Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Hamisu *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.012

JOURNAL OF STATISTICAL SCIENCES & DATA ANALYTICS
Department of Statistics, Ahmadu Bello University, Zaria, Nigeria | www.jossda.abu.edu.ng www.jossda.com.ng

Robust Machine Learning Imputation Estimator Using Multiple Auxiliary Variables for Handling Missing Data in Survey Sampling

H. A. Hamisu[1*], O. O. Ishaq[1], A. A. Osi[1], U. A. Huguma[1], N. Yusuf[1], S. S. Suleiman[1], J. S. Ogheneovo[1], F. U. Muhammad[1], A. F. Kabara[1]

[1]*Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, Nigeria*

Corresponding Author:
H. A. Hamisu, Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, Nigeria. Ahamisu437@gmail.com

Abstract

Missing data caused by survey nonresponse can bias finite-population estimates and reduce statistical efficiency, particularly when conventional imputation methods impose linearity or depend on a single auxiliary variable. This study develops a robust machine-learning imputation estimator of the finite-population mean using Gradient Boosting Machines (GBM) and multiple auxiliary variables. Under simple random sampling without replacement and a missing-at-random mechanism, a GBM model is trained on responding units and used to predict the study variable for nonrespondents. The completed-sample estimator is expressed in prediction-error form, from which its approximate bias and mean squared error (MSE) are derived. The bias is proportional to the nonresponse fraction and the average GBM prediction bias, while the MSE comprises the sampling variance, unexplained residual variance, prediction variance, and squared prediction-bias penalty. A Monte Carlo study with 10,000 replications compares the proposed estimator with mean, regression, Lee (1994), Singh and Horn (2000), Singh and Deo (2003), Kadilar and Cingi (2008), Singh (2009), Gira (2015), Prasad (2017), and Singh et al. (2022) estimators under linear, nonlinear, outlier-contaminated, and skewed data conditions. The proposed GBM estimator attained the smallest MSE under linear and nonlinear conditions, with MSEs of 1.890 and 12.494, respectively, and remained the second-best method under outlier and skewed conditions, where Prasad's estimator was marginally superior. Its MSE decreased consistently as the response rate increased from 50% to 90%. The results demonstrate that machine-learning-assisted imputation can substantially improve finite-population mean estimation when several informative auxiliary variables and complex relationships are present, although robust benchmark procedures remain important for heavily contaminated or skewed data.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

Keywords: gradient boosting machine; imputation; missing at random; multiple auxiliary variables; nonresponse; survey sampling.

1. Introduction

Sample surveys are indispensable for official statistics, public-health surveillance, market research, and social-science investigation. Their inferential quality, however, is often threatened by unit or item nonresponse. When the probability of response is associated with the study variable, complete-case analysis may produce biased estimates; even when the missingness mechanism is ignorable, discarding incomplete observations increases variance and wastes auxiliary information (Groves *et al., 2009*; Little and Rubin, 2019). Imputation replaces missing values with plausible substitutes, permitting conventional complete-data analyses and helping preserve the analytical sample.

Classical survey imputation procedures include mean, ratio, regression, and transformation-based methods. These approaches are attractive because they are transparent and computationally inexpensive, but many rely on a pre-specified parametric relationship and one auxiliary variable. A misspecified linear model can perform poorly when the response surface is nonlinear, contains interactions, or is affected by outliers and skewness. In contemporary surveys, several auxiliary variables are frequently available from administrative records, paradata, or earlier survey waves. The main methodological challenge is therefore to exploit this information without imposing an unduly restrictive model.

Machine-learning algorithms offer flexible function approximation and automatic interaction detection. Tree-based ensembles are particularly relevant because they handle nonlinearities, mixed scales, and multicollinearity without requiring a global linear equation. Random forests have been incorporated into model-assisted estimation, while boosting methods iteratively combine weak learners to reduce prediction error (Breiman, 2001; Hastie *et al*., 2009; McConville *et al*., 2017). Nevertheless, strong predictive accuracy alone does not establish the statistical quality of an imputation estimator. Its bias, MSE, design behaviour, and efficiency relative to established procedures must be examined.

This paper develops a GBM-based imputation estimator of the finite-population mean using six auxiliary variables. The contribution is threefold. First, it provides a completed-sample estimator that combines observed respondent outcomes with GBM predictions for nonrespondents. Second, it derives an explicit error representation, approximate bias, MSE, and a general efficiency condition. Third, it evaluates the estimator against ten established alternatives under linear, nonlinear, outlier-contaminated, and skewed data structures and across response rates of 50%, 70%, and 90%.

2. Related Imputation Estimators

Let y be the study variable and x an auxiliary variable observed for all sampled units. Mean imputation assigns the respondent mean to missing observations. Ratio and regression imputation use the association between y and x, while compromised and power-transformation estimators introduce constants or transformations to improve efficiency. *Lee et al*. (1994) studied variance estimation from imputed survey data; Singh and Horn (2000)

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

proposed compromised imputation; Singh and Deo (2003) used a power transformation; Kadilar and Cingi (2008), Singh (2009), Gira (2015), Prasad (2017), and Singh *et al*. (2022) developed additional ratio-type or optimised imputation procedures.

These estimators can be efficient when their working assumptions are close to the data-generating mechanism. However, the gains may deteriorate if the outcome depends on several auxiliary variables through nonlinear or interactive relationships. A multivariate linear regression can incorporate several predictors, but it remains sensitive to functional-form misspecification and influential observations and may require manual terms or transformations. The proposed approach replaces the fixed regression function by a data-adaptive GBM predictor, while retaining a clearly defined estimator of the population mean.

3. Methodology

3.1. Sampling Framework and Missingness Mechanism

Consider a finite population $U = (1,\ldots, N)$ from which a sample s of size *n* is selected by simple random sampling without replacement. Let R denote the set of r respondents and m the set of m = n − r nonrespondents. For every sampled unit i, the p-dimensional auxiliary vector x_i is completely observed, whereas y_i is observed only if *i* belongs to R. The target parameter is the finite-population mean $\bar{Y} = \frac{1}{N}\sum_{i\in U} Y_i$. The analysis assumes that the missingness mechanism is missing at random conditional on x_i , so the distribution of y_i among units having the same auxiliary information does not depend on response status.

$$y_{i=} f(x_i) + \varepsilon_i,\ E(\varepsilon_i / x_i) = 0,\ Var(\varepsilon_i / x_i) = \delta^2 \qquad (1)$$

Here $f(.)$ is an unknown and potentially complex conditional-mean function. A GBM, denoted $\hat{f}_R(\cdot)$, is trained using the respondent data $\{(x_i, y_i): i \in R\}$. The model sequentially fits shallow regression trees to current residuals and combines them through shrinkage. Hyperparameters such as the number of trees, interaction depth, learning rate, and minimum node size should be selected by cross-validation within the respondent set.

3.2. Proposed Imputation Scheme and Estimator

$$y_i^* = y_i, i \in R;\ \ y_i^* = \hat{f}_R(x_i),\ \ i \in M. \qquad (2)$$

$$\bar{\hat{y}}_{ML} = \frac{1}{n}\left[\sum_{i\in R} y_i + \sum_{i\in M} \hat{f}_R(x_i)\right]. \qquad (3)$$

Equation (3) is the sample mean of the completed data. It preserves observed responses and uses the joint predictive information in all auxiliary variables for the missing outcomes. If there is no nonresponse, $m = 0$ and the estimator reduces to the ordinary sample mean.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

3.3. Derivation of Prediction Error and Bias

For a fixed auxiliary vector, decompose the fitted predictor into the true regression function, conditional prediction bias, and zero-mean training variation:

$$\hat{f}_R(x_i) = f(x_i) + b_i + u_i,\ E_R(u_i / x_i) = 0. \tag{4}$$

Because $y_i = f(x_i) + \varepsilon_i$, the imputation error for a non-respondent is $\hat{f}_R(x_i) - y_i = b_i + u_i - \varepsilon_i$. Adding and subtracting the unobserved sample values gives the exact completed-mean representation:

$$\bar{\hat{Y}}_{ML} = \bar{y}_s + \frac{1}{n}\sum_{i \in M}\left[\hat{f}_R(x_i) - y_i\right]. \tag{5}$$

$$\bar{\hat{Y}}_{ML} - \bar{Y} = \left(\bar{y}_s - \bar{Y}\right) + \left(\bar{b}_M + \bar{u}_M - \bar{\varepsilon}_M\right). \tag{6}$$

Under SRSWOR, $E\left(\bar{y}_s - \bar{Y}\right) = 0$. Under MAR and the super-population model, $E\left(\bar{\varepsilon}_M\right) = 0$; by construction, $E\left(\bar{u}_M\right) = 0$. Taking expectations in (6) therefore yields:

$$Bias\left(\bar{\hat{Y}}_{ML}\right) = \left(\frac{m}{n}\right)\bar{B}_M,\ \ since\ E\left(\bar{b}_M\right) = \bar{B}_M. \tag{7}$$

Thus, the estimator is exactly unbiased when the average GBM prediction bias is zero. More generally, its bias is attenuated by the nonresponse fraction. If the GBM predictor is consistent, $\sup_X |b_R(X)|$ approaches zero as *r* increases, and the estimator is asymptotically unbiased under the maintained assumptions.

3.4. Derivation of Mean Squared Error

Let $A = \bar{y}_s - \bar{Y}$ *and* $D = \bar{b}_M + \bar{u}_M - \bar{\varepsilon}_M$. Squaring both sides of equation (6) and taking the expectations of the result gives:

$$MSE\left(\bar{\hat{Y}}_{ML}\right) = E\left(A^2\right) + 2\left(\frac{m}{n}\right)E(AD) + \left(\frac{m^2}{n^2}\right)E\left(D^2\right). \tag{8}$$

Under the working orthogonality approximation $E(AD) = 0$. Further, $E\left(D^2\right)$ equals the sum of the squared average prediction bias, mean prediction variance, and mean residual variance. Under SRSWOR, $E\left(A^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2$, define $\tau_M^2 = \frac{1}{m}\sum_{i \in M} Var\{u_R(x_i)\}$ and use $Var\left(\bar{u}_M\right) \approx \frac{\tau_M^2}{m}$ and $Var\left(\bar{\varepsilon}_M\right) \approx \frac{\delta^2}{m}$. The approximate MSE is therefore:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

$$MSE\left(\bar{\hat{Y}}_{ML}\right) \approx \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2+\left(\frac{m}{n^2}\right)\left(\delta^2+\tau_M^2\right)+\left(\frac{m^2}{n^2}\right)\bar{B}_M^2. \quad (9)$$

If δ^2 is represented as $S_y^2\left(1-R_f^2\right)$, where R_f^2 is the fraction of outcome variation explained by the flexible prediction function, equation 9 separates the usual sampling component from three imputation-related quantities: unexplained residual variation, GBM training variation, and squared prediction bias. Consequently, a high-quality predictor lowers MSE by increasing R_f^2 while controlling τ_M^2 and $\bar{B}_f^2$.

3.5. Efficiency Condition

For any competing estimator T_j with MSE_j, the GBM estimator is more efficient when $MSE\left(\bar{\hat{Y}}_{ML}\right) < MSE_j$. Let $C=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2$ and $Q=\delta^2+\bar{\tau}_M^2+m\bar{B}_M^2$. Substitution into (9) produces the general condition:

$$Q < \left(\frac{n^2}{m}\right)\left(MSE_j - C\right). \quad (10)$$

The condition is applicable to mean, ratio, regression, compromised, power-transformation, and recently optimised imputation estimators by inserting their respective MSE expressions. It also clarifies when GBM may fail: weak auxiliary variables inflate δ^2, overfitting inflates $\bar{\tau}_M^2$, and distributional shift between respondents and nonrespondents inflates $\bar{B}_M^2$. A higher response rate reduces m/n and usually improves the quality of GBM training, lowering all imputation penalties.

3.6. Simulation Design and Evaluation Criteria

A Monte Carlo Simulation with 10,000 replications was used to compare the proposed estimator with ten existing methods. Six auxiliary variables were incorporated into the GBM. Four data conditions were considered: linear association; nonlinear association; contamination by outliers; and skewed distributions. Response-rate sensitivity was assessed at 0.50, 0.70, and 0.90. The principal measures were empirical bias, variance, MSE, root mean square error MSE, MSE rank, and Percentage Relative Efficiency (PRE).

$$Bias\left(T\right)=\frac{1}{B}\sum_{b=1}^{B}\left(T_b-\bar{Y}\right), \quad MSE\left(T\right)=\frac{1}{B}\sum_{b=1}^{B}\left(T_b-\bar{Y}\right)^2. \quad (11)$$

$$PRE\left(T:\bar{\hat{Y}}_{ML}\right)=100\times\frac{MSE\left(T\right)}{MSE\left(\bar{\hat{Y}}_{ML}\right)}. \quad (12)$$

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

With this orientation, PRE greater than 100 indicates that the comparator has a larger MSE than the proposed estimator. Model conclusions were based primarily on MSE rather than bias alone because an estimator with small bias may still have high variance.

4. Results

4.1. Best-Performing Estimators

Table 1. Best estimator under each data condition

Condition	Best estimator	True mean	Mean estimate	Bias	Variance	MSE	Rank
Linear	Proposed GBM	155.788	155.838	0.050	2.096	1.890	1
Nonlinear	Proposed GBM	247.755	248.753	0.998	12.791	12.494	1
Outlier	Prasad (2017)	163.319	163.870	0.550	4.428	4.585	1
Skewed	Prasad (2017)	161.025	161.468	0.443	2.421	2.350	1

Table 1 shows that the proposed estimator achieved the smallest MSE under both linear and nonlinear conditions. Under the linear condition, its empirical estimate was 155.838 against a true mean of 155.788, giving a bias of only 0.050 and an MSE of 1.890. Under the nonlinear condition, it ranked first with an MSE of 12.494. These results show that the GBM estimator is not restricted to nonlinear data; shrinkage and tree aggregation can also approximate an essentially linear response surface effectively.

Prasad's (2017) estimator ranked first under outlier and skewed conditions, with MSEs of 4.585 and 2.350, respectively. The proposed GBM estimator ranked second in both cases, with corresponding MSEs of 4.889 and 2.488. These results are also shown in Figure 1. The differences were modest: approximately 6.6% under outliers and 5.9% under skewness. Accordingly, the evidence supports strong competitiveness rather than universal dominance of GBM under contaminated distributions.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

4.2. Comparative MSE and Percentage Relative Efficiency

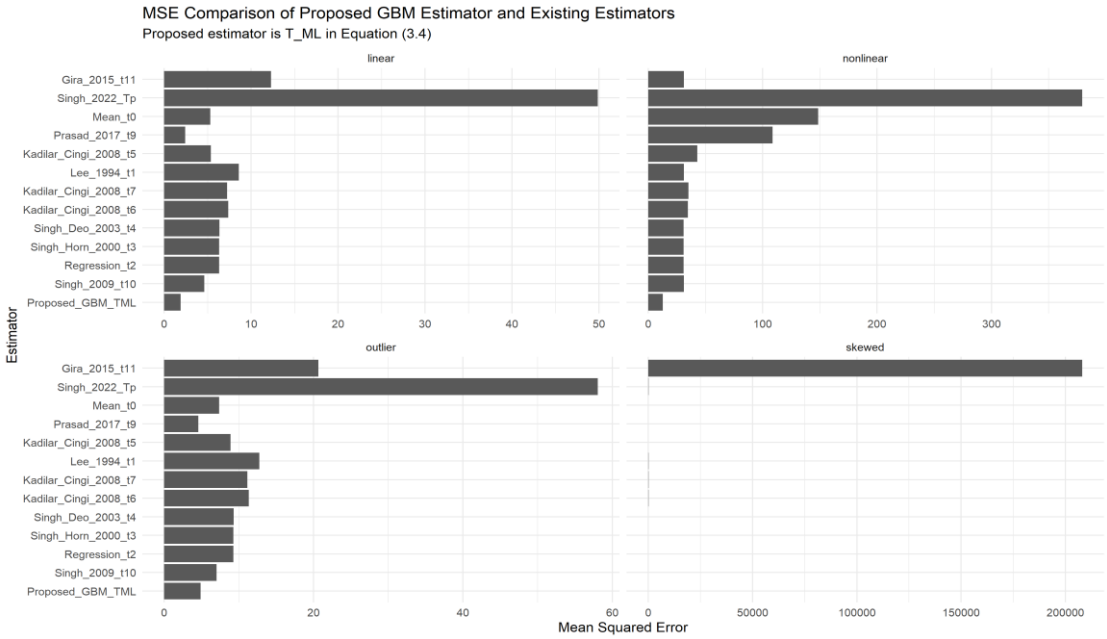

Figure 1. MSE Comparison of the Proposed GBM and Existing Estimators

Figure 1 demonstrates substantial separation between the proposed estimator and several conventional methods. GBM recorded MSE ranks of 1, 1, 2, and 2 for the linear, nonlinear, outlier, and skewed scenarios, respectively. Singh *et al*. (2022) produced comparatively large errors across all four settings, while the Gira (2015) estimator was especially unstable under skewness. The study reports PRE gains of about 160%–227% relative to mean imputation and MSE reductions of approximately 40%–56% relative to regression imputation under nonlinear relationships.

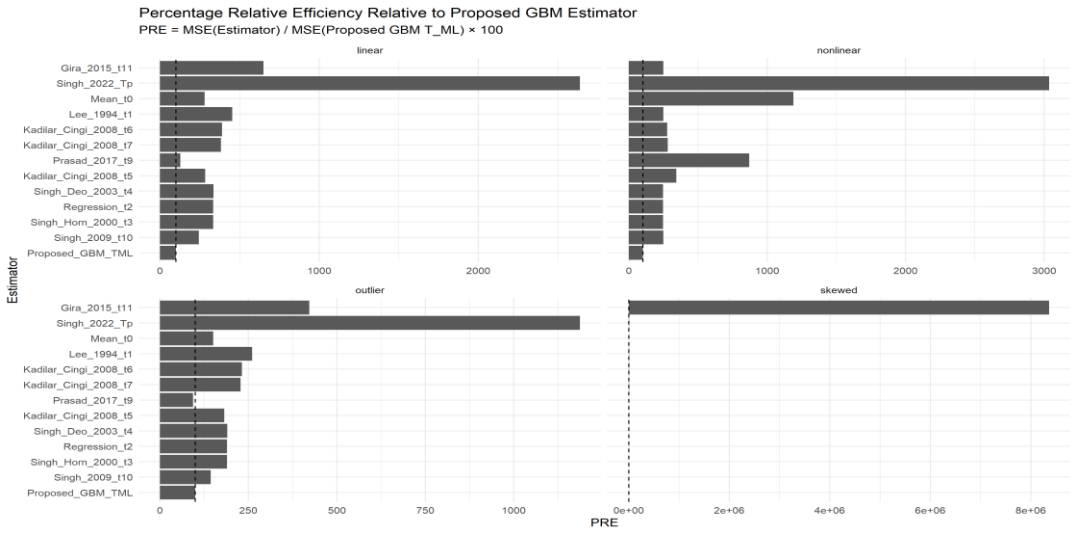

Figure 2. PRE relative to the proposed GBM estimator

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

In Figure 2, the proposed estimator is fixed at PRE = 100. Values above 100 for competing estimators indicate larger MSE and therefore lower efficiency. The largest relative gains occur where traditional procedures impose an inadequate functional form or respond poorly to skewness. In the outlier and skewed scenarios, Prasad's PRE falls slightly below 100, correctly reflecting its marginally smaller MSE.

4.3. Response-Rate Sensitivity

Table 2. MSE of the Proposed GBM Estimator by Response Rate

Condition	50% response	70% response	90% response	MSE reduction: 50%→90%
Linear	4.951	2.610	1.957	60.5
Nonlinear	99.152	28.914	13.812	86.1
Outlier	30.600	11.671	5.011	83.6
Skewed	16.237	5.508	2.285	85.9

The proposed estimator improved monotonically as response increased. From 50% to 90% response, MSE declined by 60.5% in the linear setting, 86.1% in the nonlinear setting, 83.6% under outliers, and 85.9% under skewness. The effect is strongest in the nonlinear case because additional respondents both reduce the number of imputations and give the GBM more information for learning the complex response surface.

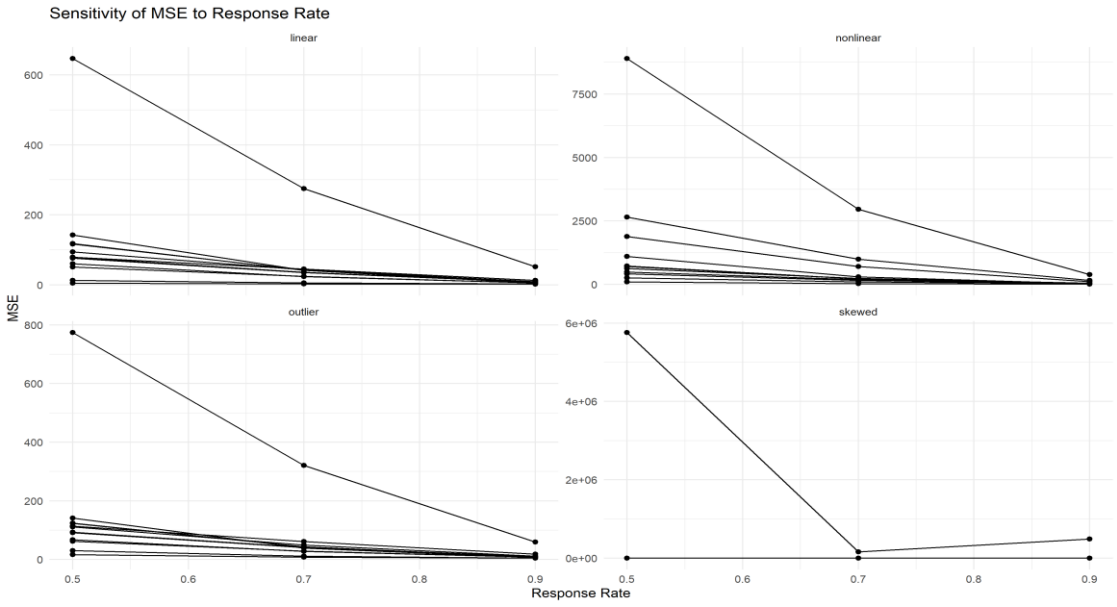

Figure 3. Response-rate sensitivity of the estimators

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

The trend in Figure 3 is consistent with the theoretical MSE in (9). As m decreases, the residual and prediction-variance term $\frac{m\left(\delta^2+\bar{\tau}_M^2\right)}{n^2}$ and the squared-bias term $\frac{m^2\bar{B}_M^2}{n^2}$ contract. The practical implication is that machine-learning imputation does not replace efforts to improve data collection: higher response remains valuable even when flexible imputation is available.

5. Discussion

The findings support the integration of flexible prediction algorithms into model-assisted survey estimation. Under nonlinear relationships, regression imputation cannot fully capture curvature and interactions unless these are specified in advance. GBM constructs an additive ensemble of trees and learns these features directly, explaining its first-place performance and the reported 40%–56% MSE reduction relative to regression imputation. Its strong linear performance also indicates that boosting can approximate simple structures without a large efficiency penalty when appropriately regularised.

The bias derivation explains why predictive tuning matters for statistical inference. The estimator is not unbiased merely because GBM has high predictive accuracy; its bias equals the nonresponse fraction multiplied by average prediction bias among nonrespondents. Cross-validation, representative respondent data, and sufficient overlap in auxiliary information are therefore essential. If response is not missing at random conditional on the available predictors, the residual mean among nonrespondents need not be zero, and the derivation no longer justifies negligible bias.

The outlier and skewed scenarios provide an important qualification. Prasad's estimator obtained slightly smaller MSEs, although GBM remained second. Tree ensembles are often described as robust because a split is local and a single extreme predictor value does not determine a global slope. However, squared-error boosting can still be influenced by extreme outcome residuals. Robust loss functions, outlier diagnostics, winsorization justified by measurement-error evidence, or quantile/Huber boosting may therefore improve performance in contaminated populations.

The results also align with established survey-sampling principles. Särndal *et al*. (1992) emphasised that auxiliary information improves estimation when the working model predicts the study variable well, while Särndal and Lundström (2005) showed the importance of using auxiliary data in nonresponse adjustment. The proposed estimator extends this model-assisted reasoning by replacing a fixed linear predictor with a flexible ensemble. It should consequently be viewed as a survey estimator supported by machine learning, not as a purely algorithmic substitute for probability sampling.

For practical use, the point estimator should be accompanied by a defensible variance estimator. Repeated-sampling procedures such as the bootstrap, jackknife, or balanced repeated replication can retrain the GBM within each replicate, thereby propagating both sampling and model-training uncertainty. Computational cost is higher than for closed-form

imputation estimators, and the apparent black-box character of GBM can be mitigated by reporting variable importance, partial dependence, and tuning details.

6. Conclusion

This study developed a robust GBM imputation estimator for estimating a finite-population mean when the study variable is missing for some sampled units and multiple auxiliary variables are fully observed. Theoretical analysis showed that its bias is governed by the nonresponse fraction and the average ML prediction bias, while its MSE comprises ordinary sampling variation, unexplained outcome variation, prediction variance, and squared prediction bias. The proposed estimator was best under linear and nonlinear conditions and remained close to the best estimator under outlier and skewed conditions. Its MSE declined substantially as the response rate increased.

The evidence supports use of the method when several informative auxiliary variables are available, nonlinearities or interactions are plausible, and the missing-at-random assumption is defensible. Practitioners should select GBM hyperparameters using nested or cross-validation procedures, examine overlap between respondents and nonrespondents, compare against robust benchmarks, and report replication-based standard errors. For highly skewed or contaminated outcomes, robust boosting losses and sensitivity analyses are recommended.

Further research should extend the estimator to stratified, cluster, and unequal-probability designs; study nonignorable response mechanisms; integrate calibration constraints so that imputed estimates reproduce known auxiliary totals; compare GBM with random forests, extreme gradient boosting, and interpretable generalised additive models; and develop multiple-imputation variants that propagate imputation uncertainty into interval estimation.

Declarations

Funding: No external funding information was reported in the source document.

Conflict of interest: The author declares no conflict of interest.

Data availability: The simulation results and analytical outputs supporting the findings are contained in the study records. Data and code may be made available by the author subject to the requirements of the proceedings and any applicable data-use restrictions.

Ethical approval: The study is methodological and simulation-based; no personally identifiable human-subject data are reported in this manuscript.

Acknowledgements

The author acknowledges Dr. O. O. Ishaq and A. A. Osi of the Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, for their academic support.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.012

References

Breiman, L. (2001). Random forests. *Machine Learning, 45(1), 5-32.*

Gira, A. A. (2015). Estimation of population mean with a new imputation method. *Applied Mathematical Sciences*, 9, 1663-1672.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). John Wiley & Sons.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.).*

Kadilar, C., and Cingi, H. (2008). Estimators for the population mean in the case of missing data. Communications in Statistics. *Theory and Methods, 37*, 2226-2236.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology, 12(1),* 1-16.

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.

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data (3rd ed.).* John Wiley & Sons.

Prasad, S. (2017). Ratio exponential type estimators with imputation for missing data in sample surveys. *Model Assisted Statistics and Applications, 12*, 95-106.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63(3),* 581–592.

Särndal, C.-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.

Singh, G. N., Jaiswal, A. K., Singh, C., & Usman, M. (2022). An improved alternative method of imputation for missing data in survey sampling. *Journal of Statistics Applications & Probability, 11(2),* 535-543.

Singh, K., Singh, A., Priyanka, & Singh, K. V. (2014). *Exponential-type compromised imputation in survey sampling.* Journal of Statistics Applications & Probability, *3(2),* 211-221.

Singh, S., & Deo, B. (2003). Imputation by power transformation. *Statistical Papers, 44(4),* 555–579.

Singh, S., & Horn, S. (2000). Compromised imputation in survey sampling. *Metrika, 51(3),* 267–276.