Heliyon 10 (2024) e33839

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Enhanced direct and synthetic estimators for domain mean with simulation and applications

Anoop Kumar [a], Shashi Bhushan [b,*], Rohini Pokhrel [c], Walid Emam [d], Yusra Tashkandy [d], M.J.S. Khan [e]

[a] *Department of Statistics, Central University of Haryana, Mahendergarh, 123031, India*
[b] *Department of Statistics, University of Lucknow, Lucknow, 226007, India*
[c] *Department of Mathematics & Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, 226017, India*
[d] *Department of Statistics and Operations Research, Faculty of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia*
[e] *Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, 202002, India*

ARTICLE INFO

Keywords:
Small area estimation
Auxiliary variable
Direct estimators
Synthetic estimators
Simple random sampling

ABSTRACT

This article considers the issue of domain mean estimation utilizing bivariate auxiliary information based enhanced direct and synthetic logarithmic type estimators under simple random sampling (SRS). The expressions of mean square error (MSE) of the proposed estimators are provided to the 1^{st} order approximation. The efficiency criteria are derived to exhibit the dominance of the proposed estimators. To exemplify the theoretical results, a simulation study is conducted on a hypothetically drawn trivariate normal population from *R* programming language. Some applications of the suggested methods are also provided by analyzing the actual data from the municipalities of Sweden and acreage of paddy crop in the Mohanlal Ganj tehsil of the Indian state of Uttar Pradesh. The findings of the simulation and real data application exhibit that the proposed direct and synthetic logarithmic estimators dominate the conventional direct and synthetic mean, ratio, and logarithmic estimators in terms of least MSE and highest percent relative efficiency.

1. Introduction

The statistical field known as small area estimation (SAE) conflates the sampling survey, statistical models, and findings on a limited population. The development of agricultural preparation for prompt yield increase has necessitated the use of SAE techniques to gather data on many economic sectors, racial and ethnic groups, medical specialties, geographic areas, cultivable land, income, health, or poverty measures, minerals, among others. Extensive surveys may yield information at large scale of aggregation at both state and national levels because of sample designs. After acknowledging the necessity for precise values at the very basic levels of accumulation, including block, tehsil, and gramme panchayat, for the efficient utilization of monetary assets, the aim of the government is shifted from macro to micro levels.

The construct of a small area comes from an extensive survey that indicates its importance to know both the characteristics of the mean population/total and the dimensions of the well known domain of subpopulations. If these estimators of domain only use domain-specific sample data, they are known as direct estimators. A direct estimator may also employ the accessible supplementary information related to the parameter of interest. A thorough explanation of the direct method of estimation was provided by [1].

* Corresponding author.
E-mail addresses: anoop.asy@gmail.com (A. Kumar), bhushan_s@lkouniv.ac.in (S. Bhushan), rohanpokhrel90@gmail.com (R. Pokhrel).

https://doi.org/10.1016/j.heliyon.2024.e33839
Received 1 October 2023; Received in revised form 24 June 2024; Accepted 27 June 2024

Available online 4 July 2024

A modified direct regression estimator of domain means was developed by [2]. Ref. [3] investigated improved direct estimators for domain mean utilizing municipalities data.

When the domain sizes are very small, parametric estimates of these domains could not be correct since the sample from the usual direct sampling approach might not appropriately represent these domains. Under these circumstances, a synthetic (indirect) estimator is used to estimate the population parameter. By utilizing other domains data that are equivalent to the domain of small size of concern, these strategies essentially aim to offer an appropriate sample for all domains. Ref. [4] proposed a generalised synthetic estimator of the domain mean for the purpose of estimating agricultural area. A generalised synthetic estimator for small regions utilising systematic sampling was suggested by [5]. The effectiveness of the generalised regression estimator for small domains was investigated by [6]. In small domains, synthetic estimators employing auxiliary information were proposed by [7]. By employing auxiliary data, [8] suggested logarithmic type direct and synthetic estimators for domain mean under SRS, while [9] studied SAE utilizing design based direct and synthetic logarithmic estimators for domain mean. For a more thorough examination of small area estimation, the reader is advised to see [10], [11], [12], [13] among others.

Utilising the auxiliary data effectively allows the survey researchers to increase the effectiveness of their suggested estimators. This information is connected to the supplementary variable, which has a high correlation with the variable being investigated. As a result, sample surveys are frequently used to estimate parameters that are pertinent in small areas. Only a small amount of work has been done on the estimation of domain mean utilizing bivariate auxiliary information (BVAI) under SRS. Ref. [14] examined the work of [4] by utilizing BVAI. However, [15] suggested BVAI-based direct and synthetic logarithmic estimators for domain mean under SRS. Most of the researchers investigated both direct and synthetic estimators utilizing BVAI separately. The goal of this paper is distinct from the above researches and discussed in the following points:

- We propose enhanced direct and synthetic logarithmic type estimators of domain mean using SRS by employing BVAI.
- Mathematically compare the efficiency of the proposed direct and synthetic estimators with the existing direct and synthetic estimators.
- Numerically compare the performance of the proposed direct and synthetic estimators against the existing direct and synthetic estimators.
- The performance of the proposed direct and synthetic estimators is illustrated utilizing some real data sets.

1.1. Notation

Suppose that the population $\Omega = (\Omega_1, \Omega_2, ..., \Omega_N)$ is made up of 'A' unique Ω_a small areas, or domains, that are each N_a in size. The domains, which might refer to a district, tehsil, or similar state-level entity, are the small domains of a population being surveyed and may consider a variety of various shapes according to the circumstances. Let y be the research variable and (x, z) be the supplementary variables. The constraint that n_a, $a = 1, 2, ..., A$ units in the s^{th} sample obtained from the domain 'a' of small size lead to the selection of a random sample of n size without replacement. Thus, $\sum_{a=1}^{A} N_a = N$ and $\sum_{a=1}^{A} n_a = n$. Let y_{ai}, x_{ai}, and z_{ai}; $i = 1, 2, ..., N_a$, denote the i^{th} unit of the small domain a of the population for the characteristics y, x, and z, respectively. The terminologies used throughout the paper are defined below:

$\bar{Y} = \sum_{i=1}^{N} y_i/N$: population mean utilizing N observations on y;
$\bar{Y}_a = \sum_{i=1}^{N_a} y_{ai}/N_a$: population mean of domain a utilizing N_a observations on y;
$\bar{X} = \sum_{i=1}^{N} x_i/N$: population mean of variable x utilizing N observations;
$\bar{X}_a = \sum_{i=1}^{N_a} x_{ai}/N_a$: population mean of variable x for domain a utilizing N_a observations;
$\bar{Z} = \sum_{i=1}^{N} z_i/N$: population mean of variable z utilizing N observations;
$\bar{Z}_a = \sum_{i=1}^{N_a} z_{ai}/N_a$: population mean of variable z for domain a utilizing N_a observations;
$\bar{x} = \sum_{i=1}^{n} x_i/n$: sample mean utilizing n observations on characteristic x;
$\bar{x}_a = \sum_{i=1}^{n_a} x_{ai}/n_a$: sample mean utilizing n_a observations on x;
$\bar{z} = \sum_{i=1}^{n} z_i/n$: sample mean utilizing n observations on characteristic z;
$\bar{z}_a = \sum_{i=1}^{n_a} z_{ai}/n_a$: sample mean utilizing n_a observations on z;
$\bar{y} = \sum_{i=1}^{n} y_i/n$: sample mean utilizing n observations on y;
$\bar{y}_a = \sum_{i=1}^{n_a} y_{ai}/n_a$: sample mean of domain a utilizing n_a observations on y;
$S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$: Population mean square of variable x;
$S_{x_a}^2 = \sum_{i=1}^{N_a}(x_{ai} - \bar{X}_a)^2/(N_a-1)$: Population mean square of variable x for the domain a;
$S_z^2 = \sum_{i=1}^{N}(z_i - \bar{Z})^2/(N-1)$: Population mean square of variable z;
$S_{z_a}^2 = \sum_{i=1}^{N_a}(z_{ai} - \bar{Z}_a)^2/(N_a-1)$: Population mean square of variable z for the domain a;
$S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$: Population mean square of variable y;
$S_{y_a}^2 = \sum_{i=1}^{N_a}(y_{ai} - \bar{Y}_a)^2/(N_a-1)$: Population mean square of variable y for the domain a;
$C_x = S_x/\bar{X}$: Population variation coefficient of variable x;
$C_{x_a} = S_{x_a}/\bar{X}_a$: Population variation coefficient of variable x for domain a;
$C_z = S_z/\bar{Z}$: Population variation coefficient of variable z;
$C_{z_a} = S_{z_a}/\bar{Z}_a$: Population variation coefficient of variable z for domain a;

$C_y = S_y/\bar{Y}$: Population variation coefficient of variable y;
$C_{y_a} = S_{y_a}/\bar{Y}_a$: Population variation coefficient of variable y for domain a;
ρ_{yx}: correlation coefficient of variables y and x;
ρ_{yz}: correlation coefficient of variables y and z;
ρ_{xz}: correlation coefficient of variables x and z;
$\rho_{y_a x_a}$: correlation coefficient of variables y and x for the domain a;
$\rho_{y_a z_a}$: correlation coefficient of variables y and z for the domain a;
$\rho_{x_a z_a}$: correlation coefficient of variables x and z for the domain a.
To obtain the characteristics of the direct estimators, we consider the following error terms:
$e_0 = (\bar{y}_a - \bar{Y}_a)/\bar{Y}_a$, $e_1 = (\bar{x}_a - \bar{X}_a)/\bar{X}_a$, $e_2 = (\bar{z}_a - \bar{Z}_a)/\bar{Z}_a$; such that $E(e_i) = 0,\ i = 0, 1, 2,\ E(e_0^2) = f_a C_{y_a}^2$, $E(e_1^2) = f_a C_{x_a}^2$, $E(e_2^2) = f_a C_{z_a}^2$, $E(e_0 e_1) = f_a \rho_{y_a x_a} C_{y_a} C_{x_a}$, $E(e_0 e_2) = f_a \rho_{y_a z_a} C_{y_a} C_{z_a}$, and $E(e_1 e_2) = f_a \rho_{x_a z_a} C_{x_a} C_{z_a}$, where $f_a = (N_a - n_a)/N_a n_a$.

Similarly, to obtain the characteristics of the synthetic estimators, we consider the following error terms:
$\varepsilon_0 = (\bar{y} - \bar{Y})/\bar{Y}$, $\varepsilon_1 = (\bar{x} - \bar{X})/\bar{X}$, $\varepsilon_2 = (\bar{z} - \bar{Z})/\bar{Z}$; such that $E(\varepsilon_i) = 0,\ i = 0, 1, 2,\ E(\varepsilon_0^2) = f C_y^2$, $E(\varepsilon_1^2) = f C_x^2$, $E(\varepsilon_2^2) = f C_z^2$, $E(\varepsilon_0 \varepsilon_1) = f \rho_{yx} C_y C_x$, $E(\varepsilon_0 \varepsilon_2) = f \rho_{yz} C_y C_z$, and $E(\varepsilon_1 \varepsilon_2) = f \rho_{xz} C_x C_z$, where $f = (N - n)/Nn$.

Further, in this section, we consider a literature review of all well-known direct and synthetic estimators of the domain mean that rely on BVAI.

1.2. Direct estimators

The direct conventional unbiased estimator is shown hereunder as

$$\bar{y}_{m,a}^d = \bar{y}_a$$

The variance of the estimator $\bar{y}_{m,a}^d$ is given by

$$V(\bar{y}_{m,a}^d) = f_a \bar{Y}_a^2 C_{y_a}^2$$

The BVAI based direct ratio estimator is given by

$$\bar{y}_{r,a}^d = \bar{y}_a \left(\frac{\bar{X}_a}{\bar{x}_a}\right)\left(\frac{\bar{Z}_a}{\bar{z}_a}\right)$$

The MSE of the estimator $\bar{y}_{r,a}^d$ is given by

$$MSE(\bar{y}_{r,a}^d) = f_a \bar{Y}_a^2 (C_{y_a}^2 + C_{x_a}^2 + C_{z_a}^2 - 2\rho_{y_a x_a} C_{y_a} C_{x_a} - 2\rho_{y_a z_a} C_{y_a} C_{z_a} + 2\rho_{x_a z_a} C_{x_a} C_{z_a})$$

Ref. [15] suggested direct logarithmic type estimator using BVAI as follows:

$$\bar{y}_{bk,a}^d = \bar{y}_a \left[1 + \lambda_a \log\left(\frac{\bar{x}_a}{\bar{X}_a}\right)\right]\left[1 + \delta_a \log\left(\frac{\bar{z}_a}{\bar{Z}_a}\right)\right]$$

where λ_a and δ_a are the constants.

The MSE of the estimator $\bar{y}_{bk,a}^d$ is given by

$$MSE(\bar{y}_{bk,a}^d) = f_a \bar{Y}_a^2 \begin{pmatrix} C_{y_a}^2 + \lambda_a^2 C_{x_a}^2 + \delta_a^2 C_{z_a}^2 + 2\lambda_a \rho_{y_a x_a} C_{y_a} C_{x_a} + 2\delta_a \rho_{y_a z_a} C_{y_a} C_{z_a} \\ +2\lambda_a \delta_a \rho_{x_a z_a} C_{x_a} C_{z_a} \end{pmatrix} \quad (1)$$

Minimize (1) with respect to λ_a and δ_a, we get

$$\lambda_{a(opt)} = \left(\frac{C_{y_a}}{C_{x_a}}\right)\left(\frac{\rho_{y_a z_a}\rho_{x_a z_a} - \rho_{y_a x_a}}{1 - \rho_{x_a z_a}^2}\right)$$

$$\delta_{a(opt)} = \left(\frac{C_{y_a}}{C_{z_a}}\right)\left(\frac{\rho_{y_a x_a}\rho_{x_a z_a} - \rho_{y_a z_a}}{1 - \rho_{x_a z_a}^2}\right)$$

The minimal MSE of the suggested direct logarithmic estimator $\bar{y}_{bk,a}^d$ is determined by using $\lambda_{a(opt)}$ and $\delta_{a(opt)}$ values in (1) as

$$MSE(\bar{y}_{bk,a}^d)_{min} = f_a \bar{Y}_a^2 C_{y_a}^2 \left(1 - R_{y_a \cdot x_a z_a}^2\right)$$

where $R_{y_a \cdot x_a z_a}^2$ denotes the multiple coefficient of correlation (y on x, z) in domain a.

1.3. Synthetic estimators

The synthetic mean per unit estimator is given by

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

The MSE of the estimator $\bar{y}^s_{m,a}$ is given by

$$MSE(\bar{y}^s_{m,a}) = (\bar{Y} - \bar{Y}_a)^2 + f\bar{Y}^2C_y^2$$

The BVAI based synthetic ratio estimator is given by

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

Under the synthetic assumption $\bar{Y}_a \cong \bar{Y}(\bar{X}_a/\bar{X})(\bar{Z}_a/\bar{Z})$, the $MSE(\bar{y}^s_{r,a})$ is given by

$$MSE(\bar{y}^s_{r,a}) = \bar{Y}_a^2 f(C_y^2 + C_x^2 + C_z^2 - 2\rho_{yx}C_yC_x - 2\rho_{yz}C_yC_z + 2\rho_{xz}C_xC_z)$$

Ref. [15] suggested BVAI based synthetic logarithmic type estimator as follows:

$$\bar{y}^s_{bk,a} = \bar{y}\left[1 + \lambda\log\left(\frac{\bar{x}}{\bar{X}_a}\right)\right]\left[1 + \delta\log\left(\frac{\bar{z}}{\bar{Z}_a}\right)\right]$$

where λ, and δ are the constants.

Using $\bar{Y}_a \cong \bar{Y}(1+\lambda A)(1+\delta B)$, $\bar{Y}_a \cong \bar{Y}(1+\lambda A)$, and $\bar{Y}_a \cong \bar{Y}(1+\delta B)$ as synthetic assumptions, the $MSE(\bar{y}^s_{bk,a})$ can be expressed as

$$MSE(\bar{y}^s_{bk,a}) = f\bar{Y}_a^2\left(C_y^2 + \lambda^2C_x^2 + \delta^2C_z^2 + 2\lambda\rho_{yx}C_yC_x + 2\delta\rho_{yz}C_yC_z + 2\lambda\delta\rho_{xz}C_xC_z \right) \tag{2}$$

The optimum values of λ and δ are obtained by minimizing (2) for these variables as

$$\lambda_{(opt)} = \left(\frac{C_y}{C_x}\right)\left(\frac{\rho_{yz}\rho_{xz} - \rho_{yx}}{1-\rho_{xz}^2}\right)$$

$$\delta_{(opt)} = \left(\frac{C_y}{C_z}\right)\left(\frac{\rho_{yx}\rho_{xz} - \rho_{yz}}{1-\rho_{xz}^2}\right)$$

We put $\lambda_{(opt)}$ and $\delta_{(opt)}$ in (2) to establish the minimal MSE of the proposed synthetic estimator $\bar{y}^s_{bk,a}$.

$$MSE(\bar{y}^s_{bk,a})_{min} = f\bar{Y}_a^2C_y^2\left(1 - R^2_{y.xz}\right)$$

where $R^2_{y.xz}$ denotes the multiple coefficient of correlation (y on x, z).

In Section 2, we propose BVAI-based enhanced direct and synthetic logarithmic type estimators for the mean of domain a. Section 3 develops the efficiency criteria for the proposed estimators. In Section 4, the efficiency criteria of the proposed estimators have been assessed by performing a simulation study. In Section 5, some applications of the suggested direct and synthetic logarithmic type estimators are presented utilizing the data of the municipalities of Sweden and the acreage of paddy crop of the Indian state of Uttar Pradesh's Mohanlal Ganj tehsil. Section 6 concisely conclude this paper.

2. Suggested estimators

In this section, the logarithmic transformation for the direct and synthetic estimators is used due to the following key points. Firstly, the logarithm transformation can stabilize variance, which is particularly useful when dealing with data that exhibit heteroscedasticity, where the variability of the data changes across the range of values. Secondly, it can normalize data distributions that are skewed, making the data more closely approximate a normal distribution, which is a common assumption for many statistical methods. Additionally, the logarithm transformation can make multiplicative relationships additive, simplifying the modelling of complex relationships within the data. These benefits enhance the interpretability and robustness of the estimators, potentially leading to more reliable and meaningful results. For more details about the advantages of logarithmic transformation, reader may see [16] and [17]. Taking inspiration from these advantages and motivated by [16], this paper suggests BVAI based enhanced direct and synthetic logarithmic type estimators for mean of domain a under SRS.

2.1. Direct estimator

The proposed novel direct logarithmic estimator is given by

$$\bar{y}^d_{kp,a} = \zeta_a\bar{y}_a\left[1 + \lambda_a\log\left(\frac{\bar{x}_a}{\bar{X}_a}\right)\right]\left[1 + \delta_a\log\left(\frac{\bar{z}_a}{\bar{Z}_a}\right)\right]$$

where ζ_a, λ_a, and δ_a are the suitably opted scalars.

To establish the mathematical expressions of the mean square error and minimum mean square error of the suggested direct logarithmic estimator $\bar{y}^d_{kp,a}$, we consider the error terms defined in Section 1.1 and express $\bar{y}^d_{kp,a}$ as

$$\begin{aligned}\bar{y}^d_{kp,a} &= \zeta_a \bar{Y}_a(1+e_0)\left[1+\lambda_a \log\left(\frac{\bar{X}_a(1+e_1)}{\bar{X}_a}\right)\right]\left[1+\delta_a \log\left(\frac{\bar{Z}_a(1+e_2)}{\bar{Z}_a}\right)\right] \\ &= \zeta_a \bar{Y}_a(1+e_0)\left[1+\lambda_a\left(e_1-\frac{e_1^2}{2}+\dots\right)\right]\left[1+\delta_a\left(e_2-\frac{e_2^2}{2}+\dots\right)\right]\end{aligned}$$

When we neglect the error components with powers higher than 2 and subtract $\bar{Y}_a$ from both sides of the previous expression, we obtain

$$\bar{y}^d_{kp,a}-\bar{Y}_a = (\zeta_a\bar{Y}_a-\bar{Y}_a)+\zeta_a\bar{Y}_a\left\{\begin{array}{l} e_0+\lambda_a\left(e_1-\frac{e_1^2}{2}\right)+\delta_a\left(e_2-\frac{e_2^2}{2}\right)+\lambda_a e_0 e_1 \\ +\delta_a e_0 e_2+\lambda_a\delta_a e_1 e_2\end{array}\right\}$$

Squaring and taking expectation on both sides provides

$$MSE(\bar{y}^d_{kp,a}) = \bar{Y}_a^2\left[\begin{array}{l} 1+\zeta_a^2\left\{1+f_a\left(\begin{array}{l} C^2_{y_a}+\lambda_a(\lambda_a-1)C^2_{x_a}+\delta_a(\delta_a-1)C^2_{z_a} \\ +4\lambda_a\rho_{y_ax_a}C_{y_a}C_{x_a}+4\delta_a\rho_{y_az_a}C_{y_a}C_{z_a} \\ +4\lambda_a\delta_a\rho_{x_az_a}C_{x_a}C_{z_a}\end{array}\right)\right\} \\ -2\zeta_a\left\{1+f_a\left(\begin{array}{l}\lambda_a\rho_{y_ax_a}C_{y_a}C_{x_a}+\delta_a\rho_{y_az_a}C_{y_a}C_{z_a} \\ +\lambda_a\delta_a\rho_{x_az_a}C_{x_a}C_{z_a}-\frac{\lambda_a}{2}C^2_{x_a}-\frac{\delta_a}{2}C^2_{z_a}\end{array}\right)\right\}\end{array}\right]$$

The above expression can further be written as

$$MSE(\bar{y}^d_{kp,a}) = \bar{Y}_a^2\left(1+\zeta_a^2 P_1-2\zeta_a Q_1\right) \tag{3}$$

where $P_1 = 1+f_a\left(\begin{array}{l} C^2_{y_a}+\lambda_a(\lambda_a-1)C^2_{x_a}+\delta_a(\delta_a-1)C^2_{z_a}+4\lambda_a\rho_{y_ax_a}C_{y_a}C_{x_a} \\ +4\delta_a\rho_{y_az_a}C_{y_a}C_{z_a}+4\lambda_a\delta_a\rho_{x_az_a}C_{x_a}C_{z_a}\end{array}\right)$ and $Q_1 = 1+f_a\left(\begin{array}{l}\lambda_a\rho_{y_ax_a}C_{y_a}C_{x_a}+\delta_a\rho_{y_az_a}C_{y_a}C_{z_a} \\ +\lambda_a\delta_a\rho_{x_az_a}C_{x_a}C_{z_a}-\frac{\lambda_a}{2}C^2_{x_a}-\frac{\delta_a}{2}C^2_{z_a}\end{array}\right)$.

Minimize (3) with respect to ζ_a gives the optimum value of ζ_a as

$$\zeta_{a(opt)} = \frac{Q_1}{P_1}$$

Using the value of $\zeta_{a(opt)}$ in (3), provides

$$min.MSE(\bar{y}^d_{kp,a}) = \bar{Y}_a^2\left(1-\frac{Q_1^2}{P_1}\right) \tag{4}$$

It is remarkable that the simultaneously optimizing $\lambda_{a(opt)}$ and $\delta_{a(opt)}$ is not possible. The optimum values of $\lambda_{a(opt)}$ and $\delta_{a(opt)}$ can be established by using $\zeta_a = 1$ in the estimator $\bar{y}^d_{kp,a}$ and minimizing the MSE expression regarding λ_a and δ_a, respectively. The optimum value of λ_a and δ_a are reported below:

$$\lambda_{a(opt)} = \left(\frac{C_{y_a}}{C_{x_a}}\right)\left(\frac{\rho_{y_az_a}\rho_{x_az_a}-\rho_{y_ax_a}}{1-\rho^2_{x_az_a}}\right)$$

$$\delta_{a(opt)} = \left(\frac{C_{y_a}}{C_{z_a}}\right)\left(\frac{\rho_{y_ax_a}\rho_{x_az_a}-\rho_{y_az_a}}{1-\rho^2_{x_az_a}}\right)$$

2.2. Synthetic estimator

The proposed novel synthetic logarithmic estimator is

$$\bar{y}^s_{kp,a} = \zeta\bar{y}\left[1+\lambda\log\left(\frac{\bar{x}}{\bar{X}_a}\right)\right]\left[1+\delta\log\left(\frac{\bar{z}}{\bar{Z}_a}\right)\right]$$

where ζ, λ, and δ are the suitably opted scalars.

We use the error terms defined in Section 1.1 to express $\bar{y}^s_{kp,a}$ and determine the mean square error and minimum mean square error of the estimator $\bar{y}^s_{kp,a}$.

$$\bar{y}^{s}_{kp,a} = \zeta \bar{Y}(1+\varepsilon_0)\left[1+\lambda \log\left(\frac{\bar{X}(1+\varepsilon_1)}{\bar{X}_a}\right)\right]\left[1+\delta \log\left(\frac{\bar{Z}(1+\varepsilon_2)}{\bar{Z}_a}\right)\right]$$
$$= \zeta \bar{Y}(1+\varepsilon_0)\left[1+\lambda \log\left(\frac{\bar{X}}{\bar{X}_a}\right)+\lambda\log\left(1+\varepsilon_1\right)\right]\left[1+\delta \log\left(\frac{\bar{Z}}{\bar{Z}_a}\right)+\delta\log\left(1+\varepsilon_2\right)\right]$$
$$= \zeta \bar{Y}(1+\varepsilon_0)\left[1+\lambda A+\lambda\left(\varepsilon_1-\frac{\varepsilon_1^2}{2}+...\right)\right]\left[1+\delta B+\delta\left(\varepsilon_2-\frac{\varepsilon_2^2}{2}+...\right)\right]$$

where $A=\log\left(\frac{\bar{X}}{\bar{X}_a}\right)$ and $B=\log\left(\frac{\bar{Z}}{\bar{Z}_a}\right)$. Further, by excluding error terms with powers higher than 2 and subtracting $\bar{Y}_a$ on both sides of the last expression, we arrive at

$$\bar{y}^{s}_{kp,a}-\bar{Y}_a = \zeta\bar{Y}\left\{\begin{array}{l}(1+\lambda A)(1+\delta B)+(1+\lambda A)\delta\left(\varepsilon_2-\frac{\varepsilon_2^2}{2}\right)+(1+\delta B)\lambda\left(\varepsilon_1-\frac{\varepsilon_1^2}{2}\right)\\ +\lambda\delta\varepsilon_1\varepsilon_2+(1+\lambda A)(1+\delta B)\varepsilon_0+(1+\lambda A)\delta\varepsilon_0\varepsilon_2+(1+\delta B)\lambda\varepsilon_0\varepsilon_1\end{array}\right\}-\bar{Y}_a \tag{5}$$

Taking square and expectation both sides to (5), we get

$$MSE(\bar{y}^{s}_{kp,a}) = E\left[\begin{array}{l}\{\zeta\bar{Y}(1+\lambda A)(1+\delta B)-\bar{Y}_a\}^2\\ +\zeta^2\bar{Y}^2\left\{\begin{array}{l}\lambda(1+\delta B)\left(\varepsilon_1-\frac{\varepsilon_1^2}{2}\right)+(1+\lambda A)\delta\left(\varepsilon_2-\frac{\varepsilon_2^2}{2}\right)\\ +\lambda\delta\varepsilon_1\varepsilon_2+(1+\lambda A)(1+\delta B)\varepsilon_0+(1+\lambda A)\delta\varepsilon_0\varepsilon_2\\ +(1+\delta B)\lambda\varepsilon_0\varepsilon_1\end{array}\right\}^2\\ +2\{\zeta\bar{Y}(1+\lambda A)(1+\delta B)-\bar{Y}_a\}\zeta\bar{Y}\times\\ \left\{\begin{array}{l}\lambda(1+\delta B)\left(\varepsilon_1-\frac{\varepsilon_1^2}{2}\right)+(1+\lambda A)\delta\left(\varepsilon_2-\frac{\varepsilon_2^2}{2}\right)+\lambda\delta\varepsilon_1\varepsilon_2\\ +(1+\lambda A)(1+\delta B)\varepsilon_0+(1+\lambda A)\delta\varepsilon_0\varepsilon_2+(1+\delta B)\lambda\varepsilon_0\varepsilon_1\end{array}\right\}\end{array}\right]$$

After simplifying and using $\bar{Y}_a \cong \bar{Y}(1+\lambda A)(1+\delta B)$, $\bar{Y}_a \cong \bar{Y}(1+\lambda A)$, and $\bar{Y}_a \cong \bar{Y}(1+\delta B)$ as the synthetic assumptions, the $MSE(\bar{y}^{s}_{kp,a})$ is expressed as

$$MSE(\bar{y}^{s}_{kp,a}) = \left[\begin{array}{l}\bar{Y}_a^2+\zeta^2\left(\begin{array}{l}\bar{Y}_a^2\left[1+f\left\{\begin{array}{l}C_y^2+\lambda(\lambda-1)C_x^2+\delta(\delta-1)C_z^2\\ +4\lambda\rho_{yx}C_yC_x+4\delta\rho_{yz}C_yC_z\\ +2\lambda\delta\rho_{xz}C_xC_z\end{array}\right\}\right]\\ +2\bar{Y}\bar{Y}_a f\lambda\delta C_xC_z\end{array}\right)\\ -2\zeta\left[\begin{array}{l}\bar{Y}_a^2\left\{1+f\left(\lambda\rho_{yx}C_yC_x+\delta\rho_{yz}C_yC_z-\frac{\lambda}{2}C_x^2-\frac{\delta}{2}C_z^2\right)\right\}\\ +\bar{Y}\bar{Y}_a f\lambda\delta\rho_{xz}C_xC_z\end{array}\right]\end{array}\right]$$

which can further be written as

$$MSE(\bar{y}^{s}_{kp,a}) = \bar{Y}_a^2+\zeta^2P_2-2\zeta Q_2 \tag{6}$$

where $P_2=\bar{Y}_a^2\left[1+f\left\{\begin{array}{l}C_y^2+\lambda(\lambda-1)C_x^2+\delta(\delta-1)C_z^2+4\lambda\rho_{yx}C_yC_x\\ +4\delta\rho_{yz}C_yC_z+2\lambda\delta\rho_{xz}C_xC_z\end{array}\right\}\right]+2\bar{Y}\bar{Y}_a f\lambda\delta C_xC_z$ and
$Q_2=\left[\begin{array}{l}\bar{Y}_a^2\left\{1+f\left(\lambda\rho_{yx}C_yC_x+\delta\rho_{yz}C_yC_z-\frac{\lambda}{2}C_x^2-\frac{\delta}{2}C_z^2\right)\right\}\\ +\bar{Y}\bar{Y}_a f\lambda\delta\rho_{xz}C_xC_z\end{array}\right]$.

Minimize (6) with respect to ζ, we get

$$\zeta_{(opt)}=\frac{Q_2}{P_2}$$

Using the value of $\zeta_{(opt)}$ in (6), provides

$$min.MSE(\bar{y}^{s}_{kp,a}) = \bar{Y}_a^2-\frac{Q_2^2}{P_2} \tag{7}$$

It is remarkable that the simultaneously optimizing $\lambda_{(opt)}$ and $\delta_{(opt)}$ is tedious. The optimum values of $\lambda_{(opt)}$ and $\delta_{(opt)}$ can be determined by using $\zeta=1$ in the estimator $\bar{y}^{s}_{kp,a}$ and minimizing the MSE expression with respect to λ and δ, respectively. The optimum value of λ and δ are reported below:

$$\lambda_{(opt)}=\left(\frac{C_y}{C_x}\right)\left(\frac{\rho_{yz}\rho_{xz}-\rho_{yx}}{1-\rho_{xz}^2}\right)$$

$$\delta_{(opt)} = \left(\frac{C_y}{C_z}\right)\left(\frac{\rho_{yx}\rho_{xz} - \rho_{yz}}{1 - \rho_{xz}^2}\right)$$

Corollary 2.1. *The suggested novel direct estimator $\bar{y}^d_{kp,a}$ performs worse than the suggested novel synthetic estimator $\bar{y}^s_{kp,a}$, if*

$$\frac{Q_2^2}{P_2} > \bar{Y}_a^2 \frac{Q_1^2}{P_1^2} \tag{8}$$

and vice versa. Otherwise, the estimators $\bar{y}^d_{kp,a}$ and $\bar{y}^s_{kp,a}$ are efficient equally provided the equivalence in (8) retains.

Proof. In order to achieve (8), we compare (4) and (7). □

3. Efficiency conditions

By evaluating the MSEs of the suggested and the current direct and synthetic estimators, this section supplies the efficiency conditions in the subsequent lemma.

Lemma 3.1.

(i). *The proposed direct estimator $\bar{y}^d_{kp,a}$ dominates the direct mean per unit estimator $\bar{y}^d_{m,a}$ if*

$$MSE(\bar{y}^d_{kp,a}) < V(\bar{y}^d_{m,a}) \implies \frac{Q_1^2}{P_1} > 1 - f_a C_{y_a}^2$$

(ii). *The proposed synthetic estimator $\bar{y}^s_{kp,a}$ dominates the synthetic mean per unit estimator $\bar{y}^s_{m,a}$ if*

$$MSE(\bar{y}^s_{kp,a}) < MSE(\bar{y}^s_{m,a}) \implies \frac{Q_2^2}{P_2} > \bar{Y}_a^2 - f\bar{Y}_a^2 C_y^2$$

Lemma 3.2.

(i). *The proposed direct estimator $\bar{y}^d_{kp,a}$ dominates the direct ratio estimator $\bar{y}^d_{r,a}$ if*

$$MSE(\bar{y}^d_{kp,a}) < MSE(\bar{y}^d_{r,a}) \implies \frac{Q_1^2}{P_1} > 1 - f_a \left(\begin{array}{l} C_{y_a}^2 + C_{x_a}^2 + C_{z_a}^2 - 2\rho_{y_a x_a} C_{y_a} C_{x_a} \\ -2\rho_{y_a z_a} C_{y_a} C_{z_a} + 2\rho_{x_a z_a} C_{x_a} C_{z_a} \end{array}\right)$$

(ii). *The proposed synthetic estimator $\bar{y}^s_{kp,a}$ dominates the synthetic ratio estimator $\bar{y}^s_{r,a}$ if*

$$MSE(\bar{y}^s_{kp,a}) < MSE(\bar{y}^s_{r,a}) \implies \frac{Q_2^2}{P_2} > \bar{Y}_a^2 - f\bar{Y}_a^2 \left(\begin{array}{l} C_y^2 + C_x^2 + C_z^2 - 2\rho_{yx} C_y C_x \\ -2\rho_{yz} C_y C_z + 2\rho_{xz} C_x C_z \end{array}\right)$$

Lemma 3.3.

(i). *The proposed direct estimator $\bar{y}^d_{kp,a}$ dominates the conventional direct logarithmic type estimator $\bar{y}^d_{bk,a}$ if*

$$MSE(\bar{y}^d_{kp,a}) < MSE(\bar{y}^d_{bk,a}) \implies \frac{Q_1^2}{P_1} > 1 - f_a C_{y_a}^2 (1 - R^2_{y_a . x_a z_a})$$

(ii). *The proposed synthetic estimator $\bar{y}^s_{kp,a}$ dominates the conventional synthetic logarithmic type estimator $\bar{y}^s_{bk,a}$ if*

$$MSE(\bar{y}^s_{kp,a}) < MSE(\bar{y}^s_{bk,a}) \implies \frac{Q_2^2}{P_2} > \bar{Y}_a^2 - f\bar{Y}_a^2 C_y^2 (1 - R^2_{y.xz})$$

The suggested direct and synthetic estimators dominate the contemporary direct and synthetic estimators under the efficiency conditions determined in the above lemmas.

4. Simulation study

To assess the effectiveness of the suggested direct and synthetic logarithmic type estimators, a simulation experiment is conducted utilizing a hypothetically drawn normal population. The normal population of $N = 16{,}000$ size is generated with the parameters

$\bar{Y}=15$, $\bar{X}=20$, $\bar{Z}=25$, $\sigma_y=76$, $\sigma_x=75$, $\sigma_z=77$ and various values of correlation coefficients ρ_{yz}, ρ_{yx}, and ρ_{xz} mentioned in Tables 1-2.

The population is divided into 8 domains each of size 2000. The required statistics of all domains are calculated for a random sample of 440 which is chosen from all domains. Based on 16,000 iterations, the direct and synthetic estimators' MSE and percent relative efficiency (PRE) are tabulated by considering the formulae given below:

$$MSE(\bar{y}^d_{*,a})=\frac{1}{16,000}\sum_{i=1}^{16,000}(\bar{y}^d_{*,a}-\bar{Y}_a)^2$$

$$MSE(\bar{y}^s_{*,a})=\frac{1}{16,000}\sum_{i=1}^{16,000}(\bar{y}^s_{*,a}-\bar{Y}_a)^2$$

$$PRE(\bar{y}^d_{m,a},\bar{y}^d_{*,a})=\frac{MSE(\bar{y}^d_{m,a})}{MSE(\bar{y}^d_{*,a})}\times 100$$

$$PRE(\bar{y}^s_{m,a},\bar{y}^s_{*,a})=\frac{MSE(\bar{y}^s_{m,a})}{MSE(\bar{y}^s_{*,a})}\times 100$$

where $\bar{y}^d_{*,a}=\bar{y}^d_{m,a}$, $\bar{y}^d_{r,a}$, $\bar{y}^d_{bk,a}$, $\bar{y}^d_{kp,a}$, and $\bar{y}^s_{*,a}=\bar{y}^s_{m,a}$, $\bar{y}^s_{r,a}$, $\bar{y}^s_{bk,a}$, $\bar{y}^s_{kp,a}$.

Tables 1-2 contain the outcomes for the direct and synthetic estimators.

4.1. Important results of simulation study

The important results of the simulation study have been described in pointwise format to provide a thorough grasp of the merits of the suggested estimators.

(i). The outcomes of Table 1 exhibit that the proposed direct logarithmic estimator $\bar{y}^d_{kp,a}$ represses the conventional direct estimators like direct unbiased estimator $\bar{y}^d_{m,a}$, direct ratio estimator $\bar{y}^d_{r,a}$, direct conventional logarithmic estimator $\bar{y}^d_{bk,a}$ having least mean square error and highest percent relative efficiency, respectively, for different amounts of ρ_{yz}, ρ_{yx}, and ρ_{xz} in all domains.

(ii). The outcomes of Table 2 exhibit that the proposed synthetic estimator $\bar{y}^s_{kp,a}$ represses the conventional synthetic estimators such as synthetic mean per unit estimator $\bar{y}^s_{m,a}$, synthetic ratio estimator $\bar{y}^s_{r,a}$, synthetic conventional logarithmic estimator $\bar{y}^s_{bk,a}$ having least mean square error and highest percent relative efficiency, respectively, for different amounts of ρ_{yz}, ρ_{yx}, and ρ_{xz} in all 8 domains.

(iii). The outcomes of Tables 1-2 exhibit that the mean square error and percent relative efficiency of the suggested direct and synthetic estimators $\bar{y}^d_{kp,a}$ and $\bar{y}^s_{kp,a}$ decrease and increase as the amounts of ρ_{yz}, ρ_{yx}, and ρ_{xz} reduce.

(iv). Furthermore, the outcomes of Table 1 and Table 2 demonstrate the better efficiency the suggested synthetic estimator over the proposed direct estimator for several combinations of correlation coefficients in each domain. This fact is supported by Corollary 2.1.

5. Real data applications

In this section, two real data sets are used for evaluating the suggested direct and synthetic estimators.

5.1. Data set 1

We have taken actual data of Swedish municipal statistics from the book of [18]. The Swedish municipalities are its smaller local governing units that are responsible to manage an important part of basic facilities, namely, schools, hospitals, emergency facilities, and planning department. There are total 284 municipalities in Sweden which are known as the MU284. It offers a lot of different features and sizes. Sweden is divided into 8 districts (domains), with sizes of 25, 48, 32, 38, 56, 41, 15, and 29 respectively: Stockholm, East Middle Sweden, Smaland and the islands, South Sweden, West Sweden, North Middle Sweden, Middle Norrland, and Upper Norrland. This study considers the 4 domains such as (1). East Middle Sweden, (2). Smaland and the islands, (3). North Middle Sweden, and (4). Middle Norrland, out of the above mentioned 8 domains. The municipalities are described in various ways by eight variables in the data set. We select $REV84$, $P75$, and $ME84$ from among these eight variables. In this data collection, the following study and supplementary variables are considered:
y: Real estate values in 1984 (REV84) evaluation (in millions of Kronor),
x: Population in 1975 (P75) (in '000'),
z: Municipality employees in 1984 (ME84),

Table 1
Simulated MSE and PRE of direct estimators for different values of correlation coefficients.

	ρ_{yx}	0.7		0.6		0.5		0.4		0.8		0.7		0.6		0.5	
	ρ_{yz}	0.8		0.7		0.6		0.5		0.7		0.6		0.5		0.4	
	ρ_{xz}	0.9		0.8		0.7		0.6		0.6		0.5		0.4		0.3	
Domains	Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	$\bar{y}^d_{m,a}$	10.02	100.00	10.00	100.00	10.62	100.00	10.64	100.00	10.18	100.00	10.26	100.00	10.25	100.00	10.24	100.00
	$\bar{y}^d_{r,a}$	7.07	141.6	8.76	114.11	9.28	114.40	10.93	97.33	4.47	227.95	7.36	139.50	9.40	109.00	11.46	89.35
	$\bar{y}^d_{bk,a}$	3.61	277.72	5.10	195.89	6.54	162.33	7.68	138.53	2.86	356.41	4.29	239.41	5.62	182.22	6.85	149.33
	$\bar{y}^d_{kp,a}$	3.59	279.39	5.07	197.20	6.34	167.56	7.29	146.00	2.85	356.64	4.26	240.69	5.55	184.69	6.71	152.64
2	$\bar{y}^d_{m,a}$	10.18	100.00	10.22	100.00	9.74	100.00	9.74	100.00	10.18	100.00	9.65	100.00	9.63	100.00	9.61	100.00
	$\bar{y}^d_{r,a}$	6.01	169.51	7.64	133.81	9.91	98.26	11.55	84.36	4.32	235.76	5.90	163.75	7.65	125.91	9.38	102.41
	$\bar{y}^d_{bk,a}$	3.62	281.00	5.14	198.90	6.18	157.66	7.22	134.88	2.87	355.00	4.16	231.79	5.44	176.96	6.60	145.59
	$\bar{y}^d_{kp,a}$	3.58	284.70	5.09	200.73	6.05	160.94	6.99	139.36	2.86	355.24	4.14	233.41	5.33	180.52	6.38	150.57
3	$\bar{y}^d_{m,a}$	10.12	100.00	10.13	100.00	10.27	100.00	10.28	100.00	10.36	100.00	10.03	100.00	10.02	100.00	10.03	100.00
	$\bar{y}^d_{r,a}$	7.41	136.61	9.30	108.99	11.00	93.34	12.88	79.84	4.52	229.14	6.69	149.76	8.67	115.63	10.63	94.28
	$\bar{y}^d_{bk,a}$	3.70	273.23	5.21	194.58	6.46	158.86	7.56	135.92	3.04	345.02	4.39	228.59	5.73	174.88	6.95	144.19
	$\bar{y}^d_{kp,a}$	3.69	274.14	5.16	196.50	6.32	162.50	7.31	140.72	3.00	345.29	4.35	230.26	5.63	178.14	6.75	148.63
4	$\bar{y}^d_{m,a}$	10.57	100.00	10.61	100.00	9.76	100.00	9.75	100.00	10.26	100.00	10.20	100.00	10.22	100.00	10.24	100.00
	$\bar{y}^d_{r,a}$	7.32	144.46	9.28	114.30	12.43	78.51	14.45	67.47	4.71	217.80	6.27	162.65	8.19	124.80	10.11	101.34
	$\bar{y}^d_{bk,a}$	3.73	283.25	5.28	201.03	6.26	155.74	7.31	133.44	2.97	345.88	4.46	228.49	5.84	174.86	7.11	144.13
	$\bar{y}^d_{kp,a}$	3.71	285.09	5.23	202.86	6.14	158.92	7.09	137.44	2.95	346.06	4.43	230.31	5.72	178.69	6.86	149.40
5	$\bar{y}^d_{m,a}$	10.43	100.00	10.43	100.00	10.73	100.00	10.75	100.00	10.44	100.00	10.64	100.00	10.65	100.00	10.65	100.00
	$\bar{y}^d_{r,a}$	6.68	156.06	8.47	123.17	10.50	102.17	12.24	87.80	5.22	199.97	6.73	158.10	8.77	121.46	10.81	98.57
	$\bar{y}^d_{bk,a}$	3.80	274.90	5.35	195.03	6.74	159.17	7.89	136.19	2.93	355.63	4.46	238.78	5.86	181.77	7.15	149.02
	$\bar{y}^d_{kp,a}$	3.76	277.06	5.29	197.21	6.59	162.73	7.61	141.15	2.92	355.79	4.43	240.36	5.75	185.26	6.92	153.86
6	$\bar{y}^d_{m,a}$	10.00	100.00	10.32	100.00	9.86	100.00	9.83	100.00	10.31	100.00	9.86	100.00	9.81	100.00	9.77	100.00
	$\bar{y}^d_{r,a}$	7.00	147.00	8.71	118.52	9.95	99.12	11.70	84.00	4.35	237.30	7.32	134.83	9.34	105.04	11.36	85.98
	$\bar{y}^d_{bk,a}$	4.00	294.83	4.99	206.92	6.04	163.26	7.07	138.88	2.82	368.03	4.10	240.51	5.37	182.85	6.53	149.72
	$\bar{y}^d_{kp,a}$	3.00	297.05	4.96	207.93	5.89	167.39	6.79	144.69	2.80	368.72	4.08	241.86	5.29	185.35	6.39	153.01
7	$\bar{y}^d_{m,a}$	10.16	100.00	10.13	100.00	10.39	100.00	10.37	100.00	10.56	100.00	10.16	100.00	10.14	100.00	10.13	100.00
	$\bar{y}^d_{r,a}$	8.40	120.88	10.60	95.61	12.78	81.31	14.88	69.66	4.36	242.08	5.99	169.58	7.82	129.64	9.64	105.12
	$\bar{y}^d_{bk,a}$	3.58	283.89	5.04	201.01	6.31	164.78	7.40	140.20	3.04	346.82	4.42	229.63	5.77	175.60	7.00	144.67
	$\bar{y}^d_{kp,a}$	3.56	285.09	5.00	202.88	6.20	167.52	7.21	143.79	3.02	347.04	4.39	231.49	5.64	179.67	6.74	150.36
8	$\bar{y}^d_{m,a}$	10.14	100.00	10.12	100.00	10.01	100.00	9.98	100.00	10.08	100.00	10.10	100.00	10.08	100.00	10.06	100.00
	$\bar{y}^d_{r,a}$	7.57	133.93	9.44	107.17	11.15	89.77	13.12	76.03	5.92	170.38	8.82	114.45	11.43	88.14	14.06	71.51
	$\bar{y}^d_{bk,a}$	3.46	293.39	4.91	205.88	6.10	164.13	7.15	139.61	2.79	361.92	4.19	240.86	5.50	183.21	6.71	149.99
	$\bar{y}^d_{kp,a}$	3.42	296.12	4.89	206.97	5.97	167.53	6.91	144.41	2.76	362.22	4.16	242.55	5.42	186.01	6.55	153.46

Table 3 displays the parameters of the domain for data set 1. The mean square error and percent relative efficiency of different direct and synthetic estimators are presented in Tables 4-5 by adapting the below equations:

$$PRE(\bar{y}^d_{m,a}, \bar{y}^d_{*,a}) = \frac{MSE(\bar{y}^d_{m,a})}{MSE(\bar{y}^d_{*,a})} \times 100 \tag{9}$$

$$PRE(\bar{y}^s_{m,a}, \bar{y}^s_{*,a}) = \frac{MSE(\bar{y}^s_{m,a})}{MSE(\bar{y}^s_{*,a})} \times 100 \tag{10}$$

5.2. Data set 2

In order to collect taxes and carry out other administrative duties, Uttar Pradesh, similar to the other Indian states, is separated into various districts. Each district is separated further into several revenue inspector circles (RICs), which consists of numerous villages. The RICs are viewed as domains of small size in the current work.

It is observed that the acreage used to grow a specific crop fluctuates each year, either growing larger or smaller. The agricultural acreage assessment issue for the RICs of the Uttar Pradesh's Mohanlal Ganj tehsil is thus taken into consideration for real data application. Sisendi, Amethi, Mohanlal Ganj, Nigoha, Khujauli, Nagram, Gosaiganj, and Behrauli are the 8 RICs of Mohanlal Ganj tehsil that we consider as small domains. We selected 4 of these 8 domains to focus on in our analysis, namely, (1). Sisendi, (2). Khujauli, (3). Gosaiganj, and (4). Behrauli. The crop acreage of paddy (measured in hectares) during the 2018-19 cultivation period is regarded as a research variable y. As supplementary variables x and z, the crop acreage of paddy for the cultivation periods 2017-18 and 2016-17, respectively, is taken into consideration. Table 6 shows the parameters of all domains for quick reference. With the use of the formulas (9) and (10), respectively, we computed the MSE and PRE of the suggested direct and synthetic logarithmic estimators using the domain parameters, and the results are shown in Tables 7-8.

Table 2
Simulated MSE and PRE of synthetic estimators for different values of correlation coefficients.

	ρ_{yx}	0.7		0.6		0.5		0.4		0.8		0.7		0.6		0.5	
	ρ_{yz}	0.8		0.7		0.6		0.5		0.7		0.6		0.5		0.4	
	ρ_{xz}	0.9		0.8		0.7		0.6		0.6		0.5		0.4		0.3	
Domains	Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	$\bar{y}^s_{m,a}$	11.196	100.00	10.853	100.00	15.711	100.00	15.675	100.00	12.537	100.00	13.91	100.00	13.754	100.00	13.587	100.00
	$\bar{y}^s_{r,a}$	1.007	1111.31	1.236	878.06	0.937	1676.90	1.097	1429.50	0.474	2645.22	1.065	1305.60	1.368	1005.47	1.667	815.29
	$\bar{y}^s_{bk,a}$	0.544	2059.00	0.753	1440.67	0.598	2625.65	0.701	2237.26	0.308	4072.85	0.699	1990.49	0.911	1510.45	1.102	1233.24
	$\bar{y}^s_{kp,a}$	0.542	2059.67	0.751	1441.97	0.597	2632.79	0.698	2246.64	0.307	4072.99	0.698	1991.89	0.909	1513.31	1.098	1237.09
2	$\bar{y}^s_{m,a}$	13.453	100.00	13.380	100.00	10.312	100.00	10.195	100.00	10.94	100.00	13.79	100.00	14.054	100.00	14.261	100.00
	$\bar{y}^s_{r,a}$	0.690	1949.93	0.866	1545.78	1.413	729.63	1.631	625.24	0.646	1693.98	0.657	2098.46	0.84	1673.73	1.021	1397.11
	$\bar{y}^s_{bk,a}$	0.372	3620.46	0.527	2541.06	0.906	1138.05	1.046	974.91	0.421	2604.62	0.431	3203.65	0.558	2517.87	0.674	2116.44
	$\bar{y}^s_{kp,a}$	0.370	3621.93	0.526	2543.51	0.904	1141.03	1.041	978.93	0.420	2604.66	0.430	3206.31	0.557	2522.93	0.672	2123.21
3	$\bar{y}^s_{m,a}$	11.236	100.00	10.893	100.00	15.814	100.00	15.785	100.00	12.529	100.00	13.997	100.00	13.843	100.00	13.678	100.00
	$\bar{y}^s_{r,a}$	1.009	1113.39	1.238	880.09	0.933	1695.14	1.092	1445.66	0.474	2644.59	1.069	1309.23	1.373	1008.30	1.673	817.57
	$\bar{y}^s_{bk,a}$	0.544	2063.57	0.754	1444.43	0.596	2651.73	0.698	2260.59	0.308	4071.12	0.701	1995.41	0.914	1514.32	1.106	1236.42
	$\bar{y}^s_{kp,a}$	0.542	2064.24	0.753	1445.73	0.595	2658.95	0.695	2270.07	0.308	4071.25	0.700	1996.82	0.912	1517.18	1.103	1240.29
4	$\bar{y}^s_{m,a}$	10.929	100.00	10.776	100.00	14.046	100.00	14.148	100.00	11.051	100.00	13.589	100.00	13.69	100.00	13.747	100.00
	$\bar{y}^s_{r,a}$	0.820	1332.27	1.051	1025.61	1.692	830.01	1.988	711.50	0.648	1704.25	0.675	2012.13	0.869	1576.02	1.062	1294.39
	$\bar{y}^s_{bk,a}$	0.442	2474.57	0.639	1686.39	1.086	1293.44	1.276	1108.39	0.422	2621.3	0.442	3073.68	0.577	2372.11	0.701	1961.74
	$\bar{y}^s_{kp,a}$	0.440	2475.48	0.638	1687.96	1.083	1296.78	1.271	1112.93	0.420	2621.34	0.441	3076.20	0.576	2376.85	0.699	1967.99
5	$\bar{y}^s_{m,a}$	13.144	100.00	13.114	100.00	11.178	100.00	11.047	100.00	12.362	100.00	12.170	100.00	12.053	100.00	11.942	100.00
	$\bar{y}^s_{r,a}$	0.706	1862.67	0.883	1485.97	1.438	777.30	1.658	666.10	0.692	1786.17	0.728	1670.62	0.946	1274.39	1.165	1024.65
	$\bar{y}^s_{bk,a}$	0.382	3462.48	0.536	2445.5	0.921	1213.36	1.063	1039.44	0.451	2745.18	0.477	2552.61	0.628	1918.38	0.769	1553.00
	$\bar{y}^s_{kp,a}$	0.379	3463.86	0.535	2447.85	0.919	1216.53	1.058	1043.72	0.450	2746.22	0.476	2554.65	0.627	1922.16	0.766	1557.93
6	$\bar{y}^s_{m,a}$	11.828	100.00	11.603	100.00	11.882	100.00	11.713	100.00	11.709	100.00	12.678	100.00	12.610	100.00	12.529	100.00
	$\bar{y}^s_{r,a}$	1.033	1144.72	1.280	906.38	1.07	1110.58	1.260	929.49	0.498	2352.42	1.036	1223.65	1.335	944.60	1.632	767.88
	$\bar{y}^s_{bk,a}$	0.558	2121.49	0.781	1487.38	0.685	1734.95	0.807	1451.43	0.323	3621.10	0.681	1864.93	0.889	1418.54	1.079	1161.15
	$\bar{y}^s_{kp,a}$	0.557	2122.17	0.778	1488.72	0.683	1739.61	0.804	1457.48	0.322	3621.20	0.679	1866.25	0.887	1421.23	1.076	1164.79
7	$\bar{y}^s_{m,a}$	10.186	100.00	10.191	100.00	17.793	100.00	18.342	100.00	10.736	100.00	15.731	100.00	16.189	100.00	16.552	100.00
	$\bar{y}^s_{r,a}$	0.871	1169.15	1.130	901.76	1.845	964.16	2.187	838.54	0.617	1738.68	0.623	2524.04	0.793	2040.52	0.962	1720.92
	$\bar{y}^s_{bk,a}$	0.472	2167.16	0.689	1479.82	1.185	1501.98	1.405	1305.71	0.401	2674.66	0.408	3855.39	0.527	3071.35	0.635	2608.47
	$\bar{y}^s_{kp,a}$	0.470	2168.00	0.687	1481.19	1.182	1505.83	1.399	1311.03	0.400	2674.70	0.407	3858.66	0.526	3077.55	0.633	2616.84
8	$\bar{y}^s_{m,a}$	10.573	100.00	10.465	100.00	10.08	100.00	10.057	100.00	10.743	100.00	13.896	100.00	14.867	100.00	15.763	100.00
	$\bar{y}^s_{r,a}$	0.965	1095.42	1.199	872.79	1.284	785.24	1.504	668.84	0.645	1664.70	1.071	1297.26	1.418	1048.46	1.774	888.76
	$\bar{y}^s_{bk,a}$	0.521	2030.43	0.732	1432.45	0.823	1225.23	0.964	1043.17	0.421	2559.18	0.703	1976.62	0.944	1574.05	1.173	1343.50
	$\bar{y}^s_{kp,a}$	0.520	2031.11	0.730	1433.76	0.821	1228.46	0.960	1047.49	0.420	2559.22	0.702	1978.01	0.943	1577.03	1.170	1347.70

Table 3
Population characteristics for various domains in data set 1.

Domains	N_a	$\bar{Y}_a$	$\bar{X}_a$	$\bar{Z}_a$	S_{y_a}	S_{x_a}	S_{z_a}	$\rho_{y_a x_a}$	$\rho_{y_a z_a}$	$\rho_{x_a z_a}$
1	48	2971.10	29.17	1658.71	3334.66	35.05	2145.20	0.96	0.97	0.99
2	32	2498.75	23.94	1317.03	2040.72	20.91	1410.55	0.95	0.93	0.95
3	41	2175.32	20.98	1099.76	1693.82	17.35	1010.17	0.98	0.98	0.99
4	15	3648.47	26.60	1533.87	2410.56	24.12	1482.13	0.84	0.84	0.99

Table 4
Direct estimators' MSE and PRE for data set 1.

Domains	$\bar{y}^d_{m,a}$		$\bar{y}^d_{r,a}$		$\bar{y}^d_{bk,a}$		$\bar{y}^d_{kp,a}$	
Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	880332.20	100.00	1438946.00	61.18	55735.76	1579.47	54653.96	1610.74
2	563945.70	100.00	1163415.00	48.47	51340.73	1098.44	51017.86	1105.39
3	288651.80	100.00	472259.40	61.12	12866.29	2243.47	12546.50	2300.66
4	1549549.00	100.00	6607974.00	23.45	461387.00	335.85	440619.30	351.68

5.3. Main findings of real data applications

The main findings of real data applications have been interpreted in pointwise format to provide a thorough grasp of the merits of the suggested estimators.

Table 5
Synthetic estimators' MSE and PRE for data set 1.

Domains	$\bar{y}^s_{m,a}$		$\bar{y}^s_{r,a}$		$\bar{y}^s_{bk,a}$		$\bar{y}^s_{kp,a}$	
Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	334172.80	100.00	953072.60	35.06	14950.62	2235.18	12819.51	2606.75
2	657827.60	100.00	674117.90	97.58	10574.73	6220.75	8924.72	7370.85
3	1136826.00	100.00	510899.40	222.51	8014.36	14184.87	6658.16	17074.17
4	648822.30	100.00	1437180.00	45.15	22544.69	2877.94	19616.01	3307.62

Table 6
Population characteristics for various domains in data set 2.

Domains	N_a	$\bar{Y}_a$	$\bar{X}_a$	$\bar{Z}_a$	S_{y_a}	S_{x_a}	S_{z_a}	$\rho_{y_a x_a}$	$\rho_{y_a z_a}$	$\rho_{x_a z_a}$
1	18	224.11	237.88	234.27	106.62	100.91	102.68	0.95	0.90	0.95
2	32	106.41	112.16	115.16	64.57	68.28	69.91	0.99	0.97	0.98
3	24	87.32	88.44	89.32	58.59	58.01	59.12	0.99	0.96	0.97
4	36	111.39	109.17	109.78	86.05	82.50	80.34	0.96	0.97	0.98

Table 7
The direct estimators' MSE and PRE for data set 2.

Domains	$\bar{y}^d_{m,a}$		$\bar{y}^d_{r,a}$		$\bar{y}^d_{bk,a}$		$\bar{y}^d_{kp,a}$	
Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	2210.49	100.00	1885.22	117.25	209.45	1055.40	189.17	1168.54
2	564.67	100.00	586.76	96.23	10.79	5235.47	6.58	8583.23
3	389.40	100.00	379.02	102.74	9.12	4269.97	5.14	7573.83
4	852.08	100.00	798.07	106.77	39.95	2132.73	35.64	2391.12

Table 8
The synthetic estimators' MSE and PRE for data set 2.

Domains	$\bar{y}^s_{m,a}$		$\bar{y}^s_{r,a}$		$\bar{y}^s_{bk,a}$		$\bar{y}^s_{kp,a}$	
Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
1	9478.01	100.00	559.56	1693.82	18.22	52018.71	16.95	55926.06
2	643.50	100.00	126.14	510.14	4.11	15666.83	3.85	16702.09
3	1821.49	100.00	84.95	2144.09	2.77	65846.74	2.60	69963.71
4	455.82	100.00	138.23	329.75	4.50	10126.99	4.22	10803.70

(i). Tables 4-5 present, respectively, the MSEs and PREs of the direct and synthetic estimators consisting of data set 1. Table 4's findings exhibit that the direct logarithmic estimator $\bar{y}^d_{kp,a}$ achieves the lowest MSEs and highest PREs among the conventional direct estimators, including direct unbiased estimator $\bar{y}^d_{m,a}$, direct ratio estimator $\bar{y}^d_{r,a}$, and direct logarithmic estimator $\bar{y}^d_{bk,a}$. Hence, the proposed direct estimator $\bar{y}^d_{kp,a}$ surpasses the usual direct estimators. The findings shown in Table 5 exhibit that the suggested synthetic estimator $\bar{y}^s_{kp,a}$ achieves the lowest MSEs and highest PREs among the current synthetic estimators like synthetic unbiased estimator $\bar{y}^s_{m,a}$, synthetic ratio estimator $\bar{y}^s_{r,a}$, and synthetic logarithmic estimator $\bar{y}^s_{bk,a}$. Hence, the proposed synthetic estimator $\bar{y}^s_{kp,a}$ surpasses the classical synthetic estimators. Additionally, Corollary 2.1 makes it easier for the suggested synthetic estimator $\bar{y}^s_{kp,a}$ to surpass the proposed direct estimator $\bar{y}^d_{kp,a}$ in every domain.

(ii). Tables 7-8 present, respectively, the MSEs and PREs of the direct and synthetic estimators consisting of data set 2. The findings of Table 7 exhibit that the suggested direct estimator $\bar{y}^d_{kp,a}$ achieves the lowest MSEs and highest PREs out of the conventional direct estimators, including direct unbiased estimator $\bar{y}^d_{m,a}$, direct ratio estimator $\bar{y}^d_{r,a}$, and direct logarithmic estimator $\bar{y}^d_{bk,a}$. Hence, the proposed direct estimator $\bar{y}^d_{kp,a}$ surpasses the available direct estimators. The findings summarized in Table 8 exhibit that the suggested synthetic estimator $\bar{y}^s_{kp,a}$ achieves the lowest MSE and highest PRE among the current synthetic estimators like synthetic unbiased estimator $\bar{y}^s_{m,a}$, synthetic ratio estimator $\bar{y}^s_{r,a}$, and synthetic logarithmic estimator $\bar{y}^s_{bk,a}$. Hence, the suggested synthetic estimator $\bar{y}^s_{kp,a}$ outperforms the traditional synthetic estimators. Additionally, Corollary 2.1 makes it easier for the suggested synthetic estimator $\bar{y}^s_{kp,a}$ to outperform the proposed direct estimator $\bar{y}^d_{kp,a}$ in every domain.

6. Conclusions

In the present article, we have suggested efficient direct and synthetic logarithmic type estimators for the domain mean under the SRS framework. The suggested estimators' MSEs are calculated, compared with the current estimators, and efficiency conditions are devised. Additionally, a simulation that employed an artificially created normal population supported the theoretical findings. The findings of simulation are demonstrated with MSEs and PREs in Tables 1-2. The key results of the simulation analysis are summarized in Subsection 4.1, from which we conclude that the offered direct and synthetic estimators are more efficient than direct and synthetic estimators currently in practice, respectively. Furthermore, real data sets from municipalities of Sweden and the area planted with paddy in Mohanlal Ganj tehsil, Uttar Pradesh, India, were used to illustrate the uses of the suggested methodologies. Tables 4, 5, 7, and 8 present the outcomes of the direct and synthetic estimators based on real populations and also demonstrate the supremacy of the suggested estimators over the traditional estimators. As a consequence, we may suggest using the direct and synthetic estimators to estimate the domain means of small regions.

CRediT authorship contribution statement

Anoop Kumar: Writing – review & editing, Writing – original draft, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Shashi Bhushan:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Rohini Pokhrel:** Validation, Software, Resources, Project administration, Funding acquisition, Formal analysis, Data curation. **Walid Emam:** Writing – review & editing, Funding acquisition. **Yusra Tashkandy:** Project administration, Funding acquisition. **M.J.S. Khan:** Writing – review & editing, Visualization, Resources.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Walid Emam reports financial support was provided by King Saud University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data included in article/supplementary material/referenced in article.

Acknowledgements

The study was funded by Researchers Supporting Project number (RSP2024R488), King Saud University, Riyadh, Saudi Arabia.

References

[1] J.N.K. Rao, I. Molina, Small Area Estimation, John Wiley & Sons, 2015.
[2] B.B. Khare, Ashutosh, S. Khare, Modified direct regression estimator for domain mean using auxiliary character, Int. J. Appl. Math. Stat. 57 (6) (2018) 10–20.
[3] S. Bhushan, A. Kumar, R. Pokhrel, Improved direct estimators for domain mean with an application to municipalities data, Int. J. Agricultur. Stat. Sci. 19 (2) (2023) 835–841.
[4] G.C. Tikkiwal, A. Ghiya, A generalized class of synthetic estimators with application to crop acreage estimation for small domains, Biom. J. 42 (2000) 865–876.
[5] K.K. Pandey, G.C. Tikkiwal, Generalized class of synthetic estimators for small area under systematic sampling design, Stat. Trans. N.S., Poland 11 (1) (2010) 75–89.
[6] G.C. Tikkiwal, P.K. Rai, A. Ghiya, On the performance of generalized regression estimator for small domains, Commun. Stat., Simul. Comput. 42 (4) (2013) 891–909.
[7] P.K. Rai, K.K. Pandey, Synthetic estimators using auxiliary information in small domains, Stat. Trans. N.S. 14 (1) (2013) 31–44.
[8] S. Bhushan, A. Kumar, R. Pokhrel, Logarithmic type direct and synthetic estimators for domain mean using simple random sampling, Rev. Colomb. Estad. 47 (1) (2024) 111–133.
[9] A. Kumar, S. Bhushan, R. Pokhrel, W. Emam, Small area estimation using design based direct and synthetic logarithmic estimators, Ain Shams Eng. J. (2024), https://doi.org/10.1016/j.asej.2024.102836.
[10] N. Tzavidis, L.C. Zhang, A. Luna, T. Schmid, N. Rojas-Perilla, From start to finish: a framework for the production of small area official statistics, J. R. Stat. Soc., Ser. A, Stat. Soc. 181 (4) (2018) 927–979.
[11] N. Rojas-Perilla, S. Pannier, T. Schmid, N. Tzavidis, Data-driven transformations in small area estimation, J. R. Stat. Soc., Ser. A, Stat. Soc. 183 (1) (2020) 121–148.
[12] S. Kumar, B.V.S. Sisodia, D. Singh, P. Basak, Calibration approach based estimation of finite population total in survey sampling under super population model when study variable and auxiliary variable are inversely related, J. Rel. Stat. Stud. 10 (2) (2017) 83–93.
[13] S. Guha, U.C. Sud, B.V.S. Sisodia, Calibration approach based chain ratio-product type estimator involving two auxiliary variables in two phase sampling, J. Indian Soc. Agric. Stat. 72 (3) (2018) 179–186.
[14] B.B. Khare, Ashutosh, Simulation study of the generalized synthetic estimator for domain mean in the sample survey, Int. J. Tomogr. Stat. 31 (3) (2018) 87–100.
[15] S. Bhushan, A. Kumar, R. Pokhrel, Logarithmic type direct and synthetic estimators using bivariate auxiliary information with an application to real data, J. Indian Soc. Agric. Stat. 77 (1) (2023) 133–148.
[16] S. Bhushan, A. Kumar, New efficient logarithmic estimators using multiauxiliary information under ranked set sampling, Concurr. Comput., Pract. Exp. 34 (27) (2022) e7337.
[17] M.N. Qureshi, Y. Faizan, A. Shetty, M.H. Ahelali, M. Hanif, O.A. Alamri, Ln-type estimators for the estimation of the population mean of a sensitive study variable using auxiliary information, Heliyon 10 (1) (2024) e23066.
[18] C.E. Sarndal, B. Swensson, J. Wretman, Model Assisted Survey Sampling, Springer-Verlag, New York, 2003.