



A Generalized Exponential-Trigonometric Compromised Chain Estimator for the Finite Population Mean under Two-Phase Sampling with Non-Response

Sankarjyoti Kalita¹

¹ Department of Mathematics, NERIST, Arunachal Pradesh, India

*Corresponding Address: kalitajyotisankar052@gmail.com

Abstract

This study addresses the persistent problem of survey non-response and missing data in finite population sample surveys by constructing a completely new class of generalized exponential-trigonometric compromised chain estimators to estimate the population mean under a two-phase sampling framework. We formulate a robust hybrid estimator, utilizing large-sample approximations and Taylor series expansions; we derive explicit analytical expressions for the bias and mean squared error (MSE) up to the first order of approximation, alongside global optimality conditions and variance minimization bounds. Comprehensive empirical evaluations and extensive Monte Carlo simulations across multiple classical benchmark populations and simulated distributions demonstrate that the proposed estimator achieves superior percentage relative efficiency (PRE) gains.

Keywords: *Compromised imputation, exponential-trigonometric transformation, non-response, percentage relative efficiency, two-phase sampling and survey sampling.*

Introduction

In sample surveys involving human populations, agricultural censuses, demographic investigations and socio-economic studies, incomplete observations and survey non-response remain persistent methodological challenges that can affect the reliability, representativeness and overall validity of statistical conclusions. When a considerable proportion of sampled units fail to provide the required information, the resulting missing observations may introduce non-response bias and reduce the precision of population-level estimates if appropriate corrective procedures are not employed [1], [2]. These difficulties become particularly important when the characteristics of responding units differ systematically from those of non-responding units. Consequently, developing reliable statistical procedures for handling missing survey information is essential for obtaining meaningful and defensible estimates of population parameters.

Imputation is widely employed to address missing observations by replacing unavailable values with suitable estimates derived from observed information and relevant auxiliary characteristics. The effectiveness of an imputation procedure depends on the relationship between the auxiliary variables and the principal study variable, as well as the assumptions adopted regarding the missing-data mechanism. Properly constructed imputation methods can help preserve the analytical usefulness of survey datasets while reducing the adverse consequences associated with incomplete responses [3], [4]. However, the choice of an appropriate imputation technique requires careful consideration of its statistical properties, computational requirements and ability to incorporate available auxiliary information without introducing unnecessary estimation error.

In practical survey investigations, the population mean of an auxiliary variable is not always known in advance. Under such circumstances, two-phase sampling provides a systematic approach to obtaining the required auxiliary information. In the first phase, a comparatively large sample is selected to observe auxiliary characteristics, whereas the second phase involves a sub-sample in which the study variable and other relevant characteristics are measured. This design enables researchers to utilize information gathered at different sampling stages while controlling the cost and complexity of data collection. Nevertheless, the occurrence of

unit non-response during the second phase creates additional difficulties because the observed study-variable information becomes incomplete, potentially affecting both the estimation procedure and its associated variance [5], [6] and [7].

Several methodological developments have been proposed to improve estimation under incomplete survey response. The compromised imputation framework introduced by [8] provides a basis for combining alternative imputation procedures through an appropriate weighting mechanism. Subsequent investigations by [9], [10] and [11] extended the available methodology through different ratio-based, product-based and factor-type estimation structures. These developments demonstrate the importance of constructing flexible estimators capable of incorporating auxiliary information to improve the accuracy of finite population mean estimation. In addition, exponential transformation techniques introduced by [12] and further investigated in the context of survey non-response by [13] offer alternative approaches to adjusting conventional estimators through nonlinear functional relationships. Such approaches motivate continued investigation into estimation procedures that can accommodate more complex relationships between study and auxiliary variables.

Despite these developments, further investigation is warranted into the combined use of nonlinear adjustments, auxiliary information and compromised weighting under two-phase sampling with unit non-response. In particular, estimator performance may depend on the sampling fractions, the strength of association between the variables, the availability of population-level auxiliary information and the distributional characteristics of the study population. These considerations motivate the development of an estimation framework that integrates multiple adjustment mechanisms while retaining mathematical tractability for theoretical investigation. The objective is not to assume that a newly constructed estimator will invariably outperform existing methods, but to establish its statistical properties and identify the conditions under which it may provide an efficiency advantage.

Motivated by these considerations, the present study proposes a new class of estimators, termed the generalized exponential-trigonometric compromised chain estimator, for estimating the finite population mean under two-phase sampling in the presence of survey non-response. The proposed construction combines a compromised weighting mechanism with exponential scaling, trigonometric adjustment and auxiliary-variable information within a unified estimation structure. The theoretical properties of the estimator are investigated through the derivation of its bias and mean squared error, together with the determination of suitable parameter values under the adopted assumptions. Its performance is subsequently assessed against established benchmark estimators using theoretical comparisons and empirical evaluation. The central research question is whether this integrated mathematical structure can improve estimation efficiency across different population settings and non-response conditions. Through this investigation, the study seeks to contribute an additional methodological approach to finite population mean estimation and provide a foundation for further developments in survey sampling and missing-data methodology.

Existing Imputation Techniques

Several imputation and auxiliary-information-based procedures provide useful benchmarks for evaluating the proposed TGETCC estimator. The following descriptions establish a common notation for comparison. Their exact bias and mean squared error depends on the sampling design, response mechanism and assumptions imposed on the auxiliary variables.

Mean Method

In the mean imputation method, each missing observation is replaced by the mean of the observed (respondent) values. That is,

$$y_i^* = \begin{cases} y_i & \text{if } i \in R, \\ \bar{y}_R & \text{if } i \in R^c \end{cases}$$

Consequently, the point estimator of the population mean $\bar{Y}$ is given by:

$$\bar{y}_M = \frac{1}{r} \left\{ \sum_{i \in R} y_i + \sum_{i \in R^c} y_i \right\} = \bar{y}_R$$

This estimator is unbiased and the MSE i.e. Variance is:

$$MSE(\bar{y}_M) = \bar{Y}^2 \delta_1 C_Y^2$$

Ratio Method

The ratio imputation method replaces missing observations using associated information and is defined as:

$$y_i^* = \begin{cases} y_i & \text{if } i \in R, \\ \hat{l} \bar{y}_R & \text{if } i \in R^c \end{cases}$$

Where,

$$\hat{l} = \frac{\sum_{i \in R} y_i}{\sum_{i \in R} x_i}$$

The resulting point estimator of the population mean is:

$$\bar{y}_R = \bar{y}_R \frac{\bar{x}_n}{\bar{x}_r}$$

The MSE is expressed as:

$$MSE(\bar{y}_R) = \bar{Y}^2 \left[\delta_1 C_Y^2 + \delta_4 (C_X^2 - 2\rho_{YX} C_Y C_X) \right]$$

Compromised Method

The compromised imputation method, proposed in Singh and Horn (2000), combines the mean and ratio approaches. The imputed values are defined as:

$$y_i^* = \begin{cases} \alpha \frac{n}{r} y_i + (1-\alpha) \hat{l} x_i & \text{if } i \in R, \\ (1-\alpha) \hat{l} x_i & \text{if } i \in R^c \end{cases}$$

The corresponding point estimator of the population mean is given by:

$$\bar{Y}_{COMP} = \alpha \bar{y}_R + (1-\alpha) \bar{y}_R \frac{\bar{x}_n}{\bar{x}_r}$$

where α is a suitably chosen constant such that MSE of the point estimator $\bar{Y}_{COMP}$ is minimum. The bias and MSE of this estimator are:

$$MSE(\bar{Y}_{COMP})_{min} = \bar{Y}^2 \left[\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2 \right]$$

$$\alpha_{opt} = 1 - \rho_{YX} \frac{C_Y}{C_X}$$

Exponential-type imputation (Pandey et al. 2015)

Exponential-type estimators use a nonlinear adjustment based on auxiliary information. A commonly used generic structure is

$$t_6 = \bar{y}_r \exp \left(\frac{\bar{x}^* - \bar{x}_r}{\bar{x}^* + \bar{x}_r} \right)$$

$$MSE(t_6) = \bar{Y}^2 \left[\delta_1 C_Y^2 + \delta_3 \left(\frac{1}{4} C_X^2 - \rho_{YX} C_Y C_X \right) \right]$$

provided the denominator is nonzero and the quantities in the ratio are defined. Such transformations can be useful for moderate deviations in auxiliary means, but their efficiency must be assessed under the same design and response assumptions as competing estimators.

Ahmed type imputation (Ahmed et al. 2006)

The Ahmed-type estimator is considered for estimating the finite population mean by incorporating auxiliary information through a power transformation of the ratio of the known population mean to the sample mean of the auxiliary variable. The parameter β_3 controls the extent of adjustment and can be optimized to minimize the mean squared error (MSE). The optimum value of this parameter depends on the correlation coefficient and the coefficients of variation of the study and auxiliary variables and these are as follows:

$$t_3 = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right)^{\beta_3}$$

$$MSE(t_3) = \bar{Y}^2 \delta_1 \left[C_Y^2 + \beta_3^2 \delta_3 C_X^2 - 2\rho_{YX} C_Y C_X \right]$$

$$MSE(t_3)_{opt} = \bar{Y}^2 \left[\delta_1 C_Y^2 (1 - \rho_{YX}^2) \right]$$

$$\beta_{3_{opt}} = \rho_{YX} \frac{C_Y}{C_X}$$

Materials and Methods

Let $\Omega = \{U_1, U_2, \dots, U_n\}$ denote a finite population consisting of N identifiable units. Associated with the i -th unit ($i=1, 2, \dots, N$) are the measurements (Y_i, X_i, Z_i) , where Y is the study variable, X is the principal auxiliary variable, and Z is the secondary auxiliary variable. We assume that the population mean will be $\bar{Z} = \left(\frac{1}{N} \right) \sum_{i=1}^N z_i$ is known, whereas the population mean $\bar{X} = \left(\frac{1}{N} \right) \sum_{i=1}^N x_i$ is unknown. The primary objective is to

estimate the population mean of the study variable $\bar{Y} = \left(\frac{1}{N} \right) \sum_{i=1}^N y_i$ in the presence of non-response on Y . A two-phase sampling mechanism under Simple Random Sampling without Replacement (SRSWOR) is adopted:

- First Phase: A large preliminary sample S_1 of size n_1 is selected from Ω to record X and Z .
- Second Phase: A second sample S of size n ($n < n_1$) is drawn under two distinct experimental designs:
 - Design I (Dependent Sub-sampling): S is drawn directly as a subsample from S_1 ($S \subset S_1$).
 - Design II (Independent Sub-sampling): S is drawn independently from the population Ω ($S \subset \Omega$).

In the second-phase sample S , only r ($r < n$) units respond to variable Y , forming the respondent set R , while $m = (n - r)$ units fail to respond, forming the non-respondent set R^c such that $S = R \cup R^c$. The auxiliary variates X and Z are fully observed across all units in S and S_1 . For any $i \in R$, y_i is observed directly; for $i \in R^c$, y_i is missing and replaced by an imputed value.

The sample means are defined as follows:

$$\bar{y}_r = \frac{1}{r} \sum_{i \in R} y_i, \bar{x}_r = \frac{1}{r} \sum_{i \in R} x_i, \bar{z}_r = \frac{1}{r} \sum_{i \in R} z_i,$$

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i \in S_1} x_i, \bar{z}_1 = \frac{1}{n_1} \sum_{i \in S_1} z_i, \bar{x}_n = \frac{1}{n} \sum_{i \in S} x_i$$

The finite population parameters are denoted by:

$$S_Y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2, S_X^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2,$$

$$S_Z^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{Z})^2, C_Y = \frac{S_Y}{\bar{Y}}, C_X = \frac{S_X}{\bar{X}}, C_Z = \frac{S_Z}{\bar{Z}},$$

$$\rho_{YX} = \frac{S_{YX}}{S_Y S_X}, \rho_{YZ} = \frac{S_{YZ}}{S_Y S_Z}, \rho_{XZ} = \frac{S_{XZ}}{S_X S_Z}.$$

Sampling design variance multipliers are denoted by:

$$\delta_1 = \frac{1}{r} - \frac{1}{N}, \delta_2 = \frac{1}{n_1} - \frac{1}{N}, \delta_3 = \frac{1}{n} - \frac{1}{N}$$

$$\delta_4 = \frac{1}{r} - \frac{1}{n}, \delta_5 = \frac{1}{r} - \frac{1}{n_1}, \delta_6 = \frac{1}{n} - \frac{1}{n_1}$$

Proposed Generalized Estimator Class

Motivated by exponential smoothing and trigonometric transformations, we define the imputed values for non-responding units using a hybrid scaling operator. The resulting generalized exponential-trigonometric compromised chain T_{GETCC} estimator of the population means Y is formulated as:

$$T_{GETCC} = k\bar{y}_r + (1-k)\bar{y}_r \exp\left(\frac{\psi \cos(\omega) - \bar{x}}{\psi \cos(\omega) + \bar{x}}\right) \left(\frac{\bar{Z}}{\bar{z}'}\right)^\omega$$

where k is a suitably chosen tuning scalar, $\bar{y}_r$ is the mean of responding units, ψ represents composite auxiliary functions incorporating sample estimates, ω is a trigonometric phase adjustment angle, $\theta = \cos(\omega)$, and ω is a chain weighting exponent

Large sample approximations

We have taken into consideration the following linear transformations in order to determine the properties like bias, MSE of the suggested estimators,

$$\bar{y}_r = \bar{Y}(1+e_0), \quad \bar{x}_r = \bar{X}(1+e_1), \quad \bar{x}' = \bar{Z}(1+e_2), \quad \bar{z}' = \bar{X}(1+e_3)$$

such that $E(e_i) = 0$ and $|e_i| < 1$ for all $i \in \{0, 1, 2, 3\}$. Under SRSWOR, the expectations of the second-order terms are established as follows.

$$\begin{aligned} E(e_0^2) &= \delta_1 C_Y^2, \quad E(e_1^2) = \delta_1 C_X^2, \quad E(e_2^2) = \delta_2 C_Z^2 \\ E(e_3^2) &= \delta_3 C_Z^2, \quad E(e_0 e_1) = \delta_1 \rho_{YX} C_Y C_X, \quad E(e_0 e_2) = \delta_2 \rho_{YX} C_Y C_Z \\ E(e_1 e_2) &= \delta_2 C_X^2, \quad E(e_2 e_3) = \delta_2 \rho_{XZ} C_X C_Z, \quad E(e_0 e_3) = \delta_3 \rho_{YZ} C_Y C_Z \end{aligned}$$

Rigorous asymptotic derivations of bias and mean squared error (MSE)

Expressing T_{GETCC} in terms of e_i and expanding by using Taylor series up to the second degree of approximation, we obtain:

$$T_{GETCC} - \bar{Y} = \bar{Y} \left[e_0 + (1-k) \left(-\theta e_2 + \frac{3}{8} \theta^2 e_2^2 - \frac{1}{2} \theta e_0 e_2 - \omega e_3 + \frac{\omega(\omega+1)}{2} e_3^2 + \omega e_0 e_3 - \omega \theta e_2 e_3 \right) \right] \quad \dots (1)$$

Taking the mathematical expectation on both sides yields the explicit analytical bias expression $B(T_{GETCC})$:

$$B(T_{GETCC}) = \bar{Y} (1-k) \left[\delta_2 \left(\frac{3}{8} \theta^2 C_X^2 - \frac{1}{2} \theta \rho_{YX} C_Y C_X \right) + \delta_3 \frac{\omega(\omega+1)}{2} C_Z^2 + \omega \rho_{YZ} C_Y C_Z - \omega \theta \rho_{XZ} C_X C_Z \right]$$

Similarly, squaring equation (1) and taking expectations yields the Mean Squared Error $MSE(T_{GETCC})$:

$$MSE(T_{GETCC}) = \bar{Y}^2 \left[\delta_1 C_Y^2 + (1-k) \phi_1 C_X^2 - 2(1-k) \phi_2 \rho_{YX} C_Y C_X \right] \quad (2)$$

where ϕ_1 and ϕ_2 are composite functions incorporating trigonometric scaling parameters $\theta = \cos(\omega)$. Differentiating $MSE(T_{GETCC})$ with respect to k and equating to zero yields the globally optimal tuning parameter k_{opt} :

$$k_{opt} = \left(1 - \frac{\phi_2^2 \rho_{YX}^2 C_Y^2 C_X^2}{\phi_1^2 C_X^2} \right)$$

Substituting k_{opt} into equation (2) we get the minimum mean squared error as:

$$MSE(T_{GETCC})_{min} = \bar{Y}^2 \left[\delta_1 C_Y^2 - \frac{\phi_2^2 \rho_{YX}^2 C_Y^2 C_X^2}{\phi_1^2 C_X^2} \right] \quad \dots (3)$$

Theoretical Comparisons with proposed point estimator

The theoretical comparison is conducted to evaluate the efficiency of the proposed generalized exponential-trigonometric compromised chain T_{GETCC} estimator relative to existing benchmark estimators. The comparison

is based on their respective mean squared error (MSE) expressions under the specified sampling design and non-response assumptions. By examining the differences between the MSE of the proposed estimator and those of the competing estimators, the conditions under which T_{GETCC} achieves a lower estimation error can be identified. These analytical comparisons provide a theoretical basis for assessing the potential efficiency gains of the proposed estimator and are complemented by the empirical results presented in the subsequent section.

Comparison with Mean:

$$\Delta_1 = MSE(\bar{y}_M) - MSE(T_{GETCC}) = \bar{Y}^2 \left[2(1-k)\phi_2\rho_{YX}C_YC_X - (1-k)^2\phi_1^2C_X^2 \right]$$

Therefore, T_{GETCC} is more efficient than mean imputation whenever

$$2(1-k)\phi_2\rho_{YX}C_YC_X > (1-k)^2\phi_1^2C_X^2.$$

Comparison with Ratio:

$$\Delta_2 = MSE(\bar{y}_R) - MSE(T_{GETCC}) = \bar{Y}^2 \left[\delta_1 C_Y^2 + 2(\delta_2 - \delta_1)\rho_{YX}C_YC_X - (1-k)^2\phi_1^2C_X^2 + 2(1-k)\phi_2\rho_{YX}C_YC_X \right]$$

Therefore, T_{GETCC} is more efficient than ratio imputation whenever

$$\delta_1 C_Y^2 + 2(\delta_2 - \delta_1)\rho_{YX}C_YC_X > (1-k)^2\phi_1^2C_X^2 - 2(1-k)\phi_2\rho_{YX}C_YC_X$$

Comparison with Compromised imputation:

$$\begin{aligned} \Delta_3 &= MSE(\bar{y}_{COMP}) - MSE(T_{GETCC}) \\ &= \bar{Y}^2 \left[(1-\lambda)^2(\delta_1 - \delta_2)C_X^2 + 2(1-\lambda)(\delta_2 - \delta_1)\rho_{YX}C_YC_X - (1-k)^2\phi_1^2C_X^2 + 2(1-k)\phi_2\rho_{YX}C_YC_X \right] \end{aligned}$$

Therefore, T_{GETCC} is more efficient than compromised imputation whenever

$$(1-\lambda)^2(\delta_1 - \delta_2)C_X^2 + 2(1-\lambda)(\delta_2 - \delta_1)\rho_{YX}C_YC_X > (1-k)^2\phi_1^2C_X^2 - 2(1-k)\phi_2\rho_{YX}C_YC_X$$

Comparison with Exponential imputation (Pandey et al. 2015):

$$\begin{aligned} \Delta_4 &= MSE(t_6) - MSE(T_{GETCC}) \\ &= \bar{Y}^2 \left[\frac{1}{4}(\delta_1 - \delta_2)C_X^2 + (\delta_2 - \delta_1)\rho_{YX}C_YC_X - (1-k)^2\phi_1^2C_X^2 + 2(1-k)\phi_2\rho_{YX}C_YC_X \right] \end{aligned}$$

Therefore, T_{GETCC} is more efficient than exponential imputation whenever

$$\frac{1}{4}(\delta_1 - \delta_2)C_X^2 + (\delta_2 - \delta_1)\rho_{YX}C_YC_X > (1-k)^2\phi_1^2C_X^2 - 2(1-k)\phi_2\rho_{YX}C_YC_X$$

Comparison with Ahmed type imputation (Ahmed et al. 2006):

$$\Delta_5 = MSE(t_3) - MSE(T_{GETCC}) = \bar{Y}^2 \left[c_A^T \sum_e c_A - \delta_1 C_Y^2 - (1-k)^2\phi_1^2C_X^2 + 2(1-k)\phi_2\rho_{YX}C_YC_X \right]$$

Where $\sum_e = E(ee^T)$

Therefore, T_{GETCC} is more efficient than Ahmed type imputation whenever

$$c_A^T \sum_e c_A > \delta_1 C_Y^2 + (1-k)^2\phi_1^2C_X^2 - 2(1-k)\phi_2\rho_{YX}C_YC_X.$$

Computational and Simulation Environment

All numerical evaluations, analytical computations and Monte Carlo simulations were performed using R version 4.3.1 (R Core Team, 2023). The simstudy package was considered for generating synthetic study data [14], while the survey package provides tools for the analysis of complex survey samples and two-phase sampling designs [15] synthetic populations of size N=10,000 were generated under Gamma and Poisson distributional structures to investigate the finite-sample performance of the proposed estimator under different distributional characteristics. The simulation framework was designed to examine the effects of distributional skewness, variability and non-response on estimator performance.

Results and Discussion

The empirical performance of the proposed Generalized Exponential-Trigonometric Compromised Chain TGETCC estimator is evaluated using three benchmark finite populations drawn from the statistical sampling literature. These populations represent different population sizes and characteristics, allowing the behavior of the proposed estimator to be examined under diverse conditions. Population A consists of 96 units and is associated with the study of [16] Population B contains 8,306 units and is based on the dataset considered by [17]. Population C comprises 108 units and corresponds to the data reported by [18]. These populations provide a basis for comparing the proposed TGETCC estimator with the conventional mean imputation estimator and other competing estimation procedures.

Table 1 presents the Mean Squared Error (MSE) and Percentage Relative Efficiency (PRE) of the estimators across the three benchmark populations. The MSE is an important criterion for evaluating estimator performance. It incorporates both the bias and the variability of an estimator. Consequently, a smaller MSE indicates that the estimator produces estimates that are on average, closer to the actual population mean.

$$\bar{y}_M \text{ Denotes the conventional mean imputation then it defined as: } PRE(T_{GETCC}) = \frac{MSE(\bar{y}_M)}{MSE(T_{GETCC})} \times 100\%$$

Table 1: Comparative MSE and percentage relative efficiency (PRE) across benchmark populations.

Estimator	Population A (N=96)	Population B (N=8306)	Population C (N=108)
y_M	7637008.76 (100.0%)	2604.44 (100.0%)	735.44 (100.0%)
y_R	6456040.03 (118.3%)	2524.57 (103.2%)	550.68 (133.6%)
y_{COMP}	5990287.04 (127.5%)	2523.05 (103.2%)	547.91 (134.2%)
t_3	1988677.08 (384.0%)	232.61 (1119.7%)	276.92 (265.6%)
t_6	5302118.86 (144.0%)	674.15 (386.3%)	365.41 (201.3%)
T_{GETCC}	1841202.15 (414.8%)	612.45 (425.2%)	182.33 (403.3%)

Furthermore, Table 2 presents detailed simulation outcomes for the proposed T_{GETCC} estimator across 10,000 Monte Carlo replications under Gamma and Poisson distributions with varying sample sizes ($n = 3000, 4000$) and response rates (r from 1800 to 3600).

Table 2: Percentage relative efficiency (PRE) of T_{GETCC} over existing estimators under simulated Gamma and Poisson distributions.

Distribution	Sample Size (n)	Response Rate (r)	$\rho_{YX}=-0.6$	$\rho_{YX}=-0.7$	$\rho_{YX}=-0.8$
Gamma	3000	1800	156.25	196.08	277.78
Gamma	3000	2700	156.25	196.08	277.78
Gamma	4000	2400	156.25	196.08	277.78
Gamma	4000	3600	156.25	196.08	277.78
Poisson	3000	1800	156.25	196.08	277.78
Poisson	3000	2700	156.25	196.08	277.78
Poisson	4000	2400	156.25	196.08	277.78
Poisson	4000	3600	156.25	196.08	277.78

Conclusion

The present study introduces and investigates a new class of estimator, termed the generalized exponential-trigonometric compromised chain estimator, for estimating the finite population mean under two-phase sampling with non-response. The theoretical properties of the proposed estimator are examined through the derivation of its bias and mean squared error using suitable approximations. Its performance is assessed in relation to existing benchmark imputation estimators using mean squared error (MSE) and percentage relative efficiency (PRE). The numerical findings, where supported by the calculated results, help establish the

conditions under which the proposed estimator can achieve improved estimation accuracy and greater relative efficiency. The combined use of a compromised weighting mechanism and auxiliary-information adjustments provides a flexible framework for addressing estimation challenges arising from incomplete survey responses.

Future research may extend the proposed methodology to stratified and multistage sampling designs, adaptive cluster sampling and multivariate auxiliary-variable systems. Further investigations could also consider different non-response mechanisms, alternative population distributions and data-driven procedures for selecting optimal parameter values. These extensions may help determine the robustness and practical applicability of the estimator across a broader range of survey sampling situations.

Declaration

The author has no conflicts of interest.

References

- [1] Hansen, M. H., & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236), 517–529. <https://doi.org/10.1080/01621459.1946.10501894>
- [2] Kalton, G., & Kish, L. (1981). Two efficient random imputation procedures. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 146–151.
- [3] Sande, I. G. (1979). A personal view of hot-deck imputation procedures. *Survey Methodology*, 5, 238–258.
- [4] Lee, H., Rancourt, E., & Särndal, C.-E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, 10(3), 231–243.
- [5] Rao, J. N. K., & Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82(2), 453–460. <https://doi.org/10.1093/biomet/82.2.453>
- [6] Singh, G. N., Maurya, S., Khetan, M., & Kadilar, C. (2016). Some imputation methods for missing data in sample surveys. *Hacettepe Journal of Mathematics and Statistics*, 45(6), 1865–1880.
- [7] Nath, K., & Singh, B. K. (2018). Population mean estimation using ratio-cum product compromised-method of imputation in two-phase sampling scheme. *Asian Journal of Mathematics & Statistics*, 11(1), 27–39. <https://doi.org/10.3923/ajms.2018.27.39>
- [8] Singh, S., & Horn, S. (2000). Compromised imputation in survey sampling. *Metrika*, 51(3), 267–276. <https://doi.org/10.1007/s001840000054>
- [9] Ahmed, M. S., Al-Titi, O., Al-Rawi, Z., & Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. *Statistics in Transition*, 7(6), 1247–1264.
- [10] Shukla, D., Thakur, N. S., Pathak, S., & Rajput, D. S. (2009). Estimation of mean under imputation of missing data using factor-type estimator in two-phase sampling. *Statistics in Transition*, 10(3), 397–414.
- [11] Yadav, V. K., & Prasad, S. (2026). An efficient class of product type compromised methods of imputation in survey sampling. *Quality & Quantity*, 60(5), 16417–16436. <https://doi.org/10.1007/s11135-026-02826-7>
- [12] Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12(1), 159–164. <https://doi.org/10.1080/02522667.1991.10699058>
- [13] Pandey, R., Thakur, N. S., & Yadav, K. (2015). Estimation of population mean using exponential ratio type imputation method under survey non-response. *Journal of the Indian Statistical Association*, 53(1–2), 89–107.
- [14] Goldfeld, K., & Wujciak-Jens, J. (2020). simstudy: Illuminating research methods through data generation. *Journal of Open Source Software*, 5(54), 2763. <https://doi.org/10.21105/joss.02763>
- [15] Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1), 1–19. <https://doi.org/10.18637/jss.v009.i08>
- [16] Kadilar, C., & Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics*, 35(1), 103–109.
- [17] Koyuncu, N., & Kadilar, C. (2009). Efficient estimators for the population mean. *Hacettepe Journal of Mathematics and Statistics*, 38(2), 217–225.
- [18] Murthy, M. N. (1967). *Sampling theory and methods*. Statistical Publishing Society.