symmetry

MDPI

*Article*

# Advancing Survey Sampling Efficiency under Stratified Random Sampling and Post-Stratification: Leveraging Symmetry for Enhanced Estimation Accuracy in the Prediction of Exam Scores

**Gullinkala Ramya Venkata Triveni [1], Faizan Danish [1,*] and Olayan Albalawi [2]**

[1] Department of Mathematics, School of Advanced Sciences, VIT-AP University, Inavolu, Beside AP Secretariat, Amaravati 522237, Andhra Pradesh, India; gullinkala.22phd7110@vitap.ac.in
[2] Department of Statistics, Faculty of Science, University of Tabuk, Tabuk 47713, Saudi Arabia
* Correspondence: faizan.danish@vitap.ac.in

**Abstract:** This pioneering investigation introduces two innovative estimators crafted to evaluate the finite population distribution function of a study variable, employing auxiliary variables within the framework of stratified random sampling and post-stratification while emphasizing symmetry in the sampling process. The derivation of mathematical expressions for bias and the mean square error up to the first degree of approximation fortifies the credibility of the proposed estimators. Drawing from three distinct datasets, including real-world data capturing student behaviors and exam performances from 500 students, this research highlights the superior efficiency of the proposed estimators compared to existing methods across both sampling schemes. Employing the proposed estimator, we effectively forecast students' exam scores based on their study hours, backed by empirical evidence showcasing its precision in terms of mean square error and percentage relative efficiency. This study not only introduces inventive solutions to enduring challenges in survey sampling but also provides practical insights into enhancing predictive accuracy in educational assessments.

**Keywords:** cumulative distribution function; stratified sampling; post-stratification; percentage relative efficiency; symmetry

**MSC:** 62D05; 94A20

check for updates

**Citation:** Triveni, G.R.V.; Danish, F.; Albalawi, O. Advancing Survey Sampling Efficiency under Stratified Random Sampling and Post-Stratification: Leveraging Symmetry for Enhanced Estimation Accuracy in the Prediction of Exam Scores. *Symmetry* **2024**, *16*, 604. https://doi.org/10.3390/sym16050604

Academic Editor: Qing-Wen Wang

Received: 27 March 2024
Revised: 4 May 2024
Accepted: 6 May 2024
Published: 13 May 2024



## 1. Introduction

In the pursuit of estimating population parameters, auxiliary variables are crucial tools. These variables, which are different but intricately linked to the variable of interest, offer a dependable method for improving the consistency and validity of statistical estimations. Based on survey sampling theory, the importance of auxiliary variables becomes clear, especially in guaranteeing symmetry in the sampling process. Given the impracticality or difficulty of collecting comprehensive data from the entire population in survey research, researchers use sampling—an intentional selection of a smaller subset—to collect data, aiming for symmetry in representation. The goal is to extrapolate findings from this sample to the entire population, using auxiliary variables that promote symmetry in the estimate. Using auxiliary variables in the sample and estimating process appears to be a powerful method for boosting estimation accuracy while maintaining symmetry in the representation of population features. Previous research has clarified a variety of population parameters, including mean, median, total, distribution function, etc., each of which requires supportive variable data in addition to the variable of interest, adding to the symmetrical representation of population traits. In stratified sampling, a distribution function estimator is useful for estimating the cumulative distribution function (CDF) within each stratum. These estimates are then combined to generate a comprehensive estimate for the entire population, ensuring symmetry throughout the estimating process.

Over the years, numerous scholars have delved into different facets of estimators in stratified random sampling (St RS), enriching our understanding and refining the methodologies in this critical domain. The author of [1] addressed the credibility of the approximate formula for computing variance, while [2] discussed estimation methods and the dual property of ratio estimation. Eventually, ref. [3] focused on techniques in post-stratification, and later, ref. [4] demonstrated methods to improve ratio and regression estimators. Further, ref. [5] explored the characteristics of estimators for finite population distribution functions. Later, ref. [6] proposed a Bayesian model-based theory for post-stratification, and [7] presented calibration estimators using auxiliary data. Further, refs. [8,9] also contributed with their estimators for distribution functions in post-stratification.

Later, ref. [10] extended estimators with readily accessible supporting variables. An efficient ratio estimator for stratified sampling was introduced by [11]. Further, ref. [12] presented a family of estimators for the population mean, which was validated empirically. Later, ref. [13] proposed superior exponential ratio estimators. Further, ref. [14] derived a diligent ratio and product estimator, which outperformed others. Thereafter, ref. [15] devised exponential ratio estimators based on supporting variables. Additionally, ref. [16] suggested reliable ratio and difference estimators for population distributions. Researchers have made significant advancements in sampling estimation. Later, ref. [17] improved exponential estimators for post-stratification. Ref. [18] introduced superior estimators for SRS and stratified sampling with two supporting variables. Following that, ref. [19] enhanced the difference cum exponential estimator. Thereafter, ref. [20] proposed efficient estimators. Ref. [21] suggested a ratio estimator for post-stratification. Subsequently, ref. [22] improved the generalized population mean estimator, while [23] developed estimators for the finite population mean in SRS. Later, ref. [24] presented a two-parameter ratio product ratio estimator. Afterward, ref. [25] suggested an innovative family of exponential estimators using supporting attributes and actual data sets.

In recent times, several authors have focused on distribution function estimators using supporting variables. Ref. [26] proposed finite population distribution function estimators, which have outperformed others in simple random sampling (SRS) and stratified sampling. Following that, ref. [27] introduced imputation methods for calculating the population mean in two-occasion successive sampling. Eventually, ref. [28] recommended exponential-type estimators for finite population mean, demonstrating superiority with four data sets. Thereafter, ref. [29] devised estimators for the population mean and efficiently combined and separate estimators in stratified sampling. Additionally, ref. [30] proposed a ratio estimator with the highest effectiveness via empirical and simulation studies. Afterward, ref. [31] developed an estimator for estimating the population distribution function and proved its efficiency by using a simulation study. Further, ref. [32] discussed the efficiency of the ratio estimator in stratified sampling and proved its efficiency by utilizing empirical studies. Later, ref. [33] suggested robust-type estimators for population variance, outperforming existing methods in simple and St RS. A hybrid estimator for the population mean was proposed by [34], showing superior efficiency through empirical and simulated experiments. Later, ref. [35] proposed a log-type estimator in stratified ranked set sampling. A new approach to the mean estimators in ranked set sampling was introduced by [36].

The literature on the estimation of CDFs is notably sparse, highlighting a significant gap in research. In response, this article is committed to advancing this field by introducing innovative CDF estimators. Our focus lies in proposing two distinctive classes of estimators that harness auxiliary variable information to accurately estimate the CDF of a specific variable under examination. By leveraging auxiliary variables, our proposed estimators aim to fill this gap and provide enhanced methods for estimating CDFs, thus contributing to the broader advancement of statistical estimation techniques. The paper has been systematically organized to enhance clarity and coherence in presenting the research on stratified and post-stratified sampling methods. Beginning with an introduction that sets the stage for the study, Section 2 elucidates key terms and concepts essential for understanding the subsequent discussion. The literature review delves into existing estimators in both stratified and post-stratified sampling, laying the groundwork for the methodology section, where novel estimators for each method are proposed. The theoretical framework provides a theoretical underpinning for both sampling techniques, while Section 6 details the implementation and outcomes of empirical investigations conducted for each method. Section 7 brings together the findings from both empirical studies, facilitating a comprehensive analysis of the proposed estimators and their implications. Finally, Section 8 offers a concise summary of the study's key findings and their significance for future research and practical applications. This organization ensures a logical flow of ideas and a clear delineation of the contributions made in the domains of both stratified and post-stratified sampling methodologies.

## 2. Background and Notations

### *2.1. Notations in Stratified Random Sampling*

To evaluate the finite population distribution function, regarding a finite population, $\Omega = 1,\ 2,\ 3, \ldots\ N$ of $N$ distinct units is distributed to k homogeneous strata, $N_h$ is the size of $h^{th}$ stratum such that $\sum_{h=1}^{k}\ N_h\ = N$. A sample size $n_h\left(\sum_{h=1}^{k} n_h = n\right)$ is taken from the $h^{th}$ stratum by utilizing SRS without replacement.

Let $F_{st}(y) = F(y) = \sum_{h=1}^{k} W_h F_h(y)$ and $F_{st}(x) = F(x) = \sum_{h=1}^{k} W_h F_h(x)$ be the population distribution function of the variables $Y$ (study variable) and $X$ (auxiliary variable) under St RS, respectively. Let $\hat{F}_{st}(y) = \hat{F}(y) = \sum_{h=1}^{k} W_h \hat{F}_h(y)$ and $\hat{F}_{st}(x) = \hat{F}(x) = \sum_{h=1}^{k} W_h \hat{F}_h(x)$ be the sample distribution functions of the variables $Y$ and $X$, respectively.

where $h = 1,2,3\ldots k$ and $i = 1,2,3\ldots N_h$ and $n =$ sample size.

- $W_h = \frac{N_h}{N}$ denotes the stratum weight of $h^{th}$ stratum.
- $F_h(y) = \sum_{i=1}^{N_h} \Delta(Y_{ih} \le y)/N_h$ and $\hat{F}_h(y) = \sum_{i=1}^{n_h} \Delta(Y_{ih} \le y)/n_h$ represents population and sample distribution functions of $Y$ for the $h^{th}$ stratum and $\Delta(Y_{ih} \le y)$ is the indicator variable of $Y$.
- $F_h(x) = \sum_{i=1}^{N_h} \Delta(X_{ih} \le x)/N_h$ and $\hat{F}_h(x) = \sum_{i=1}^{n_h} \Delta(X_{ih} \le x)/n_h$ represents the population and sample distribution functions of $X$ for the $h^{th}$ stratum and $\Delta(X_{ih} \le x)$ is the indicator variable of $X$.

Here, we consider error terms for finding bias and MSE of the estimators.

Let $e_y = \frac{\hat{F}_{st}(y) - F(y)}{F(y)}$, $\hat{F}_{st}(y) = F(y) + e_y F(y) = F(y)(1 + e_y)$

and $e_x = \frac{\hat{F}_{st}(x) - F(x)}{F(x)}$, $\hat{F}_{st}(x) = F(x) + e_x F(x) = F(x)(1 + e_x)$

$$E(e_y) = E(e_x) = 0$$

$$E\left(e_y^2\right) = \sum_{h=1}^{k} w_h^2 \lambda_h C_{yh}^2 = V_{0st}(say)$$

$$E\left(e_x^2\right) = \sum_{h=1}^{k} w_h^2 \lambda_h C_{xh}^2 = V_{1st}(say),$$

$$E(e_x e_y) = \sum_{h=1}^{k} w_h^2 \lambda_h R_{xyh} C_{xh} C_{yh} = V_{01st}(say)$$

where

$$C_{yh} = \frac{S_{yh}}{F(y)},\ C_{xh} = \frac{S_{xh}}{F(x)},\ \lambda_h = \frac{1}{n_h} - \frac{1}{N_h} \text{ and } R_{xyh} = \frac{S_{xyh}}{S_{yh} S_{xh}}.$$

$$S_{yh}^2 = \sum_{i=1}^{N_h} (\Delta(Y_{ih} \le y) - F(y))^2/(N-1),$$

$$S_{xh}^2 = \sum_{i=1}^{N_h} (\Delta(X_{ih} \le x) - F(x))^2/(N-1),$$

$$S_{xyh} = \sum_{i=1}^{N_h} (\Delta(Y_{ih} \le y))(\Delta(X_{ih} \le x))/(N-1).$$

*2.2. Notations in Post-Stratification*

Post-stratification in survey sampling addresses missing crucial attributes by dividing the population into subgroups based on known auxiliary variables. Survey weights are adjusted to account for variations in the distribution of these variables, mitigating biases from nonresponse and small sample sizes. By using CDF estimators, researchers achieve more precise distribution estimates within specific subgroups, enhancing the understanding of the study variable's characteristics.

In post-stratification, the traditional unbiased estimator of the population distribution function is referred to as

$$\hat{F}_{ps}(y) = \sum_{h=1}^{K} W_h \hat{F}_h(y)$$

where $\hat{F}_{ps}(y)$ is the post-stratified empirical distribution function at $y$

$K$ is the number of post-strata.

$\hat{F}_h(y)$ is the distribution function of $y$ for the $h^{th}$ stratum.

Variance of $\hat{F}_{ps}(y)$ is formulated as

$$Var(\hat{F}_{ps}(y)) = \left[\frac{1}{n} - \frac{1}{N}\right] \sum_{h=1}^{K} W_h S_{yh}^2 - \frac{1}{n^2} \sum_{h=1}^{K} (1 - W_h) S_{yh}^2$$

Consider the error terms below to obtain the bias and MSE of our proposed estimator,

$$e_0 = \frac{\sum_{h=1}^{K} W_h F_h(y) e_{0h}}{F(y)},\ e_1 = \frac{\sum_{h=1}^{K} W_h F_h(x) e_{1h}}{F(x)} \text{ and } e_2 = \frac{\sum_{h=1}^{K} W_h F_h(\overline{x}) e_{2h}}{F(\overline{x})}$$

where

$$e_{0h} = \frac{\hat{F}_{ps}(y) - F_h(y)}{F_h(y)},\ e_{1h} = \frac{\hat{F}_{ps}(x) - F_h(x)}{F_h(x)} \text{ and } e_{2h} = \frac{\hat{F}_{ps}(\overline{x}) - F_h(\overline{x})}{F_h(\overline{x})}$$

$$E(e_{0h}) = E(e_{1h}) = E(e_{2h}) = 0$$

$$E(e_{0h}^2) = \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] C_{yh}^2,\ E(e_{1h}^2) = \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] C_{xh}^2,\ E(e_{2h}^2) = \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] C_{\overline{x}h}^2$$

$$E(e_{0h}e_{1h}) = \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] R_{xyh} C_{xh} C_{yh}, E(e_{0h}e_{1h}) = \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] R_{\overline{x}yh} C_{\overline{x}h} C_{yh}$$

$$E(e_{0h}e_{1h}) = \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] R_{x\overline{x}h} C_{xh} C_{\overline{x}h}.$$

Now, we will find the expected values of error terms:

$$E(e_0) = E\left(\frac{\sum_{h=1}^{K} W_h F_h(y) e_{0h}}{F(y)}\right) = \frac{1}{F(y)}\left(\sum_{h=1}^{K} W_h F_h(y) E(e_{0h})\right) = 0$$

Similarly,

$$E(e_0) = E(e_1) = E(e_2) = 0$$

$$E(e_0^2) = E\left(\frac{\sum_{h=1}^{K} W_h F_h(y) e_{0h}}{F(y)}\right)^2 = \frac{1}{F^2(y)} \sum_{h=1}^{K} W_h^2 F_h^2(y) E(e_{0h}^2)$$

$$= \frac{1}{F^2(y)} \sum_{h=1}^{K} W_h^2 F_h^2(y) \left[\frac{1}{nW_h} - \frac{1}{N_h}\right] C_{yh}^2$$

$$E(e_0^2) = \frac{1}{F^2(y)} \left[\frac{1}{n} - \frac{1}{N}\right] \sum_{h=1}^{K} W_h S_{yh}^2 = V_{0ps}(say)$$

Similarly,

$$E(e_1^2) = \frac{1}{F^2(x)} \left[\frac{1}{n} - \frac{1}{N}\right] \sum_{h=1}^{K} W_h S_{xh}^2 = V_{1ps}(say)$$

$$E(e_2^2) = \frac{1}{\overline{X}^2} \left[\frac{1}{n} - \frac{1}{N}\right] \sum_{h=1}^{K} W_h S_{\overline{x}h}^2 = V_{2ps}(say)$$

and

$$E(e_0e_1) = \frac{1}{F(x)F(y)}\left[\frac{1}{n} - \frac{1}{N}\right]\sum_{h=1}^{K} W_h R_{xyh} C_{yh} C_{xh} = V_{01ps}(say)$$
$$E(e_0e_2) = \frac{1}{F(y)\overline{X}}\left[\frac{1}{n} - \frac{1}{N}\right]\sum_{h=1}^{K} W_h R_{y\overline{x}h} C_{yh} C_{\overline{x}h} = V_{02ps}(say)$$
$$E(e_1e_2) = \frac{1}{F(x)\overline{X}}\left[\frac{1}{n} - \frac{1}{N}\right]\sum_{h=1}^{K} W_h R_{x\overline{x}h} C_{\overline{x}h} C_{xh} = V_{12ps}(say)$$

## 3. Literature Review

### *3.1. Pre-Existing Estimators under Stratified Random Sampling*

Several authors have suggested estimators for calculating the finite population mean in stratified sampling. We adopted them in the cumulative distribution function estimators under stratified sampling to evaluate the population cumulative distribution function of the study variable by utilizing the knowledge of information of supporting variable. Until the first degree of approximation, we obtained bias and MSE equations for the following pre-existing estimators, which were given by prominent authors.

1. The usual unbiased estimator of $F(y)$ is given as

$$\hat{F}_{SRS_{st}}(y) = \frac{1}{n}\sum_{i=1}^{n} \Delta(Y_i \leq y)$$

The MSE of $\hat{F}_{SRS_{st}}(y)$ is

$$MSE\left(\hat{F}_{SRS_{st}}(y)\right) = F^2(y)\sum_{h=1}^{k} w_h^2 \lambda_h C_{yh}^2 = F^2(y)V_{0st} \tag{1}$$

2. The classical ratio estimator of $F(y)$, according to [1], is given by

$$\hat{F}_{Re}(y) = \hat{F}_{st}(y)\left[\frac{F(x)}{\hat{F}_{st}(x)}\right]$$

And its bias and MSE become

$$\begin{aligned} Bias\left(\hat{F}_{Re}(y)\right) &= F(y)(V_{1st} - V_{01st}) \\ MSE\left(\hat{F}_{Re}(y)\right) &= F^2(y)(V_{0st} + V_{1st} - 2V_{01st}) \end{aligned} \tag{2}$$

3. The product estimator of $F(y)$ was proposed by [2]

$$\hat{F}_{Pe}(y) = \hat{F}_{st}(y)\left[\frac{\hat{F}_{st}(x)}{F(x)}\right]$$

Its bias and MSE become approximately constant until the first order:

$$\begin{aligned} Bias\left(\hat{F}_{Pe}(y)\right) &= F(y)V_{01st} \\ MSE\left(\hat{F}_{Pe}(y)\right) &= F^2(y)(V_{0st} + V_{1st} + 2V_{01st}) \end{aligned} \tag{3}$$

4. A difference-type estimator was proposed by [4]:

$$\hat{F}_{De}(y) = m_1\hat{F}_{st}(y) + m_2\left[F(x) - \hat{F}_{st}(x)\right]$$

where $m_1$ and $m_2$ are unknown fixed values.

$$Bias\left(\hat{F}_{De}(y)\right) = F(y)(m_1 - 1)$$
$$MSE\left(\hat{F}_{De}(y)\right) = F^2(y) - 2m_1F^2(y) + m_1^2F^2(y) + m_1^2F^2(y)V_{0st} - 2m_1m_2F(x)F(y)V_{01st} + m_2^2F^2(x)V_{1st}$$

After minimizing $MSE\left(\hat{F}_{De}(y)\right)$, we get optimum values as:

$m_1 = \frac{V_{1st}}{V_{1st}V_{0st} - (V_{01st})^2 + V_{0st}}$ and $m_2 = \frac{F(y)V_{01st}}{F(x)\left(V_{1st}V_{0st} - (V_{01st})^2 + V_{0st}\right)}$

Now, $MSE\ \left(\hat{F}_{De}(y)\right)$. Can we rewrite it as

$$MSE_{min}\left(\hat{F}_{De}(y)\right) = \frac{F^2(y)\left(V_{0st}V_{1st} - (V_{01st})^2\right)}{V_{1st}V_{0st} - (V_{01st})^2 + V_{0st}} \tag{4}$$

5. A generalized ratio type exponential estimator was adopted by [13]:

$$\hat{F}_{REe}(y) = \hat{F}_{st}(y)exp\left(\frac{a_{st}\left(F(x) - \hat{F}_{st}(x)\right)}{a_{st}\left(F(x) + \hat{F}_{st}(x)\right) + 2b_{st}}\right)$$

Here, $a_{st}$ and $b_{st}$ are fixed values, and bias and MSE will be

$$\begin{aligned} Bias\left(\hat{F}_{REe}(y)\right) &= F(y)\left(\tfrac{3}{8}\textstyle\sum_{h=1}^{k} V_{1st} - \tfrac{1}{2}V_{01st}\right) \\ MSE\left(\hat{F}_{REe}(y)\right) &= F^2(y)\left(V_{0st} + \tfrac{1}{4}V_{1st} - V_{01st}\right) \end{aligned} \tag{5}$$

6. Ref. [34] proposed a general class of estimators:

$$\hat{F}_{t_k}(y) = \left[t_1\hat{F}_{st}(x) + t_2\left(F(x) - \hat{F}_{st}(x)\right)\right]\left[\frac{a_{st}F(x) + b_{st}}{c_{st}\hat{F}_{st}(x) + d_{st}}\right]^{\alpha}\left[exp\left(\frac{F(x) - \hat{F}_{st}(x)}{F(x) + \hat{F}_{st}(x)}\right)\right]^{\beta}$$

Here, $t_1, t_2, \alpha, \beta$ are suitable fixed values, and $a_{st}, b_{st}, c_{st}, d_{st}$ are either functions of the known parameters of $x$ or fixed values. The bias and MSE are

$$\begin{aligned} Bias\left(\hat{F}_{t_k}(y)\right) = \ & F(y)[(t_1\varphi_{st} - 1) \\ & +\left\{\varphi_{st}\left(\left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)t_2 r + \left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)\tfrac{t_1}{2} + \left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)^2\tfrac{t_1}{2}\right)V_{1st} - \left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)t_1 V_{01st}\right\}\Big] \end{aligned}$$

where $\varphi_{st} = \left[\frac{a_{st}F(x)+b_{st}}{c_{st}F(x)+d_{st}}\right]^{\alpha}, \eta_{st} = \frac{c_{st}F(x)}{c_{st}F(x)+d_{st}}$ and $r = \frac{F(x)}{F(y)}$.

$$\begin{aligned} MSE\left(\ \hat{F}_{t_k}(y)\right) = \ & F^2(y)\Big[(t_1\ \varphi_{st} - 1)^2 \\ & +\varphi_{st}{}^2\left\{t_1^2 V_{0st} - \left(t_2 r + t_1\left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)\right)^2 V_{1st} - 2\left(t_1 t_2 r + t_1^2\left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)\right)V_{01st}\right\} \\ & +2\varphi_{st}((t_1\ \varphi_{st} - 1)\ \left\{\left(\left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)t_2 r + \left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)\tfrac{t_1}{2} + \left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)^2\tfrac{t_1}{2}\right)\ V_{1st} \right. \\ & \left. -\left(\tfrac{\beta}{2} + \alpha\eta_{st}\right)t_1\ V_{01st}\right\}] \end{aligned}$$

We can rewrite the above equation as

$$MSE\left(\hat{F}_{t_k}(y)\right) = F^2(y)\left[1 - \gamma_1 t_1 + \gamma_2 t_1^2 - 2\gamma_3 t_2 + \gamma_4 t_2^2 + 2\gamma_5 t_1 t_2\right] \tag{6}$$

Such that

$$\begin{aligned} \gamma_1 &= \varphi_{st}\left[2 + (V_{1st} - 2V_{01st})\left(\frac{\beta}{2} + \alpha\eta_{st}\right) + \left(\frac{\beta}{2} + \alpha\eta_{st}\right)^2 V_{1st}\right] \\ \gamma_2 &= \varphi_{st}\left[1 + V_{0st} + (V_{1st} - 4V_{01st})\left(\frac{\beta}{2} + \alpha\eta_{st}\right) + 2\left(\frac{\beta}{2} + \alpha\eta_{st}\right)^2 V_{1st}\right] \\ \gamma_3 &= r\varphi_{st}V_{1st}\left(\frac{\beta}{2} + \alpha\eta_{st}\right) \\ \gamma_4 &= \varphi_{st}{}^2 r^2 V_{1st} \\ \gamma_5 &= r\varphi_{st}{}^2 V_{1st}\left[2\left(\frac{\beta}{2} + \alpha\eta_{st}\right)V_{1st} - V_{01st}\right] \end{aligned}$$

We determine the optimal values by computing the derivatives of the MSE with respect to $t_1$ and $t_2$.

$t_1 = \frac{\gamma_1\gamma_4 - 2\gamma_3\gamma_5}{2\gamma_2\gamma_4 - 2\gamma_5^2}$ and $t_2 = \frac{2\gamma_2\gamma_3 - \gamma_1\gamma_5}{2\gamma_2\gamma_4 - 2\gamma_5^2}$.

### *3.2. Existing Estimators in Post-Stratification*

We have transformed the following stratified estimators into post-stratified estimators as follows:

1. The usual unbiased estimator of $F(y)$ is given as

$$\hat{F}_{(ps)}(y) = \frac{1}{n}\sum_{i=1}^{n} \Delta(Y_i \le y)$$

The variance of $\hat{F}_{SRS_{ps}}(y)$ is

$$\begin{aligned} ar(\hat{F}_{(ps)}(y)) &= \left[\tfrac{1}{n} - \tfrac{1}{N}\right]\sum_{h=1}^{K} W_h S_{yh}^2 - \tfrac{1}{n^2}\sum_{h=1}^{K}(1 - W_h)S_{yh}^2 \\ &= V_{0ps}F^2(y) - \tfrac{1}{n^2}\sum_{h=1}^{K}(1 - W_h)S_{yh}^2 \end{aligned} \tag{7}$$

2. The classical ratio estimator of $F(y)$, according to [1], is given by

$$\hat{F}_{Re(ps)}(y) = \hat{F}_{ps}(y)\left[\frac{F(x)}{\hat{F}_{ps}(x)}\right]$$

and its bias and MSE become

$$\begin{aligned} Bias\left(\hat{F}_{Re(ps)}(y)\right) &= F(y)\left(V_{1ps} - V_{01ps}\right) \\ MSE\left(\hat{F}_{Re(ps)}(y)\right) &= F^2(y)\left(V_{0ps} + V_{1ps} - 2V_{01ps}\right) \end{aligned} \tag{8}$$

3. The product estimator of $F(y)$ was proposed by [2]:

$$\hat{F}_{Pe(ps)}(y) = \hat{F}_{ps}(y)\left[\frac{\hat{F}_{ps}(x)}{F(x)}\right]$$

Its bias and MSE become approximately constant until the first order:

$$\begin{aligned} Bias\left(\hat{F}_{Pe(ps)}(y)\right) &= F(y)V_{01ps} \\ MSE\left(\hat{F}_{Pe(ps)}(y)\right) &= F^2(y)\left(V_{0ps} + V_{1ps} + 2V_{01ps}\right) \end{aligned} \tag{9}$$

4. A difference-type estimator was proposed by [4]:

$$\hat{F}_{De(ps)}(y) = m_1\hat{F}_{ps}(y) + m_2\left[F(x) - \hat{F}_{ps}(x)\right]$$

where $m_1$ and $m_2$ are unknown fixed values.

$$\begin{aligned} Bias\left(\hat{F}_{De(ps)}(y)\right) &= F(y)(m_1 - 1) \\ MSE\left(\hat{F}_{De(ps)}(y)\right) &= F^2(y) - 2m_1F^2(y) + m_1^2F^2(y) + m_1^2F^2(y)V_{0ps} - 2m_1m_2F(x)F(y)V_{01ps} + m_2^2F^2(x)V_{1ps} \end{aligned}$$

After minimizing $MSE\left(\hat{F}_{De(ps)}(y)\right)$, we obtain optimum values as

$m_1 = \frac{V_{1ps}}{V_{0ps}V_{1ps} + V_{0ps} - V_{01ps}^2}$ and $m_2 = \frac{F(y)V_{01ps}}{F(x)\left(V_{0ps}V_{1ps} + V_{0ps} - V_{01ps}^2\right)}$

Now, $MSE(\hat{F}_{De(ps)}(y))$. We can we rewrite it as

$$MSE_{min}(\hat{F}_{De(ps)}(y)) = \frac{\left(V_{0ps}V_{1ps} - V_{01ps}^2\right)}{V_{0ps}V_{1ps} + V_{0ps} - V_{01ps}^2} \tag{10}$$

5. A generalized ratio-type exponential estimator was adopted by [13]:

$$\hat{F}_{REe(ps)}(y) = \hat{F}_{ps}(y) exp\left(\frac{a_{ps}(F(x) - \hat{F}_{ps}(x))}{a_{ps}(F(x) + \hat{F}_{ps}(x)) + 2b_{ps}}\right)$$

Here, $a_{ps}$ and $b_{ps}$ are fixed values, and bias and MSE will be

$$\begin{gathered} ias(\hat{F}_{REe(ps)}(y)) = F(y)\left(\frac{3}{8}V_{1ps} - \frac{1}{2}V_{01ps}\right) \\ MSE\left(\hat{F}_{REe(ps)}(y)\right) = F^2(y)\left(V_{0ps} + \frac{1}{4}V_{1ps} - V_{01ps}\right) \end{gathered} \tag{11}$$

6. Ref. [34] proposed a general class of estimators given by

$$\hat{F}_{t_k(ps)}(y) = \left[t_{1ps}\hat{F}_{ps}(x) + t_{2ps}\left(F(x) - \hat{F}_{ps}(x)\right)\right]\left[\frac{a_{ps}F(x) + b_{ps}}{c_{ps}\hat{F}_{ps}(x) + d_{ps}}\right]^{\alpha}\left[exp\left(\frac{F(x) - \hat{F}_{ps}(x)}{F(x) + \hat{F}_{ps}(x)}\right)\right]^{\beta}$$

Here, $t_{1ps}, t_{2ps}, \alpha, \beta$ are suitable fixed values, and $a_{ps}, b_{ps}, c_{ps}, d_{ps}$ are either fixed values or functions of the known parameters of $x$. Bias and MSE are

$$\begin{aligned} Bias\left(\hat{F}_{t_k(ps)}(y)\right) & \\ &= F(y)\big[\left(t_{1ps}\varphi_{ps} - 1\right) \\ &+\left\{\varphi_{ps}\left(\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)t_{2ps}r + \left(\frac{\beta}{2} + \alpha\eta_{ps}\right)\frac{t_{1ps}}{2}\right.\right. \\ &\left.\left.+\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)^2\frac{t_{1ps}}{2}\right)V_{1ps} - \left(\frac{\beta}{2} + \alpha\eta_{ps}\right)t_{1ps}V_{01ps}\right\}\big] \end{aligned}$$

where $\varphi_{ps} = \left[\frac{a_{ps}F(x)+b_{ps}}{c_{ps}\hat{F}_{ps}(x)+d_{ps}}\right]^{\alpha}, \eta_{ps} = \frac{c_{ps}F(x)}{c_{ps}F(x)+d_{ps}}$ and $r = \frac{F(x)}{F(y)}$.

$$\begin{aligned} &MSE\left(\hat{F}_{t_k(ps)}(y)\right) \\ &\quad= F^2(y)\left[\left(t_{1ps}\,\varphi_{ps} - 1\right)^2 + \varphi_{ps}^2\left\{t_{1ps}^2V_{0ps} - \left(t_{2ps}r + t_{1ps}\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)\right)^2V_{1ps}\right.\right. \\ &\quad\left.-2\left(t_{1ps}t_{2ps}r + t_{1ps}^2\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)\right)V_{01ps}\right\} + 2\varphi_{ps}\left(\left(t_{1ps}\,\varphi_{ps}\right.\right. \\ &\quad\left.-1\right)\left\{\left(\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)t_{2ps}r + \left(\frac{\beta}{2} + \alpha\eta_{ps}\right)\frac{t_{1ps}}{2} + \left(\frac{\beta}{2} + \alpha\eta_{ps}\right)^2\frac{t_{1ps}}{2}\right)V_{1ps}\right. \\ &\quad\left.\left.-\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)t_{1ps}\,V_{01ps}\right\}\right] \end{aligned}$$

We can rewrite the above equation as

$$MSE\left(\hat{F}_{t_k(ps)}(y)\right) = F^2(y)\left[1 - \gamma_{1ps}t_{1ps} + \gamma_{2ps}t_{1ps}^2 - 2\gamma_{3ps}t_{2ps} + \gamma_{4ps}t_{2ps}^2 + 2\gamma_{5ps}t_{1ps}t_{2ps}\right] \tag{12}$$

Such that

$$\gamma_{1ps} = \varphi_{ps}\left[2 + (V_{1ps} - 2V_{01ps})\left(\frac{\beta}{2} + \alpha\eta_{ps}\right) + \left(\frac{\beta}{2} + \alpha\eta_{ps}\right)^2 V_{1ps}\right]$$

$$\gamma_{2ps} = \varphi_{ps}\left[1 + V_{0ps} + (V_{1ps} - 4V_{01ps})\left(\frac{\beta}{2} + \alpha\eta_{ps}\right) + 2\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)^2 V_{1ps}\right]$$

$$\gamma_{3ps} = r\varphi_{ps}V_{1ps}\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)$$

$$\gamma_{4ps} = \varphi_{ps}^2 r^2 V_{1ps}$$

$$\gamma_{5ps} = r\varphi_{ps}^2 V_{1ps}\left[2\left(\frac{\beta}{2} + \alpha\eta_{ps}\right)V_{1ps} - V_{01ps}\right]$$

We obtain the optimal values of $t_{1ps}$ and $t_{2ps}$ by differentiating $MSE\left(\hat{F}_{t_k(ps)}(y)\right)$ with respect to $t_{1ps}$ and $t_{2ps}$.

$t_{1ps} = \frac{\gamma_{1ps}\gamma_{4ps} - 2\gamma_{3ps}\gamma_{5ps}}{2\gamma_{2ps}\gamma_{4ps} - 2\gamma_{5ps}^2}$ and $t_{2ps} = \frac{2\gamma_{2ps}\gamma_{3ps} - \gamma_{1ps}\gamma_{5ps}}{2\gamma_{2ps}\gamma_{4ps} - 2\gamma_{5ps}^2}$.

## 4. Proposed Estimators

### *4.1. Proposed Estimator in Stratified Random Sampling*

Inspired by [34], we have proposed a compound of difference, ratio, product, and exponential type of estimator to evaluate the population cumulative distribution function of the study variable as

$$\hat{F}_{stp}(y) = \left[n_1\hat{F}_{st}(y) + n_2\left(F(x) - \hat{F}_{st}(x)\right)\right]\left[\frac{a_{st}F(x) + b_{st}}{c_{st}\hat{F}_{st}(x) + d_{st}}\right]^{\alpha}\left[\frac{a_{st}F(x) + b_{st}}{c_{st}\hat{F}_{st}(x) + d_{st}}\right]^{-\gamma} exp\left[\frac{F(x) - \hat{F}_{st}(x)}{F(x) + \hat{F}_{st}(x)}\right]^{\beta} \quad (13)$$

where $n_1, n_2, \alpha,\ \beta, \gamma$ are suitable constants and $a_{st}, b_{st}, c_{st}, d_{st}$ denote the functions or constants of known parameters of auxiliary variable $x$.

We have considered six estimators from the literature. By substituting suitable values of $n_1,\ n_2,\ a_{st},\ b_{st},\ c_{st},\ d_{st},\ \alpha,\ \gamma,$ and $\beta$ in our proposed estimator, i.e., (13), we obtained the above-mentioned estimators and represented them in Table 1 as follows:

**Table 1.** Several recognized estimators from the proposed class.

| S. No | $n_1$ | $n_2$ | $a_{st}$ | $b_{st}$ | $c_{st}$ | $d_{st}$ | $\alpha$ | $\gamma$ | $\beta$ | Converge Estimator |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\hat{F}_{SRS_{st}}(y)$ |
| 2. | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | $\hat{F}_{Re}(y)$ |
| 3. | 1 | 0 | 1 | 0 | 1 | 0 | -1 | 0 | 0 | $\hat{F}_{Pe}(y)$ |
| 4. | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | $\hat{F}_{Pe}(y)$ |
| 5. | $n_1$ | $n_2$ | - | - | - | - | 0 | 0 | 0 | $\hat{F}_{De}(y)$ |
| 6. | 1 | 0 | - | - | - | - | 0 | 0 | 1 | $\hat{F}_{REe}(y)$ |
| 7. | $n_1$ | $n_2$ | $a_{st}$ | $b_{st}$ | $c_{st}$ | $d_{st}$ | $\alpha$ | 0 | $\beta$ | $\hat{F}_{t_k}(y)$ |

Bias and MSE of Proposed Estimator $\hat{\mathrm{F}}_{\mathrm{stp}}(\mathrm{y})$

By converting Equation (13) as $e_i^{\prime s}(i = x,\ y)$, we have

$$\hat{F}_{stp}(y) = \left[n_1F(y) + n_1F(y)e_y - n_2F(x)e_x\right]\left[\varphi_{st}(1 + \eta_{st}e_x)^{-\alpha}\right]\left[\varphi_{st1}(1 + \eta_{st}e_x)^{\gamma}\right]\left[1 - \frac{\beta}{2}e_x + \frac{(2+\beta)\beta}{8}e_x^2\right]$$

where $\varphi_{st} = \left(\frac{a_{st}F(x)+b_{st}}{c_{st}F(x)+d_{st}}\right)^{\alpha}$, $\varphi_{st1} = \left(\frac{a_{st}F(x)+b_{st}}{c_{st}F(x)+d_{st}}\right)^{-\gamma}$, $\eta_{st} = \frac{c_{st}F(x)}{c_{st}F(x)+d_{st}}$ and $r = \frac{F(x)}{F(y)}$

$$\begin{aligned}\hat{F}_{stp}(y) - F(y) = \;& \varphi_{st}\varphi_{st1}F(y)\left[n_1 + n_1 e_y - 1\right] \\ & +\varphi_{st}\varphi_{st1}e_x\left\{n_1\eta_{st}\left(\eta_{st}\frac{\alpha(\alpha+1)}{2} - \alpha + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right) - rn_2\right\} \\ & +\varphi_{st}\varphi_{st1}e_x^2\left\{n_1\eta_{st}\left(-\alpha^2\gamma - \frac{\alpha\gamma(\gamma+1)}{2}\eta_{st} + \frac{\alpha(\alpha+1)}{2}\frac{\gamma(\gamma+1)}{2}\eta_{st}^3 + \frac{\alpha\gamma(\gamma+1)}{2}\eta_{st}^2 + \frac{\alpha\beta}{2}\right.\right. \\ & \left. -\frac{\beta}{2}\frac{\alpha(\alpha+1)}{2}\eta_{st} - \frac{\beta\gamma}{2} - \frac{\beta}{2}\frac{\gamma(\gamma+1)}{2}\eta_{st}\right) \\ & \left. +n_2\eta_{st}\left(r\alpha - \frac{\gamma(\gamma+1)}{2}\eta_{st} - r\gamma - r\frac{\gamma(\gamma+1)}{2}\eta_{st} - r\frac{\beta}{2\gamma\eta_{st}}\right)\right\} \\ & +\varphi_{st}\varphi_{st1}e_x e_y\left\{n_1\eta_{st}\left(-\alpha + \frac{\alpha(\alpha+1)}{2}\eta_{st} + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right\}\end{aligned} \tag{14}$$

By considering expectations on each side of (14), we acquire the proposed estimator's bias:

$$\begin{aligned}Bias\left(\hat{F}_{stp}(y)\right) = \;& \varphi_{st}\varphi_{st1}F(y)[n_1 - 1] \\ & +\varphi_{st}\varphi_{st1}V_{1st}\left\{n_1\eta_{st}\left(-\alpha^2\gamma - \frac{\alpha\gamma(\gamma+1)}{2}\eta_{st} + \frac{\alpha(\alpha+1)}{2}\frac{\gamma(\gamma+1)}{2}\eta_{st}^3 + \frac{\alpha\gamma(\gamma+1)}{2}\eta_{st}^2 + \frac{\alpha\beta}{2}\right.\right. \\ & \left. -\frac{\beta}{2}\frac{\alpha(\alpha+1)}{2}\eta_{st} - \frac{\beta\gamma}{2} - \frac{\beta}{2}\frac{\gamma(\gamma+1)}{2}\eta_{st}\right) \\ & \left. +n_2\eta_{st}\left(r\alpha - \frac{\gamma(\gamma+1)}{2}\eta_{st} - r\gamma - r\frac{\gamma(\gamma+1)}{2}\eta_{st} - r\frac{\beta}{2\gamma\eta_{st}}\right)\right\} \\ & +\varphi_{st}\varphi_{st1}V_{01st}\left\{n_1\eta_{st}\left(-\alpha + \frac{\alpha(\alpha+1)}{2}\eta_{st} + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right\}\end{aligned}$$

Squaring on both sides of (14) and eliminating higher powers of $e_i^{\prime s}$, we acquire

$$\begin{aligned}&\left[\hat{F}_{stp}(y) - F(y)\right]^2 \\ &\quad = \varphi_{st}^2\varphi_{st1}^2F^2(y)\left[(n_1 - 1)^2 + n_1^2 e_y^2\right] \\ &\quad +{\varphi_{st}}^2{\varphi_{st1}}^2 e_x^2\left\{n_1\eta_{st}\left(\frac{\alpha(\alpha+1)}{2} - \alpha + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right) - rn_2\right\}^2 \\ &\quad +{\varphi_{st}}^2{\varphi_{st1}}^2\left(e_x e_y\right)^2\left\{n_1\eta_{st}\left(-\alpha + \frac{\alpha(\alpha+1)}{2}\eta_{st} + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right\}^2\end{aligned} \tag{15}$$

MSE is obtained by considering expectations on both sides of (15):

$$\begin{aligned}MSE\left(\hat{F}_{stp}(y)\right) = \;& \varphi_{st}^2\varphi_{st1}^2F^2(y)\{1 \\ & +n_1^2\left(1 + V_{0st} + \eta_{st}^2 V_{1st}\left(\left(\eta_{\,st}\frac{\alpha(\alpha+1)}{2} - \alpha\right) + \left(\gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right)^2\right. \\ & \left. +2V_{01st}\left(\left(\eta_{st}\frac{\alpha(\alpha+1)}{2} - \alpha\right) + \left(\gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right)^2\right) - 2n_1 + n_2^2 r^2 \\ & \left. -2rn_1n_2\eta_{st}\left(\eta_{st}\frac{\alpha(\alpha+1)}{2} - \alpha + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right\}\end{aligned}$$

$$MSE\left(\hat{F}_{stp}(y)\right) = \varphi_{st}^2\varphi_{st1}^2F^2(y)\left[1 + l_1 n_1^2 - 2n_1 + n_2^2 r^2 - 2n_1 n_2 l_2\right] \tag{16}$$

where

$$\begin{aligned}l_1 &= 1 + V_{0st} + \eta_{st}^2 V_{1st}\left(\left(\eta_{st}\frac{\alpha(\alpha+1)}{2} - \alpha\right) + \left(\gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right)^2 \\ &\quad +2V_{01st}\left(\left(\eta_{st}\frac{\alpha(\alpha+1)}{2} - \alpha\right) + \left(\gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\right)^2 \\ l_2 &= r\eta_{st}\left(\eta_{st}\frac{\alpha(\alpha+1)}{2} - \alpha + \gamma + \frac{\gamma(\gamma+1)}{2}\eta_{st} - \frac{\beta}{2\eta_{st}}\right)\end{aligned}$$

We obtained optimum values by differentiating (16) separately with respect to $n_1$ and $n_2$ to equate them with zero. We obtain
$n_1 = \frac{r^2}{r^2 l_1 - l_2^2}$ and $n_2 = \frac{l_2}{r^2 l_1 + l_2^2}$.

### 4.2. Proposed Estimator in Post-Stratification

We have proposed a post-stratified estimator by taking the combination of different types of estimators. The proposed post-stratified estimator is

$$\hat{F}_{psp}(y) = \hat{F}_{ps}(y) + \left[t_3\left(F(x) - \hat{F}_{ps}(x)\right) + t_4\chi\left(\overline{X} - \overline{x}_{ps}\right)\right] exp\left[\frac{F(x) - \hat{F}_{ps}(x)}{F(x) + \hat{F}_{ps}(x)}\right]^{\psi} \tag{17}$$

Here, $\hat{F}_{psp}(y)$ denotes the post-stratified proposed estimator of the distribution function of $y$. Here, $t_3, t_4, \chi$ and $\psi$ are suitable constants.

Expressing Equation (17) in terms of $e_i^{\prime s}$, we have

$$\hat{F}_{psp}(y) = F(y)(1+e_0) + \left[t_3(F(x) - F(x)(1+e_1)) + t_4\chi\left(\overline{X} - \overline{X}(1+e_2)\right)\right] exp\left[\tfrac{F(x)-F(x)(1+e_1)}{F(x)+F(x)(1+e_1)}\right]$$

$$\hat{F}_{psp}(y) = F(y) + F(y)e_0 - t_3F(x)e_1 - t_4\chi\overline{X}e_2 + t_3F(X)\tfrac{\psi}{2}e_1^2 + \overline{X}t_4\chi\psi\tfrac{e_1e_2}{2}$$

where $F(y) = \sum_{h=1}^{K} W_hF_h(y)$, $F(x) = \sum_{h=1}^{K} W_hF_h(x)$, $\hat{F}_{ps}(y) = \sum_{h=1}^{K} W_h\hat{F}_h(y)$, $\hat{F}_{ps}(x) = \sum_{h=1}^{K} W_h\hat{F}_h(x)$, $\hat{F}_h(y) = F_h(y)(1+e_{0h})$ and $\hat{F}_h(x) = F_h(x)(1+e_{1h})$.

Bias and MSE of Proposed Estimator $\hat{\mathrm{F}}_{\mathrm{psp}}(\mathrm{y})$

$$\hat{F}_{psp}(y) - F(y) = F(y)e_0 - t_3F(x)e_1 - t_4\chi\overline{X}e_2 + t_3F(x)\frac{\psi}{2}e_1^2 + \overline{X}t_4\chi\psi\frac{e_1e_2}{2} \tag{18}$$

Therefore, the bias of the proposed post-stratified estimator is obtained by applying expectation on both sides of (18):

$$Bias\left(\hat{F}_{psp}(y)\right) = t_3F(x)\frac{\psi}{2}V_{1ps} + \overline{X}t_4\chi\psi\frac{V_{12ps}}{2}$$

Squaring on both sides of (18) and eliminating higher powers of $e_i^{\prime s}$, we acquire

$$\begin{aligned}\left[\hat{\mathrm{F}}_{\mathrm{psp}}(\mathrm{y}) - \mathrm{F}(\mathrm{y})\right]^2 &\\ &= \mathrm{F}^2(\mathrm{y})e_0^2 + \mathrm{F}^2(\mathrm{x})t_3^2e_1^2 + \overline{\mathrm{X}}^2t_4^2\chi^2e_2^2 - 2F(x)F(y)t_3e_0e_1 \\ &+2F(x)t_3t_4\overline{X}\chi e_1e_2 - 2F(y)t_4\overline{X}\chi e_0e_2\end{aligned} \tag{19}$$

By considering expectations on each side of (19), the MSE of our proposed post-stratified estimator will be

$$\begin{aligned}MSE\left(\hat{F}_{psp}(y)\right) = &\ \mathrm{F}^2(\mathrm{y})V_{0ps} + \mathrm{F}^2(\mathrm{x})t_3^2V_{1ps} + \overline{\mathrm{X}}^2t_4^2\chi^2V_{2ps} - 2F(x)F(y)t_3V_{01ps} \\ &+2F(x)t_3t_4\overline{X}\chi V_{12ps} - 2F(y)t_4\overline{X}\chi V_{02ps}\end{aligned} \tag{20}$$

We obtain values for $t_3$ and $t_4$ by applying the differentiation of Equation (20) separately with respect to $t_3$ and $t_4$ and equate them with zero. We obtain

$$t_3 = \frac{F(y)\left[V_{01ps}V_{2ps} - V_{12ps}V_{02ps}\right]}{F(x)\left[V_{1ps}V_{2ps} - V_{12ps}^2\right]}$$

and

$$t_4 = \frac{F(y)\left[V_{1ps}V_{02ps} - V_{01ps}V_{12ps}\right]}{\overline{X}\chi\left[V_{1ps}V_{2ps} - V_{12ps}^2\right]}$$

After substituting $t_3$ and $t_4$ in (20), we have

$$\begin{aligned} &MSE\big(\hat{F}_{psp}(y)\big) \\ &= F^2(y)\Big\{V_{0ps} + \left[\frac{V_{1ps}w_1^2+V_{2ps}w_2^2-2V_{01ps}w_1w_3+2V_{12ps}w_1w_2-2V_{02ps}w_2w_3}{w_3^2}\right]\Big\} \\ &\qquad = F^2(y)\big(V_{0ps}+\mathfrak{R}\big)(\text{say}) \end{aligned} \tag{21}$$

where $\mathfrak{R} = \left[\frac{V_{1ps}w_1^2+V_{2ps}w_2^2-2V_{01ps}w_1w_3+2V_{12ps}w_1w_2-2V_{02ps}w_2w_3}{w_3^2}\right]$

$$\begin{aligned} w_1 &= V_{01ps}V_{2ps} - V_{12ps}V_{02ps} \\ w_2 &= V_{1ps}V_{02ps} - V_{01ps}V_{12ps} \\ w_3 &= V_{1ps}V_{2ps} - V_{12ps}^2. \end{aligned}$$

## 5. Theoretical Framework

### *5.1. Efficiency Comparison of Existing Estimators and Proposed Estimator under St RS*

By comparing Equation (16) with Equations (1)–(6), we discover the following conditions.

$$\begin{aligned} &MSE\big(\hat{F}_{stp}(y)\big) < MSE\big(\hat{F}_{SRS_{st}}(y)\big) \\ &\varphi_{st}^2\varphi_{st1}^2\left[1+l_1n_1^2-2n_1+n_2^2r^2-2n_1n_2l_2\right] < V_{0st} \end{aligned} \tag{22}$$

$$\begin{aligned} &SE\big(\hat{F}_{stp}(y)\big) < MSE\big(\hat{F}_{SRS_{st}}(y)\big) \\ &\varphi_{st}^2\varphi_{st1}^2\left[1+l_1n_1^2-2n_1+n_2^2r^2-2n_1n_2l_2\right] < V_{0st} \end{aligned} \tag{23}$$

$$\begin{aligned} &MSE\hat{F}_{stp}(y) < MSE\big(\hat{F}_{Pe}(y)\big) \\ &\varphi_{st}^2\varphi_{st1}^2\left[1+l_1n_1^2-2n_1+n_2^2r^2-2n_1n_2l_2\right] < (V_{0st}+V_{1st}+2V_{01st}) \end{aligned} \tag{24}$$

$$\begin{aligned} &MSE\hat{F}_{stp}(y) < MSE\big(\hat{F}_{De}(y)\big) \\ &\varphi_{st}^2\varphi_{st1}^2\left[1+l_1n_1^2-2n_1+n_2^2r^2-2n_1n_2l_2\right] < \frac{\big(V_{0st}V_{1st}-(V_{01st})^2\big)}{V_{1st}V_{0st}-(V_{01st})^2+V_{0st}} \end{aligned} \tag{25}$$

$$\begin{aligned} &MSE\hat{F}_{stp}(y) < MSE\big(\hat{F}_{REe}(y)\big) \\ &\varphi_{st}^2\varphi_{st1}^2\left[1+l_1n_1^2-2n_1+n_2^2r^2-2n_1n_2l_2\right] < \Big(V_{0st}+\tfrac{1}{4}V_{1st}-V_{01st} \end{aligned} \tag{26}$$

$$\begin{aligned} &MSE\hat{F}_{stp}(y) < MSE\big(\hat{F}_{t_k}(y)\big) \\ &\varphi_{st}^2\varphi_{st1}^2\left[1+l_1n_1^2-2n_1+n_2^2r^2-2n_1n_2l_2\right] < \left[1-\gamma_1t_1+\gamma_2t_1^2-2\gamma_3t_2+\gamma_4t_2^2+2\gamma_5t_1t_2\right] \end{aligned} \tag{27}$$

### *5.2. Theoretical Conditions under Post-Stratification*

The following conditions are derived when comparing the MSE of the proposed estimator with the MSEs of other considered existing estimators.

From Equations (21) and (7), we have

$$\begin{aligned} &MSE\big(\hat{F}_{psp}(y)\big) - Var\Big(\hat{F}_{(ps)}(y)\Big) < 0 \\ &\frac{F^2(y)\big(V_{0ps}+\mathfrak{R}\big)}{V_{0ps}F^2(y)-\frac{1}{n^2}\sum_{h=1}^{K}(1-W_h)S_{yh}^2} < 1 \end{aligned} \tag{28}$$

From Equations (21) and (8),

$$\begin{aligned} &MSE\big(\hat{F}_{psp}(y)\big) - MSE\Big(\hat{F}_{Re(ps)}(y)\Big) < 0 \\ &\frac{\big(V_{0ps}+\mathfrak{R}\big)}{\sum_{h=1}^{k} w_h^2\lambda_h\Big(C_{yh}^2+C_{xh}^2-2R_{xyh}C_{xh}C_{yh}\Big)} < 1 \end{aligned} \tag{29}$$

From Equations (21) and (9),

$$MSE\left(\hat{F}_{psp}(y)\right) - MSE\left(\hat{F}_{Pe(ps)}(y)\right) < 0$$
$$\frac{\left(V_{0ps}+\mathfrak{R}\right)}{\sum_{h=1}^{k} w_h^2 \lambda_h \left(C_{yh}^2+C_{xh}^2+2R_{xyh}C_{xh}C_{yh}\right)} < 1 \tag{30}$$

From Equations (21) and (10),

$$SE\left(\hat{F}_{psp}(y)\right) - MSE_{min}\left(\hat{F}_{De(ps)}(y)\right) < 0$$
$$\frac{\left(V_{0ps}+\mathfrak{R}\right)}{\frac{\left(V_{0ps}V_{1ps}-V_{01ps}^2\right)}{V_{0ps}V_{1ps}+V_{0ps}-V_{01ps}^2}} < 1 \tag{31}$$

From Equations (21) and (11),

$$SE\left(\hat{F}_{psp}(y)\right) - MSE\left(\hat{F}_{Ree(ps)}(y)\right) < 0$$
$$\frac{\left(V_{0ps}+\mathfrak{R}\right)}{\left(V_{0ps}+\frac{1}{4}V_{1ps}-V_{01ps}\right)} < 1 \tag{32}$$

From Equations (21) and (12),

$$MSE\left(\hat{F}_{psp}(y)\right) - MSE\left(\hat{F}_{t_k(ps)}(y)\right) < 0$$
$$\frac{\left(V_{0ps}+\mathfrak{R}\right)}{\left[1-\gamma_{1ps}t_{1ps}+\gamma_{2ps}t_{1ps}^2-2\gamma_{3ps}t_{2ps}+\gamma_{4ps}t_{2ps}^2+2\gamma_{5ps}t_{1ps}t_{2ps}\right]} < 1 \tag{33}$$

## 6. Empirical Studies

### *6.1. Empirical Study in Stratified Random Sampling*

Data Set-I: We have used the data from [35] to evaluate the suggested estimator's relative effectiveness. The data consist of six strata. A sample of 180 observations is taken from a total of 923 observations. Table 2 presents estimations of the data as follows:

**Table 2.** Summary statistics for the Data Set-I.

| $h$ | $N_h$ | $n_h$ | $W_h$ | $\lambda_h$ | $F_h(y)$ | $F_h(x)$ | $C_{yh}$ | $C_{xh}$ | $R_{xyh}$ | $S_{yxh}$ | $F(y)$ | $F(x)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 127 | 31 | 0.1375 | 0.0244 | 0.3543 | 0.3779 | 0.197955 | 0.21084 | 4.0597 | 0.0103 | 0.0487 | 0.0520 |
| 2 | 117 | 21 | 0.1267 | 0.039 | 0.4188 | 0.4872 | 0.575675 | 0.669312 | 10.8358 | 0.0356 | 0.0531 | 0.0617 |
| 3 | 103 | 29 | 0.1115 | 0.0248 | 0.4272 | 0.466 | 0.275157 | 0.300381 | 5.7725 | 0.0143 | 0.0476 | 0.0520 |
| 4 | 170 | 38 | 0.1841 | 0.0204 | 0.5765 | 0.6118 | 0.19553 | 0.20751 | 1.8412 | 0.0220 | 0.1061 | 0.1126 |
| 5 | 205 | 22 | 0.2221 | 0.0406 | 0.6146 | 0.6537 | 0.663306 | 0.705382 | 4.8587 | 0.0963 | 0.1365 | 0.1452 |
| 6 | 201 | 39 | 0.2177 | 0.0207 | 0.5025 | 0.3532 | 0.154403 | 0.108518 | 1.4113 | 0.0119 | 0.1094 | 0.0769 |

The functions of auxiliary variables are

$$\sum_{h=1}^{k} W_h S_{hx} = 0.486847, \sum_{h=1}^{k} W_h C_{hx} = 6.2793, \sum_{h=1}^{k} W_h R_{hxy} = 0.852239.$$

By using the above functions of known auxiliary variables, we have calculated the MSE and percentage relative efficiency (PRE) of the estimators in both Tables 3 and 4.

**Table 3.** A comparison of the MSEs and PREs of considered pre-existing estimators and our proposed estimator.

| S. No | Estimator | MSE | PRE |
|---|---|---|---|
| 1. | $\hat{F}_{SRS_{st}}(y)$ | 0.0488 | 100 |
| 2. | $\hat{F}_{Re}(y)$ | 0.0928 | 52.58 |
| 3. | $\hat{F}_{Pe}(y)$ | 0.0972 | 50.21 |
| 4. | $\hat{F}_{De}(y)$ | 0.0414 | 117.87 |
| 5. | $\hat{F}_{REe}(y)$ | 0.0593 | 82.30 |
| 6. | $\hat{F}_{t_k}(y)$ | 0.0223 | 218.83 |
| 7. | $\hat{F}_{stp}(y)$ | 0.0056 | 871.42 |

**Table 4.** MSEs of our proposed estimator.

| $a_{st}$ | $b_{st}$ | $c_{st}$ | $d_{st}$ | $\alpha$ | $\gamma$ | $\beta$ | Estimator | MSE | PRE |
|---|---|---|---|---|---|---|---|---|---|
| 0.8522 | 0.4868 | 0.8522 | 0.4868 | 0 | −1 | 0 | $\hat{F}_{stp1}(y)$ | 0.0092 | 530.43 |
| 6.2793 | 1 | 6.2793 | 1 | −1 | 0 | 0 | $\hat{F}_{stp2}(y)$ | 0.0085 | 574.12 |
| 6.2793 | 1 | 6.2793 | 1 | 0 | −1 | 0 | $\hat{F}_{stp3}(y)$ | 0.0085 | 574.12 |
| 1 | 0.4868 | 1 | 0.4868 | 1 | 0 | 1 | $\hat{F}_{stp4}(y)$ | 0.0055 | 887.27 |
| 0.8522 | 0.4868 | 0.8522 | 0.4868 | 1 | 0 | 1 | $\hat{F}_{stp5}(y)$ | 0.0053 | 920.75 |
| 1 | 6.2793 | 1 | 6.2793 | −1 | −1 | −1 | $\hat{F}_{stp6}(y)$ | 0.0052 | 938.46 |
| 0.8522 | 0.4868 | 0.8522 | 0.4868 | 0 | 1 | 1 | $\hat{F}_{stp7}(y)$ | 0.0052 | 938.46 |
| 1 | 6.2793 | 1 | 6.2793 | 0 | 0 | −1 | $\hat{F}_{stp8}(y)$ | 0.0052 | 938.46 |
| 0.4868 | 0.8522 | 0.4868 | 0.8522 | −1 | 1 | −1 | $\hat{F}_{stp9}(y)$ | 0.0052 | 938.46 |
| 1 | 0.4868 | 1 | 0.4868 | 0 | −1 | 0 | $\hat{F}_{stp10}(y)$ | 0.0044 | 1109.09 |
| 0.8522 | 1 | 0.8522 | 1 | 1 | −1 | 0 | $\hat{F}_{stp11}(y)$ | 0.0042 | 1161.90 |
| 0.4868 | 1 | 0.4868 | 1 | 1 | −2 | 0 | $\hat{F}_{stp12}(y)$ | 0.0039 | 1251.28 |
| 0.4868 | 6.2793 | 0.4868 | 6.2793 | 1 | 0 | 1 | $\hat{F}_{stp13}(y)$ | 0.0010 | 4800.00 |
| 1 | 0.8522 | 1 | 0.8522 | 1 | 0 | 1 | $\hat{F}_{stp14}(y)$ | 0.00088 | 5545.45 |
| 0.4868 | 6.2793 | 0.4868 | 6.2793 | 0 | −1 | 1 | $\hat{F}_{stp15}(y)$ | 0.00086 | 5674.42 |

By using the data in Table 2, we calculated MSE and PRE values for the pre-existing estimator and our proposed stratified estimator, and they are shown in Table 3 with suitable values for $\alpha = 1$, $\beta = 1$, $\gamma = 0$, and $a_{st} = c_{st} = 1$, $b_{st} = d_{st} = 0.4868$.

Additionally, by substituting different values in relevant variables in our suggested estimator, we obtained the following types of estimators, ratio, product, etc. The MSE and PRE values of some estimators of the proposed class of estimators are presented in Table 4.

Data Set-II

In this numerical investigation, we utilized the data [37] detailing student behaviors and exam performances. The dataset encompasses information for 500 students, ensuring a diverse range of study patterns and their exam performances. Here, we maintained symmetry in our sampling process by dividing the data into six strata, as presented in Table 5. We selected a sample of 120 by using the Neyman allocation method. We focused on exam scores as the study variable, reflecting the student's score in an exam, while study hours served as an auxiliary variable, indicating the number of hours a student dedicated to exam preparation. We aim to predict students' exam scores based on their study hours, thereby emphasizing the symmetry of representation across different strata in our analysis.

**Table 5.** Data statistics for Data Set-II.

| $h$ | $N_h$ | $n_h$ | $W_h$ | $\lambda_h$ | $F_h(y)$ | $F_h(x)$ | $S_{yh}$ | $S_{xh}$ | $S_{yxh}$ | $F(y)$ | $F(x)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 91 | 23 | 0.182 | 0.032 | 0.110 | 0.143 | 0.139 | 0.157 | 0.0103 | 0.0487 | 0.0520 |
| 2 | 82 | 19 | 0.164 | 0.040 | 0.110 | 0.146 | 0.132 | 0.151 | 0.0356 | 0.0531 | 0.0617 |
| 3 | 89 | 22 | 0.178 | 0.034 | 0.124 | 0.101 | 0.145 | 0.132 | 0.0143 | 0.0476 | 0.0520 |
| 4 | 89 | 21 | 0.178 | 0.036 | 0.112 | 0.112 | 0.139 | 0.139 | 0.0220 | 0.1061 | 0.1126 |
| 5 | 72 | 18 | 0.144 | 0.042 | 0.083 | 0.139 | 0.097 | 0.139 | 0.0963 | 0.1365 | 0.1452 |
| 6 | 77 | 18 | 0.154 | 0.043 | 0.117 | 0.117 | 0.132 | 0.132 | 0.0119 | 0.1094 | 0.0769 |

The functions of auxiliary variables are

$$\sum_{h=1}^{k} W_h^2 S_{hx} = 0.023, \sum_{h=1}^{k} W_h^2 C_{hx} = 1.126.$$

Based on the data in Table 6, it is evident that the MSE of our proposed estimator is lower compared to all other existing estimators. Additionally, the PRE of our proposed estimator is notably high in comparison. This suggests that our proposed estimator demonstrates superior precision when compared to other estimators.

**Table 6.** MSE and PRE values of the estimators.

| S. No | Estimator | MSE | PRE |
|---|---|---|---|
| 1. | $\hat{F}_{SRS_{st}}(y)$ | 0.000110 | 100.000000 |
| 2. | $\hat{F}_{Re}(y)$ | 0.000073 | 151.118987 |
| 3. | $\hat{F}_{Pe}(y)$ | 0.000341 | 32.373953 |
| 4. | $\hat{F}_{De}(y)$ | 0.000249 | 44.335351 |
| 5. | $\hat{F}_{REe}(y)$ | 0.000068 | 163.389178 |
| 6. | $\hat{F}_{t_k}(y)$ | 0.000085 | 129.674930 |
| 7. | $\hat{F}_{stp}(y)$ | 0.000057 | 194.876102 |

The line graphs illustrating the PRE results for Data Set I and Data Set II, obtained from Tables 3 and 6, are displayed in Figures 1 and 2, respectively.

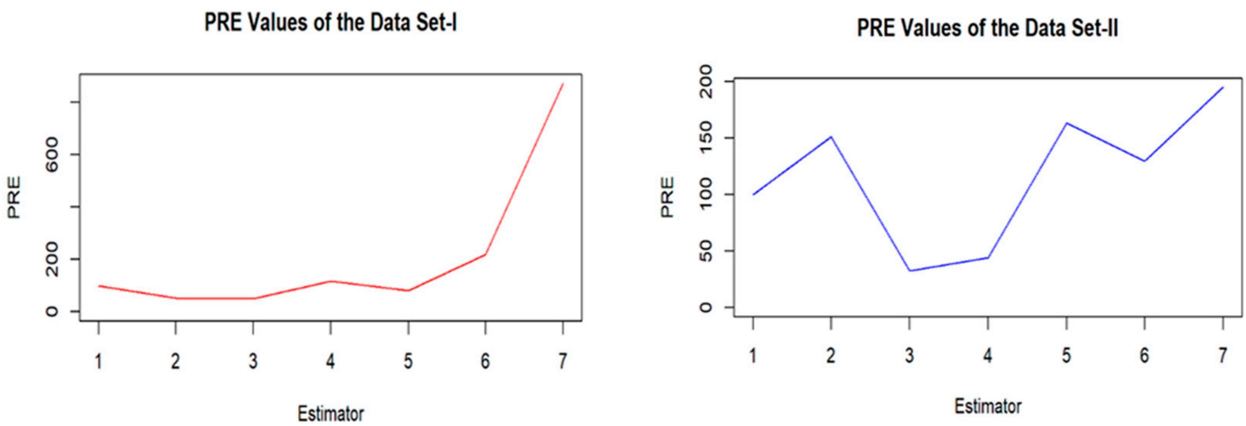


**Figure 1.** PRE values of the estimators for Data Sets I and II.

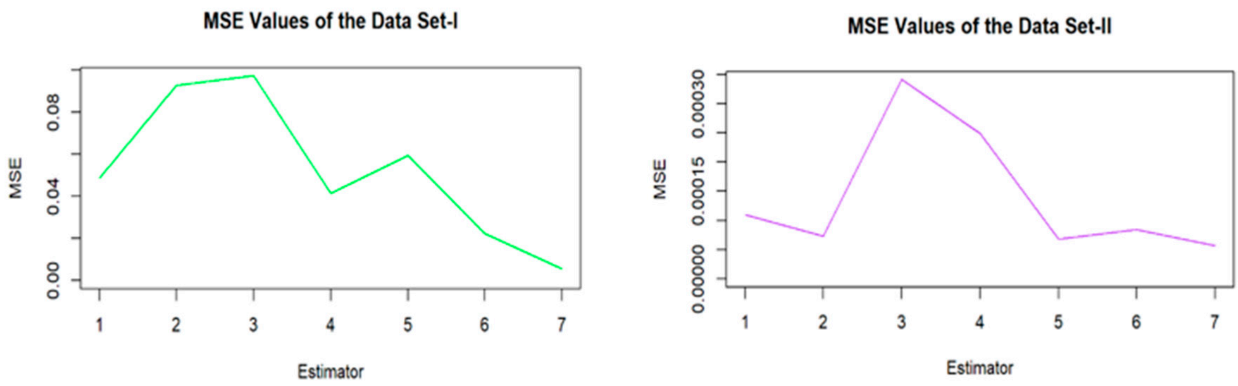


**Figure 2.** MSE values of the estimators for Data Sets I and II.

### 6.2. Empirical Validation under Post-Stratification

Data Set-III: Source [38].

In Data Set, $Y$ represents the apple production amount in 1999, and $X$ represents the number of apple trees in 1999. The data statistics are available in Table 7. We consider $Y$ as the study variable and $X$ as the auxiliary variable.

**Table 7.** Data statistics for Data Set-III.

| $h$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $N_h$ | 106 | 106 | 94 | 171 | 204 | 173 |
| $n_h$ | 9 | 17 | 38 | 67 | 7 | 2 |
| $W_h$ | 0.1241 | 0.1241 | 0.1101 | 0.2002 | 0.2389 | 0.2026 |
| $\lambda_h$ | 0.1017 | 0.0494 | 0.0157 | 0.0091 | 0.138 | 0.4942 |
| $F_h(y)$ | 0.5872 | 0.5189 | 0.3298 | 0.3684 | 0.4657 | 0.7052 |
| $F_h(x)$ | 0.5472 | 0.566 | 0.3404 | 0.3801 | 0.4657 | 0.7225 |
| $\overline{X}_h$ | 24376 | 27422 | 72410 | 74365 | 26442 | 9844 |
| $S_{yh}$ | 0.495 | 0.502 | 0.4727 | 0.4838 | 0.5 | 0.4573 |
| $S_{xh}$ | 0.5001 | 0.4979 | 0.4764 | 0.4868 | 0.4965 | 0.4490 |
| $S_{\overline{x}h}$ | 49189 | 5746 | 160757 | 285603 | 45403 | 18794 |
| $R_{xyh}$ | 0.7722 | 0.8330 | 0.7854 | 0.7755 | 0.6750 | 0.7319 |
| $R_{y\overline{x}h}$ | −0.4470 | −0.4370 | −0.2957 | −0.1848 | −0.3929 | −0.5598 |
| $R_{x\overline{x}h}$ | −0.4523 | −0.4816 | −0.3087 | −0.1936 | −0.4129 | −0.6102 |

Utilizing the statistical data provided in Table 7, we computed the MSE and PRE values, which are summarized in Table 8. This table allows us to assess the effectiveness of the proposed estimators compared to others.

**Table 8.** MSE and PRE values of the estimators.

| S. No | Estimator | Data Set-II | | Data Set-III | |
|---|---|---|---|---|---|
| | | MSE | PRE | MSE | PRE |
| 1. | $\hat{F}_{(ps)}(y)$ | 0.037 | 100.000 | 0.759 | 100.000 |
| 2. | $\hat{F}_{Re(ps)}(y)$ | 0.018 | 205.556 | 0.383 | 198.018 |
| 3. | $\hat{F}_{Pe(ps)}(y)$ | 0.126 | 29.365 | 2.615 | 29.013 |
| 4. | $\hat{F}_{De(ps)}(y)$ | 0.016 | 231.250 | 0.341 | 222.647 |

**Table 8.** *Cont.*

| S. No | Estimator | Data Set-II | | Data Set-III | |
|---|---|---|---|---|---|
| | | MSE | PRE | MSE | PRE |
| 5. | $\hat{F}_{REe(ps)}(y)$ | 0.019 | 194.737 | 0.386 | 196.642 |
| 6. | $\hat{F}_{t_k(ps)}(y)$ | 0.018 | 205.556 | 0.370 | 204.822 |
| 7. | $\hat{F}_{psp}(y)$ | 0.016 | 231.250 | 0.331 | 229.412 |

In both Data Sets II and III, we can observe the efficiency of the proposed post-stratified estimator compared to other considered estimators in terms of MSE and PRE.

The line graphs depicting the PRE results for Data Set II and Data Set III, derived from Table 8, are showcased in Figures 3 and 4, correspondingly.

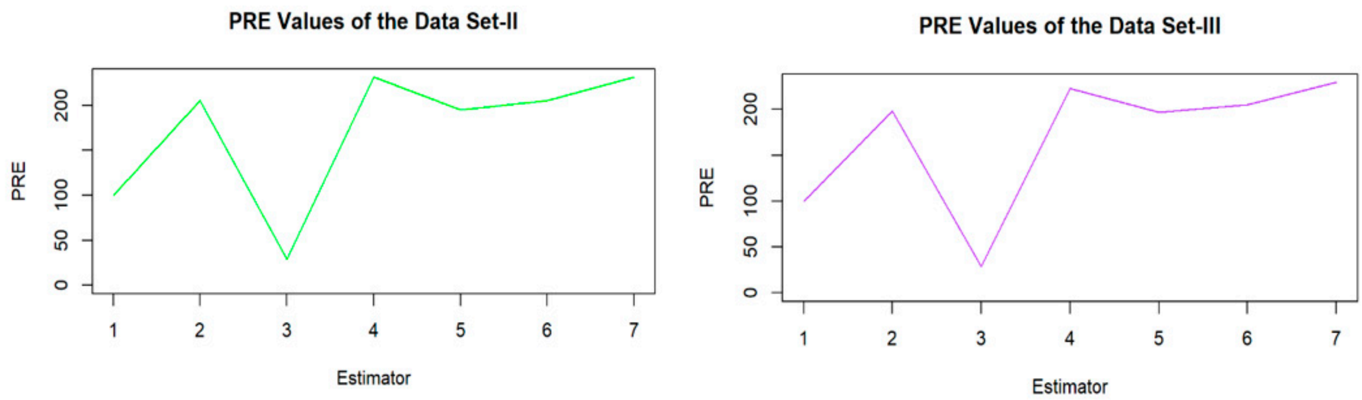


**Figure 3.** PRE values of the estimators for Data Sets II and III.

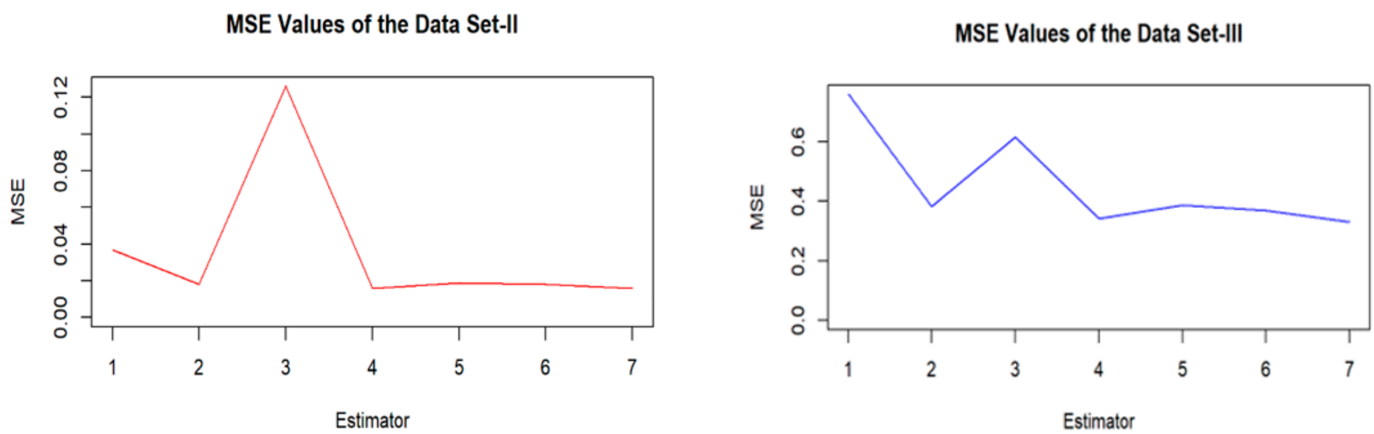


**Figure 4.** MSE values of the estimators for Data Sets II and III.

## 7. Results and Discussion

In this study, we have proposed two novel estimators to estimate the CDF of a study variable by employing the auxiliary variables' information under stratified random sampling and in post-stratification.

The first estimator is proposed under St RS, which contains a combination of estimators presented in Equation (13). By taking suitable constants in place of $n_1, n_2, \alpha, \beta, \gamma$ and functions of auxiliary variable $x$ or constants in places of $a_{st}, b_{st}, c_{st}, d_{st}$, we obtain efficient results for our proposed estimators. Because the proposed estimator contains a class of estimators, it has several existing estimators in it. Because we used suitable values in the functions or constants, we have different estimators, which are represented in Table 1. Here, two data sets were used to prove the efficiency of the proposed estimator. The derived conditions are available in Equations (22)–(27). Data Set-I is taken from [35], and the numerical study is presented in Section 6. From Table 3, we can observe the results; the proposed estimator $\hat{F}_{stp}(y)$ Outperform other estimators in terms of MSE and PRE. Data Set-II is extracted from the website https://doi.org/10.34740/KAGGLE/DSV/7623777 (accessed on 29 February 2024). Table 5 presents all the values needed for the calculation of MSEs. From Table 6, among the estimators, $\hat{F}_{stp}(y)$ stands out with its remarkably low MSE of 0.000057 and a high PRE of 194.876102, underscoring its superior predictive accuracy and demonstrating greater symmetry than usual unbiased [1,2,4,13,34] estimators.

The second estimator we proposed in this study is under post-stratification with constants $t_3, t_4, \chi$ and $\psi$. We derived the equations of bias and MSE up to the first degree of approximation, and can find the theoretical conditions in Section 5 from Equations (28)–(33). To prove the efficiency of the proposed estimator in post-stratification, we have utilized two data sets. We have taken the information of $Y$, $X$, and $\overline{X}$ from the Data Set-II and III. We can observe from Table 8, that the MSE value of the proposed estimator is low and the relative efficiency values are high compared with the considered estimators, which is the same as we can observe from the figures. From Table 8, the comparative analysis of various estimators applied to Data Set-II and Data Set-III reveals distinct performance characteristics. Notably, the estimator $\hat{F}_{psp}(y)$ consistently exhibits superior predictive accuracy, as evidenced by its low MSE value and consistently high PRE across both datasets. Conversely, $\hat{F}_{Pe(ps)}(y)$ demonstrates poor performance, with significantly higher MSE values and lower PRE, suggesting limited predictive capability. Among the estimators, $\hat{F}_{Re(ps)}(y)$, $\hat{F}_{De(ps)}(y)$ , and $\hat{F}_{REe(ps)}(y)$ present moderate performance, displaying relatively lower MSE and higher PRE compared to $\hat{F}_{Pe(ps)}(y)$ but not reaching the levels of $\hat{F}_{psp}(y)$.

From Figures 1 and 3, a striking trend emerges as the plotted trend line gracefully ascends, embodying our recommended estimator's trajectory. In contrast, Figure 2 reveals a consistent decline in MSE values, notably showcasing the diminishing errors of both [34] and our proposed estimator, labeled as 6 and 7, respectively. Figure 4 accentuates the nearly identical MSE values of the second and fourth estimators, hinting at commendable performance, albeit not surpassing the prowess demonstrated by our proposed estimator. Examining Figure 3, a clear victor emerges as the proposed estimator outshines its counterparts in both Dataset-II and Dataset-III, closely trailed by [4], in both datasets. Conversely, ref. [2] presents a lackluster performance across both datasets, marking it as the weakest contender. Figure 4 mirrors this pattern, with our proposed method boasting the lowest MSE followed closely by estimator [4] across both datasets. Notably, ref. [2] and the classical estimator struggle to keep pace, recording notably higher MSE values in Dataset-I and Dataset-II, respectively. Hence, the evidence from Figures 3 and 4 unequivocally supports the superiority of our proposed estimator over its counterparts, a conclusion further reinforced by the insights gleaned from Figures 1 and 2. Table 3 serves as a comprehensive showcase of MSE and PRE metrics for existing estimators juxtaposed with our proposed solution, listed as serial No. 7. Notably, our proposed estimator garners the lowest MSE and the highest PRE, setting a benchmark closely followed by [34]. Table 4 corroborates this finding, further establishing the pre-eminence of our proposed estimator. Additionally, Table 6 unveils the performance metrics for Dataset-II, highlighting once more the supremacy of our proposed method, trailed by the estimator [13]. This consistent dominance across datasets underscores the inconsistency plaguing existing estimators, a testament to the robustness and reliability of our proposed solution.

## 8. Conclusions

This study introduces two unique estimators that are precisely built to assess the limited population distribution function within the realms of stratified random sampling and post-stratification, ensuring symmetry in the sampling process. The study illustrates the outstanding efficiency of these estimators in comparison to conventional approaches across both sampling schemes by exploiting three unique datasets, including real-world data encompassing student behavior and exam results. Through complete empirical validation, the estimators routinely beat their counterparts in terms of both mean square error and percentage relative efficiency, demonstrating their ability to perform in practical settings. Furthermore, the study provides important insights into the predictive accuracy of educational assessments, as demonstrated by the successful prediction of students' exam scores based on study hours using the proposed estimator. This study not only introduces novel approaches to long-standing survey sample challenges but also reveals avenues for improving prediction accuracy in educational assessments. The convergence of theoretical derivations and empirical validations emphasizes the proposed estimators' resilience and versatility, ensuring symmetry in their potential use across a diverse range of sampling settings. The study of Figures 1–4 demonstrates the estimators' higher efficiency, establishing their place as pioneering contributions to the field of survey sampling. Additionally, fundamental ideas such as non-response analysis and calibration approaches are proposed to improve the resilience of estimators across different data sets and settings. In conclusion, the findings reflect a substantial advancement in the field of survey sampling, with major implications for future research efforts. As research into potential expansions and modifications of the estimators continues, there is a concerted attempt to improve their effectiveness and usability in practical contexts, maintaining the trend of innovation and advancement in survey sampling procedures.

**Author Contributions:** Methodology, G.R.V.T. and F.D.; Software, G.R.V.T. and F.D.; Formal analysis, O.A.; Data curation, F.D.; Writing—original draft, G.R.V.T.; Writing—review & editing, F.D. and O.A.; Visualization, F.D.; Supervision, F.D.; Funding acquisition, O.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** No primary data has been used in the article.

**Acknowledgments:** We highly appreciate the efforts of the reviewers and editors for their efforts towards the improvement of the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cochran, W.G. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *J. Agric. Sci.* **1940**, *30*, 262–275. [CrossRef]
2. Murthy, M.N. Product method of estimation. Sankhya. *Indian J. Stat. Ser. A* **1964**, *26*, 69–74.
3. Holt, D.; Smith, T.F. Post-stratification. *J. R. Stat. Soc. Ser. A (Gen.)* **1979**, *142*, 33–46. [CrossRef]
4. Rao, T.J. On certail methods of improving ratio and regression estimators. *Commun. Stat.-Theory Methods* **1991**, *20*, 3325–3340. [CrossRef]
5. Chambers, R.L.; Dorfman, A.H.; Hall, P. Properties of estimators of the finite population distribution function. *Biometrika* **1992**, *79*, 577–582. [CrossRef]
6. Little, R.J. Post-stratification: A modeler's perspective. *J. Am. Stat. Assoc.* **1993**, *88*, 1001–1012. [CrossRef]
7. Rao, J.N.K. Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Off. Stat.* **1994**, *10*, 153–165.
8. Silva, P.L.D.; Skinner, C.J. Estimating distribution functions with auxiliary information using poststratification. *J. Off. Stat.* **1995**, *11*, 277–294.
9. Wang, S.; Dorfman, A.H. A new estimator for the finite population distribution function. *Biometrika* **1996**, *83*, 639–652. [CrossRef]
10. Abu-Dayyeh, W.A.; Ahmed, M.S.; Ahmed, R.A.; Muttlak, H.A. Some estimators of a finite population mean using auxiliary information. *Appl. Math. Comput.* **2003**, *139*, 287–298. [CrossRef]
11. Kadilar, C.; Cingi, H. A new ratio estimator in stratified random sampling. *Commun. Stat. Theory Methods* **2005**, *34*, 597–602. [CrossRef]
12. Koyuncu, N.; Kadilar, C. Family of estimators of population mean using two auxiliary variables in stratified random sampling. *Commun. Stat. Theory Methods* **2009**, *38*, 2398–2417. [CrossRef]
13. Singh, R.; Chauhan, P.; Sawan, N.; Smarandache, F. Improvement in estimating the population mean using an exponential estimator in simple random sampling. *Int. J. Stat. Econ.* **2009**, *3*, 13–18.
14. Singh, H.P.; Solanki, R.S. Efficient ratio and product estimators in stratified random sampling. *Commun. Stat.-Theory Methods* **2013**, *42*, 1008–1023. [CrossRef]

15. Grover, L.K.; Kaur, P. A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. *Commun. Stat.-Simul. Comput.* **2014**, *43*, 1552–1574. [CrossRef]
16. Munoz, J.F.; Arcos, A.; Alvarez, E.; Rueda, M. New ratio and difference estimators of the finite population distribution function. *Math. Comput. Simul.* **2014**, *102*, 51–61. [CrossRef]
17. Tailor, R.; Tailor, R.; Chouhan, S. Improved ratio-and product-type exponential estimators for population mean in case of post-stratification. *Commun. Stat.-Theory Methods* **2017**, *46*, 10387–10393. [CrossRef]
18. Muneer, S.; Shabbir, J.; Khalil, A. Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. *Commun. Stat.-Theory Methods* **2017**, *46*, 2181–2192. [CrossRef]
19. Shabbir, J.; Gupta, S. Estimation of finite population mean in simple and stratified random sampling using two auxiliary variables. *Commun. Stat.-Theory Methods* **2017**, *46*, 10135–10148. [CrossRef]
20. Singh, G.N.; Khalid, M.; Sharma, A.K. Some efficient classes of estimators of population mean in two-phase successive sampling under random nonresponse. *Commun. Stat.-Theory Methods* **2017**, *46*, 12194–12209. [CrossRef]
21. Vishwakarma, G.K. An alternative to ratio estimator in post-stratification. *Commun. Stat.-Theory Methods* **2018**, *47*, 1989–2000. [CrossRef]
22. Gupta, R.K.; Yadav, S.K. Improved estimation of population mean using information on size of the sample. *Am. J. Math. Stat.* **2018**, *8*, 27–35.
23. Haq, A.; Shabbir, J. An improved class of estimators of finite population mean in simple random sampling using an auxiliary attribute. *J. Stat. Theory Pract.* **2018**, *12*, 282–289. [CrossRef]
24. Pal, S.K.; Singh, P. Estimation of finite population mean using auxiliary information in presence of non-response. *Commun. Stat.-Simul. Comput.* **2018**, *47*, 143–165. [CrossRef]
25. Zaman, T.; Kadilar, C. Novel family of exponential estimators using information of auxiliary attribute. *J. Stat. Manag. Syst.* **2019**, *22*, 1499–1509. [CrossRef]
26. Hussain, S.; Ahmad, S.; Saleem, M.; Akhtar, S. Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. *PLoS ONE* **2020**, *15*, e0239098. [CrossRef] [PubMed]
27. Singh, G.N.; Khalid, M. Some imputation methods to compensate with non-response for estimation of population mean in two-occasion successive sampling. *Commun. Stat.-Theory Methods* **2020**, *49*, 3329–3351. [CrossRef]
28. Ahmad, S.; Arslan, M.; Khan, A.; Shabbir, J. A generalized exponential-type estimator for population mean using auxiliary attributes. *PLoS ONE* **2021**, *16*, e0246947. [CrossRef] [PubMed]
29. Bhushan, S.; Kumar, A.; Singh, S. Some efficient classes of estimators under stratified sampling. *Commun. Stat.-Theory Methods* **2021**, *52*, 1767–1796. [CrossRef]
30. Mradula Yadav, S.K.; Varshney, R.; Dube, M. Efficient estimation of population mean under stratified random sampling with a linear cost function. *Commun. Stat.-Simul. Comput.* **2021**, *50*, 4364–4387. [CrossRef]
31. Ahmad, S.; Hussain, S.; Zahid, E.; Iftikhar, A.; Hussain, S.; Shabbir, J.; Aamir, M. A Simulation Study: Population Distribution Function Estimation Using Dual Auxiliary Information under Stratified Sampling Scheme. *Math. Probl. Eng.* **2022**, *2022*, 3263022. [CrossRef]
32. Triveni, G.R.V.; Danish, F. Heuristical Approach for Optimizing Population Mean Using Ratio Estimator in Stratified Random Sampling. *J. Reliab. Stat. Stud.* **2023**, *16*, 137–152. [CrossRef]
33. Zaman, T.; Bulut, H. An efficient family of robust-type estimators for the population variance in simple and stratified random sampling. *Commun. Stat.-Theory Methods* **2023**, *52*, 2610–2624. [CrossRef]
34. Tiwari, K.K.; Bhougal, S.; Kumar, S. A general class of estimators in stratified random sampling. *Commun. Stat.-Simul. Comput.* **2023**, *52*, 442–452. [CrossRef]
35. Triveni, G.R.V.; Danish, F. Exploring the dependability of Combined Ratio Estimators in Stratified Ranked Set Sampling: Insights from COVID-19 data. *Alex. Eng. J.* **2024**, *92*, 267–272. [CrossRef]
36. Kocyigit, E.G.; Kadilar, C. Information theory approach to ranked set sampling and new sub-ratio estimators. *Commun. Stat.-Theory Methods* **2024**, *53*, 1331–1353. [CrossRef]
37. MrSimple07. Student Exam Performance Prediction [Data Set]. Kaggle. 2024. Available online: https://www.kaggle.com/datasets/mrsimple07/student-exam-performance-prediction (accessed on 29 February 2024).
38. Kadilar, C.; Cingi, H. Ratio estimators in stratified random sampling. *Biom. J.* **2003**, *45*, 218–225. [CrossRef]