Journal of Hunan University

(Natural Sciences)

Vol. 52 No. 9

September 2025

Available online at

https://jonuns.com

Open Access Article

https://doi.org/10.55463/issn.1674-2974.52.9.1

Multiple-Constraint-Based Calibration Approach for Cumulative Distribution Function Estimation

Putao FAN*[1]

[1]School of Economics and Statistics, Xingzhi College of Xi'an University of Finance and Economics, China

* Corresponding author: grape0913@163.com

Article History:

Received: August 18, 2025
Revised: September 25, 2025
Accepted: October 6, 2025
Published: October 30, 2025

Abstract: This paper presents a novel calibration framework for estimating cumulative distribution functions (CDFs) within the context of stratified sampling (StRS) by utilizing multiple calibration constraints and Lata's distance function. Traditional CDF estimators often neglect the incorporation of auxiliary information, resulting in inefficiencies and biased estimates, particularly for heterogeneous populations. To address this limitation, we propose a new class of CDF estimators that extends existing calibration methodologies by incorporating multiple constraint conditions, thereby enhancing precision and reducing estimation error. Methodological advancement lies in the strategic integration of additional auxiliary information through a generalized calibration mechanism tailored to the structural characteristics of stratified sampling design. This generalization represents a significant extension of the calibration theory into the domain of CDF estimation, an area that remains relatively underexplored in the survey sampling literature. The performance of the proposed estimators was assessed via a comprehensive simulation study involving two real-world datasets: COVID-19 incidence data and apple yield statistics from the agricultural sector. Across both applications, the proposed estimators consistently outperform benchmark estimators in terms of mean squared error (MSE) and precision relative efficiency (PRE) across a range of quantiles.

These findings underscore the practical utility and theoretical relevance of the proposed approach in complex survey environments marked by population heterogeneity. The new estimator family provides a robust and flexible tool for improving the accuracy of CDF estimation in applied survey statistics.

Keywords: CDF; StRS; Calibration Approach; Simulation Study; Relative Efficiency.

多约束校准累积分布函数估计

摘要：本文探究分层抽样（StRS）下基于多约束的累积分布函数（CDF）校准估计过程。传统 CDF 估计通常不引入额外信息，在非齐性总体中易致效率低、结果有偏差。文中提出特殊 CDF 估计量，结合多个校准约束与 Lata 距离函数，专为提升 StRS 中表现构建，将辅助信息纳入校准过程，使推断更为精确。且将现有校准过程推广至调查抽样文献中研究较不足的 CDF 估计领域，意义显著。

通过新冠肺炎统计、苹果产量这两项实际应用的数值模拟评估该估计量，对比调整后估计量发现，其在不同分位数的均方误差（MSE）与精度相对效率（PRE）上均更优，证实其在总体异质性显著的实际调查中具适用性与实用价值，为提升调查统计中 CDF 估计准确性提供了有效工具。

关键词：累积分布函数（CDF）；分层抽样（StRS）；校准方法；模拟研究

---

1. Introduction

When we want to determine the proportion of values in research variable Y that fall within a specific range, we encounter the issue of estimating the countable population CDF. Estimating CDF is considered important under certain circumstances, such as when a soil scientist needs to know the percentage of people experiencing poverty and resource scarcity in a developing country. In general, we are interested in the proportion of y_i values in the population, but in some cases, CDF is more important. For instance, users of sample information from surveys may need to determine the population CDF or proportion in certain units with values less than or equal to a particular threshold value, t_y . For example, we may wish to determine the proportion of agricultural fields with a pesticide toxicity level below zero or the proportion of filtration plants with arsenic levels in drinking water below a certain threshold. This proportion represents a specific cumulative distribution function value for the population.

$\mathcal{F}_Y(t_y) = \frac{1}{M}\sum_{i=1}^{M}\gamma(y_i \leq t_y)$,

where $\gamma(y_i \leq t_y) = 1$ for $y_i \leq t_y$ and $\gamma(y_i \leq t_y) = 0$ for $y_i > t_y$. Note that γ represents the indicator function. Conventional methods for the cumulative distribution function (CDF) depend entirely on the selected sampling technique and specific subset of the population included in the sample. The estimation of $\mathcal{F}_Y(t_y)$ can be accomplished using $\hat{\mathcal{F}}_y(t_y) = \frac{1}{m}\sum_{i=1}^{m}\gamma(y_i \leq t_y)$.

Many studies have calculated CDF for a population by incorporating data from single or multiple additional variables. Some important contributions to CDF estimation come from [1], including the development of a technique to estimate CDF by leveraging information from both primary and supplementary variables. Similarly, additional variables, differences, and ratio-type CDF estimators were provided in [2]. Using the model-calibration pseudo-empirical likelihood method, an estimator for CDF was defined in [3]. The CDF estimator was developed using the calibration approach described in [4]. A new class of estimators was constructed in [5] to estimate the CDF when the research variable was confronted with nonresponse. CDF estimators were developed in [6] by leveraging auxiliary information in both SRS and StRS' schemes. Moreover, [7] presented a more adaptable estimator of the cumulative distribution function (CDF) involving exponential factors that used both the mean and rank of an auxiliary variable. However, much of the current literature has not focused on the development of calibration-based CDF estimators despite these developments. In this study, we develop multiple constraint-based CDF estimators.

The remainder of the manuscript is arranged in such a way that it logically takes readers through conceptual grounds for empirical validation. Section 2 discusses how the adapted estimator and associated calibration methodology are developed. Section 3 presents the recently developed CDF estimators, which focus on the theoretical development of such estimators and their structural features. Section 4 illustrates the practical usefulness of the proposed approach by providing a numerical and comparative analysis. Finally, Section 5 concludes the study, summarizing the main findings and outlining their general implications.

2. Material Method

2.1. Calibration Technique and Adapted Estimator

The calibration method leveraged additional information to correct the initial design weights and improve the accuracy of the estimates. Weights were optimized through this technique to be better adjusted while still satisfying the constraints defined in the supplementary data. The authors of [8] are pioneers in this discipline. Calibration estimation methods for the StRS estimator were developed in [9]. A calibration estimator was introduced for ratio-type estimation techniques in the StRS [10]. Novel multivariate calibrated mean estimators were developed in the StRS design by [11, 12], which incorporated information available from more than one additional dataset. A few novel weights using different distance functions in StRS were theoretically developed by [13]. A simulation was performed to compare the performances of the weights. A novel calibration estimation method using stratified double sampling (DSRS) to estimate the population mean was developed by [14]. A group of calibration estimation methods has been suggested for calculating the population mean using a number of calibration conditions in StRS (see [15] and [16]). A distance function and calibrated mean estimator in the StRS were proposed in [17]. Novel calibration estimators for the population mean in the presence of two supplementary information in the StRS were proposed in [18]. A calibration estimator for the population mean in the StRS with a set of new calibration conditions using the known CV was suggested by [19]. Studies [20-23] improved on an earlier study in this field and presented a novel method to estimate the dispersion mean by utilizing the properties of linear moments. These estimators use the L-mean, L-standard deviation, and L-CV properties. They showed that these descriptive limitations can be utilized to provide better estimates of dispersion measures under StRS in the presence of extreme observations. Using the robust properties of linear moments, novel calibration-based CV estimators that provide appropriate CV estimates for the research variable DSRS were created in [24]. For both symmetric and asymmetric data, a calibrated empirical neutrosophic CDF estimation was proposed in [25, 26]. Taking inspiration from these studies, we expanded this work to estimate CDF under StRS.

The heterogeneous population was transformed into homogenous groups using the StRS. When the variance between strata is much larger than the variance within strata, efficiency is improved [27]. Assume a countable population $G = n_1, n_2, \dots, n_{N^*}$ of size N^* and let (Y, X) respectively, be the research and supplementary information associated with each unit n_j $(j = 1,2,\dots,N^*)$ of the population, respectively. Let the population of size, N^*, which is divided into $\mathcal{L}$ homogeneous strata where the size of ϑ^{th} stratum is N_ϑ^*, for $\vartheta = 1,2,..,\mathcal{L}$ in such that $\sum_{\vartheta=1}^{\mathcal{L}} N_\vartheta^* = N^*$. Now, a simple random sampling without replacement (SRSWOR) with size n_ϑ is drawn from the ϑ^{th} stratum such that $\sum_{\vartheta=1}^{\mathcal{L}} n_\vartheta = n$. The stratum weights is defined as $W_\vartheta^* = \frac{N_\vartheta^*}{N^*}$.

Under StRS scheme, the conventional unbiased estimator of the CDF is given as

$T_o = \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^* \hat{\mathcal{F}}_{y\vartheta}(t_y)$,

where $\hat{\mathcal{F}}_{y\vartheta}(t_y) = \frac{1}{m}\sum_{i=1}^{m} \gamma(y_i \leq t_y)$ is the sample CDF estimate of Y in the ϑ^{th} stratum.

2.2. Adapted Family of CDF Estimators

According to the methodological framework suggested in [24], the modified family of estimators is as follows:

$$T_A = \sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* \hat{\mathcal{F}}_{y\vartheta}(t_y) \tag{1}$$

where Ω_ϑ^* is the calibration weight, and the study adopts the Lata distance function [17], which is expressed as follows:

$$\sum_{\vartheta=1}^{\mathcal{L}} \frac{S_{\vartheta x}^2}{\psi_\vartheta^*} (\Omega_\vartheta^* - W_\vartheta^*)^2 \tag{2}$$

The adjusted weights Ω_ϑ^* in the StRS are determined such that they satisfy three calibration constraints:

$$\sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* \hat{\mathcal{F}}_{x\vartheta}(t_x) = \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^* \mathcal{F}_{x\vartheta}(t_x) \tag{3}$$

$$\sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* \, c_{x\vartheta} = \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^* \, C_{x\vartheta} \tag{4}$$

$$\sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* = \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^* \tag{5}$$

Note that $W_\vartheta^* = \frac{N_\vartheta^*}{N^*}$ denote the traditional stratum weight, where $\left(\hat{\mathcal{F}}_{x\vartheta}(t_x), \mathcal{F}_{x\vartheta}(t_x)\right)$ demonstrate the sample and population CDF of the X in ϑ^{th} stratum and $(c_{x\vartheta}, C_{x\vartheta})$ demonstrate the sample and population CV of the X in ϑ^{th} stratum, and ψ_ϑ^* are selected weights. Hence, the function of Lagrange with multipliers λ_1, λ_2 and λ_3, denoted by Δ, is given by

$$\Delta = \sum_{\vartheta=1}^{\mathcal{L}} \frac{S_{\vartheta x}^2}{\psi_\vartheta^*} (\Omega_\vartheta^* - W_\vartheta^*)^2 - 2\lambda_1 \left(\sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* \hat{\mathcal{F}}_{x\vartheta}(t_x) - \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^* \mathcal{F}_{x\vartheta}(t_x)\right) - 2\lambda_2 \left(\sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* \, c_{x\vartheta} - \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^* \, C_{x\vartheta}\right) - 2\lambda_3 \left(\sum_{\vartheta=1}^{\mathcal{L}} \Omega_\vartheta^* - \sum_{\vartheta=1}^{\mathcal{L}} W_\vartheta^*\right) \tag{6}$$

where λ_1, λ_2 and λ_3 are the Lagrange's multipliers, setting $\frac{\partial \Delta}{\partial \Omega_\vartheta^*}=0$, we get

$$2\frac{S_{\vartheta x}^2}{\psi_\vartheta^*}(\Omega_\vartheta^*-W_\vartheta^*)-2\lambda_1\hat{\mathcal{F}}_{x\vartheta}(t_x)-2\lambda_2 c_{x\vartheta}-2\lambda_3=0$$

Consequently, the calibration weight can be represented as

$$\Omega_\vartheta^*=W_\vartheta^*+\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\left(\lambda_1\hat{\mathcal{F}}_{x\vartheta}(t_x)+\lambda_2 c_{x\vartheta}+\lambda_3\right) \quad (7)$$

Substituting Equation (7) into Equations (3), (4), and (5), we obtain

$$\begin{bmatrix}\breve{\mathbb{W}}_{11} & \breve{\mathbb{W}}_{12} & \breve{\mathbb{W}}_{13}\\ \breve{\mathbb{W}}_{21} & \breve{\mathbb{W}}_{22} & \breve{\mathbb{W}}_{23}\\ \breve{\mathbb{W}}_{31} & \breve{\mathbb{W}}_{32} & \breve{\mathbb{W}}_{33}\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{bmatrix}=\begin{bmatrix}\breve{\mathbb{W}}_{10}\\ \breve{\mathbb{W}}_{20}\\ 0\end{bmatrix} \quad (8)$$

By solving Equation (8), we obtain

$$\lambda_i=\frac{A_i}{B}\ ;\ i=1,2,3 \quad (9)$$

With

$$A_1=\breve{\mathbb{W}}_{10}\left(\breve{\mathbb{W}}_{33}\breve{\mathbb{W}}_{22}-\breve{\mathbb{W}}_{23}^2\right)+\breve{\mathbb{W}}_{20}(\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{23}-\breve{\mathbb{W}}_{33}\breve{\mathbb{W}}_{12})$$

$$A_2=\breve{\mathbb{W}}_{20}\left(\breve{\mathbb{W}}_{33}\breve{\mathbb{W}}_{11}-\breve{\mathbb{W}}_{13}^2\right)-\breve{\mathbb{W}}_{10}(\breve{\mathbb{W}}_{12}\breve{\mathbb{W}}_{33}-\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{23})$$

$$A_3=\breve{\mathbb{W}}_{10}(\breve{\mathbb{W}}_{23}\breve{\mathbb{W}}_{12}-\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{22})+\breve{\mathbb{W}}_{20}(\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{12}-\breve{\mathbb{W}}_{11}\breve{\mathbb{W}}_{22})$$

$$B=\breve{\mathbb{W}}_{11}\breve{\mathbb{W}}_{22}\breve{\mathbb{W}}_{33}-\breve{\mathbb{W}}_{13}^2\breve{\mathbb{W}}_{22}-\breve{\mathbb{W}}_{12}^2\breve{\mathbb{W}}_{33}-\breve{\mathbb{W}}_{23}^2\breve{\mathbb{W}}_{11}+2\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{23}\breve{\mathbb{W}}_{12}$$

where

$$\breve{\mathbb{W}}_{10}=\sum_{\vartheta=1}^{\mathcal{L}}W_\vartheta^*\left(\mathcal{F}_{x\vartheta}(t_x)-\hat{\mathcal{F}}_{x\vartheta}(t_x)\right),$$

$$\breve{\mathbb{W}}_{20}=\sum_{\vartheta=1}^{\mathcal{L}}W_\vartheta^*\left(C_{x\vartheta}-c_{x\vartheta}\right)$$

$$\breve{\mathbb{W}}_{11}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}^2(t_x),$$

$$\breve{\mathbb{W}}_{12}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)c_{x\vartheta},$$

$$\breve{\mathbb{W}}_{13}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)$$

$$\breve{\mathbb{W}}_{21}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)c_{x\vartheta},$$

$$\breve{\mathbb{W}}_{22}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}c_{x\vartheta}^2,$$

$$\breve{\mathbb{W}}_{23}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}c_{x\vartheta}$$

$$\breve{\mathbb{W}}_{32}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}c_{x\vartheta},$$

$$\breve{\mathbb{W}}_{31}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x),\quad \breve{\mathbb{W}}_{33}=\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}$$

Substituting these values in Equations (7) and (1), we have

$$\begin{aligned}T_A &=T_o+\lambda_1\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)\\ &+\lambda_2\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)c_{x\vartheta}\right)\\ &+\lambda_3\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)\end{aligned} \quad (10)$$

Further, the final compressed form of the T_A estimator in Equation (10) can be expressed as

$$T_A=\sum_{\vartheta=1}^{\mathcal{L}}W_\vartheta^*\hat{\mathcal{F}}_{y\vartheta}(t_y)+\hat{\beta}_1\left(\sum_{\vartheta=1}^{\mathcal{L}}W_\vartheta^*\left(\mathcal{F}_{x\vartheta}(t_x)-\hat{\mathcal{F}}_{x\vartheta}(t_x)\right)\right)+\hat{\beta}_2\left(\sum_{\vartheta=1}^{\mathcal{L}}W_\vartheta^*\left(C_{x\vartheta}-c_{x\vartheta}\right)\right) \quad (11)$$

where new betas are given by

$$\hat{\beta}_i=\frac{A_j}{B}\ ;(i,j)=(1,4)\text{ and }(i,j)=(2,5) \quad (12)$$

with

$$\begin{aligned}A_4=&\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)\left(\breve{\mathbb{W}}_{33}\breve{\mathbb{W}}_{22}-\breve{\mathbb{W}}_{23}^2\right)\\ &-\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)c_{x\vartheta}\right)(\breve{\mathbb{W}}_{12}\breve{\mathbb{W}}_{33}-\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{23})\\ &+\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)(\breve{\mathbb{W}}_{23}\breve{\mathbb{W}}_{12}-\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{22})\end{aligned}$$

$$\begin{aligned}A_5=&\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)(\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{23}-\breve{\mathbb{W}}_{33}\breve{\mathbb{W}}_{12})\\ &+\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)c_{x\vartheta}\right)\left(\breve{\mathbb{W}}_{11}\breve{\mathbb{W}}_{33}-\breve{\mathbb{W}}_{13}^2\right)\\ &+\left(\sum_{\vartheta=1}^{\mathcal{L}}\psi_\vartheta^*(S_{\vartheta x}^2)^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)(\breve{\mathbb{W}}_{13}\breve{\mathbb{W}}_{12}-\breve{\mathbb{W}}_{11}\breve{\mathbb{W}}_{23})\end{aligned}$$

All the members of T_A are listed in Table 1. Where $\bar{x}_\vartheta, s_{x\vartheta}, s_{x\vartheta}^2, c_{x\vartheta}, \tilde{x}_{x\vartheta}, G_{x\vartheta}, mad_{x\vartheta}, iqr_{x\vartheta}$ and $hm_{x\vartheta}$ are the sample mean, standard deviation, variance, coefficient of variation, median, geometric mean, median absolute deviation, inter-quartile range and harmonic mean of the supplementary variable in the ϑ^{th} stratum, respectively. The family members are listed in Table 1.

Table 1. Adapted family of CDF estimators

A	T_A	ψ_ϑ^*
1	T_1	1
2	T_2	$1/\bar{x}_\vartheta$
3	T_3	$1/s_{x\vartheta}$
4	T_4	$1/s_{x\vartheta}^2$
5	T_5	$1/c_{x\vartheta}$
6	T_6	$1/\tilde{x}_{x\vartheta}$
7	T_7	$1/G_{x\vartheta}$
8	T_8	$1/mad_{x\vartheta}$

9	T_9	$1/iqr_{x\vartheta}$
10	T_{10}	$1/hm_{x\vartheta}$

3. Proposed Family of CDF Estimators

Using supplementary information, the calibration estimation methods increase the precision of their estimates by modifying the original design weights. Taking motivation from the adapted estimators, we propose the following class of CDF estimators:

$$P_A = \sum_{\vartheta=1}^{\mathcal{L}} \Omega_{p\vartheta}^{*} \hat{\mathcal{F}}_{y\vartheta}(t_y) \tag{13}$$

According to the methodological framework suggested in [24], the modified family of estimators is as follows:

$$\sum_{\vartheta=1}^{\mathcal{L}} \frac{S_{\vartheta x}^{2}}{\psi_{\vartheta}^{*}} \left(\Omega_{p\vartheta}^{*} - W_{\vartheta}^{*}\right)^{2} \tag{14}$$

and satisfy the calibration constraint:

$$\sum_{\vartheta=1}^{\mathcal{L}} \Omega_{p\vartheta}^{*} = \sum_{\vartheta=1}^{\mathcal{L}} W_{\vartheta}^{*} \tag{15}$$

$$\sum_{\vartheta=1}^{\mathcal{L}} \Omega_{p\vartheta}^{*} \hat{\mathcal{F}}_{x\vartheta}(t_x) = \sum_{\vartheta=1}^{\mathcal{L}} W_{\vartheta}^{*} \mathcal{F}_{x\vartheta}(t_x) \tag{16}$$

where ψ_{ϑ}^{*} are wisely selected weights as listed in Table 2. The Lagrange function is given by:

$$\Delta = \sum_{\vartheta=1}^{\mathcal{L}} \frac{S_{\vartheta x}^{2}}{\psi_{\vartheta}^{*}} \left(\Omega_{p\vartheta}^{*} - W_{\vartheta}^{*}\right)^{2} - 2\lambda_{p_1}\left(\sum_{\vartheta=1}^{\mathcal{L}} \Omega_{p\vartheta}^{*} - \sum_{\hbar=1}^{\mathcal{L}} W_{\vartheta}^{*}\right) - 2\lambda_{p_2}\left(\sum_{\vartheta=1}^{\mathcal{L}} \Omega_{p\vartheta}^{*} \hat{\mathcal{F}}_{x\vartheta}(t_x) - \sum_{\vartheta=1}^{\mathcal{L}} W_{\vartheta}^{*} \mathcal{F}_{x\vartheta}(t_x)\right) \tag{17}$$

where λ_{p_1} and λ_{p_2} are the Lagrange's multipliers, setting $\frac{\partial \Delta}{\partial \Omega_{p\vartheta}^{*}} = 0$, we get

$$2\frac{S_{\vartheta x}^{2}}{\psi_{\vartheta}^{*}}\left(\Omega_{p\vartheta}^{*} - W_{\vartheta}^{*}\right) - 2\lambda_{p_1} - 2\lambda_{p_2}\hat{\mathcal{F}}_{x\vartheta}(t_x) = 0$$

Thus

$$\Omega_{p\vartheta}^{*} = W_{\vartheta}^{*} + \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\left(\lambda_{p_1} + \lambda_{p_2}\hat{\mathcal{F}}_{x\vartheta}(t_x)\right) \tag{18}$$

Substituting Equation (18) in Equation (15)

$$\lambda_{p_1}\sum_{\vartheta=1}^{\mathcal{L}} \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1} + \lambda_{p_2}\sum_{\vartheta=1}^{\mathcal{L}} \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x) = 0 \tag{19}$$

Substituting Equation (18) in Equation (16)

$$\lambda_{p_1}\sum_{\vartheta=1}^{\mathcal{L}} \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x) + \lambda_{p_2}\sum_{\vartheta=1}^{\mathcal{L}} \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\hat{\mathcal{F}}_{x\vartheta}^{2}(t_x) = \sum_{\vartheta=1}^{\mathcal{L}} W_{\vartheta}^{*}\left(\mathcal{F}_{x\vartheta}(t_x) - \hat{\mathcal{F}}_{x\vartheta}(t_x)\right) \tag{20}$$

Solve Equations (20) and (19), we get lambda values

$$\lambda_{p_1} = \frac{-(\mathbb{W}_{10})(\mathbb{W}_{13})}{(\mathbb{W}_{33})(\mathbb{W}_{11}) - (\mathbb{W}_{13})^{2}} \tag{21}$$

and

$$\lambda_{p_2} = \frac{(\mathbb{W}_{10})(\mathbb{W}_{33})}{(\mathbb{W}_{33})(\mathbb{W}_{11}) - (\mathbb{W}_{13})^{2}} \tag{22}$$

Substituting Equations (22) and (21) into Equation (18), we obtain the weights asfollows:

$$\Omega_{p\vartheta}^{*} = W_{\vartheta}^{*} + \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\left[\frac{-(\mathbb{W}_{10})(\mathbb{W}_{13})}{(\mathbb{W}_{33})(\mathbb{W}_{11}) - (\mathbb{W}_{13})^{2}}\right] + \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\hat{F}_{x\vartheta}(t_x)\left[\frac{(\mathbb{W}_{10})(\mathbb{W}_{33})}{(\mathbb{W}_{33})(\mathbb{W}_{11}) - (\mathbb{W}_{13})^{2}}\right] \tag{23}$$

Substituting Equation (23) in Equation (13), we get

$$P_A = \sum_{\vartheta=1}^{L} W_{\vartheta}^{*}\hat{\mathcal{F}}_{y\vartheta}(t_y) + \hat{\beta}\sum_{\vartheta=1}^{L} W_{\vartheta}^{*}\left(\mathcal{F}_{x\vartheta}(t_x) - \hat{\mathcal{F}}_{x\vartheta}(t_x)\right) \tag{24}$$

where

$$\hat{\beta} = \frac{\left\{(\mathbb{W}_{33})\left(\sum_{\vartheta=1}^{\mathcal{L}} \psi_{\vartheta}^{*}(S_{\hbar x}^{2})^{-1}\hat{\mathcal{F}}_{x\vartheta}(t_x)\hat{\mathcal{F}}_{y\vartheta}(t_y)\right) - (\mathbb{W}_{13})\left(\sum_{\vartheta=1}^{\mathcal{L}} \psi_{\vartheta}^{*}(S_{\vartheta x}^{2})^{-1}\hat{\mathcal{F}}_{y\vartheta}(t_y)\right)\right\}}{(\mathbb{W}_{33})(\mathbb{W}_{11}) - (\mathbb{W}_{13})^{2}}$$

Table 2. Proposed family of CDF estimators

A	P_A	ψ_{ϑ}^{*}
1	P_1	1
2	P_2	$1/\bar{x}_{\vartheta}$
3	P_3	$1/s_{x\vartheta}$
4	P_4	$1/s_{x\vartheta}^{2}$
5	P_5	$1/c_{x\vartheta}$
6	P_6	$1/\tilde{x}_{x\vartheta}$
7	P_7	$1/G_{x\vartheta}$
8	P_8	$1/mad_{x\vartheta}$
9	P_9	$1/iqr_{x\vartheta}$
10	P_{10}	$1/hm_{x\vartheta}$

4. Numerical Study

For assess the performance of the new CDF proposals with adapted ones, we calculated the MSE and PRE using the following simulation phases:

Phase-1: From stratum ϑ, sample of size n_{ϑ} is selected based on SRSWOR.

Phase-2: Calculate CDF estimates (say) $\hat{\tau} = T_A$, P_A.

Phase-3: Repeat the procedure $R = 5000$ times and, attained $\hat{\tau}_1,\ \hat{\tau}_2, \ldots,\ \hat{\tau}_R$.

Phase-4: Compute the MSE as, $MSE(\hat{\tau}) = \frac{1}{R}\sum_{i=1}^{R}\left(\hat{\tau} - F_y(t_y)\right)^2$.

Phase-5: Compute the PRE as, $PRE(\hat{\tau}) = \frac{MSE(T_A)}{MSE(P_A)} \times 100$.

MSEs and PREs are listed in Tables 3–10. Interestingly, we used six quantile points to evaluate the performance of all four populations.
i.e.,$(t = 0.25, 0.30, 0.35, 0.40, 0.45, 0.50)$ in upcoming sub-sections.

4.1 Apple data (population 1 and 2)

In this subsection, we analyze a collection of apple-fruit data to illustrate the effectiveness of the proposed estimators. See [25] for further details.

- **Population 1:** We considered the quantity of apples manufactured in 1999 as y, and the number of apple plants in 1999 as x.
- **Population 2:** We consider the quantity of apples manufactured in 1998 as x, and the quantity of apples manufactured in 1999 as y.

4.2 COVID-19 (population 3 and 4)

In this subsection, we analyze a collection of COVID data to illustrate the effectiveness of the proposed estimators. See [25] for further details.

- **Population 3:** We took the overall number of cases per million x, and the overall number of deaths per million y.
- **Population 4:** We take the overall number of cases per million as x, and the overall number of recoveries per million as y.

4.3. Results

Results of Tables 3,4,5, and 6 indicate that:

- For the first population, the adapted estimator T_A for $A = 10$ has the minimum value of MSE for $T_{10} = 9.030717 \times 10^{-4}$ at the quantile $(t = 0.40)$. The proposed class P_A for $A = 9$ has the minimum MSE value for $P_9 = 2.421297 \times 10^{-5}$ *at* quantile $(t = 0.35)$.
- For the second population, the adapted estimator T_A for $A = 10$ has a min MSE for $T_{10} = 1.818683 \times 10^{-3}$ at quantile $(\mathrm{t} = 0.45)$. The proposed class P_A for $A = 10$ has the minimum MSE value for $P_{10} = 1.315255 \times 10^{-5}$ at quantile $(t = 0.45)$.
- For the third population, the adapted estimator T_A for $A = 5$ has a min MSE for $T_5 = 1.583063 \times 10^{-3}$ at quantile $(t = 0.50)$. The proposed class P_A for $A = 2$ has the minimum MSE value for $P_2 = 5.748056 \times 10^{-4}$ at quantile $(t = 0.50)$.
- For the last population, the adapted estimator T_A for $\mathrm{A} = 10$ has a min MSE for $T_{10} = 1.422993 \times 10^{-4}$ at quantile $(\mathrm{t} = 0.25)$. The proposed class P_A for $A = 5$ has the minimum MSE value for $P_5 = 8.35568 \times 10^{-7}$ at quantile $(t = 0.35)$.

These results clearly indicate the superior performance of the proposed estimators across all datasets and quantiles.

4.4. Discussion

The simulation results prove that all estimators of the proposed type are better than the adapted estimators with respect to smaller MSE and greater validity in all the considered populations and quantiles. This implies that the multiple calibration constraints and the distance function of Lata significantly enhance the accuracy of the estimation within the framework of stratified sampling.

This study proposes a more flexible and robust estimation framework compared to previous works [25, 26], which used either individual auxiliary variables or less flexible calibration schemes. The inclusion of different auxiliary statistics as constraints (CV, MAD, IQR, etc.) improves the flexibility of the estimators, particularly when the populations to be studied are heterogeneous.

The generalizability of this study, as evidenced by a variety of datasets, and the theoretical concepts based on calibration estimation are the strengths of this research. However, there are limitations to the study: it is conducted in simulated settings and on known datasets. Future research should confirm the estimators in more dynamic or real-time survey situations.

The proposed estimators provide a strong solution to the original research problem, that is, how to enhance the CDF estimation accuracy when using stratified designs and multiple auxiliary sources. Practically, the results can be of interest to survey statisticians who have to work with noisy or heterogeneous data, for example, when monitoring the environment or epidemiological research.

Table 3. Estimators Evaluation using MSE (Population-1)

Estimator	0.25	0.30	0.35
T_o	2.594613×10^{-4}	4.301423×10^{-5}	1.956457×10^{-5}
T_1	1.922548×10^{-3}	5.20586×10^{-3}	3.619154032×10^{9}
T_2	1.976991×10^{-3}	5.211069×10^{-3}	3.475671145×10^{9}
T_3	2.03579×10^{-3}	5.29622×10^{-3}	5.104594106×10^{9}
T_4	2.352276×10^{-3}	5.563346×10^{-3}	$1.03700867504\times 10^{11}$
T_5	1.955756×10^{-3}	5.275128×10^{-3}	3.114741285×10^{9}
T_6	1.963876×10^{-3}	5.150312×10^{-3}	1.918244699×10^{9}

T_7	1.940073×10^{-3}	5.239954×10^{-3}	2.076158175×10^{9}
T_8	1.972221×10^{-3}	5.059714×10^{-3}	3.557156414×10^{9}
T_9	1.994037×10^{-3}	5.229386×10^{-3}	4.549273069×10^{9}
T_{10}	1.977707×10^{-3}	5.469972×10^{-3}	1.837003656×10^{9}
P_1	1.418536×10^{-4}	4.271179×10^{-5}	2.429075×10^{-5}
P_2	1.568435×10^{-4}	4.495924×10^{-5}	2.425946×10^{-5}
P_3	1.715253×10^{-4}	4.849984×10^{-5}	2.431017×10^{-5}
P_4	2.191739×10^{-4}	6.889722×10^{-5}	2.464085×10^{-5}
P_5	1.502574×10^{-4}	4.358367×10^{-5}	2.433336×10^{-5}
P_6	1.487301×10^{-4}	4.427208×10^{-5}	2.422389×10^{-5}
P_7	1.421342×10^{-4}	4.364024×10^{-5}	2.421579×10^{-5}
P_8	1.569342×10^{-4}	4.465132×10^{-5}	2.425851×10^{-5}
P_9	1.64924×10^{-4}	4.603624×10^{-5}	2.421297×10^{-5}
P_{10}	1.309247×10^{-4}	4.452302×10^{-5}	2.421688×10^{-5}
Estimator	0.40	0.45	0.50
T_o	8.896124×10^{-5}	1.30131×10^{-5}	1.641263×10^{-4}
T_1	1.252875×10^{3}	2.02195×10^{-3}	2.369157×10^{3}
T_2	9.777718×10^{-4}	2.275561×10^{-3}	2.370159×10^{-3}
T_3	1.06207×10^{-3}	2.568345×10^{-3}	2.481978×10^{-3}
T_4	1.292101×10^{-3}	3.405087×10^{-3}	2.767964×10^{-3}
T_5	1.032064×10^{-3}	2.216873×10^{-3}	2.505688×10^{-3}
T_6	9.053303×10^{-4}	1.933698×10^{-3}	2.276341×10^{-3}
T_7	9.167604×10^{-4}	1.955577×10^{-3}	2.211969×10^{-3}
T_8	9.391274×10^{-4}	2.047945×10^{-3}	2.450866×10^{-3}
T_9	9.790786×10^{-4}	2.23709×10^{-3}	2.491127×10^{-3}
T_{10}	9.030717×10^{-4}	1.765097×10^{-3}	1.994424×10^{-3}
P_1	1.654678×10^{-1}	3.823043×10^{-5}	7570531×10^{-1}
P_2	4.812882×10^{-5}	3.844595×10^{-5}	9.186628×10^{-5}
P_3	5.269049×10^{-5}	3.860046×10^{-5}	9.346188×10^{-5}
P_4	5.999749×10^{-5}	3.916897×10^{-5}	9.708767×10^{-5}
P_5	4.949526×10^{-5}	3.831195×10^{-5}	9.183469×10^{-5}
P_6	4.298961×10^{-5}	3.85833×10^{-5}	8.951493×10^{-5}
P_7	4.171629×10^{-5}	3.837737×10^{-5}	8.971941×10^{-5}
P_8	4.667148×10^{-5}	3.854635×10^{-5}	9.095483×10^{-5}
P_9	4.874069×10^{-5}	3.854259×10^{-5}	9.226298×10^{-5}
P_{10}	3.665484×10^{-5}	3.833467×10^{-5}	8.726829×10^{-5}

Table 4. Estimators Evaluation using MSE (Population-2)

Estimator	0.25	0.30	0.35
T_o	2.594613×10^{-4}	4.301423×10^{-5}	1.956457×10^{-5}
T_1	3.539607×10^{-3}	4.113929×10^{-3}	1.25592537×10^{8}
T_2	3.593323×10^{-3}	4.477022×10^{-3}	1.672178570×10^{9}
T_3	3.718912×10^{-3}	4.611594×10^{-3}	$1.1559569398\times 10^{10}$
T_4	4.872951×10^{-3}	5.956713×10^{-3}	1.208722×10^{2}
T_5	3.572055×10^{-3}	4.165031×10^{-3}	5.50791703×10^{8}
T_6	3.713379×10^{-3}	4.365173×10^{-3}	2.24901055×10^{8}
T_7	3.537516×10^{-3}	4.38255×10^{-3}	2.33119062×10^{8}
T_8	3.873746×10^{-3}	4.34982×10^{-3}	2.96939418×10^{8}
T_9	3.472731×10^{-3}	4.922749×10^{-3}	4.27728358×10^{8}
T_{10}	2.318759×10^{-3}	5.575869×10^{-3}	3.13769139×10^{8}
P_1	1.278528×10^{-4}	4.568748×10^{-5}	1.700507×10^{-5}
P_2	1.497745×10^{-4}	4.714533×10^{-5}	1.722204×10^{-5}
P_3	1.995869×10^{-4}	5.4104×10^{-5}	1.739544×10^{-5}
P_4	8.320634×10^{-4}	1.026146×10^{-4}	2.068805×10^{-5}
P_5	1.382963×10^{-4}	4.523643×10^{-5}	1.702661×10^{-5}
P_6	1.272662×10^{-4}	4.66333×10^{-5}	1.714229×10^{-5}
P_7	1.267306×10^{-4}	4.691398×10^{-5}	1.715556×10^{-5}
P_8	1.311213×10^{-4}	4.59703×10^{-5}	1.718189×10^{-5}
P_9	1.403327×10^{-4}	4.605732×10^{-5}	1.734911×10^{-5}

P_{10}	1.730649×10^{-4}	4.72352×10^{-5}	1.732055×10^{-5}
Estimator	0.40	0.45	0.50
T_o	8.896124×10^{-5}	1.30131×10^{-5}	1.641263×10^{-4}
T_1	1.781255×10^{2}	2.331357×10^{-3}	2.259865×10^{4}
T_2	3.477215×10^{-3}	2.362158×10^{-3}	3.274457×10^{-3}
T_3	3.659816×10^{-3}	2.34868×10^{-3}	3.302588×10^{-3}
T_4	4.613548×10^{-3}	3.233755×10^{-3}	3.915853×10^{-3}
T_5	3.375139×10^{-3}	2.290925×10^{-3}	3.230237×10^{-3}
T_6	3.362902×10^{-3}	2.463512×10^{-3}	3484018×10^{-3}
T_7	3.306161×10^{-3}	2.286904×10^{-3}	3.346969×10^{-3}
T_8	3.407272×10^{-3}	2.63735×10^{-3}	3.52468×10^{-3}
T_9	3.515165×10^{-3}	3.435726×10^{-3}	3.913814×10^{-3}
T_{10}	2.953103×10^{-3}	1.818683×10^{-3}	2.826864×10^{-3}
P_1	4.336456×10^{-1}	1.376278×10^{-5}	8.736454×10^{0}
P_2	4.513953×10^{-5}	1.459016×10^{-5}	1.004088×10^{-4}
P_3	4.911268×10^{-5}	1.546533×10^{-5}	1.043795×10^{-4}
P_4	5.932739×10^{-5}	1.797074×10^{-5}	1.180862×10^{-4}
P_5	4.3951×10^{-5}	1.426442×10^{-5}	9.79122×10^{-5}
P_6	4.434061×10^{-5}	1.374879×10^{-5}	9.40055×10^{-5}
P_7	4.45527×10^{-5}	1.364973×10^{-5}	9.462139×10^{-5}
P_8	4.377123×10^{-5}	1.399799×10^{-5}	9.580312×10^{-5}
P_9	4.406783×10^{-5}	1.442699×10^{-5}	9.921578×10^{-5}
P_{10}	4.887962×10^{-5}	1.315255×10^{-5}	9.555782×10^{-5}

Table 5. Estimators Evaluation using MSE (Population-3)

Estimator	0.25	0.30	0.35
T_o	3.380796×10^{-3}	2.188525×10^{-3}	7.189942×10^{-4}
T_1	1.005615×10^{-2}	6.535063×10^{-3}	2.1667959×10^{7}
T_2	1.219118×10^{-2}	5.705098×10^{-3}	1.9604134×10^{7}
T_3	1.228735×10^{-2}	5.746267×10^{-3}	1.6792399×10^{7}
T_4	1.460684×10^{-2}	6.187202×10^{-3}	4.33786629×10^{8}
T_5	1.017091×10^{-2}	6.539128×10^{-3}	1.4853803×10^{7}
T_6	1.168259×10^{-2}	5.780392×10^{-3}	2.5998181×10^{7}
T_7	1.234218×10^{-2}	5.687896×10^{-3}	19584888×10^{7}
T_8	1.091879×10^{-2}	5.86535×10^{-3}	21448644×10^{7}
T_9	1.216478×10^{-2}	5.636077×10^{-3}	25442790×10^{7}
T_{10}	1.303682×10^{-2}	5.732834×10^{-3}	47706659×10^{7}
P_1	2.154961×10^{-3}	1.302922×10^{-3}	8.975721×10^{-4}
P_2	2.295363×10^{-3}	1.277611×10^{-3}	8.998204×10^{-4}
P_3	2.295661×10^{-3}	1.293377×10^{-3}	9.054203×10^{-4}
P_4	2.671897×10^{-3}	1.34432×10^{-3}	9.318249×10^{-4}
P_5	2.132335×10^{-3}	1.310065×10^{-3}	8.985292×10^{-4}
P_6	2.420951×10^{-3}	1.2652×10^{-3}	9.034731×10^{-4}
P_7	2.344513×10^{-3}	1.2715×10^{-3}	8.972617×10^{-4}
P_8	2.402808×10^{-3}	1.267768×10^{-3}	9.061196×10^{-4}
P_9	2.346333×10^{-3}	1.27568×10^{-3}	9.053963×10^{-4}
P_{10}	2.353763×10^{-3}	1.283176×10^{-3}	8.962257×10^{-4}
Estimator	0.40	0.45	0.50
T_o	1.465426×10^{-3}	4.045465×10^{-4}	9.081624×10^{-4}
T_1	4.188404×10^{3}	5.562473×10^{-3}	1.533144×10^{5}
T_2	9.48156×10^{-3}	5.484497×10^{-3}	1.701507×10^{-3}
T_3	9.669151×10^{-3}	5.672633×10^{-3}	1.743335×10^{-3}
T_4	1.090167×10^{-2}	1.154211×10^{-2}	2.114105×10^{-3}
T_5	9.74932×10^{-3}	5.528837×10^{-3}	1.583063×10^{-3}
T_6	9.369587×10^{-3}	5.494831×10^{-3}	1.739088×10^{-3}
T_7	9.468511×10^{-3}	5.514235×10^{-3}	1.695176×10^{-3}
T_8	9.212778×10^{-3}	5.776133×10^{-3}	1.737307×10^{-3}
T_9	9.919978×10^{-3}	6.245507×10^{-3}	1.799077×10^{-3}
T_{10}	9.661442×10^{-3}	5.593035×10^{-3}	1.67166×10^{-3}

P_1	2.97533×10^{-3}	1.825799×10^{-3}	2.268892×10^{-1}
P_2	3.620941×10^{-3}	2.019822×10^{-3}	5.748056×10^{-4}
P_3	3.67324×10^{-3}	2.054964×10^{-3}	5.767978×10^{-4}
P_4	4.378646×10^{-3}	2.423028×10^{-3}	5.877469×10^{-4}
P_5	3.328308×10^{-3}	1.818604×10^{-3}	5.760489×10^{-4}
P_6	3.732958×10^{-3}	2.064624×10^{-3}	5.761676×10^{-4}
P_7	3.633589×10^{-3}	2.014014×10^{-3}	5.749843×10^{-4}
P_8	3.78×10^{-3}	2.091094×10^{-3}	$5..769543\times 10^{-4}$
P_9	3.740482×10^{-3}	2.059011×10^{-3}	5.757456×10^{-4}
P_{10}	3.637431×10^{-3}	2.000691×10^{-3}	5.783546×10^{-4}

Table 6. Estimators Evaluation using MSE (Population-4)

Estimator	0.25	0.30	0.35
T_o	5.746636×10^{-4}	4.454928×10^{-4}	1.962437×10^{-5}
T_1	1.482002×10^{-4}	6.143895×10^{-3}	5.62798661×10^{8}
T_2	1.593488×10^{-4}	5.752197×10^{-3}	$7.5995107227\times 10^{10}$
T_3	1.677971×10^{-4}	6.133123×10^{-3}	$2.41799377753\times 10^{11}$
T_4	1.927349×10^{-4}	6.294663×10^{-3}	1.405908×10^{14}
T_5	1.588787×10^{-4}	6.473225×10^{-3}	9.26864558×10^{8}
T_6	1.453586×10^{-4}	5.27655×10^{-3}	4.084557991×10^{9}
T_7	1.489418×10^{-4}	5.234615×10^{-3}	8.117432529×10^{9}
T_8	1.535836×10^{-4}	5.829154×10^{-3}	3.179022966×10^{9}
T_9	1.659095×10^{-4}	5.16841×10^{-3}	$2.1649877151\times 10^{10}$
T_{10}	1.422993×10^{-4}	4.763554×10^{-3}	2.943291×10^{13}
P_1	1.321193×10^{-6}	4.675102×10^{-6}	8.376335×10^{-7}
P_2	1.770935×10^{-6}	4.967124×10^{-6}	1.640673×10^{-6}
P_3	2.109698×10^{-6}	4.604603×10^{-6}	1.85656×10^{-6}
P_4	2.902736×10^{-6}	4.942793×10^{-6}	3.173627×10^{-6}
P_5	1.450006×10^{-6}	4.377889×10^{-6}	8.35568×10^{-7}
P_6	1.207517×10^{-6}	5.409869×10^{-6}	1.486868×10^{-6}
P_7	1.262451×10^{-6}	5.51836×10^{-6}	1.571471×10^{-6}
P_8	1.312633×10^{-6}	5.143096×10^{-6}	1.418977×10^{-6}
P_9	1.349516×10^{-6}	4.764726×10^{-6}	1.307684×10^{-6}
P_{10}	1.323993×10^{-6}	5.639438×10^{-6}	1.71975×10^{-6}
Estimator	0.40	0.45	0.50
T_o	9.654232×10^{-4}	4.302393×10^{-4}	2.178278×10^{-4}
T_1	4.435296×10^{3}	7.141291×10^{-4}	1.930336×10^{4}
T_2	8.743236×10^{-4}	8.321064×10^{-4}	2.35021×10^{-3}
T_3	8.979253×10^{-4}	8.550942×10^{-4}	2.453582×10^{-3}
T_4	1.387437×10^{-3}	1.315344×10^{-3}	3.095104×10^{-3}
T_5	7.338006×10^{-4}	7.09225×10^{-4}	2.969962×10^{-3}
T_6	7.788378×10^{-4}	7.63556×10^{-4}	2.391973×10^{-3}
T_7	8.578879×10^{-4}	8.294365×10^{-4}	2.381159×10^{-3}
T_8	7.474386×10^{-4}	7.271146×10^{-4}	2.634411×10^{-3}
T_9	8.927886×10^{-4}	8.29276×10^{-4}	2.335489×10^{-3}
T_{10}	8.677821×10^{-4}	8.458473×10^{-4}	2.400434×10^{-3}
P_1	5.984922×10^{-3}	1.810083×10^{-5}	1.245991×10^{-3}
P_2	1.772586×10^{-5}	1.823174×10^{-5}	2.83337×10^{-5}
P_3	1.933322×10^{-5}	1.815514×10^{-5}	2.893277×10^{-5}
P_4	2.390742×10^{-5}	1.814561×10^{-5}	3.311264×10^{-5}
P_5	1.769403×10^{-5}	1.80517×10^{-5}	2.619863×10^{-5}
P_6	1.631438×10^{-5}	1.821074×10^{-5}	2.560608×10^{-5}
P_7	1.633232×10^{-5}	1.828391×10^{-5}	2.664522×10^{-5}
P_8	1.637488×10^{-5}	1.811504×10^{-5}	2.480254×10^{-5}
P_9	1.75483×10^{-5}	1.814774×10^{-5}	2.667181×10^{-5}
P_{10}	1.712843×10^{-5}	1.832674×10^{-5}	2.789366×10^{-5}

Table 7. Estimators Evaluation using PRE (Population-1)

(P_A, T_A)	0.25	0.30	0.35
(P_1, T_1)	1.355305×10^3	1.218834×10^4	1.489931×10^{16}
(P_2, T_2)	1.260486×10^3	1.159065×10^4	1.432708×10^{16}
(P_3, T_3)	1.186874×10^3	1.092008×10^4	2.099777×10^{16}
(P_4, T_4)	1.073247×10^3	8.074848×10^3	4.208493×10^{17}
(P_5, T_5)	1.301604×10^3	1.210345×10^4	1.280029×10^{16}
(P_6, T_6)	1.32043×10^3	1.163332×10^4	7.918813×10^{15}
(P_7, T_7)	1.364959×10^3	1.200716×10^4	8.573573×10^{15}
(P_8, T_8)	1.256718×10^3	1.133161×10^4	1.466354×10^{16}
(P_9, T_9)	1.209064×10^3	1.135928×10^4	1.878858×10^{16}
(P_{10}, T_{10})	1.510569×10^3	1.228572×10^4	7.585633×10^{15}
(P_A, T_A)	0.40	0.45	0.50
(P_1, T_1)	7.571717×10^5	5.288848×10^3	3.129446×10^5
(P_2, T_2)	2.031572×10^3	5.918857×10^3	2.58001×10^3
(P_3, T_3)	2.015677×10^3	6.653664×10^3	2.655605×10^3
(P_4, T_4)	2.153592×10^3	8.693329×10^3	2.850994×10^3
(P_5, T_5)	2.085177×10^3	5.786375×10^3	2.728477×10^3
(P_6, T_6)	2.105928×10^3	5.011749×10^3	2.542974×10^3
(P_7, T_7)	2.197608×10^3	5.095652×10^3	2.46543×10^3
(P_8, T_8)	2.012208×10^3	5.312943×10^3	2.694597×10^3
(P_9, T_9)	2.00875×10^3	5.804202×10^3	2.700028×10^3
(P_{10}, T_{10})	2.463717×10^3	4.60444×10^3	2.285393×10^3

Table 8. Estimators Evaluation using PRE (Population-2)

(P_A, T_A)	0.25	0.30	0.35
(P_1, T_1)	2.768503×10^3	9.0045×10^3	7.385592×10^{14}
(P_2, T_2)	2.399155×10^3	9.496214×10^3	9.709526×10^{15}
(P_3, T_3)	1.863304×10^3	8.523574×10^3	6.645172×10^{16}
(P_4, T_4)	585.6465	5.804935×10^3	5.842608×10^{18}
(P_5, T_5)	2.582901×10^3	9.207251×10^3	3.234888×10^{15}
(P_6, T_6)	2.917805×10^3	9.360635×10^3	1.311966×10^{15}
(P_7, T_7)	2.791368×10^3	9.341671×10^3	1.358854×10^{15}
(P_8, T_8)	2.954321×10^3	9.462238×10^3	1.728212×10^{15}
(P_9, T_9)	2.474641×10^3	1.068831×10^4	2.46542×10^{15}
(P_{10}, T_{10})	1.339821×10^3	1.180448×10^4	1.811542×10^{15}
(P_A, T_A)	0.40	0.45	0.50
(P_1, T_1)	4.107627×10^4	1.693958×10^4	2.586708×10^5
(P_2, T_2)	7.703258×10^3	1.619008×10^4	3.261126×10^3
(P_3, T_3)	7.451877×10^3	1.518674×10^4	3.164019×10^3
(P_4, T_4)	7.776421×10^3	1.799456×10^4	3.316095×10^3
(P_5, T_5)	7.679322×10^3	1.606042×10^4	3.299117×10^3
(P_6, T_6)	7.584249×10^3	1.791803×10^4	3.706185×10^3
(P_7, T_7)	7.420787×10^3	1.675421×10^4	3.537222×10^3
(P_8, T_8)	7.784274×10^3	1.884092×10^4	3.679087×10^3
(P_9, T_9)	7.976716×10^3	2.381457×10^4	3.944749×10^3
(P_{10}, T_{10})	6.041582×10^3	1.38276×10^4	2.958276×10^3

Table 9. Estimators Evaluation using PRE (Population-3)

(P_A, T_A)	0.25	0.30	0.35
(P_1, T_1)	4.666512×10^2	5.015698×10^2	2.414063×10^{12}
(P_2, T_2)	5.31122×10^2	4.465444×10^2	2.178672×10^{12}
(P_3, T_3)	5.352423×10^2	4.44284×10^2	1.854652×10^{12}
(P_4, T_4)	5.466841×10^2	4.602478×10^2	4.655237×10^{13}
(P_5, T_5)	4.769844×10^2	4.991451×10^2	1.653124×10^{12}
(P_6, T_6)	4.825621×10^2	4.568759×10^2	2.877582×10^{12}
(P_7, T_7)	5.264281×10^2	4.473373×10^2	2.18274×10^{12}
(P_8, T_8)	4.544179×10^2	4.626518×10^2	2.367088×10^{12}

(P_9, T_9)	5.184594×10^2	4.418098×10^2	2.810127×10^{12}
(P_{10}, T_{10})	5.538713×10^2	4.467691×10^2	5.323063×10^{12}
(P_A, T_A)	0.40	0.45	0.50
(P_1, T_1)	1.407711×10^5	3.046596×10^2	6.7572371×10^7
(P_2, T_2)	2.618535×10^2	2.715337×10^2	2.960144×10^2
(P_3, T_3)	2.632322×10^2	2.760454×10^2	3.022436×10^2
(P_4, T_4)	2.489735×10^2	4.763507×10^2	3.596965×10^2
(P_5, T_5)	2.929212×10^2	3.040154×10^2	2.748141×10^2
(P_6, T_6)	2.509963×10^2	2.66142×10^2	3.018371×10^2
(P_7, T_7)	2.605829×10^2	2.737932×10^2	2.948213×10^2
(P_8, T_8)	2.437243×10^2	2.762255×10^2	3.01117×10^2
(P_9, T_9)	2.652059×10^2	3.033256×10^2	3.124778×10^2
(P_{10}, T_{10})	2.656117×10^2	2.795552×10^2	2.890372×10^2

Table 10. Estimators Evaluation using PRE (Population-4)

(P_A, T_A)	0.25	0.30	0.35
(P_1, T_1)	1.121715×10^4	1.314173×10^5	6.718913×10^{16}
(P_2, T_2)	8.998006×10^3	1.158054×10^5	4.631947×10^{18}
(P_3, T_3)	7.953604×10^3	1.331955×10^5	1.302405×10^{19}
(P_4, T_4)	6.639765×10^3	1.273503×10^5	4.429973×10^{21}
(P_5, T_5)	1.09571×10^4	1.478618×10^5	1.109263×10^{17}
(P_6, T_6)	1.203781×10^4	9.753564×10^4	2.747088×10^{17}
(P_7, T_7)	1.179783×10^4	9.485816×10^4	5.1655×10^{17}
(P_8, T_8)	1.170043×10^4	1.133394×10^5	2.240363×10^{17}
(P_9, T_9)	1.2294×10^4	1.084723×10^5	1.655589×10^{18}
(P_{10}, T_{10})	1.074773×10^4	8.44686×10^4	1.711465×10^{21}
(P_A, T_A)	0.40	0.45	0.50
(P_1, T_1)	7.4107824×10^3	3.945285×10^3	1.549238299×10^9
(P_2, T_2)	4.932476×10^3	4.564053×10^3	8.294752×10^3
(P_3, T_3)	4.644468×10^3	4.70993×10^3	8.480287×10^3
(P_4, T_4)	5.803376×10^3	7.248828×10^3	9.347197×10^3
(P_5, T_5)	4.147166×10^3	3.928855×10^3	1.133633×10^4
(P_6, T_6)	4.773933×10^3	4.192888×10^3	9.341423×10^3
(P_7, T_7)	5.252702×10^3	4.536428×10^3	8.936533×10^3
(P_8, T_8)	4.564543×10^3	4.013872×10^3	1.062154×10^4
(P_9, T_9)	5.087607×10^3	4.569583×10^3	8.756394×10^3
(P_{10}, T_{10})	5.066327×10^3	4.615374×10^3	8.605661×10^3

5. Conclusion

This study developed new families of estimators to estimate the population cumulative distribution function (CDF) through a calibration strategy in stratified sampling (StRS) with several calibration constraints and the distance function of Lata [17]. The proposed estimators were empirically tested using simulations on four real-life datasets, and they were found to be the best in terms of mean squared error (MSE) and precision, especially in heterogeneous populations. The main academic contribution of this work is the application of the calibration framework to estimate CDFs by combining a range of auxiliary information as well as a new distance function, which provides a more versatile and powerful estimation strategy than others. The findings suggest that the estimators should be used in real-life scenarios, including official statistics, environmental surveillance, and studies of public health, where distributional estimation accuracy and auxiliary data are crucial. Future studies can be carried out to make the study more advanced to two-stage or adaptive sampling and to examine the theoretical properties of the proposed estimators, such as their asymptotic properties and behavior under model misspecification.

Declarations

Author Contributions

Conceptualization: P.F. and P.F.; methodology: P.F.; software: P.F.; validation: P.F., P.F., and P.F.; formal analysis: P.F.; investigation: P.F.; resources: P.F., data curation; P.F., writing—original draft preparation; P.F., writing—review and editing; P. F., visualization; P. F., supervision; P. F., project administration; P.F., funding acquisition; P. F., all authors have read and agreed to the published version of the manuscript.

Data Availability Statement

All the dataset information is already available in Reference [25].

Funding

No funding was provided by any organization for research or for the preparation and presentation of this article.

Institutional Review Board Statement

"*Not Applicable*".

Informed Consent Statement

"*Not Applicable*".

Conflicts of Interest

The authors declare that they have no conflicts of interest during the development of this research.

References

[1] Chambers, R. L., & Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika, 73*(3), 597-604.

[2] Rao, J. N. K., Kovar, J. G., & Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika,77*, 365-375.

[3] Chen, J., & Wu, C. (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistica Sinica,12,* 1223-1239.

[4] Rueda, M., Martínez, S., Martínez, H., & Arcos, A. (2007). Estimation of the distribution function with calibration methods. *Journal of statistical planning and inference, 137*(2), 435-448.

[5] Yaqub, M., & Shabbir, J. (2018). Estimation of population distribution function in the presence of non-response. *Hacettepe Journal of Mathematics and Statistics, 47*(2), 471-511.

[6] Hussain, S., Ahmad, S., Saleem, M., & Akhtar, S. (2020). Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. *Plos one, 15*(9), e0239098.

[7] Ahmad, S., Aamir, M., Hussain, S., Shabbir, J., Zahid, E., Subkrajang, K., & Jirawattanapanit, A. (2022a). A new generalized class of exponential factor-type estimators for population distribution function using two auxiliary variables. *Mathematical Problems in Engineering,* 2022.

[8] Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association, 87*(418), 376-382.

[9] Tracy, D. S., Singh, S., & Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology, 29*(1), 99-104.

[10] Kim, J. M., Sungur, E. A., & Heo, T. Y. (2007). Calibration approach estimators in stratified sampling. *Statistics & probability letters, 77*(1), 99-103.

[11] Rao, D. K., Khan, M. G., & Khan, S. (2012). Mathematical programming on multivariate calibration estimation in stratified sampling. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, 6*(12), 58-62.

[12] Rao, D. K., Khan, M., & Singh, G. (2017). On calibrated weights in stratified sampling. *ANZIAM Journal, 59*, 190-204.

[13] Koyuncu, N., & Kadilar, C. (2013). Calibration estimator using different distance measures in stratified random sampling. *International Journal of Modern Engineering Research, 3*(1), 415-419.

[14] Koyuncu, N., & KADILAR, C. (2014). A new calibration estimator in stratified double sampling. *Hacettepe Journal of Mathematics and Statistics, 43*(2), 337-346.

[15] Nidhi, Sisodia, B. V. S., Singh, S., & Singh, S. K. (2017). Calibration approach estimation of the mean in stratified sampling and stratified double sampling. *Communications in Statistics-Theory and Methods, 46*(10), 4932-4942.

[16] Koyuncu, N. (2018). Calibration estimator of population mean under stratified ranked set sampling design. *Communications in Statistics-Theory and Methods, 47*(23), 5845-5853.

[17] Lata, A. S., Rao, D., & Khan, M. G. (2017, December). Calibration estimation using proposed distance function. In *2017 4th Asia-Pacific World Congress on Computer Science and Engineering,* 162-166.

[18] Ozgul, N. (2019). New calibration estimator based on two auxiliary variables in stratified sampling. *Communications in Statistics-Theory and Methods, 48*(6), 1481-1492.

[19] Garg, N., & Pachori, M. (2020). Use of coefficient of variation in calibration estimation of population mean in stratified sampling. *Communications in Statistics-Theory and Methods, 49*(23), 5842-5852.

[20] Shahzad, U., Ahmad, I., Almanjahie, I., Hanif, M., & Al-Noor, N. H. (2021). L-Moments and calibration-based variance estimators under double stratified random sampling scheme: Application of Covid-19 pandemic. *Scientia Iranica*, *30*(2), 814-821.

[21] Shahzad, U., Ahmad, I., Almanjahie, I., & Al-Noor, N. H. (2021). Utilizing L-moments and calibration method to estimate the variance based on covid-19 data. *Fresenius Environmental Bulletin, 30*(7), 8988-8994.

[22] Shahzad, U., Ahmad, I., Hanif, M., & Al-Noor, N. H. (2022). Estimation of coefficient of variation using linear moments and calibration approach for nonsensitive and sensitive variables. *Concurrency and Computation: Practice and Experience*, *34*(18), e7006.

[23] Shahzad, U., Ahmad, I., Al-Noor, N. H., & Benedict, T. J. (2022). Use of calibration constraints and

linear moments for variance estimation under stratified adaptive cluster sampling. *Soft Computing*, *26*(21), 11185-11196.
[24] Shahzad U, Ahmad I, García-Luengo AV, Zaman T, Al-Noor NH, Kumar A. (2023). Estimation of Coefficient of Variation Using Calibrated Estimators in Double Stratified Random Sampling. *Mathematics. 11*(1),252.
[25] Abbasi, H., Hanif, M., Shahzad, U., Emam, W., Tashkandy, Y., Iftikhar, S., & Shahzadi, S. (2023). Calibration Estimation of Cumulative Distribution Function Using Robust Measures. *Symmetry*, *15*(6), 1157.
[26] Abbasi, H., Shahzad, U., Emam, W., Hanif, M., Ali, N., Mukhtar, M. (2024). Calibrated empirical neutrosophic cumulative distribution function estimation for both symmetric and asymmetric data. *Symmetry, 16*(5), 633.
[27] Zaman, T., & Kadilar, C. (2021). Exponential ratio and product type estimators of the mean in stratified two-phase sampling. *AIMS Mathematics*, *6*(5), 4265-4279.

参考文献:

[1] Chambers, R. L., & Dunstan, R. (1986). 基于调查数据的分布函数估计 Biometrika, 73 (3), 597-604.
[2] Rao, J. N. K., Kovar, J. G., & Mantel, H. J. (1990). 利用辅助信息从调查数据中估计分布函数和分位数 Biometrika, 77, 365-375.
[3] Chen, J., & Wu, C. (2002). 基于模型校准伪经验似然法的分布函数与分位数估计 Statistica Sinica, 12, 1223-1239.
[4] Rueda, M., Martínez, S., Martínez, H., & Arcos, A. (2007). 基于校准方法的分布函数估计 Journal of Statistical Planning and Inference, 137 (2), 435-448.
[5] Yaqub, M., & Shabbir, J. (2018). 存在无回答情况下的总体分布函数估计 Hacettepe Journal of Mathematics and Statistics, 47 (2), 471-511.
[6] Hussain, S., Ahmad, S., Saleem, M., & Akhtar, S. (2020). 简单随机抽样和分层随机抽样下双重利用辅助信息的有限总体分布函数估计 Plos One, 15 (9), e0239098.
[7] Ahmad, S., Aamir, M., Hussain, S., Shabbir, J., Zahid, E., Subkrajang, K., & Jirawattanapanit, A. (2022a). 利用两个辅助变量的总体分布函数指数因子型估计的新广义类 Mathematical Problems in Engineering, 2022.
[8] Deville, J. C., & Särndal, C. E. (1992). 抽样调查中的校准估计量 Journal of the American Statistical Association, 87 (418), 376-382.
[9] Tracy, D. S., Singh, S., & Arnab, R. (2003). 关于分层抽样和二重抽样中校准的注记 Survey Methodology, 29 (1), 99-104.
[10] Kim, J. M., Sungur, E. A., & Heo, T. Y. (2007). 分层抽样中的校准方法估计量 Statistics & Probability Letters, 77 (1), 99-103.
[11] Rao, D. K., Khan, M. G., & Khan, S. (2012). 分层抽样中多元校准估计的数学规划 International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, 6 (12), 58-62.
[12] Rao, D. K., Khan, M., & Singh, G. (2017). 论分层抽样中的校准权数 ANZIAM Journal, 59, 190-204.
[13] Koyuncu, N., & Kadilar, C. (2013). 分层随机抽样中使用不同距离测度的校准估计量 International Journal of Modern Engineering Research, 3 (1), 415-419.
[14] Koyuncu, N., & Kadilar, C. (2014). 分层二重抽样中的一种新校准估计量 Hacettepe Journal of Mathematics and Statistics, 43 (2), 337-346.
[15] Nidhi, Sisodia, B. V. S., Singh, S., & Singh, S. K. (2017). 分层抽样和分层二重抽样中均值的校准方法估计 Communications in Statistics-Theory and Methods, 46 (10), 4932-4942.
[16] Koyuncu, N. (2018). 分层有序抽样设计下总体均值的校准估计量 Communications in Statistics - Theory and Methods, 47 (23), 5845-5853.
[17] Lata, A. S., Rao, D., & Khan, M. G. (2017, December). 基于所提距离函数的校准估计In 2017 4th Asia-Pacific World Congress on Computer Science and Engineering, 2017年第四届亚太计算机科学与工程世界大会，162-166.
[18] Ozgul, N. (2019). 分层抽样中基于两个辅助变量的新校准估计量 Communications in Statistics - Theory and Methods, 48 (6), 1481-1492.
[19] Garg, N., & Pachori, M. (2020). 变异系数在分层抽样总体均值校准估计中的应用 Communications in Statistics - Theory and Methods, 49 (23), 5842-5852.
[20] Shahzad, U., Ahmad, I., Almanjahie, I., Hanif, M., & Al-Noor, N. H. (2021). 双重分层随机抽样方案下基于 L-矩和校准的方差估计量：新冠疫情的应用 Scientia Iranica, 30 (2), 814-821.
[21] Shahzad, U., Ahmad, I., Almanjahie, I., & Al-Noor, N. H. (2021). 利用 L-矩和校准方法基于新冠疫情数据估计方差 Fresenius Environmental Bulletin, 30 (7), 8988-8994.
[22] Shahzad, U., Ahmad, I., Hanif, M., & Al-Noor, N. H. (2022).

对非敏感和敏感变量利用线性矩和校准方法估计变异系数 Concurrency and Computation: Practice and Experience, 34 (18), e7006.
[23] Shahzad, U., Ahmad, I., Al-Noor, N. H., & Benedict, T. J. (2022). 分层自适应整群抽样下方差估计中校准约束与线性矩的应用 Soft Computing, 26 (21), 11185-11196.
[24] Shahzad U, Ahmad I, García-Luengo AV, Zaman T, Al-Noor NH, Kumar A. (2023). 双重分层随机抽样中利用校准估计量估计变异系数 Mathematics, 11 (1), 252.
[25] Abbasi, H., Hanif, M., Shahzad, U., Emam, W., Tashkandy, Y., Iftikhar, S., & Shahzadi, S. (2023). 基于稳健测度的累积分布函数校准估计 Symmetry, 15 (6), 1157.
[26] Abbasi, H., Shahzad, U., Emam, W., Hanif, M., Ali, N., Mukhtar, M. (2024). 对称和非对称数据的校准经验中智累积分布函数估计 Symmetry, 16 (5), 633.
[27] Zaman, T., & Kadilar, C. (2021). 分层两阶段抽样中均值的指数比和乘积型估计量 AIMS Mathematics, 6 (5), 4265-4279.

Word count: 5,529 words, excluding references.

Peer-review record:
Fast-track status: Not fast-tracked
First-round reviews received: 3 reports
Revision cycles completed: 3 rounds
Final version submitted: October 6, 2025

Disclaimer/Publisher’s Note:

The views, opinions and data expressed in this article are solely those of the authors and do not necessarily reflect those of the **Journal of Hunan University (Natural Sciences)** or its editors. The journal and its editorial staff accept no responsibility for any injury to persons or damage to property resulting from the ideas, methods, instructions or products discussed herein.