Thailand Statistician
July 2022; 20(3): 575-584
http://statassoc.or.th
Contributed paper

A Predictive Approach for Finite Population Mean when Auxiliary Variables are Attributes

Ashish Kumar and Monika Saini*
Department of Mathematics and Statistics, Manipal University Jaipur, Jaipur (Rajasthan), India.
*Corresponding author; e-mail: drmnksaini4@gmail.com

Received: 13 March 2020
Revised: 6 July 2020
Accepted: 4 August 2020

Abstract

In this paper, the issue of estimating the finite population mean utilizing predictive approach has been deliberated when data on auxiliary attribute is accessible. Exponential ratio cum product estimator are proposed utilizing Singh et al. (2007) ratio and product type exponential estimators as a predictor of the mean unobserved units of the population. The bias and mean square error (MSE) of proposed estimators have been gained up to the first order of approximation. The theoretical conditions under which the proposed estimator is more efficient than the usual unbiased, Naik and Gupta (1996) and Singh et al. (2007) estimators. A practical study is carried out in support of theoretical results.

Keywords: Exponential estimators, point bi-serial correlation, finite population, mean square error.

1. Introduction

It is notable that the proper utilization of auxiliary data gives the more productive estimators when associated with study variable for estimating the population mean. There are several applied circumstances, when auxiliary data is existing in the form of attribute, which are associated with study variable. The instances of this state are: sex is viewed as a good auxiliary attribute in estimating the tallness of people, bread of a dairy animals is a worthy auxiliary attribute while estimating milk production, crop assortment is utilized as an auxiliary attribute in estimating the yield of wheat (see Naik and Gupta (1996), Jhajj et al. (2006), etc.), in all of these models point bi-serial correlation coefficient between the auxiliary attribute and the study variable exists. A great variety of approaches/techniques is available to construct more efficient design based and model based methods for utilizing auxiliary information (variable/attribute).

For estimation of population mean of study variable utilizing auxiliary data on model based approach is also notorious as predictive approach. Predictive approach for sampling surveys can be considered as general framework for statistical inference on the character of finite population. In a predictive approach a model is expressed for the population values and is utilized to predict the non-sampled values. Well-known estimators of population totals encounter in the classical theory, as expansion, ratio, regression, another estimators, can be predictors is a general prediction theory, under some special model. Numerous authors have utilized ratio, product and regression type estimators under predictive approach for population parameters when auxiliary data is present in from of variables.

Srivastava (1983) prescribed the predictive approach for product estimator also exploited the prediction criterion given by Basu (1971) and affirmation that if the standard product estimator is utilized as a predictor for the mean of the unobserved units of the population, the resulting estimator of the mean of the entire population is unique in relation to the conventional product estimator. Biradar and Singh (1998), Agarwal and Roy (1999) and Nayak and Sahoo (2012) conveyed some predictive estimators for finite population variance. Saini (2013) proposed a class of predictive estimators for two phase sampling entailing ratio, regression and product estimators. Singh et al. (2014) proposed ratio and product type exponential estimators of the population mean of a study variable through predictive approach using Bahl and Tuteja (1991) ratio and product type exponential estimators as a predictor, Yadav and Mishra (2015) recommended upgraded ratio cum product type predictive estimators to estimate the population mean. Yadav et al. (2018) suggested predictive estimation of finite population mean using coefficient of kurtosis and median of an auxiliary variable under simple random sampling scheme. Mulwa and Kilai (2019) proposed estimation of finite population mean under model based approach using auxiliary variables. Kumar and Saini (2020) recommended exponential ratio and product estimators for finite population mean using auxiliary attributes under predictive approach.

Motivated by Singh et al. (2014) and Yadav and Mishra (2015), here we attempt to examine the existing Singh et al. (2007) exponential estimators under data on auxiliary attribute as predictor of the mean of the unobserved units of the population using the information of the observed units in the sample. The remaining part of the paper has been organized as follows. In Section 2 we discussed the material and methods. In Section 3, the proposed estimators and their properties are created. Bias and efficiency comparison between proposed and existing estimators are carried out in Section 4 and Section 5. In Section 6, we observe the performance of different estimators using real data sets. Last section sketches some concluding remarks.

2. Methodology

Let us consider a review population $\tau=(\tau_1,\tau_2,...,\tau_M)$ which consists of M identifiable units. Let y_i and Δ_i be the observations on the study variable y and the auxiliary attribute for Δ respectively for the i^{th} unit of the population, for $i=1,2,...,M$. Assume that there is a complete contrast in the population with respect to the presence $(\Delta=1)$ or absence $(\Delta=0)$ of an attribute Δ. We take the following notations which we will be used along the paper.

Let $m(<M)$ be the size of the sample selected using simple random sampling without replacement (SRSWOR) from a population. Let $H=\sum_{i=1}^{M}\Delta_i$ and $h=\sum_{i=1}^{m}\Delta_i$ denoted the total number of units in the population and sample, respectively, that possess the attribute Δ, respectively. Let the corresponding population and sample proportions be $Z=\sum_{i=1}^{M}\Delta_i/M=H/M$ and $z=\sum_{i=1}^{m}\Delta_i/m=h/m,$ respectively. We are interested in estimating the population mean $\overline{Y}$ of the study variable on the basis of observed values of y on units in a sample from τ. Let S denotes the set of all possible samples from τ. Let s be a member of S (i.e., $s\in S$), let $\eta(s)$ denotes the

effective sample size that is the number of distinct units in s and $\tilde{s}$ denote the range of all those units of τ which are not is s. We make

$$\overline{Y}_s = \frac{1}{\eta(s)}\sum_{i\in s} y_i \text{ and } \overline{Y}_{\tilde{s}} = \frac{1}{M-\eta(s)}\sum_{i\in \tilde{s}} y_i. \tag{1}$$

Under the usual predictive set-up, it is possible to express $\overline{Y}$ as

$$\overline{Y} = \frac{1}{M}\sum_{i=1}^{M} y_i = \frac{1}{M}\left[\sum_{i\in s} y_i + \sum_{i\in \tilde{s}} y_i\right]. \tag{2}$$

For any given $s \in S$, using Equations (1) and (2), can be written as

$$\overline{Y} = \frac{\eta(s)}{M}\overline{Y}_s + \frac{(M-\eta(s))}{M}\overline{Y}_{\tilde{s}}. \tag{3}$$

Basu (1971) debated that in this representation of $\overline{Y}$, the sample mean $\overline{Y}_{\tilde{s}}$, being based on the units in the sample s whose y values have been observed, is known. The statistician, therefore, should effort a prediction of the mean $\overline{Y}_{\tilde{s}}$ of the unobserved units of the population on the basis of observed units in s. While admitting that a decision-theorist might object to making the choice of estimator after looking at the data, Basu (1971) nevertheless considered such an approach to represent the "heart of the matter" in estimating the finite population mean (see Cessal et al. (1977), pp.110).

In simple random sampling procedure, the sample mean for the sample of size n (i.e.,$\eta(s) = m$) is

$$\overline{Y}_s = \frac{1}{m}\sum_{i\in s} y_i = \overline{y} \text{ and } \overline{Y}_{\tilde{s}} = \frac{1}{M-m}\sum_{i\in \tilde{s}} y_i. \tag{4}$$

Using (4), $\overline{Y}$ in (3) can be written as

$$\overline{Y} = \frac{m}{M}\overline{y} + \frac{(M-m)}{M}\overline{Y}_{\tilde{s}}. \tag{5}$$

In the light of (5), an appropriate estimator of population mean $\overline{Y}$ can be written as

$$p = \left[\frac{m}{M}\overline{y} + \frac{(M-m)}{M}P\right], \tag{6}$$

where P is painstaking as a predictor of $\overline{Y}_{\tilde{s}}$. If we implement the prediction approach, proposed by Srivastava (1983), for the population mean $\overline{Y}_{\tilde{s}}$ of the unobserved units of the population:

In case no additional information on π is available, an obvious choice of P is $\overline{y}$, i.e.

$$P = \frac{1}{m}\sum_{i=\tilde{s}} y_i = \overline{y}. \tag{7}$$

Using (7), we can write (6) as

$$p = \left[\frac{m}{M}\overline{y} + \frac{(M-m)}{M}\overline{y}\right] = \overline{y},$$

the routine mean per unit estimator of $\overline{Y}$. Captivating the point bi-serial correlation coefficient between auxiliary attribute and study variable, Naik and Gupta (1996) proposed ratio estimator of population mean when the former information of population proportion of units, holding the same attribute is existing.

Ratio estimator proposed by Naik and Gupta (1996) in the form of an obvious choice of P is

$$P = \pi_{\tilde{R}} = \left(\frac{\bar{y}}{z}\right) Z_{\tilde{s}},$$

where $z = \frac{1}{m}\sum_{i=s}\Delta_i$, $Z = \frac{1}{M}\sum_{i=1}^{M}\Delta_i$ and $Z_{\tilde{s}} = \frac{1}{M-m}\sum_{i\in\tilde{s}}\Delta_i = \frac{MZ-mz}{M-m}.$

For this choice of P,

$$p = \frac{m}{M}\bar{y} + \frac{(M-m)}{M}\left(\frac{\bar{y}}{z}\right) Z_{\tilde{s}} = \frac{m}{M}\bar{y} + \frac{M-m}{M}\frac{MZ-mz}{M-m}\left(\frac{\bar{y}}{z}\right) = \left(\frac{\bar{y}}{z}\right) Z,$$

which is the customary ratio estimator of $\bar{Y}$ proposed by Naik and Gupta (1996).

If the information on the auxiliary attribute Δ is negatively correlated with study variable y is available and one propose to use this information in the form of product estimator in the form of Naik and Gupta (1996), an obvious choice of P is $P = \pi_{\tilde{P}} = \left(\frac{\bar{y}}{Z_{\tilde{s}}}\right) z.$

For the obvious choice of P,

$$p = \frac{m}{M}\bar{y} + \frac{M-m}{M}\frac{\bar{y}\,z}{Z_{\tilde{s}}} = \frac{m}{M}\bar{y} + \frac{M-m}{M}\frac{M-m}{MZ-mz}\bar{y}\,z = \bar{y}\frac{MZ+(M-2m)z}{MZ-mz} = \pi_{Pp},$$

which is not the existing product estimator $\pi_P = \left(\frac{\bar{y}}{Z}\right) z$ of $\bar{Y}$ proposed by Naik and Gupta (1996).

Thus we find that if we adopt the prediction approach proposed by Srivastava (1983), use of mean per unit, ratio estimator for predicting the mean $\bar{Y}_{\tilde{s}}$ of the unobserved units of the population outcomes in the usual estimators of the population mean $\bar{Y}$. However, if the product estimator is used with such an approach, the resulting estimator of $\bar{Y}$ is not the customary product estimator π_P.

The biases and mean squared errors of the estimators π_R, π_P and π_{Pp}, up to the first order of approximations are found as:

$$Bias(\pi_R) = \Theta\bar{Y}C_Z^2\left(1-\xi\right), \tag{8}$$

$$Bias(\pi_P) = \Theta\bar{Y}C_Z^2\xi, \tag{9}$$

$$Bias(\pi_{P_p}) = \Theta\bar{Y}C_Z^2\left(\xi + f(1-f)^{-1}\right), \tag{10}$$

$$MSE(\pi_R) = \Theta\bar{Y}^2\left[C_y^2 + C_Z^2\left(1-2\xi\right)\right], \tag{11}$$

$$MSE(\pi_P) = MSE(\pi_{P_p}) = \Theta\bar{Y}^2\left[C_y^2 + C_Z^2\left(1+2\xi\right)\right], \tag{12}$$

where $\Theta = m^{-1}(1-f)$, $f = (m/M)$ (sample fraction), $C_y^2 = S_y^2/\bar{Y}^2$ (population coefficient of variation of y), $S_y^2 = (M-1)^{-1}\sum_{i=1}^{M}(y_i - \bar{Y})$ (population mean square of y), $C_z^2 = S_z^2/Z^2$ (population coefficient of variation of Δ), $S_z^2 = (M-1)^{-1}\sum_{i=1}^{M}(\Delta_i - Z)^2$ (population mean square of Δ), $\xi = \rho\, C_y/C_z$, $\rho = S_{yz}/(S_y S_z)$ (bi-serial correlation coefficient between y and Δ), $S_{yz} = \sum_{i=1}^{M}(y_i - \bar{Y})(\Delta_i - Z)$ (population covariance between y and Δ).

Following Bahl and Tuteja (1991), Singh et al. (2007) suggested the following ratio and product type exponential estimators using information on auxiliary attribute

$$\pi_{\mathrm{Re}} = \bar{y}\exp\left(\frac{Z-z}{Z+z}\right) \text{ and } \pi_{P\mathrm{e}} = \bar{y}\exp\left(\frac{z-Z}{z+Z}\right).$$

The biases and mean squared errors of π_{Re} and $\pi_{P\mathrm{e}}$ up to the first degree of approximation are

$$Bias(\pi_{\mathrm{Re}}) = \frac{\Theta}{8}\bar{Y}C_z^2(3-4\xi), \quad Bias(\pi_{P\mathrm{e}}) = \frac{\Theta}{8}\bar{Y}C_z^2(4\xi-1),$$

$$MSE(\pi_{\mathrm{Re}}) = \frac{\Theta}{8}\bar{Y}^2\left[C_y^2 + \frac{C_z^2}{4}(1-4\xi)\right], \tag{13}$$

$$MSE(\pi_{P\mathrm{e}}) = \frac{\Theta}{8}\bar{Y}^2\left[C_y^2 + \frac{C_z^2}{4}(1+4\xi)\right]. \tag{14}$$

In this paper, we projected the ratio and product type exponential estimators of population mean $\bar{Y}$ by using π_{Re} and $\pi_{P\mathrm{e}}$ as a predictor P of $\bar{Y}_{\tilde{s}}$ of the unobserved units of the population τ on the bases of observed units in s. The bias and mean squares error of proposed ratio and product type exponential estimators up to the first order of approximation have obtained.

3. Proposed Estimators and their Properties

In case, information on an auxiliary attribute Δ positively correlated with the study variable y is available and one plans to use this in the form of Singh et al. (2007) ratio type exponential estimator π_{Re}, an obvious choice of P is $\pi_{\tilde{\mathrm{Re}}} = \bar{y}\exp\left(\frac{Z_{\tilde{s}}-z}{Z_{\tilde{s}}+z}\right)$. For this choice of P:

$$p = p_{\mathrm{Re}} = \left[\frac{m}{M}\bar{y} + \left(\frac{M-m}{M}\right)\bar{y}\exp\left(\frac{Z_{\tilde{s}}-z}{Z_{\tilde{s}}+z}\right)\right] = \left[\frac{m}{M}\bar{y} + \left(\frac{M-m}{M}\right)\bar{y}\exp\left(\frac{M(Z-z)}{M(Z-z)-2mz}\right)\right], \tag{15}$$

which is not the Singh et al. (2007) ratio type exponential estimator using auxiliary attribute.

If the information of auxiliary attribute Δ is negatively correlated with the study variable y and one wants to use this information in the form of Singh et al. (2007) product type exponential estimator $\pi_{P\mathrm{e}}$ for an obvious choice of P is

$$\pi_{\tilde{P}\mathrm{e}} = \bar{y}\exp\left(\frac{z-Z_{\tilde{s}}}{z+Z_{\tilde{s}}}\right),$$

$$p = p_{P\mathrm{e}} = \left[\frac{m}{M}\bar{y} + \left(\frac{M-m}{M}\right)\bar{y}\exp\left(\frac{z-Z_{\tilde{s}}}{z+Z_{\tilde{s}}}\right)\right] = \left[\frac{m}{M}\bar{y} + \left(\frac{M-m}{M}\right)\bar{y}\exp\left(\frac{M(z-Z)}{MZ+(M-2m)z}\right)\right], \tag{16}$$

which is not the Singh et al. (2007) product type exponential estimator using auxiliary attribute.

Motivated by Singh et al. (2014) and Yadav and Mishra (2015), we have proposed the following ratio cum product estimator in predictive estimation approach as:

$$\theta = \kappa p_{\mathrm{Re}} + (1-\kappa)p_{Pe}, \tag{17}$$

where p_{Re} and p_{Pe} are the estimators defined in (15) and (16) and is characterizing scalar to be determined such that the mean square error of the proposed estimator θ is minimum.

To study the properties of the proposed estimator θ, we define $\Sigma_0 = \left(\dfrac{\bar{y}-\bar{Y}}{\bar{Y}}\right)$ and $\Sigma_1 = \left(\dfrac{z-Z}{Z}\right)$ such that $E(\Sigma_0) = E(\Sigma_1) = 0$ and up to the first order of approximation $E(\Sigma_0^2) = \Theta C_y^2$, $E(\Sigma_1^2) = \Theta C_z^2$, $E(\Sigma_0\Sigma_1) = \Theta\xi C_z^2$.

Expressing (15) in terms of Σ 's, we have

$$p_{\mathrm{Re}} = \bar{Y}(1+\Sigma_0)\left[f+(1-f)\exp\left(-\frac{\Sigma_1}{2(1-f)}\left\{1+\frac{(1-2f)}{2(1-f)}\Sigma_1\right\}^{-1}\right)\right],$$

Escalating the right hand side of above equation, simplifying after multiplication and neglecting terms of Σ 's having power greater than two, we have

$$p_{\mathrm{Re}} \approx \bar{Y}\left[1+\Sigma_0-\frac{\Sigma_1}{2}-\frac{\Sigma_0\Sigma_1}{2}+\frac{\Sigma_1^2}{8}(3-4f)\right]. \tag{18}$$

Now expressing (16) in terms of Σ 's, we have

$$p_{P\mathrm{e}} = \bar{Y}(1+\Sigma_0)\left[f+(1-f)\exp\left(\frac{\Sigma_1}{2(1-f)}\left\{1+\frac{\Sigma_1}{2(1-f)}\right\}^{-1}\right)\right].$$

Expanding the right hand side of above equation, multiplying out and neglecting terms of Σ 's having power greater than two we have

$$p_{P\mathrm{e}} \approx \bar{Y}\left[1+\Sigma_0+\frac{\Sigma_1}{2}+\frac{\Sigma_0\Sigma_1}{2}-\frac{\Sigma_1^2}{8(1-f)}\right]. \tag{19}$$

Now expressing θ from (17) in terms of Σ 's, using (18) and (19), we have

$$\theta = \bar{Y}\left[\kappa\left\{1+\Sigma_0-\frac{\Sigma_1}{2}-\frac{\Sigma_0\Sigma_1}{2}+\frac{\Sigma_1^2}{8}(3-4f)\right\}+(1-\kappa)\left\{1+\Sigma_0+\frac{\Sigma_1}{2}+\frac{\Sigma_0\Sigma_1}{2}-\frac{\Sigma_1^2}{8(1-f)}\right\}\right].$$

After simplification, we get

$$\theta = \bar{Y}\left[1+\Sigma_0-(2\kappa-1)\frac{\Sigma_1}{2}-(2\kappa-1)\frac{\Sigma_0\Sigma_1}{2}+\frac{\Sigma_1^2}{8}\left\{\frac{4\kappa-1-7f+4f^2}{(1-f)}\right\}\right],$$

$$\theta-\bar{Y} \approx \bar{Y}\left[\Sigma_0-(2\kappa-1)\frac{\Sigma_1}{2}-(2\kappa-1)\frac{\Sigma_0\Sigma_1}{2}+\frac{\Sigma_1^2}{8}\left\{\frac{4\kappa-1-7f+4f^2}{(1-f)}\right\}\right]. \tag{20}$$

Taking expectation on both sides, we the bias of θ as

$$Bias(\theta) = \bar{Y}\left[E(\Sigma_0)-(2\kappa-1)\frac{E(\Sigma_1)}{2}-(2\kappa-1)\frac{E(\Sigma_0\Sigma_1)}{2}+\frac{E(\Sigma_1^2)}{8}\left\{\frac{4\kappa-1-7f+4f^2}{(1-f)}\right\}\right].$$

Substituting the values of different expectations, we get

$$Bias(\theta) = \bar{Y}\left[-(2\kappa-1)\frac{1}{2}\Theta\xi C_z^2+\frac{1}{8}\Theta C_z^2\left\{\frac{4\kappa-1-7f+4f^2}{(1-f)}\right\}\right].$$

Squaring both sides of (20) and neglecting terms of Σ 's having power greater than two we have

$$(\theta-\bar{Y})^2 \approx \bar{Y}^2\left[\Sigma_0-(2\kappa-1)\frac{\Sigma_1^2}{2}\right],$$

Taking expectation on both side and we get the MSE of θ up to the first order of approximation as:

$$MSE(\theta) = E(\theta - \bar{Y})^2 = \bar{Y}^2 E\left[\Sigma_0 - (2\kappa - 1)\frac{\Sigma_1^2}{2}\right]^2 = \bar{Y}^2 E\left[\Sigma_0 - \kappa_1 \frac{\Sigma_1^2}{2}\right]^2,$$

where $\kappa_1 = (2\kappa - 1)$,

$$MSE(\theta) = E(\theta - \bar{Y})^2 = \bar{Y}^2\left[E(\Sigma_0^2) + \kappa_1^2 \frac{E(\Sigma_1^4)}{4} - \kappa_1 E(\Sigma_0 \Sigma_1)\right].$$

Substituting the values of different expectations, we get:

$$MSE(\theta) = \bar{Y}^2\left[\Theta C_y^2 + \kappa_1^2 \frac{1}{4}\Theta C_z^2 - \kappa_1 \Theta \xi C_z^2\right]. \tag{21}$$

Partially differentiating equation (21) with respect to κ_1 and equation to zero, we get the optimum value of κ_1 as

$$\kappa_1 = 2\xi \text{ or } \kappa = \frac{1}{2}(1 + 2\xi),$$

We get the minimum MSE of θ as

$$MSE(\theta)_{\min} = \Theta \bar{Y}^2 (C_y^2 - \xi C_z^2). \tag{22}$$

4. Efficiency Comparison of Estimators

In this section, we have obtained the conditions under which the proposed estimator θ (ratio-cum product estimator) is better than the π_R (ratio estimator), π_P (product estimator) proposed by Naik and Gupta (1996) and π_{Re} (ratio-type exponential estimator), π_{Pe} (product-type exponential estimator) proposed by Singh et al. (2007).

It is well-known that the sample mean $\bar{y}$ is an unbiased estimator of the population mean $\bar{Y}$ under simple random sampling without replacement, we have

$$Var(\bar{y}) = MSE(\bar{y}) = \Theta \bar{Y}^2 C_y^2. \tag{23}$$

and if an estimator is unbiased, its MSE is equal to its variance. From (23), (22), (11), (12), (13) and (14), we have

Condition (i) $MSE(\bar{y}) > MSE(\theta)_{\min}$ if and only if $\xi C_z^2 > 0$. (24)

Condition (ii) $MSE(\pi_R) > MSE(\theta)_{\min}$ if and only if $C_z^2(1 - \xi) > 0$. (25)

Condition (iii) $[MSE(\pi_P) = MSE(\pi_{Pp})] > MSE(\theta)_{\min}$ if and only if $C_z^2(1 + 3\xi) > 0$.
(26)

Condition (iv) $MSE(\pi_{\text{Re}}) > MSE(\theta)_{\min}$ if and only if $\frac{1}{8}\left[C_y^2 + \frac{C_z^2}{4}(1 - 4\xi)\right] > \left[C_y^2 - \xi C_z^2\right]$. (27)

Condition (v) $MSE(\pi_{Pe}) > MSE(\theta)_{\min}$ if and only if $\frac{1}{8}\left[C_y^2 + \frac{C_z^2}{4}(1 + 4\xi)\right] > \left[C_y^2 - \xi C_z^2\right]$. (28)

Thus from conditions (24)-(28), we conclude that the proposed ratio cum product estimator θ is efficient than $\bar{y}$, π_R (ratio estimator), π_P (ratio estimator proposed by Naik and Gupta (1996) and π_{Re} (ratio-type exponential estimator), π_{Pe} (product-type exponential estimator) proposed by Singh et al. (2007).

5. Empirical Study

To evaluate the performance of the proposed ratio cum product estimator θ over the estimators $\bar{y}$, π_R, π_P, π_{Re} and $\pi_{P\text{e}}$ of estimation of population mean $\bar{Y}$ we have considered the following natural population dataset:

Data 1: (Source: Mukhopadhyaya (2000), p. 44)

The variables are defined as: y is the household size and Δ is a household that availed an agricultural loan from a bank. For this data, we have $N = 25$, $\bar{Y} = 9.44$, $Z = 0.40$, $\rho = -0.387$, $C_y = 0.17028$, $C_Z = 1.27478$ and $n = 7$.

Data 2: (Source: Gujarati et al. (2009), p. 577)

The variables are defined as: y is the income (thousands of dollars) and Δ is a home ownership. For this data, we have $N = 40$, $\bar{Y} = 14.40$, $Z = 0.53$, $\rho = 0.897$, $C_y = 0.955$, $C_Z = 0.384$ and $n = 10$.

Data 3: (Source: Chatterjee and Hadi (2013), p. 159)

The variables are defined as: y is the height (inches) and Δ is the gender. For this data, we have $N = 60$, $\bar{Y} = 66.90$, $Z = 0.30$, $\rho = -0.449$, $C_y = 0.044$, $C_Z = 0.621$ and $n = 10$.

To compare the efficiency of different estimators of the population mean $\bar{Y}$, we have figured the percentage relative efficiency (PREs) using R (R Core Team 2013) of different estimators with deference to $\bar{y}$ by using the following way:

$$PRE(\pi_R, \bar{y}) = \frac{MSE(\bar{y})}{MSE(\pi_R)} \times 100 = \frac{C_y^2}{[C_y^2 + C_z^2(1-2\xi)]} \times 100,$$

$$PRE(\pi_P, \bar{y}) = \frac{MSE(\bar{y})}{MSE(\pi_P)} \times 100 = \frac{C_y^2}{[C_y^2 + C_z^2(1+2\xi)]} \times 100,$$

$$PRE(\pi_{\text{Re}}, \bar{y}) = \frac{MSE(\bar{y})}{MSE(\pi_{\text{Re}})} \times 100 = \frac{C_y^2}{[C_y^2 + \frac{C_z^2}{4}(1-4\xi)]} \times 100,$$

$$PRE(\pi_{P\text{e}}, \bar{y}) = \frac{MSE(\bar{y})}{MSE(\pi_{P\text{e}})} \times 100 = \frac{C_y^2}{[C_y^2 + \frac{C_z^2}{4}(1+4\xi)]} \times 100,$$

$$PRE(\theta, \bar{y}) = \frac{MSE(\bar{y})}{MSE(\theta)_{\min}} \times 100 = \frac{C_y^2}{[C_y^2 - \xi C_z^2]} \times 100.$$

It is observed from Table 1 that the percentage relative efficiency (PRE) of the proposed ratio cum product estimator θ with respect to the traditional unbiased estimator $\bar{y}$ is larger than the percentage relative efficiency (PREs) of estimators π_R, π_P, π_{P_p}, π_{Re} and π_{Pe} of population mean in predictive estimation approach using population data set 1, 2 and 3 when bi-serial correlation coefficient exist between study variable and auxiliary attribute.

Table 1 Percentage relative efficiency (PREs)

$PRE(s)$	Population		
	Data 1	Data 2	Data 3
$PRE(\pi_{\mathrm{R}},\bar{y})$	1.591	109.567	0.463
$PRE(\pi_{P},\bar{y})=PRE(\pi_{Pp},\bar{y})$	1.951	59.381	0.525
$PRE(\pi_{\mathrm{Re}},\bar{y})$	5.583	135.82	1.727
$PRE(\pi_{P\mathrm{e}},\bar{y})$	8.254	75.992	2.215
$PRE(\theta,\bar{y})$	25.659	140.812	13.545

Note: Intrepid number indicate the highest percentage relative efficiency in the relevant data set

6. Concluding Remarks

The foremost objective of statistician is to minimize the mean square error in estimation to ideally infer the parameter of the given population. To this end, we have made comparisons of proposed estimator with existing estimators. In this paper we projected the ratio cum product estimator when auxiliary information is accessible in form of auxiliary attribute. The characteristics of the proposed estimator are derived up to first order of approximation. We derived the theoretical conditions under which the proposed estimators are less biased and more efficient than the traditional unbiased estimator, ratio and product estimators proposed by Naik and Gupta (1996) and Singh et al. (2007) ratio and product type exponential estimators. This fact has been also supported through an empirical study using real data sets. The results of this article is quite informative both theoretically and empirically. In this manner, we recommend the use of proposed estimators in practice.

Acknowledgements

Authors are thankful to the reviewers for their valuable comments that enhance the quality of the paper.

References

Agrawal MC, Roy DC. Efficient estimators of population variance with regression-type and ratio-type predictor-inputs. Metron. 1999; 57(3): 4-13.

Bahl S, Tuteja R. Ratio and product type exponential estimators. J Inf Optim Sci. 1991; 12(1): 159-164.

Basu D. An essay on the logical foundations of survey sampling, Part I. Foundations of Statistical Inferences, VP Godambe and DA Sprott. Holt, Rinehart and Winston, Toronto, Canada; 1971.

Cassel C-M, Sarndal C-E, Wretman JH. Foundations of inference in survey sampling. New York: John Wiley & Sons; 1977.

Biradar R, Singh H. A class of estimators for population parameters using supplementary information. Aligarh J Stat. 1998; 17: 54-71.

Chatterjee S, Hadi AS. Regression analysis by example. New York: John Wiley & Sons; 2013.

Gujarati DN, Porter DC, Gunasekar S. Basic econometrics. New York: McGraw-Hill; 2009.

Kumar A, Saini M. Exponential estimators for finite population mean using auxiliary attributes: A predictive approach. Rev Investig Operacional. 2020 (In Press).

Jhajj HS, Sharma MK, Grover LK. A family of estimators of population mean using information on auxiliary attribute. Pakistan J Stat Ser. 2006; 22(1): 43-50.

Mukhopadhyay P. Theory and methods of survey sampling. New Delhi: Prentice-Hall; 2000.

Damaris Mulwa F, Kilai M. Estimation of Finite Population Mean under Model Based Approach using Auxiliary Variables. Int J Res Eng Sci Manag. 2019; 2(5): 685-688.

Naik VD, Gupta PC. A note on estimation of mean with known population proportion of an auxiliary character. J Ind Soc Agric Stat. 1996; 48(2): 151-158.

Nayak R, Sahoo L. Some alternative predictive estimators of population variance. Rev Colomb Estad. 2012; 35(3): 509-521.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020 [cited 2020 Mar 10]; Available from: http://www.r-project.org/.

Saini M. A class of predictive estimators in two-stage sampling when auxiliary character is estimated at SSU level. Int J Pure Appl Math. 2013; 8 5(2):285-295.

Singh HP, Solanki RS, Singh AK. Predictive estimation of finite population mean using exponential estimators. Stat Stat Econ J. 2014; 94(1): 41-53.

Singh R, Chauhan P, Sawan N, Smarandache F. Ratio-product type exponential estimator for estimating finite population mean using Information on auxiliary attribute, Ann Arbor, USA. 2007 [cited 2020 Mar 10]; Available from: http://fs.unm.edu/Stat/RatioProductTypeExponential.pdf.

Srivastava SK. Predictive estimation of finite population mean using product estimator. Metrika. 1983; 30(1): 93-99.

Sukhatme P V, Sukhatme B V. Sampling Theory of Surveys with Applications. Iowa State Univ Ames, IA, USA. 1970;

Yadav DK, Shukla AK, Tomer S, Kumar B. Predictive estimation of finite population mean using coefficient of kurtosis and median of an auxiliary variable under simple random sampling scheme. Int J Math Arch. 2018; 5046(9):137-143.

Yadav SK, Mishra SS. Developing improved predictive estimator for finite population mean using auxiliary information. Stat Econ J. 2015; 95(1): 76-85.