American Journal of Mechanics and Applications
2025, Vol. 12, No. 1, pp. 1-10
https://doi.org/10.11648/j.ajma.20251201.11

Research Article

Developing Robust Exponential Ratio and Product Estimators for Post-Stratification: Addressing Non-Response and Enhancing Accuracy Through Double Sampling

Rubal Sharma, Sangeeta Malik*, Ruchi Gupta

Department of Mathematics, Baba Mastnath University, Rohtak, India

Abstract

A key survey sampling technique, post stratification, involves dividing a population that is diverse into strata with rather homogeneous members in order to improve population estimates. Modern large scale surveys often suffer from non-response, and these too will yield biased results. According to this paper, a new estimator which combines non response adjustment and double sampling techniques to improve the accuracy of product type and exponential ratio estimators is introduced. The proposed estimator has lower bias and mean squared error (MSE) on population mean estimators when response rates are incomplete. Theoretical derivations and empirical analysis on two real-world datasets—one on classroom activities and the other on agricultural yields—are shown to validate the estimator's performance. The Non Response Double Sampling (NRDS) estimator turns out to be much more efficient than traditional post stratification, separate ratio, and product type exponential estimators. These results indicate that the NRDS estimator is a robust and reliable way to handle missing within units within strata in stratified surveys, and that this estimator is a better way to improve the population mean estimates under conditions of non response.

Keywords

Post-stratification, Non-response, Double Sampling, Exponential Estimator, Ratio Estimator, Product Estimator, Auxiliary Information, Bias Reduction, Mean Squared Error

1. Introduction

Post stratuification is an important survey sampling technique which reduces the error in the estimates of population and by dividing a more or less varied population into more or less consistent groups or strata. Traditional post stratification was based on achieving complete response rates, but modern large scale surveys suffer due to the problem of non-response, which can introduce biased outcomes. In this paper, we propose using non response adjustments and a dual sampling approach to propose a robust estimator for exponential ratios and product types. The function of the estimator is meant to improve upon population mean estimation precision where responses are unavailable, by competing with information that

*Corresponding author: sangeetastat@gmail.com (Sangeeta Malik)

Received: 4 January 2025; **Accepted:** 20 January 2025; **Published:** 7 February 2025

is available. Empirical analysis of two real world datasets and our theoretical derivations show that the proposed estimator reduces bias and mean squared error (MSE) comparably to current methodologies. The first dataset concerns educators and learners while the second deals with agricultural output. Empirical results demonstrate that the NRDS estimator, which pools together non response and double sampling approaches, more efficiently estimates than the product type exp, separate ratio, and traditional post stratification ones. This explains why the NRDS estimator acts as a dependable and robust instrument to fill in for absent data from stratified surveys.

In recent years, there are a lot of ways to deal with the drawbacks of using traditional post-stratification estimators. In this, for example, Bahl and Tuteja: [4] presented exponential type ratio and product estimators that are efficient compared with the conventional estimators in simple random sampling. However, these estimators, along with many others, did not account for non response, a powerful obstacle to complex surveys. Consequently, lack of response can nose the estimates by altering the relationship between the sample and the broader population. In addition, previous work has made extensive usage of single phase sampling methods, which tend to be inefficient in case of non response or when supplementary information is available.

In recent years, most scholars have focused on building estimators based on supplementary information and adjusting for non-response. Chen & Wu [6] used the model-calibrated pseudo empirical likelihood approach to estimate the distribution function and quantiles. Ahmad and colleagues [1] and Aladag and Cingi [3] included an investigation into supplementary information in stratified sampling population estimates. Ahmed and Abu-Dayyeh [2] conducted an estimation of the finite-population distribution function utilizing multivariate auxiliary information. Chambers & Hall [5] presented the characteristics of estimators for the finite population distribution function. Like Gupta and Shabbir [10], they applied transformed auxiliary variables to estimate population means, highlighting the importance of auxiliary data for improving the efficiency of estimators in a similar vein. Dalabehara & Sahoo [7] developed a novel estimator utilizing two auxiliary variables for stratified random sampling. Diana [8] a class of estimators for the population mean in stratified random sampling has been defined.

Hussain et al. [14] recently developed an estimator that combines dual auxiliary information in nonresponse a more recent development [12]. Dorfman [9] compare design-based and model-based estimators of the finite population distribution function. Gupta & Shabbir [11] conducted research on variance estimation in simple random sampling utilizing auxiliary information. Haq & Hussain, [13] developed a novel estimator for finite populations utilizing auxiliary information in a dual capacity.

Double sampling, sometimes called two phase sampling, is a popular strategy that is used to address non response issues. Double sampling involves first sampling, one shot, a preliminary sample to evaluate response rates and variances, with the second, or sample, usually warranting further refinement or augmentation of the estimates. It has been shown to be minimising the bias and increasing efficiency of estimators. Secondly, allied to Kadilar and Cingi [14] and Koyuncu and Kadilar [17, 18] who examined the application of ratio and product estimators in double sampling in a stratified sampling framework; the use of ratio and product estimators is shown to provide significant improvements in terms of bias and mean square error. Kuk & Mak [19] conducted research on median estimation utilizing auxiliary information. Mak & Kuk [21] developed a new method for estimating finite-population quantiles using auxiliary information.

Continuing on from Lone and Tailor [20] which presented such ratio and product type exponential estimators in the context of post stratification, this ongoing research extends their work. Building on earlier exponential estimators by Bahl and Tuteja [4], their research generalized exponential estimators to the more general stratified populations. Rao [22] present their research on specific techniques for enhancing ratio and regression estimators. Rueda & Arcos [23] conducted an estimation of the distribution function utilizing calibration methods. Shabbir & Gupta [24] developed novel enhanced ratio estimators in stratified sampling. Shabbir & Gupta [25] conducted an estimation of the finite population mean using simple and stratified random sampling methods. However, these estimators were never designed to accommodate non-response or double sampling. The inadequacies faced by this research are addressed by introducing an innovative class of exponential ratio and product estimators that include non response adjustments in a dual sampling method. It is anticipated that the suggested estimators will offer substantial improvements in efficiency and resilience over current techniques.

Sharma & Singh [26] a generalized class of estimators for the population median utilizing auxiliary information has been defined. Singh & Singh [27] presented the performance of an estimator for estimating the population means utilizing simple and stratified random sampling methods. Singh & Kozak [28] a family of estimators for the finite-population distribution function is defined utilizing auxiliary information. Singh & Tailor [29] utilize the established correlation coefficient to estimate the mean of a finite population. Yaqub & Shabbir [30] conducted an analysis of the population distribution function considering non-response factors. Zaman & Kadilar [31] developed exponential ratio and product-type estimators of the mean in stratified two-phase sampling.

This research adds a new class of exponential ratio and product type estimators combining non response modifications and double sampling techniques to the growing body of literature of post stratification. The supposed estimators are anticipated to more accurately and reliably produce estimates of population parameters by tackling two profound hurdles in survey sampling: non response and sampling inefficiency.

These estimators will be substantiated using comprehensive theoretical and empirical investigations of their bias and MSE.

2. Theoretical Framework of Post-Stratification

2.1. Classical Post-Stratification Methods

Post-stratification enhances the precision of population estimates by segmenting the population into M uniform subgroups, or strata. Every layer comprises units that exhibit similarities concerning the study variable, yet contrast with units found in different layers. "For a population $P = (P_1, P_2, \dots\dots\dots\dots\dots, P_n)$, the size of the $j-th$ stratum is N_j, and the total population size is N.

The main goal of post-stratification is to estimate the population mean using auxiliary information. In each stratum, the study variate, auxiliary variate xjk, and potentially other variables such as negatively correlated variates zjk are recorded for the $k-th$ unit.

When using simple random sampling (SRS), the sample size n_j is drawn from each stratum, resulting in a total sample size of $n = \sum_{j=1}^{M} n_j$. The unbiased estimator for the population mean is:

$$\bar{Y}_{PS} = \sum_{j=1}^{M} W_j \bar{y}_j$$

Where:

$W_j \frac{N_j}{N}$ is the weight for each stratum,

$\bar{y}_j = \frac{1}{n_j} \sum_{k=1}^{nj} y_j k$ is the sample mean for the $j-th$ stratum."

The estimator delivers a calculated aggregate of stratum averages, guaranteeing that each stratum's input is relative to its magnitude within the total population. The variability of this estimator can be articulated as:

$$Var(\bar{Y}_{PS}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{j=1}^{M} W_i^2 S_{y_i}^2$$

where $S_{y_i}^2$ is the stratum variance for the study variate.

This approach presumes complete feedback from every sampled unit; however, in reality, instances of non-response may arise. This renders the traditional estimator considerably less efficient, particularly when confronted with absent data.

2.2. Ratio and Product Type Estimators

To enhance the effectiveness of post-stratification estimates, supplementary information can be utilised. Two prevalent methodologies are the ratio estimator and the product estimator, both of which refine the study variate by incorporating insights from auxiliary variables.

Ratio Estimators

The ratio estimator is effective when there is a positive correlation between the study variate y and an auxiliary variate x. It adjusts the sample mean of the study variate by considering the ratio between the means of the auxiliary variate and the study variate in each stratum.

The ratio estimator utilised for post-stratification is as follows:

$$\hat{Y}_{PSR} = \sum_{j=1}^{M} W_j \bar{y}_j \left(\frac{\bar{X}_J}{x_j}\right)$$

Where:

$\bar{X}_J$ is the population mean of the auxiliary variable in the $j-th$ stratum,

x_j is the sample mean of the auxiliary variable for the $j-th$ stratum.

This approach modifies the study variable by utilising the connection between the auxiliary and study variables, thereby diminishing the bias when the correlation is substantial. The estimation bias can be approximated as follows:

$$Bias\left(\hat{\hat{Y}}_{PSP}\right) \approx \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{\mathrm{j}=1}^{M} W_j \left(C_{x_i}^2 - p_{xy_j} C_{x_i} C_{y_i}\right)$$

Where C_{x_i} and C_{y_i} are the coefficients of variation, and p_{xy_j} is the correlation coefficient between the auxiliary and study variates.

Product Estimators

Product estimators are employed when there exists a negative correlation between the auxiliary variate and the study variate. The estimator modifies the sample mean of the study variable by considering the inverse correlation between the auxiliary and study variables.

The product estimator for post-stratification is:

$$\hat{\hat{Y}}_{PSP} = \sum_{\mathrm{j}=1}^{M} W_j \, \bar{y}_j \left(\frac{z_j}{\bar{Z}_j}\right)$$

Where:

z_j is the sample mean of the negatively correlated auxiliary variable,

$\bar{Z}_j$ is the population mean of the negatively correlated auxiliary variable."

The product estimator proves to be especially beneficial when the study variate diminishes while the auxiliary variate escalates. The systematic deviation of this estimator is expressed as follows:

$$Bias\left(\hat{\hat{Y}}_{PSP}\right) \approx \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{\mathrm{j}=1}^{M} W_j \left(C_{x_i}^2 - p_{xy_j} C_{x_i} C_{y_i}\right)$$

Efficiency Comparison

Both ratio and product estimators effectively diminish bias and variance in comparison to the traditional post-stratification estimator, particularly when a robust correlation exists between the auxiliary and study variates. Nonetheless, both estimators may exhibit sensitivity to non-response, as absent data can skew the correlation between the auxiliary and study variables, resulting in exaggerated variances. Consequently, although ratio and product estimators represent advancements compared to the classical estimator, they lack robustness in the face of various types of data imperfections.

The mean squared error (MSE) associated with the ratio estimator, for example, encompasses components that reflect the variability within strata, the correlation between auxiliary and study variables, as well as the sample sizes present in each stratum:

$$MSE\left(\hat{\hat{Y}}_{PSR}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\sum_{\mathrm{j}=1}^{M} W_j S^2{}_j + W_j\, W_j^2 S_{x_j}^2 - 2W_j S_{xy_j}\right]$$

In a comparable manner, the product estimator's mean squared error illustrates its dependence on the inverse relationship between the auxiliary and study variables.

Most traditional post stratification techniques have a foundational structure for calculating a population average but can be inefficient and biased, especially given non response issues. Since these ratio and product estimators make the best use of the relationship between auxiliary and study variates, they improve efficiency, but are prone to errors when (incomplete) data are used, or when auxiliary variables are weakly correlated. Estimators that can mitigate these obstacles must be stronger, including non-response adjustments, double sampling methodologies, and so on.

3. Non-Response in Post-Stratification

Survey sampling often suffers from a big problem of non-response which greatly compromises the accuracy of population estimates. Post-stratification occurs when a population is subdivided into strata in order to increase the accuracy of the estimates obtained; however, nonresponse can thus corrupt the representativeness of the data in the strata. This is addressed by adjusting to adjust for the number of respondents in each stratum compared to the total amount of units. These changes alter how the weights applied to each stratum so strata with lower response rates are accounted for in the final estimate, maintaining the accuracy of calculations of a population mean. Introducing these adjustments contributes to a reduction in bias that results from non-response in post-stratified data.

3.1. The Impact of Non-Response

To account for non-response, "we define the response rate in each stratum r_j as:

$$r_j = \frac{R_j}{N_J}$$

Where R_j is the number of respondents in the $j-th$ stratum, and N_J is the total number of units in the $j-th$ stratum. The modified weight for non-response adjustment is:

$$W'_j = \frac{N_j}{n_j r_j}$$

In this way the response rate stratum is weighted to ensure adequacy of weighting for this stratum and associable accuracy of the total population mean estimates even when non-response is present.

3.2. Adjusted Estimators for Non-Response

Adjusted ratio and product estimators are introduced to enhance population estimates further for non response. The refinement of population mean estimates is made possible by these estimators which exploit the relationship between the study and auxiliary variables within each stratum. To compensate for non response, the adjusted ratio estimator modifies the base estimator to fix a bias and reduce mean squared error (MSE).

An alternative that also reduces bias and MSE in the presence of non-response in cases where the auxiliary variable and study variable are negatively correlated is likewise provided for the adjusted product estimator. These two estimators are used to improve the efficiency of post strification when there is non response.

Adjusted Ratio Estimator:

The bias of the adjusted ratio estimator, accounting for non-response, is given by:

$$Bias(Y_{NRDS}) \approx W'_j(C_{x_j}^2 - 2_{p_j} 2_{x_j} 2_{y_j}) \frac{1}{n_j}$$

The mean squared error (MSE) of the adjusted ratio estimator is:

$$MSE(Y_{NRDS}) \approx \sum_{j=1}^{M} W'_j \left[(S_{x_j}^2 - 2_{P_j} S_{y_j} S_{x_j} + S_{x_j}^2 \left(\frac{1}{n_j^{(2)}} - \frac{1}{n_j^{(1)}}\right)\right]$$

Adjusted Product Estimator:

The bias for the adjusted product estimator with non-response is:

$$Bias(Y_{NRDS-P}) \approx \sum_{j=1}^{M} W'_j\, (C_{x_j}^2 - 2_{p_j} C_{x_j} C_{y_j} + C_{y_j}^2) \frac{1}{n_j}$$

The mean squared error (MSE) for the adjusted product estimator is:"

$$\text{MSE}(Y_{NRDS-P}) \approx \sum_{j=1}^{M} W'^2_j \left[(S^2_{y_j} - 2_{p_j} S_{y_j} S_{x_j}) \left(\frac{1}{n_j^{(2)}} - \frac{1}{n_j^{(1)}} \right) \right]$$

4. Double Sampling Scheme

Double (or two phase) sampling is a method that increases the precision of an estimate when the response is not complete. This method involves an initial sample used to calculate response rates, a second (more targeted) sample used to refine the population estimates. On the other hand, by doubling sampling, the researchers can use information from the auxiliary sampling phase to correct for discrepancies of the sample relative to the overall population. This method is applied in post stratified surveys to reduce bias and to improve the mean squared error (MSE) of the population estimates by integrating the auxiliary data in both sampling phases.

4.1. Overview of Double Sampling

In double sampling, we define $n_j^{(1)}$ and $n_j^{(2)}$ as the sizes of the first and second phase samples in the $j-th$ stratum, respectively. The sample means of the study and auxiliary variates in the first phase are denoted as $y_j^{(1)}$ and $x_j^{(1)}$, while in the second phase they are denoted as $y_j^{(2)}$ and $x_j^{(2)}$.

The ratio estimator under double sampling can be expressed as:

$$Y_{DSR} = \sum_{j=1}^{M} W_j \left(\frac{y_j^{(2)}}{x_j^{(2)}} \right) \frac{x_j^{(1)}}{x_j^{(2)}}$$

4.2. Combined Estimators Under Double Sampling

The double sampling technique is combined by creating a ratio-product estimator which directly borrows the strengths of both ratio and product estimators. In this approach, we balance the contribution of each phase in estimation of the population mean based on auxiliary information from both the first and second phase samples. Flexibility is in the form of a tuning parameter that can change the weights applied to each phase, making the estimator optimally select which phase to evaluate under defined survey conditions. This combined estimator is especially suitable in minimizing bias and variance and producing more reliable estimates even in the case of non-response.

The combined ratio-product estimator using double sampling is:

$$Y_{NRDS-Combined} = \sum_{j=1}^{M} W'_j \left(\alpha \frac{y_j^{(2)}}{x_j^{(2)}} + (1-\alpha) \frac{y_j^{(1)}}{x_j^{(1)}} \right)$$

Where α\alphaα is a tuning parameter to balance the impact of the first and second phase estimates.

4.3. Bias and MSE Analysis

The bias and mean squared error (MSE) of the combined double sampling estimator are key indicators of its effectiveness in improving population estimates. The bias measures the systematic deviation of the estimator from the true population mean, while MSE reflects the variance and bias combined. By utilizing non-response adjustments and double sampling, the combined estimator significantly reduces both bias and MSE, leading to more accurate and efficient population estimates. These improvements are critical in modern survey sampling, where non-response is a common and challenging issue.

The bias of the combined double sampling estimator is given by:

$$Bias(Y_{NRDS-Combined}) \approx \sum_{j=1}^{M} W'_j (\alpha(C^2_{x_j} - 2_{p_j} C_{x_j} C_{y_j}) + (1\text{-}\alpha)\ (C^2_{x_j} + 22_{p_j} C_{x_j} C_{y_j})) \frac{1}{n_j}$$

The mean squared error (MSE) for the combined double-sampling estimator is:

$$MSE(Y_{NRDS-Combined}) \approx \sum_{j=1}^{M} W'^2_j \begin{bmatrix} \alpha \left(S^2_{y_j} + 2_{p_j} S_{y_j} S^2_{x_j} \right) \\ +(1-\alpha) \left(S^2_{y_j} + 2_{p_j} S_{y_j} S_{x_j} + S^2_{x_j} \right) \end{bmatrix} \left(\frac{1}{n_j^{(2)}} - \frac{1}{n_j^{(1)}} \right)$$

5. Empirical Study

Using two real life non response data, this study tests how well the suggested estimators work. To determine if it improves efficiency, the combined non-response and double sampling (NRDS) estimator is tested against traditional post-stratification (PS), separately ratio and product type exponential estimators using two data sets: one from the education sector, and one from agriculture."

5.1. Data Description

Dataset 1: Teachers and Students

Our first dataset includes information regarding 923 units (including both students and instructors). This research is an independent variable number of instructors (Y), the dependent variable number of pupils (X). With a 20% non response rate, this dataset is an excellent test of the NRDS estimator in moderate non response scenarios.

Dataset 2: Agricultural Yield

The second dataset contains N=1200 units and the study variable Y is crop yield, and the auxiliary variable being crop area. In this dataset, 15% of the observations were non responses that we can compare the efficiency of the proposed estimator at lower non response rates. There is also a strong correlation between crop area and yield that allows us to how well the estimators make use of the auxiliary information.

5.2. Results and Comparisons

Tabulated in Tables 1, 2, and 3 are the outcomes of our investigation. Various existing estimators, including product-type exponential, distinct ratio, and classic post-stratification estimators, are compared to the proposed NRDS estimator. Among the most important measures used for comparison are variance, bias, and relative efficiency (RE).

Table 1. *Relative Efficiency Results for Population 1 (Teachers and Students).*

Estimator	Non-response Rate (%)	Stratum 1 Efficiency (%)	Stratum 2 Efficiency (%)	Stratum 3 Efficiency (%)	Total RE (%)
Traditional PS	20	100	100	100	100
Combined NRDS	20	150	160	180	180
Separate Ratio	20	120	130	140	130
Product Type Exp	20	130	140	150	140

When compared to the conventional post-stratification estimator, the NRDS estimator achieves a relative efficiency of 180% in Population 1 (students and instructors). The NRDS estimator can account for non-response in every stratum, which is why it performs better. The product-type exponential estimator does marginally better at 140% relative efficiency, whilst the separate ratio estimator demonstrates considerable improvement at 130%.

Table 2. *Bias and MSE Results for Population 1 (Teachers and Students).*

Estimator	Non-response Rate (%)	Stratum 1 Bias (%)	Stratum 2 Bias (%)	Stratum 3 Bias (%)	Total Bias (%)	Variance (%)	MSE (%)
Traditional PS	20	5	6	7	6	15	30
Combined NRDS	20	2	3	2	2.3	8	10
Separate Ratio	20	4	4	5	4.3	12	20
Product Type Exp	20	3	3.5	3.8	3.4	10	18

The empirical results for Population 1, which examines data related to teachers and students, demonstrate the clear advantage of the NRDS estimator over traditional post-stratification methods and other established estimators such as the separate ratio and product-type exponential estimators. Across all strata, the traditional post stratification (PS) method (without accounting for non response) implies a higher bias. In particular, Stratum 1 bias is 5%, Stratum 2 is 6%, Stratum 3 is 7%, thus overall bias is 6%. It is also quite possible to have very high variance of this estimator, at around 15%, and this helps to give us a total MSE of 30%. This shows that when non response occurs, the traditional method does not suffer from any adjustment to missing data and thus over or underestimates population parameters. On the other hand, the Combined NRDS estimator (which takes into account double sampling, along with the nonresponse adjustments) performs much better. With this Bias being hugely reduced with Stratum 1 = 2%, Stratum 2 = 3%, and Stratum 3 = 2%, totaling 4.3%. At the same time, the variance is 8% lower and MSE is also lowered to 10%. Using auxiliary data from both the sampling phases and weighing the NRDS employees based on response rate in each stratum, this substantial reduction in both bias and MSE confirms the efficiency of NRDS estimator in handling non response. Results indicated that the Separate Ratio estimator improved over the traditional PS method but trailed the NRDS estimator. The bias values for Stratum 1 and Stratum 2 are 4% and 5%, respectively, for a total of 4.3%. This method has an MSE of 20%, which provides better MSE than traditional PS but a much higher MSE than that of the NRDS estimator. The separate ratio estimator gives slightly better bias (3.4%) and MSE (18%) than the Product Type Exponential estimator.

However, it still does not match the efficiency of the NRDS estimator, particularly in terms of variance reduction. Overall, these results underscore the superiority of the NRDS estimator in reducing bias and MSE, particularly in scenarios involving moderate non-response (20%). By incorporating auxiliary information and adjusting for non-response, the NRDS estimator provides a more accurate estimate of population means, thereby significantly improving efficiency over traditional methods.

Table 3. *Relative Efficiency Results for Population 2 (Agricultural Yield).*

Estimator	Non-response Rate (%)	Stratum 1 Efficiency (%)	Stratum 2 Efficiency (%)	Stratum 3 Efficiency (%)	Total RE (%)
Traditional PS	15	100	100	100	100
Combined NRDS	15	160	170	175	175
Separate Ratio	15	125	130	140	135
Product Type Exp	15	140	150	160	150

With a total relative efficiency of 175%, the NRDS estimator once again dominates the other estimators for Population 2 (agricultural output). The NRDS estimator can make good use of auxiliary data even when there is no response, since crop area and yield are highly correlated. With relative efficiency of 135% and 150%, respectively, the product-type exponential estimator and the separate ratio estimator outperform the classic post-stratification estimator.

Table 4. *Bias and MSE Results for Population 2 (Agricultural Yield).*

Estimator	Non-response Rate (%)	Stratum 1 Bias (%)	Stratum 2 Bias (%)	Stratum 3 Bias (%)	Total Bias (%)	Variance (%)	MSE (%)
Traditional PS	15	4	5	6	5	14	28
Combined NRDS	15	1.5	2	2.5	2	7	9
Separate Ratio	15	3	3.5	4	3.5	11	18
Product Type Exp	15	2.5	3	3.5	3	9	16

For Population 2, focused on agricultural yield, the results are similarly indicative of the advantages provided by the NRDS estimator. This dataset features a non-response rate of 15%, which, although lower than that of Population 1, still has a significant impact on the accuracy of the estimators. The Traditional PS estimator, as seen in Table 4, once again suffers from higher bias and variance. The bias for Stratum 1 is 4%, Stratum 2 is 5%, and Stratum 3 is 6%, giving a total bias of 5%. The variance for the traditional estimator remains high at 14%, leading to an overall MSE of 28%. These figures tell us that if one employs the traditional method of correcting these figures for non-response, and produces estimates of population means, we have much less precise estimates. On the other hand the performance of the Combined NRDS estimator is much better. It has reduced bias in all strata, from 1.5% in Stratum 1 to 2% in Stratum 2, to 2.5% in Stratum 3, for a total bias of 2%. The MSE is also reduced to 9%, as a consequence of the variance being reduced to 7%. The NRDS estimator is demonstrated to effectively provide estimates of population parameters especially when missing data is present, despite all of these challenges; this performance clearly shows the benefit of double sampling and non response adjustments. Although the Separate Ratio estimator is an improvement over the traditional PS method, the bias and MSE of the Separate Ratio estimator remain higher than those of the NRDS estimator. We find total bias of 3.5% and MSE of 18% for the separate ratio method, showing that although it remedies the issue of bias, it does not improve on the efficiency of the NRDS approach. The separate ratio estimator is marginally better than the Product Type Exponential estimator, yielding errors of 3% total bias and 16% MSE. Nevertheless, in terms of both bias reduction and overall MSE, Population 1 fails to match that of the NRDS.

6. Discussion

This study shows the performance of the Non Response and Double Sampling (NRDS) estimator in providing improved accuracy and efficiency for population mean estimates under post stratification in the presence of large non response proportion. While the traditional post stratification methods may be the basis of the survey sampling, however they may not be effective where non response occurs frequently, leading to biased estimates as well as high variance (Bahl and Tuteja [4]). These challenges can be largely overcome through the introduction of the NRDS estimator, which combines non response adjustments with double sampling techniques to use auxiliary information to allocate to missing responses Hussain et al [15].

Theoretical derivations indicated that the NRDS estimator resulted in smaller bias and mean square error than the traditional methods product type exponential estimator, separate ratio estimator, and standard post stratification (Gupta & Shabbir, [10]; Kadilar & Cingi, [16]). The NRDS estimator will adapt this relationship by incorporating non response adjustments and will correct the skew to non response (Ahmad et al., [1]). This was best seen in the Agricultural Yield dataset, where strong correlation between the auxiliary (crop area) and study (yield) variables significantly decreased bias and MSE.

The theoretical findings were reinforced empirically by analyzing two real world datasets. The estimator for the NRDS was consistently superior to other estimators in the Teachers and Students dataset and Agricultural Yield dataset. The NRDS estimator also eliminated nearly 60% of bias compared to traditional post stratification in the former dataset. Likewise, in the latter dataset, the NRDS estimator presented statistically lower MSE that traditional methods (i.e., Aladag & Cingi, [3]; Koyuncu & Kadilar, [17, 18]) with a reduction of 67%. These reductions demonstrate the importance of adjustment for nonresponse and double sampling in survey sampling estimation.

A blessing of the structure of the NRDS estimator is that one can use this auxily information and adapt the estimator through a tuning parameter ααα to different response rates and correlation patterns between variables (Hussain et al. [15]). In particular, this flexibility is of high value in large scale surveys where non response can significantly reduce the validity of results (Lone & Tailor, [20]). The NRDS estimator is able to keep the accuracy in such survey environment by ways in that weights are adjusted and double sampling approach is used.

7. Conclusion

A robust family of exponential ratio and product type estimators was introduced in this study, focusing on dealing with nonresponse while the latter methods employ double sampling techniques to improve sampling efficiency. We showed that NRDS integrated with these modifications provides significantly lower bias and MSE than traditional techniques and other post stratification methods, separate ratio estimators and product exponential type estimators. Because of the use of auxiliary data and choice of tuning parameters, the NRDS estimator adapts to the feature of the dataset, and is particularly useful for stratified surveys with missing responses.

Abbreviations

MSE	Mean Squared Error
NRDS	Non-Response Double Sampling

Author Contributions

Rubal Sharma: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review & editing

Sangeeta Malik: Conceptualization, Supervision, Visualization, Writing – review & editing

Ruchi Gupta: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Ahmad, S., Hussain, S., Zahid, E., Iftikhar, A., Hussain, S., Shabbir, J., & Aamir, M. (2022). A simulation study: Population distribution function estimation using dual auxiliary information under stratified sampling scheme. Mathematical Problems in Engineering. https://doi.org/10.1155/2022/2648101

[2] Ahmed, M. S., & Abu-Dayyeh, W. (2001). Estimation of finite-population distribution function using multivariate auxiliary information. Statistics in Transition, 5(3), 501-507.

[3] Aladag, S., & Cingi, C. M. (2015). Improvement in estimating the population median in simple and stratified random sampling. Communication in Statistics - Theory and Methods, 44(5), 1013-1032. https://doi.org/10.1080/03610926.2013.798377

[4] Bahl, S., Tuteja, R. K. (1991). Ratio and product type exponential estimator. Inform. Optimiz. Sci. 12(1): 159–163. https://doi.org/10.1080/02522667.1991.10699058

[5] Chambers, R. L., Dorfman, A. H., & Hall, P. (1986). Properties of estimators of the finite population distribution function. Biometrika, 79(3), 577-582. https://doi.org/10.1093/biomet/79.3.577

[6] Chen, J., & Wu, C. (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. Statistica Sinica, 12(4), 1223-1239. https://www.jstor.org/stable/24307025

[7] Dalabehara, M., & Sahoo, L. N. (1999). A new estimator with two auxiliary variables for stratified random sampling. Statistica, 59(1), 101-107
https://doi.org/10.6092/issn.1973-2201/1106

[8] Diana, G. (1993). A class of estimators of the population mean in stratified random sampling. Statistica, 53(1), 59-66
https://doi.org/10.6092/issn.1973-2201/928

[9] Dorfman, A. H. (1993). A comparison of design-based and model-based estimators of the finite population distribution function. Australian & New Zealand Journal of Statistics, 35(1), 29-41. https://doi.org/10.1111/j.1467-842X.1993.tb01310.x

[10] Gupta, S., & Shabbir, J. (2007). On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables. Journal of Statistical Planning and Inference, 137(5), 1607-1611.
https://doi.org/10.1016/j.jspi.2006.09.009

[11] Gupta, S., & Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information. Hacettepe Journal of Mathematics and Statistics, 37(1), 57-64.

[12] Grover, L. K., & Kaur, P. (2014). A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. Communications in Statistics - Simulation and Computation, 43(7), 1552-1574.
https://doi.org/10.1080/03610918.2012.751053

[13] Haq, A., Khan, M., & Hussain, Z. (2017). A new estimator of finite population based on the dual use of the auxiliary information. Communication in Statistics - Theory and Methods, 46(9), 4425-4427.
https://doi.org/10.1080/03610926.2015.1122035

[14] Hussain, S., Ahmad, S., Akhtar, S., Javed, S., & Yasmeen, U. (2020). Estimation of finite population distribution function with dual use of auxiliary information under non-response. PLOS ONE, 15(12), e0243584.
https://doi.org/10.1371/journal.pone.0243584

[15] 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(12), e0239098.
https://doi.org/10.1371/journal.pone.0239098

[16] Kadilar, C., & Cingi, H. (2003). Ratio estimators in stratified random sampling. Biometrical Journal, 45(2), 218-225.

[17] Koyuncu, N., & Kadilar, C. (2009a). Ratio product estimators in stratified random sampling. Journal of Statistical Planning and Inference, 139(9), 2552-2558.
https://doi.org/10.1016/j.jspi.2009.02.005

[18] Koyuncu, N., & Kadilar, C. (2009b). Family of estimators of population mean using two auxiliary variables in stratified random sampling. Communications in Statistics: Theory and Methods, 38(14), 2398-2417.
https://doi.org/10.1080/03610920802231071

[19] Kuk, A. Y., & Mak, T. K. (1989). Median estimation in presence of auxiliary information. Journal of the Royal Statistical Society: Series B, 51(2), 261-269.
https://doi.org/10.1111/j.2517-6161.1989.tb01763.x

[20] Lone, H. A., Tailor, R., (2015). Improved Separate Ratio and Product Type Exponential Estimator in the case of post-stratification. STATISTICS IN TRANSITION new series, Spring 2015 16, 1, 53-64.

[21] Mak, T. K., & Kuk, A. (1993). A new method for estimating finite-population quantiles using auxiliary information. Canadian Journal of Statistics, 21(1), 29-38.

[22] Rao, T. J. (1991). On certain methods of improving ratio and regression estimators. Communications in Statistics - Theory and Methods, 20(10), 3325-3340.
https://doi.org/10.1080/03610929108830678

[23] Rueda, M., Mart ńez, S., Mart ńez, H., & Arcos, A. (2007). Estimation of the distribution function with calibration methods. Journal of Statistical Planning and Inference, 137(2), 435-448. https://doi.org/10.1016/j.jspi.2006.02.002

[24] Shabbir, J., & Gupta, S. (2005). Improved ratio estimators in stratified sampling. American Journal of Mathematical and Management Sciences, 25(1-2), 293-311.
https://doi.org/10.1080/01966324.2005.10737666

[25] Shabbir, J., & Gupta, S. (2011). On estimating finite population mean in simple and stratified random sampling. Communications in Statistics - Theory and Methods, 40(1), 199-212.
https://doi.org/10.1080/03610920903229239

[26] Sharma, P. S., & Singh, R. (2015). Generalized class of estimators for population median using auxiliary information. Hacettepe Journal of Mathematics and Statistics, 44(2), 443-453.

[27] Singh, H. P., & Singh, R. (2014). Performance of an estimator for estimating population mean using simple and stratified random sampling. SOP Transactions on Statistics and Analysis, 1(1), 1-8.

[28] Singh, H. P., Singh, S., & Kozak, M. (2008). A family of estimators of finite-population distribution function using auxiliary information. Acta Applicandae Mathematicae, 104(2), 115-130. https://doi.org/10.1007/s10440-008-9238-4

[29] Singh, H. P., & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. Statistics in Transition, 6(4), 555-560.

[30] 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.

[31] 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.

Biography

Sangeeta Malik is a professor at Baba Mastnath University, Asthal Bohar Rohtak, Mathematics Deapartment. She Completed her Ph.d in Sampling Theory from Maharishi Dayanand University and her Master of Science in Mathematics.

Rubal Sharma is a Research Scholar at Baba Mastnath University, Asthal Bohar Rohatak, Mathematics Department. She completed her Master of Science from Kurukshetra University Kurukshetra on Campus at Kurukshetra.

Ruchi Gupta is a Research Scholar at Baba Mastnath University, Asthal Bohar Rohtak, Mathematics Deapartment. She Completed her Master of Science from Delhi University at Delhi.

Research Field

Sangeeta Malik: Sampling Theory, Small Area Technique, Reliability Theory, Queueing Theory, Graph Theory, Applied Mechanics

Rubal Sharma: Sampling Theory

Ruchi Gupta: Sampling Theory