Heliyon 10 (2024) e40878

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Optimizing population mean estimation in stratified sampling using linear cost: A simulation study

Poonam Singh [a], Prayas Sharma [b,*], Rajesh Singh [a], Badr Aloraini [c], Aysha Akhtar [d]

[a] Department of Statistics, Banaras Hindu University, Varanasi, India
[b] Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, India
[c] Department of Mathematics, Shaqra University, Shaqra, Saudi Arabia
[d] Department of Mathematics, Bahria Foundation College, Peshawar Road Campus, Rawalpindi, Pakistan

ARTICLE INFO

Keywords:
Auxiliary information
Cost
Lagrange's multiplier
Integer programming problems
Optimization
Percentage relative efficiency (PRE)
Mean square error (MSE)
Stratified sampling
Simulation

ABSTRACT

Improving efficiency has long been a focal challenge in sampling literature. However, simultaneously enhancing estimator efficacy and optimizing survey costs is a practical necessity across various fields such as medicine, agriculture, and transportation. In this study, we present a comprehensive family of generalized exponential estimators specifically designed for estimating population means within stratified sampling frameworks. Optimizing the survey cost is one the major challenges in the stratified sampling because the cost of the survey is fixed and decided before the survey. To optimize survey costs, we employ integer programming and Lagrange multipliers. We have carefully derived the Mean Square Error (MSE) of the proposed estimators and addressed this as an optimization problem to further refine estimator performance in light of cost constraints and optimal sample sizes. The results have been rigorously validated using real-world datasets, and both theoretical and empirical evaluations show that the proposed estimators significantly outperform existing alternatives. These findings underscore the estimators' practical relevance and theoretical robustness.

1. Introduction

Stratified sampling is often used in research and surveys because it makes population figures more accurate than simple random sampling, especially when the population is not all the same. Stratified sampling is a method used in poll research and statistical analysis where a population is split into separate groups, or strata, and samples are taken from each group separately. This method works especially well when the population is heterogeneous because it makes population figures like the mean or average more accurate by taking into account differences within and between groups (Singh et al., 1996).

Extra information about the population, which is sometimes called an "auxiliary variable," is often available in real life. These factors are linked to the variable of interest and can be used to make estimators work better. Researchers can get more accurate predictions of population statistics without raising the sample size by using extra information during the estimating process.

[1] pioneered the discussion of the ratio estimation method utilizing auxiliary information. Building on this foundation, numerous researchers have expanded on this approach. Key contributions have been made by Ref. [2–4], and more recently [5]. These scholars have collectively advanced the field of population parameter estimation through the innovative use of auxiliary data. Stratified

* Corresponding author.
E-mail address: prayassharma02@gmail.com (P. Sharma).

https://doi.org/10.1016/j.heliyon.2024.e40878
Received 16 October 2024; Received in revised form 1 December 2024; Accepted 2 December 2024
Available online 3 December 2024
2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

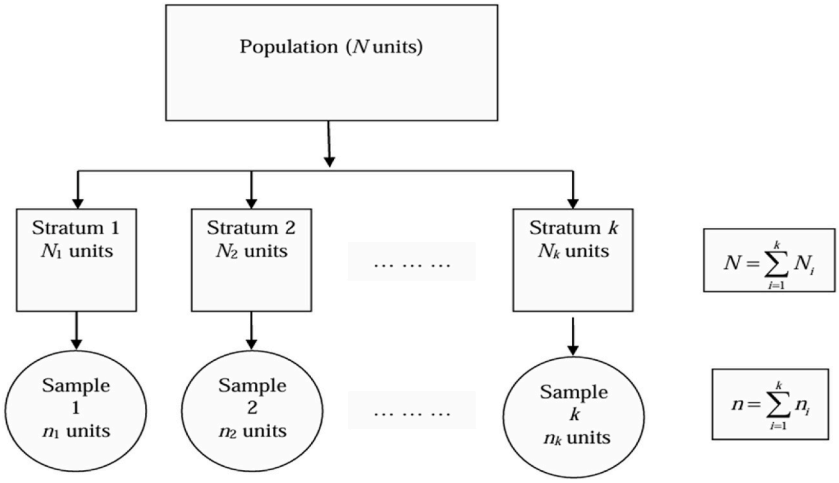

Fig. 1. Samples using Stratified Sampling.

sampling is a widely utilized sampling design, particularly effective for handling heterogeneous data. It typically requires a smaller sample size compared to Simple Random Sampling while offering greater precision in estimates. The authors [6–13] have contributed seminal works to the literature on stratified sampling, offering foundational insights and advancing key theoretical frameworks in the field However, limitations like time, money, and resource shortages are common in real-world sampling. In these situations, figuring out the best way to divide the overall sample size across the strata is essential to getting precise estimates while keeping expenses to a minimum.

This problem is generally approached in two primary ways.

(i) Minimizing variance for a fixed cost.
(ii) Minimizing cost for a fixed variance.

[14] was the pioneer in addressing the problem of optimal sample allocation in stratified sampling. Since then, significant contributions have been made by various researchers, including [15–21]. Notwithstanding its promise, careful methodological design and optimization are necessary for the efficient use of auxiliary variables in stratified sampling when budgetary limitations are present.

The **motivation** for this work lies in addressing the following challenges.

1. Heterogeneity Across Strata: The sizes, costs of data gathering, and degrees of variability may differ throughout strata. If these distinctions are ignored, the allocation may be inefficient, either under sampling highly variable strata or oversampling less variable ones.
2. Financial Limitations: In actuality, the cost of sampling often increases in direct proportion to the sample size. One such restriction is that the whole cost of sampling cannot go beyond a certain spending limit. Sampling procedures must thus strike a compromise between statistical accuracy and cost effectiveness.
3. Precision Maximization: By include auxiliary variables in the estimation procedure, stratified sampling aims to reduce the variance of the population mean estimate. The advantages of stratified sampling, such as lower variance as compared to ordinary random sampling, could not be fully realized if resources are not allocated appropriately.

The study attempts to give an allocation technique that minimizes the variation of the population mean estimate while remaining within budget by using the cost and variance features of each stratum. In domains like economics, medicine, and environmental studies, where effective resource usage is essential for extensive surveys and research initiatives, this optimization has immediate applications.

In order to do this, we provide a new family of generalized exponential estimators for population mean estimation that includes auxiliary variables and an economical sampling technique for the best distribution across strata. To estimate the population mean more precisely, the suggested technique makes use of the link between the study and auxiliary variables. Our method not only reduces the estimate's variance within financial limits, but it also shows how useful auxiliary data can be in improving stratified sampling's effectiveness. To find the ideal sample sizes for every stratum, we frame an optimization problem and use the Lagrange multiplier approach. We compare the MSEs of the suggested and current estimators at these optimized sample sizes after applying this optimization approach to both.

A thorough framework for enhancing the planning and execution of stratified sampling in resource-constrained environments is provided by this combined emphasis on cost optimization and auxiliary variable use.

Let us consider a population having N units. It is divided into L homogeneous subgroups called strata and k^{th} strata consist of N_k units where $k = 1, 2, 3, \ldots, L$ such that $\sum_{k=1}^{L} N_k = N$. Then n_k sample units are drawn from the k^{th} stratum by simple random sampling without replacement scheme. And $\sum_{k=1}^{L} n_k = n$. Fig. 1 shows how samples are drawn from the population using stratified sampling.

Let.

Y: be the study variable.

X: be the auxiliary variable.

$\overline{Y}_k$: Population mean of the study variable Y for the k^{th} stratum.

$\overline{X}_k$: Population mean of the auxiliary variable X for the k^{th} stratum.

$\overline{y}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki}$: Sample means of study variable from the k^{th} stratum and $\overline{y}_{ki}$ be the i^{th} unit in the k^{th} stratum.

$\overline{x}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}$: Sample means of auxiliary variable from the k^{th} stratum and x_{ki} be the i^{th} unit in the k^{th} stratum.

Then let us define the approximations as:

$$\xi_{k0} = \frac{\overline{y}_k}{\overline{Y}_k} - 1 \text{ and } \xi_{k1} = \frac{\overline{x}_k}{\overline{X}_k} - 1$$

Such that $E(\xi_{k0}) = E(\xi_{k1}) = 0$

$$E\left(\xi_{k0}^2\right) = \left(\frac{N_k - n_k}{N_k n_k}\right) C_{ky}^2$$

$$E\left(\xi_{k1}^2\right) = \left(\frac{N_k - n_k}{N_k n_k}\right) C_{kx}^2$$

$$E(\xi_0 \xi_1) = \left(\frac{N_k - n_k}{N_k n_k}\right) \rho_{kxy} C_{kx} C_{ky}$$

where

$$\overline{Y}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} Y_{ki}\ ;\ \overline{X}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} X_{ki};\ S_{ky}^2 = \frac{1}{(N_k-1)}\sum_{i=1}^{N_k}(Y_{ki} - \overline{Y})^2\ ;\ S_{kx}^2 = \frac{1}{(N_k-1)}\sum_{i=1}^{N_k}(X_{ki} - \overline{X})^2$$

$$S_{kxy} = \frac{1}{(N_k-1)}\sum_{i=1}^{N_k}(Y_{ki} - \overline{Y})(X_{ki} - \overline{X})\ ;\ C_{ky}^2 = \frac{S_{ky}^2}{\overline{Y}_k^2}\ ;\ C_{kx}^2 = \frac{S_{kx}^2}{\overline{X}_k^2}\ ;\ \rho_{kxy} = \frac{S_{kxy}}{S_{kx}S_{ky}}\ ;\ f_k = \left(\frac{N_k - n_k}{N_k n_k}\right)$$

2. Existing estimators

Separate Usual mean estimator T_0 of Population mean $\overline{Y}$

$$T_0 = \sum_{k=1}^{L} W_k \overline{y}_k \tag{1}$$

Where $W_k = \frac{N_k}{N}$ is the known proportion of population units.

The Variance of the estimator T_0 is given by:

$$V(T_0) = \sum_{k=1}^{L} W_k^2 \left(\frac{N_k - n_k}{N_k n_k}\right) \overline{Y}_k^2 C_{ky}^2 \tag{2}$$

Separate Ratio estimator T_1 of Population mean $\overline{Y}$

$$T_1 = \sum_{k=1}^{L} W_k \overline{y}_k \left(\frac{\overline{X}_k}{\overline{x}_k}\right) \tag{3}$$

The Mean Square Error of the estimator T_1 is:

$$MSE(T_1) = \sum_{k=1}^{L} W_k^2 \overline{Y}_k^2 \left(\frac{N_k - n_k}{N_k n_k}\right)\left(C_{ky}^2 + C_{kx}^2 - 2\rho_{kxy} C_{kx} C_{ky}\right) \tag{4}$$

Separate Exponential Estimator T_2 of Population mean $\overline{Y}$

$$T_2 = \sum_{k=1}^{L} W_k \overline{y}_k e^{\left(\frac{\overline{X}_k - \overline{x}_k}{\overline{X}_k + \overline{x}_k}\right)} \tag{5}$$

The Mean Square Error of T_2 is:

$$MSE(T_2) = \sum_{k=1}^{L} W_k^2 \overline{Y}_k^2 \left(\frac{N_k - n_k}{N_k n_k}\right)\left(C_{ky}^2 + \left(\frac{1}{4}\right)C_{kx}^2 - \rho_{kxy} C_{kx} C_{ky}\right) \tag{6}$$

Separate Regression estimator T_{reg} of population mean $\overline{Y}$

$$T_{reg} = \sum_{k=1}^{L} W_k(\overline{y}_k + \beta_k(\overline{X}_k - \overline{x}_k)) \tag{7}$$

The Mean Square Error of T_{reg} is:

$$Min.MSE(T_{reg}) = \sum_{k=1}^{L} W_k^2 \left(\frac{N_k - n_k}{N_k n_k}\right)\overline{Y}_k^2\left(1 - \rho_{kxy}^2\right)C_{ky}^2 \tag{8}$$

Estimators given in equations (1), (3), (5) and (7) are the existing estimators considered in this study and equations (2), (4), (6) and (8) represents their MSE expression respectively.

3. Proposed estimators

Motivated by Ref. [5] we proposed generalized exponential estimator T_3 for estimating population mean $\overline{Y}$ under stratified sampling:

$$T_3 = \sum_{k=1}^{L} W_k \overline{y}_k a^{\left(\frac{\overline{X}_k - \overline{x}_k}{\overline{X}_k + \overline{x}_k}\right)} \text{ Where } a > 0. \tag{9}$$

Estimator T_3 is same as estimator T_2 for a = 2.718
Using approximations we write equation (9) as:

$$T_3 = \sum_{k=1}^{L} W_k \overline{Y}_k (1 + \xi_{ko}) a^{\left(\frac{\overline{X}_k - \overline{X}_k(1+\xi_{k1})}{\overline{X}_k + \overline{X}_k(1+\xi_{k1})}\right)} \tag{10}$$

$$= \sum_{k=1}^{L} W_k \overline{Y}_k (1 + \xi_{ko}) a^{\left[-\frac{\xi_{k1}}{2}\left(1+\frac{\xi_{k1}}{2}\right)^{-1}\right]}$$

Expanding $\left(1 + \frac{\xi_{k1}}{2}\right)^{-1}$ and ignoring higher order terms as they become very small.
We write equation (10) as:

$$T_3 = \sum_{k=1}^{L} W_k \overline{Y}_k \left(1 + \xi_{ko} - \frac{\xi_{k1}}{2} log\, a + \frac{\xi_{k1}^2}{4} log\, a + \frac{\xi_{ko}\xi_{k1}}{2} log\, a + \frac{\xi_{k1}^2}{8}(log\, a)^2\right) \tag{11}$$

Subtracting $\overline{Y} = \sum_{k=1}^{L} W_k \overline{Y}_k$ from both the sides of equation (11) and squaring it we get:

$$(T_3 - \overline{Y})^2 = \sum_{k=1}^{L} W_k^2 \overline{Y}_k^2 \left(\xi_{ko}^2 + \frac{\xi_{k1}^2}{4}(log\, a)^2 - \xi_{ko}\xi_{k1}\, log\, a\right) \tag{12}$$

Taking expectation on both the sides of equation (12)

$$MSE(T_3) = E(T_3 - \overline{Y})^2 = \sum_{k=1}^{L} W_k^2 \left(\frac{N_h - n_h}{N_h n_h}\right)\overline{Y}_k^2 \left(C_{ky}^2 + \frac{C_{kx}^2}{4}(log\, a)^2 - \rho_{kxy} C_{kx} C_{ky}\, log\, a\right) \tag{13}$$

To get Min. MSE we Differentiate equation (13) with respect to log a and equate it to zero

$$a_k = exp\left(\frac{2\rho_{kxy} C_{ky}}{C_{kx}}\right) \tag{14}$$

Substituting this value given in equation (14) in MSE expression we get the min. MSE of estimator T_3 as:

$$Min.MSE(T_3) = \sum_{k=1}^{L} W_k^2 \left(\frac{N_k - n_k}{N_k n_k}\right)\overline{Y}_k^2\left(1 - \rho_{kxy}^2\right)C_{ky}^2 \tag{15}$$

Min. MSE (T_3) expression given in equation (15) is same as MSE expression of the regression estimator given in equation (8).

Table 1
Showing members of generalized family of estimators T_{prop}

λ_1	λ_2	α_1	α_2	β_1	β_2	Members of estimator T_{prop}
0	1	–	0	–	0	$T_0 = \sum_{k=1}^{L} W_k \bar{y}_k$
0	1	–	1	–	0	$T_1 = \sum_{k=1}^{L} W_k \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)$
0	1	–	1	–	1	$T_{21} = \sum_{k=1}^{L} W_k \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}$
0	1	–	α_2	–	β_2	$T_{22} = \sum_{k=1}^{L} W_k \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)^{\alpha_2} a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)^{\beta_2}}$
1	0	0	–	1	–	$T_{31} = \sum_{k=1}^{L} W_k \bar{y}_k a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}$
λ_1	0	0	–	1	–	$T_{32} = \sum_{k=1}^{L} W_k \lambda_{1k} \bar{y}_k a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}$
λ_1	0	1	–	1	–	$T_{33} = \sum_{k=1}^{L} W_k \lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}$
λ_1	0	−1	–	1	–	$T_{34} = \sum_{k=1}^{L} W_k \lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}$
λ_1	0	1	–	−1	–	$T_{35} = \sum_{k=1}^{L} W_k \lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{x}_k - \bar{X}_k}{\bar{X}_k + \bar{x}_k}\right)}$
λ_1	0	−1	–	−1	–	$T_{36} = \sum_{k=1}^{L} W_k \lambda_{1k} \bar{y}_k \left(\frac{\bar{x}_k}{\bar{X}_k}\right) a^{\left(\frac{\bar{x}_k - \bar{X}_k}{\bar{X}_k + \bar{x}_k}\right)}$
λ_1	0	α_1	–	β_1	–	$T_{37} = \sum_{k=1}^{L} W_k \lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)^{\alpha_1} a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)^{\beta_1}}$
λ_1	λ_2	0	0	1	−1	$T_{38} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)} + \lambda_{2k} \bar{y}_k a^{\left(\frac{\bar{x}_k - \bar{X}_k}{\bar{X}_k + \bar{x}_k}\right)}\right)$
λ_1	λ_2	1	1	1	0	$T_{41} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)} + \lambda_{2k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)\right)$
λ_1	λ_2	1	0	1	0	$T_{42} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)} + \lambda_{2k} \bar{y}_k\right)$
λ_1	λ_2	1	0	0	1	$T_{43} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) + \lambda_{2k} \bar{y}_k a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}\right)$
λ_1	λ_2	1	1	−1	0	$T_{44} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)} + \lambda_{2k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)\right)$
λ_1	λ_2	1	1	1	−1	$T_{45} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)} + \lambda_{2k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)}\right)$
λ_1	λ_2	1	−1	−1	1	$T_{46} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{x}_k - \bar{X}_k}{\bar{X}_k + \bar{x}_k}\right)} + \lambda_{2k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right) a^{\left(\frac{\bar{x}_k - \bar{X}_k}{\bar{X}_k + \bar{x}_k}\right)}\right)$
λ_1	λ_2	α_1	α_2	β_1	β_2	$T_{47} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)^{\alpha_1} a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)^{\beta_1}} + \lambda_{2k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)^{\alpha_2} a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)^{\beta_2}}\right)$
λ_1	1-λ_1	α_1	α_2	β_1	β_2	$T_{48} = \sum_{k=1}^{L} W_k \left(\lambda_{1k} \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)^{\alpha_1} a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)^{\beta_1}} + (1 - \lambda_{1k}) \bar{y}_k \left(\frac{\bar{X}_k}{\bar{x}_k}\right)^{\alpha_2} a^{\left(\frac{\bar{X}_k - \bar{x}_k}{\bar{X}_k + \bar{x}_k}\right)^{\beta_2}}\right)$

Motivated by Ref. [11], we proposed generalized family of estimators given as:

$$T_{prop}=\sum_{k=1}^{L}W_k\left(\lambda_{1k}\overline{y}_k\left(\frac{\overline{X}_k}{\overline{x}_k}\right)^{\alpha_1}a^{\left[\frac{\overline{X}-\overline{x}}{\overline{X}+\overline{x}}\right]^{\beta_1}}+\lambda_{2k}\overline{y}_k\left(\frac{\overline{X}_k}{\overline{x}_k}\right)^{\alpha_2}a^{\left[\frac{\overline{X}-\overline{x}}{\overline{X}+\overline{x}}\right]^{\beta_2}}\right) \tag{16}$$

where $\lambda_{1k}+\lambda_{2k}\neq 1$ and $(\alpha_1,\alpha_2,\beta_1,\beta_2\ and\ a)$ are suitably chosen constants which reduces estimator T_{prop} into different forms. For $a=2.718$ all the members of generalized family of estimators reduces to exponential estimators. Members of the proposed class of estimators are mentioned in Table 1. To study the properties of estimator T_{prop}, we calculate the bias and mean square error of the estimator. For deriving MSE (T_{prop}), equation (16) can be written as:

$$\begin{aligned}T_{prop}=\sum_{k=1}^{L}W_k\Bigg\{&\lambda_{1k}\overline{Y}_k(1+\xi_0)\left(1+\alpha_1\xi_1+\frac{\alpha_1(\alpha_1+1)}{2}\xi_1^2\right)\left(1-\frac{\beta_1\xi_1}{2}log\,a+\frac{\beta_1\xi_1^2}{4}log\,a+\frac{\beta_1^2\xi_1^2}{8}(log\,a)^2\right)\\&+\lambda_{2k}\overline{Y}_k(1+\xi_0)\left(1+\alpha_2\xi_1+\frac{\alpha_2(\alpha_2+1)}{2}\xi_1^2\right)\left(1-\frac{\beta_2\xi_1}{2}log\,a+\frac{\beta_2\xi_1^2}{4}log\,a+\frac{\beta_2^2\xi_1^2}{8}(log\,a)^2\right)\Bigg\}\end{aligned} \tag{17}$$

$$T_{prop}=\sum_{k=1}^{L}W_k\left\{\lambda_{1k}\overline{Y}_k\left(1+\xi_0-a_0\xi_1-a_0\xi_0\xi_1+a_1\xi_1^2\right)+\lambda_{2k}\overline{Y}_k\left(1+\xi_0-b_0\xi_1-b_0\xi_0\xi_1+b_1\xi_1^2\right)\right\} \tag{18}$$

where $a_0=\alpha_1+\frac{\beta_1}{2}log\,a$

$$b_0=\alpha_2+\frac{\beta_2}{2}log\,a$$

$$a_1=\frac{\alpha_1(\alpha_1+1)}{2}+\frac{\alpha_1\beta_1}{2}log\,a+\frac{\beta_1}{4}log\,a+\frac{\beta_1^2}{8}(log\,a)^2$$

$$b_1=\frac{\alpha_2(\alpha_2+1)}{2}+\frac{\alpha_2\beta_2}{2}log\,a+\frac{\beta_2}{4}log\,a+\frac{\beta_2^2}{8}(log\,a)^2$$

subtracting $\overline{Y}$ from both the sides of equation (18) and squaring we get

$$\begin{aligned}\left(T_{prop}-\overline{Y}\right)^2=\sum_{k=1}^{L}W_h^2\overline{Y}_h^2\Big\{&1+\lambda_{1k}^2\left(1+\xi_0^2+\left(a_0^2+2a_1\right)\xi_1^2-4a_0\xi_0\xi_1\right)+\lambda_{2k}^2\left(1+\xi_0^2+\left(b_0^2+2b_1\right)\xi_1^2-4b_0\xi_0\xi_1\right)\\&+2\lambda_{1k}\lambda_{2k}\left(1+\xi_0^2+(a_1+a_0b_0+b_1)\xi_1^2-2(a_0+b_0)+\xi_0\xi_1\right)-2\lambda_{1k}\left(1+a_1\xi_1^2-a_0\xi_0\xi_1\right)-2\lambda_{2k}\left(1-b_0\xi_0\xi_1+b_1\xi_1^2\right)\Big\}\end{aligned} \tag{19}$$

taking expectation on both the sides of equation (19)

$$MSE\left(T_{prop}\right)=\sum_{k=1}^{L}W_k^2\overline{Y}_k^2\left\{1+\lambda_{1k}^2A_{1k}+\lambda_{2k}^2A_{2k}+2\lambda_{1k}\lambda_{2k}A_{3k}-2\lambda_{1k}A_{4k}-2\lambda_{2k}A_{5k}\right\} \tag{20}$$

where $A_{1k}=1+f_k\left(C_{yk}^2+\left(a_0^2+2a_1\right)C_{xk}^2-4a_0\rho_{xyk}C_{yk}C_{xk}\right)$

$$A_{2k}=1+f_k\left(C_{yk}^2+\left(b_0^2+2b_1\right)C_{xk}^2-4b_0\rho_{xyk}C_{yk}C_{xk}\right)$$

$$A_{3k}=1+f_k\left(C_{yk}^2+(a_1+a_0b_0+b_1)C_{xk}^2-2(a_0+b_0)\rho_{xyk}C_{yk}C_{xk}\right)$$

$$A_{4k}=1+f_k\left(a_1C_{xk}^2-a_0\rho_{xyk}C_{yk}C_{xk}\right)$$

$$A_{5k}=1+f_k\left(b_1C_{xk}^2-b_0\rho_{xyk}C_{yk}C_{xk}\right)$$

differentiating equation (20) with respect to λ_{1k} and λ_{2k}, equate it to zero we get:

$$\lambda_{1k}=\frac{A_{3k}A_{5k}-A_{2k}A_{4k}}{A_{3k}^2-A_{1k}A_{2k}}$$

$$\lambda_{2k}=\frac{A_{3k}A_{4k}-A_{1k}A_{5k}}{A_{3k}^2-A_{1k}A_{2k}}$$

substituting value of λ_{1k} and λ_{2k} in equation (20) we get minimum mean square error expression of T_{prop} as:

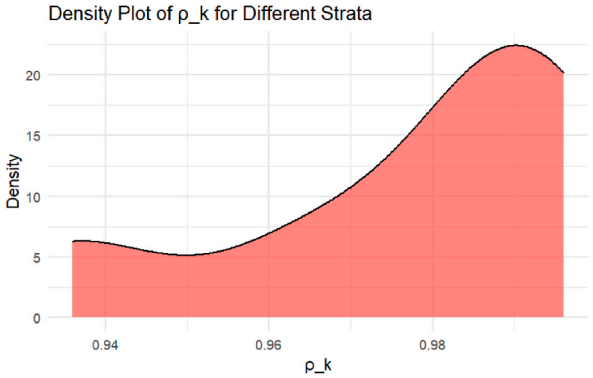

Fig. 2. Density plot of the data used for estimation.

Table 2
Population I [6].

Stratum	Population Parameters							
	N_k	n_k	h^{th}	h^{th}	S_{yk}	S_{xk}	S_{yxk}	ρ_k
1	127	31	703.74	20804.59	883.835	30486.75	25237154	0.936
2	117	21	413	9211.79	644.922	15180.77	9747943	0.996
3	103	29	573.17	14309.3	1033.467	27549.7	28294397	0.994
4	170	38	424.66	9478.85	810.585	18218.93	14523886	0.983
5	205	22	267.03	5569.95	403.654	8497.776	3393592	0.989
6	201	39	393.84	12997.59	711.723	23094.14	15864574	0.965

Table 3
Population II [22].

Stratum k	Population Parameters							
	N_k	n_k	h^{th}	h^{th}	S_{yk}	S_{xk}	S_{yxk}	ρ_k
1	106	9	1536	24,375	6425	49,189	259152246.5	0.82
2	106	17	2212	27,421	11,552	57,461	570858945.9	0.86
3	94	38	9384	72,409	29,907	16,0757	4326983639	0.90
4	171	67	5588	74,365	28,643	2,85,603	8098721462	0.99
5	204	7	967	26,441	2390	45,403	77044350.7	0.71
6	173	2	404	9844	946	18,794	15823420.36	0.89

$$Min.MSE(T_{prop}) = \sum_{k=1}^{L} W_k^2 \overline{Y}_k^2 \left(1 - \frac{A_{1k}A_{5k}^2 + A_{2k}A_{4k}^2 - 2A_{3k}A_{4k}A_{5k}}{A_{1k}A_{2k} - A_{3k}^2}\right)$$

Particular case of T_{prop}: When $\lambda_{1k} + \lambda_{2k} = 1$, then T_{prop} becomes

$$T_{48} = \sum_{k=1}^{L} W_k \left(\lambda_{1k}\overline{y}_k \left(\frac{\overline{X}_k}{\overline{x}_k}\right)^{\alpha_1} a^{\left(\frac{\overline{X}_k - \overline{x}_k}{\overline{X}_k + \overline{x}_k}\right)^{\beta_1}} + (1-\lambda_{1k})\overline{y}_k \left(\frac{\overline{X}_k}{\overline{x}_k}\right)^{\alpha_2} a^{\left(\frac{\overline{X}_k - \overline{x}_k}{\overline{X}_k + \overline{x}_k}\right)^{\beta_2}}\right) \tag{21}$$

proceeding in the same way as for the estimator T_{prop}, we obtain MSE for T_{48} given in equation (21) as follows:

$$MSE(T_{48}) = \sum_{k=1}^{L} W_k^2 \overline{Y}_k^2 \left(1 + \lambda_{1k}^2 (A_{1k} + A_{2k} - 2A_{3k}) - 2\lambda_{1k}(A_{2k} - A_{3k} + A_{4k} - A_{5k}) + (A_{2k} - 2A_{5k})\right) \tag{22}$$

differentiating equation (22) with respect to λ_{1k} and equating it to zero we get:

$$\lambda_{1k} = \frac{(A_{2k} - A_{3k} + A_{4k} - A_{5k})}{(A_{1k} + A_{2k} - 2A_{3k})} = \lambda_{1k}^{*}(Say) \tag{23}$$

substituting value of λ_{1k} given in equation (23) in equation (22) we get minimum mean square error expression of T_{48} same as the regression estimator:

Table 4
MSE table for the population I and population II

MSE of estimators for population I			MSE of estimators for population II			Rank
Estimator	MSE		Estimator	MSE		
T_0	2229.266		T_0	697800.2		
T_1	129.8228		T_1	165100.1		XI
T_2	571.6983		T_2	352641		
T_{reg}	107.341		T_{reg}	111918		VIII
T_3	107.341		T_3	111918		VIII
Members of T_{prop}	a	**MSE**	**Members of T_{prop}**	a	**MSE**	
T_{21} (Performs better for 0<a<2.718 …)	**1.5**	297.0492	T_{21} (Performs better for 0<a<2.718 …)	**1.5**	133969.52	
	2	536.9064		2	127777.67	
	2.718 …	903.6387		2.718 …	135177.39	
	4	1528.073		**4**	165822.15	
T_{31}	2.718 …	571.6983	T_{31}	2.718 …	352641	VIII
	$exp\left(2\rho_k\frac{C_{yk}}{C_{xk}}\right)$	107.341		$exp\left(2\rho_k\frac{C_{yk}}{C_{xk}}\right)$	111918	
T_{32} (Performs better for 2.718<a<11)	2.718 …	567.9373	T_{32} (Performs better for 2.718<a<10)	2.718 …	285388.18	VII
	3	471.9605		5	184770.8	
	5	156.9212		**10**	106942.7	
	11	106.54		20	111876.7	
T_{33} (Performs better for 0<a<2.718)	2.718	783.8644	T_{33} (Performs better for 0<a<2.718)	2.718	112021.59	
	2	485.4033		2	106958.09	
	0.5	306.1075		**0.5**	**105223.78**	
T_{34} (Performs better for a>2.718)	2.718	4171.273	T_{34} (Performs better for a>2.718)	2.718	660450.04	X
	5	2859.802		20	286184.6	
	40	120.9056		**65**	107710.2	
T_{35} (Performs better for 0<a<2.718)	2.718	567.8754	T_{35} (Performs better for 0<a<2.718)	2.718	285388.18	
	2	306.1075		1.5	186842	
	1.5	161.6897		**0.5**	106958.1	
T_{37} (Performs better for a>2.718)	2.718	198.1782	T_{37} (Performs better for a>2.718)	2.718	200508.27	IX
(for $\alpha = 0.5$ and $\beta = 0.5$)	4	135.1029	(for $\alpha = 0.5$ and $\beta = 0.5$)	4	173578.7	
	6	116.9228		**28**	107057	
T_{38} (Performs better for 1<a<2.718)	2.718	92.3099	T_{38} (Performs better for 1<a<2.718)	2.718	96832.33	III
	2	84.44973		2	93004.3	
	1.5	78.725485		1.5	90345.86	
T_{41} (very small effect of a, no variation in MSE)	2.718	106.353322	T_{41} (very small effect of a, no variation in MSE)	2.718	97436.754	VI
T_{42} (Performs better for a>2.718)	2.718	95.65255	T_{42} (Performs better for a>2.718)	2.718	96014.664	II
	4	75.881426		4	93502.15	
	11	31.60157		11	78978.52	
T_{43} (Performs better for 0<a<2.718)	2.718	105.78164	T_{43} (Performs better for a>2.718 …)	2.718	96122.5	V
	2	105.57572		2	92119.85	
	1.1	105.1318		1.5	82665.47	
T_{44} (Performs better for a>2.718)	2.718	105.78177	T_{44} (Performs better for a>2.718)	2.718	96124.3	IV
	4	105.5200		4	90391.73	
	6	105.22009		**6**	62028.99	
T_{45} (Performs better for a>2.718)	2.718	100.606181	T_{45} (Performs better for a>2.718)	2.718	97737.5	**I**
	5	80.15791254		11	74972.25	
	10	2.73783011		20	49452.56	
T_{46} (Performs better for a>2.718)	2.718	92.3154242	T_{46} (Performs better for a>2.718)	2.718	96834.14	III
	5	78.49649273		5	90240.99	
	11	78.60822805		11	89695.98	
T_{48} (MSE does not depends on value of a)	2.718	107.341	T_{48} (MSE does not depends on value of a)	2.718	111918	VIII

$$Min.MSE(T_{48}) = \sum_{k=1}^{L} W_k^2\left(\frac{N_k - n_k}{N_k n_k}\right)\bar{Y}_k^2\left(1-\rho_{kxy}^2\right)C_{ky}^2$$

4. Empirical study

For the empirical study we have considered the following two real data sets and density plot has been presented in Fig. 2, providing insights into the shape of the data used for the empirical study.

Data SetI [6]: For elementary and secondary schools spread throughout 923 districts in six regions of Turkey in 2007, the number of teachers is the study variable, and the number of students is the auxiliary variable. With the designations of Stratum 1, 2, 3, 4, 5, and 6, the six areas are regarded as strata. Table 2 contains information on Data *Set*I.

Data Set II: The amount of apples produced in 854 Turkish villages in 1999 is the main variable, while the number of apple trees is the auxiliary variable. The data is categorised according to Turkey's regions, and neyman allocation is used to choose random samples (villages) from each stratum (region). Table 3 contains information on Data Set II.

Table 5
PRE table for the proposed and existing estimators.

Estimators	PRE of Estimators ($\boldsymbol{n}$ = **30**)		PRE of Estimators ($\boldsymbol{n}$ = **60**)		PRE of Estimators ($\boldsymbol{n}$ = **90**)	
T_0	100		100		100	
T_1	128.1927		129.2832		131.4731	
T_2	240.0782		241.8273		242.8073	
T_{reg}	373.8135		373.9876		374.4867	
T_3	373.8135		373.9876		374.4867	
Members of T_{prop}	a	**PRE**	a	**PRE**	a	**PRE**
T_{21} (Performs better for 0<a<2.718)	**1.5**	**227.7501**	**1.5**	**227.9872**	**1.5**	228.2462
	2	205.1442	2	205.8332	2	205.8445
	2.718	182.6632	2.718	183.2213	2.718	183.4563
	4	157.4030	4	158.2321	4	158.3452
T_{31}	2.718	240.0782	2.718	240.8976	2.718	241.9232
	$\exp\left(2\rho_k\frac{C_{yk}}{C_{xk}}\right)$	373.8135	$\exp\left(2\rho_k\frac{C_{yk}}{C_{xk}}\right)$	374.3421	$\exp\left(2\rho_k\frac{C_{yk}}{C_{xk}}\right)$	374.4322
T_{32} (Performs better for 2.718<a<11)	2.718	240.7384	2.718	241.4321	2.718	241.5621
	3	244.524	3	245.4325	3	245.5432
	5	256.1329	**5**	256.8732	**5**	256.9992
	11	258.4824	11	259.2321	11	259.4022
T_{33} (Performs better for 0<a<2.718)	**1.5**	233.6703	**1.5**	234.2132	**1.5**	234.5422
	2.718	189.143	2.718	189.9872	2.718	189.9922
	4	164.2214	4	164.7823	4	164.8228
T_{34} (Performs better for a>2.718)	2.718	105.5624	2.718	106.3421	2.718	106.4452
	5	110.4826	5	110.8982	5	110.9923
	40	120.4687	40	120.8762	40	120.8882
T_{35} (Performs better for 0<a<2.718)	2.718	240.7452	2.718	241.2323	2.718	241.2672
	2	242.4734	2	242.8762	2	242.9642
	1.5	248.1682	1.5	248.8672	1.5	248.8852
T_{37} (Performs better for a>2.718)	2.718	375.6724	2.718	376.8222	2.718	376.8972
(for $\alpha = 0.5$ and $\beta = 0.5$)	4	379.8578	4	380.2342	4	380.4521
	6	380.4025	**6**	380.8976	**6**	380.9234
T_{38} (Performs better for 1<a<2.718)	2.718	378.4762	2.718	378.8765	2.718	378.9987
	2	380.5445	2	380.8876	2	380.8976
	1.5	382.6754	1.5	382.7656	1.5	382.8765
T_{41} (very small effect of a, small variation in MSE)	2.718	374.6294	2.718	374.6543	2.718	374.7865
T_{42} (Performs better for a>2.718)	2.718	377.3478	2.718	377.2245	2.718	377.8876
	4	378.3715	4	379.2321	4	378.8865
	11	381.8026	11	382.4532	11	381.8976
T_{43} (Performs better for 0<a<2.718)	2.718	374.6139	2.718	375.3212	2.718	374.7265
	2	374.8008	2	375.2132	2	374.8976
	1.5	375.0213	1.5	375.7632	1.5	375.1113
T_{44} (Performs better for a>2.718)	2.718	374.6139	2.718	375.2343	2.718	374.7842
	4	374.8573	4	375.6782	4	374.8982
	6	375.1978	6	375.9872	6	375.2445
T_{45} (Performs better for a>2.718)	2.718	374.9811	2.718	375.3242	2.718	374.9921
	5	377.158	5	377.9872	5	378.1232
	10	384.3127	10	384.7621	10	384.3127
T_{46} (Performs better for a>2.718)	2.718	374.1115	2.718	374.1243	2.718	374.1248
	5	374.5115	5	374.6222	5	374.6245
	11	374.5146	11	374.6573	11	374.6643
T_{48} (PRE does not depends on value of a)	2.718	373.8135	2.718	373.9876	2.718	374.4867

Utilizing the data sets presented in Tables 2 and 3, we have computed the MSE of the estimators defined in Table 1 to assess the efficiency of our proposed estimators. The MSE values for these estimators are summarized in Table 4. In our analysis, we varied the parameter a around 2.718, recording the corresponding values of a and minimum MSE obtained. We have opted not to include MSE values for all possible a to avoid unnecessarily lengthening Table 4.

From Table 4, we draw the following conclusions.

- **Performance of Generalized Exponential Estimators**: The members of the proposed family of generalized exponential estimators consistently outperform the existing estimators. This is evidenced by the fact that the Mean Square Error (MSE) of the proposed estimators is notably lower compared to that of the existing estimators.
- **Comparison of MSE**: Our comparison of the MSEs reveals that the proposed estimators offer superior precision. The reduced MSE indicates that the proposed estimators provide more accurate estimates of the population mean under stratified sampling.
- **Convergence to Exponential Estimators**: It is observed that all proposed estimators converge to exponential estimators when the parameter a approaches the value 2.718 (the base of the natural logarithm, e). This demonstrates the flexibility and generalizability of the proposed estimators within the framework of exponential estimators.

Table 6
Data set for the empirical analysis from [6].

Stratum	Population Parameters						
	N_k	$\overline{Y}_k$	$\overline{X}_k$	S_{yk}	S_{xk}	S_{yxk}	ρ_h
1	127	703.74	20804.59	883.835	30486.75	25237154	0.936
2	117	413	9211.79	644.922	15180.77	9747943	0.996
3	103	573.17	14309.3	1033.467	27549.7	28294397	0.994
4	170	424.66	9478.85	810.585	18218.93	14523886	0.983
5	205	267.03	5569.95	403.654	8497.776	3393592	0.989
6	201	393.84	12997.59	711.723	23094.14	15864574	0.965

- The estimators T_{21}, T_{33}, T_{35}, T_{38} and T_{43} demonstrate greater efficiency compared to the exponential estimators for values of $0 < a < 2.718$. This indicates that within this range of a, the proposed estimators yield lower MSE and hence provide more accurate and reliable estimates of the population mean in stratified sampling.
- The estimators $T_{22}, T_{31}, T_{32}, T_{34}, T_{36}, T_{37}, T_{41}, T_{42}.T_{43}, T_{44}, T_{45}$ *and* T_{46} exhibit superior efficiency compared to the exponential estimators for values of $2.718 < a < exp\left(2\rho_k \frac{C_{yk}}{C_{xk}}\right)$.This means that within this range of a, the proposed estimators achieve lower MSE and thus offer more precise and effective estimates of the population mean in stratified sampling.
- The estimator T_{48} is unique in that it does not depend on the value of a. Consequently, its MSE remains consistent regardless of a, and is equivalent to the MSE of the regression estimator. This property highlights T_{48}'s robustness and stability in estimation, making it a reliable choice when compared to other estimators that vary with a.
- MSE of the ratio cum dual to ratio estimator T_{45} is the smallest among all the estimators discussed. Consequently, this estimator is identified as the best performer in terms of accuracy and reliability, providing the most precise estimates of the population mean

Overall, the results affirm that the generalized exponential estimators proposed in this study deliver enhanced estimation performance compared to traditional methods.

5. Simulation study

To confirm our findings, simulation research was carried out.

Step 1. Using the multivariate normal distribution, we created a population of size N = 1050 for the simulation research (study variable Y and auxiliary variable X with a specific correlation coefficient).

Step 2. Since stratified sampling is what we are working with, we create random samples for every stratum using various factors to create a heterogeneous stratum. To create the population for the three stratums $k = 1,2,$ *and* 3, we utilized the following parameters:

$$N_1 = 350; N_2 = 350 \text{ and } N_3 = 350;\ \overline{Y}_1 = 50; \overline{Y}_2 = 40 \text{ and } \overline{Y}_3 = 30\ ;$$

$$\overline{X}_1 = 51; \overline{X}_2 = 41 \text{ and } \overline{X}_3 = 31\ ;\ \sigma_{x1}^2 = 121; \sigma_{x2}^2 = 49 \text{ and } \sigma_{x3}^2 = 81$$

$$\sigma_{y1}^2 = 25; \sigma_{y2}^2 = 36 \text{ and } \sigma_{y3}^2 = 64\ ;\ \rho_1 = 0.70; \rho_2 = 0.80 \text{ and } \rho_3 = 0.90$$

Step 3. Next, we use basic random sampling without replacement to choose 1500 bi-variate samples of sizes n = 30,60, and 90 from each stratum.

Step 4. We calculate the MSE of estimators for each sample drawn from the population using this sample data. To get the MSE of the estimator, we average the total of the MSEs of the three strata.

$$MSE(T_i) = \frac{1}{1500}\sum_{j=1}^{1500} MSE\left(T_i^j\right)$$

Where j is number of iterations and T_i is i^{th} estimator and $MSE\left(T_i^j\right) = \sum_{k=1}^{3} MSE\left(T_{ik}^j\right)$.

Percent Relative Efficiency of the estimators is calculated as $PRE\left(T_i^j\right) = \frac{MSE(T_0)}{MSE\left(T_i^j\right)} * 100$where $T_i^j = T_0, T_1, T_2, T_{reg}, T_3, T_{21}, T_{22}, T_{31}, T_{32}, T_{34}, T_{35}, T_{36}, T_{37}, T_{41}, T_{42}.T_{43}, T_{44}, T_{45}$ *and* T_{46}

Results of simulation study are illustrated in Table 5.

From the analysis of Tables 4 and 5, it is evident that both the empirical study and the simulation study produce consistent and comparable results. This similarity in outcomes suggests a strong alignment for results discussed for the proposed estimators. Consequently, within the specified range, the proposed estimators demonstrate superior performance and reliability. As a result, they are recommended over the traditional exponential estimators for achieving more accurate and efficient estimates.

6. Optimization problem

Perpetually, budget for a survey is fixed and, in such cases, statistician has to minimize the variance of the estimator due to the cost curtailment or he has to minimize the cost for the acceptable value of variance.

Therefore, in stratified sampling to get the better results the allocation of sample sizes to different strata is made in any one of the following ways.

- Sampling variance is minimized for a given cost, or
- The cost is minimized for specified precision.

Here we will discuss the problem of sample allocation to different strata to minimize the variance for the fixed cost.

Our objective is to minimize the mean square error of the discussed estimators for the fixed cost. So, we set an optimization problem as:

Minimize MSE (T_i) ; where $i = 0, 1, 2, 3, \ldots$

Subject to $\sum_{k=1}^{L} c_k n_k = C_0$

$n_k \geq 2$

$n_k \leq N_k, n_k$ are the integers.and we solve this problem using Lagrange's multiplier method. It is a technique that is used to find the local minima or maxima of a function subject to constraints. This method follows these simple steps.

Step 1. for the given multivariate function i.e. MSE (T_i) and the constraints $\sum_{k=1}^{L} c_k n_k = C_0$, the Lagrange's function is defined as: $\varphi(n_k) = MSE(T_i) - \lambda\left(\sum_{k=1}^{L} c_k n_k - C_0\right)$ where λ is a Lagrange's multiplier.

Step 2. differentiate $\varphi(n_k)$ with respect to n_k to get the partial derivative and set it equal to zero.

Step 3. Search for any immediate solution of n_k, if not then precede further drop out λ and equate the equation to get the value of n_k.

To solve our problem, we have considered following data set given in table:

Using above data, we will derive the integer programming problem for minimizing the mean square error of our proposed estimator T_3 by taking fixed cost as $c_1 = 2, c_2 = 3, c_3 = 4, \ c_5 = 5, c_4 = 6, c_5 = 7$ and $C_0 = 400$. The optimization problem for estimator T_4 is written as:

Minimize $\frac{96789.38}{n_1} + \frac{3320.74}{n_2} + \frac{12778.2}{n_3} + \frac{22149.75}{n_4} + \frac{3564.889}{n_5} + \frac{34837.95}{n_6}$

Subject to $2n_1 + 3n_2 + 4n_3 + 5n_4 + 6n_5 + 7n_6 \leq 400$

$n_k \geq 2 n_k \leq N_k$

where $k = 1, 2, 3, 4, 5$ *and* 6

Similarly, we write optimization problem and derive optimum sample size for the existing estimators T_0, T_1 *and* T_2. To get the optimum values of n_k we use the Lagrange's multiplier technique. We define the Lagrange's function as:

$$\varphi(n_k) = \sum_{k-1}^{L} \frac{1}{n_k} \overline{Y}_k^2 C_{yk}^2 \left(1 - \rho_k^2\right) + \lambda\left(\sum_{k=1}^{L} c_k n_k - C_0\right) \tag{24}$$

differentiating equation (24) with respect to n_k and equation it to zero, we get:

$$n_k = \sqrt{\frac{\overline{Y}_k^2 C_{yk}^2 \left(1 - \rho_k^2\right)}{\lambda c_k}} \tag{25}$$

and differentiating equation (24) with respect to λ and equating it to zero we get:

$$\sum_{k=1}^{L} c_k n_k = C_0$$

substituting value of n_k from equation (25) in equation (24) we get

$$\sum_{k=1}^{L} c_k \sqrt{\frac{\overline{Y}_k^2 C_{yk}^2 \left(1 - \rho_k^2\right)}{\lambda c_k}} = C_0$$

$$\Rightarrow \sqrt{\lambda} = \sum_{k=1}^{L} \sqrt{\frac{\overline{Y}_k^2 C_{yk}^2 \left(1 - \rho_k^2\right) c_k}{C_0}} \tag{26}$$

put this value of equation (26) in equation (25), we get:

Table 7
MSE and PRE table for the Population given in Table 6 for optimized sample size (using Lagrange's multiplier) and fix cost.

Estimators	Allocations						MSE	PRE
	n_1	n_2	n_3	n_4	n_5	n_6		
T_0 (Usual Mean estimator)	27	16	23	16	7	12	6084.404	100
T_1 (Ratio Estimator)	55	8	14	14	5	15	272.3166	2234.313
T_2 (Exponential estimator)	27	15	21	16	7	13	1571.178	387.2511
T_3 (Proposed Estimator)	51	8	13	15	6	16	**237.6476**	**2560.263**

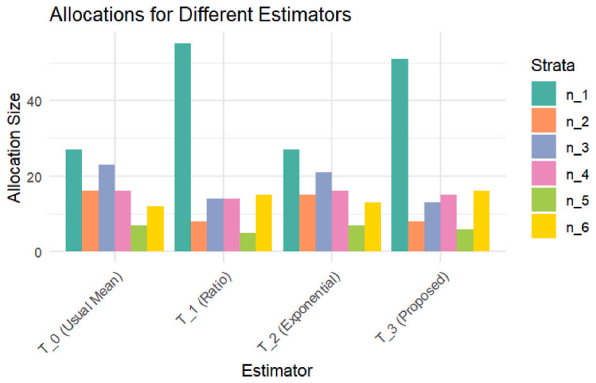

Fig. 3. MSEs of the different estimators.

$$n_k = \frac{C_0\sqrt{\frac{\overline{Y}_k^2 C_{yk}^2(1-\rho_k^2)}{c_k}}}{\sum_{k=1}^{L}\sqrt{\overline{Y}_k^2 C_{yk}^2(1-\rho_k^2)c_k}}$$

n_k is the optimum sample size for k^{th} stratum for fixed cost C_0 for the estimator T_3. In the same way using Lagrange's multiplier method we have computed optimum sample sizes for each estimator and the values are given in Table 7. MSE and PRE for the remaining members of the proposed estimator can be obtained in the similar manner.

The MSE and PRE for both existing and proposed estimators are presented in Table 7 numerically and allocations for the different estimators is visualized in Fig. 3. This table clearly depicts that the MSE of the proposed estimator has been significantly reduced when considering fixed cost and the optimal sample sizes for the stratum determined using Lagrange's method.

7. Conclusion

In this article, we have introduced a family of estimators T_{prop} that, for various values of appropriately chosen constants $(\alpha_1, \beta_1, \alpha_2, and\ \beta_2)$, T_{prop} reduces to different forms, collectively referred to as members of the generalized exponential family of estimators. We compared the MSE of these proposed estimators with those of traditional estimators, including the usual mean estimator, ratio estimator, exponential estimator, and the more general exponential estimators. Our analysis demonstrates that the proposed estimators consistently outperform the existing alternatives, offering superior accuracy and efficiency in estimating the population mean under stratified sampling.

Then, we addressed an optimization problem for improving the performance of the estimators by minimizing the MSE through the determination of optimal sample sizes. The solution is obtained using the Lagrange multiplier technique, and the results are thoroughly validated using real-world dataset. Our findings reveal that the proposed estimators outperform existing methods by a significant margin, highlighting their practical effectiveness and strong theoretical foundation.

The analysis makes the somewhat simplistic assumption that sampling costs follow a linear relationship. In actuality, costs may have nonlinear elements like fixed costs or stratified logistical costs. This may be the study's future focus. Additional research on the non-linear cost may be done.

CRediT authorship contribution statement

Poonam Singh: Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation. **Prayas Sharma:** Visualization, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Rajesh Singh:** Writing – review & editing. **Badr Aloraini:** Writing – review & editing, Validation, Funding acquisition. **Aysha Akhtar:** Writing – review & editing, Validation,

Software, Data curation.

Ethical statement

There are no human/animal subjects in this article therefore an ethics statement is not applicable because this study is applied on already published data.

Data availability statement

All data applied is included in the manuscript.

Funding

The authors declare that no funds or other grants were received for the preparation of this manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Chochran, The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce, J. Agric. Sci. 30 (2) (1940) 262–275.
[2] A. Das, Use of auxiliary information in estimating the finite population variance, Sankhya 40 (1978) 139–148.
[3] H. Singh, R. Singh, M. Espejo, M. Pineda, S. Nadarajah, On the efficiency of a dual to ratio-cum-product estimator in sample surveys, Math. Proc. Roy. Ir. Acad. 105 (2) (2005) 51–56.
[4] M. Khoshnevisan, R. Singh, P. Chauhan, N. Sawan, F. Smarandache, A general family of estimators for estimating population means using known value of some population parameter(s), Far East J. Statis. 22 (2) (2007) 181–191.
[5] P. Singh, C. Bouza, R. Singh, Generalized exponential estimator for estimating the population mean using auxiliary variables, J. Sci. Res. 63 (1) (2019).
[6] N. Koyuncu, C. Kadilar, Ratio and product estimators in stratified random sampling, J. Stat. Plann. Inference 139 (8) (2009) 2552–2558.
[7] N. Koyuncu, C. Kadilar, On improvement in estimating population mean in stratified sampling, J. Appl. Stat. 37 (6) (2010) 999–1013.
[8] P. Sharma, R. Singh, Efficient estimator of population mean in stratified random sampling using auxiliary attribute, World Appl. Sci. J. 27 (12) (2013) 1786–1791.
[9] P. Sharma, R. Singh, Improved Estimators for Simple random sampling and stratified random sampling under Second order of Approximation, Statis. Trans.New Series 14 (3) (2013).
[10] H.P. Singh, R.S. Solanki, Efficient ratio and product estimators in stratified randomsampling, Commun. Stat. Theor. Methods 42 (6) (2013) 1008–1023, https://doi.org/10.1080/03610926.2011.592257.
[11] R.S. Solanki, H.P. Singh, An improved estimation in stratified random sampling, Commun. Stat. Theor. Methods 45 (7) (2016) 2056–2070.
[12] S. Bhushan, A. Kumar, S. Singh, Some efficient classes of estimators under stratified sampling, Commun. Stat. Theor. Methods 52 (6) (2023) 1767–1796.
[13] P. Singh, R. Singh, M. Mishra, Almost unbiased optimum ratio-type estimator for estimating population mean in stratified sampling in presence of non-response, Braz. J. Biomet. 42 (1) (2024) 68–77.
[14] J. Neyman, On the two different aspects of the representative methods. The method of stratified sampling and the method of purposive selection, J. Roy. Stat. Soc. 97 (1934) 558–606.
[15] S. Chatterjee, A note on optimum allocation, Scand. Actuar. J. 1967 (1–2) (1967) 40–44.
[16] N. Jahan, M. Khan, M. Ahsan, A generalized compromise allocation, J. Indian Statis. Associat. 32 (1994) 95–101.
[17] J.A. de Moura Brito, G.S. Semaan, A.C. Fadel, L.R. Brito, An optimization approach applied to the optimal stratification problem, Commun. Stat. Simulat. Comput. 46 (6) (2017) 4419–4451.
[18] R. Varshney, Mradula, Optimum allocation in multivariate stratified sampling design in the presence of non-response with gamma cost function, J. Stat. Comput. Simulat. 89 (13) (2019) 2454–2467.
[19] Mradula, S.K. Yadav, R. Varshney, M. Dube, Efficient estimation of population mean under stratified random sampling with linear cost function, Commun. Stat. Simulat. Comput. 50 (12) (2021) 4364–4387.
[20] S.K. Yadav, G.K. Vishwakarma, R. Varshney, A. Pal, Improved memory type product estimator for population mean in stratified random sampling under linear cost function, SN Comput. Sci. 4 (3) (2023) 235.
[21] A. Pal, R. Varshney, S.K. Yadav, T. Zaman, Improved memory-type ratio estimator for population mean in stratified random sampling under linear and non-linear cost functions, Soft Comput. 28 (13) (2024) 7739–7754.
[22] C. Kadilar, H. Cingi, New ratio estimators using correlation coefficient, Inter 4 (March) (2006) 1–11.