Revista Colombiana de Estadística - Theoretical Statistics
July 2024, volume 47, issue 2, pp. 193–210
http://doi.org/10.15446/rce.v47n2.112265

Two-Step Calibrated Designed Weighted Estimators of Finite Population Variance for a Mailed Survey Design Characterized by Non-response

Estimadores ponderados diseñados y calibrados en dos pasos de la varianza de la población finita para un diseño de encuesta enviada por correo caracterizado por la falta de respuesta

A. Audu[1,2,a], M. Lekganyane[3,b], O. O. Ishaq[4,c], K. O. Aremu[1,5,d]

[1]Department of Mathematics and Applied Mathematics, School of Science and Technology, Sefako Makgatho Health Sciences University, Pretoria, South Africa

[2]Department of Statistics, Faculty of Physical and Computing Sciences, Usmanu Danfodiyo University, Sokoto, Nigeria

[3]Department of Statistical Sciences, School of Science and Technology, Sefako Makgatho Health Sciences University, Pretoria, South Africa

[4]Department of Statistics, Faculty of Physical and Computing Sciences, Aliko Dangote University of Science and Technology, Wudil, Nigeria

[5]Department of Mathematics, Faculty of Physical and Computing Sciences, Usmanu Danfodiyo University, Sokoto, Nigeria

Abstract

In this paper, a new class of variance estimators based on a two-step designed weights technique in the presence of non-response is proposed. The proposed estimator is designed to be robust against extreme values or outliers. In the first step, the calibration weights of the new class of estimators are set proportional to the design weights of the existing finite population variance estimator for a mailed survey design characterized by the presence of non-response. In the second step, the constants of proportionality are determined based on different objectives of the investigator such as bias reduction or minimum mean squared error. Many estimators available in the literature can be shown to be special cases of the proposed two-step calibrated estimator. The properties of the proposed estimators are studied

[a]Ph.D. E-mail: ahmed.audu@udusok.edu.ng
[b]M.Sc. E-mail: maria.lekganyane@smu.ac.za
[c]Ph.D. E-mail: ishaq.o.o@kustwudil.edu.ng
[d]Ph.D. E-mail: aremukazeemolalekan@gmail.com

theoretically and numerically. Empirical studies were conducted using ten simulated data to illustrate the performance of proposed estimators over existing ones. The results of the numerical comparison depicted the superiority of two members of the proposed estimator in all cases of data considered.

***Key words*:** Variance estimator; Calibrated estimator; Auxiliary information; Mailed survey; Two-step calibration.

Resumen

En este artículo se propone una nueva clase de estimadores de varianza basados en una técnica de ponderaciones diseñadas en dos pasos en presencia de falta de respuesta. El estimador propuesto está diseñado para ser robusto frente a valores extremos o valores atípicos. En el primer paso, las ponderaciones de calibración de la nueva clase de estimadores se establecen proporcionales a las ponderaciones de diseño del estimador de varianza de población finita existente para un diseño de encuesta enviada por correo caracterizado por la presencia de falta de respuesta. En el segundo paso, las constantes de proporcionalidad se determinan en función de diferentes objetivos del investigador, como la reducción del sesgo o el error cuadrático medio mínimo. Se puede demostrar que muchos estimadores disponibles en la literatura son casos especiales del estimador calibrado de dos pasos propuesto. Se estudian teórica y numéricamente las propiedades de los estimadores propuestos. Se realizaron estudios empíricos utilizando diez datos simulados para ilustrar el desempeño de los estimadores propuestos sobre los existentes. Los resultados de la comparación numérica mostraron la superioridad de dos miembros del estimador propuesto en todos los casos de datos considerados.

***Palabras clave*:** Estimador de varianza; Estimador calibrado; Información auxiliar; Encuesta enviada por correo; Calibración en dos pasos.

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 declined 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 situations, 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 & Hurwitz (1946), and Filion (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), De Leeuw (1992), Bishop et al. (1987), Messer (2009), Virtanen et al. (2007), Armstrong (1975), Maheux et al. (1989), Bergk et al. (2005), Gendall & Healey (2008), Audu, Singh, Khare & Dauran (2021), Audu et al. (2022). Little & Rubin (2019) and Lohr (2021) 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). Recently, Sinha & Kumar (2015) studied the concept used by Okafor & Lee (2000) to define an unbiased estimator of population variance when the study variable is characterized by the presence of non-response. In this paper, we have considered only the MCAR situation.

Calibration approaches have been utilized by several authors like Tracy et al. (2003), Singh (2003), Estevao & Säarndal (2006), Särndal (2007), Kim et al. (2007), Kim & Park (2010), Clement & Enang (2015), Rao et al. (2016), Koyuncu & Kadilar (2016), Rao et al. (2012), Ozgul (2019), Audu, Singh, Muhammed, Nakone & Ishaq (2020), Audu, Danbaba, Abubakar, Ishaq & Zakari (2020), Audu, Singh & Khare (2021) to enhance the efficiency of estimators for estimating the population mean in different sampling schemes.

Recently, Dykes et al. (2015) utilized the concept of calibration to enhance the precision of Hansen & Hurwitz (1946) method for estimating population mean in the presence of non-response. In this paper, new unbiased and calibration estimators for estimating the population variance of mail surveys characterized by the presence of non-response were considered using the concept of the maximum likelihood estimation method.

This manuscript is structured in five sections. Section one consists of an introduction and background to the study. Section two reviewed some related existing estimators as well as definitions of basic terms used. In section three, the procedures for obtaining the new estimator as well as its properties and members of the estimator were discussed. Section four discusses the empirical study which includes data simulation and the computation of MSEs and PREs to assess the efficiency and efficiency gained of the proposed estimator. Discussion of the result and conclusion inferred from the result were presented in section five.

2. Some Related Existing Estimators

Let Y and X be study and auxiliary variables of the population of interest with pair units. Assume that N_1 and N_2 are the numbers of units in the population such that $N_1 + N_2 = N$, belongs 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 $\bar{Y} = N^{-1}\sum_{i=1}^{N} Y_i$, $S_Y^2 = (N-1)^{-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$ 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 $\bar{Y}_2 = N_2^{-1}\sum_{i=1}^{N_2} Y_{2i}$, $S_{y_2}^2 = (N_2-1)^{-1}\sum_{i=1}^{N_2}\left(Y_{2i} - \bar{Y}_2\right)^2$ 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 $m_{pq(2)} = N_2^{-1}\sum_{i=1}^{N_2}\left(Y_{2i} - \bar{Y}_2\right)^p \left(X_{2i} - \bar{X}_2\right)^q$. Similarly, for an auxiliary variable X, let $\bar{X} = N^{-1}\sum_{i=1}^{N} X_i$, $S_X^2 = (N-1)^{-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$ and $\beta_2(x) = \frac{m_{04}}{m_{02}^2}$ be the population mean, variance and kurtosis respectively and let $\bar{X}_2 = N_2^{-1}\sum_{i=1}^{N_2} X_{2i}$, $S_{x_2}^2 = (N_2-1)^{-1}\sum_{i=1}^{N_2}\left(X_{2i} - \bar{X}_2\right)^2$ and $\beta_2(x_2) = \frac{m_{04(2)}}{m_{02(2)}^2}$ be the population mean, variance and kurtosis within the respectively. For study variable Y, let y_{1i} be the response of i^{th}, $i = 1, 2, 3, \ldots, n_1$ respondent sample unit drawn by SRSWOR method from the response group C_R and $y_{h_2 i}$ be the response of i^{th}, $i = 1, 2, 3, \ldots, h_2$ 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 $\bar{y}_{h_2} = h_2^{-1}\sum_{i=1}^{h_2} y_{h_2 i}$ and $s_{yh_2}^2 = h_2^{-1}\sum_{i=1}^{h_2}(y_{h_2 i} - \bar{y}_{h_2})^2$ 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}, $i = 1, 2, 3, \ldots, n_1$ respondent sample unit drawn by SRSWOR method from the response group C_R and $x_{h_2 i}$ be the value of ßth, $i = 1, 2, 3, \ldots, h_2$ 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 $\bar{x}_{h_2} = h_2^{-1}\sum_{i=1}^{h_2} x_{h_2 i}$ and $s_{xh_2}^2 = h_2^{-1}\sum_{i=1}^{h_2}(x_{h_2 i} - \bar{x}_{h_2})^2$ be the sample mean and maximum likelihood estimator of population variance based on sub-sample units of size h_2.

Sinha & Kumar (2015) followed the concept utilized by Okafor & 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 in 1, The variance of the estimator is given in 2

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

$$\operatorname{var}\left(s_{y(*)}^2\right) = \theta_1 S_y^4\left(\beta_2(y) - 1\right) + \theta_2 S_{y_2}^4\left(\beta_2(y_2) - 1\right) \tag{2}$$

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

Sinha & Kumar (2015) defined conventional ratio and regression for estimating population variance in the case under non-response are defined as in 3 and 4 respectively.

$$t_{RT(*)} = s^2_{y(*)} s^{-2}_{x(*)} S^2_x, \tag{3}$$

$$t_{RG(*)} = s^2_{y(*)} + \alpha \left(S^2_x - s^2_{x(*)} \right) \tag{4}$$

where α is the real-value function to obtained by minimizing the variance of estimator $t_{RG(*)}$ and $s^2_{x(*)} = \frac{1}{n-1} \left(\sum_{i=1}^{n_1} x^2_{1i} + k \sum_{i=1}^{h_2} x^2_{h_2 i} - n \left(\frac{n_1 \bar{x}_1 + n_2 \bar{x}_{h_2}}{n} \right)^2 \right)$.

The variances of the estimators $t_{RT(*)}$ and $t_{RG(*)}$ are given as in 5 and 6 respectively.

$$\begin{aligned} \text{var}\left(t_{RT(*)}\right) &= \theta_1 \left(S^4_y \left(\beta_2(y) - 1\right) + R^2 S^4_x \left(\beta_2(x) - 1\right) - 2RS^2_y S^2_x \left(\frac{m_{22}}{m_{20} m_{02}} - 1 \right) \right) \\ &+ \theta_2 \Bigg(S^4_{y_2} \left(\beta_2(y_2) - 1\right) + R^2 S^4_{x_2} \left(\beta_2(x_2) - 1\right) \\ &- 2RS^2_{y_2} S^2_{x_2} \left(\frac{m_{22(2)}}{m_{20(2)} m_{02(2)}} - 1 \right) \Bigg), \end{aligned} \tag{5}$$

$$\text{var}\left(t_{RG(*)}\right) = \left(\begin{array}{c} \theta_1 S^4_y \left(\beta_2(y) - 1\right) + \theta_2 S^4_{y_2} \left(\beta_2(y_2) - 1\right) \\ - \frac{\left(\theta_1 S^2_y S^2_x \left(\frac{m_{22}}{m_{20} m_{02}} - 1 \right) + \theta_2 S^2_{y_2} S^2_{x_2} \left(\frac{m_{22(2)}}{m_{20(2)} m_{02(2)}} - 1 \right) \right)^2}{\theta_1 S^4_x (\beta_2(x) - 1) + \theta_2 S^4_{x_2} (\beta_2(x_2) - 1)} \end{array} \right) \tag{6}$$

where $R = \frac{S^2_y}{S^2_x}$, $\alpha = \frac{\theta_1 S^2_y S^2_x \left(\frac{m_{22}}{m_{20} m_{02}} - 1 \right) + \theta_2 S^2_{y_2} S^2_{x_2} \left(\frac{m_{22(2)}}{m_{20(2)} m_{02(2)}} - 1 \right)}{\theta_1 S^4_x (\beta_2(x) - 1) + \theta_2 S^4_{x_2} (\beta_2(x_2) - 1)}$.

Audu et al. (2024), as an alternative to Sinha & Kumar (2015) estimator, suggested a traditional estimator of population variance in the presence of non-response using the call-back approach to obtain information for estimating statistics of the non-response group as in 7. They demonstrated favourably, the efficiency of their estimators over the estimators of Sinha & Kumar (2015) using real life and simulated data.

$$s^2_{y(NR)} = \sum_{i=1}^{n_1} w^*_1 \left(y_i - \bar{y}_1 \right)^2 + \sum_{i=1}^{h_2} w^*_2 \left(y_{h_2 i} - \bar{y}_{h_2} \right)^2, \tag{7}$$

where $w^*_1 = \frac{1}{n-1}$, $w^*_2 = \frac{n_2}{h_2(n-1)}$.

The proposed estimator in 7 was modified using a calibration approach as in 8 based on the chi-square distance defined in 9

$$t^* = \sum_{i=1}^{n_1} w^*_{1i} \left(y_i - \bar{y}_1 \right)^2 + \sum_{i=1}^{h_2} w^*_{2i} \left(y_{h_2 i} - \bar{y}_{h_2} \right)^2, \tag{8}$$

where w_{1i}^*, w_{2i}^* are the new calibration weights to be obtained by minimizing the chi-square distance Z^* defined as in 9.

$$\left.\begin{array}{ll} \min & Z^* = \frac{\sum_{i=1}^{n_1}((n-1)w_{1i}^*-1)^2/\varphi_{1i}}{2(n-1)} + \frac{\sum_{i=1}^{h_2}((n-1)(h_2-1)w_{2i}^*-n_2)^2/\varphi_{2i}}{2n_2(n-1)(h_2-1)} \\ s.t. & \sum_{i=1}^{n_1} w_{1i}^*\left(x_i-\bar{x}_1\right)^2 + \sum_{i=1}^{h_2} w_{2i}^*\left(x_{h_2 i}-\bar{x}_{h_2}\right)^2 = S_X^2 \\ & \sum_{i=1}^{n_1} w_{1i}^* + \sum_{i=1}^{h_2} w_{2i}^* = \sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^* \end{array}\right\} \tag{9}$$

The calibrated weights w_{1i}^*, $i = 1, 2, \ldots, n_1$ were obtained as in 10 and 11 respectively.

$$w_{1i} = \frac{1}{n-1} + \frac{\varphi_{1i}\left(T_3\left(x_{1i}-\bar{x}_1\right)^2 - T_2\right)\left(S_X^2 - s_{x(NR)}^2\right)}{(n-1)\left(T_1T_3 - T_2^2\right)}, \tag{10}$$

$$w_{2i} = \frac{r}{n-1} + \frac{\varphi_{2i}\left(T_3\left(x_{h_2 i}-\bar{x}_{h_2}\right)^2 - rT_2\right)\left(S_X^2 - s_{x(NR)}^2\right)}{(n-1)\left(T_1T_3 - T_2^2\right)} \tag{11}$$

where $s_{x(NR)}^2 = \frac{(n_1-1)s_{x1}^2 + n_2 s_{xh_2}^2}{n-1}$.

The generalized as well as the members of the calibration estimator t^* were obtained as in 12.

$$t^* = s_{y(NR)}^2 + \beta_t\left(S_X^2 - s_{x(NR)}^2\right), \tag{12}$$

where $\beta_t = \frac{T_1T_2 - T_3T_4}{T_5T_1 - T_3^2}$, $T_1 = \sum_{i=1}^{n_1}\varphi_{1i} + r\sum_{i=1}^{h_2}\varphi_{2i}$,
$T_2 = \sum_{i=1}^{n_1}\varphi_{1i}\left(x_{1i}-\bar{x}_1\right)^2\left(y_{1i}-\bar{y}_1\right)^2 + r\sum_{i=1}^{h_2}\varphi_{2i}\left(x_{h_2 i}-\bar{x}_{h_2}\right)^2\left(y_{h_2 i}-\bar{y}_{h_2}\right)^2$,
$T_3 = \sum_{i=1}^{n_1}\varphi_{1i}\left(x_{1i}-\bar{x}_1\right)^2 + r\sum_{i=1}^{h_2}\varphi_{2i}\left(x_{h_2 i}-\bar{x}_{h_2}\right)^2$,
$T_4 = \sum_{i=1}^{n_1}\varphi_{1i}\left(y_{1i}-\bar{y}_1\right)^2 + r\sum_{i=1}^{h_2}\varphi_{2i}\left(y_{h_2 i}-\bar{y}_{h_2}\right)^2$,
$T_5 = \sum_{i=1}^{n_1}\varphi_{1i}\left(x_{1i}-\bar{x}_1\right)^4 + r\sum_{i=1}^{h_2}\varphi_{2i}\left(x_{h_2 i}-\bar{x}_{h_2}\right)^4$.

Case 1: When $\varphi_{1i} = \varphi_{2i} = 1$, Equation 13 is obtained.

$$, t_1 = s_{y(NR)}^2 + \beta_{t1}\left(S_X^2 - s_{x(NR)}^2\right), \tag{13}$$

where $\beta_{t1} = \frac{(n_1+rh_2)(n_1\mu_{22}+rh_2\mu_{22}^*) - \left((n_1-1)s_{x_1}^2 + r(h_2-1)s_{xh_2}^2\right)\left((n_1-1)s_{y_1}^2 + r(h_2-1)s_{yh_2}^2\right)}{(n_1+rh_2)\left(n_1\mu_{04}+rh_2\mu_{04}^*\right) - \left((n_1-1)s_{x_1}^2 + r(h_2-1)s_{xh_2}^2\right)^2}$,
$\mu_{pq} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_{1i}-\bar{y}_1\right)^p\left(x_{1i}-\bar{x}_1\right)^q$, $\mu_{pq}^* = \frac{1}{h_2}\sum_{i=1}^{h_2}\left(y_{h_2 i}-\bar{y}_{h_2}\right)^p\left(x_{h_2 i}-\bar{x}_{h_2}\right)^q$.

Case 2: When $\varphi_{1i} = 1/w_1^*$, $\varphi_{2i} = 1/w_2^*$, Equation 14 is obtained.

$$t_2 = s_{y(NR)}^2 + \beta_{t2}\left(S_X^2 - s_{x(NR)}^2\right), \tag{14}$$

where $\beta_{t2} = \frac{(n_1+h_2)(n_1\mu_{22}+h_2\mu_{22}^*) - \left((n_1-1)s_{x_1}^2 + (h_2-1)s_{xh_2}^2\right)\left((n_1-1)s_{y_1}^2 + (h_2-1)s_{yh_2}^2\right)}{(n_1+h_2)\left(n_1\mu_{04}+h_2\mu_{04}^*\right) - \left((n_1-1)s_{x_1}^2 + (h_2-1)s_{xh_2}^2\right)^2}$.

Case 3: When $\varphi_{1i} = 1/\left(x_{1i} - \bar{x}_1\right)^2$, $\varphi_{2i} = 1/\left(x_{h_2 i} - \bar{x}_{h_2}\right)^2$, equation 15 is obtained.

$$t_3 = s^2_{y(NR)} + \beta_{t3}\left(S^2_X - s^2_{x(NR)}\right), \tag{15}$$

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

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

3. Proposed Two-Step Calibration Estimator

Let new calibration weights be ψ^*_{1i}, ψ^*_{2i}. To obtain the new calibration weights ψ^*_{1i}, ψ^*_{2i}, we set ψ^*_{1i}, ψ^*_{2i}. to be directly proportional to the design weights as in 16 and 17, respectively.

$$\psi^*_{1i} \propto w^*_1, \tag{16}$$

$$\psi^*_{2i} \propto w^*_2 \tag{17}$$

Since w^*_1, w^*_2 are fixed for all sample units $i \in n_1$ and $i \in h_2$ respectively, then 16 and 17 can be written as in 18 and 19, respectively.

$$\psi^*_{1i} = \theta_{1i} w^*_1, \tag{18}$$

$$\psi^*_{2i} = \theta_{2i} w^*_2, \tag{19}$$

where θ_{1i} and θ_{2i} are the constants of proportionality to be decided by the users based on the available information.

Taking summation of 18 and 19 over sampling units and $i \in n_1$ and $i \in h_2$ respectively, we obtained 20 and 21, respectively.

$$\sum_{i=1}^{n_1}\psi^*_{1i} = \sum_{i=1}^{n_1}\theta_{1i} w^*_1, \tag{20}$$

$$\sum_{i=1}^{h_2}\psi^*_{2i} = \sum_{i=1}^{h_2}\theta_{2i} w^*_2 \tag{21}$$

Using the results of 20 and 21 in 8, two-step calibration estimator of population variance is proposed as in 22.

$$t^*_p = \sum_{i=1}^{n_1}\psi^*_{1i}\left(y_{1i} - \bar{y}_1\right)^2 + \sum_{i=1}^{h_2}\psi^*_{2i}\left(y_{h_2 i} - \bar{y}_{h_2}\right)^2, \tag{22}$$

where ψ^*_{1i}, ψ^*_{2i} are to be obtained by minimizing the chi-square distances Z^{**}_p defined in 23.

$$\left.\begin{array}{ll} \min & Z^{**}_p = \frac{\sum_{i=1}^{n_1}\left(\psi^*_{1i} - w^*_1\right)^2/w^*_1\varphi_{1i}}{2} + \frac{\sum_{i=1}^{h_2}\left(\psi^*_{2i} - w^*_2\right)^2/w^*_2\varphi_{2i}}{2} \\ s.t. & \sum_{i=1}^{n_1}\psi^*_{1i}\left(x_{1i} - \bar{x}_1\right)^2 + \sum_{i=1}^{h_2}\psi^*_{2i}\left(x_{h_2 i} - \bar{x}_{h_2}\right)^2 = S^2_X \\ & \sum_{i=1}^{n_1}\psi^*_{1i} + \sum_{i=1}^{h_2}\psi^*_{2i} = \sum_{i=1}^{n_1}\theta_{1i} w^*_1 + \sum_{i=1}^{h_2}\theta_{2i} w^*_2 \end{array}\right\} \tag{23}$$

To compute new calibrated weights w_{1i}^*, $i = 1, 2, \ldots, n_1$ and w_{2i}^*, $i = 1, 2, \ldots, h_2$, we defined Lagrange function L of the form given in 24.

$$\begin{aligned} L = & \sum_{i=1}^{n_1} \frac{\left(\psi_{1i}^* - w_1^*\right)^2}{2 w_1^* \varphi_{1i}} + \sum_{i=1}^{h_2} \frac{\left(\psi_{2i}^* - w_2^*\right)^2}{2 w_2^* \varphi_{2i}} \\ & - \lambda_1 \left(\sum_{i=1}^{n_1} \psi_{1i}^* \left(x_i - \bar{x}_1\right)^2 + \sum_{i=1}^{h_2} \psi_{2i}^* \left(x_{h_2 i} - \bar{x}_{h_2}\right)^2 - S_X^2 \right) \\ & - \lambda_2 \left(\sum_{i=1}^{n_1} \psi_{1i}^* + \sum_{i=1}^{h_2} \psi_{2i}^* - \sum_{i=1}^{n_1} \theta_{1i} w_1^* - \sum_{i=1}^{h_2} \theta_{2i} w_2^* \right) \end{aligned} \tag{24}$$

Differentiate 24 partially with respect to ψ_{1i}^*, ψ_{2i}^*, λ_1, λ_2 and equates the results to zeros, we obtained 25, 26, 27 and 28, respectively.

$$\psi_{1i}^* = w_1^* + \lambda_1 w_1^* \varphi_{1i} \left(x_i - \bar{x}_1\right)^2 + \lambda_2 w_1^* \varphi_{1i}, \tag{25}$$

$$\psi_{2i}^* = w_2^* + \lambda_1 w_2^* \varphi_{2i} \left(x_{h_2 i} - \bar{x}_{h_2}\right)^2 + \lambda_2 w_2^* \varphi_{2i}, \tag{26}$$

$$\sum_{i=1}^{n_1} \psi_{1i}^* \left(x_{1i} - \bar{x}_1\right)^2 + \sum_{i=1}^{h_2} \psi_{2i}^* \left(x_{h_2 i} - \bar{x}_{h_2}\right)^2 = S_X^2, \tag{27}$$

$$\sum_{i=1}^{n_1} \psi_{1i}^* + \sum_{i=1}^{h_2} \psi_{2i}^* = \sum_{i=1}^{n_1} \theta_{1i} w_1^* + \sum_{i=1}^{h_2} \theta_{2i} w_2^* \tag{28}$$

Substitute both 25 and 26 in 27 and 28, we obtained a system of linear equations in 29.

$$\left. \begin{aligned} \lambda_1 \hat{M}_1 + \lambda_2 \hat{M}_2 &= S_X^2 - s_{x(NR)}^2 \\ \lambda_1 \hat{M}_2 + \lambda_2 \hat{M}_3 &= \sum_{i=1}^{n_1} \left(\theta_{1i} - 1\right) w_1^* + \sum_{i=1}^{h_2} \left(\theta_{2i} - 1\right) w_2^* \end{aligned} \right\}, \tag{29}$$

where $\hat{M}_1 = (n-1)^{-1} \left(\sum_{i=1}^{n_1} \varphi_{1i} \left(x_{1i} - \bar{x}_1\right)^4 + k \sum_{i=1}^{h_2} \varphi_{2i} \left(x_{h_2 i} - \bar{x}_{h_2}\right)^4 \right)$,
$\hat{M}_2 = (n-1)^{-1} \left(\sum_{i=1}^{n_1} \varphi_{1i} \left(x_{1i} - \bar{x}_1\right)^2 + k \sum_{i=1}^{h_2} \varphi_{2i} \left(x_{h_2 i} - \bar{x}_{h_2}\right)^2 \right)$,
$\hat{M}_3 = (n-1)^{-1} \left(\sum_{i=1}^{n_1} \varphi_{1i} + k \sum_{i=1}^{h_2} \varphi_{2i} \right)$.

Solve 29 for λ_1 and λ_2, the expression for λ_1 and λ_2 were obtained as in 30 and 31 respectively.

$$\lambda_1 = \frac{\hat{M}_3 \left(S_X^2 - s_{x(NR)}^2\right) - \hat{M}_2 \left(\sum_{i=1}^{n_1} \left(\theta_{1i} - 1\right) w_1^* + \sum_{i=1}^{h_2} \left(\theta_{2i} - 1\right) w_2^* \right)}{\left(\hat{M}_1 \hat{M}_3 - \hat{M}_2^2\right)}, \tag{30}$$

$$\lambda_2 = \frac{\hat{M}_1 \left(\sum_{i=1}^{n_1} \left(\theta_{1i} - 1\right) w_1^* + \sum_{i=1}^{h_2} \left(\theta_{2i} - 1\right) w_2^* \right) - \hat{M}_2 \left(S_X^2 - s_{x(NR)}^2\right)}{\left(\hat{M}_1 \hat{M}_3 - \hat{M}_2^2\right)} \tag{31}$$

Substitute the expression for λ_1 and λ_2 obtained in 30 and 31 in 25 and 26 respectively, expression for new calibration weights were obtained as in 32 and 33 respectively.

$$\psi_{1i}^{*}=w_1^{*}\left(1+\frac{1}{\hat{M}_1\hat{M}_3-\hat{M}_2^2}\left(\begin{array}{l}\varphi_{1i}\left(\hat{M}_3\left(x_{1i}-\bar{x}_1\right)^2-\hat{M}_2\right)\left(S_X^2-s_{x(NR)}^2\right)+\\ \varphi_{1i}\left(\hat{M}_1-\hat{M}_2\left(x_{1i}-\bar{x}_1\right)^2\right)\left(\begin{array}{l}\sum\limits_{i=1}^{n_1}\left(\theta_{1i}-1\right)w_1^{*}\\ +\sum\limits_{i=1}^{h_2}\left(\theta_{2i}-1\right)w_2^{*}\end{array}\right)\end{array}\right)\right), \tag{32}$$

$$\psi_{2i}^{*}=w_2^{*}\left(1+\frac{1}{\hat{M}_1\hat{M}_3-\hat{M}_2^2}\left(\begin{array}{l}\varphi_{2i}\left(\hat{M}_3\left(x_{h_2i}-\bar{x}_{h_2}\right)^2-\hat{M}_2\right)\left(S_X^2-s_{x(NR)}^2\right)+\\ \varphi_{2i}\left(\hat{M}_1-\hat{M}_2\left(x_{h_2i}-\bar{x}_{h_2}\right)^2\right)\left(\begin{array}{l}\sum\limits_{i=1}^{n_1}\left(\theta_{1i}-1\right)w_1^{*}\\ +\sum\limits_{i=1}^{h_2}\left(\theta_{2i}-1\right)w_2^{*}\end{array}\right)\end{array}\right)\right) \tag{33}$$

Therefore, the proposed calibration estimator becomes 34.

$$t_p^{*}=s_{y(NR)}^2+\hat{b}_{t1}\left(S_X^2-s_{x(NR)}^2\right)+\hat{b}_{t2}\left(\sum_{i=1}^{n_1}\left(\theta_{1i}-1\right)w_1^{*}+\sum_{i=1}^{h_2}\left(\theta_{2i}-1\right)w_2^{*}\right), \tag{34}$$

where $\hat{b}_{t1}=\frac{\hat{M}_3\hat{M}_5-\hat{M}_2\hat{M}_4}{\hat{M}_1\hat{M}_3-\hat{M}_2^2}$, $\hat{b}_{t2}=\frac{\hat{M}_1\hat{M}_4-\hat{M}_2\hat{M}_5}{\hat{M}_1\hat{M}_3-\hat{M}_2^2}$,
$\hat{M}_4=(n-1)^{-1}\left(\sum_{i=1}^{n_1}\varphi_{1i}\left(y_{1i}-\bar{y}_1\right)^2+k\sum_{i=1}^{h_2}\varphi_{2i}\left(y_{h_2i}-\bar{y}_{h_2}\right)^2\right)$,
$\hat{M}_5=(n-1)^{-1}\left(\sum_{i=1}^{n_1}\varphi_{1i}\left(x_{1i}-\bar{x}_1\right)^2\left(y_{1i}-\bar{y}_1\right)^2+k\sum_{i=1}^{h_2}\varphi_{2i}\left(x_{h_2i}-\bar{x}_{h_2}\right)^2\left(y_{h_2i}-\bar{y}_{h_2}\right)^2\right)$.

To obtain the MSE of t_p^{*}, the following error terms e_i, $i=0,1,2,3,4$ were defined as $e_0=\frac{s_{y(NR)}^2-S_y^2}{S_y^2}$, $e_1=\frac{s_{x(NR)}^2-S_x^2}{S_x^2}$, $e_2=\frac{\hat{b}_{t1}-B_{t1}}{B_{t1}}$, $e_3=\frac{\hat{b}_{t2}-B_{t2}}{B_{t2}}$, $e_4=\frac{\sum\limits_{i=1}^{n_1}(\theta_{1i}-1)w_1^{*}+\sum\limits_{i=1}^{h_2}(\theta_{2i}-1)w_2^{*}-\sum\limits_{i=1}^{N_1}(\theta_{1i}-1)-\sum\limits_{i=1}^{N_2}(\theta_{2i}-1)}{\sum\limits_{i=1}^{N_1}(\theta_{1i}-1)+\sum\limits_{i=1}^{N_2}(\theta_{2i}-1)}$ such that $|e_i|\sim 0$, $i=0,1,2,3,4$,
where $B_{t1}=\frac{M_3M_5-M_2M_4}{M_1M_3-M_2^2}$, $B_{t2}=\frac{M_1M_4-M_2M_5}{M_1M_3-M_2^2}$,
$M_2=(N-1)^{-1}\left(\sum_{i=1}^{N_1}\left(X_{1i}-\bar{X}_1\right)^2+\sum_{i=1}^{N_2}\left(X_{2i}-\bar{X}_2\right)^2\right)$,
$M_3=(N-1)^{-1}(N_1+N_2)$,
$M_1=(N-1)^{-1}\left(\sum_{i=1}^{N_1}\left(X_{1i}-\bar{X}_1\right)^4+\sum_{i=1}^{N_2}\left(X_{2i}-\bar{X}_2\right)^4\right)$,
$M_4=(N-1)^{-1}\left(\sum_{i=1}^{N_1}\left(Y_{1i}-\bar{Y}_1\right)^2+\sum_{i=1}^{N_2}\left(Y_{2i}-\bar{Y}_2\right)^2\right)$,
$M_5=(N-1)^{-1}\left(\sum_{i=1}^{N_1}\left(X_{1i}-\bar{X}_1\right)^2\left(Y_{1i}-\bar{Y}_1\right)^2+\sum_{i=1}^{N_2}\left(X_{2i}-\bar{X}_2\right)^2\left(y_{h_2i}-\bar{y}_{h_2}\right)^2\right)$.
Express in terms of e_i, $i=0,1,2,3,4$, we have 35.

$$\begin{aligned}t_p^{*}-S_y^2&=S_y^2e_0-S_x^2B_{t1}e_1-S_x^2B_{t1}e_1e_2\\&\quad+B_{t2}\left(\sum_{i=1}^{N_1}\theta_{1i}+\sum_{i=1}^{N_2}\theta_{2i}-N\right)\left(1+e_3+e_4+e_3e_4\right)\end{aligned} \tag{35}$$

Since $w_1^*,\ w_2^*$ are fixed for all sample units and respectively, then, for simplicity, we assume that $\theta_{1i} = \theta_{2i} = \theta$. Under this assumption, we obtained $e_4 = \frac{\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^* - (N_1+N_2)}{N_1+N_2} = \frac{\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^* - N}{N}$ and 35 becomes 36.

$$t_p^* - S_y^2 = S_y^2 e_0 - S_x^2 B_{t1} e_1 - S_x^2 B_{t1} e_1 e_2 + B_{t2} N (\theta - 1)(1 + e_3 + e_4 + e_3 e_4) \quad (36)$$

Square 36 up to second-degree approximation, we obtained 37.

$$\begin{aligned}\left(t_p^* - S_y^2\right)^2 = {} & S_y^4 e_0^2 + S_x^4 B_{t1}^2 e_1^2 + B_{t2}^2 N^2 (\theta - 1)^2 \left(1 + e_3^2 + e_4^2 + 4e_3 e_4\right) 6 \\ & - 2B_{t1} S_y^2 S_x^2 e_0 e_1 - 2B_{t1} B_{t2} N (\theta - 1) S_x^2 (e_1 e_2 + e_1 e_3 + e_1 e_4) \\ & + 2B_{t2} N (\theta - 1)(e_0 e_3 + e_0 e_4)\end{aligned} \quad (37)$$

Take expectation of 37, we obtained 38.

$$\begin{aligned}MSE\left(t_p^*\right) = {} & \operatorname{var}\left(s_{y(NR)}^2\right) + B_{t1}^2 \operatorname{var}\left(s_{x(NR)}^2\right) - 2B_{t1} \operatorname{cov}\left(s_{x(NR)}^2 s_{y(NR)}^2\right) + B_{t2}^2 N^2 (\theta - 1)^2 \\ & \left(1 + var\left(\hat{b}_{t2}\right)/B_{t2}^2 + \operatorname{var}\left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)/N^2 + 4\operatorname{cov}\left(\hat{b}_{t2}\left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)\right)/B_{t2} N\right) \\ & - 2B_{t1}(\theta - 1) N \operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t2}\right) - 2B_{t1} B_{t2} (\theta - 1) \operatorname{cov}\left(s_{x(NR)}^2 \left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)\right) \\ & - 2B_{t2}(\theta - 1) N \operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t1}\right) + 2N(\theta - 1)\operatorname{cov}\left(s_{y(NR)}^2 \hat{b}_{t2}\right) \\ & + 2B_{t2}(\theta - 1)\operatorname{cov}\left(s_{y(NR)}^2 \left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)\right)\end{aligned} \quad (38)$$

Since the design considered in the study is simple random sampling,
$\operatorname{var}\left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right) = 0$, $\operatorname{cov}\left(\hat{b}_{t2}\left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)\right) = 0$,
$\operatorname{cov}\left(s_{x(NR)}^2\left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)\right) = 0$,
$\operatorname{cov}\left(s_{y(NR)}^2\left(\sum_{i=1}^{n_1} w_1^* + \sum_{i=1}^{h_2} w_2^*\right)\right) = 0$, then 38 becomes 39.

$$\begin{aligned}MSE\left(t_p^*\right) = {} & \operatorname{var}\left(s_{y(NR)}^2\right) + B_{t1}^2 \operatorname{var}\left(s_{x(NR)}^2\right) - 2B_{t1}\operatorname{cov}\left(s_{x(NR)}^2 s_{y(NR)}^2\right) + B_{t2}^2 N^2 (\theta - 1)^2 \\ & \left(1 + var\left(\hat{b}_{t2}\right)/B_{t2}^2\right) - 2B_{t1}(\theta - 1) N \operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t2}\right) - 2B_{t2}(\theta - 1) N \\ & \operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t1}\right) + 2N(\theta - 1)\operatorname{cov}\left(s_{y(NR)}^2 \hat{b}_{t2}\right)\end{aligned} \quad (39)$$

On setting $\frac{\partial MSE\left(t_p^*\right)}{\partial(\theta - 1)} = 0$, we obtained 40.

$$\theta = 1 + \frac{\left(B_{t1}\operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t2}\right) + B_{t2}\operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t1}\right) - \operatorname{cov}\left(s_{y(NR)}^2 \hat{b}_{t2}\right)\right)}{N\left(B_{t2}^2 + var\left(\hat{b}_{t2}\right)\right)} \quad (40)$$

Substituting 40 in 39, MSE of the proposed estimator is obtained in 41.

$$\begin{aligned}MSE\left(t_p^*\right)_{\min} = {} & \operatorname{var}\left(s_{y(NR)}^2\right) + B_{t1}^2 \operatorname{var}\left(s_{x(NR)}^2\right) - 2B_{t1}\operatorname{cov}\left(s_{x(NR)}^2 s_{y(NR)}^2\right) \\ & - \frac{\left(B_{t1}\operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t2}\right) + B_{t2}\operatorname{cov}\left(s_{x(NR)}^2 \hat{b}_{t1}\right) - \operatorname{cov}\left(s_{y(NR)}^2 \hat{b}_{t2}\right)\right)^2}{B_{t2}^2\left(1 + var\left(\hat{b}_{t2}\right)/B_{t2}^2\right)}\end{aligned} \quad (41)$$

As $n \to N$, $\lim\limits_{n\to N} \hat{b}_{t1} = B_{t1}$, $\lim\limits_{n\to N} \hat{b}_{t2} = B_{t2}$, $\lim\limits_{n\to N} var\left(\hat{b}_{t2}\right) = 0$,
$\lim\limits_{n\to N} \operatorname{cov}\left(s^2_{x(NR)}\hat{b}_{t1}\right) = B^2_{t1}\operatorname{var}\left(s^2_{x(NR)}\right)$,
$\lim\limits_{n\to N} \operatorname{cov}\left(s^2_{x(NR)}\hat{b}_{t2}\right) = B^2_{t2}\operatorname{var}\left(s^2_{x(NR)}\right)$,
$\lim\limits_{n\to N} \operatorname{cov}\left(s^2_{y(NR)}\hat{b}_{t2}\right) = B^2_{t2}\operatorname{var}\left(s^2_{y(NR)}\right)$

Then, the expression for θ becomes as in 42.

$$\theta = 1 + \frac{\left(B_{t1}B^2_{t2}\operatorname{var}\left(s^2_{x(NR)}\right) + B^2_{t1}B_{t2}\operatorname{var}\left(s^2_{x(NR)}\right) - B^2_{t2}\operatorname{var}\left(s^2_{y(NR)}\right)\right)}{NB^2_{t2}} \tag{42}$$

The estimates θ denoted by $\hat{\theta}$ is given by 43.

$$\hat{\theta} = 1 + \frac{\left(\hat{b}_{t1}\left(\hat{b}_{t1} + \hat{b}_{t2}\right)\operatorname{var}\widehat{\left(s^2_{x(NR)}\right)} - \hat{b}_{t2}\operatorname{var}\widehat{\left(s^2_{y(NR)}\right)}\right)}{N\hat{b}_{t2}}, \tag{43}$$

where
$\widehat{var\left(s^2_{x(NR)}\right)} = \theta_1 s^4_{x1}\left(b_2(x_1) - 1\right) + \theta_3 s^4_{xh_2}\left(\frac{(n_2-2)(n_2-3)}{n^2-1}b_2(x_2) - 1 + \frac{3(n_2-1)}{n_2+1}\right)$,
$\widehat{var\left(s^2_{y(NR)}\right)} = \theta_1 s^4_{y1}\left(b_2(y_1) - 1\right) + \theta_3 s^4_{yh_2}\left(\frac{(n_2-2)(n_2-3)}{n^2-1}b_2(y_2) - 1 + \frac{3(n_2-1)}{n_2+1}\right)$
$b_2(x_1) = \frac{(n_1-1)^{-1}\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)^4}{\left((n_1-1)^{-1}\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)^2\right)^2}$, $b_2(y_1) = \frac{(n_1-1)^{-1}\sum_{i=1}^{n_1}(y_{1i}-\bar{y}_1)^4}{\left((n_1-1)^{-1}\sum_{i=1}^{n_1}(y_{1i}-\bar{y}_1)^2\right)^2}$
$b_2(x_2) = \frac{(h_2-1)^{-1}\sum_{i=1}^{h_2}\left(x_{h_2i}-\bar{x}_{h_2}\right)^4}{\left((h_2-1)^{-1}\sum_{i=1}^{h_2}\left(x_{h_2i}-\bar{x}_{h_2}\right)^2\right)^2}$, $b_2(y_2) = \frac{(h_2-1)^{-1}\sum_{i=1}^{h_2}\left(y_{h_2i}-\bar{y}_{h_2}\right)^4}{\left((h_2-1)^{-1}\sum_{i=1}^{h_2}\left(y_{h_2i}-\bar{y}_{h_2}\right)^2\right)^2}$
Substitute 43 in 34, the suggested estimator becomes 44.

$$\begin{aligned} t^*_p = &\, s^2_{y(NR)} + \hat{b}_{t1}\left(S^2_X - s^2_{x(NR)}\right) \\ &+ \tfrac{n}{(n-1)N}\left(\hat{b}_{t1}\left(\hat{b}_{t1} + \hat{b}_{t2}\right)\operatorname{var}\widehat{\left(s^2_{x(NR)}\right)} - \hat{b}_{t2}\operatorname{var}\widehat{\left(s^2_{y(NR)}\right)}\right) \end{aligned} \tag{44}$$

4. Empirical 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, 10 different populations (6 nonlinear and 4linear) of size $N = 1000$ were generated for study and auxiliary variables using distributions defined in Tables 1 and 2. Sample of sizes $n = 100$ is selected 1000 times by method SRSWOR. The sizes of the responding and non-responding groups are $N_1 = 600$ and $N_2 = 400$. The values of MSEs and PREs of the estimators were computed 1000 times and average of the results were obtained and presented in Tables 3 and 4.

The MSEs and Percentage relative efficiencies (PREs) of members of proposed estimator and other estimators as well as classical estimators have been calculated based on population data sets I-IV given in Table 2.

Table 1: Distributions of Non-linear Populations for Empirical Study

Population	**Auxiliary variable** x	**Study variable** y
I	$Z_1 \sim abs\,(Normal\,(10,40))$ $Z_2 \sim gamma\,(1/3, 0.7)$ $X = \rho_{z_1z_2} Z_1 + \sqrt{1-\rho^2_{z_1z_2}} Z_2$	$Y_i = 0.5X_i + 0.5X_i^2 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
II	$Z_1 \sim abs\,(Normal\,(10,40))$ $Z_2 \sim gamma\,(1/3, 0.7)$ $X = \rho_{z_1z_2} Z_1 + \sqrt{1-\rho^2_{z_1z_2}} Z_2$	$Y_i = 0.5X_i^2 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
III	$Z_1 \sim \frac{1}{abs(Normal(10,40))}$ $Z_2 \sim gamma\,(1/3, 0.7)$ $X = \rho_{z_1z_2} Z_1 + \sqrt{1-\rho^2_{z_1z_2}} Z_2$	$Y_i = 0.5X_i + 0.5X_i^2 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
IV	$Z_1 \sim \frac{1}{abs(Normal(10,40))}$ $Z_2 \sim gamma\,(1/3, 0.7)$ $X = \rho_{z_1z_2} Z_1 + \sqrt{1-\rho^2_{z_1z_2}} Z_2$	$Y_i = 0.5X_i^2 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
V	$Z_1 \sim \frac{1}{abs(Normal(10,40))}$ $Z_2 \sim gamma\,(1/3, 0.7)$ $X = \rho_{z_1z_2} Z_1 + \sqrt{1-\rho^2_{z_1z_2}} Z_2$	$Y_i = 0.5X_i^{1/5} + 0.5X_i^2 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
VI	$Z_1 \sim abs\,(Normal\,(10,40))$ $Z_2 \sim gamma\,(1/3, 0.7)$ $X = \rho_{z_1z_2} Z_1 + \sqrt{1-\rho^2_{z_1z_2}} Z_2$	$Y_i = 0.5X_i^{1/5} + 0.5X_i^2 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$

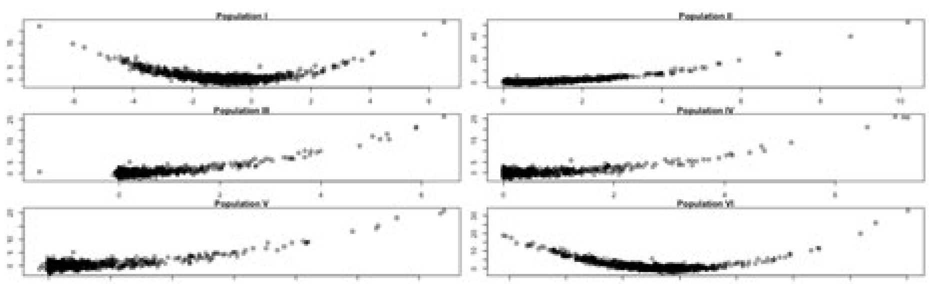

Figure 1: Plots of the Distributions for Non-linear Populations considered in the Study.

The simulation procedure for generating data and computations of MSEs and PRE is presented as follows.

Step 1: Generate population data of size N for auxiliary Variable X using probability functions defined in Table 1 for non-linear populations and Table 2 for linear populations and use the relations in the tables to generate study variable Y.

Step 2: Combine Y and X as pairs (Y, X).

Step 3: Divide the Information on (Y, X) randomly into two non-overlapping groups of sizes N1 (Response group (RG)) and $N2$ (Non-response group (NRG)).

Step 4: Compute parameters of both study and auxiliary variables required for the estimators in computing MSEs from the data in step 4 for response and non-response

Table 2: Distributions of Linear Populations used for Empirical Study

Population	Auxiliary variable x	Study variable y
I	$Z_1 \sim \frac{1}{abs(rnorm(N,10,40))}$ $Z_2 \sim rgamma\,(N, 1/3, 0.7)$ $X = \rho_{z_1 z_2} Z_1 + \sqrt{1-\rho^2_{z_1 z_2}} Z_2$	$Y_i = \rho_{z_1 z_2} X_1 + \sqrt{1-\rho^2_{z_1 z_2}} X_1 + \varepsilon_i$ $\varepsilon \sim N\,(0,1)$
II	$Z_1 \sim rpois\,(N, 1)$ $Z_2 \sim rgamma\,(N, 1/3, 0.7)$ $X = \rho_{z_1 z_2} Z_1 + \sqrt{1-\rho^2_{z_1 z_2}} Z_2$	$Y_i = \rho_{z_1 z_2} X_1 + \sqrt{1-\rho^2_{z_1 z_2}} X_1 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
III	$Z_1 \sim \frac{1}{abs(rnorm(N,10,40))}$ $Z_2 \sim rpois\,(N, 1)$ $X = \rho_{z_1 z_2} Z_1 + \sqrt{1-\rho^2_{z_1 z_2}} Z_2$	$Y_i = \rho_{z_1 z_2} X_1 + \sqrt{1-\rho^2_{z_1 z_2}} X_1 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$
IV	$Z_1 \sim \frac{1}{abs(rnorm(N,10,40))}$ $Z_2 \sim rchisq\,(N, 2)$ $X = \rho_{z_1 z_2} Z_1 + \sqrt{1-\rho^2_{z_1 z_2}} Z_2$	$Y_i = \rho_{z_1 z_2} X_1 + \sqrt{1-\rho^2_{z_1 z_2}} X_1 + \varepsilon_i,$ $\varepsilon \sim N\,(0,1)$

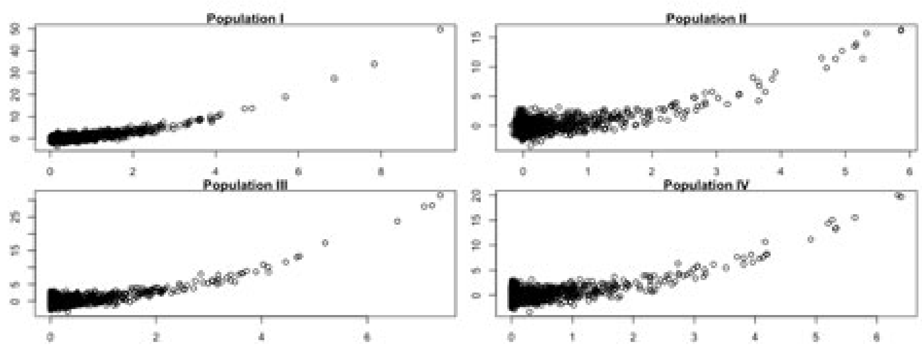

Figure 2: Plots of the Distributions for Linear Populations considered in the Study.

groups.
Step 5: Take a pair of random samples of size $n1$ and $h2$ from RG and NRG by the SRSWOR method.
Step 6: Compute the corresponding sample statistics the estimators require from the two sample data obtained in Step 5.
Step 7: Use the parameters and sample statistics obtained in Step 4 and Step 6 respectively to compute the values of the estimators.
Step 8: Subtract the value of population variance for Y from each value of the estimators and square each of the results.
Step 9: Repeat steps 5 to 8 1,000 times.
Step 10: For each estimator, compute the MSE by averaging corresponding results in step 9 as defined in 45 below.

$$MSE\,(t_i) = \frac{1}{1000} \sum_{i=1}^{1000} \left(t_i - S_Y^2\right)^2 \tag{45}$$

where $t_i$is any estimator in the study

Step 11: Compute PRE for each estimator using the corresponding value of MSE in step 10 as defined in 46.

$$PRE\left(t_i\right) = \frac{MSE\left(t_i\right)}{MSE\left(t_0\right)} \times 100 \tag{46}$$

where $t_0$is the sample variance estimator in the study
Step 12: Display the results of MSE and PRE for each estimator

Table 3: MSEs and PREs of the proposed and existing estimators under non-linear populations in table 1

Estimators	**Pop I.**		**Pop II**	
	MSEs	**PREs**	**MSEs**	**PREs**
Sample var $S^2_{y(NR)}$	0.0110543	100	0.007311865	100
Audu et al. (2024) t^*				
t_1	0.0065961	167.5892	0.004517959	161.84
t_2	0.0065869	167.8224	0.004511843	162.0594
t_3	0.0112174	98.54635	0.007450296	98.14194
Proposed estimators t^*_p				
t^*_1	1883.473	0.0005869	216.1245	0.003383
t^*_2	0.0002843	3887.953	0.0009255868	789.9707
t^*_3	2.115e-05	52266.57	0.0001646828	4439.969
Estimators	**Pop III**		**Pop IV**	
	MSEs	**PREs**	**MSEs**	**PREs**
Sample var $S^2_{y(NR)}$	0.0170276	100	0.0119407	100
Audu et al. (2024) t^*				
t_1	0.0094101	180.9497	0.00688844	173.344
t_2	0.0094006	181.1341	0.006881621	173.5158
t_3	0.0175774	96.87257	0.0123831	96.42738
Proposed estimators t^*_p				
t^*_1	9454.251	0.0001801	1328.129	0.000899
t^*_2	0.0014203	1198.856	0.0001967116	6070.157
t^*_3	0.0042802	397.8197	0.001540973	774.8807
Estimators	**Pop V**		**Pop VI**	
	MSEs	**PREs**	**MSEs**	**PREs**
Sample var $S^2_{y(NR)}$	0.0079877	100	0.01286999	100
Audu et al. (2024) t^*				
t_1	0.0049022	162.9407	0.007361355	174.8318
t_2	0.0048955	163.1629	0.007354019	175.0062
t_3	0.0081311	98.23625	0.01333207	96.53404
Proposed estimators t^*_p				
t^*_1	341.4105	0.0023396	2006.721	0.000641
t^*_2	0.0007824	1020.914	0.0003341724	3851.301
t^*_3	8.7969e-05	9080.057	0.001965261	654.8741

Table 4: MSEs and PREs of the proposed and existing estimators under linear populations in table 2

Estimators	Pop I.		Pop II	
	MSEs	**PREs**	**MSEs**	**PREs**
Sample var $S^2_{y(NR)}$	9.85916	100	9.660349e-05	100
Audu et al. (2024) t^*				
t_1	0.03369	29264.55	1.269929e-04	76.07
t_2	0.03457	28517.10	9.218536e-05	104.7927
t_3	10.20315	96.62857	0.0004358157	22.16613
Proposed estimators t_p^*				
t_1^*	0.02452817	5.0801e-13	5.135344e-05	188.1149
t_2^*	0.01462239	40195.25	0.006821902	1.416079
t_3^*	0.02452817	67425.09	0.006601051	1.463456
Estimators	**Pop III**		**Pop IV**	
	MSEs	**PREs**	**MSEs**	**PREs**
Sample var $S^2_{y(NR)}$	0.0006199	100	0.0005609202	100
Audu et al. (2024) t^*				
t_1	5.73520e-05	1081.036	4.500634e-04	124.6314
t_2	3.29548e-05	1881.349	6.063634e-05	925.0561
t_3	0.00475951	13.02646	0.001808004	31.02428
Proposed estimators t_p^*				
t_1^*	1.56052e-05	3972.998	5.700757e-05	983.9398
t_2^*	1.93122e-05	3210.387	6.634633e-06	8454.426
t_3^*	6.9989e-06	8858.445	6.86205e-06	8174.236

5. Discussions and Conclusion

The paper introduces a new class of variance estimators designed to address the issue of non-response in survey designs. These estimators are developed to be robust against extreme values or outliers in the data. The approach involves a two-step process: in the first step, calibration weights are set proportional to the design weights of an existing finite population variance estimator for a survey with non-response. In the second step, constants of proportionality are determined based on various investigator objectives, such as bias reduction or minimizing mean squared error. The paper demonstrates that many existing estimators in the literature can be considered special cases of the proposed two-step calibrated estimator. The properties of these estimators t_2^* andt_3^* are explored both theoretically and numerically, and empirical studies using simulated data are conducted to evaluate their performance. In section 3 of the paper, ten real data sets are simulated and used to assess the efficiency of the proposed estimators t_1^*, t_2^* andt_3^* compare to existing ones. The evaluation is based on measures such as the mean squared error (MSE) and the percentage relative efficiency (PRE). The results are presented in Table 3 and 4, and the conclusion drawn from the results obtained from Tables 3 and 4 is that the proposed estimators t_1^*, t_2^* andt_3^*, specifically outperform their counterparts under the two scenarios with the exception of t_1^* that performed below existing estimators under non-linear populations. This suggests that the two-step calibrated approach is a promising method for improving the accuracy and efficiency of variance estimation in survey designs, particularly when dealing with non-response.

This study is limited to the use of a two-step calibration transformation approach, however, other approach like power calibration transformation and calibrated maximum likelihood design weight approaches can be used for further study.

Acknowledgements

The Authors would like to thank the Editor-in-Chief and the two anonymous referees for their valuable comments and very constructive suggestions which resulted in significant improvement over the quality of the article.

[Received: December 2023 — Accepted: May 2024]

References

Alwin, D. F. (1977), 'Making errors in surveys: An overview', *Sociological Methods & Research* **6**(2), 131–150.

Armstrong, J. S. (1975), 'Monetary incentives in mail surveys', *The Public Opinion Quarterly* **39**(1), 111–116.

Audu, A., Danbaba, A., Abubakar, A., Ishaq, O. O. & Zakari, Y. (2020), 'on the efficiency of calibration ratio estimators of population mean in stratified random sampling', *Proceeding of Royal Statistics Society Nigeria Local Group* pp. 247–261.

Audu, A., Singh, R. & Khare, S. (2021), 'Developing calibration estimators for population mean using robust measures of dispersion under stratified random sampling', *Statistics in Transition New Series* **22**(2), 125–142.

Audu, A., Singh, R., Khare, S. & Dauran, N. (2021), 'Almost unbiased estimators for population mean in the presence of non-response and measurement error', *Journal of Statistics and Management Systems* **24**(3), 573–589.

Audu, A., Singh, R., Khare, S. & Dauran, N. (2022), 'Almost unbiased variance estimators under the simultaneous influence of non-response and measurement errors', *Journal of Statistical Theory and Practice* **16**(2), 15.

Audu, A., Singh, R. V. K., Ishaq, O. O., Khare, S., Singh, R. & Adewara, A. A. (2024), '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* **53**(3), 848–864.

Audu, A., Singh, R. V. K., Muhammed, S., Nakone, B. & Ishaq, O. O. (2020), 'Efficiency of calibration ratio-cum-product estimators of population mean', *Proceeding of Royal Statistics Society Nigeria Local Group* pp. 234–246.

Bergk, V., Gasse, C., Schnell, R. & Haefeli, W. E. (2005), 'Mail surveys: obsolescent model or valuable instrument in general practice research?', *Swiss Medical Weekly* **135**(1314), 189–189.

Bishop, G. E., Hippler, H.-J., Schwarz, N. & Strack, F. (1987), 'A comparison of response effects in self-administered and telephone surveys'.

Church, A. H. (1993), 'Estimating the effect of incentives on mail survey response rates: A meta-analysis', *Public opinion quarterly* **57**(1), 62–79.

Clement, E. P. & Enang, E. I. (2015), 'Calibration approach alternative ratio estimator for population mean in stratified sampling', *International Journal of Statistics and Economics* **16**(1), 83–93.

De Leeuw, E. D. (1992), *Data quality in mail, telephone and face to face surveys.*, ERIC.

Dykes, L., Singh, S., A. sedory, S. & Louis, V. (2015), 'Calibrated estimators of population mean for a mail survey design', *Communications in Statistics-Theory and Methods* **44**(16), 3403–3427.

Estevao, V. M. & Säarndal, C.-E. (2006), 'Survey estimates by calibration on complex auxiliary information', *International Statistical Review* **74**(2), 127–147.

Filion, F. L. (1975), 'Estimating bias due to nonresponse in mail surveys', *Public Opinion Quarterly* **39**(4), 482–492.

Gendall, P. & Healey, B. (2008), 'Alternatives to prepaid monetary incentives in mail surveys', *International Journal of Public Opinion Research* **20**(4), 517–527.

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.

Kim, J. K. & Park, M. (2010), 'Calibration estimation in survey sampling', *International Statistical Review* **78**(1), 21–39.

Kim, J.-M., Sungur, E. A. & Heo, T.-Y. (2007), 'Calibration approach estimators in stratified sampling', *Statistics & probability letters* **77**(1), 99–103.

Koyuncu, N. & Kadilar, C. (2016), 'Calibration weighting in stratified random sampling', *Communications in Statistics-Simulation and Computation* **45**(7), 2267–2275.

Little, R. J. & Rubin, D. B. (2019), *Statistical analysis with missing data*, Vol. 793, John Wiley & Sons.

Lohr, S. L. (2021), *Sampling: design and analysis*, Chapman and Hall/CRC.

Maheux, B., Legault, C. & Lambert, J. (1989), 'Increasing response rates in physicians' mail surveys: an experimental study', *American Journal of Public Health* **79**(5), 638–639.

Messer, B. L. (2009), Improving survey response in mail and Internet general public surveys using address-based sampling and mail contact procedures, PhD thesis, Washington State University.

Okafor, F. C. & Lee, H. (2000), 'Double sampling for ratio and regression estimation with sub-sampling the non-respondents', *Survey Methodology* **26**(2), 183–188.

Ozgul, N. (2019), 'New calibration estimator based on two auxiliary variables in stratified sampling', *Communications in Statistics-Theory and Methods* **48**(6), 1481–1492.

Rao, D. K., Tekabu, T. & Khan, M. G. (2016), New calibration estimators in stratified sampling, *in* '2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)', IEEE, pp. 66–70.

Rao, D., Khan, M. G. & Khan, S. (2012), 'Mathematical programming on multivariate calibration estimation in stratified sampling', *World Academy of Science, Engineering and Technology* **72**, 58–62.

Särndal, C.-E. (2007), 'The calibration approach in survey theory and practice', *Survey methodology* **33**(2), 99–119.

Singh, S. (2003), *Advanced Sampling Theory With Applications: How Michael Selected Amy*, Vol. 2, Springer Science & Business Media.

Sinha, R. & Kumar, V. (2015), 'Families of estimators for finite population variance using auxiliary character under double sampling the non-respondents', *National Academy Science Letters* **38**, 501–505.

Tracy, D., Singh, S. & Arnab, R. (2003), 'Note on calibration in stratified and double sampling', *Survey Methodology* **29**(1), 99–104.

Virtanen, V., Sirkiä, T. & Jokiranta, V. (2007), 'Reducing nonresponse by sms reminders in mail surveys', *Social Science Computer Review* **25**(3), 384–395.