STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–26.
Published online in International Academic Press (www.IAPress.org)

Estimating the population mean of product type using calibration in stratified random sampling

Saja Mohammad Hussein [1,*], Mohammad Adel Kazem [2]

[1] *Department of Statistics, College of Administration and Economics, University of Baghdad Iraq, Baghdad, Iraq*
[2] *Al-Kindi Hospital, Baghdad Al Russafa Health, Ministry of Health, Baghdad, Iraq*

Abstract For estimating the population mean in stratified random sampling, there is a wealth of literature on the subject. In this research, three new combinations of product-type estimators are proposed based on a calibration technique. This is achieved by using different distance functions and leveraging information from the auxiliary variable to improve estimation efficiency. Mathematical formulas for the proposed estimators and their Biases, as well as estimates of the variances of these means were derived. The calibration weights for these estimators were based on certain constraints imposed on the auxiliary variable information, such as constraints involving the variance of the estimated population mean in the stratified random sample and constraints related to the coefficient of variation. To evaluate the performance of the proposed estimators, two simulation studies were conducted. The first simulation study considered populations of different sizes $(100, 300, and 900)$. The second simulation dealt with a population of 2000 units distributed across five strata, using three sets of different population characteristics in the two experiments. The estimators were compared using the mean squared error (MSE) and absolute bias and relative efficiency (RE) criteria. The results showed that the first proposed estimator, based on the Hellinger-type distance function, outperformed the other estimators, recording the lowest mean squared error and the highest relative efficiency in most simulations compared to the traditional estimator. The second proposed estimator also showed improvement in some cases, while the third estimator, based on the Kullback–Leibler distance function, did not achieve the expected performance under the study conditions. The proposed estimators were applied to two real-word datasets, and the practical application revealed that the estimators' performance varied depending on the evaluation criterion. The simulation and practical application results confirmed a difference in the absolute experimental bias values between the estimators depending on the population size and the scenario used. The results showed that the performance of the estimators varies depending on the evaluation criterion and data characteristics. These findings highlight the importance of selecting a distance function and calibration constraints that are appropriate to the nature of the population and the available ancillary information, which contributes to improving the efficiency of population mean estimation in stratified random sampling.

Keywords Stratified random sampling, calibration, product type estimator, Hellinger-type distance, relative quadratic distance, kullback–Leibler distance, mean squared error, bias, relative efficiency.

DOI: 10.19139/soic-2310-5070-5124

1. Introduction

Stratified random sampling is one of the most widely used sampling methods in statistical surveys. It can improve precision by dividing the population into relatively homogeneous strata, thereby reduces variance and icreasing estimation efficiency compared to simple random sampling. To achieve higher levels of precision, researchers often leverage auxiliary information available about the statistical population when constructing estimators [1]. Auxiliary information is an important tool for improving estimation efficiency in survey sampling and can be incorporated into estimators construction to reduce bias and variance, thereby enhancing the accuracy of the results. Prominent examples of estimators that utilize auxiliary information include regression estimators, ratio estimators, and product

*Correspondence to: Saja Mohammad Hussein (Email: saja@coadec.uobaghdad.edu.iq). Department of Statistics, College of Administration and Economics, University of Baghdad Iraq, Baghdad.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

estimators. Consequently, many researchers have focused on developing these estimators to improve efficiency and minimize the mean squared error (MSE). Regression and ratio estimators are typically employed when the relationship between the study variable and the auxiliary variable is positive, whereas product estimators are preferred when this relationship is negative. Identifying the nature of this relationship allows for the derivation of more efficient estimates compared to the traditional estimator. Among the modern methods developed to enhance estimators that utilize auxiliary information is the calibration approach, introduced by Deville and Särndal in 1992 [2]. This approach involves adjusting the original design weights of a sample to derive new calibration weights that satisfy constraints based on known auxiliary information about the population. This technique has gained significant attention in the literature on statistical sampling due to its ability to improve the accuracy of estimates while preserving the properties of the original design weights as much as possible. Calibration relies on minimizing an appropriate distance function between the original weights and the adjusted weights, subject to the constraints imposed by the auxiliary information. Many researchers have addressed the calibration method and proposed different estimators using various distance functions and constraints. These efforts have contributed to the development of a large number of calibration estimators that have proven efficient in improving the estimation of population parameters. Notably, Singh et al. (1998) [3] presented a study on an advanced calibration method for estimating the variance of the generalized regression estimator. The research highlights the need to develop a more effective method for estimating the variance of the Generalized Regression (GREG) estimator. The researchers demonstrate that the traditional approach based on ”low-level calibration” is less effective than a new method known as ”higher-level calibration,” which incorporates both an auxiliary variable and its known variance. Theoretical and empirical results indicate that this approach yields more accurate and stable estimates. It encompasses special cases of estimators already established in the literature, such as the estimator proposed by Deng and Wu (1987). Furthermore, the method has been extended to accommodate stratified sampling designs. Rao et al.(2012) [4] extended the use of calibration to scenarios involving multiple auxiliary variables within stratified sampling, employing a chi-square distance function for a regression-type estimator. Singh et al.(2015) [5] proposed a regression-type estimator for the population mean that utilizes auxiliary information specifically the mean and stratified variance of the auxiliary variable along with a chi-square distance function. Simulation results demonstrated that this proposed estimator is more efficient than several other existing calibration estimators. Also in 2015, Rao et al. [6] proposed a regression-type estimator for the population mean using multiple auxiliary variables, the calibration weights were determined via the Lagrange multiplier method, and the resulting estimator was derived in a manner analogous to the linear regression (LREG) estimator. Koyuncu and Kadilar (2016) [7] presented a new regression-type estimator for the population mean by incorporating three constraints based on the mean and variance of the auxiliary variable, while also refining the weight formula originally presented by Tracy et al. (2003). Furthermore, Nidhi et al. (2017) [8] presented a study on the application of calibration techniques for estimating the population mean in stratified and double-stage stratified sampling. A new mean estimator based on the calibration method was developed, relying on a set of constraints that contribute to improving the accuracy of this estimation. Simulation results showed that the proposed estimator outperformed the estimator presented by Tracy et al. (2003). The results also demonstrated the effectiveness of the new estimator in both stratified and double-stage stratified sampling. In a study by Ozgul (2018) [9], new weights were proposed for estimating the population mean using a regression-type estimator within a stratified sampling framework. This approach incorporated an additional constraint alongside existing ones, utilizing three constraints based on auxiliary information specifically the mean and the coefficient of variation in addition to the constraint that the sum of the weights must equal a specific value. The study showed an improvement in performance compared to certain existing calibration estimators. Subsequently, Ozgul (2019) [10] extended the approach from using a single auxiliary variable to two, proposing regression-type calibration estimators for the population mean and deriving optimal calibration weights. The research noted that alternative distance functions could be employed when difficulties arise in obtaining positive weights. Simulation results indicated that the proposed estimators outperform existing calibration estimators, Rai et al.(2020) [11] suggested using several distance functions, not just chi-square, with two auxiliary variables to determine how the estimator results differ when the distance function is changed. The study, through simulation, showed that the performance of the calibration estimators depends on the distance function used in constructing the weights, and this was confirmed using real data. , Iseh and Bassey(2023) [12] utilized an adjustment to calibration weights to develop an exponential-ratio-type estimator for the population median within

the contexts of simple random sampling and stratified sampling, leveraging known auxiliary population information. The authors proposed ratio-type estimators for the population median in stratified sampling using a calibration approach, with the auxiliary information centered on the median. The study indicated that adjusting the weights of the exponential-ratio estimator through calibration resulted in improved efficiency compared to existing estimators. In the same year, Shahzad et al. [13] introduced designs for estimating the median using multistage ranked set sampling (MRSS). This method was subsequently extended to two-stage stratified sampling, employing various calibration weights and constraints based on the sum of weights, the mean, and the coefficient of variation which led to enhanced performance of the calibration estimators within the MRSS design. Gupta and Trisandhya (2024) [14] addressed the estimation of a population mean under stratified sampling in the presence of measurement error, while also examining the impact of different distance functions on calibration weights ,testing three such functions. The study concluded that measurement error should be taken into account when evaluating calibration weights. Jabeen et al.(2024) [15] shifted the focus from estimating the population means to estimate the population variance. Their study introduced three types of calibration-based estimators a standard variance estimator, a ratio-type estimator, and an exponential-type estimator aiming to improve population variance estimation by deriving these estimators under a set of calibration constraints. In the same year, Pandey et al. (2024) [16] presented a more advanced approach, utilizing calibration weights to estimate population variance through ratio, regression, and exponential-type estimators within a two-phase stratified sampling design. Their numerical analysis and simulation results demonstrated that the proposed estimators outperformed the standard variance estimator. Alomair et al. (2025) [17] proposed an estimator for the population cumulative distribution function (CDF) using calibration weights based on a chi-square distance function, subject to several constraints. This approach incorporated auxiliary information specifically the CDF and the coefficient of variation of the auxiliary variable, in addition to its mean and addressed the challenge of bandwidth selection by developing a kernel-based nonparametric estimator. Also in 2025, Babatunde and Oladugba [18] addressed the issue of outliers in the auxiliary variable by incorporating its median into the calibration constraints. Their findings indicated that the proposed estimator performed better than the other calibration estimators examined in the study based on the RMSE and ABS comparison criteria. In the same year, Rai et al. [19] proposed composite calibration-based estimators for the population total under a successive sampling design, utilizing composite weights. Simulation results demonstrated that the proposed composite calibration estimators could be more efficient than traditional estimators across various sampling fractions and weight selections. Furthermore, a recent study by Garg et al. (2026) [20] focused on estimating the population mean using multiple auxiliary variables, incorporating the coefficient of variation; the various estimators were compared through a simulation study. In the same year, KAzim and Hussein [21] conducted a study comparing several calibration estimators within a stratified random sampling framework to estimate average wheat production across Iraqi governorates, using cultivated area as an auxiliary variable. The results indicated that calibration methods improved estimation accuracy compared to the traditional estimator with the Kadilar and Koyuncu method demonstrating superior efficiency. Given the importance of the product estimator in populations characterized by a negative correlation between the study variable and an auxiliary variable, it has garnered significant attention from researchers who have worked to develop and extend it for various sampling designs, including stratified random sampling. Studies have demonstrated that leveraging auxiliary information within strata can lead to a notable improvement in the precision of population mean estimation compared to traditional estimators [22]. Koyuncu and Kadilar (2009) [23] proposed an advanced family of ratio and product estimators for stratified random sampling, deriving formulas for bias and mean squared error (MSE) and comparing their efficiency against traditional estimators. The results indicated that the proposed estimators yielded lower MSE values, demonstrating their superiority in estimating the population mean. In a study by Alam et al. (2019) [24], new calibration estimators based on multiple auxiliary variables within a stratified random sampling framework were proposed; numerical results showed a significant improvement in estimation efficiency compared to traditional estimators. Furthermore, Ojua et al. (2021) [25] introduced calibration-based product-type estimators for the population mean in stratified random sampling, utilizing a single auxiliary variable subject to two calibration constraints. This approach incorporated three distance functions: chi-square, Kullback–Leibler distance function, and modified chi-square. The findings revealed that the proposed estimators outperformed the traditional product estimator in most cases, with the estimator based on the chi-square distance function proving to be the most efficient. A review of previous studies indicates that statistical calibration methods have contributed

to improving the efficiency of estimating population means in stratified sampling by employing various distance functions and calibration constraints. Although prior research has focused on developing product-type estimators that use auxiliary information and applying calibration techniques within stratified random sampling, most studies have been limited to specific types of distance functions or a restricted set of calibration constraints. Furthermore, the performance of these estimators varies depending on the nature of the data and the relationship between the study variable and the auxiliary variables. Many existing studies, such as those by Alam et al. (2019) [24] and Ojua et al. (2021) [25], have relied primarily on known population means of auxiliary variables. In contrast, the use of additional information specifically regarding the relative dispersion of auxiliary variables, such as the use of coefficient of variation remains relatively limited within the context of calibrated ratio-type estimators for stratified random sampling. Yet, incorporating such information could enhance the efficiency of calibration weights and improve estimation accuracy. The present study aims to address this gap by developing three calibrated product-type estimators within the framework of stratified random sampling to estimate the population mean using different and new combinations. These methods employ various distance functions, specifically the Hellinger-type distance function and relative quadratic distance function, which have not been used by any previous researchers in product-type estimators. These estimators are integrated with auxiliary information represented by constraints. Specifically, the means and weights of the strata in the population are used, along with the coefficient of variation under one or two constraints. The auxiliary information in this research goes beyond simply the mean of the auxiliary variables to include information about their relative dispersion. Furthermore, the research goes beyond proposing point estimators and includes deriving their mathematical formulas, associated calibration weights, and bias expressions by also developing variance estimators for each proposed estimator. This provides a means to quantify sampling uncertainty and compare the precision of the proposed estimators against the traditional product-type estimator $\bar{y}_{p0}$. The derived variance formulas account for both the within-stratum variance of the product-type estimator and the calibration weights. To evaluate the efficiency of these estimators, a simulation study and an empirical application were conducted, with the results compared against those of the traditional product-type estimator. The main scientific contribution of the research is to propose three new combinations of calibrated product-type estimators using three different distance functions that played a role in modifying the weights of the strata and auxiliary information of a different nature, and to provide an integrated theoretical and applied framework for studying their performance, and to study how these choices affect the accuracy and efficiency of estimating the population mean when the auxiliary variable is negatively correlated with the study variable, thus contributing to providing more efficient and accurate alternatives for estimating the population mean and enhancing the applications of statistical calibration in stratified sampling. The research is organized as follows: Section two presents the notation and definitions used in stratified random sampling, derives the proposed formulas for the population mean estimators, and calculates the corresponding bias for each estimator. Section three derives the variances for the proposed population mean estimators. Section four outlines the criteria for comparison. Section five details the simulation experiment, the application to real-world data, and the discussion of the results. Finally, section six presents the conclusions and section seven presents recommendations.

2. Proposed calibrated estimators for the population mean of product–type in stratified random sampling

Before presenting the proposed estimators, we outline the key symbols and terms used in stratified random sampling throughout the formulas and derivations, as follows Table 1.

N_h	Number of population units in stratum h.
n_h	Number of sample units in stratum h.
w_h	Weight of stratum h.
$Y_{h1}, Y_{h2}, \ldots, Y_{hN_h}$	Population values for the study variable in stratum h.
$y_{h1}, y_{h2}, \ldots, y_{hn_h}$	Sample values for the study variable in stratum h.
$f_h = \dfrac{n_h}{N_h}$	Sampling fraction for stratum h.

$\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}$	Population mean of the study variable in stratum h.
$\overline{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} X_{hi}$	Population mean of the auxiliary variable in stratum h.
$\overline{Y} = \sum_{h=1}^{L} \frac{N_h}{N}\overline{Y}_h$	Population mean of the study variable.
$\overline{X} = \sum_{h=1}^{L} \frac{N_h}{N}\overline{X}_h$	Population mean of the auxiliary variable.
$\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$	Sample mean of the study variable in stratum h.
$\overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$	Sample mean of the auxiliary variable in stratum h.
$S_{hy}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(Y_{hi}-\overline{Y}_h\right)^2$	Population variance of the study variable in stratum h.
$s_{hy}^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}\left(y_{hi}-\overline{y}_h\right)^2$	Sample variance of the study variable in stratum h.
$S_{hx}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(X_{hi}-\overline{X}_h\right)^2$	Population variance of the auxiliary variable in stratum h.
$s_{hx}^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}\left(x_{hi}-\overline{x}_h\right)^2$	Variance of the sample for stratum h for the auxiliary variable.
$\overline{y}_{st} = \frac{\sum_{h=1}^{L} N_h\overline{y}_h}{N}$	Mean of stratified sampling for the study variable.
$\overline{y} = \frac{\sum_{h=1}^{L} n_h\overline{y}_h}{n}$	Sampling mean.
$\gamma_h = \frac{1}{n_h} - \frac{1}{N_h}$	Finite population correction term for stratum h.
$C_{hx} = \frac{S_{hx}}{\overline{X}_h}$	Coefficient of variation for the auxiliary variable in stratum h.
$c_{hx} = \frac{s_{hx}}{\overline{x}_h}$	Coefficient of variation for the auxiliary variable for sample drawn from stratum h.
$\rho_{xy} = \frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\right]\left[\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2\right]}}$	Correlation coefficient between the auxiliary variable and the study variable.
ρ_{hxy}	Correlation coefficient between the auxiliary variable and the study variable in stratum h.
ρ_{hxs}	Correlation coefficient between the mean of the auxiliary variable and the variance of the auxiliary variable in stratum h.
ρ_{hys}	Correlation coefficient between the mean of the study variable and the variance of the study variable in stratum h.

2.1. The traditional population mean estimator of the product – type in stratified random sampling

The population mean of product – type was first used in sampling literature when Robson (1957) [26] proposed it as an alternative to the ratio estimator in cases where the relationship between the study variable and the auxiliary variable is inverse (negative). Later, Murthy (1964) [27] further developed the product estimator method and derived its statistical properties, making it one of the fundamental estimators used in auxiliary sampling theory. The traditional population mean estimator of the product – type in stratified random sampling [2] is given by the following equation:

$$\bar{y}_{p0} = \sum_{h=1}^{L} w_h \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h} \tag{1}$$

The bias of the traditional population mean estimator of the product-type is given by the following equation:

$$\text{Bias}\left(\overline{y}_{p0}\right) = \sum_{h=1}^{L} w_h \left(\frac{N_h - n_h}{N_h n_h}\right) \frac{S_{hxy}}{\overline{X}_h} \tag{2}$$

The variance of the traditional population mean estimator of the product –type is also given by the following equation:

$$\widehat{V}\left(\overline{y}_{p0}\right) = \sum_{h=1}^{L} w_h^2 \gamma_h \left(s_{hy}^2 + R^2 s_{hx}^2 + 2R s_{hxy}\right) \tag{3}$$

Where:

$$s_p = \left(s_{hy}^2 + R^2 s_{hx}^2 + 2R s_{hxy}\right), \qquad s_{hxy} = \rho_{xy} s_{hx} s_{hy}, \qquad R = \frac{\overline{Y}}{\overline{X}}.$$

The researchers proposed several estimators of product – type for estimating the population mean using the calibration technique, subject to several constraints, as follows:

2.2. proposed calibrated estimator of the population mean of product – type based on the Hellinger-type distance function

Following the method of derivation used by Ojua et al. [25], the proposed calibrated product- type estimator of the population mean is given by the following formula:

$$\bar{y}_{sp1} = \sum_{h=1}^{L} \Omega_{h1}\, \bar{x}_h \bar{y}_h / \bar{X}_h$$

Based on the Hellinger-type distance function as shown in formula Equation (4). This function is used in statistical calibration to determine weights that satisfy the calibration constraints while minimizing the variance between the original and calibrated weights according to equation Equation (4). It is characterized by its ability to keep positive weights within the function's validity range; it also depends on the square roots of the weights.

$$\sum_{h=1}^{L} \frac{2\left(\sqrt{\Omega_{h1}} - \sqrt{w_h}\right)^2}{Q_h} \tag{4}$$

Where Ω_{h1} are the assumed calibration weights chosen so that the sum of the Hellinger-type distance functions given in Equation (4), is as small as possible, and subject to the calibration constraint given in Equation (5).
Q_h :constants chosen appropriately to obtain different forms of the estimator, most often $Q_h = 1$.

$$\sum_{h=1}^{L} \Omega_{h1} \bar{x}_h = \sum_{h=1}^{L} w_h \bar{X}_h. \tag{5}$$

The Lagrange multiplier is used to minimum Equation (4) when considering the constraint in Equation (5). Defining the Lagrange multiplier λ , the Lagrange function is given by the following:

$$L = \sum_{h=1}^{L} \frac{2\left(\sqrt{\Omega_{h1}} - \sqrt{w_h}\right)^2}{Q_h} - 2\lambda \left[\sum_{h=1}^{L} \Omega_{h1} \bar{x}_h - \sum_{h=1}^{L} w_h \bar{X}_h\right]. \tag{6}$$

To obtain the adjusted calibration weights Ω_{h1}, the Lagrange function is derived with respect to Ω_{h1} and set equal to zero:

$$\frac{\partial}{\partial \Omega_{h1}} \sum_{h=1}^{L} \frac{2\left(\sqrt{\Omega_{h1}} - \sqrt{w_h}\right)^2}{Q_h} - 2\lambda \left[\sum_{h=1}^{L} \Omega_{h1} \bar{x}_h - \sum_{h=1}^{L} w_h \overline{X}_h\right].$$

$$\frac{2}{Q_h}\left(\frac{\sqrt{\Omega_{h1}} - \sqrt{w_h}}{\sqrt{\Omega_{h1}}}\right) - 2\lambda \bar{x}_h = 0.$$

$$\left(\frac{\sqrt{\Omega_{h1}} - \sqrt{w_h}}{\sqrt{\Omega_{h1}}}\right) = \lambda Q_h \bar{x}_h.$$

$$1 - \frac{\sqrt{w_h}}{\sqrt{\Omega_{h1}}} = \lambda Q_h \bar{x}_h.$$

$$\frac{\sqrt{w_h}}{\sqrt{\Omega_{h1}}} = 1 - \lambda Q_h \bar{x}_h.$$

$$\sqrt{\Omega_{h1}} = \frac{\sqrt{w_h}}{1 - \lambda Q_h \bar{x}_h}.$$

The new weights are given by the following equation:

$$\Omega_{h1} = \frac{w_h}{\left(1 - \lambda Q_h \bar{x}_h\right)^2}. \tag{7}$$

To find λ, substitute Equation (7) into the constraint given in Equation (5)

$$\sum_{h=1}^{L} \frac{w_h \bar{x}_h}{\left(1 - \lambda Q_h \bar{x}_h\right)^2} = \sum_{h=1}^{L} w_h \overline{X}_h. \tag{8}$$

The value of λ can be found using numerical analysis methods, and Newton-Raphson method was used to find λ using MATLAB.
As a result, the first proposed calibration estimator is:

$$\bar{y}_{sp1} = \sum_{h=1}^{L} \frac{w_h\, \bar{x}_h\, \bar{y}_h}{\left(1 - \lambda Q_h \bar{x}_h\right)^2 \bar{X}_h}. \tag{9}$$

To find the bias of the estimator $\left(\overline{y}_{sp1}\right)$

$$\text{Bias}\left(\overline{y}_{sp1}\right) = E\left(\overline{y}_{sp1}\right) - \overline{Y} \tag{10}$$

We used numerical methods and mathematical analysis to obtain an approximate formula for the bias:

$$\text{Bias}\left(\overline{y}_{sp1}\right) = \sum_{h=1}^{L} w_h \overline{Y}_h \gamma_h \rho_{hxy} C_{hx} C_{hy} + 2M \left[\frac{\sum_{h=1}^{L} w_h^2 Q_h \overline{X}_h^3 \gamma_h C_{hx}^2}{B^2} - \frac{3D \sum_{h=1}^{L} w_h^2 \overline{X}_h^2 \gamma_h C_{hx}^2}{8B^3} \right]$$
$$- \frac{\sum_{h=1}^{L} w_h^2 Q_h \overline{X}_h^2 \overline{Y}_h \gamma_h \left[2C_{hx}^2 + \rho_{hxy} C_{hx} C_{hy}\right]}{B} + \frac{3N}{4B^2} \sum_{h=1}^{L} w_h^2 \overline{X}_h^2 \gamma_h C_{hx}^2. \tag{11}$$

Where:

$$B = \sum_{h=1}^{L} w_h Q_h \overline{x}_h^2, \qquad M = \sum_{h=1}^{L} w_h Q_h \overline{X}_h \overline{Y}_h.$$

This is the second-order $\overline{y}_{sp1}$ estimator bias formula.

2.3. proposed calibrated estimator of the population mean of product – type based on the relative quadratic distance function

The second estimator was proposed in the following form:

$$\bar{y}_{sp2} = \sum_{h=1}^{L} \Omega_{h2} \frac{\bar{x}_h\, \bar{y}_h}{\bar{X}_h}$$

Based on the relative quadratic distance function as shown in Equation (12). This function aims to reduce the relative deviations between the calibrated weights and the original weights, instead of focusing on the absolute differences between them. This allows for the evaluation of the change in the weight of each layer relative to its original weight. The function is characterized by its non-negative value, given its reliance on the squares of the deviations and

positive weighting coefficients. It also provides a clear mathematical criterion for determining the calibrated weights that reduce the distance between the new and original weights.

$$\sum_{h=1}^{L} \frac{1}{Q_h} \left(\frac{\Omega_{h2}}{w_h} - 1 \right)^2 . \tag{12}$$

Ω_{h2} represents the assumed calibration weights, chosen so that the sum of the relative quadratic distance function given in Equation (12) is as small as possible, subject to the calibration constraint given in Equation (13).

$$\sum_{h=1}^{L} \Omega_{h2} \left(\frac{\bar{x}_h}{c_{xh}} \right) = \sum_{h=1}^{L} w_h \left(\frac{\bar{X}_h}{C_{xh}} \right) . \tag{13}$$

Where

$$c_{xh} = \frac{s_{xh}}{\bar{x_h}}, \qquad C_{xh} = \frac{S_{xh}}{\bar{X}_h}, \qquad S_{hx}^2 = \frac{\sum_{i=1}^{N_h} \left(X_{hi} - \bar{X}_h \right)^2}{N_h - 1}.$$

The Lagrange multiplier is used to minimum Equation (12) when considering the constraint in Equation (13). Defining the Lagrange multiplier λ , the Lagrange function is given by the following:

$$L = \sum_{h=1}^{L} \frac{1}{Q_h} \left(\frac{\Omega_{h2}}{w_h} - 1 \right)^2 - 2\lambda \left[\sum_{h=1}^{L} \Omega_{h2} \left(\frac{\bar{x}_h}{c_{xh}} \right) - \sum_{h=1}^{L} w_h \left(\frac{\bar{X}_h}{C_{xh}} \right) \right] . \tag{14}$$

To obtain the adjusted calibration weights Ω_{h2}, the Lagrange function is derived with respect to Ω_{h2} and set equal to zero.

$$\frac{\partial}{\partial \Omega_{h2}} \sum_{h=1}^{L} \frac{1}{Q_h} \left(\frac{\Omega_{h2}}{w_h} - 1 \right)^2 - 2\lambda \left[\sum_{h=1}^{L} \Omega_{h2} \left(\frac{\overline{x}_h}{c_{x_h}} \right) - \sum_{h=1}^{L} w_h \left(\frac{\overline{X}_h}{C_{x_h}} \right) \right]$$

$$\frac{2}{Q_h} \left(\frac{\Omega_{h2}}{w_h} - 1 \right) \left(\frac{1}{w_h} \right) - 2\lambda \left(\frac{\overline{x}_h}{c_{x_h}} \right) = 0.$$

$$\frac{1}{Q_h w_h} \left(\frac{\Omega_{h2}}{w_h} - 1 \right) = \lambda \left(\frac{\overline{x}_h}{c_{x_h}} \right) .$$

$$\frac{\Omega_{h2}}{w_h} = 1 + \lambda w_h Q_h \left(\frac{\overline{x}_h}{c_{x_h}} \right) .$$

The new weights are given by the following equation:

$$\Omega_{h2} = w_h + \lambda w_h^2 Q_h \left(\frac{\overline{x}_h}{c_{x_h}} \right) . \tag{15}$$

To find λ, we substitute Equation (15) into the constraint given in Equation (13):

$$\sum_{h=1}^{L} \left[w_h + \lambda w_h^2 Q_h \left(\frac{\overline{x}_h}{c_{x_h}} \right) \right] \left(\frac{\overline{x}_h}{c_{x_h}} \right) = \sum_{h=1}^{L} w_h \left(\frac{\overline{X}_h}{C_{x_h}} \right) .$$

$$\lambda = \frac{\sum_{h=1}^{L} w_h \left(\frac{\overline{X}_h}{C_{x_h}}\right) - \sum_{h=1}^{L} w_h \left(\frac{\overline{x}_h}{c_{x_h}}\right)}{\sum_{h=1}^{L} w_h^2 Q_h \left(\frac{\overline{x}_h^2}{c_{x_h}^2}\right)}. \tag{16}$$

We substitute Equation (16) into Equation (15), we get:

$$\Omega_{h2} = \&w_h + \frac{w_h^2 Q_h \left(\frac{\overline{x}_h}{c_{x_h}}\right)}{\sum_{h=1}^{L} w_h^2 Q_h \left(\frac{\overline{x}_h^2}{c_{x_h}^2}\right)} \sum_{h=1}^{L} w_h \left(\frac{\overline{X}_h}{C_{x_h}}\right) - \sum_{h=1}^{L} w_h \left(\frac{\overline{x}_h}{c_{x_h}}\right). \tag{17}$$

Consequently, the second proposed calibration estimator is:

$$\bar{y}_{sp2} = \sum_{h=1}^{L} w_h \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h} + \frac{\sum_{h=1}^{L} w_h \left(\frac{\bar{X}_h}{C_{xh}}\right) - \sum_{h=1}^{L} w_h \left(\frac{\bar{x}_h}{c_{xh}}\right)}{\sum_{h=1}^{L} w_h^2 Q_h \left(\frac{\bar{x}_h^2}{c_{xh}^2}\right)} w_h^2 Q_h \left(\frac{\bar{x}_h}{c_{xh}}\right) \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h}. \tag{18}$$

To find the bias of the estimator $\left(\overline{y}_{sp2}\right)$

$$\text{Bias}\left(\overline{y}_{sp2}\right) = E\left(\overline{y}_{sp2}\right) - \overline{Y} \tag{19}$$

We used numerical methods and mathematical analysis to obtain an approximate formula for the bias:

$$\begin{aligned}
\text{Bias}\left(\overline{y}_{sp2}\right) = & \sum_{h=1}^{L} w_h \overline{Y}_h \gamma_h \rho_{hxy} C_{hx} C_{hy} \\
& - \frac{1}{D_0} \sum_{h=1}^{L} w_h^3 Q_h Z_h^2 \overline{Y}_h \gamma_h \left(6C_{hx}^2 + 2\rho_{hxy} C_{hx} C_{hy} - \frac{5}{2}\rho_{hxs} C_{hx} C_{hs} \right. \\
& \qquad \left. - \frac{1}{2}\rho_{hsy} C_{hy} C_{hs} + \frac{1}{4} C_{hs}^2 \right) \\
& - \frac{P_0}{D_0} \sum_{h=1}^{L} w_h Z_h \left(\gamma_h C_{hx}^2 + \frac{3}{8}\gamma_h C_{hs}^2 - \gamma_h \rho_{hxs} C_{hx} C_{hs}\right) \\
& + \frac{2P_0}{D_0^2} \sum_{h=1}^{L} w_h^3 Q_h Z_h^3 \gamma_h \left(4C_{hx}^2 - 2\rho_{hxs} C_{hx} C_{hs} + \frac{1}{4} C_{hs}^2\right).
\end{aligned} \tag{20}$$

Where,

$$D_0 = \sum_{h=1}^{L} w_h^2 Q_h Z_h^2, \qquad P_0 = \sum_{h=1}^{L} w_h^2 Q_h Z_h \overline{Y}_h, \qquad Z_h = \frac{\overline{X}_h^2}{S_{hx}}.$$

2.4. proposed calibrated estimator of the population mean of product – type based on the Kullback–Leibler distance function

The third proposed estimator was formulated as follows:

$$\bar{y}_{sp3} = \sum_{h=1}^{L} \Omega_{h3} \frac{\bar{x}_h \, \bar{y}_h}{\bar{X}_h}. \tag{21}$$

Based on the Kullback–Leibler distance function as shown in formula Equation (22). This function is used in statistical calibration to adjust original weights proportionally to achieve calibration constraints using known auxiliary information. This method is characterized by producing positive weights when the original weights are positive, and by the ability to adjust weights proportionally rather than adding absolute differences. However, it may produce large weights in some cases.

$$\sum_{h=1}^{L} \frac{1}{Q_h} \left[\Omega_{h3} \log\left(\frac{\Omega_{h3}}{w_h} \right) - \Omega_{h3} + w_h \right]. \tag{22}$$

Ω_{h3} are assumed calibration weights chosen so that the sum of the distance scale of type (Kullback–Leibler distance function) given in Equation (22) is as small as possible. These are subject to the calibration constraints given in Equations (23) and (24).

$$\sum_{h=1}^{L} \Omega_{h3} \, \bar{x}_h = \bar{X}. \tag{23}$$

$$\sum_{h=1}^{L} \Omega_{h3} \, c_{xh} = \sum_{h=1}^{L} w_h \, C_{xh}. \tag{24}$$

The Lagrange multiplier is used to minimum Equation (22) when considering the constraints in Equations (23) and (24). Defining the Lagrange multipliers λ_1, λ_2 , the Lagrange function is given by the following:

$$L = \sum_{h=1}^{L} \frac{1}{Q_h} \left[\Omega_{h3} \log\left(\frac{\Omega_{h3}}{w_h} \right) - \Omega_{h3} + w_h \right] - \lambda_1 \left(\sum_{h=1}^{L} \Omega_{h3} \bar{x}_h - \bar{X} \right) - \lambda_2 \left(\sum_{h=1}^{L} \Omega_{h3} c_{xh} - \sum_{h=1}^{L} w_h C_{xh} \right). \tag{25}$$

To obtain the adjusted calibration weights Ω_{h3}, the Lagrange function is derived with respect to the calibrated weight and set equal to zero.

$$\frac{\partial}{\partial \Omega_{h3}} \sum_{h=1}^{L} \frac{1}{Q_h} \left[\Omega_{h3} \log\left(\frac{\Omega_{h3}}{w_h} \right) - \Omega_{h3} + w_h \right] - \lambda_1 \left(\sum_{h=1}^{L} \Omega_{h3} \overline{x}_h - \overline{X} \right) - \lambda_2 \left(\sum_{h=1}^{L} \Omega_{h3} c_{xh} - \sum_{h=1}^{L} w_h C_{xh} \right).$$

$$\frac{1}{Q_h} \log\left(\frac{\Omega_{h3}}{w_h} \right) - \lambda_1 \overline{x}_h - \lambda_2 c_{xh} = 0.$$

$$\log\left(\frac{\Omega_{h3}}{w_h} \right) = \lambda_1 \overline{x}_h Q_h + \lambda_2 c_{xh} Q_h.$$

$$\frac{\Omega_{h3}}{w_h} = \exp\left(\lambda_1 \overline{x}_h Q_h + \lambda_2 c_{xh} Q_h \right).$$

The new weights are given by the following equation:

$$\Omega_{h3} = w_h \exp(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{xh} Q_h)\,. \tag{26}$$

To find λ_1, λ_2, we substitute Equation (26) into the calibration constraints given in Equations (23) and (24).

$$\sum_{h=1}^{L} w_h \exp(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{xh} Q_h)\, \bar{x}_h = \bar{X}. \tag{27}$$

$$\sum_{h=1}^{L} w_h \exp(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{xh} Q_h)\, c_{xh} = \sum_{h=1}^{L} w_h C_{xh}. \tag{28}$$

λ_1, λ_2 can be found using numerical analysis methods, and the Newton-Raphson method was used to find them using MATLAB. Consequently, the third proposed calibration estimator is:

$$\bar{y}_{sp3} = \sum_{h=1}^{L} w_h \exp(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{xh} Q_h)\, \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h}. \tag{29}$$

To find the bias of the estimator $\left(\overline{y}_{sp3}\right)$

$$\text{Bias}\left(\overline{y}_{sp3}\right) = E\left(\overline{y}_{sp3}\right) - \overline{Y}. \tag{30}$$

We used numerical methods and mathematical analysis to obtain an approximate formula for the bias:

$$\begin{aligned} \text{Bias}\left(\overline{y}_{sp3}\right) = \sum_{h=1}^{L} w_h \overline{Y}_h \Bigg[& \gamma_h \rho_{hxy} C_{hx} C_{hy} + r_h w_h \overline{X}_h \gamma_h C_{hx}^2 + s_h w_h C_{hx} K_{hx} \\ & + r_h w_h \overline{X}_h \gamma_h \rho_{hxy} C_{hx} C_{hy} + s_h w_h C_{hx} K_{hy} \\ & + \frac{1}{2}\left(r_h^2 V_1 + 2 r_h s_h V_2 + s_h^2 V_3\right) \\ & + Q_h \overline{X}_h \left(\frac{-C R_1 + B R_2}{D}\right) + Q_h C_{hx} \left(\frac{B R_1 - A R_2}{D}\right) \Bigg]. \end{aligned} \tag{31}$$

Where,

$$r_h = \frac{Q_h\left(B C_{hx} - C \overline{X}_h\right)}{D}, \qquad s_h = \frac{Q_h\left(B \overline{X}_h - A C_{hx}\right)}{D}.$$

$$K_{hy} = \gamma_h C_{hx} \left[-\rho_{hxy} C_{hx} C_{hy} + \frac{1}{2} \rho_{hys} C_{hy} C_{hs}\right].$$

$$K_{hx} = \gamma_h C_{hx} \left[-C_{hx}^2 + \frac{1}{2} \rho_{hxs} C_{hx} C_{hs}\right].$$

$$K_{hc} = \gamma_h C_{hx}^2 \left[C_{hx}^2 - \rho_{hxs} C_{hx} C_{hs} - \frac{1}{4} C_{hs}^2\right].$$

Let

$$A = \sum_{h=1}^{L} w_h Q_h \overline{X}_h^2, \quad B = \sum_{h=1}^{L} w_h Q_h \overline{X}_h C_{hx},$$
$$C = \sum_{h=1}^{L} w_h Q_h C_{hx}^2, \quad D = AC - B^2.$$

$$V_1 = \sum_{h=1}^{L} w_h^2 \overline{X}_h^2 \gamma_h C_{hx}^2,$$
$$V_2 = \sum_{h=1}^{L} w_h^2 \overline{X}_h C_{hx} K_{hx},$$
$$V_3 = \sum_{h=1}^{L} w_h^2 C_{hx}^2 K_{hc}.$$

$$R_1 = \sum_{h=1}^{L} w_h \left[\overline{X}_h E\left(T_h e_{hx}\right) + \frac{1}{2} \overline{X}_h E\left(T_h^2\right) \right].$$

$$R_2 = \sum_{h=1}^{L} w_h \left[M_h E\left(T_h d_h\right) + \frac{1}{2} C_{hx} E\left(T_h^2\right) \right].$$

$$E\left(T_h^2\right) = r_h^2 V_1 + 2 r_h s_h V_2 + s_h^2 V_3.$$

$$E\left(T_h d_h\right) = w_h \overline{X}_h K_{hx} + s_h w_h C_{hx} K_{hc}.$$

$$E\left(T_h e_{hx}\right) = r_h w_h \overline{X}_h \gamma_h C_{hx}^2 + s_h w_h C_{hx} K_{hx}.$$

$$M_h = \gamma_h C_{hx} \left[C_{hx}^2 - \frac{1}{2} \rho_{hxs} C_{hx} C_{hs} - \frac{1}{8} C_{hs}^2 \right].$$

This is the bias formula for the second-order $\overline{y}_{sp3}$ estimator.

3. proposed Variances of calibraited estimators for the population mean of product – type in Stratified Random Sampling

3.1. Variance of the calibrated estimator of the population mean of product – type based on the Hellinger-type distance function

proposed Variance of the calibrated estimator of the population mean of product type is given by the following formula:

$$\hat{V}(\bar{y}_{sp1}) = \sum_{h=1}^{L} \frac{\Omega_{h1}^{*} \, \Omega_{h1}^{2}}{w_h^2} \, s_p. \tag{32}$$

Based on the Hellinger-type distance function,

$$\sum_{h=1}^{L} \frac{2\left(\sqrt{\Omega_{h1}^{*}}-\sqrt{K_h}\right)^2}{Q_h}. \tag{33}$$

The hypothetical calibration weights $\Omega_{h1}^{*}$are chosen so that the sum of the Hellinger-type distance functions given in Equation (33) is as small as possible, and are subject to the calibration constraint given in Equation (34).

$$\sum_{h=1}^{L} \Omega_{h1}^{*}\, s_{xh}^{2} = V(\bar{x}_{st})\,. \tag{34}$$

Where

$$V(\bar{x}_{st}) = \sum_{h=1}^{L} w_h^2 \frac{s_{hx}^2}{n_h}\,(1-f_h)\,.$$

The Lagrange multiplier is used to minimum Equation (33) when considering the constraint in Equation (34). Defining the Lagrange multiplier λ , the Lagrange function is given by the following:

$$L = \sum_{h=1}^{L} \frac{2\left(\sqrt{\Omega_{h1}^{*}}-\sqrt{K_h}\right)^2}{Q_h} - 2\lambda\left[\sum_{h=1}^{L}\Omega_{h1}^{*}s_{xh}^{2} - V(\bar{x}_{st})\right]. \tag{35}$$

To obtain the optimal value of Ω_{h1}^{*}, the Lagrange function is derived with respect to Ω_{h1}^{*} and equal to zero. The new weights are given by the following equation:

$$\Omega_{h1}^{*} = \frac{K_h}{\left(1-\lambda Q_h s_{xh}^{2}\right)^2}. \tag{36}$$

where:

$$K_h = \frac{w_h^2\,(1-f_h)}{n_h}$$

To find λ, we substitute Equation (36) into the constraint given in Equation (34).

$$\sum_{h=1}^{L} \frac{K_h\, s_{xh}^{2}}{\left(1-\lambda Q_h s_{xh}^{2}\right)^2} = V(\bar{x}_{st})\,. \tag{37}$$

λ can be found using numerical analysis methods and the Newton-Raphson method was used to find it using MATLAB. Consequently, the variance of the first proposed calibration estimator is:

$$\hat{V}(\bar{y}_{sp1}) = \sum_{h=1}^{L} \frac{K_h\,\Omega_{h1}^{2}\,s_p}{w_h^2\left(1-\lambda Q_h s_{xh}^{2}\right)^2}. \tag{38}$$

3.2. Variance of calibrated estimator of the population mean of product – type based on the relative quadratic distance function

Proposed variance of calibrated estimator of the population mean of product – type according to the following formula:

$$\hat{V}(\bar{y}_{sp2}) = \sum_{h=1}^{L} \frac{\Omega_{h2}^{*}\,\Omega_{h2}^{2}}{w_h^2}\,S_p \tag{39}$$

Based on the relative quadratic distance function,

$$\sum_{h=1}^{L} \frac{1}{Q_h} \left(\frac{\Omega_{h2}^*}{K_h} - 1 \right)^2 . \tag{40}$$

The hypothetical calibration weights Ω_{h2}^* are chosen so that the sum of the relative quadratic distance functions (RMPD) given in Equation (40) is as small as possible, subject to the calibration constraint given in Equation (41).

$$\sum_{h=1}^{L} \Omega_{h2}^* \left(\frac{s_{hx}^2}{c_{hx}^2} \right) = \sum_{h=1}^{L} \frac{1}{C_{hx}^2} V(\bar{x}_{st}) . \tag{41}$$

The Lagrange multiplier is used to minimum Equation (40) when considering the constraint in Equation (41). Defining the Lagrange multiplier λ , the Lagrange function is given by the following:

$$L = \sum_{h=1}^{L} \frac{1}{Q_h} \left(\frac{\Omega_{h2}^*}{K_h} - 1 \right)^2 - 2\lambda \left[\sum_{h=1}^{L} \Omega_{h2}^* \left(\frac{s_{hx}^2}{c_{hx}^2} \right) - \sum_{h=1}^{L} \frac{1}{C_{hx}^2} V(\bar{x}_{st}) \right] . \tag{42}$$

To obtain the optimal value of Ω_{h2}^*, the Lagrange function is derived with respect to Ω_{h2}^*, and equal to zero. The new weights are given by the following equation:

$$\Omega_{h2}^* = D_h + \lambda K_h^2 Q_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right) . \tag{43}$$

To find λ, we substitute Equation (43) into the constraint given in Equation (41).

$$\sum_{h=1}^{L} \left[K_h + \lambda K_h^2 Q_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right) \right] \left(\frac{s_{hx}^2}{c_{hx}^2} \right) = \sum_{h=1}^{L} \frac{1}{C_{hx}^2} V(\bar{x}_{st}) . \tag{44}$$

$$\lambda = \frac{\displaystyle\sum_{h=1}^{L} \frac{1}{C_{hx}^2} V(\bar{x}_{st}) - \sum_{h=1}^{L} K_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right)}{\displaystyle\sum_{h=1}^{L} K_h^2 Q_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right)^2} . \tag{45}$$

$$\Omega_{h2}^* = K_h + \frac{K_h^2 Q_h \left(\frac{s_{hx}^2}{C_{hx}^2} \right)}{\displaystyle\sum_{h=1}^{L} K_h^2 Q_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right)^2} \left[\sum_{h=1}^{L} \frac{1}{C_{hx}^2} V(\bar{x}_{st}) - \sum_{h=1}^{L} K_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right) \right] . \tag{46}$$

Consequently, the variance of the second proposed calibration estimator is:

$$\hat{V}(\bar{y}_{sp2}) = \sum_{h=1}^{L} \frac{\Omega_{h2}^2}{w_h^2} \left[K_h + \frac{K_h^2 Q_h \left(\frac{s_{hx}^2}{C_{hx}^2} \right)}{\displaystyle\sum_{h=1}^{L} K_h^2 Q_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right)} \left(\sum_{h=1}^{L} \frac{1}{C_{hx}^2} V(\bar{x}_{st}) - \sum_{h=1}^{L} K_h \left(\frac{s_{hx}^2}{c_{hx}^2} \right) \right) \right] s_p . \tag{47}$$

3.3. Variance of the calibrated estimator of the population mean of product type based on the Kullback–Leibler distance function

proposed variance of the calibrated estimator of the population mean of product type is givin by the following formula:

$$\hat{V}(\bar{y}_{sp3}) = \sum_{h=1}^{L} \frac{\Omega_{h3}^{*}\,\Omega_{h3}^{2}}{w_h^2}\, s_p. \tag{48}$$

Based on the Kullback–Leibler distance function,

$$\sum_{h=1}^{L} \frac{1}{Q_h} \left[\Omega_{h3}^{*} \log\left(\frac{\Omega_{h3}^{*}}{K_h}\right) - \Omega_{h3}^{*} + K_h \right]. \tag{49}$$

The calibration weights Ω_{h3}^{*} are chosen so that the sum of the Kullback–Leibler distance function given in Equation (49) is as small as possible, subject to the calibration constraint given in Equations (50) and (51).

$$\sum_{h=1}^{L} \Omega_{h3}^{*}\, \bar{x}_h = V(\bar{x}_{st})\,. \tag{50}$$

$$\sum_{h=1}^{L} \Omega_{h3}^{*}\, c_{hx}^{2} = \sum_{h=1}^{L} K_h\, C_{hx}^{2}. \tag{51}$$

The Lagrange multiplier is used to minimum Equation (49) when considering the constraints in Equations (50) and (51). Defining the Lagrange multipliers λ_1, λ_1 , the Lagrange function is given by the following:

$$L = \sum_{h=1}^{L} \frac{1}{Q_h} \left[\Omega_{h3}^{*} \log\left(\frac{\Omega_{h3}^{*}}{K_h}\right) - \Omega_{h3}^{*} + K_h \right] - \lambda_1 \left(\sum_{h=1}^{L} \Omega_{h3}^{*} \bar{x}_h - V(\bar{x}_{st}) \right) - \lambda_2 \left(\sum_{h=1}^{L} \Omega_{h3}^{*} c_{hx}^{2} - \sum_{h=1}^{L} K_h C_{hx}^{2} \right) \tag{52}$$

To obtain the optimal value of Ω_{h3}^{*}, the Lagrange function is derived with respect to Ω_{h3}^{*} and equal to zero. The new weights are given by the following equation:

$$\Omega_{h3}^{*} = K_h \exp\left(\lambda_1\, \bar{x}_h\, Q_h + \lambda_2\, c_{hx}^{2}\, Q_h\right) \tag{53}$$

To find λ_1, λ_2 we substitute Equation (53) into the constraints given in Equations (50) and (51).

$$\sum_{h=1}^{L} \left[K_h \exp\left(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{hx}^{2} Q_h\right) \right] \bar{x}_h = V(\bar{x}_{st})\,. \tag{54}$$

$$\sum_{h=1}^{L} \left[K_h \exp\left(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{hx}^{2} Q_h\right) \right] c_{hx}^{2} = \sum_{h=1}^{L} K_h C_{hx}^{2}. \tag{55}$$

λ_1, λ_2 can be found using numerical analysis methods. The Newton-Raphson method was used to find them using MATLAB.
Consequently, the variance of the proposed third calibration estimator is:

$$\hat{V}(\bar{y}_{sp3}) = \sum_{h=1}^{L} \frac{K_h \exp\left(\lambda_1 \bar{x}_h Q_h + \lambda_2 c_{hx}^2 Q_h\right) \Omega_{h3}^2}{w_h^2} s_p. \tag{56}$$

4. Comparison Criterions

4.1. Mean Squared Error (MSE)

The proposed calibration estimators are compared based on the mean squared error (MSE) criterion, which can be calculated using the following equation [28]:

$$\text{MSE} = \frac{1}{R} \sum_{i=1}^{R} \left(\bar{y}_i - \bar{Y}\right)^2, \qquad i = 1, 2, 3, \ldots, R. \tag{57}$$

Where: $\bar{y}_i$: sample mean, $\bar{Y}$: population mean, and R: number of times the experiment is repeated (No. of samples).

4.2. Relative Efficiency

The relative efficiency of any $\bar{y}_{spj}$ estimator in stratified random sampling is defined as follows [29]:

$$RE = \frac{\text{MSE}(\bar{y}_{p0})}{\text{MSE}\left(\bar{y}_{sp(j)}\right)} \times 100, \qquad j = 1, 2, 3. \tag{58}$$

Where: $\text{MSE}(\bar{y}_{p0})$ and $\text{MSE}\left(\bar{y}_{sp(j)}\right)$ are the mean squared errors of the conventional estimator and the estimator under comparison, respectively.

5. Results and Discussion

5.1. Simulation Study

5.1.1. Simulation Design:

For the purpose of comparing the proposed calibrated estimators of product-type for estimating the population $mean$ in stratified random sampling, a population consisting of two variables, M and Z, were generated from a standard normal distribution with $mean$ zero and $variance$ of one, and with different sizes $N = 100, 300, 900$. The population was divided into three strata as shown in Table 2. The study variable Y and the auxiliary variable X were then generated according to the following formulas (N. Koyuncu & Kadilar, C.) [30], the two experiments used different sample sizes

$$Y_{hi} = a_h + Z_{hi} \tag{59}$$

$$X_{hi} = b_h + \sqrt{1 - P_{XYh}^2}\, M_{hi} + P_{xy} \left(\frac{M_h}{Z_h}\right) Z_{hi} \quad , i = 1, 2, 3, \rightarrow n, \quad h = 1, 2, 3 \tag{60}$$

First experiment: Data were generated for populations of size $N = 100, 300, and 900$, and the population were divided into three strata, as shown in Table 2.

Table 2. Population and sample allocation across strata in simulation study.

Population Size	**Stratum 1**	**Stratum 2**	**Stratum 3**
$N = 100$	$N_1 = 34$	$N_2 = 33$	$N_3 = 33$
	$n_1 = 17$	$n_2 = 12$	$n_3 = 20$
$N = 300$	$N_1 = 100$	$N_2 = 100$	$N_3 = 100$
	$n_1 = 50$	$n_2 = 30$	$n_3 = 40$
$N = 900$	$N_1 = 300$	$N_2 = 300$	$N_3 = 300$
	$n_1 = 150$	$n_2 = 90$	$n_3 = 110$

Three cases were studied using different standard deviations and correlations for the two variables, as shown in Table (3). The third case represents the actual statistics of the data used in this research (Data1).

Table 3. Parameter values for generating the study and auxiliary variables across the three cases and strata.

Stratum	a_h	b_h	**Case 1**			**Case 2**			**Case 3**		
			S_m	S_z	ρ_{XY}	S_m	S_z	ρ_{XY}	S_m	S_z	ρ_{XY}
N_1	1	2	1.20	3.40	-0.30	1.20	3.40	-0.70	6.00	3.51	-0.75
N_2	2	1	2.10	2.80	-0.50	2.10	2.80	-0.80	13.83	6.11	-0.73
N_3	2	2	1.80	3.20	-0.70	1.80	3.20	-0.90	19.48	8.94	-0.50

Second Experiment: To evaluate the performance of the proposed calibration estimators as the population size and the number of strata increase, a simulation experiment was conducted using generated data with a population size of $N = 2000$, divided into five strata. The behavior of the estimators was examined in terms of bias, MSE, and relative efficiency. Three scenarios were studied, employing different standard deviations and correlations for the two variables, as shown in Table (4). The third scenario represents the actual statistics of the data used in this research (Data2). Simple random samples without replacement were drawn $10,000$ times, using various sample sizes as indicated in Table (4). The estimators were compared based on the mean squared error (MSE), bias, and relative efficiency, utilizing the MATLAB software.

Table 4. Parameter Values for Generating the Study and Auxiliary Variables Across the Three Cases and Strata

Stratum	N_h	n_h	a_h	b_h	**Case 1**			**Case 2**			**Case 3**		
					S_m	S_z	ρ_{XY}	S_m	S_z	ρ_{XY}	S_m	S_z	ρ_{XY}
N_1	347	139	1	2	1.20	3.40	-0.30	1.20	3.40	-0.70	3.03	16.87	-0.64
N_2	612	245	2	1	2.10	2.80	-0.50	2.10	2.80	-0.80	1.05	10.09	-0.60
N_3	121	48	2	2	1.80	3.20	-0.70	1.80	3.20	-0.90	0.94	12.75	-0.26
N_4	200	80	1	2	4.80	3.10	-0.60	4.80	3.10	-0.90	1.28	11.83	-0.26
N_5	720	288	2	1	2.60	2.70	-0.70	2.60	2.70	-0.95	3.87	15.29	-0.60

*5.1.2. **Simulation Results and Discussion:***

5.1.2.1. Results of First Experiment Simulation

Table (5) shows that $\bar{y}_{sp1}$ achieved the lowest mean squared error (MSE) values across all population sizes and the three scenarios, indicating its superiority in terms of accuracy. This can be explained by the fact that the Hellinger-type distance function maintains adjusted weights close to the original weights while satisfying the calibration constraint, thereby striking a balance between reducing bias and preventing a significant increase in variance. Furthermore, the use of a single constraint renders the calibration process more stable and reduces weight fluctuations across different samples. As for the second estimator, based on the relative quadratic distance, it showed limited improvement in some cases, while in others, it underperformed compared to the traditional estimator, $\bar{y}_{sp0}$. This is because the relative quadratic distance function relies on the relative errors of the weights, making it more sensitive to variations in the coefficient of variation across strata; consequently, its performance was not consistent across all scenarios. This implies that employing the relative quadratic distance function does not, by itself, guarantee an improvement in accuracy, rather, performance depends on how well the chosen function and constraints align with the characteristics of the strata. In contrast, $\bar{y}_{sp3}$ which utilizes the Kullback–Leibler distance function and two calibration constraints did not achieve the best MSE values. In fact, it recorded the highest MSE among the four estimators across all cases and population sizes presented in the table. While increasing the number of constraints forces the weights to satisfy more conditions, it can also lead to greater volatility when samples vary, which is reflected in the higher MSE observed in most simulation scenarios. The results in the table indicate that the MSE values for the traditional estimator $\bar{y}_{sp0}$ and the estimator $\bar{y}_{sp1}$ decrease as the population size N increases from $100 to 300$, and subsequently to 900, across all three scenarios. This decline is attributed to the fact that the sampling designs employed involve larger sample sizes as the population size grows, which contributes to a reduction in sampling error. In contrast, $\bar{y}_{sp2}$ did not exhibit a consistent decrease across all cases when moving from one population size to another. Meanwhile, the MSE values for the third estimator decreased as the population size increased; however, they remained higher than those of the traditional estimator $\bar{y}_{sp0}$ in all the comparisons presented.

Table 5. MSE of the proposed calibrated estimators for the first simulation experiment.

Estimator	$N = 100$			$N = 300$			$N = 900$		
	Case 1	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**
$\bar{y}_{p0}$	0.02158	0.01165	0.06340	0.01113	0.00587	0.01877	0.00341	0.00176	0.00747
$\bar{y}_{sp1}$	0.01536	0.00891	0.03646	0.00778	0.00452	0.01059	0.00248	0.00140	0.00390
$\bar{y}_{sp2}$	0.02111	0.01625	0.04757	0.01620	0.01333	0.02165	0.00516	0.00395	0.00703
$\bar{y}_{sp3}$	0.07779	0.04332	0.21363	0.02631	0.01847	0.09177	0.00743	0.00493	0.02422

The results of the first simulation presented in Table (6), show that the traditional estimator $\bar{y}_{p0}$ achieved the lowest empirical absolute bias in most of the studied cases; this indicates that $\bar{y}_{p0}$ was closest to the population mean in terms of bias. While the use of the Hellinger-type distance function in $\bar{y}_{sp1}$ did not generally reduce bias compared to $\bar{y}_{p0}$ despite the difference being relatively small $\bar{y}_{sp1}$ did achieve the lowest bias for a population size of N=900 in both the first and second cases. This suggests that employing the Hellinger-type distance function with a calibration constraint may help reduce bias under specific conditions, though it does not guarantee superiority in every instance, in this specific case, $\bar{y}_{sp1}$ demonstrated a superior ability to reduce bias compared to $\bar{y}_{p0}$. This outcome is linked to the suitability of the Hellinger-type distance function and the calibration constraint for the statistical characteristics of the scenario, with the population size, sample design, and scenario parameters collectively influencing the result. Conversely, $\bar{y}_{sp2}$ recorded higher bias in most comparisons, while the performance of $\bar{y}_{sp3}$ varied depending on the scenario and population size, generally ranking third except in case (1) with sample sizes of N=100 and N=900. The higher bias observed for $\bar{y}_{sp3}$ indicates that its calibration mechanism failed to yield improvements regarding this criterion; such variations can likely be attributed to the interaction between the distance function and calibration constraints with the population characteristics and sample design.

Table 6. Absolute bias of the proposed estimators for the first simulation experiment

Estimators	$N = 100$			$N = 300$			$N = 900$		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
$\bar{y}_{p0}$	0.00380116	0.00898250	0.03870440	0.00529574	0.00668442	0.01269685	0.00261765	0.00361348	0.00986481
$\bar{y}_{sp1}$	0.00484527	0.00933456	0.04628726	0.00646244	0.00744455	0.01682217	0.00232229	0.00316594	0.00993071
$\bar{y}_{sp2}$	0.01629038	0.04111112	0.06345000	0.01601166	0.01764353	0.03337958	0.00464758	0.00656610	0.01480738
$\bar{y}_{sp3}$	0.04762864	0.01881766	0.04887298	0.01435121	0.01375571	0.03280194	0.00474411	0.00367684	0.01372761

Table (7) presents the relative efficiency of the traditional estimator and the proposed estimators across various population sizes. The results demonstrate that the first proposed estimator,$\bar{y}_{sp1}$, achieved the highest relative efficiency values in all cases; its value exceeded unity, indicating its superiority over the traditional estimator $\bar{y}_{p0}$ in terms of efficiency. This reflects the suitability of the Hellinger-type distance function combined with the calibration constraint for improving mean estimation. While the estimator $\bar{y}_{sp2}$ also showed improvement in some instances, its efficiency was inconsistent, falling below that of the traditional estimator $\bar{y}_{p0}$ in other cases. In contrast, the estimator $\bar{y}_{sp3}$recorded values below unity across all simulation scenarios. This indicates that employing the Kullback–Leibler distance function with two calibration constraints failed to improve relative efficiency for this dataset compared to the traditional estimator $\bar{y}_{p0}$, suggesting that increasing the number of constraints does not necessarily enhance efficiency; rather, success depends on the compatibility of the constraints and the distance function with the characteristics of the population under study.

Table 7. The relative efficiency of the proposed estimators for the first simulation experiment.

Estimator	$N = 100$			$N = 300$			$N = 900$		
	Case 1	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**
$\bar{y}_{p0}$	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
$\bar{y}_{p1}$	1.40475	1.30792	1.73890	1.42959	1.29834	1.77256	1.37420	1.25206	1.91221
$\bar{y}_{p2}$	1.02197	0.71706	1.33288	0.68683	0.44025	0.86697	0.66023	0.44494	1.06248
$\bar{y}_{p3}$	0.27734	0.26900	0.29677	0.42287	0.31775	0.20450	0.45881	0.35614	0.30831

5.1.2.2. Results of Second Experiment Simulation

In this experiment, a larger population was used, divided into five strata rather than three with varying stratum and sample sizes. The experiment also incorporates three different scenarios regarding standard deviations and correlations, as shown in Table (4). This setup helps verify whether the primary findings hold true when population sizes increase and when transitioning to a more extensive stratified design. The results inTable (8) indicate that $\bar{y}_{sp1}$ yielded the lowest MSE across all three scenarios. As for the second estimator, it outperformed the traditional estimator, $\bar{y}_{p0}$, in terms of MSE in the first and second scenarios, but performed worse (resulting in a higher MSE) in the third scenario; this confirms that its efficiency varies according to the characteristics of the population. Meanwhile, $\bar{y}_{sp0}$ recorded the highest MSE values in all three scenarios reaching 0.945900 in the third case. Consequently, in this experiment, the use of the Kullback–Leibler distance function, subject to the two specified constraints, did not improve accuracy compared to the traditional estimator,$\bar{y}_{p0}$. The results in Table (8) show that the traditional estimator,$\bar{y}_{p0}$, achieved the lowest empirical absolute bias across all three scenarios, whereas $\bar{y}_{sp2}$ exhibited the highest bias. $\bar{y}_{sp1}$ ranked second after the traditional estimator in all cases, while $\bar{y}_{sp3}$ ranked third except in the first scenario (for population sizes N=100 and N=900). In the third scenario, the absolute bias of the second estimator was significantly higher than that of the traditional estimator; similarly, the absolute bias of the third estimator was notably high compared to its performance in the first and second scenarios. This increase may be linked to the characteristics of the third scenario, specifically the standard deviation values and correlation coefficients used to generate the two variables. If the nature of the relationship between the variables, or the variances within the strata,

does not align well with the weight-adjustment mechanism, then calibration may fail to reduce bias and could, in fact, increase it. These results indicate that the mean of the estimates derived from simulation iterations for the conventional estimator, $\bar{y}_{p0}$, was the closest to the true population mean. Although $\bar{y}_{sp1}$ relies on calibration using the Hellinger - type distance function, this calibration did not result in a lower empirical absolute bias compared to the conventional estimator,$\bar{y}_{p0}$. This can be explained by the fact that adjusting calibration weights does not necessarily guarantee a reduction in bias in every instance; the effect of such an adjustment depends on how well the calibration constraint and the distance function align with the characteristics of the simulated population. The higher bias of $\bar{y}_{sp2}$ compared to both the conventional estimator and $\bar{y}_{sp1}$ suggests that the weights derived from the relative quadratic distance function failed to adequately correct for deviations in this scenario, potentially due to the sensitivity of the calibration weights to relative variations between strata. Furthermore, the higher bias of $\bar{y}_{sp3}$ compared to $\bar{y}_{sp1}$ suggest that employing the Kullback–Leibler distance function with two constraints offered no advantage in terms of bias reduction in this case. The observed higher bias may be attributed to the nature of the constraints, the characteristics of the data, the sizes of the strata, the sample sizes drawn from them, and the degree of intra-stratum variance all of which can influence the results. As for relative efficiency, the results in Table (8) support the previous conclusion, with $\bar{y}_{sp1}$ achieving the highest relative efficiency in all cases. In contrast, the results for the second estimator show inconsistent performance, while the third estimator $\bar{y}_{sp3}$ exhibited an efficiency of less than one in all three cases. These findings indicate that the primary conclusion from the first experiment holds true for the second: $\bar{y}_{sp1}$ is the most efficient estimator in terms of both MSE and relative efficiency among those compared, whereas $\bar{y}_{sp3}$ failed to outperform the conventional estimator $\bar{y}_{p0}$ in any of the scenarios examined.

Table 8. MSE, absolute bias, and relative efficiency of the proposed estimators for the second experiment ($N = 2000$)

Estimators	**MSE**			**Absolute bias**			**Relative efficiency**		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
$\bar{y}_{p0}$	0.001438	0.000743	0.032017	0.000053	0.001490	0.005409	1.000000	1.000000	1.000000
$\bar{y}_{sp1}$	0.000845	0.000622	0.007461	0.002305	0.003139	0.016789	1.703223	1.194114	4.291273
$\bar{y}_{sp2}$	0.003807	0.002782	0.027866	0.010611	0.010889	0.045121	0.377884	0.267169	1.148973
$\bar{y}_{sp3}$	0.006181	0.004052	0.094590	0.004647	0.003759	0.040742	0.232736	0.183408	0.338483

5.2. Real Data Application

5.2.1. Data Description:

To estimate the population mean, the variance of the estimated population mean, and the absolute bias of the product-type estimators under stratified random sampling using calibration, all proposed methods were applied to two datasets. The first dataset (Data1), derived from Singh (2003) [31] with a population size of N=96, consisted of the study variable (Y), "life expectancy at birth,which followed a generalized extreme value distribution, and an auxiliary variable (X), "crude birth rate," which followed a beta distribution, as assessed using the Kolmogorov-Smirnov test in EasyFit-3 software. The population was divided into three strata, from which samples were drawn, and the total sample size represented approximately 40% of the population size. The second dataset (Data2) contained data on small enterprises in several Iraqi governorates, collected from the Authority of Statistics and Geographic Information Systems. It included two variables: (Y), representing actual production volume, and (X), representing the number of employees, for a population of N=500 enterprises. The Kolmogorov- Smirnov test using the Easy Fit-3 software showed that both variables follow a normal distribution. The population was divided into five strata according to the number of workers: stratum1 (8 to 18 workers), stratum2 (18 to 22 workers), stratum 3 (22 to 25 workers), stratum4 (25 to 29 workers), and stratum5 (29 to 47 workers). Samples were drawn from each stratum, as shown in Table (9). The sample size for Data2 was determined as follows:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \qquad n_0 = \frac{\sum w_h s_h^2}{V}$$

$$V = \left(\frac{d}{t}\right)^2$$

Using proportional allocation to find the sample size for the stratum:

$$n_h = \frac{n}{N} N_h = n \cdot w_h$$

Table 9. Stratum sizes, sample sizes, and descriptive statistics for the two empirical data sets

Stratum	**Data 1**					**Data 2**				
	N_h	n_h	s_x	s_y	ρ_{XY}	N_h	n_h	s_x	s_y	ρ_{XY}
1	33	17	6.01	3.51	-0.75	80	48	3.03	16.87	-0.64
2	31	12	13.83	6.11	-0.73	75	45	1.05	10.09	-0.6
3	32	20	19.48	8.94	-0.5	95	57	0.94	12.75	-0.26
4						115	68	1.28	11.83	-0.26
5						135	80	3.87	15.29	-0.6

5.2.2. ***Empirical Results and Discussion for Data 1:***

Table (9) presents the characteristics of the three population strata. A key finding is that the correlation coefficients between the study variable and the auxiliary variable are negative across all strata, with values of $-0.75, -0.73, and -0.5$. These values indicate a moderate to strong negative relationship, particularly in the first and second strata implying that as the values of the auxiliary variable increase, the values of the study variable decrease. This supports the use of product-type calibration estimators, which are specifically designed to handle negative correlations and are expected to yield higher precision than the traditional estimator, $\bar{y}_{p0}$.

Table (10) displays the results for the population mean estimate, absolute bias, and the estimated variance of the estimators under study. A comparison of the estimates with the true population mean of 63.9760, shows that the third proposed estimator yielded the estimate closest to the true value (63.7614), followed by the conventional estimator, $\bar{y}_{p0}$. Meanwhile, $\bar{y}_{sp1}$ was relatively further from the true mean, and $\bar{y}_{sp2}$ exhibited the largest deviation from the true value.

Regarding absolute bias, the results indicate that the conventional estimator, $\bar{y}_{p0}$, achieved the lowest value (0.04453928), followed by $\bar{y}_{sp2}$ in second place, and then $\bar{y}_{sp3}$. The first estimator, based on the Hellinger-type distance function, recorded the highest absolute bias reaching 1.16785025 which is clearly greater than the bias values of the other estimators. Although $\bar{y}_{sp1}$ performed well in the simulation results in terms of mean squared error and relative efficiency, it exhibited the highest theoretical bias when applied to the Data 1. This demonstrates that an estimator's superiority according to one statistical criterion does not necessarily imply its superiority across all criteria. Such performance may be linked to the specific characteristics of the relationship between the study variable and the auxiliary variable in this dataset, as well as the suitability of the calibration constraint and the distance function employed.

Regarding variance, $\bar{y}_{sp3}$ yielded the lowest value, followed by the first estimator and then the traditional estimator $\bar{y}_{p0}$, while $\bar{y}_{sp2}$ exhibited the highest variance. This can be explained by the nature of the distance functions and the constraints employed in deriving the calibration weights. Specifically, $\bar{y}_{sp1}$ utilized the Hellinger-type distance function with a single constraint, which allowed for a smooth adjustment of the weights and resulted in a lower variance than the traditional estimator, a finding consistent with the simulation results, where it demonstrated superior performance in terms of MSE. In contrast, $\bar{y}_{sp2}$ relied on the relative quadratic distance function, which is more sensitive to relative changes between strata; this sensitivity manifested as higher variance the highest among all estimators indicating that the calibration weights associated with this estimator failed to achieve variance reduction in this empirical application. This performance may be attributed to the sensitivity of the distance function formulation to relative shifts between the original and calibrated weights, which can influence the extent of weight

adjustment when stratum characteristics vary a factor reflected in the first application, where $\bar{y}_{sp2}$ recorded the highest variance.

As for $\bar{y}_{sp3}$, it employed the Kullback–Leibler distance function with two calibration constraints, enabling the incorporation of additional auxiliary information. This resulted in the lowest variance and an estimate closest to the true mean for the given dataset. While this is a positive outcome, it contrasts with the estimator's performance in the simulation, where its MSE values were higher than those of all other estimators across all scenarios. The difference in the performance ranking of the estimators between the simulation study and the practical application does not constitute a contradiction in the research findings; rather, it reflects the fact that the performance of calibration estimators depends on population characteristics and the nature of the auxiliary information used to construct the constraints. In the simulation study, virtual populations were generated based on specific parameters including population sizes, standard deviations, and correlation coefficients and the estimators were evaluated using the MSE across a large number of samples. The simulation results indicated that $\bar{y}_{sp1}$ which relies on the Hellinger-type distance function, yielded the lowest MSE values and the highest relative efficiency in the cases examined, whereas $\bar{y}_{sp3}$ based on the Kullback–Leibler distance function, proved less efficient in those scenarios.

In the practical application, by contrast, the analysis involved a single real-world population possessing characteristics distinct from those of the virtual populations used in the simulation and a single sample drawn from it. Consequently, one cannot expect an estimator that achieved the lowest MSE in a simulation to necessarily yield the lowest variance for every real-world population. The empirical data Data1 reveal a negative correlation between the study variable and the auxiliary variable across all three strata as previously noted along with significant variation in the coefficients of variation for the auxiliary variable between these strata. Furthermore, the differences in the probability distributions of variables Y and X have shaped the actual characteristics of the population,tra specifically its dispersion, skewness, and coefficient of variation thereby exerting an indirect influence on the performance of the calibrated product estimators. In addition to information regarding the mean of the auxiliary variable, the estimator $\bar{y}_{sp3}$ incorporates another constraint based on the coefficient of variation. This allows it to leverage further insights into the relative dispersion of the auxiliary variable. The effectiveness of this additional constraint depends on how well it aligns with the actual structure of the data. A key factor explaining the superior performance of $\bar{y}_{sp3}$ is the coefficient of variation; notably, this coefficient is not constant across strata but instead shows a clear, decreasing trend $(0.603 > 0.324 > 0.197)$.
Since $\bar{y}_{sp3}$ employs two calibration constraints one based on the mean of the auxiliary variable and the other on its coefficient of variation it effectively captures additional, substantive information about the variation in the relative dispersion of the auxiliary variable across strata. By utilizing this information,$\bar{y}_{sp3}$ is able to adjust the weights in a manner that is more consistent with the actual data structure. This explains why the value produced by $\bar{y}_{sp3}$ was the closest to the true population mean, whereas the estimates derived from other methods were further removed from it. In simulations, we do not deal with the actual society itself, but rather with virtual societies as previously noted. The additional information represented by the coefficient of variation may not necessarily be consistent with the structure generated for the society. Regarding the third estimator, the inclusion of an additional constraint does not automatically lead to a reduction in the MSE. The weights must satisfy both constraints simultaneously; if these constraints do not align well with the relationship between $XandY$ within the simulated society, forcing the weights to satisfy both may result in larger shifts in the weights themselves, thereby increasing the estimator's variance and, consequently, its MSE.

5.2.3. ***Empirical Results and Discussion for Data 2:***
The results in Table (10) indicate that the four estimators yielded relatively close values for the actual mean of the variable under study, with the true value of the population mean being 112.092. The traditional estimator $\bar{y}_{p0}$ produced the highest estimate, while $\bar{y}_{sp1}$ produced the lowest. The second $\bar{y}_{sp2}$ and third $\bar{y}_{sp3}$ estimators ranked second and third, respectively, following the traditional estimator $\bar{y}_{p0}$. Notably, the mean obtained using $\bar{y}_{sp1}$ is the closest to the population mean.
The results show that the traditional estimator, $\bar{y}_{p0}$, recorded the lowest absolute bias (0.00471948), followed by $\bar{y}_{sp1}$ based on the Hellinger-type distance function and then $\bar{y}_{sp3}$ conversely, $\bar{y}_{sp2}$ exhibited the highest bias (1.27646374).

This suggests that the traditional estimator,$\bar{y}_{p0}$ performed best in terms of theoretical bias for this dataset, while $\bar{y}_{sp1}$ was the most effective among the proposed calibrated estimators according to the bias criterion. The bias value of $\bar{y}_{sp1}$ is relatively close to that of the traditional estimator,$\bar{y}_{p0}$ with a difference of approximately 0.00191711. This indicates that $\bar{y}_{sp1}$ achieved a low theoretical bias for the Data2 dataset compared to the second and third estimators. The higher bias observed for $\bar{y}_{sp2}$ suggests that it was less suitable for this dataset; this may be related to the degree of alignment between the distance function and the constraints employed, and the specific characteristics of the data. The empirical results indicate that the ranking of estimators based on theoretical bias is not consistent across the two datasets. This underscores the importance of considering the population distribution, its characteristics, and the nature of the relationship between the study variable and the auxiliary variable when evaluating calibration estimators.
The table shows that $\bar{y}_{sp1}$ yielded the lowest estimated variance (0.22512), followed by $\bar{y}_{sp3}$ (0.22673), and then the traditional estimator $\bar{y}_{p0}$. In contrast,$\bar{y}_{sp2}$ recorded a variance of 2.14946 significantly higher than that of the other estimators. This implies that $\bar{y}_{sp1}$ achieved a marginal improvement in variance compared to the traditional estimator $\bar{y}_{p0}$, while $\bar{y}_{sp3}$ showed a slight reduction in variance. Conversely, $\bar{y}_{sp2}$ performed poorly by these criteria; the combination of high bias and high variance suggests that $\bar{y}_{sp2}$ was less suitable for this dataset. This may be linked to how well the distance function and the applied constraints align with the data's characteristics. These findings indicate that the use of a relative quadratic distance function did not lead to a consistent improvement in estimation performance, given the specific data characteristics and constraints employed in this study. The high bias and variance observed for $\bar{y}_{sp2}$ are generally consistent with simulation results, which also point to the suboptimal performance of this estimator. Nevertheless, $\bar{y}_{sp2}$ did demonstrate improvement in certain simulation scenarios, suggesting that its efficiency depends on population characteristics, sampling conditions, and the appropriateness of both the distance function and the calibration constraints. However, it differs from some simulation results regarding the ranking of $\bar{y}_{sp1}$ and $\bar{y}_{sp3}$ in terms of variance; this confirms that the performance of estimators may change when transitioning from a virtual generated population to real-world data.

The results from applying the methods to the second real-world dataset (Data2) align with the findings from the simulation study. Specifically, $\bar{y}_{sp1}$ demonstrated the best performance among the estimators examined the simulation results showed that this estimator yielded the lowest empirical mean squared error (MSE) across the scenarios studied, while the application results indicated that it produced the lowest variance and provided an estimate closest to the true population mean. This consistency suggests that the combination of the Hellinger-type distance function and the calibration constraint used in $\bar{y}_{sp1}$ was well-suited to the data characteristics and conditions in both cases. The formulation of the distance function allowed for the determination of the calibrated weights by reducing the distance between them and the original weights while adhering to the imposed calibration constraint. The variance result of the third estimator was consistent with the simulation results. However, the results indicated that $\bar{y}_{sp3}$, based on the Kullback-Liebler distance function and two calibration constraints, and may be more suitable than the second estimator.

The superior performance of $\bar{y}_{sp1}$ over all other estimators in the simulation experiment can be attributed to the nature of the Hellinger-type distance function, which focuses on minimizing the distance between the original weights and the calibrated weights. This process results in an appropriate adjustment of the weights while ensuring they remain close to their original values. Alternatively, this may be attributed to the calibration constraint employed; $\bar{y}_{sp1}$ relies on a single calibration constraint involving information from the auxiliary variable. This constraint may be sufficient to achieve an appropriate balance between utilizing auxiliary information and adjusting the weights, whereas increasing the number of constraints in another estimator does not necessarily guarantee improved performance under all conditions. Another possibility is that its superiority stems from its interaction with the characteristics of the simulation scenarios, as the efficiency of an estimator depends on stratum sizes, sample sizes, standard deviations, and negative correlation coefficients and these conditions were relatively favorable for the performance of $\bar{y}_{sp1}$ within the scenarios. $\bar{y}_{sp1}$ Demonstrated strong performance in both the simulations and the practical application to this dataset, which reinforces its suitability given the population characteristics and conditions under study.

Table 10. Empirical results for Data 1 and Data 2: mean estimates, absolute bias, and estimated variances

Estimators	**Data 1**			**Data 2**		
	Estimates	**Absolute bias**	**Estimates of variance**	**Estimates**	**Absolute bias**	**Estimates of variance**
$\bar{y}_{p0}$	64.587081302	0.044539285	2.221816991	112.5263056	0.00471948	0.22857952
$\bar{y}_{sp1}$	65.317497631	1.167850257	2.187718062	112.2594916	0.00663659	0.22511622
$\bar{y}_{sp2}$	67.501318796	0.079684067	6.054651976	112.5205667	1.27646374	2.14946154
$\bar{y}_{sp3}$	63.761410550	0.132619702	2.163212450	112.3666869	0.01885341	0.22673489

6. Conclusions

Based on the research findings from both the simulation experiments and the empirical application, it can be concluded that employing a calibration approach to construct product-type estimators for the population mean within the context of stratified random sampling can enhance both the accuracy and efficiency of estimation. This improvement depends on the distance function utilized, the calibration constraints imposed on the weights, and the specific characteristics of the statistical population, as evidenced by the simulation results across varying population sizes, numbers of strata, and variable properties. The estimator $\bar{y}_{sp1}$, which relies on a Hellinger-type distance function, yielded the lowest MSE values and the highest relative efficiency in the simulation experiments; this indicates its effectiveness in improving the accuracy of population mean estimation compared to the traditional product–type estimator, $\bar{y}_{p0}$, under the studied conditions. Conversely, the estimator $\bar{y}_{sp2}$ based on the relative quadratic distance function, exhibited variable performance depending on the population characteristics, while $\bar{y}_{sp3}$, based on the Kullback–Leibler distance function, failed to demonstrate any improvement in relative efficiency across the simulation scenarios. The empirical results further indicate that the ranking of the estimators varies according to the dataset and the comparison criterion used. In the first dataset, $\bar{y}_{sp3}$ was closest to the true population mean and recorded the lowest estimated variance, while $\bar{y}_{sp1}$ also performed well in terms of variance. In the second dataset, $\bar{y}_{sp1}$ recorded the lowest estimated variance, followed by $\bar{y}_{sp3}$ and then the traditional estimator $\bar{y}_{p0}$, whereas $\bar{y}_{sp2}$ exhibited the highest variance. The discrepancy between the simulation study results and the findings from the initial application to real-world data can be explained by differences in the evaluation context and the criteria used to measure the performance of the estimators. In the simulation study, estimator performance was assessed using the empirical mean squared error (MSE) calculated as the mean of the squared differences between the true population mean and the estimated values across simulation iterations thereby reflecting the average performance of the estimators under the specific simulation scenarios and data-generation mechanisms employed. Furthermore, the proposed estimators rely on different distance functions and calibration constraints, which lead to variations in the resulting calibrated weights and, consequently, differences in estimation performance. Specifically, $\bar{y}_{sp1}$ utilizes a Hellinger-type distance function with a single calibration constraint, whereas $\bar{y}_{sp3}$ employs a Kullback–Leibler distance function with two constraints. In contrast, the application to real-world data involves a finite population with its own unique distributional characteristics. Although one of the simulation scenarios incorporated the stratum-specific standard deviations and correlation coefficients derived from the first real-world dataset, this alignment does not guarantee that the joint distributions of the two variables are identical, nor does it ensure that all characteristics of the actual population are captured. This issue is particularly significant given that the variables X and Y in the $Data1$ do not follow a normal distribution; factors such as skewness, the presence of outliers, or substantial variance differences across strata can affect performance in finite-sample settings.

The simulation results show that absolute empirical bias varied across estimators, population sizes, and scenarios. The reduction in bias was not uniform across all estimators or in all scenarios as the population size increased; the empirical analysis of the two datasets confirmed differences in the ranking of estimators between the datasets, demonstrating that the superiority of an estimator may shift depending on the characteristics of the data employed. Furthermore, the discrepancy between the simulation and the application can be attributed to the calculation methods used, as two distinct approaches were adopted based on the nature of each study: the a empirical application used the derived theoretical bias formulas derived for the three proposed estimators, whereas the simulation study calculated

bias based on the difference between the true population mean and the estimated value across simulation iterations. Consequently, the simulation results reflect the empirical performance of the estimators, while the results from the applied component represent bias calculated according to the derived theoretical formulas. Taken together, these results indicate that increasing the number of calibration constraints or employing a more complex distance function does not necessarily improve estimation efficiency. Rather, any improvement depends on how well the chosen distance function and constraints align with the characteristics of the population and its strata. Consequently, $\overline{y}_{sp1}$ emerges as the most consistent option based on the simulation results within the scope of this study and the $Data2$ dataset though the specific nature of the applied data and the chosen selection criteria must always be considered when determining the most appropriate estimator.

7. Recommendations

1. Conduct further research to analyze how increasing the number of calibration constraints and altering their nature affects the accuracy and efficiency of estimation.

2. Apply the first proposed estimator in other empirical setting, particularly those involving negatively correlated variables, given its strong performance regarding mean squared error, relative efficiency, and variance across most simulation scenarios and in practical applications involving normally distributed data.

3. Further develop the proposed estimators and test them within alternative sampling designs, while exploring how calibration constraints and distance functions might be adapted to suit those designs.

4. Evaluate the performance of the proposed estimators alongside other estimators such as regression and ratio estimators under various correlation patterns when estimating the population mean within the context of stratified random sampling. Selecting an appropriate distance function based on population characteristics and the nature of the relationship between the dependent and auxiliary variables, rather than assuming the superiority of a single function in all cases.

5. Investigate the theoretical properties of the proposed estimators in greater depth, including asymptotic unbiasedness, consistency, and asymptotic distribution.

REFERENCES

1. M. H. Mahdi and S. M. Hussein, "Estimating the population mean in stratified random sampling using combined regression with the presence of outliers," *Journal of Economics and Administrative Sciences*, vol. 29, no. 136, pp. 70–80, 2023.
2. J.-C. Deville and C.-E. Särndal, "Calibration estimators in survey sampling," *Journal of the American statistical Association*, vol. 87, no. 418, pp. 376–382, 1992.
3. S. Singh, "Estimation of variance of the general regression estimator: Higher level calibration approach," *Survey methodology*, vol. 24, no. 1, pp. 41–50, 1998.
4. D. K. Rao, M. G. M. Khan, and S. Khan, "Mathematical programming on multivariate calibration estimation in stratified sampling," *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, vol. 6, no. 12, pp. 58–62, 2012.
5. G. K. Singh, D. K. Rao, and M. G. M. Khan, "Calibration estimator of population mean in stratified random sampling," in *Proceedings of the 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, 2015.
6. D. K. Rao, M. G. M. Khan, and K. G. Reddy, "Stratified calibration estimator of population mean using multi-auxiliary information," in *2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, 2015.
7. N. Koyuncua and M. Kadilar, "Calibration weighting in stratified random sampling," *Communications in Statistics - Simulation and Computation*, vol. 45, no. 7, pp. 2267–2275, 2016.
8. S. B. Nidhi, "Calibration approach estimation of the mean in stratified sampling and stratified double sampling," *Communications in Statistics - Theory and Methods*, vol. 46, no. 10, pp. 4932–4942, 2017.
9. N. Özgül, "New calibration estimator in stratified sampling," *Journal of Statistical Computation and Simulation*, vol. 88, no. 13, pp. 2561–2572, 2018.
10. N. Ozgul, "New calibration estimator based on two auxiliary variables in stratified sampling," *Communications in Statistics-Theory and Methods*, vol. 48, no. 6, pp. 1481–1492, 2019.
11. P. K. Rai, A. Singh, and M. Qasim, "Calibration-based estimators using different distance measures under two auxiliary variables: A comparative study," *Journal of Modern Applied Statistical Methods*, vol. 19, no. 1, 2020.

12. M. J. Iseh and K. Bassey, "Improving the efficiency of exponential ratio-type estimator for population median: A calibration weight adjustment approach," *Journal of Applied Statistics*, vol. 14, no. 1, pp. 1–23, 2023.
13. U. Shahzad, I. Ahmad, F. Alshahrani, I. Almanjahie, and S. Iftikhar, "Calibration-based mean estimators under stratified median ranked set sampling," *Mathematics*, vol. 11, no. 8, 2023.
14. S. Gupta and P. Trisandhya, "Calibration estimators with different types of distance measures under stratified sampling in the presence of measurement error," *Axioms*, vol. 13, p. 737, 2024.
15. R. Jabeen, A. Zaka, M. Nagy, H. Al-Mofleh, and A. Z. Afify, "A new modified estimator of population variance in calibrated survey sampling," *Scientific Reports*, vol. 14, no. 1, 2024.
16. M. K. Pandey, G. N. Singh, T. Zaman, A. Al Mutairi, and M. S. Mustafa, "A general class of improved population variance estimators under non-sampling errors using calibrated weights in stratified sampling," *Scientific Reports*, vol. 14, no. 1, 2024.
17. A. M. Alomair, W. Zhu, U. Shahzad, and F. K. Alarfaj, "Non-parametric calibration estimation of distribution function under stratified random sampling," *AIMS Mathematics*, vol. 10, no. 2, pp. 4457–4472, 2025.
18. O. T. Babatunde and A. Oladugba, "Robust calibration estimation of population mean in stratified sampling in the presence of outlier," *Quality & Quantity: International Journal of Methodology*, vol. 59, no. 3, pp. 2925–2940, 2025.
19. P. K. Rai, S. Tiwari, and Alka, "Calibration approach-based estimator of population total under successive sampling design," *Communications in Statistics - Theory and Methods*, vol. 54, no. 11, pp. 3307–3321, 2025.
20. N. Garg, Sneha, and M. Pachori, "Calibration-enhanced estimation of population mean with multi-auxiliary information in stratified sampling," *Journal of the Indian Society of Agricultural Statistics*, vol. 80, no. 1, pp. 49–59, 2026.
21. M. A. Kazim and S. M. Hussein, "Comparison of some calibration estimators for estimating the mean wheat production in the iraqi governorates using cultivated area as an auxiliary variable," *Al-Turath University College*, vol. 43, no. 1, pp. 287–300, 2026.
22. Y. Tillé, *Sampling and Estimation from Finite Populations*. John Wiley & Sons, 2020.
23. N. Koyuncu and C. Kadilar, "Ratio and product estimators in stratified random sampling," *Journal of statistical planning and inference*, vol. 139, no. 8, pp. 2552–2558, 2009.
24. S. Alam, S. Singh, and J. Shabbir, "New methodology of calibration in stratified random sampling," in *Survey Research Methods Section*, vol. 3, 2019, pp. 90–95.
25. D. Ojua, J. Abuchu, E. Ojua, and E. Enang, "Comparing calibration product type estimators of population mean in stratified sampling under two constraints using different distance measures," *Journal of Advances in Mathematics and Computer Science*, vol. 36, no. 9, pp. 1–16, 2021.
26. D. Robson, "Applications of multivariate polykays to the theory of unbiased ratio-type estimation," *Journal of the American Statistical Association*, vol. 52, no. 280, pp. 511–522, 1957.
27. M. Murthy, "Product method of estimation," *Sankhyā: The Indian journal of statistics, series A*, pp. 69–74, 1964.
28. T. Hastie, R. Tibshirani, J. Friedman *et al.*, "The elements of statistical learning," 2009.
29. E. P. Clement and A. E. Anieting, "Improvement on calibration weightings in stratified random sampling," *International Journal of Statistical Sciences*, vol. 24, no. 2, pp. 125–136, 2024.
30. N. Koyuncu and C. Kadilar, "Calibration estimator using different distance measures in stratified random sampling," *International Journal of Modern Engineering Research (IJMER)*, vol. 3, no. 1, pp. 415–419, 2013.
31. S. Singh, *Advanced Sampling Theory With Applications: How Michael"" Selected"" Amy*. Springer Science Business Media, 2003, vol. 2.