



On the estimation of finite population variance for a mail survey design in the presence of non-response using new conventional and calibrated estimators

Ahmed Audu, Ran Vijay Kumar Singh, Olatunji Olawoyin Ishaq, Supriya Khare, Rajesh Singh & Adedayo Amos Adewara

To cite this article: Ahmed Audu, Ran Vijay Kumar Singh, Olatunji Olawoyin Ishaq, Supriya Khare, Rajesh Singh & Adedayo Amos Adewara (2022): On the estimation of finite population variance for a mail survey design in the presence of non-response using new conventional and calibrated estimators, Communications in Statistics - Theory and Methods, DOI: [10.1080/03610926.2022.2093906](https://doi.org/10.1080/03610926.2022.2093906)

To link to this article: <https://doi.org/10.1080/03610926.2022.2093906>



Published online: 07 Jul 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



On the estimation of finite population variance for a mail survey design in the presence of non-response using new conventional and calibrated estimators

Ahmed Audu^a, Ran Vijay Kumar Singh^b, Olatunji Olawoyin Ishaq^c, Supriya Khare^d, Rajesh Singh^d, and Adedayo Amos Adewara^e

^aDepartment of Statistics, Usmanu Danfodiyo University, Sokoto, Nigeria; ^bDepartment of Mathematics, Kebbi State Uni. Sci. and Tech, Aliero, Nigeria; ^cDepartment of Statistics, Kano University of Science and Technology, Wudil, Nigeria; ^dDepartment of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India; ^eDepartment of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria

ABSTRACT

In this paper, we proposed new conventional unbiased and calibration estimators for estimating finite population variance for a Mail Survey Design characterized by the presence of non-response in practice. The properties of the proposed estimators are studied theoretically and numerically. Empirical studies were conducted using two each of existing and simulated data to illustrate the performance of proposed estimators over existing ones. The results of both theoretical and numerical comparison depicted the superiority of proposed estimators for estimating population variance for a Mail Survey Design characterized by non-response.

ARTICLE HISTORY

Received 26 July 2021
Accepted 20 June 2022

KEYWORDS

Variance estimator;
calibrated estimator;
auxiliary information; mail
survey; efficiency

1. Introduction

For almost all questionnaire designs, the problem of non-response in human-related survey sampling is prevalent. Non-response occurs when certain respondents in the survey decline to answer a full survey or when interviewers fail to follow up with non-respondents in the survey. The missing information induces the non-response bias, which often proves to be a significant source of error in the survey. It should be noted that the nature of the mail survey involves postal service, e-mail surveys and online surveys and, thus, the study of the associated problems is relevant. Therefore, in practical situation, the importance of mail surveys using internet services can never be ignored. As low-cost alternatives for data collection, government, business organizations, telemarketers, and other researchers continually shift to mail and online surveys. As a result of consequences resulting from survey non-response for basic decision making in various fields, previous researchers have, attempted to find different ways to prevent non-response during survey designs or to minimize their effect after the data has been collected and processed. For example, in the past researchers explored procedures in self-administered mail surveys to reduce non-response error as in Hansen and Hurwitz (1946), Fillion (1975), as well as procedures for sample

adjustment due to non-response and error of respondents as in Alwin (1977). Different methods have been developed and studied to minimize the problem of non-response; for instance, see: Church (1993), DeLeeuw (1992), Bishop et al. (1988), Messer (2009), Virtanen, Sirkia, and Jokiranta (2007), Armstrong (1975), Maheux, Legault, and Lambert (1989), Bergk et al. (2005), and Gendall and Healey (2008), Audu et al. (2021, 2022). Little and Rubin (2002) and Lohr (2010) may be referred to for specifics of the different terminologies commonly used in the missing data problem including Missing absolutely at random (MCAR), Missing at random (MAR) and Not-Missing at random (NMAR). In this paper, we have considered only the MCAR situation. Recently, Sinha and Kumar (2015) studied the concept used by Okafor and Lee (2000) to define an unbiased estimator of population variance when the study variable is characterized by the presence of non-response.

Calibration approach have been utilized by several authors like Tracy, Singh, and Arnab (2003), Singh (2003), Estavao and Samdal (2016), Sarndal (2007), Kim, Sungur, and Heo (2007), Kim and Park (2010), Clement and Enang (2015), Rao, Tekabu, and Khan (2016), Koyuncu and Kadilar (2016), Rao, Khan, and Khan (2012), Ozgul (2018), Audu, Singh, et al. (2020), Audu, Danbaba, et al. (2020), Audu, Singh, and Khare (2021) to enhance the efficiency of estimators for estimating population mean in different sampling schemes. Recently, Dykes et al. (2015) utilized the concept of calibration to enhance the precision of Hansen and Hurwitz (1946) method for estimating population mean in the presence of non-response. In this paper, new unbiased and calibration estimators for estimating population variance of mail survey characterized by presence of non-response were considered using the concept of maximum likelihood estimation method.

2. Notations and methodology

Let Y and X be study and auxiliary variables of population of interest with pair units (X_i, Y_i) , $i = 1, 2, 3, \dots, N$. Assume that N_1 and N_2 are the number of units in the population such that $N_1 + N_2 = N$, belonging to the response class C_R and non-response class C_{NR} , respectively. Let n_1 and n_2 be the random sample size from C_R and C_{NR} , respectively, such that $n_1 + n_2 = n$, the total fixed sample size for the study. Let h_2 denote the size of the sub-sample from the n_2 non-respondents who respond on being re-interviewed so that $n_2 = kh_2$, where $k = \frac{n_2}{h_2} > 1$. For the study variable Y , let $\bar{Y} = N^{-1} \sum_{i=1}^N Y_i$, $S_Y^2 = (N-1)^{-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$ and $\beta_2(y) = \frac{m_{40}}{m_{20}^2}$ be the population mean, variance and kurtosis, respectively, where $m_{pq} = N^{-1} \sum_{i=1}^N (Y_i - \bar{Y})^p (X_i - \bar{X})^q$ and let $\bar{Y}_2 = N_2^{-1} \sum_{i=1}^{N_2} Y_{2i}$, $S_{Y_2}^2 = (N_2 - 1)^{-1} \sum_{i=1}^{N_2} (Y_{2i} - \bar{Y}_2)^2$ and $\beta_2(y_2) = \frac{m_{40(2)}}{m_{20(2)}^2}$ be the population mean, variance and kurtosis within the C_{NR} , respectively, where $m_{pq(2)} = N_2^{-1} \sum_{i=1}^{N_2} (Y_{2i} - \bar{Y}_2)^p (X_{2i} - \bar{X}_2)^q$. Similarly, for an auxiliary variable X , let $\bar{X} = N^{-1} \sum_{i=1}^N X_i$, $S_X^2 = (N-1)^{-1} \sum_{i=1}^N (X_i - \bar{X})^2$ and $\beta_2(x) = \frac{m_{04}}{m_{02}^2}$ be the population mean, variance and kurtosis, respectively, and let $\bar{X}_2 = N_2^{-1} \sum_{i=1}^{N_2} X_{2i}$, $S_{X_2}^2 = (N_2 - 1)^{-1} \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2$ and $\beta_2(x_2) = \frac{m_{04(2)}}{m_{02(2)}^2}$ be the population mean, variance and

kurtosis within the C_{NR} , respectively. For study variable Y , let y_{1i} be the response of i^{th} , $i = 1, 2, 3, \dots, n_1$ respondent sample unit drawn by SRSWOR method from the response group C_R and y_{h_2i} be the response of i^{th} , $i = 1, 2, 3, \dots, h_2$ respondent sub-sample unit drawn by SRSWOR method from the non-response group C_{NR} who respond after been re-interviewed. Let $\bar{y}_1 = n_1^{-1} \sum_{i=1}^{n_1} y_{1i}$ and $s_{y1}^2 = (n_1 - 1)^{-1} \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2$ be the sample mean and variance based on sample units of size n_1 selected by SRSWOR method from C_R , let $\bar{y}_{h_2} = h_2^{-1} \sum_{i=1}^{h_2} y_{h_2i}$ and $s_{y_{h_2}}^2 = h_2^{-1} \sum_{i=1}^{h_2} (y_{h_2i} - \bar{y}_{h_2})^2$ be the sample mean and maximum likelihood estimator of population variance based on sub-sample units of size h_2 . Similarly, for an auxiliary variable X , let x_{1i} be the value of i^{th} , $i = 1, 2, 3, \dots, n_1$ respondent sample unit drawn by SRSWOR method from the response group C_R and x_{h_2i} be the value of i^{th} , $i = 1, 2, 3, \dots, h_2$ respondent sub-sample unit drawn by SRSWOR method from the non-response group C_{NR} who respond after been re-interviewed. Let $\bar{x}_1 = n_1^{-1} \sum_{i=1}^{n_1} x_{1i}$ and $s_{x1}^2 = (n_1 - 1)^{-1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$ be the sample mean and variance based on sample units of size n_1 selected by SRSWOR method from C_R , let $\bar{x}_{h_2} = h_2^{-1} \sum_{i=1}^{h_2} x_{h_2i}$ and $s_{x_{h_2}}^2 = h_2^{-1} \sum_{i=1}^{h_2} (x_{h_2i} - \bar{x}_{h_2})^2$ be the sample mean and maximum likelihood estimator of population variance based on sub-sample units of size h_2 .

3. Some existing estimators

Recently, Sinha and Kumar (2015) followed the concept utilized by Okafor and Lee (2000) in defining the sample variance of an auxiliary variable in the presence of non-response to define the following unbiased estimator for estimating population variance when the study variable is characterized by non-response as;

$$s_{y(*)}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n_1} y_{1i}^2 + k \sum_{i=1}^{h_2} y_{h_2i}^2 - n \left(\frac{n_1 \bar{y}_1 + n_2 \bar{y}_{h_2}}{n} \right)^2 \right) \quad (3.1)$$

$$\text{var}(s_{y(*)}^2) = \theta_1 S_y^4 (\beta_2(y) - 1) + \theta_2 S_{y_2}^4 (\beta_2(y_2) - 1) \quad (3.2)$$

where $\theta_1 = n^{-1} - N^{-1}$, $\theta_2 = \frac{N_2(k-1)}{Nn}$, $k = \frac{n_2}{h_2} > 1$

The conventional ratio and regression for estimating population variance in this case under non-response are defined as in (3.3) and (3.4), respectively;

$$t_{RT(*)} = s_{y(*)}^2 s_{x(*)}^{-2} S_x^2 \quad (3.3)$$

$$t_{RG(*)} = s_{y(*)}^2 + \alpha (S_x^2 - s_{x(*)}^2) \quad (3.4)$$

where α is the real-value function to obtained by minimizing the variance of estimator

$$t_{RG(*)} \text{ and } s_{x(*)}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n_1} x_{1i}^2 + k \sum_{i=1}^{h_2} x_{h_2i}^2 - n \left(\frac{n_1 \bar{x}_1 + n_2 \bar{x}_{h_2}}{n} \right)^2 \right),$$

The variances of the estimators $t_{RT(*)}$ and $t_{RG(*)}$ are given as in (3.5) and (3.6), respectively.

$$\begin{aligned} \text{var}(t_{RT(*)}) = & \theta_1 \left(S_y^4(\beta_2(y) - 1) + R^2 S_x^4(\beta_2(x) - 1) - 2RS_y^2 S_x^2 \left(\frac{m_{22}}{m_{20}m_{02}} - 1 \right) \right) \\ & + \theta_2 \left(S_{y_2}^4(\beta_2(y_2) - 1) + R^2 S_{x_2}^4(\beta_2(x_2) - 1) - 2RS_{y_2}^2 S_{x_2}^2 \left(\frac{m_{22(2)}}{m_{20(2)}m_{02(2)}} - 1 \right) \right) \end{aligned} \quad (3.5)$$

where $R = \frac{S_y^2}{S_x^2}$

$$\text{var}(t_{RG(*)}) = \left(\frac{\theta_1 S_y^4(\beta_2(y) - 1) + \theta_2 S_{y_2}^4(\beta_2(y_2) - 1)}{\left(\theta_1 S_y^2 S_x^2 \left(\frac{m_{22}}{m_{20}m_{02}} - 1 \right) + \theta_2 S_{y_2}^2 S_{x_2}^2 \left(\frac{m_{22(2)}}{m_{20(2)}m_{02(2)}} - 1 \right) \right)^2} \right) \quad (3.6)$$

$$\alpha = \frac{\theta_1 S_y^2 S_x^2 \left(\frac{m_{22}}{m_{20}m_{02}} - 1 \right) + \theta_2 S_{y_2}^2 S_{x_2}^2 \left(\frac{m_{22(2)}}{m_{20(2)}m_{02(2)}} - 1 \right)}{\theta_1 S_x^4(\beta_2(x) - 1) + \theta_2 S_{x_2}^4(\beta_2(x_2) - 1)} \quad (3.7)$$

4. Estimation of population variance using new conventional estimator

4.1. Proposed estimators

Having studied Sinha and Kumar (2015) estimator, based on the assumptions defined in section 2, we proposed alternative variance estimator of population variance under non-response as;

$$s_{y(NR)}^2 = \frac{(n_1 - 1)s_{y1}^2 + n_2 s_{yh_2}^2}{n - 1} \quad (4.1)$$

4.2. Properties of the proposed estimators

Theorem 4.1. *The proposed alternative estimators $s_{y(NR)}^2$ is unbiased.*

Proof of Theorem 4.1:

$$E(s_{y(NR)}^2) = E_1 E_2 \left(\frac{(n_1 - 1)s_{y1}^2 + n_2 s_{yh_2}^2}{n - 1} \right) = E_1 \frac{(n_1 - 1)E_2(s_{y1}^2) + n_2 E_2(s_{yh_2}^2)}{n - 1} \quad (4.2)$$

$$E(s_{y(NR)}^2) = \frac{(n_1 - 1)E_1(s_{y1}^2) + n_2 E_1(s_{y_2}^2)}{n - 1} = \frac{(n_1 - 1)S_y^2 + n_2 S_y^2}{n - 1} = S_y^2 \quad (4.3)$$

Hence, the proof

Theorem 4.2. *The variance of $s_{y(NR)}^2$ is*

$$\text{var}(s_{y(NR)}^2) = \theta_1 S_y^4(\beta_2(y) - 1) + \theta_3 S_{y_2}^4 \left(\frac{(n_2 - 2)(n_2 - 3)}{n^2 - 1} \beta_2(y_2) - 1 + \frac{3(n_2 - 1)}{n_2 + 1} \right) \quad (4.4)$$

where $\theta_1 = \frac{1}{n} - \frac{1}{N}$, $\theta_3 = \frac{N_2 n(k-1)}{N(n-1)^2}$

Proof of Theorem 4.2:

$$\text{var}\left(s_{y(NR)}^2\right) = V_1\left(E_2\left(s_{y(NR)}^2 \mid n_1, n_2\right)\right) + E_1\left(V_2\left(s_{y(NR)}^2 \mid n_1, n_2\right)\right) \quad (4.5)$$

$$\begin{aligned} \text{var}\left(s_{y(NR)}^2\right) &= V_1\left(E_2\left(\frac{(n_1 - 1)s_{y1}^2 + n_2 s_{yh2}^2}{n - 1} \mid n_1, n_2\right)\right) \\ &\quad + E_1\left(V_2\left(\frac{(n_1 - 1)s_{y1}^2 + n_2 s_{yh2}^2}{n - 1} \mid n_1, n_2\right)\right) \end{aligned} \quad (4.6)$$

$$\text{var}\left(s_{y(NR)}^2\right) = V_1\left(s_y^2\right) + E_1\left(\frac{n_2^2}{(n - 1)^2} V_2\left(s_{yh2}^2\right)\right) \quad (4.7)$$

$$\text{var}\left(s_{y(NR)}^2\right) = \theta_1 S_y^4 (\beta_2(y) - 1) + E_1\left(\frac{n_2^2}{(n - 1)^2} \left(\frac{1}{h_2} - \frac{1}{n_2}\right) s_{y2}^2 \left(\hat{\beta}_2(y_{n_2}) - 1\right)\right) \quad (4.8)$$

$$\text{var}\left(s_{y(NR)}^2\right) = \theta_1 S_y^4 (\beta_2(y) - 1) + \frac{n(k - 1)}{(n - 1)^2} \left(E_1\left(\frac{n_2}{n}\right) E_1\left(s_{y2}^2\right) \left(E_1\left(\hat{\beta}_2(y_{n_2})\right) - 1\right)\right) \quad (4.9)$$

Lemma 1. Singh (2003) defined the unbiased estimators of N_1 and N_2 are, respectively,

$$\hat{N}_1 = \frac{n_1}{n} N \text{ and } \hat{N}_2 = \frac{n_2}{n} N \quad (4.10)$$

Using Lemma 1,

$$E_1\left(\frac{n_2}{n}\right) = \frac{N_2}{N} \quad (4.11)$$

Lemma 2. Cramér (1946) defined the unbiased estimator of population kurtosis based on sample size n as;

$$\nu = \frac{n^2 - 1}{(n - 2)(n - 3)} \frac{\hat{m}_{40}}{\hat{m}_{20}^2} - \frac{3(n - 1)^2}{(n - 2)(n - 3)} \quad (4.12)$$

where $\hat{m}_{pq} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^p (x_i - \bar{x})^q$

Using Lemma 2,

$$E\left(\hat{\beta}_2(y_{n_2})\right) = \frac{(n_2 - 2)(n_2 - 3)}{n_2^2 - 1} \beta_2(y_2) + \frac{3(n_2 - 1)}{(n_2 + 1)} \quad (4.13)$$

(See appendix for the proof of $E(\hat{\beta}_2(y_{n_2}))$)

Substituting (4.13) and (4.9), we obtained (4.4), hence the proof.

Following Cochran (1940) with assumptions that $s_{x(NR)}^2 \neq 0$, $s_{x(NR)}^2 \neq S_x^2$, we defined ratio and difference type estimators of population variance under non-response as in (4.14) and (4.15), respectively;

$$t_{RT(NR)} = s_{y(NR)}^2 \frac{S_x^2}{s_{x(NR)}^2} \quad (4.14)$$

$$t_{RG(NR)} = s_{y(NR)}^2 + \kappa \left(S_x^2 - s_{x(NR)}^2\right) \quad (4.15)$$

where $s_{x(NR)}^2 = \frac{(n_1-1)s_{x1}^2 + n_2 s_{x2}^2}{n-1}$, κ is the real-value function to be obtained by minimizing the MSE of estimator $t_{RG(NR)}$

To obtain the biases and MSEs of the estimators $t_{RT(NR)}$ and $t_{RG(NR)}$, the following error terms Δ_0 and Δ_1 are defined as $\Delta_0 = \frac{s_{y(NR)}^2 - S_y^2}{S_y^2}$, $\Delta_1 = \frac{s_{x(NR)}^2 - S_x^2}{S_x^2}$ such that $s_{y(NR)}^2 = S_y^2(1 + \Delta_0)$, $s_{x(NR)}^2 = S_x^2(1 + \Delta_1)$. Since $s_{y(NR)}^2$ and $s_{x(NR)}^2$ are unbiased of S_y^2 and S_x^2 , respectively, the expectations of Δ_0 and Δ_1 to first order approximation are given in (4.16).

$$\left. \begin{aligned} E(\Delta_0) = E(\Delta_1) = 0, E(\Delta_0^2) &= \frac{\text{var}(s_{y(NR)}^2)}{S_y^4} \\ E(\Delta_1^2) &= \frac{\text{var}(s_{x(NR)}^2)}{S_x^4}, E(\Delta_0\Delta_1) = \frac{\text{cov}(s_{y(NR)}^2, s_{x(NR)}^2)}{S_y^2 S_x^2} \end{aligned} \right\} \quad (4.16)$$

where $\text{var}(s_{x(NR)}^2) = \theta_1 S_x^4 (\beta_2(x) - 1) + \theta_3 S_{x_2}^4 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \beta_2(x_2) - 1 + \frac{3(n_2-1)}{n_2+1} \right)$, $\text{cov}(s_{y(NR)}^2, s_{x(NR)}^2) = \theta_1 S_y^2 S_x^2 \left(\frac{m_{22}}{m_{20}m_{02}} - 1 \right) + \theta_3 S_{y_2}^2 S_{x_2}^2 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \frac{m_{22(2)}}{m_{20(2)}m_{02(2)}} - 1 + \frac{3(n_2-1)}{n_2+1} \right)$

By expressing $t_{RT(NR)}$ and $t_{RG(NR)}$ as functions of Δ_i , $i = 0, 1$ up to first order approximation, we obtained (4.17) and (4.18), respectively, as

$$t_{RT(NR)} - S_y^2 = S_y^2 (\Delta_0 - \Delta_1 + \Delta_1^2 - \Delta_0\Delta_1) \quad (4.17)$$

$$t_{RG(NR)} - S_y^2 = S_y^2 \Delta_0 - \kappa S_x^2 \Delta_1 \quad (4.18)$$

By squaring both sides of (4.17) and (4.18) and then take the expectation of the results, we obtained the MSEs of $t_{RT(NR)}$ and $t_{RG(NR)}$, respectively, up to first order approximation as in (4.19) and (4.20), respectively.

$$\begin{aligned} MSE(t_{RT(NR)}) &= \theta_1 S_y^4 (\beta_2(y) - 1) + \theta_3 S_{y_2}^4 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \beta_2(y_2) - 1 + \frac{3(n_2-1)}{n_2+1} \right) \\ &\quad + R^2 \left(\theta_1 S_x^4 (\beta_2(x) - 1) + \theta_3 S_{x_2}^4 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \beta_2(x_2) - 1 + \frac{3(n_2-1)}{n_2+1} \right) \right) \\ &\quad - 2R \left(\theta_1 S_y^2 S_x^2 \left(\frac{m_{22}}{m_{20}m_{02}} - 1 \right) + \theta_3 S_{y_2}^2 S_{x_2}^2 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \frac{m_{22(2)}}{m_{20(2)}m_{02(2)}} - 1 + \frac{3(n_2-1)}{n_2+1} \right) \right) \end{aligned} \quad (4.19)$$

where $R = \frac{S_y^2}{S_x^2}$

$$\begin{aligned} MSE(t_{RT(NR)}) &= \theta_1 S_y^4 (\beta_2(y) - 1) + \theta_3 S_{y_2}^4 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \beta_2(y_2) - 1 + \frac{3(n_2-1)}{n_2+1} \right) \\ &\quad - \frac{\left(\theta_1 S_y^2 S_x^2 \left(\frac{m_{22}}{m_{20}m_{02}} - 1 \right) + \theta_3 S_{y_2}^2 S_{x_2}^2 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \frac{m_{22(2)}}{m_{20(2)}m_{02(2)}} - 1 + \frac{3(n_2-1)}{n_2+1} \right) \right)^2}{\theta_1 S_x^4 (\beta_2(x) - 1) + \theta_3 S_{x_2}^4 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \beta_2(x_2) - 1 + \frac{3(n_2-1)}{n_2+1} \right)} \end{aligned} \quad (4.20)$$

where the optimum value of κ is

$$\kappa = \frac{\theta_1 S_y^2 S_x^2 \left(\frac{m_{22}}{m_{20} m_{02}} - 1 \right) + \theta_3 S_{y_2}^2 S_{x_2}^2 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \frac{m_{22(2)}}{m_{20(2)} m_{02(2)}} - 1 + \frac{3(n_2-1)}{n_2+1} \right)}{\theta_1 S_x^4 (\beta_2(x) - 1) + \theta_3 S_{x_2}^4 \left(\frac{(n_2-2)(n_2-3)}{n^2-1} \beta_2(x_2) - 1 + \frac{3(n_2-1)}{n_2+1} \right)} \quad (4.21)$$

In real life applications, the value of κ can be obtained by using the values of the corresponding unbiased statistics for all the unknown parameters involved in its computation.

4.3. Efficiency comparison

In this section efficiency of the proposed variance estimators are compared with their counterpart theoretically and empirically.

4.3.1. Theoretical comparison

The proposed estimator $s_{y(NR)}^2$ is more efficient than $s_{y(*)}^2$ if

$$\text{var}(s_{y(*)}^2) - \text{var}(s_{y(NR)}^2) > 0 \quad (4.22)$$

(4.19) can further be simplify as

$$\beta_2(y_2) > \frac{(n^2 - 1) \left((n-1)^2 (n_2 + 1) + 2n^2 (n_2 - 2) \right)}{(n_2 + 1) \left((n-1)^2 (n^2 - 1) - n^2 (n_2 - 2)(n_2 - 3) \right)} \quad (4.23)$$

The inequality (4.20) is the efficiency condition of $s_{y(NR)}^2$ over $s_{y(*)}^2$. To verify this condition in real life cases, the value of sample kurtosis from non-respondent group can be used as an estimate of $\beta_2(y_2)$.

4.3.2. Empirical study

In this section, we study the performance of the proposed estimator as compared to estimators considered in Section 3 using two real data set for different rate of sub-sampling. The Data description is as given below.

4.3.2.1. Data Set 1: Source (Sinha and Kumar (2015)). In this data set, we consider 109 village population of urban area under Police Statio Baria, Tehsil Champa, Orissa. The data is obtained from District Census Handbook, 1981, Orissa published by government of India. For numerical calculation 25% (i.e. 27 village) from bottom of the list are considered non-respondents. Here, the total population of the village is taken as study character (y) and number of agriculture labors in the village is considered as auxiliary character (x). The parameters of the population are as follows:

$$\begin{aligned}
N &= 109, \quad n = 35, \quad \bar{Y} = 485.92, \quad \bar{X} = 41.24, \quad S_Y^2 = 101593.31, \quad S_X^2 = 2156.05, \quad \mu_{03} \\
&= 198328.59, \quad \mu_{04} = 34235945.30, \quad \mu_{40} = 67105355398.07, \quad \mu_{21} = 7368525.75, \quad \mu_{22} \\
&= 288564214, \quad S_{Y(2)}^2 = 54070.78, \quad S_{X(2)}^2 = 2314.59, \quad \mu_{03(2)} = 241364.95, \quad \mu_{04(2)} \\
&= 42008577.04, \quad \mu_{40(2)} = 8734465562.43, \quad \mu_{21(2)} = 2509197.09, \quad \mu_{22(2)} = 498531365.8
\end{aligned}$$

4.3.2.2. Data Set 2: Singh and Chaudhary (1986). In this data set we consider 35% (i.e. 16 village) from bottom of the list are non-respondents. Here, Study character (y) denotes area under guava crops and area under fresh fruits as auxiliary character (x). The population parameters are as follows:

$$\begin{aligned}
N &= 47, \quad n = 27, \quad \bar{Y} = 29.75, \quad \bar{X} = 23.83, \quad S_Y^2 = 1474.72, \quad S_X^2 = 742.39, \quad \mu_{03} \\
&= 39249.27, \quad \mu_{04} = 3569542.53, \quad \mu_{40} = 15216459.36, \quad \mu_{21} = 81626.62, \quad \mu_{22} \\
&= 7047068.18, \quad S_{Y(2)}^2 = 83.70, \quad S_{X(2)}^2 = 36.93, \quad \mu_{03(2)} = 128.58, \quad \mu_{04(2)} = 3567.71, \quad \mu_{40(2)} \\
&= 18100.73, \quad \mu_{21(2)} = 326.63, \quad \mu_{22(2)} = 7373.25
\end{aligned}$$

5. Estimation of population variance using calibrated estimators

5.1. Proposed calibration estimator

Equation (4.1) can be written as,

$$s_{y(NR)}^2 = \sum_{i=1}^{n_1} w_1^* (y_i - \bar{y}_1)^2 + \sum_{i=1}^{h_2} w_2^* (y_{h_2 i} - \bar{y}_{h_2})^2 \quad (5.1)$$

where $w_1^* = \frac{1}{n-1}$, $w_2^* = \frac{n_2}{h_2(n-1)}$.

The proposed calibrated estimator of population variance under non-response is given in (5.2).

$$t^* = \sum_{i=1}^{n_1} w_{1i}^* (y_i - \bar{y}_1)^2 + \sum_{i=1}^{h_2} w_{2i}^* (y_{h_2 i} - \bar{y}_{h_2})^2 \quad (5.2)$$

where w_{1i}^* , w_{2i}^* are the new calibration weights to be obtained by minimizing the chi-square distance Z^* defined as;

$$\left. \begin{aligned}
\min Z^* &= \frac{\sum_{i=1}^{n_1} ((n-1)w_{1i}^* - 1)^2 / \phi_{1i}}{2(n-1)} + \frac{\sum_{i=1}^{h_2} ((n-1)(h_2-1)w_{2i}^* - n_2)^2 / \phi_{2i}}{2n_2(n-1)(h_2-1)} \\
&\quad s.t. \sum_{i=1}^{n_1} w_{1i}^* (x_i - \bar{x}_1)^2 + \sum_{i=1}^{h_2} w_{2i}^* (x_{h_2 i} - \bar{x}_{h_2})^2 = S_X^2 \\
&\quad \sum_{i=1}^{n_1} w_{1i}^* + \sum_{i=1}^{h_2} w_{2i}^* = \sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*
\end{aligned} \right\} \quad (5.3)$$

To compute new calibrated weights w_{1i}^* , $i = 1, 2, \dots, n_1$ and w_{2i}^* , $i = 1, 2, \dots, h_2$, we defined Lagrange function L of the formed;

$$L = \frac{\sum_{i=1}^{n_1} ((n-1)w_{1i}^* - 1)^2 / \phi_{1i}}{2(n-1)} + \frac{\sum_{i=1}^{h_2} ((n-1)(h_2-1)w_{2i}^* - n_2)^2 / \phi_{2i}}{2n_2(n-1)(h_2-1)} - \lambda_1 \left(\sum_{i=1}^{n_1} w_{1i}^* (x_i - \bar{x}_1)^2 + \sum_{i=1}^{h_2} w_{2i}^* (x_{h_2i} - \bar{x}_{h_2})^2 - S_X^2 \right) - \lambda_2 \left(\sum_{i=1}^{n_1} w_{1i}^* + \sum_{i=1}^{h_2} w_{2i}^* - \sum_{i=1}^{n_1} w_1^* - \sum_{i=1}^{h_2} w_2^* \right) \quad (5.4)$$

Differentiate (5.4) partially with respect to w_{1i}^* , w_{2i}^* , λ_1 , λ_2 and equates the results to zeros, we obtained (5.5)-(5.8)

$$w_{1i}^* = \frac{1 + \lambda_1 \phi_{1i} (x_{1i} - \bar{x}_1)^2 + \lambda_2 \phi_{1i}}{n-1} \quad (5.5)$$

$$w_{2i}^* = \frac{r(1 + \lambda_1 \phi_{2i} (x_{h_2i} - \bar{x}_{h_2})^2 + \lambda_2 \phi_{2i})}{n-1} \quad (5.6)$$

$$\sum_{i=1}^{n_1} w_{1i}^* (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{h_2} w_{2i}^* (x_{h_2i} - \bar{x}_{h_2})^2 = S_X^2 \quad (5.7)$$

$$\sum_{i=1}^{n_1} w_{1i}^* + \sum_{i=1}^{h_2} w_{2i}^* = \frac{n_1 + rh_2}{(n-1)} \quad (5.8)$$

where $r = \frac{n_2}{h_2-1}$.

Substitute both (5.5) and (5.6) in (5.7) and (5.8), we obtained system of linear equations in (5.9).

$$\left. \begin{aligned} \lambda_1 T_1 + \lambda_2 T_2 &= S_X^2 - s_{NR}^2 \\ \lambda_1 T_2 + \lambda_2 T_3 &= 0 \end{aligned} \right\} \quad (5.9)$$

where

$$\begin{aligned} T_1 &= (n-1)^{-1} \left(\sum_{i=1}^{n_1} \phi_{1i} (x_{1i} - \bar{x}_1)^4 + r \sum_{i=1}^{h_2} \phi_{2i} (x_{h_2i} - \bar{x}_{h_2})^4 \right), \\ T_2 &= (n-1)^{-1} \left(\sum_{i=1}^{n_1} \phi_{1i} (x_{1i} - \bar{x}_1)^2 + r \sum_{i=1}^{h_2} \phi_{2i} (x_{h_2i} - \bar{x}_{h_2})^2 \right), \\ T_3 &= (n-2)^{-1} \left(\sum_{i=1}^{n_1} \phi_{1i} + r \sum_{i=1}^{h_2} \phi_{2i} \right), \end{aligned}$$

Solve for λ_1 and λ_2 , we obtained (5.10).

$$\lambda_1 = \frac{T_3 (S_X^2 - s_{x(NR)}^2)}{(T_1 T_3 - T_2^2)}, \lambda_2 = \frac{-T_2 (S_X^2 - s_{x(NR)}^2)}{(T_1 T_3 - T_2^2)} \quad (5.10)$$

Substitute the expression for λ_1 and λ_2 in (5.5) and (5.6), we obtained new calibration weights as in (5.11) and (5.12).

$$w_{1i} = \frac{1}{n-1} + \frac{\phi_{1i} (T_3 (x_{1i} - \bar{x}_1)^2 - T_2) (S_X^2 - s_{x(NR)}^2)}{(n-1) (T_1 T_3 - T_2^2)} \quad (5.11)$$

Table 1. MSEs and PREs of the estimators using Data Sets I and II.

Estimators	1/5	1/k 1/4	1/3
Data Set I			
$s_{y(*)}^2$	1309552792 (100)	1263581176 (100)	1217609560 (100)
$t_{RT(*)}$	1264166847 (103.5902)	1228677196 (102.8408)	1193187545 (102.0468)
$t_{RG(*)}$	1209161554 (108.3025)	1190902935 (106.1028)	1171451391 (103.9403)
$s_{y(NR)}^2$	1256066505 (104.2582)	1223466461 (103.2788)	1190866417 (102.2457)
$t_{RT(NR)}$	1168928728 (112.0302)	1157248607 (109.1884)	1145568486 (106.2887)
$t_{RG(NR)}$	646959370 (202.4165)	835911674 (151.162)	986721113 (123.3996)
Data Set II			
$s_{y(*)}^2$	216767.9 (100)	216596.5 (100)	216425.2 (100)
$t_{RT(*)}$	21623.6 (1002.46)	21583.55 (1003.526)	21543.5 (1004.596)
$t_{RG(*)}$	21613.24 (1002.94)	21573.21 (1004.007)	21533.17 (1005.078)
$s_{y(NR)}^2$	216699.7 (100.0314)	216545.4 (100.0236)	216391.1 (100.0157)
$t_{RT(NR)}$	19790.44 (1095.316)	20208.69 (1071.799)	20626.93 (1049.236)
$t_{RG(NR)}$	[1] 19789.08 (1095.391)	20205.87 (1071.948)	20622.14 (1049.48)

Table 2. Population used for empirical study.

Population	Auxiliary variable x	Study variable y
I	$X \sim \text{Gamma}(1.2, 1.8)$	$Y_i = B_0 + B_1 X_i + \varepsilon_{i1}, \varepsilon \sim N(0, 1) \quad B_0 = 30, \quad B_1 = 0.2, 0.4, 0.6, \dots, 3.0$
II	$X \sim \exp(4)$	$Y_i = B_0 + B_1 X_i + \varepsilon_{i1}, \varepsilon \sim N(0, 1) \quad B_0 = 30, \quad B_1 = 0.5, 0.1, 0.1, 1.5, \dots, 7.5$

$$w_{2i} = \frac{r}{n-1} + \frac{\phi_{2i}(T_3(x_{h2i} - \bar{x}_{h2})^2 - rT_2)(S_X^2 - s_{x(NR)}^2)}{(n-1)(T_1 T_3 - T_2^2)} \quad (5.12)$$

$$\text{where } s_{x(NR)}^2 = \frac{(n_1-1)s_{x1}^2 + n_2 s_{xh2}^2}{n-1}$$

Therefore, the proposed calibration estimator becomes

$$t^* = s_{y(NR)}^2 + \beta_t (S_X^2 - s_{x(NR)}^2) \quad (5.13)$$

where

$$\beta_t = \frac{\left(\sum_{i=1}^{n_1} \phi_{1i} + r \sum_{i=1}^{h_2} \phi_{2i} \right) \left(\sum_{i=1}^{n_1} \phi_{1i} (x_{1i} - \bar{x}_1)^2 (y_{1i} - \bar{y}_1)^2 + r \sum_{i=1}^{h_2} \phi_{2i} (x_{h2i} - \bar{x}_{h2})^2 (y_{h2i} - \bar{y}_{h2})^2 \right)}{\left(\sum_{i=1}^{n_1} \phi_{1i} (x_{1i} - \bar{x}_1)^2 + r \sum_{i=1}^{h_2} \phi_{2i} (x_{h2i} - \bar{x}_{h2})^2 \right) \left(\sum_{i=1}^{n_1} \phi_{1i} (y_{1i} - \bar{y}_1)^2 + r \sum_{i=1}^{h_2} \phi_{2i} (y_{h2i} - \bar{y}_{h2})^2 \right)} \\ - \frac{\left(\sum_{i=1}^{n_1} \phi_{1i} (x_{1i} - \bar{x}_1)^4 + r \sum_{i=1}^{h_2} \phi_{2i} (x_{h2i} - \bar{x}_{h2})^4 \right) \left(\sum_{i=1}^{n_1} \phi_{1i} + r \sum_{i=1}^{h_2} \phi_{2i} \right)}{\left(\sum_{i=1}^{n_1} \phi_{1i} (x_{1i} - \bar{x}_1)^2 + r \sum_{i=1}^{h_2} \phi_{2i} (x_{h2i} - \bar{x}_{h2})^2 \right)^2}$$

Case 1: When $\phi_{1i} = \phi_{2i} = 1$

$$t_1^* = s_{y(NR)}^2 + \beta_{t1} (S_X^2 - s_{x(NR)}^2) \quad (5.14)$$

Table 3. PRE and MSE of estimators using population I.

β	ρ_{xy}	$s_{y(NR)}^2$	$t_{RT(NR)}$	$t_{RG(NR)}$	t_1^*	t_2^*	t_3^*
0.2	0.1171	100 (4.6747e-06)	32.89218 (1.4212e-05)	146.7344 (3.1859e-06)	262.7894 (1.7789e-06)	144.8137 (3.2281e-06)	18.08717 (0.00002585)
0.4	0.2313	100 (1.1366e-05)	125.411 (9.0629e-06)	159.1967 (7.1396e-06)	350.923 (3.2389e-06)	174.3142 (6.520396e-06)	31.14642 (0.00003649)
0.6	0.3367	100 (2.277088e-05)	449.3827 (5.067146e-06)	208.5725 (1.0917e-05)	441.7367 (5.1549e-06)	207.3291 (1.098296e-05)	44.04085 (0.00005170)
0.8	0.4308	100 (4.0497e-05)	182.233 (2.224828e-05)	294.4884 (1.3752e-05)	537.0691 (7.5404e-06)	243.476 (1.6633e-05)	55.6198 (0.00007281)
1.0	0.5127	100 (6.637167e-05)	123.8213 (5.360279e-05)	434.4286 (1.5278e-05)	637.6272 (1.0409e-05)	282.5859 (2.348725e-05)	65.47894 (0.0001013634)
1.2	0.5827	100 (0.0001024415)	137.37994 (7.456804e-5)	665.769 (1.5387e-05)	743.674 (1.3775e-05)	324.5587 (3.156333e-05)	73.62623 (0.0001391)
1.4	0.6418	100 (0.000150973)	243.9056 (6.189812e-05)	1067.746 (1.4139e-05)	855.272 (1.7652e-05)	369.3221 (4.0878e-05)	80.24952 (0.0001881)
1.6	0.6913	100 (0.000214452)	897.0132 (2.390735e-05)	1826.901 (1.1739e-05)	972.3845 (2.2054e-05)	416.8173 (5.144989e-05)	85.58907 (0.0002505601)
1.8	0.7326531	100 (0.0002955842)	556.0269 (5.316006e-05)	3463.378 (8.5346e-06)	1094.922 (2.6996e-05)	466.993 (6.3295e-05)	89.87823 (0.0003289)
2.0	0.7672	100 (0.0003972947)	422.8881 (9.394795e-05)	7882.887 (5.0399e-06)	1222.766 (3.2491e-05)	519.8016 (7.643198e-05)	93.32071 (0.0004257)
2.2	0.7962	100 (0.0005227281)	3573.696 (1.46271e-05)	26860.19 (1.946107e-06)	1355.783 (3.855543e-05)	575.199 (9.087779e-05)	9.608549 (0.00544024)
2.4	0.8206	100 (0.000675249)	3213.493 (2.101293e-05)	495032.7 (1.364049e-07)	1493.832 (4.520248e-05)	633.1424 (0.0001066504)	9.830901 (0.006868637)
2.6	0.8412	100 (0.0008584409)	3006.559 (2.855227e-05)	123336.6 (6.960149e-07)	1636.764 (5.244743e-05)	693.5911 (0.0001237676)	10.01 (0.008575832)
2.8	0.8587	100 (0.001076108)	2889.258 (3.724513e-05)	21878.89 (4.918474e-06)	1784.435 (6.030524e-05)	756.5051 (0.0001422473)	10.15446 (0.01059739)
3.0	0.8736	100 (0.001332272)	2829.114 (4.709151e-05)	9310.022 (1.431009e-05)	1936.696 (6.8791e-05)	821.8455 (0.0001621074)	10.27108 (0.01297111)

$$\text{where } \beta_{t1} = \frac{(n_1 + rh_2)(n_1\mu_{22} + rh_2\mu_{22}^*) - ((n_1 - 1)s_{x1}^2 + r(h_2 - 1)s_{xh_2}^2)((n_1 - 1)s_{y1}^2 + r(h_2 - 1)s_{yh_2}^2)}{(n_1 + rh_2)(n_1\mu_{04} + rh_2\mu_{04}^*) - ((n_1 - 1)s_{x1}^2 + r(h_2 - 1)s_{xh_2}^2)^2}$$

$$\mu_{pq} = \frac{1}{n_1} \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^p (x_{1i} - \bar{x}_1)^q, \mu_{pq}^* = \frac{1}{h_2} \sum_{i=1}^{h_2} (y_{h_2i} - \bar{y}_{h_2})^p (x_{h_2i} - \bar{x}_{h_2})^q$$

Case 2: When $\phi_{1i} = 1/w_1^*, \phi_{2i} = 1/w_2^*$

$$t_2^* = s_{y(NR)}^2 + \beta_{t2} \left(S_X^2 - s_{x(NR)}^2 \right) \quad (5.15)$$

$$\text{where } \beta_{t2} = \frac{(n_1 + h_2)(n_1\mu_{22} + h_2\mu_{22}^*) - ((n_1 - 1)s_{x1}^2 + (h_2 - 1)s_{xh_2}^2)((n_1 - 1)s_{y1}^2 + (h_2 - 1)s_{yh_2}^2)}{(n_1 + h_2)(n_1\mu_{04} + h_2\mu_{04}^*) - ((n_1 - 1)s_{x1}^2 + (h_2 - 1)s_{xh_2}^2)^2}$$

Case 3: When $\phi_{1i} = 1/(x_{1i} - \bar{x}_1)^2, \phi_{2i} = 1/(x_{h_2i} - \bar{x}_{h_2})^2$

$$t_3^* = s_{y(NR)}^2 + \beta_{t3} \left(S_X^2 - s_{x(NR)}^2 \right) \quad (5.16)$$

$$\text{where } \beta_{t3} = \frac{(n_1\mu_{(0/2)} + rh_2\mu_{(0/2)}^*)(\mu_{20} + r\mu_{20}^*) - (n_1 + rh_2)(n_1\mu_{(2/2)} + rh_2\mu_{(2/2)}^*)}{(n_1\mu_{02} + rh_2\mu_{02}^*)(n_1\mu_{(0/2)} + rh_2\mu_{(0/2)}^*) - (n_1 + rh_2)^2}$$

$$\mu_{(p/q)} = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{(y_{1i} - \bar{y}_1)^p}{(x_{1i} - \bar{x}_1)^q}, \mu_{(p/q)}^* = \frac{1}{h_2} \sum_{i=1}^{h_2} \frac{(y_{h_2i} - \bar{y}_{h_2})^p}{(x_{h_2i} - \bar{x}_{h_2})^q}$$

Table 4. PRE and MSE of estimators using population II.

β	ρ_{xy}	$s^2_{y(NR)}$	$t_{RT(NR)}$	$t_{RG(NR)}$	t_1^*	t_2^*	t_3^*
0.5	0.1037	100 (2.720751e-05)	58.2202 (4.673207e-05)	101.7181 (2.674796e-05)	105.2606 (2.584775e-05)	103.0667 (2.639795e-05)	125.200 (2.173114e-05)
1.0	0.2351	100 (2.319958e-05)	55.18288 (4.204125e-05)	101.4277 (2.287302e-05)	105.8292 (2.192171e-05)	103.2123 (2.247752e-05)	127.7558 (1.815932e-05)
1.5	0.3548174	100 (1.907439e-05)	50.73184 (3.759847e-05)	96.60892 (1.974393e-05)	104.2419 (1.82982e-05)	101.1105 (1.886489e-05)	131.242 (1.453376e-05)
2.0	0.4593	100 (1.496586e-05)	44.80297 (3.340372e-05)	86.30399 (1.734087e-05)	99.885 (1.498309e-05)	96.16906 (1.556203e-05)	136.2285 (1.098585e-05)
2.5	0.5477823	100 (1.102267e-05)	37.41951 (2.945701e-05)	70.42962 (1.565062e-05)	91.99144 (1.198228e-05)	87.68395 (1.257091e-05)	143.8753 (7.661264e-06)
3.0	0.6210459	100 (7.40834e-06)	28.76093 (2.575834e-05)	50.53618 (1.465948e-05)	79.645 (9.301701e-06)	74.88078 (9.893512e-06)	156.9587 (4.71993e-06)
3.5	0.6810	100 (4.301189e-06)	19.28117 (2.230771e-05)	29.94996 (1.436125e-05)	61.91142 (6.947327e-06)	57.10689 (7.531822e-06)	184.1222 (2.336051e-06)
4.0	0.7298	100 (1.894349e-06)	9.9154 (1.910512e-05)	12.82452 (1.477131e-05)	38.46267 (4.925162e-06)	34.51911 (5.487826e-06)	271.3587 (6.980977e-07)
4.5	0.7695	100 (3.9576e-07)	2.450441 (1.615057e-05)	2.48289 (1.593949e-05)	12.21012 (3.241246e-06)	10.51569 (3.763519e-06)	4491.977 (8.810376e-09)
5.0	0.8020	100 (2.817222e-08)	0.2095516 (1.344405e-05)	0.1568705 (1.795891e-05)	1.481459 (1.901654e-06)	1.193285 (2.360896e-06)	5.80633 (4.851985e-07)
5.5	0.8286	100 (1.029145e-06)	9.36815 (1.098557e-05)	4.907556 (2.097062e-05)	112.7836 (9.124952e-07)	80.27914 (1.281958e-06)	43.63482 (2.35854e-06)
6.0	0.8507	100 (3.651047e-06)	41.60674 (8.775135e-06)	14.50818 (2.516545e-05)	1304.337 (2.799159e-07)	690.5573 (5.287103e-07)	62.152 (5.874384e-06)
6.5	0.8689	100 (8.161058e-06)	119.7912 (6.812736e-06)	26.51062 (3.078411e-05)	80835.21 (1.009592e-08)	7910.988 (1.03161e-07)	72.26942 (1.129255e-05)
7.0	0.8842	100 (1.484117e-05)	291.096 (5.098375e-06)	38.93619 (3.811663e-05)	13584.52 (1.092506e-07)	202652.1 (7.323469e-09)	78.57826 (1.888711e-05)
7.5	0.8971	100 (2.398817e-05)	660.4577 (3.632052e-06)	50.49986 (4.750146e-05)	4110.165 (5.836302e-07)	9862.963 (2.432146e-07)	82.87087 (2.894644e-05)

5.2. Simulation study

In this section, we conducted a simulation study to examine the superiority of the proposed estimators over other estimators considered in the present study. For this purpose, we generated random populations each of size $N=1000$ for auxiliary variable using Gamma and Exponential distributions and the study populations were generated using linear regression function as defined in Table 2. Sample of sizes $n=100$ is selected 1000 times by method SRSWOR by drawing 60 and 40 samples from the responding and non-responding groups respectively. The sizes of the responding and non-responding groups are $N_1 = 600$ and $N_2 = 400$.

6. Conclusion

In section 4, we proposed alternative unbiased variance estimator of population variance under non-response as $s^2_{y(NR)}$, $t_{RT(NR)}$, $t_{RG(NR)}$. Two real data sets were used to evaluate the efficiency of proposed estimators in comparison to some existing estimators by computing the MSEs and PREs of the estimators and the results are presented in Table 1. Having studied the results obtained from Table 1, we conclude that the proposed estimators $s^2_{y(NR)}$, $t_{RT(NR)}$, $t_{RG(NR)}$ are more efficient than their counterparts $s^2_{y(*)}$, $t_{RT(*)}$, $t_{RG(*)}$ under non-response. In Subsection 5.1, we proposed the members of the proposed estimators for population variance t_1^* , t_2^* , t_3^* under a calibration scheme. To study the performance of

these estimators over conventional estimators, we conducted a simulation study on Gamma and Exponential Populations of size 1000 and the results are summarized in Tables 3 and 4, respectively. From the results of Table 3 and 4, we see that the proposed estimators possess more efficiency and lesser value of MSE for different values of correlation coefficient. The inconsistency for some values of beta may be seen in Tables 3 and 4, as the simulated data were generated from a skewed distribution, especially for the exponential distribution data.

References

- Audu, A., R. Singh, S. Khare, and N. S. Dauran. 2022. Almost unbiased variance estimators under the simultaneous influence of non-response and measurement errors. *Journal of Statistical Theory and Practice* 16 (15):1–14. doi:[10.1007/s42519-021-00226-8](https://doi.org/10.1007/s42519-021-00226-8).
- Audu, A., R. Singh, S. Khare, and N. S. Dauran. 2021. Almost unbiased estimators for population mean in the presence of non-response and measurement error. *Journal of Statistics and Management Systems* 24 (3):573–89. doi:[10.1080/09720510.2020.1759209](https://doi.org/10.1080/09720510.2020.1759209).
- Audu, A., R. Singh, and S. Khare. 2021. Developing calibration estimators for population mean using robust measures of dispersion under stratified random. *Statistics in Transition New Series* 22 (2):125–42. doi:[10.21307/stattrans-2021-019](https://doi.org/10.21307/stattrans-2021-019).
- Audu, A., R. V. K. Singh, S. Muhammed, B. Nakone, and O. O. Ishaq. 2020. On the efficiency of calibration ratio-cum-product estimators of population mean. *Proceeding of Royal Statistics Society Nigeria Local Group* 1:234–46.
- Audu, A., A. Danbaba, A. Abubakar, O. O. Ishaq, and Y. Zakari. 2020. On the efficiency of calibration ratio estimators of population mean in stratified random sampling. *Proceeding of Royal Statistics Society Nigeria Local Group* 1:247–61.
- Alwin, D. F. 1977. Making errors in surveys. *Sociological Methods & Research* 6 (2):131–50. doi:[10.1177/004912417700600201](https://doi.org/10.1177/004912417700600201).
- Armstrong, J. S. 1975. Monetary incentives in mail surveys. *Public Opinion Quarterly* 39 (1): 111–6. doi:[10.1086/268203](https://doi.org/10.1086/268203).
- Cramér, H. 1946. *Mathematical methods of statistics*. Princeton: Princeton University Press.
- Clement, E. P., and E. I. Enang. 2015. Calibration approach alternative ratio estimator for population mean in stratified sampling. *International Journal of Statistics and Economics* 16 (1): 83–93.
- Bergk, V., C. Gasse, R. Schnell, and W. E. Haefeli. 2005. Mail surveys: Obsolescent model or valuable instrument in general practice research? *Swiss Medical Weekly* 135:189–91.
- Bishop, G. G., Hippler, H., and F. Schwarz. 1988. A comparison of response effects in self-administered and telephone surveys. In: Groves, R. M., et. al. Eds., *Telephone Survey Methodology*, New York: Wiley.
- Church, A. H. 1993. Estimating effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly* 57 (1):62–79. doi:[10.1086/269355](https://doi.org/10.1086/269355).
- Cochran, W. G. 1940. Some properties of estimators based on sampling scheme with varying probabilities. *Australian Journal of Statistics* 17:22–8.
- DeLeeuw, E. 1992. Data quality in mail. In *Telephone and face to face surveys*. Amsterdam: TTPublikaties.
- Dykes, L., Singh, S., Stephen, A. S. and V. Louis. 2015. Calibrated Estimators of Population Mean for a Mail Survey Design. *Communications in Statistics - Theory and Methods* 44 (16):3403–27. doi: [10.1080/03610926.2013.841932](https://doi.org/10.1080/03610926.2013.841932).
- Estavao, V. M., and C. E. Samdal. 2016. Survey estimates by calibration on complete auxiliary information. *International Statistical Review* 74:127–47.
- Filion, F. L. 1975. Estimating bias due to non-response in mail surveys. *Public Opinion Quarterly* 39 (4):482–92. doi:[10.1086/268245](https://doi.org/10.1086/268245).

- Gendall, P., and B. Healey. 2008. Alternatives to prepaid monetary incentives in mail surveys. *International Journal of Public Opinion Research* 20 (4):517–27. doi:[10.1093/ijpor/edn048](https://doi.org/10.1093/ijpor/edn048).
- Hansen, M. H., and W. N. Hurwitz. 1946. The problem of non-response in sample surveys. *Journal of the American Statistical Association* 41 (236):517–29. doi:[10.1080/01621459.1946.10501894](https://doi.org/10.1080/01621459.1946.10501894).
- Kim, J. K., and M. Park. 2010. Calibration estimation in survey sampling. *International Statistical Review* 78 (1):21–9. doi:[10.1111/j.1751-5823.2010.00099.x](https://doi.org/10.1111/j.1751-5823.2010.00099.x).
- Kim, J. M., E. A. Sungur, and T. Y. Heo. 2007. Calibration approach estimators in stratified sampling. *Statistics & Probability Letters* 77 (1):99–103. doi:[10.1016/j.spl.2006.05.015](https://doi.org/10.1016/j.spl.2006.05.015).
- Koyuncu, N., and C. Kadilar. 2016. Calibration weighting in stratified random sampling. *Communications in Statistics - Simulation and Computation* 45 (7):2267–75. doi:[10.1080/03610918.2014.901354](https://doi.org/10.1080/03610918.2014.901354).
- Little, R. J. A., and D. B. Rubin. 2002. *Statistical analysis with missing data*. 2nd ed. New York: Wiley Series in Probability and Statistics.
- Lohr, S. L. 2010. *Sampling: Design and analysis*. 2nd ed. Boston: Brooks/Cole.
- Maheux, B., C. Legault, and J. Lambert. 1989. Increasing response rates in physicians' mail surveys: An experimental study. *American Journal of Public Health* 79 (5):638–9. doi:[10.2105/AJPH.79.5.638](https://doi.org/10.2105/AJPH.79.5.638).
- Messer, B. L. 2009. Improving survey response in mail and internet general public survey using address-based sampling and mail contact procedures. A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in Sociology, Washington State University, May 2009. *Sankhya A* 26:69–74.
- Ozgul, N. 2018. New calibration estimator based on two auxiliary variables in stratified sampling. *Communication in Statistics-Theory and Methods* 88 (13):2561–72.
- Okafor, F. C., and H. Lee. 2000. Double sampling for ratio and regression estimation with sub-sampling then non-respondents. *Survey Methodology* 26 (2):183–8.
- Rao, D., M. G. Khan, and S. Khan. 2012. Mathematical programming on multivariate calibration estimation in stratified sampling. *World Academy of Science, Engineering and Technology* 72:12–27.
- Rao, D. K., T. Tekabu, and M. G. M. Khan. 2016. New calibration estimators in stratified sampling. *Asia-Pacific World Congress on Computer Science and Engineering* :66–9.
- Sarndal, C. E. 2007. The calibration approach in survey theory and practice. *Survey Methodology* 33:99–119.
- Singh, S. 2003. *Advanced Sapling theory with applications*. Dorrdrechi: Kluwer Academic Publisher.
- Singh, D., and F. S. Chaudhary. 1986. *Theory and Analysis of Sample Designs*, p. 108. New Delhi: Wiley Eastern Limited.
- Sinha, R., and V. Kumar. 2015. Families of estimators for finite population variance using auxiliary character under double sampling the non-respondents. *National Academy Science Letters* 38 (6):501–5. doi:[10.1007/s40009-015-0380-6](https://doi.org/10.1007/s40009-015-0380-6).
- Tracy, D. S., S. Singh, and R. Arnab. 2003. Note on calibration in stratified and double sampling. *Survey Methodology* 29:99–104.
- Virtanen, V., T. Sirkia, and V. Jokiranta. 2007. Reducing non-response by SMS reminders in mail surveys. *Social Science Computer Review* 25 (3):384–95. doi:[10.1177/0894439307299588](https://doi.org/10.1177/0894439307299588).

Appendices

Appendix I: the proof of $E(\hat{\beta}_2(y_{n_2}))$

$\hat{\beta}_2(y_{n_2})$ can be expressed as in (A1)

$$\hat{\beta}_2(y_{n_2}) = \frac{(n_2 - 2)(n_2 - 3)}{n_2^2 - 1} \left(\frac{(n_2^2 - 1)\hat{\beta}_2(y_{n_2}) - 3(n - 1)^2}{(n_2 - 2)(n_2 - 3)} \right) + \frac{3(n_2 - 1)^2}{(n_2^2 - 1)} \quad (\text{A1})$$

Take expectation of (A1) and using Lemma 2, we obtain (A2) as

$$E\left(\hat{\beta}_2(y_{n_2})\right) = \frac{(n_2 - 2)(n_2 - 3)}{n_2^2 - 1} (\beta_2(y_2)) + \frac{3(n_2 - 1)^2}{(n_2^2 - 1)} \quad (\text{A2})$$

$$E\left(\hat{\beta}_2(y_{n_2})\right) = \frac{(n_2 - 2)(n_2 - 3)}{n_2^2 - 1} (\beta_2(y_2)) + \frac{3(n_2 - 1)}{(n_2 + 1)} \quad (\text{A3})$$

Appendix II: R codes for section 4 numerical illustrations

```

N=109; n=35; Ybar = 485.92; Xbar = 41.24; Sy2=101593; Sx2=2156.05;
m04=34235945.3; m40=67105355398.07; m22=288564214; Sy22=54070.78;
Sx22=2314.59; m042=42008577.04;
m402=8734465562.43; m222=498531365.8; n2=0.25*n; h2=n2*5;
N2=27; N1=N-N2; k=n2/h2;

N=47; n=27; Ybar = 29.75; Xbar = 23.83; Sy2=1474.72; Sx2=742.39;
m04=3569542.53; m40=15216459.36; m22=7047068.18; Sy22=83.7;
Sx22=36.93; m042=3567.71;
m402=18100.73; m222=7373.25; n2=0.35*n; h2=n2/4; R=Sy2/Sx2;
N2=16; N1=N-N2; k=5;

m20=Sy2*(N-1)/N; m02=Sx2*(N-1)/N; m202=Sy22*(N2-1)/N2; m022=Sx22*
(N2-1)/N2;
B2y=m40/m20^2; B2x=m04/m02^2; B2y2=m402/m202^2; B2x2=m042/m022^2;
th1=1/n-1/N; th2=N2*(k-1)/(N*n); th3=N2*n*(k-1)/(N*(n-1)^2);

mm=m22/(m20*m02); mmm=m222/(m202*m022);
a1=th1*(Sy2^2*(B2y-1)+R^2*Sx2^2*(B2x-1)-2*R*Sy2*Sx2*(mm-1));
a2=th2*(Sy22^2*(B2y2-1)+R^2*Sx22^2*(B2x2-1)-2*R*Sy22*Sx22*(mmm-1));
b1=th1*Sy2*Sx2*(mm-1)+th2*Sy22*Sx22*(mmm-1);
b2=th1*Sx2^2*(B2x-1)+th2*Sx22^2*(B2x2-1);
cc=(n2-2)*(n2-3)/(n2-1); ccc=3*(n2-1)/(n2+1);
c2=th1*Sy2*Sx2*(mm-1)+th3*Sy22*Sx22*(mmm-1+ccc);
c1=th1*Sx2^2*(B2x-1)+th3*Sx22^2*(cc*B2x2-1+ccc);

t0=th1*Sy2^2*(B2y-1)+th2*Sy22^2*(B2y2-1);
t1=a1+a2;
t2=t0-b1^2/b2;
t0s=th1*Sy2^2*(B2y-1)+th3*Sy22^2*(cc*B2y2-1+ccc);
t1s=t0s+R^2*c1-2*R*c2;
t2s=t0s-c2^2/c1;

pt0=t0/t0*100;
pt1=t0/t1*100;
pt2=t0/t2*100;
pt0s=t0/t0s*100;
pt1s=t0/t1s*100;
pt2s=t0/t2s*100;

t0;t1;t2;t0s;t1s;t2s;
pt0;pt1;pt2;pt0s;pt1s;pt2s;

```

Appendix III: R codes for the simulation studies in section 5

```

# Data for Exponential Distribution X=rexp(N,4);
# Data for Gamma Distribution: X=gamma(N,1.2,1.8)
set.seed(1000);
library(MASS)
library(e1071)
N=1000; n=100; f=1/n-1/N; b1=7.5; W2=0.4; N2=W2*N; N1=N-N2;
n2=0.4*n; n1=n-n2; h2=0.8*n2; k=n2/h2; f2=N2*(k-1)/(N*n); f3=N2*
n*(k-1)/(N*(n-1)^2);
e=rnorm(N,0,1); X=rexp(N,4); Y=30+b1*X+e;
YX=as.matrix(cbind(Y,X)); YX1=sort(sample(nrow(YX),nrow(YX)*W2));
NRYX=YX[YX1,]; RYX=YX[-YX1,];
Xbar=mean(X); Ybar=mean(Y); Vx=var(X)*N/(N-1); Vy=var(Y)*N/(N-1);
Vx2=var(NRYX[,2])*N/(N-1); Vy2=var(NRYX[,1])*N/(N-1);
Sx=sqrt(Vx); Sy=sqrt(Vy); rho=cor(X,Y,method="pearson");
slop=rho*Sy/Sx; kty=kurtosis(YX[,1]); ktx=kurtosis(YX[,2]);
ktx2=kurtosis(NRYX[,2]);
u40=NA;u20=NA;u04=NA;u02=NA;u22=NA;
for(i in 1:N){
u40[i]=(Y[i]-Ybar)^4; u20[i]=(Y[i]-Ybar)^2; u04[i]=(X[i]-Xbar)^4;
u02[i]=(X[i]-Xbar)^2;
u22[i]=(Y[i]-Ybar)^2*(X[i]-Xbar)^2;
}
l40=sum(u40)/N; l04=sum(u04)/N; l22=sum(u22)/N; l02=sum(u02)/N;
l20=sum(u20)/N;

t0=NA;t1=NA;t2=NA;t3=NA;t4=NA;t5=NA;t6=NA;t7=NA;t8=NA;
for(i in 1:1000){
smp1=c(sample(1:N1,n1,replace=F));
smp2=c(sample(1:N2,h2,replace=F));
ryxdata=RYX[smp1,]; nryxdata=NRYX[smp2,];
ry=ryxdata[,1]; rx=ryxdata[,2] rybar=mean(ry); rxbar=mean(rx);
nry=nryxdata[,1]; nrx=nryxdata[,2] nrybar=mean(nry); nrxbar=mean
(nrx);
v40=NA;v20=NA;v04=NA;v02=NA;v22=NA;
z40=NA;z20=NA;z04=NA;z02=NA;z22=NA;
for(i in 1:n1){
v40[i]=(ry[i]-rybar)^4; v20[i]=(ry[i]-rybar)^2; v04[i]=(rx[i]-rxbar)^4;
v02[i]=(rx[i]-rxbar)^2; v22[i]=(ry[i]-rybar)^2*(rx[i]-rxbar)^2;
}
m40=sum(v40)/n1; m04=sum(v04)/n1; m22=sum(v22)/n1; m02=sum(v02)/n1;
m20=sum(v20)/n1;
for(i in 1:h2){
z40[i]=(nry[i]-nrybar)^4; z20[i]=(nry[i]-nrybar)^2; z04[i]=(nrx[i]
-nrxbar)^4;
z02[i]=(nrx[i]-nrxbar)^2; z22[i]=(nry[i]-nrybar)^2*(nrx[i]-nrxbar)^2;
}
m40s=sum(z40)/h2; m04s=sum(z04)/h2; m22s=sum(z22)/h2; m02s=sum
(z02)/h2;
m20s=sum(z20)/h2;
f02=NA;f22=NA; g02=NA;g22=NA;
for(i in 1:n1){
f02[i]=1/(rx[i]-rxbar)^2; f22[i]=(ry[i]-rybar)^2/(rx[i]-rxbar)^2;

```

```

}
q22=sum(f22)/n1; q02=sum(f02)/n1;
for(i in 1:h2){
g02[i]=1/(nrx[i]-nrxbar)^2;
g22[i]=(nry[i]-nrybar)^2/(nrx[i]-nrxbar)^2;
}
q22s=sum(z22)/h2; q02s=sum(z02)/h2; q20s=sum(z20)/h2;
yyr=NA;xxr=NA;
for(i in 1:n1){
yyr=ry[i]^2; xxr=rx[i]^2;
}
yyrt=sum(yyr); xxrt=sum(xxr);
yynr=NA; xxnr=NA;
for(i in 1:h2){
yynr=nry[i]^2; xxnr=nrx[i]^2;
}
yyprt=sum(yynr); xxprt=sum(xxnr);
l=n2/h2;
ybarh=(n1*rybar+n2*nrybar)/n; xbarh=(n1*rxbar+n2*nrxbar)/n;
ssyh=(yyrt+l*yyprt-n*ybarh^2)/(n-1); ssxh=(xxrt+l*xxprt-n*xbarh^2)/(n-1);
sxx1=var(rx)*n1/(n1-1); syy1=var(ry)*n1/(n1-1); sxx2=var(nrx)*h2/(h2-1);
syy2=var(nry)*h2/(h2-1); ssyNR=((n1-1)*syy1+n2*syy2)/(n-1);
ssxNR=((n1-1)*sxx1+n2*sxx2)/(n-1);

a1=(n1+r*h2)*(n1*m22+r*h2*m22s)-(n1*sxx1+r*h2*sxx2)*(n1*syy1+r*h2*syy2);
a2=(n1+h2)*(n1*m22+h2*m22s)-(n1*sxx1+h2*sxx2)*(n1*syy1+h2*syy2);
a3=(n1*q02+r*h2*q02s)*(m20+r*m20s)-(n1+r*h2)*(n1*q22+r*h2*q22s);
b1=(n1+r*h2)*(n1*m04+r*h2*m04s)-(n1*m02+r*h2*m02s)^2;
b2=(n1+h2)*(n1*m04+h2*m04s)-(n1*m02+h2*m02s)^2;
b3=(n1*m02+r*h2*m02s)*(n1*q02+r*h2*q02s)-(n1+r*h2)^2;
c1=f*Vy*Vx*(m22/(m20*m02)-1)+f2*Vy2*Vx2*(m22s/(m20s*m02s)-1);
d1=f*Vx^2*(ktx-1)+f2*Vx2^2*(ktx2-1);

o1=(n2-2)*(n2-3)/(n^2-1); o2=3*(n2-1)/(n2+1);
c2=f*Vy*Vx*(m22/(m20*m02)-1)+f3*Vy2*Vx2*(o1*m22s/(m20s*m02s)-1+o2);
d2=f*Vx^2*(ktx-1)+f3*Vx2^2*(o1*ktx2-1+o2);

B1=c1/d1; Bt=c2/d2; Bt1=a1/b1; Bt2=a2/b2; Bt3=a3/b3;
t3=(ssyNR-Vy)^2; t4=(ssyNR*Vx/ssxNR-Vy)^2; t5=(ssyNR+Bt*(Vx-ssxNR)-Vy)^2;
t6=(ssyNR+Bt1*(Vx-ssxNR)-Vy)^2; t7=(ssyNR+Bt2*(Vx-ssxNR)-Vy)^2;
t8=(ssyNR+Bt3*(Vx-ssxNR)-Vy)^2;

}
mt3=sum(t3)/1000; mt4=sum(t4)/1000; mt5=sum(t5)/1000; mt6=sum(t6)/1000;
mt7=sum(t7)/1000; mt8=sum(t8)/1000;
p3=mt3/mt3*100; p4=mt3/mt4*100; p5=mt3/mt5*100;
p6=mt3/mt6*100; p7=mt3/mt7*100; p8=mt3/mt8*100;

mt3;mt4;mt5;mt6;mt7;mt8;
p3;p4;p5;p6;p7;p8;

```