Revista Colombiana de Estadística - Applied Statistics
January 2024, volume 47, issue 1, pp. 111–133
http://doi.org/10.15446/rce.v47n1.109325

Logarithmic Type Direct and Synthetic Estimators for Domain Mean Using Simple Random Sampling

Estimadores directos y sintéticos de tipo logarítmico para la media del dominio mediante muestreo aleatorio simple

Shashi Bhushan[1,a], Anoop Kumar[2,b], Rohini Pokhrel[3,c]

[1]Department of Statistics, University of Lucknow, Lucknow, India

[2]Department of Statistics, Central University of Haryana, Mahendergarh, India

[3]Department of Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, India

Abstract

In this article, we propose logarithmic type direct and synthetic estimators for the estimation of domain mean under simple random sampling. The properties such as bias and mean square error of the proposed direct and synthetic estimators are obtained up to first order approximation. The efficiency conditions are obtained under which the proposed direct and synthetic estimators outperform their conventional counterparts. The performance of the proposed direct and synthetic estimators is examined with the help of comprehensive computational study using real and artificially drawn populations. Some appropriate suggestions are also provided to the surveyors.

***Key words*:** Direct and synthetic estimators; Mean square error; Small area estimation; Simple random sampling.

Resumen

En este artículo se proponen estimadores directos y sintéticos de tipo logarítmico para la estimación de la media del dominio bajo muestreo aleatorio simple. Las propiedades como sesgo y error cuadrático medio de los estimadores directos y sintéticos propuestos se obtienen hasta aproximación de primer orden. Se obtienen las condiciones de eficiencia bajo las cuales los estimadores directos y sintéticos propuestos superan a sus contrapartes convencionales. El desempeño de los estimadores directos y sintéticos propuestos se examina con la ayuda de un estudio computacional integral que utiliza poblaciones reales y extraídas artificialmente. Algunas sugerencias apropiadas también se proporcionan a los encuestadores.

***Palabras clave*:** Estimación de área pequeña; Estimadores directos y sintéticos; Error cuadrático medio; Muestreo aleatorio simple.

[a]Ph.D. E-mail: shashi.py@gmail.com
[b]Ph.D. E-mail: anoop.asy@gmail.com
[c]Ph.D. E-mail: rohanpokhrel90@gmail.com

1. Introduction

In surveys, a sample of sufficient size is required to provide an accurate estimate of the population parameters. Typically, sample surveys are intended to provide accurate estimates for the population or large domains, and sample sizes are set in accordance with this design. The majority of surveys are designed to provide accurate direct estimates at the national and state levels. Further smaller regions, such as district, sub-district, or block levels are not reliably estimated by these surveys. This is due to the possibility that a sample survey intended for a wide population may only choose a few units, if any, from the small area of interest. The sample size for a given small area may be further reduced by non-response, slow response, etc. Consequently, because of these factors, direct sample estimates cannot be used with sample sizes from small domains.

The National Sample Survey Office (NSSO) surveys are intended to produce data on socio-economic factors at the state and national levels. Due to small sample numbers and a high level of sampling variability, these surveys generate credible direct estimates at the state and national levels but not at the district or block levels. Small area statistics are crucial in the estimation of the parameters of small domains in such circumstances. These small area data are frequently referred to as small area statistics. Small region often refers to a part of the population for whom the sample survey yielded limited information. These subsets can be classified as either a demographic group such as a certain age-sex-race group of individuals within a vast geographic region or a small geographic area such as a municipality, census division, block, tehsil, gram panchayat, etc.

When an estimate of the parameter of concern for a specific sub-population is based solely on sample data from the sub-population itself, it is referred to as a direct estimator. Nevertheless, the sample size for the majority of surveys is insufficient to ensure accurate direct estimates for all sub-populations. Any sub-population for which a direct estimate with the necessary accuracy is unavailable is referred to as a "small area" or "small domain".

Small area estimation (SAE) methods are an ad hoc family of techniques that may be used to circumvent the issue when direct estimates cannot be distributed due to poor quality. These techniques, which are also known as indirect estimators, make up for the weak information in each domain by drawing robustness from sample data from other domains, increasing the useful sample size in each small area. Gonzalez (1973) defined a synthetic estimator as one that uses a trustworthy direct estimator for a larger domain, spanning a number of smaller areas, to generate estimates for the smaller areas while assuming that all smaller areas share the same features as the larger area.

With the use of auxiliary data, a number of direct and synthetic estimators have been presented to estimate the domain mean. A generalised family of synthetic estimators was proposed by Tikkiwal & Ghiya (2000) with an application to the estimate of agricultural area for small domains. Under Lahiri-Midzuno and systematic sampling strategies, the generalized class of synthetic estimators was addressed by Pandey (2007). Among others, Pandey & Tikkiwal (2010) and

Tikkiwal et al. (2013) showed how the addition of supplementary data enhances the efficiency of the estimators. Additionally, Khare & Ashutosh (2017) suggested a generalised synthetic estimator for domain mean utilising bivariate auxiliary information and showed that the estimators outperform under synthetic conditions. The interested reader may also refer the studies of Sisodia & Singh (2001), Sisodia & Chandra (2012), Rai & Pandey (2013), Sharma & Sisodia (2016), Khare et al. (2018), and Bhushan, Kumar & Pokhrel (2023).

The authors in the above-mentioned manuscripts have studied direct estimators and synthetic estimators, separately. But, our objective is different from these authors. In this article, we propose the logarithmic type direct and synthetic estimators simultaneously and compare them with the corresponding conventional direct and synthetic estimators.

Section 2 provides a brief about the set-up of the problem and puts forth the notation used. In Section 3, the review of the existing literature and prominent works for both direct and synthetic methods of estimation are considered. In Section 4, we propose logarithmic type direct and synthetic estimators for estimating the domain mean using auxiliary character and a comparative study is performed theoretically. The performance of the proposed estimators has been evaluated in Section 5 with the help of real data based on Sweden municipality 'MU284' and simulated data. In Section 6, we provide a thorough conclusion of this work.

2. Problem Formulation and Notation

Let us consider that a finite population $\aleph=(\aleph_1, \aleph_2, \ldots, \aleph_N)$ is divided into non overlapping 'A' small areas, i.e., domains $\aleph_a$ of size N_a for which estimates are needed. The domains might be various and could constitute small areas of a sampled population, such as a district, tehsil, or other state-level subdivision, depending on the situation. Let y denote the characteristic under study. Furthermore, assume that the auxiliary information is also available and denoted by x. A simple random sample $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ of size n is chosen without replacement such that n_a, $a = 1, 2, \ldots, A$ units in the sample $\mathbf{s}$ come from the small area 'A'. As a result, $\sum_{a=1}^{A} N_a = N$ and $\sum_{a=1}^{A} n_a = n$. The i^{th} observation of the small domain a of the population for the characteristics x and y are denoted by X_{ai} and Y_{ai}, $a = 1, 2, \ldots, A$ and $i = 1, 2, \ldots, N_a$, respectively.

The notation taken for the population and domain are as follows:
$\bar{X}$: population mean of auxiliary characteristic x based on N observations;
$\bar{X}_a$: the population mean of characteristic x for a small domain based on N_a observations;
$\bar{x}$: sample mean based on n observations on characteristic x;
$\bar{x}_a$: the sample mean based on n_a observations on x;
$\bar{Y}$: the population mean based on N observations on y;
$\bar{Y}_a$: the population mean of domain a based on N_a observations on y;
$\bar{y}$: a sample mean based on n observations on y;
$\bar{y_a}$: a sample mean of domain a based on n_a observations on y.
$C_{x_a} = S_{x_a}/\bar{X}_a$: the population coefficient of variation of auxiliary variable in the

domain a;
$C_x = S_x/\bar{X}$: the population coefficient of variation of auxiliary variable;
$C_{y_a} = S_{y_a}/\bar{Y}_a$: the population coefficient of variation of study variable in the domain a;
$C_y = S_y/\bar{Y}$: the population coefficient of variation of study variable;
$S^2_{x_a} = (N_a - 1)^{-1}\sum_{i=1}^{N_a}(X_{ai} - \bar{X}_a)^2$: the population mean square of the auxiliary variable in the domain a;
${S_x}^2 = (N - 1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$: the population mean square of the auxiliary variable;
$S^2_{y_a} = (N_a - 1)^{-1}\sum_{i=1}^{N_a}(Y_{ai} - \bar{Y}_a)^2$: the population mean square of the study variable in the domain a;
${S_y}^2 = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$: the population mean square of the study variable;
$S_{x_a y_a} = (N_a - 1)^{-1}\sum_{i=1}^{N_a}(X_{ai} - \bar{X}_a)(Y_{ai} - \bar{Y}_a)$: the population covariance between the auxiliary and study variables in domain a;
$S_{yx} = (N - 1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})$: the population covariance between the auxiliary and study variables;
$\rho_{y_a x_a}$: the population correlation coefficient between the study and auxiliary variables in the domain a;
ρ_{yx}: the population correlation coefficient between the study and auxiliary variables;
$f_a = n_a/N_a$: the sampling fraction in domain a;
$f = n/N$: the sampling fraction.

To obtain the properties of the direct estimators, we assume the following notation:
$\bar{y}_a = \bar{Y}_a(1 + \epsilon_1)$, $\bar{x}_a = \bar{X}_a(1 + \epsilon_2)$ such that $E(\epsilon_1) = 0$, $E(\epsilon_2) = 0$, $|\epsilon_i| < 1; i = 1, 2$, $E(\epsilon_1^2) = \lambda_a C^2_{y_a}$, $E(\epsilon_2^2) = \lambda_a C^2_{x_a}$, and $E(\epsilon_1\epsilon_2) = \lambda_a C_{y_a x_a}$, where $\lambda_a = (1 - f_a)/n_a$, and $C_{y_a x_a} = \rho_{y_a x_a} C_{y_a} C_{x_a}$.

Similarly, to obtain the properties of the synthetic estimators, we assume that $\bar{y} = \bar{Y}(1 + \epsilon_3)$, $\bar{x} = \bar{X}(1 + \epsilon_4)$, such that $E(\epsilon_3) = 0$, $E(\epsilon_4) = 0$, where $|\epsilon_i| < 1; i = 3, 4$, $E(\epsilon_3^2) = \lambda C_y^2$, $E(\epsilon_4^2) = \lambda C_x^2$, and $E(\epsilon_3\epsilon_4) = \lambda C_{yx}$, where $C_{yx} = \rho_{yx} C_y C_x$ and $\lambda = (1 - f)/n$.

3. Existing Estimators for Domain Mean

When information on the auxiliary variables is not available, then the usual mean estimator is the obvious choice. Thus, under SAE, the direct and synthetic usual mean estimators are given, respectively, by

$$t^d_{m,a} = \bar{y}_a$$
$$t^s_{m,a} = \bar{y}$$

where the superscripts "d" and "s" in the estimators stand for the "direct" and "synthetic" throughout the article.

To find the bias and MSE of the direct usual mean estimator $t^d_{m,a}$, we utilize the notation provided in Section 2 and rewrite the estimator $t^d_{m,a}$ as

$$\begin{aligned} t^d_{m,a} &= \bar{Y}_a(1+\epsilon_1) \\ t^d_{m,a} - \bar{Y}_a &= \bar{Y}_a\epsilon_1 \end{aligned} \tag{1}$$

Taking expectation both the sides of (1), we get bias of the direct usual mean estimator $t^d_{m,a}$ as

$$Bias(t^d_{m,a}) = 0$$

Squaring and taking expectation both the sides of (1), we get MSE of the direct usual mean estimator $t^d_{m,a}$ as

$$MSE(t^d_{m,a}) = \lambda_a \bar{Y}_a^2 C^2_{y_a}$$

Again, to find the bias and MSE of the synthetic usual mean estimator $t^s_{m,a}$, we utilize the notation provided in Section 2 and rewrite the estimator $t^s_{m,a}$ as

$$\begin{aligned} t^s_{m,a} &= \bar{Y}(1+\epsilon_3) \\ t^s_{m,a} - \bar{Y}_a &= \bar{Y} - \bar{Y}_a + \bar{Y}\epsilon_3 \end{aligned} \tag{2}$$

Taking expectation both the sides of (2), we get bias of the synthetic usual mean estimator $t^s_{m,a}$ as

$$Bias(t^s_{m,a}) = \bar{Y} - \bar{Y}_a$$

Squaring and taking expectation both the sides of (2), we get MSE of the synthetic usual mean estimator $t^s_{m,a}$ as

$$MSE(t^s_{m,a}) = (\bar{Y} - \bar{Y}_a)^2 + \lambda \bar{Y}^2 C^2_y$$

The ratio estimator provides efficient results when the study and auxiliary variables are positively correlated. Under SAE, the direct and synthetic ratio estimators are given, respectively, by

$$t^d_{r,a} = \bar{y}_a\left(\frac{\bar{X}_a}{\bar{x}_a}\right)$$

$$t^s_{r,a} = \bar{y}\left(\frac{\bar{X}_a}{\bar{x}}\right)$$

To find the bias and MSE of the direct ratio estimator $t^d_{r,a}$, we utilize the notation provided in Section 2 and rewrite the estimator $t^d_{r,a}$ as

$$\begin{aligned} t^d_{r,a} &= \bar{Y}_a(1+\epsilon_1)\frac{\bar{X}_a}{\bar{X}_a(1+\epsilon_2)} \\ &= \bar{Y}_a(1+\epsilon_1)(1+\epsilon_2)^{-1} \\ t^d_{r,a} - \bar{Y}_a &= \bar{Y}_a(\epsilon_1 - \epsilon_2 + \epsilon_2^2 - \epsilon_1\epsilon_2) \end{aligned} \tag{3}$$

Taking expectation both the sides of (3), we get bias of the direct ratio estimator $t^d_{r,a}$ as

$$Bias(t^d_{r,a}) = \bar{Y}_a \lambda_a (C^2_{x_a} - C_{y_a x_a})$$

Squaring and taking expectation both the sides of (3), we get MSE of the direct ratio estimator $t^d_{r,a}$ as

$$MSE(t^d_{r,a}) = \bar{Y}^2_a \lambda_a (C^2_{y_a} + C^2_{x_a} - 2C_{y_a x_a})$$

To find the bias and MSE of the synthetic ratio estimator $t^s_{r,a}$, we utilize the notation provided in Section 2 and rewrite the estimator $t^s_{r,a}$ as

$$\begin{aligned} t^s_{r,a} &= \bar{Y}(1+\epsilon_3)\frac{\bar{X}_a}{\bar{X}(1+\epsilon_4)} \\ &= \bar{Y}\frac{\bar{X}_a}{\bar{X}}(1+\epsilon_3)(1+\epsilon_4)^{-1} \\ t^s_{r,a} - \bar{Y}_a &= \left(\bar{Y}\frac{\bar{X}_a}{\bar{X}} - \bar{Y}_a\right) + \bar{Y}\frac{\bar{X}_a}{\bar{X}}(\epsilon_3 - \epsilon_4 + \epsilon_4^2 - \epsilon_3\epsilon_4) \end{aligned} \tag{4}$$

Using synthetic ratio assumption $\bar{Y}_a = \bar{Y}(\bar{X}_a/\bar{X})$ and taking expectation both the sides of (4), we get bias of the synthetic ratio estimator as

$$Bias(t^s_{r,a}) = \bar{Y}_a \lambda (C^2_x - C_{yx})$$

Using synthetic ratio assumption $\bar{Y}_a = \bar{Y}(\bar{X}_a/\bar{X})$, squaring, and taking expectation both the sides of (4), we get MSE of the synthetic ratio estimator as

$$MSE(t^s_{r,a}) = \bar{Y}^2_a \lambda (C^2_x + C^2_y - 2C_{yx})$$

Bahl & Tuteja (1991) developed an exponential relationship between the study and auxiliary variables and proposed exponential ratio estimators for the estimation of population mean. Under SAE, following the work of Bahl & Tuteja (1991), we suggest the exponential ratio type direct and synthetic estimators, respectively, as

$$\begin{aligned} t^d_{bt,a} &= \bar{y}_a \exp\left(\frac{\bar{X}_a - \bar{x}_a}{\bar{X}_a + \bar{x}_a}\right) \\ t^s_{bt,a} &= \bar{y}\ \exp\left(\frac{\bar{X}_a - \bar{x}}{\bar{X}_a + \bar{x}}\right) \end{aligned}$$

To find the bias and MSE of the exponential ratio type direct estimator $t^d_{bt,a}$, we utilize the notation provided in Section 2 and rewrite the estimator $t^d_{bt,a}$ as

$$\begin{aligned} t^d_{bt,a} &= \bar{Y}_a(1+\epsilon_1)\exp\left\{\frac{\bar{X}_a - \bar{X}_a(1+\epsilon_2)}{\bar{X}_a + \bar{X}_a(1+\epsilon_2)}\right\} \\ &= \bar{Y}_a(1+\epsilon_1)\exp\left\{-\frac{\epsilon_2}{2}\left(1+\frac{\epsilon_2}{2}\right)^{-1}\right\} \\ t^d_{bt,a} - \bar{Y}_a &= \bar{Y}_a(\epsilon_1 - \frac{\epsilon_2}{2} + \frac{3}{8}\epsilon_2^2 - \frac{1}{2}\epsilon_1\epsilon_2) \end{aligned} \tag{5}$$

Taking expectation both the sides of (5), we get bias of the estimator $t^d_{bt,a}$ as

$$Bias(t^d_{bt,a}) = \bar{Y}_a \lambda_a \left(\frac{3}{8} C^2_{x_a} - \frac{1}{2} \rho_{y_a x_a} C_{y_a} C_{x_a} \right)$$

Squaring and taking expectation both the sides of (5), we get MSE of the estimator $t^d_{bt,a}$ as

$$MSE(t^d_{bt,a}) = \bar{Y}^2_a \lambda_a \left(C^2_{y_a} + \frac{1}{4} C^2_{x_a} - C_{y_a x_a} \right)$$

Again, to find the bias and MSE of the exponential ratio type synthetic estimator $t^s_{bt,a}$, we utilize the notation provided in Section 2 and rewrite the estimator $t^s_{bt,a}$ as

$$\begin{aligned} t^s_{bt,a} &= \bar{Y}(1+\epsilon_3) \exp\left\{ \frac{\bar{X}_a - \bar{X}(1+\epsilon_4)}{\bar{X}_a + \bar{X}(1+\epsilon_4)} \right\} \\ &= \bar{Y}(1+\epsilon_3) \exp\left\{ -\frac{\epsilon_4}{2} \left(1 + \frac{\epsilon_4}{2}\right)^{-1} \right\} \\ t^s_{bt,a} - \bar{Y}_a &= \bar{Y} K_{1a} - \bar{Y}_a + \bar{Y}(K_{1a}\epsilon_3 - K_{2a}\epsilon_4 + K_{3a}\epsilon^2_3 - K_{2a}\epsilon_3\epsilon_4) \end{aligned} \tag{6}$$

where $K_{1a} = 1 + A_a + A^2_a/2 + A^3_a/6$, $K_{2a} = \left(2A_a + 2A^2_a + A^3_a \bar{X}\bar{X}_a\right) / \left(\bar{X}^2_a - \bar{X}^2\right)$ $K_{3a} = \left[2\{(\bar{X}_a - \bar{X})A_a + (2\bar{X}_a - \bar{X})A^2_a + A^3_a\}\bar{X}^2\bar{X}_a/(\bar{X}^2_a - \bar{X}^2)^2\right]$, and $A_a = \left(\bar{X} - \bar{X}_a\right)/\left(\bar{X} + \bar{X}_a\right)$. Taking expectation both the sides of (6), we get bias of the estimator $t^s_{bt,a}$ as

$$Bias(t^s_{bt,a}) = K_{1a}\bar{Y} - \bar{Y}_a + \bar{Y}\lambda \left(K^2_{3a} C^2_x - K_{2a} C_{yx}\right)$$

Squaring and taking expectation both the sides of (6), we get MSE of the estimator $t^s_{bt,a}$ as

$$MSE(t^s_{bt,a}) = \left\{ \begin{array}{l} (K_{1a}\bar{Y} - \bar{Y}_a)^2 + \bar{Y}^2 \lambda (K^2_{1a} C^2_y + K^2_{2a} C^2_x - 2K_{1a}K_{2a}C_{yx}) \\ +2(\bar{Y}K_{1a} - \bar{Y}_a)\lambda\bar{Y}\left(K_{3a}C^2_x - K_{2a}C_{yx}\right) \end{array} \right\}$$

4. Proposed Logarithmic Type Direct and Synthetic Estimators

The objectives of this article are as follows:

(i) To achieve efficient estimates of domain mean $\bar{Y}_a$ via both direct and synthetic estimation methods.

(ii) To effectively use the information on the auxiliary variable to obtain efficient estimates.

Taking the above objectives under consideration, following Bhushan & Kumar (2020, 2022a), we propose some logarithmic type direct and synthetic estimators for domain mean $\bar{Y}_a$ under SRS as

$$T_a^d = \bar{y}_a \left\{ 1 + \alpha \log\left(\frac{\bar{x}_a}{\bar{X}_a}\right) \right\}$$
$$T_a^s = \bar{y} \left\{ 1 + \beta \log\left(\frac{\bar{x}}{\bar{X}_a}\right) \right\}$$

where α and β are suitably chosen constants.

4.1. Bias and Mse of Proposed Estimators

The results of the bias and MSE for the proposed estimators are given in the following theorems.

Theorem 1. *The bias, MSE, and minimum MSE of the proposed direct estimator T_a^d up to 1^{st} order approximation are, respectively, given below.*

$$\begin{aligned} Bias(T_a^d) &= \lambda_a \bar{Y}_a \alpha \left(\rho_{y_a x_a} C_{y_a} C_{x_a} - \frac{C_{x_a}^2}{2} \right) \\ MSE(T_a^d) &= \bar{Y}_a^2 \lambda_a \left(C_{y,a}^2 + \alpha^2 C_{x,a}^2 + 2\alpha \rho_{y_a x_a} C_{y_a} C_{x_a} \right) \\ minMSE(T_a^d) &= \bar{Y}_a^2 \lambda_a C_{y_a}^2 \left(1 - \rho_{y_a x_a}^2 \right) \end{aligned} \quad (7)$$

Proof. To find the bias, MSE, and minimum MSE of the proposed direct estimator T_a^d, we utilize the notation provided in Section 2 and rewrite the estimator T_a^d as

$$\begin{aligned} T_a^d &= \bar{y}_a \left\{ 1 + \alpha \log\left(\frac{\bar{x}_a}{\bar{X}_a}\right) \right\} \\ &= \bar{Y}_a (1+\epsilon_1) \left\{ 1 + \alpha \log\left(\frac{\bar{X}_a(1+\epsilon_2)}{\bar{X}_a}\right) \right\} \\ &= \bar{Y}_a (1+\epsilon_1) \left\{ 1 + \alpha \log(1+\epsilon_2) \right\} \\ &= \bar{Y}_a (1+\epsilon_1) \left\{ 1 + \alpha \left(\epsilon_2 - \frac{\epsilon_2^2}{2} + \ldots \right) \right\} \end{aligned}$$

After simplifying and substracting $\bar{Y}_a$ on both sides of the above equation, we get

$$T_a^d - \bar{Y}_a = \left\{ \alpha \left(\epsilon_1 - \frac{\epsilon_1^2}{2} \right) + \epsilon_1 + \alpha \epsilon_1 \epsilon_2 \right\} \quad (8)$$

Taking expectation on both sides, we get

$$Bias(T_a^d) = \lambda_a \bar{Y}_a \alpha \left(\rho_{y_a x_a} C_{y_a} C_{x_a} - \frac{C_{x_a}^2}{2} \right)$$

Squaring and taking expectation on both sides of (8), we get

$$MSE(T_a^d) = \bar{Y}_a^2 \lambda_a \left(C_{y,a}^2 + \alpha^2 C_{x,a}^2 + 2\alpha \rho_{y_a x_a} C_{y_a} C_{x_a} \right) \tag{9}$$

Differentiating (9) partially w.r.t. parameter α and equating to zero, we get

$$\alpha_{(opt)} = -\rho_{y_a x_a} \left(\frac{C_{y_a}}{C_{x_a}} \right)$$

Putting the value of $\alpha_{(opt)}$ in (9), we get

$$minMSE(T_a^d) = \bar{Y}_a^2 \lambda_a C_{y_a}^2 \left(1 - \rho_{y_a x_a}^2 \right)$$

□

Theorem 2. *The bias, MSE, and minimum MSE of the proposed synthetic estimator T_a^s up to 1^{st} order approximation under the synthetic assumption $\bar{Y}_a = \bar{Y}\{1 + \beta \log(\bar{X}/\bar{X}_a)\}$ are, respectively, given below.*

$$\begin{aligned} Bias(T_a^s) &= \bar{Y}\beta f \left(\rho_{yx} C_y C_x - \frac{C_x^2}{2} \right) \\ MSE(T_a^s) &= \lambda \left(\bar{Y}_a^2 C_y^2 + \bar{Y}^2 \beta^2 C_x^2 + 2\bar{Y}\bar{Y}_a \beta \rho_{yx} C_y C_x \right) \\ minMSE(T_a^s) &= \bar{Y}_a^2 \lambda C_y^2 \left(1 - \rho_{yx}^2 \right) \end{aligned} \tag{10}$$

Proof. To obtain the bias, MSE and minimum MSE of the proposed synthetic estimator T_a^s, we utilize the notation provided in Section 2 and rewrite the estimator T_a^s as

$$\begin{aligned} T_a^s &= \bar{y} \left\{ 1 + \beta \log \left(\frac{\bar{x}}{\bar{X}_a} \right) \right\} \\ &= \bar{Y}(1+\epsilon_3) \left\{ 1 + \beta \log \left(\frac{\bar{X}(1+\epsilon_4)}{\bar{X}_a} \right) \right\} \\ &= \bar{Y}(1+\epsilon_3) \left\{ 1 + \beta \log \left(\frac{\bar{X}}{\bar{X}_a} \right) + \beta \log(1+\epsilon_4) \right\} \\ &= \bar{Y}(1+\epsilon_3) \left\{ 1 + \beta B_a + \beta \left(\epsilon_4 - \frac{\epsilon_4^2}{2} \right) \right\} \end{aligned}$$

where $B_a = \log(\bar{X}/\bar{X}_a)$. Subtracting $\bar{Y}_a$ on both sides of the above equation, we get

$$T_a^s - \bar{Y}_a = \bar{Y}(1+\epsilon_3) \left\{ 1 + \beta B_a + \beta \left(\epsilon_4 - \frac{\epsilon_4^2}{2} \right) \right\} - \bar{Y}_a \tag{11}$$

Taking expectation on both sides to (11), we get

$$B(T_a^s) = (\bar{Y} - \bar{Y}_a) + \bar{Y}\beta B_a + \bar{Y}\beta f \left(\rho_{yx} C_y C_x - \frac{C_x^2}{2} \right)$$

Under the synthetic assumption $\bar{Y}_a/\bar{Y} = 1 + \beta \log(\bar{X}/\bar{X}_a)$, the bias becomes

$$Bias(T_a^s) = \bar{Y}\beta f \left(\rho_{yx} C_y C_x - \frac{C_x^2}{2}\right)$$

Squaring and taking the expectation on both sides to (11), we get

$$MSE(T_a^s) = \left[\begin{array}{l} \bar{Y}^2 \left\{ \begin{array}{l} \beta\left\{B_a^2 + (1-B_a)\lambda C_x^2 + B_a^2 \lambda C_y^2 + 4B_a f \rho_{yx} C_y C_x\right\} \\ +\lambda C_y^2 + 2\beta f \rho_{yx} C_y C_x (1+B_a) \end{array} \right\} \\ +(\bar{Y}-\bar{Y}_a)^2 + 2\beta \bar{Y}(\bar{Y}-\bar{Y}_a)\left(B_a - f\frac{C_x^2}{2} + f\rho_{yx} C_y C_x\right) \end{array} \right]$$

Under the synthetic assumption $\bar{Y}_a/\bar{Y} = 1 + \beta \log(\bar{X}/\bar{X}_a)$, the above MSE expression reduces as

$$MSE(T_a^s) = \lambda\left(\bar{Y}_a^2 C_y^2 + \bar{Y}^2 \beta^2 C_x^2 + 2\bar{Y}\bar{Y}_a \beta \rho_{yx} C_y C_x\right) \tag{12}$$

By differentiating (12) partially w.r.t. β and equating to zero, we get

$$\beta_{(opt)} = -\rho_{xy} \frac{\bar{Y}_a C_y}{\bar{Y} C_x}$$

Putting the value of $\beta_{(opt)}$ in (12), we get

$$minMSE(T_a^s) = \bar{Y}_a^2 \lambda C_y^2 \left(1 - \rho_{yx}^2\right)$$

□

Corollary 1. *The proposed synthetic estimator T_a^s outperforms the proposed direct estimator T_a^d iff*

$$\frac{(1-\rho_{yx}^2)}{(1-\rho_{y_a x_a}^2)} < \frac{\lambda_a C_{y_a}^2}{\lambda C_y^2} \tag{13}$$

and contrariwise. Otherwise, both are equally efficient when the equality holds in (13).

***Proof*.** By comparing the minimum MSEs of the proposed synthetic and direct estimators from (7) and (10), we obtain (13). □

Furthermore, we compare the minimum MSE of the proposed direct and synthetic estimators with the minimum MSE of the existing direct and synthetic estimators as follows:

$$MSE(T_a^d) < MSE(t_{m,a}^d) \implies (1 - \rho_{y_a x_a}^2) < 1$$

$$MSE(T_a^s) < MSE(t_{m,a}^s) \implies \rho_{yx}^2 > \left(1 - \frac{\bar{Y}^2}{\bar{Y}_a^2}\right) - \frac{1}{\lambda C_y^2}\left(\frac{\bar{Y}}{\bar{Y}_a} - 1\right)^2$$

$$MSE(T_a^d) < MSE(t_{r,a}^d) \implies \rho_{y_a x_a}^2 > 2\rho_{y_a x_a}\left(\frac{C_{x_a}}{C_{y_a}}\right) - \left(\frac{C_{x_a}}{C_{y_a}}\right)^2$$

$$MSE(T_a^s) < MSE(t_{r,a}^s) \implies \rho_{yx}^2 > 2\rho_{yx}\left(\frac{C_x}{C_y}\right) - \left(\frac{C_x}{C_y}\right)^2$$

$$MSE(T_a^d) < MSE(t_{bt,a}^d) \implies \rho_{y_a x_a}^2 > \rho_{y_a x_a}\left(\frac{C_{x_a}}{C_{y_a}}\right) - \frac{1}{4}\left(\frac{C_{x_a}}{C_{y_a}}\right)^2$$

$$MSE(T_a^s) < MSE(t_{bt,a}^s) \implies \rho_{yx}^2 > 1 - \frac{1}{\bar{Y}_a^2 \lambda C_y^2}\left\{\begin{array}{l} \bar{Y}^2\lambda\left(\begin{array}{c} K_{1a}^2 C_y^2 + K_{2a}^2 C_x^2 \\ -2K_{1a}K_{2a}C_{yx} \end{array}\right) \\ +2(\bar{Y}K_{1a} - \bar{Y}_a)\lambda\bar{Y} \\ \times\left(\begin{array}{c} K_{3a}C_x^2 \\ -K_{2a}C_{yx} \end{array}\right) \\ +(K_{1a}\bar{Y} - \bar{Y}_a)^2 \end{array}\right\}$$

The proposed direct and synthetic estimators outperform their conventional counterparts under the above efficiency conditions.

5. Numerical and Simulation Studies

In this section, we consider numerical and simulation studies based on real and artificially generated populations, respectively, and discussed the results.

5.1. Numerical Study

To perform a numerical study, we consider a real data of Sweden municipalities reported in Särndal et al. (2003) that is divided into 284 municipalities referred as MU284. It varies considerably in size and other characteristics. The data set consists of a few variables that explain the municipalities in different ways. This data set consists of eight variables out of which we select only two variables, namely, REV84 and P85. For numerical study, we consider eight domains (regions) as small areas. The study and auxiliary variables defined in this data set are as follows.

Y: REV84 = Real state values according to 1984 assessment (in million of Kronar) and X: P85 = 1985 population (in thousand).

The domain parameters are given in Table 1 for ready reference.

Table 1: Population parameters for different domains

Domains	N_a	$\bar{Y}_a$	$\bar{X}_a$	$S^2_{y,a}$	$S^2_{x,a}$	$\rho_{yx,a}$
1	25	6413.320	59.520	128075863	16564.760	0.993
2	48	2971.104	29.167	11119986	1228.227	0.997
3	32	2498.750	23.938	4164522	437.157	0.949
4	38	2915.526	30.631	9575690	1721.158	0.982
5	56	3046.464	28.714	27860176	3565.117	0.979
6	41	2175.317	20.976	2869024	300.924	0.975
7	15	3648.467	26.600	5810807	581.543	0.837
8	29	2269.103	17.138	7755964	405.909	0.807

Using the domain parameters given in Table 1, we have computed MSE and percent relative efficiency (PRE) of the proposed logarithmic type direct and synthetic estimators with respect to the direct and synthetic mean estimators, respectively, by using the following formulae:

$$PRE^d = \frac{MSE(t^d_{m,a})}{MSE(T^*)} \times 100$$

$$PRE^s = \frac{MSE(t^s_{m,a})}{MSE(T^{**})} \times 100$$

where T^*=$t^d_{m,a}$, $t^d_{r,a}$, $t^d_{bt,a}$, T^d_a, and T^{**}= $t^s_{m,a}$, $t^s_{r,a}$, $t^s_{bt,a}$, and T^s_a.

The results of the numerical study for direct and synthetic estimators are reported in Tables 2-3, respectively.

Table 2: MSE and PRE of direct estimators using real population

Estimators	$t^d_{m,a}$		$t^d_{r,a}$		$t^d_{bt,a}$		$\mathbf{T^d_a}$	
Domains	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	20492138	100	1374840	1491	3240677	632	**271419**	**7550**
2	1779198	100	67037	2654	221439	804	**57549**	**3092**
3	666324	100	64041	1041	152729	436	**55875**	**1193**
4	1532110	100	115286	1329	145444	1053	**33603**	**4559**
5	4457628	100	185286	2406	377368	1181	**85518**	**5212**
6	459044	100	16160	2841	71000	647	**13992**	**3281**
7	929729	100	906942	103	498875	186	**463490**	**201**
8	1240954	100	380026	327	467473	266	**356832**	**348**

Table 3: MSE and PRE of synthetic estimators using real population

Estimators	$t^s_{m,a}$		$t^s_{r,a}$		$t^s_{bt,a}$		$\mathbf{T^s_a}$	
Domains	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	11443323	100	132706	8623	4508276	254	**68936**	**16600**
2	327118	100	28481	1149	76519	428	**14795**	**2211**
3	650772	100	20145	3230	145233	448	**10465**	**6219**
4	342036	100	27426	1247	128543	266	**14247**	**2401**
5	316757	100	29944	1058	61557	515	**315555**	**2036**
6	1129771	100	15268	7400	253017	447	**7931**	**14245**
7	641767	100	42948	1494	539507	119	**22310**	**2877**
8	969337	100	16612	5835	56291	1722	**8630**	**11233**

5.2. Simulation Study

To extrapolate the results of the numerical study, we conduct a simulation study by following Singh & Horn (1998), Bhushan & Kumar (2022*b*), Bhushan et al. (2022), Bhushan, Kumar, Zaman & Al Mutairi (2023), and Bhushan, Kumar, Lone, Anwar & Gunaime (2023). In the process of simulation study, we hypothetically draw some symmetric and asymmetric populations using the following models:

$$y = 7.8 + \sqrt{(1-\rho_{xy}^2)}\; y^* + \rho_{xy}\left(\frac{S_y}{S_x}\right)x^*$$
$$x = 7.2 + x^*$$

where x^* and y^* are independent variables for the corresponding distributions. Using the models mentioned above, we generated the populations shown below:

(1) A Normal population of size $N = 6000$ using $x^* \sim N(20, 35)$ and $y^* \sim N(15, 20)$ with varying correlation coefficients $\rho_{xy} = 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$.

(2) An exponential population of size $N = 6000$ using $x^* \sim \exp(0.2)$ and $y^* \sim \exp(0.5)$ with varying correlation coefficients $\rho_{xy} = 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$.

The above populations are divided into 6 equal domains of size 1000. A random sample of size 36 is being selected from each domain and the descriptive statistics are computed. Performing 10 000 iterations, we have computed MSE and percent relative efficiency (PRE) of the proposed logarithmic type direct and synthetic estimators with respect to the direct and synthetic mean estimators, respectively, by using the following formulae:

$$MSE(T^*) = \frac{1}{10\,000}\sum_{s=1}^{10\,000}(T^* - \bar{Y}_a)^2$$
$$PRE^d = \frac{MSE(t_{m,a}^d)}{MSE(T^*)} \times 100$$
$$PRE^s = \frac{MSE(t_{m,a}^s)}{MSE(T^*)} \times 100$$

where $T^* = t_{m,a}^d,\ t_{r,a}^d,\ t_{bt,a}^d,\ T_a^d,\ t_{m,a}^s,\ t_{r,a}^s,\ t_{bt,a}^s$, and T_a^s.

The simulation results of the direct estimators for normal and exponential populations are reported in Tables 4-5, respectively, whereas the simulation results of the synthetic estimators for normal and exponential populations are reported in Tables 6-7, respectively.

5.3. Discussion of Numerical and Simulation Results

Following a thorough examination of the outcomes of the numerical and simulation studies, we outline the discussion of the results in point-wise style.

Table 4: MSE and PRE of direct estimators for artificially generated normal population

Estimators		$t^d_{m,a}$		$t^d_{r,a}$		$t^d_{bt,a}$		$\mathbf{T^d_a}$	
Domains	$\rho_{y_a x_a}$	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	0.1	32.39	100	45.79	70.74	34.97	92.63	**31.36**	**103.30**
	0.3	32.39	100	34.78	93.14	30.18	107.33	**29.12**	**111.23**
	0.5	32.39	100	25.75	125.80	26.05	124.35	**24.16**	**134.06**
	0.7	32.39	100	17.52	184.87	22.04	146.94	**16.44**	**197.05**
	0.9	32.39	100	8.69	372.79	17.26	187.66	**6.05**	**535.56**
2	0.1	32.16	100	45.63	70.47	34.75	92.53	**31.13**	**103.3**
	0.3	32.16	100	34.63	92.85	29.97	107.28	**28.92**	**111.19**
	0.5	32.16	100	25.62	125.53	25.85	124.37	**24.00**	**133.96**
	0.7	32.16	100	17.41	184.74	21.87	147.06	**16.34**	**196.79**
	0.9	32.16	100	8.61	373.5	17.11	187.92	**6.02**	**534.00**
3	0.1	32.21	100	45.94	70.12	34.91	92.27	**31.18**	**103.30**
	0.3	32.21	100	34.85	92.43	30.10	107.03	**29.02**	**110.99**
	0.5	32.21	100	25.77	125.02	25.95	124.14	**24.14**	**133.44**
	0.7	32.21	100	17.50	184.12	21.93	146.87	**16.46**	**195.70**
	0.9	32.21	100	8.64	372.89	17.15	187.85	**6.06**	**531.15**
4	0.1	32.27	100	45.78	70.49	34.88	92.53	**31.23**	**103.33**
	0.3	32.27	100	34.74	92.90	30.08	107.3	**29.01**	**111.24**
	0.5	32.27	100	25.69	125.62	25.94	124.4	**24.08**	**134.04**
	0.7	32.27	100	17.46	184.88	21.94	147.08	**16.38**	**196.99**
	0.9	32.27	100	8.64	373.73	17.17	187.95	**6.03**	**535.05**
5	0.1	32.29	100	45.92	70.33	34.93	92.46	**31.28**	**103.25**
	0.3	32.29	100	34.81	92.78	30.10	107.27	**29.05**	**111.16**
	0.5	32.29	100	25.72	125.55	25.96	124.42	**24.11**	**133.96**
	0.7	32.29	100	17.46	184.90	21.95	147.14	**16.40**	**196.86**
	0.9	32.29	100	8.64	373.96	17.17	188.04	**6.04**	**534.46**
6	0.1	32.29	100	46.08	70.06	35.00	92.24	**31.25**	**103.33**
	0.3	32.29	100	34.94	92.40	30.17	107.02	**29.07**	**111.05**
	0.5	32.29	100	25.82	125.02	26.01	124.14	**24.17**	**133.59**
	0.7	32.29	100	17.53	184.16	21.98	146.87	**16.47**	**196.03**
	0.9	32.29	100	8.66	372.97	17.19	187.85	**6.07**	**532.12**

TABLE 5: MSE and PRE of direct estimators for artificially generated exponential population

Estimators		$t^d_{m,a}$		$t^d_{r,a}$		$t^d_{bt,a}$		$\mathbf{T^d_a}$	
Domains	$\rho_{y_a x_a}$	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	0.1	0.67	100	0.78	85.67	0.69	97.88	**0.65**	**102.82**
	0.3	0.67	100	0.64	104.42	0.62	108.56	**0.59**	**113.06**
	0.5	0.67	100	0.51	131.25	0.55	121.06	**0.48**	**139.80**
	0.7	0.67	100	0.38	174.63	0.49	136.50	**0.32**	**206.97**
	0.9	0.67	100	0.25	265.12	0.43	157.90	**0.12**	**538.26**
2	0.1	0.68	100	0.78	86.17	0.69	98.09	**0.66**	**102.9**
	0.3	0.68	100	0.64	105.10	0.62	108.81	**0.59**	**113.52**
	0.5	0.68	100	0.51	132.15	0.56	121.34	**0.48**	**140.98**
	0.7	0.68	100	0.38	175.80	0.49	136.78	**0.32**	**209.81**
	0.9	0.68	100	0.25	266.47	0.43	158.13	**0.12**	**549.24**
3	0.1	0.68	100	0.79	85.74	0.69	97.86	**0.66**	**102.82**
	0.3	0.68	100	0.65	104.58	0.62	108.57	**0.60**	**113.08**
	0.5	0.68	100	0.51	131.51	0.56	121.10	**0.48**	**140.09**
	0.7	0.68	100	0.39	175.03	0.49	136.56	**0.32**	**208.23**
	0.9	0.68	100	0.25	265.72	0.43	157.99	**0.12**	**544.46**
4	0.1	0.67	100	0.78	85.88	0.69	97.96	**0.65**	**102.89**
	0.3	0.67	100	0.64	104.72	0.62	108.66	**0.59**	**113.24**
	0.5	0.67	100	0.51	131.67	0.56	121.18	**0.48**	**140.22**
	0.7	0.67	100	0.38	175.20	0.49	136.63	**0.32**	**208.11**
	0.9	0.67	100	0.25	265.85	0.43	158.03	**0.12**	**543.68**
5	0.1	0.68	100	0.79	86.15	0.69	98.06	**0.66**	**102.8**
	0.3	0.68	100	0.64	105.08	0.62	108.78	**0.60**	**113.28**
	0.5	0.68	100	0.51	132.13	0.56	121.30	**0.48**	**140.67**
	0.7	0.68	100	0.39	175.77	0.50	136.75	**0.32**	**209.57**
	0.9	0.68	100	0.25	266.41	0.43	158.10	**0.12**	**549.95**
6	0.1	0.67	100	0.78	86.06	0.69	98.03	**0.65**	**102.85**
	0.3	0.67	100	0.64	104.96	0.62	108.74	**0.59**	**113.41**
	0.5	0.67	100	0.51	131.95	0.55	121.26	**0.48**	**140.80**
	0.7	0.67	100	0.38	175.53	0.49	136.70	**0.32**	**209.37**
	0.9	0.67	100	0.25	266.10	0.43	158.05	**0.12**	**547.31**

Table 6: MSE and PRE of synthetic estimators for artificially generated normal population

Estimators		$t^s_{m,a}$		$t^s_{r,a}$		$t^s_{bt,a}$		$\mathbf{T^s_a}$	
Domains	ρ_{yx}	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	0.1	33.36	100	7.24	460.65	38.02	87.75	**5.37**	**621.36**
	0.3	33.36	100	5.55	601.41	41.05	81.26	**4.94**	**675.98**
	0.5	33.36	100	4.14	806.03	44.12	75.61	**4.07**	**820.19**
	0.7	33.36	100	2.84	1175.18	47.62	70.06	**2.77**	**1206.16**
	0.9	33.36	100	1.41	2362.55	52.85	63.13	**1.03**	**3237.60**
2	0.1	32.82	100	7.24	453.13	38.08	86.19	**5.37**	**610.79**
	0.3	32.82	100	5.55	591.40	41.01	80.01	**4.94**	**664.49**
	0.5	32.82	100	4.14	792.40	43.98	74.62	**4.07**	**806.24**
	0.7	32.82	100	2.84	1155.05	47.35	69.30	**2.77**	**1185.65**
	0.9	32.82	100	1.41	2321.62	52.36	62.68	**1.03**	**3182.54**
3	0.1	32.15	100	7.28	441.44	37.18	86.46	**5.40**	**594.83**
	0.3	32.15	100	5.58	576.07	40.01	80.36	**4.97**	**647.12**
	0.5	32.15	100	4.17	771.76	42.86	75.02	**4.09**	**785.18**
	0.7	32.15	100	2.86	1124.87	46.09	69.75	**2.78**	**1154.67**
	0.9	32.15	100	1.42	2260.79	50.90	63.17	**1.04**	**3099.38**
4	0.1	32.39	100	7.26	445.99	37.19	87.09	**5.38**	**601.43**
	0.3	32.39	100	5.56	582.17	40.10	80.76	**4.95**	**654.30**
	0.5	32.39	100	4.15	780.18	43.04	75.25	**4.08**	**793.89**
	0.7	32.39	100	2.85	1137.55	46.37	69.84	**2.77**	**1167.48**
	0.9	32.39	100	1.42	2287.66	51.34	63.09	**1.03**	**3133.76**
5	0.1	34.17	100	7.28	469.44	39.15	87.27	**5.40**	**632.97**
	0.3	34.17	100	5.58	612.78	42.27	80.83	**4.96**	**688.62**
	0.5	34.17	100	4.16	821.13	45.42	75.22	**4.09**	**835.53**
	0.7	32.39	100	2.85	1137.55	46.37	69.84	**2.77**	**1167.48**
	0.9	34.17	100	1.42	2406.12	54.35	62.86	**1.04**	**3298.13**
6	0.1	32.50	100	7.30	445.29	37.58	86.48	**5.41**	**600.38**
	0.3	32.50	100	5.59	581.22	40.45	80.34	**4.98**	**653.16**
	0.5	32.50	100	4.17	778.85	43.35	74.97	**4.10**	**792.50**
	0.7	32.50	100	2.85	1197.02	49.00	69.72	**2.78**	**1228.72**
	0.9	32.50	100	1.42	2282.89	51.53	63.06	**1.04**	**3128.29**

Table 7: MSE and PRE of synthetic estimators for artificially generated exponential population

Estimators		$t^s_{m,a}$		$t^s_{r,a}$		$t^s_{bt,a}$		$\mathbf{T^s_a}$	
Domains	ρ_{yx}	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	0.1	0.68	100	0.13	533.91	0.73	92.66	**0.11**	**619.56**
	0.3	0.68	100	0.10	649.85	0.78	86.55	**0.10**	**674.02**
	0.5	0.68	100	0.08	814.77	0.83	81.32	**0.08**	**817.82**
	0.7	0.68	100	0.06	1079.20	0.88	76.58	**0.06**	**1202.67**
	0.9	0.68	100	0.04	1622.62	0.94	71.81	**0.02**	3**228.22**
2	0.1	0.70	100	0.13	551.82	0.76	92.25	**0.11**	**640.39**
	0.3	0.70	100	0.10	671.67	0.81	86.15	**0.10**	**696.69**
	0.5	0.70	100	0.08	842.17	0.86	80.93	**0.08**	**845.32**
	0.7	0.7	100	0.06	1115.57	0.92	76.23	**0.06**	**1243.12**
	0.9	0.7	100	0.04	1677.39	0.98	71.54	**0.02**	**3336.78**
3	0.1	0.70	100	0.13	554.76	0.76	92.16	**0.11**	**643.74**
	0.3	0.70	100	0.10	675.23	0.82	86.08	**0.10**	**700.34**
	0.5	0.7	100	0.08	846.58	0.87	80.89	**0.08**	**849.74**
	0.7	0.7	100	0.06	1121.35	0.92	76.20	**0.06**	**1249.62**
	0.9	0.7	100	0.04	1686.13	0.98	71.53	**0.02**	**3354.24**
4	0.1	0.68	100	0.13	537.98	0.74	91.53	**0.11**	**624.32**
	0.3	0.68	100	0.10	654.83	0.80	85.58	**0.10**	**679.20**
	0.5	0.68	100	0.08	821.04	0.85	80.51	**0.08**	**824.10**
	0.7	0.68	100	0.06	1087.53	0.9	75.95	**0.06**	**1211.91**
	0.9	0.68	100	0.04	1635.19	0.95	71.43	**0.02**	**3253.03**
5	0.1	0.69	100	0.13	545.43	0.75	92.08	**0.11**	**632.94**
	0.3	0.69	100	0.10	663.88	0.80	86.02	**0.10**	**688.59**
	0.5	0.69	100	0.08	832.38	0.85	80.85	**0.08**	**835.49**
	0.7	0.69	100	0.06	1102.57	0.91	76.19	**0.06**	**1228.65**
	0.9	0.69	100	0.04	1657.84	0.97	71.55	**0.02**	**3297.97**
6	0.1	0.69	100	0.13	543.24	0.75	92.11	**0.11**	**630.39**
	0.3	0.69	100	0.10	661.21	0.80	86.06	**0.10**	**685.81**
	0.5	0.69	100	0.08	829.01	0.85	80.89	**0.08**	**832.12**
	0.7	0.69	100	0.06	1098.06	0.90	76.23	**0.06**	**1223.70**
	0.9	0.69	100	0.04	1657.84	0.97	71.55	**0.02**	**3297.97**

(i) From the results of Table 2 based on the real population, the proposed logarithmic type direct estimator T_a^d outperforms the direct mean estimator $t_{m,a}^d$, direct ratio estimator $t_{r,a}^d$, direct exponential ratio estimator $t_{bt,a}^d$ envisaged on the lines of Bahl and Tuteja (1991) in each domain by minimum MSE and maximum PRE. The dominance of the proposed direct estimators over the existing direct estimators can easily be seen from Figure 1.

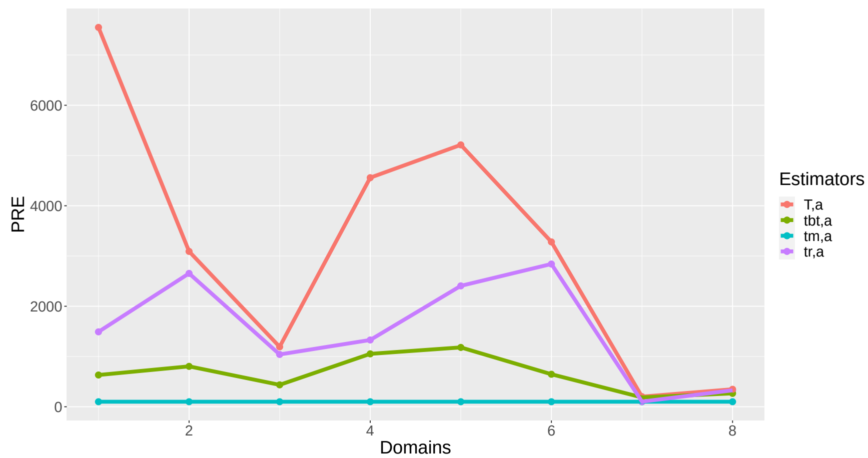

Figure 1: Graph of PRE of direct estimators for real population

(ii) From the results of Table 3 based on the real population, the proposed logarithmic type synthetic estimator T_a^s outperforms the synthetic mean estimator $t_{m,a}^s$, synthetic ratio estimator $t_{r,a}^s$, synthetic exponential ratio estimator $t_{bt,a}^s$ envisaged on the lines of Bahl & Tuteja (1991) in each domain by minimum MSE and maximum PRE. The dominance of the proposed synthetic estimators over the existing synthetic estimators can easily be seen from Figure 2.

(iii) From the results of Table 4 based on artificially generated normal population, the proposed logarithmic type direct estimator T_a^d dominate the existing direct estimators for each value of the correlation coefficient in each domain by minimum MSE and maximum PRE. The PRE values of domain 1 are presented through the line diagram in Figure 3. The line diagram for the PRE values of other domains can be provided, if required.

(iv) The similar tendency as observed from the results of Table 4 can also be observed from the results of Table 5 based on an artificially generated exponential population. This tendency can also be observed from the line diagram given in Figure 4 for the PRE values of domain 1.

(v) From the results of Table 6 consisting of the artificially generated normal population, the proposed logarithmic type synthetic estimator T_a^s surpass the existing synthetic estimators for each value of correlation coefficient in each domain by minimum MSE and maximum PRE. The PRE values of domain 1 are presented through the line diagram in Figure 5. The line diagram for the PRE values of other domains can be provided, if required.

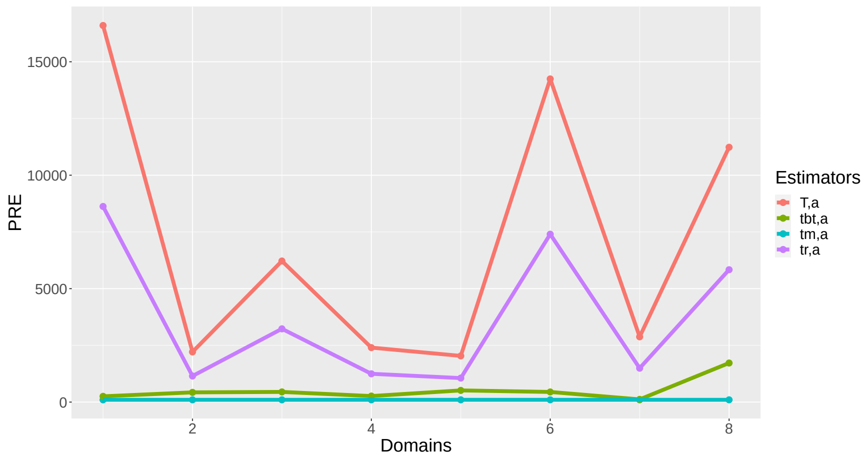

FIGURE 2: Graph of PRE of synthetic estimators for real population

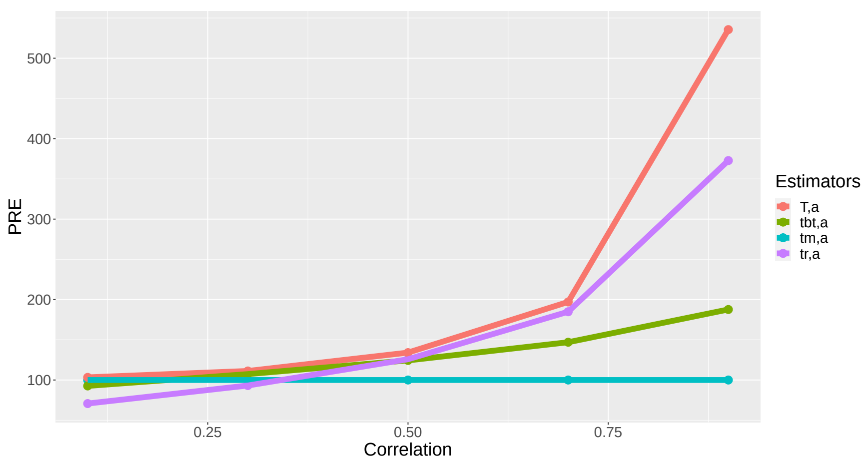

FIGURE 3: Graph of PRE values of domain 1 for direct estimators based on artificially generated normal population.

(vi) The similar tendency as observed from the results of Table 6 can also be observed from the results of Table 7 based on the artificially generated exponential population. This tendency can also be observed from the line diagram given in Figure 6 for the PRE values of domain 1.

(vii) From the results of Tables 4-5, the MSE and PRE of the proposed direct estimator T_a^d decrease and increase, respectively, as the correlation coefficient increases (see Figures 3-4). The similar tendency can be observed from the results of the proposed synthetic estimator T_a^s reported in Tables 6-7 (see Figures 5-6).

(viii) In the real population, the proposed synthetic estimator T_a^s outperforms the proposed direct estimator T_a^d in almost all domains except 2, 4, and 5, while in simulated populations, the proposed synthetic estimator T_a^s outperforms the proposed direct estimator T_a^d in all domains.

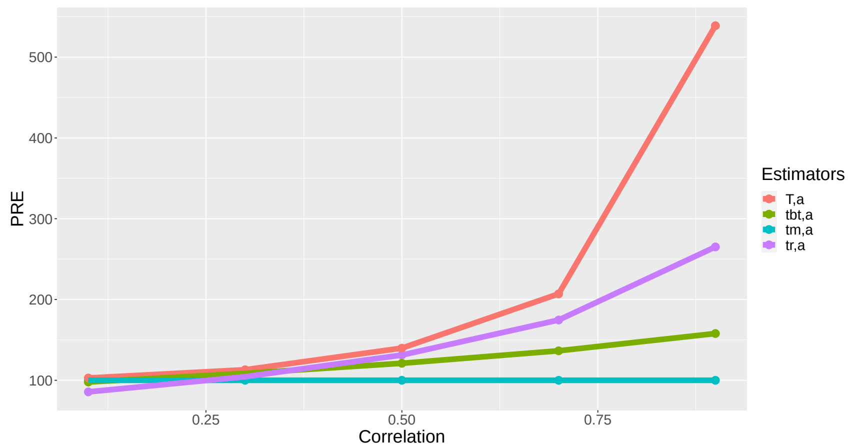

Figure 4: Graph of PRE values of domain 1 for direct estimators based on artificially generated exponential population

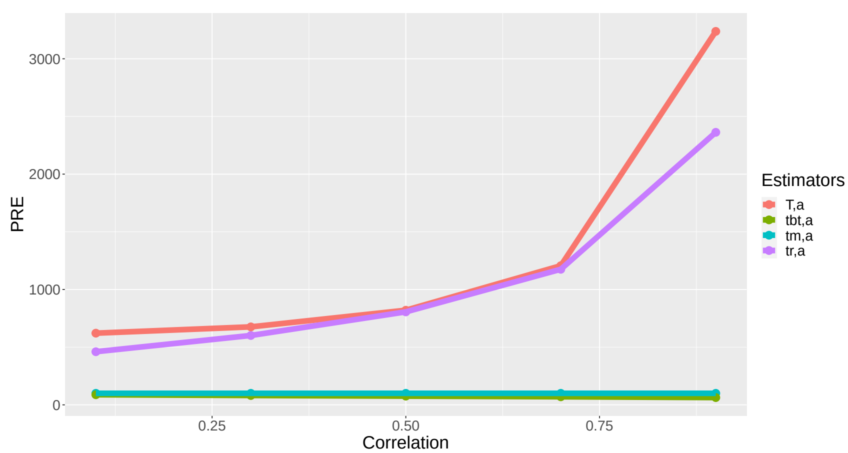

Figure 5: Graph of PRE values of domain 1 for synthetic estimators based on artificially generated normal population

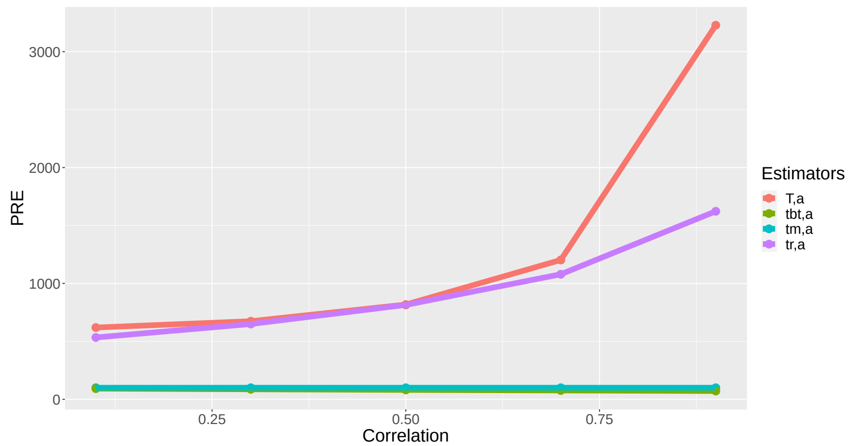

FIGURE 6: Graph of PRE values of domain 1 for synthetic estimators based on artificially generated exponential population

6. Conclusion

In this paper, we have proposed logarithmic type direct and synthetic estimators for the domain mean under SRS. The properties of the proposed estimators are obtained to the first-order approximation. Under some efficiency conditions, the proposed direct and synthetic estimators outperform the conventional direct and synthetic estimators. The theoretical results are supported with a numerical study using real data based on 6 domains and the results are reported in Tables 2-3. The results of the numerical study are further extrapolated by a simulation study using artificially generated symmetric and asymmetric populations based on 6 domains and the results are reported in Tables 4-7. The results of both numerical and simulation studies exhibit that the proposed direct and synthetic estimators dominate their corresponding conventional counterparts such as direct and synthetic usual mean estimators, direct and synthetic ratio estimators, and exponential ratio type direct and synthetic estimators in each domain of the real and simulated populations. The dominance of the proposed direct and synthetic estimators over their counterparts can also be seen from Figures 1-6. Furthermore, from the results of Tables 2-7, the proposed synthetic estimators are found to be better than the proposed direct estimators in almost all domains of the real population while in all domains of the simulated populations. Moreover, from the simulation results reported in Tables 4-7, the MSE and PRE of the proposed direct and synthetic estimators T_a^d and T_a^s decrease and increase, respectively, as the correlation coefficient increases (see Figures 3-6). Therefore, for the estimation of the domain mean, survey practitioners are strongly encouraged to use the proposed direct and synthetic logarithmic estimators. In the future, we hope to extend the present work for the estimation of population mean using several other sampling designs.

[Received: June 2023 — Accepted: December 2023]

References

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

Bhushan, S. & Kumar, A. (2020), Log type estimators of population mean under ranked set sampling, *in* 'Predictive Analytics using Statistics and Big Data: Concepts and Modeling', Vol. 28, pp. 47–74.

Bhushan, S. & Kumar, A. (2022*a*), 'New efficient logarithmic estimators using multi-auxiliary information under ranked set sampling', *Concurrency and Computation: Practice and Experience* **34**(27), e7337.

Bhushan, S. & Kumar, A. (2022*b*), 'Novel log type class of estimators under ranked set sampling', *Sankhya B* **84**(1), 421–447.

Bhushan, S., Kumar, A., Akhtar, M. T. & Lone, S. A. (2022), 'Logarithmic type predictive estimators under simple random sampling', *AIMS Mathematics* **7**(7), 11992–12010.

Bhushan, S., Kumar, A., Lone, S. A., Anwar, S. & Gunaime, N. M. (2023), 'An efficient class of estimators in stratified random sampling with an application to real data', *Axioms* **12**(6), 576.

Bhushan, S., Kumar, A. & Pokhrel, R. (2023), 'Logarithmic type direct and synthetic estimators using bivariate auxiliary information with an application to real data', *Journal of the Indian Society of Agricultural Statistics* **77**(1), 133–148.

Bhushan, S., Kumar, A., Zaman, T. & Al Mutairi, A. (2023), 'Efficient difference and ratio-type imputation methods under ranked set sampling', *Axioms* **12**(6), 558.

Gonzalez, M. E. (1973), *Use and evaluation of synthetic estimates*, US Census Bureau [custodian].

Khare, B., Ashutosh & Khare, S. (2018), 'Comparative study of synthetic estimators with ratio synthetic estimator for domain mean in survey sampling using auxiliary character', *International Journal of Applied Mathematics & Statistics* **57**(3), 152–161.

Khare, B. B. & Ashutosh (2017), 'Synthetic estimator for domain mean using auxiliary character', *International Journal of Statistics and Economics* **18**(4), 32–44.

Pandey, K. K. (2007), 'On synthetic and composite estimators for small area estimation under lahiri-midzuno sampling scheme', *Statistics in Transition - New Series* **8**(1), 111–123.

Pandey, K. & Tikkiwal, G. (2010), 'Generalized class of synthetic estimators for small area under systematic sampling design', *Statistics in Transition - New Series* **11**(1), 75–89.

Rai, P. K. & Pandey, K. (2013), 'Synthetic estimators using auxiliary information in small domains', *Statistics in Transition - New Series* **14**(1), 31–44.

Särndal, C.-E., Swensson, B. & Wretman, J. (2003), *Model assisted survey sampling*, Springer Science & Business Media.

Sharma, M. & Sisodia, B. (2016), 'Estimation of crop production for smaller geographical area: An application of discriminant function analysis', *Journal of Statistical Theory and Practice* **10**, 444–455.

Singh, S. & Horn, S. (1998), 'An alternative estimator for multi-character surveys', *Metrika* **48**, 99–107.

Sisodia, B. V. S. & Chandra, H. (2012), 'Estimation of crop production at smaller geographical level in India', *Journal of the Indian Society of Agricultural Statistics* **66**(2), 313–319.

Sisodia, B. V. S. & Singh, A. (2001), 'On small area estimation-an empirical study', *Journal of the Indian Society of Agricultural Statistics* **54**(3), 303–316.

Tikkiwal, G. C. & Ghiya, A. (2000), 'A generalized class of synthetic estimators with application to crop acreage estimation for small domains', *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **42**(7), 865–876.

Tikkiwal, G. C., Rai, P. K. & Ghiya, A. (2013), 'On the performance of generalized regression estimator for small domains', *Communications in Statistics-Simulation and Computation* **42**(4), 891–909.