} > \widehat{Y}_{SHG} > \widehat{Y}_{GK} > \widehat{Y}_{R,D} > \widehat{Y}_{SR,P} = \widehat{Y}_{reg} > \widehat{Y}_P > \widehat{Y}_{BT,P} > \widehat{Y}_{SRS} > \widehat{Y}_{BT,R} > \widehat{Y}_R$										
2	$\widehat{Y}_{Pr} > \widehat{Y}_{SHG} > \widehat{Y}_{GK} > \widehat{Y}_{R,D} > \widehat{Y}_{SR,P} = \widehat{Y}_{reg} > \widehat{Y}_P > \widehat{Y}_{BT,P} > \widehat{Y}_{SRS} > \widehat{Y}_R > \widehat{Y}_{BT,R}$										
3	$\widehat{Y}_{Pr} > \widehat{Y}_{SHG} > \widehat{Y}_{GK} > \widehat{Y}_{R,D} > \widehat{Y}_{SR,P} = \widehat{Y}_{reg} > \widehat{Y}_P > \widehat{Y}_{BT,R} > \widehat{Y}_{SRS} > \widehat{Y}_{BT,P} > \widehat{Y}_R$										

4. CONCLUSION

Our proposed procedure outperforms the contemporary techniques with respect to all considered parametric settings. Improved performance of the newly suggested estimator is attributed to its ability to incorporate auxiliary information more rigorously, whereas the performance ordering of other methods depends on varying parametric settings. As expected, $\hat{Y}_{SR,S}$ outperforms the ratio estimator and its variant when there is negative correlation between the study and auxiliary variable. Similarly, minimal performance is attributed with \hat{Y}_P and its variant, in the case of a positive linear relationship between study and auxiliary variable, see Cochran (1977). The performance of the regression estimator always equates to that of the $\hat{Y}_{SR,P}$, as shown mathematically by Singh *et al.* (2008). In future, we aim to extend the proposed scheme for other sampling schemes, such as stratified random sampling and two-phase systematic sampling. The applicability of the suggested estimator in situations when group structure is prevalent in the population, and its generalization to multi-auxiliary variables, or where significant missingness exists remain issues for future works.

REFERENCES

- Bahl, S. and Tuteja, R. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*. 12(1), 159–164.
- Bai, Y., Liang, S. and Yuan, W. (2021). Estimating global gross primary production from sun-induced chlorophyll fluorescence data and auxiliary information using machine learning methods. *Remote Sensing*. 13, 963. <https://doi.org/10.3390/rs13050963>.
- Biemer, P. P. and Peytchev, A. (2013). Using geocoded census data for nonresponse bias correction: An assessment. *Journal of Survey Statistics and Methodology*. 1(1), 24-44.
- Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*. 30(2), 262–275.
- Cochran, W. G. (1977). *Sampling techniques*. 3rd ed., John Wiley and Sons, New York.
- Chou, W., Imai, K. and Rosenfeld, B. (2017). Sensitive survey questions with auxiliary information. *Sociological Research & Methods*. 49(2), 418-454.
- Grover, L. K. and Kaur, P. (2011). An improved estimator of the finite population mean in simple random sampling. *Model Assisted Statistics and Applications*. 6(1), 47–55.
- Gujarati, D. N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education.
- Gupta, S. and Shabbir, J. (2008). On improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics*. 35(5), 559–566.
- Haberman, S. J. (1984). Adjustment by minimum discriminant information. *Annals of Statistics*. 12, 121-140.
- Haq, A., Munir, T. and Shah, B. A. (2020). Dual multivariate CUSUM charts with auxiliary information for process mean. *Quality and Reliability Engineering*, 36(3), 83-89.
- Holt, D. and Elliot, D. (1991). Methods of weighting for unit non-response. *The Statistician, Special Issue: Survey Design, Methodology and Analysis*. 40, 333–342.
- Hussain S, Ahmad S, Saleem M and Akhtar S (2020). Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. *PLoS ONE* 15(9): e0239098. <https://doi.org/10.1371/journal.pone.0239098>
- Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev A., Groves, R.M., and Raghunathan, T. E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *Journal of the Royal Statistical Society: Series A*. 173(2), 389-407.
- Murthy, M. (1964). Product method of estimation. *Sankhya A*, 26(1), 69–74.
- Olson, K. (2013). Paradata for nonresponse adjustment. *The Annals of American Academy of Political and Social Sciences*, 645(1), 142-170.
- Rao, T.J. (1991). On certain methods of improving ratio and regression estimators. *Communications in Statistics-Theory and Methods*. 20(10), 3325–3340.
- Rao, J. N. K., Kovar, J. G., and Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*. 77, 365–375.
- Shabbir, J., Haq, A., and Gupta, S. (2014). A new Difference-Cum-Exponential type estimator of finite population mean in simple random sampling. *Revista Colombiana de Estadística*. 37(1), 199–211.
- Singh, R., Chauhan, P., and Sawan, N. (2008). On linear combination of ratio and product type exponential estimator for estimating the finite population mean. *Statistics in Transition*. 9(1), 105–115.
- S. K. Srivastava (1971). A generalized estimator for the mean of a finite population using multi-auxiliary information, *Journal of American Statistical Association*, 66, 404-407.
- Zhang, L. and Chambers, R. L. (2004). Small area estimates for cross-classifications. *Journal of the Royal Statistical Society, Series B*. 66(2), 479–496.