

Calibration for efficiency of ratio estimator in domains of study with sub-sampling the nonrespondents

Abstract

In sample survey, it is expected that the information would be collected from all the selected units in the sample, but practically, it is generally not possible because of non-response. Some of the units may not respond or may not be contacted during the survey period. This work focuses on domain estimation of population mean with sub-sampling the non-respondents. In this study, we consider calibration technique as a method of correcting non-response in domains of study by minimizing the chi-square distance function between the weight of the main estimator and the calibrated weight subject to the formulated constraint on the auxiliary variable. As a result, two estimators are proposed; these are the ratio estimator for domain mean and a ratio estimator for double sampling. Bias and Mean Square Error (MSE) for the proposed estimators are derived.

We have used an auxiliary variable to estimate the population mean assuming that the non-response is observed only on the study variable. The proposed estimators and the existing estimators were compared empirically in the domains with small sampling units and two populations where considered in terms of the MSE and Percentage Relative Efficiency (PRE). We considered two cases where non-responses are uniform in the two strata at approximately (30%) and a case where the non-response rates are different with 20% and 40% in strata 1 and 2 respectively. The proposed estimators are more efficient than the existing estimators.

Keywords: auxiliary variable, calibration, non-response, ratio estimator, sub-sampling, domain

Volume 13 Issue 2 - 2024

Ikot Ekemini E, Iseh Matthew Joshua

Department of Statistics, Akwa Ibom State University, Mkpato Enin, Nigeria

Correspondence: Iseh Matthew Joshua, Department of Statistics, Akwa Ibom State University, Mkpato Enin, Nigeria, Email mathewise@aksu.edu.ng

Received: April 15, 2024 | **Published:** May 20, 2024

Introduction

It is obvious that society cannot run effectively on the basis of hunches or trial and error. Decisions based on data will provide better results than those based on intuitions or gut feelings. Statistics is a range of procedures for gathering, organizing, analyzing and presenting quantitative data. In the modern society, the need for statistical information seems endless. In particular, data are regularly collected to satisfy the need for information about specified sets of elements, called as finite population. Statistics helps us to turn data into information. One of the most important modes of data collection for satisfying such needs is sample survey, that is, a partial investigation of the finite population and on the basis of such partial information (sample information) one tries to inference about the finite population characteristics (parameters). Sample survey is less expensive than a complete enumeration, it is usually less time consuming, and may even be more accurate than the method of complete enumeration. The term sample is used for the set of units or portion of the aggregate of material which has been selected with the belief that it will be representative of the whole aggregate. The sampling theory deals with scientific and objective procedure of choosing an appropriate sampling design, i.e. selecting a sample from the population which is representative of the population as a whole and also provides suitable estimation procedure to estimate the population parameters. Most challenging about the sample representation of the population is the effect of non-response on the estimation of the population parameter. Different authors have suggested different techniques for a reliable and efficient estimator, among which is the calibration technique.

Calibration estimation in sample surveys has since its introduction by Deville JC, et al.¹ developed an established theory and method for estimation of finite population parameter. Calibration of weights is a technique that uses population data on auxiliary variables to

improve estimates in sample surveys. If auxiliary data are available, some improvement in the precision of estimate may be achieved. Incorporation of auxiliary data in the estimation process is known as calibration. In stratified random sampling, calibration approach is used to obtain optimum strata weights for improving the precision of survey estimates of population parameters. Koyuncu N, et al.² defined some calibration estimators in stratified random sampling for population characteristics and Clement EP, et al.³ applied the concept of calibration estimators for domain totals in stratified random sampling. Clement EP, et al.⁴ combined some scalars with the mean of the auxiliary variable and proposed calibration alternative ratio estimator of mean in stratified sampling.

When a researcher is interested in obtaining information from a local or small area, it becomes challenging with small sample size in some of the areas of interest and even very difficult when non-response occurs. Several authors have made attempts to obtain reliable estimates in such areas of interests popularly called domains of study. Among them is Godwin A, et al.⁵ The author considers modifications of some of the procedures for global ratio estimation in single-phase sampling with sub-sampling the non-respondents proposed by Rao P⁶ to obtain an estimate of mean for a small domain that cuts across constituent strata of a population with unknown weights. The bias and mean-square error of each of the modified estimators were obtained for comparison. However, the estimators were not subjected numerical test to validate the analytical claims most importantly in areas of small/zero sample sizes. Unlike,⁶ the population mean of the auxiliary variable adopted by Godwin A, et al.⁵ is assumed to be unknown before the start of the survey and hence double sampling was applied under stratified simple random sampling.

In a bid to improve on the efficiency of the estimators under non-response,^{7,8} adopted the concept of calibration with a single constraint

to estimate the population mean and the result was encouraging. Cochran WG⁹ showed that knowledge of N_j , of domain j that is of interest reduces the variance of the estimator of domain mean in a single-phase simple random design. The reduction in variance is shown to be greater when the proportion of non-domain elements in the population is large and the study variable varies little among the domain elements. Ashutosh¹⁰ proposed estimators for domain mean utilizing stratified sampling with non-response. The proposed estimator was compared to a direct ratio estimator for domain mean utilizing stratified sampling with non-response. Clement EP, et al.⁴ stated that in the presence of powerful auxiliary variables, the calibration estimation meets the objective of reducing both non-response bias and the sampling error. Etebong P¹¹ develops a new approach to ratio estimation that produces a more efficient class of ratio estimators that do not depend on any optimality conditions for optimum performance using calibration weightings. Iseh MJ, et al.,¹² Iseh MJ, et al.,¹³ Iseh MJ, et al.¹⁴ considered the challenges of population mean estimation in small area that is characterized by small or no sample size and in the presence of unit non-response and presents a calibration estimator that produces reliable estimates under stratified random sampling from a class of synthetic estimators using calibration approach with alternative distance measure. To overcome the challenges of poor performance of the ratio estimator in small area occasioned with small/no sample size as a result of non-response, this work considers the calibration approach using the constraints of equal weights adjustment criteria, unbiased estimator of the population mean and variance of the auxiliary variable.

In this paper, based on the attempt by Godwin A, et al.⁵ who suggested the global ratio estimation in single-phase sampling with sub-sampling the non-respondents to obtain an estimate of mean for a small domain that cuts across constituent strata of a population with unknown weights, a new improved ratio estimator for population mean in stratified random sampling is suggested using the theory of calibration estimation with three constraints to achieve optimal precision and efficiency.

Some existing estimator and theoretical underpinnings

This section considers some existing ratio estimators for estimation of domain population mean and the theoretical underpinnings for the proposed ratio estimator. Though not much have been done in the area of domains of study in the presence of non-response probably due to the intricate nature of the estimation, this paper highlights some existing estimators as applicable to domain estimation which applied the concept of sub-sampling the non-respondents.

Some existing estimator

Study notations and definitions:

- N = population size under study
- N_d = population size for the d^{th} domain
- N_{dh} = population size of h^{th} stratum in d^{th} domain
- n_{dh} = sample size for the d^{th} domain in the h^{th} stratum
- n_{dh} = domain sample Size
- n_{1dh} = sample size for respondent units for the d^{th} domain in the h^{th} Stratum
- n_{2dh} = Sample size for nonrespondents units for the d^{th} domain in the h^{th} Stratum

- $W_d h^*$ = The calibration weight
- W_{dh} = Stratum weight
- W_{dh1} = Response rate of the d^{th} domain in the h^{th} Stratum
- W_{dh2} = Non-response rate of the d^{th} domain in the h^{th} Stratum
- λ_1, λ_2 and λ_3 = the LaGrange multipliers
- X = Auxiliary variable
- Y = Study variable
- \bar{x}_{dh} = Sample mean for the d^{th} domain in the h^{th} Stratum of the auxiliary variable
- \bar{y}_{dh}^* = Unbiased estimator of the population mean for the d^{th} domain in the h^{th} Stratum of the study variable
- \bar{X}_{dh} = Population mean for d^{th} domain of the auxiliary variable in the h^{th} Stratum
- \bar{Y}_{dh} = Population mean for d^{th} domain of the study variable in the h^{th} Stratum
- \bar{X}_d = Population mean for d^{th} domain of the auxiliary variable
- \bar{Y}_d = Population mean for d^{th} domain of the study variable
- S_{ydh}^2 = Mean square of the d^{th} domain in the h^{th} Stratum of the study variable
- S_{xdh}^2 = Mean square of the d^{th} domain in the h^{th} Stratum of the auxiliary variable
- C_{xdh} = Coefficient of variation for the d^{th} domain in the h^{th} Stratum of the auxiliary variable
- C_{ydh} = Coefficient of variation for the d^{th} domain in the h^{th} Stratum of the study variable
- S_{ydh2}^2 = Mean square of non-response of the d^{th} domain in the h^{th} Stratum of the study variable
- k_{dh} = Inverse sampling rate
- Q_{dh} = Tuning parameter

Udofia (2004) estimator

An alternative ratio estimator for domain mean was suggested by⁵ is as follows:

$$t_{2j} = \sum_h^k W_h \frac{\bar{y}_h^*}{\bar{x}_h} \bar{X}_h \tag{1}$$

With

$$Bias(t_{2j}) = \sum_{h=1}^k W_h \frac{1-f_h}{n_h \bar{X}_h} (R_h S_{xh}^2 - S_{x_h y_h})$$

and

$$MSE(t_{2j}) = \sum_{h=1}^k W_h^2 \left[\frac{1-f_h}{n_h} \left(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{x_h y_h} + \frac{W_{2h}(k-1)}{n_h} S_{2yh}^2 \right) \right] \tag{2}$$

where

$$R_h = \frac{\bar{y}_h}{\bar{x}_h}, \quad f_h = \left(\frac{1}{n_h} - \frac{1}{N_h} \right)$$

Pal and Singh HP estimator

Pal and Singh¹⁵ proposed a class of ratio-cum-ratio-type exponential estimators for population mean with sub sampling the non-respondents. The estimator and the mean square error is given as:

$$t_{ps1} = \alpha \bar{y}^* \left(\frac{\bar{X}}{\bar{x}} \right) + (1 - \alpha) \bar{y}^* \exp \left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right)$$

And

$$MSE(t_{ps1}) = \bar{Y}^2 \left(\lambda C_y^2 (1 - \rho_{xy}^2) + \frac{W_2(Z-1)}{n} C_{y(2)}^2 \right) \tag{3}$$

Where

$$W_2 = \frac{n_2}{n}, \lambda = \frac{1-f}{n}, f = \frac{n}{N} \text{ and } \alpha \text{ is a constant}$$

Ashutosh estimator

Ashutosh¹⁰ proposed a direct ratio generalized estimator for domain mean through stratified sampling with non-response as;

$$T_{DG.st.\beta.d} = \bar{y}_{st.d} \left[\frac{\bar{X}_{st.d}}{\bar{X}_{st.d}} \right]^\beta$$

Where β is a chosen constant of d^{th} domain mean of x and the value of y respondents can be written as;

$$\bar{y}_{st.d} = \sum_{h=1}^H W_{h.d} \bar{y}_{h.d}$$

$$\bar{x}_{st.d} = \sum_{h=1}^H W_{h.d} \bar{x}_{h.d}$$

Members of the proposed estimators $T_{DG.st.\beta.d}^*$

$$T_{DG.st.\beta.d}^* = \bar{y}_{st.a}^* \text{ if } \beta = 0$$

$$T_{DG.st.-1.a}^* = \frac{\bar{y}_{st.a}^*}{\bar{x}_{st.a}^*} \bar{X}_{h.a} \text{ if } \beta = -1$$

$$T_{DG.st.1.a}^* = \bar{y}_{st.a}^* \frac{\bar{x}_{st.a}^*}{\bar{x}_{st.a}^*} \text{ if } \beta = 1$$

$$T_{DG.st.2.a}^* = \bar{y}_{st.a}^* \left[\frac{\bar{x}_{st.a}^*}{\bar{x}_{st.a}^*} \right]^2 \text{ if } \beta = 2$$

Bias and Mean Square Error of $T_{DG.st.-1.a}^*$ is given as;

$$Bias(T_{DG.st.-1.a}^*) = \sum_{h=1}^H W_{h.a} \bar{y}_{h.a} \left[\frac{N_{h.a} - n_{h.a}}{N_{h.a} n_{h.a}} C_{yha}^2 + \frac{(g_{h.a} - 1) W_{2ha}}{n_{h.a}} C_{2yha}^2 \right] - \bar{Y}_a$$

$$MSE(T_{DG.st.-1.a}^*) = \sum_{h=1}^H W_{h.a}^2 \bar{Y}_a^2 \left[\frac{N_{h.a} - n_{h.a}}{N_{h.a} n_{h.a}} (C_{yha}^2 + C_{yha}^2 - 2C_{yha}) + \frac{(g_{h.a} - 1) W_{2ha}}{n_{h.a}} (C_{2yha}^2 + C_{2yha}^2 - 2C_{2yha}) \right] \tag{4}$$

Sampling design in single phase

Let $\pi = \{U_1, U_2, \dots, U_N\}$ denote a finite population, the elements of which fall into L known strata with N_{dh} elements the h^{th} stratum, $h = 1, 2, \dots, L, \sum h N_{dh} = N_d$. It is assumed that π can also be partitioned according to the distribution of variable Z into exhaustive set of D sub-populations or domains of study that is denoted by $\{A_d^*; d = 1, 2, \dots, D\}$. Each stratum consist of a substratum of N_{1dh} respondents and a substratum of N_{2dh} non-respondents, $N_{1dh} + N_{2dh} = N_{dh}$ for all h . Let A_{dh}^* denote the part of domain $d(A_d^*)$ in stratum h and N_{dhj} the unknown number of elements in A_{dh}^* . Let y_{dhj} denote the value of characteristic Y for element i in A_{dh}^* .

Proposed estimator

Calibration has been proven to be an estimation technique to smoothen an existing estimator for a better precision and an improved efficiency. For household survey and other economic data that requires knowledge of the supplementary information, a new ratio estimator is suggested to enhance efficiency in domains of study even in the presence of non-response. Motivated by⁵ in an Alternative Ratio Estimator for domain mean, we proposed the following estimator:

$$t_{cal}^* = \sum_{h=1}^L W_{dh}^* \frac{\bar{y}_{dh}}{\bar{x}_{dh}} \bar{X}_{dh} \tag{5}$$

(5) can be written as

$$t_{cal}^* = \sum_{h=1}^L W_{dh}^* \bar{y}_{dhr} \tag{6}$$

Where

$$\bar{y}_{dhr} = r_{dh} \bar{X}_{dh}$$

and

$$r_{dh} = \frac{\bar{y}_{dh}^*}{\bar{x}_{dh}}, \bar{X}_{dh} \text{ is assume to be known and } W_{dh}^* \text{ is the calibration weight aimed at adjusting the existing weight in }^5 \text{ estimators using a chi-square distance measure.}$$

Thus the optimization problem is given by:

$$\varphi = \frac{\sum_{h=1}^L (W_{dh}^* - W_{dh})^2}{Q_{dh} W_{dh}}$$

Subject to the following constraints

$$\sum_{h=1}^L W_{dh}^* = 1$$

$$\sum_{h=1}^L W_{dh}^* \bar{x}_{dh} = \sum_{h=1}^L W_{dh} \bar{X}_{dh}$$

$$\sum_{h=1}^L W_{dh}^* S_{dh}^2 = \sum_{h=1}^L W_{dh} S_{dh}^2$$

Thus the optimization problem is given by:

$$\varphi = \frac{\sum_{h=1}^L (W_{dh}^* - W_{dh})^2}{Q_{dh} W_{dh}} - 2\lambda_1 \left(\sum_{h=1}^L W_{dh}^* - 1 \right) - 2\lambda_2 \left(\sum_{h=1}^L W_{dh}^* \bar{x}_{dh} - \sum_{h=1}^L W_{dh} \bar{X}_{dh} \right) - 2\lambda_3 \left(\sum_{h=1}^L W_{dh}^* S_{dh}^2 - \sum_{h=1}^L W_{dh} S_{dh}^2 \right)$$

where λ_1, λ_2 and λ_3 are the Lagrange multipliers such that

$$\frac{\partial \varphi}{\partial W_{dh}^*} = \frac{2(W_{dh}^* - W_{dh})}{Q_{dh} W_{dh}} - 2\lambda_1 - 2\lambda_2 \bar{x}_{dh} - 2\lambda_3 S_{dh}^2 = 0$$

$$\Rightarrow W_{dh}^* = W_{dh} + Q_{dh} W_{dh} (\lambda_1 + \lambda_2 \bar{x}_{dh} + \lambda_3 S_{dh}^2)$$

Substituting W_{dh}^* in Eq. 6 gives

$$\hat{t}_{cal} = \sum_{h=1}^L W_{dh} \bar{y}_{dhr} + \beta_{1(dh)} \left(1 - \sum_{h=1}^L W_{dh} \right) + \beta_{2(dh)} \left(\sum_{h=1}^L W_{dh} (\bar{X}_{dh} - \bar{x}_{dh}) \right) + \beta_{3(dh)} \left(\sum_{h=1}^L W_{dh} (S_{dh}^2 - s_{dh}^2) \right) \tag{7}$$

Where

$$\beta_{(dh)} = \frac{\left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^4 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right)^2 - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} s_{dh}^4 \right) + \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) + \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) \right)}{\left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \bar{y}_{dhr} \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dhr} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{x}_{dh} s_{dh}^2 \right) \right)}$$

$$\beta_{3(d_h)} = \frac{\left[\left(\sum_{h=1}^L Q_{dh} W_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^4 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^4 \right) + \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \right]}{\left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right)} - \left(\sum_{h=1}^L Q_{dh} W_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^4 \right) + \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \right]}{\left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) - \left(\sum_{h=1}^L Q_{dh} W_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} s_{dh}^2 \right) \left(\sum_{h=1}^L Q_{dh} W_{dh} \bar{y}_{dh} \right)}$$

$$E[e_0 e_1] = \frac{COV(\bar{x}_{dh}, \bar{y}_{dh}^*)}{\bar{X}_{dh} \bar{Y}_{dh}} = \frac{1}{\bar{X}_{dh} \bar{Y}_{dh}} \left[C \left(E[\bar{x}_{dh}], E[\bar{y}_{dh}^*] \right) \right]$$

$$= \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} C_{ydh} C_{xdh}$$

$$= \frac{1}{\bar{X}_{dh} \bar{Y}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh}$$

$$\therefore E[e_0] = E[e_1] = E[e_2] = 0$$

$$E[e_1 e_2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) C_{xdh} \lambda_{03} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \frac{S_{xdh}}{\bar{X}_{dh}} \lambda_{03}$$

where

$$\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2} \mu_{02}^{s/2}}$$

And

$$\mu_{rs} = \frac{1}{N_{dh} - 1} \sum_{i=1}^N (Y_{dhi} - \bar{Y}_{dh})^r (X_{dhi} - \bar{X}_{dh})^s$$

$$\mu_{20} = S_{ydh}^2$$

$$\mu_{02} = S_{xdh}^2$$

Hence

$$\lambda_{03} = \frac{\mu_{03}}{\mu_{20}^{0/2} \mu_{02}^{3/2}}$$

$$E[e_0 e_2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) C_{ydh} \lambda_{12} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \frac{S_{ydh}}{\bar{Y}_{dh}} \lambda_{12}$$

Where

$$\lambda_{12} = \frac{\mu_{12}}{\mu_{20}^{1/2} \mu_{02}} \bar{y}_{dhr} = \frac{\bar{y}_{dh}}{\bar{x}_{dh}} \bar{X}_{dh} = \bar{Y}_{dh} (1 + e_0 - e_1 + e_1^2 - e_0 e_1)$$

To obtain the bias

$$B(t_{cal}^*) = E[t_{cal}^* - \bar{Y}_d]$$

$$= E \left[\sum_{h=1}^L W_{dh} [\bar{Y}_{dh} (1 + e_0 - e_1 + e_1^2 - e_0 e_1)] - \beta_{2(dh)} \sum_{h=1}^L W_{dh} \bar{X}_{dh} e_1 - \beta_{3(dh)} \sum_{h=1}^L W_{dh} S_{xdh}^2 e_2 - \bar{Y}_d \right]$$

$$= E \left[\sum_{h=1}^L W_{dh} [\bar{Y}_{dh} (e_0 - e_1 + e_1^2 - e_0 e_1)] - \beta_{2(dh)} \sum_{h=1}^L W_{dh} \bar{X}_{dh} e_1 - \beta_{3(dh)} \sum_{h=1}^L W_{dh} S_{xdh}^2 e_2 \right]$$

$$= \sum_{h=1}^L W_{dh} \bar{Y}_{dh} \left[\left(E(e_1^2) - E(e_0 e_1) \right) \right]$$

$$B(t_{cal}^*) = \sum_{h=1}^L W_{dh} \bar{Y}_{dh} \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \frac{S_{ydh}^2}{\bar{X}_{dh}^2} \right] - \sum_{h=1}^L W_{dh} \bar{Y}_{dh} \left[\frac{1}{\bar{X}_{dh} \bar{Y}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh} \right] \quad (8)$$

$$MSE(t_{cal}^*) = E[t_{cal}^* - \bar{Y}_{dh}]^2$$

$$= E \left[\sum_{h=1}^L W_{dh} [\bar{Y}_{dh} (1 + e_0 - e_1 + e_1^2 - e_0 e_1)] - \beta_{2(dh)} \sum_{h=1}^L W_{dh} \bar{X}_{dh} e_1 - \beta_{3(dh)} \sum_{h=1}^L W_{dh} S_{xdh}^2 e_2 - \bar{Y}_d \right]^2$$

ignoring terms with power >2

$$MSE(t_{cal}^*) = \sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{ydh}^2 + \frac{(K_{dh} - 1)}{n_{dh}} W_{dh2} S_{ydh2}^2 \right] - 2 \sum_{h=1}^L W_{dh}^2 \left[\frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh} \right] - 2 \beta_{2(dh)} \sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh} \right] - 2 \beta_{3(dh)} \sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 S_{ydh} \lambda_{12} \right]$$

Bias and variance of the proposed estimator

From the proposed estimator above

$$\text{Let } \left(\frac{\bar{y}_{dh}^* - \bar{Y}_{dh}}{\bar{Y}_{dh}} \right) = e_0$$

$$\Rightarrow W_{dh}^* = W_{dh} + Q_{dh} W_{dh} (\lambda_1 + \lambda_2 \bar{x}_{dh} + \lambda_3 s_{dh}^2)$$

$$e_1 = \frac{(\bar{x}_{dh} - \bar{X}_{dh})}{\bar{X}_{dh}}$$

$$e_2 = \frac{(s_{xdh}^2 - S_{xdh}^2)}{S_{xdh}^2}$$

Where

$$\bar{y}_{dh}^* = \frac{\sum_{h=1}^L y_{dh}^*}{n_{dh}}, \quad \bar{x}_{dh} = \frac{\sum_{h=1}^L x_{dh}}{n_{dh}}, \quad \bar{X}_{dh} = \frac{\sum_{h=1}^L X_{dh}}{N_{dh}}, \quad s_{dh}^2 = \frac{\sum_{h=1}^L (\bar{x}_{dh} - \bar{X}_{dh})^2}{n_{dh} - 1}$$

$$\text{and } S_{dh}^2 = \frac{\sum_{h=1}^L (\bar{X}_{dh} - \bar{X})^2}{N_{dh} - 1}$$

Also,

$$\bar{y}_{dh}^* = \bar{Y}_{dh} (1 + e_0)$$

$$\bar{x}_{dh} = \bar{X}_{dh} (1 + e_1)$$

$$s_{xdh}^2 = S_{xdh}^2 (1 + e_2)$$

Let

$$E[e_0^2] = \frac{Var(\bar{y}_{dh}^*)}{\bar{Y}_{dh}^2} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) C_{ydh}^2 + \frac{(K_{dh} - 1)}{n_{dh2}} W_{dh2} C_{ydh2}^2$$

$$= \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \frac{S_{ydh}^2}{\bar{Y}_{dh}^2} + \frac{(K_{dh} - 1)}{n_{dh} \bar{Y}_{dh}^2} W_{dh2} S_{ydh2}^2$$

$$E[e_1^2] = \frac{Var(\bar{x}_{dh})}{\bar{X}_{dh}^2} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) C_{xdh}^2 = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \frac{S_{xdh}^2}{\bar{X}_{dh}^2}$$

$$E[e_2^2] = \frac{Var(s_{xdh}^2)}{S_{xdh}^2} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \frac{S_{xdh}^4}{S_{xdh}^2} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2$$

$$\begin{aligned}
 & + \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}^2}{\bar{X}_{dh}^2} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 + \\
 & 2\beta_{2dh} \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 + 2\beta_{3dh} \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^3 \lambda_{03} \right] \\
 & + \beta_{2dh}^2 \sum_{h=1}^L W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 + \\
 & \beta_{3dh}^2 \sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{dh}^4 (\lambda_{04} - 1) \right] \tag{9}
 \end{aligned}$$

To obtain minimum variance, we differentiate (9) partially with respect to \hat{a}_{2dh} and \hat{a}_{3dh}

Such that

$$\hat{a}_{2(dh)} = \frac{\sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh} \right] - \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2}{\sum_{h=1}^L W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2} \tag{10}$$

$$\hat{a}_{3(dh)} = \frac{\sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 S_{ydh} \lambda_{12} \right] - \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^3 \lambda_{03} \right]}{\sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{dh}^4 (\lambda_{04} - 1) \right]} \tag{11}$$

$$\min MSE(t_{cal}^*) = \sum_{h=1}^L W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{ydh}^2 + \frac{(K_{dh} - 1)}{n_{dh2}} W_{dh2}^2 S_{ydh2}^2 \right] - 2 \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh} +$$

$$\begin{aligned}
 & \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 - \frac{\left[\sum_{h=1}^L W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) \rho_{xy} S_{xdh} S_{ydh} - \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 \right]^2}{\sum_{h=1}^L W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2} \\
 & \left[\frac{\sum_{h=1}^L W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 S_{ydh} \lambda_{12} - \sum_{h=1}^L W_{dh}^2 \frac{\bar{Y}_{dh}}{\bar{X}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^3 \lambda_{03}}{\sum_{h=1}^L W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{dh}^4 (\lambda_{04} - 1)} \right]^2 \tag{12}
 \end{aligned}$$

Equation (12) is the minimum variance for the proposed estimator

Percentage relative efficiency of the estimators

The percentage relative efficiency of the proposed estimators with respect to the existing estimators is given as:

$$PRE = \frac{MSE(P)}{MSE(E)} \times 100$$

Empirical study

We take the Sweden municipalities MU284,¹⁶ (appendix B). The

population is geographically sub-divided (domain) into eight different parts 1, 2, 3, 4, 5, 6, 7 and 8 having their sizes 25, 48, 32, 38, 56, 41, 15 and 29 respectively. However, we considered only four domains 1, 3, 7 and 8 because these domains have small units compared to other domains. The proposed estimator is a calibration estimator. Variables like n_{dh1} and n_{dh} were computed based on existing information from the populations. Then each of the domains is classified into homogeneous groups according to our convenient into two strata: value of below 1500 (millions of kronor) and above 1500 (millions of kronor). We consider two cases 1 and 2 of non-response (in both Population I and Population II).

Case 1: If non-respondents are available in both strata (1 and 2) as well as in the domains (approximately 30%).

Case 2: If different non-respondents are available in both strata 1 and 2 approximately 20% and 40% respectively.

Population I

Y: Real estate values according to 1984 assessment (in millions of kronor).

X: Total number of municipal employees in 1984.

Population II

Another population is considered ([16] appendix B) which is classified in to four domains with stratum 1 and 2 according to the revenues less than 100 (in millions of kronor) and revenues above 100 (in millions of kronor).

Y: Revenues of 1985 municipal taxation assessment (in millions of kronor).

X: 1985 population (in thousands).

Discussion

This discussion is based on the empirical analysis carried out and results presented in Tables 1–7. From Table 7 (Populations I and II) with respect to single stage sampling (MSE of estimators for domain mean), it is observed that the mean square error of the proposed estimator t_{cal}^* is less than the MSE of the existing estimators in all the domains. This is seen in both cases of non-response where the non-response rate was uniform across the strata and where it was non-uniform as specified in the data. The Average Mean Squared Errors (AMSE) also confirms the behavior of the MSE in both populations and cases. From Table 8 (Populations I and II) with respect to single stage sampling, it is observed that the Percentage Relative Efficiency (PRE) for the proposed estimators kept at a benchmark of 100% had greater gains in efficiency than the existing estimators for all the domains.

Table 1 Value of parameters of the strata (1 and 2) and domains

Domain Parameter	Domain							
Domain Size	25		32		15		29	
Stratum	1	2	1	2	1	2	1	2
N_{dh}	2	23	12	20	2	13	18	11
W_{dh}	0.080	0.920	0.375	0.625	0.133	0.867	0.620	0.379
\bar{Y}_{dh}	955.50	6888	1056.9	3364	1231	4020	723.2	4799

Table 1 Continued...

Domain Parameter	Domain							
	25	32	15	29	25	32	15	29
\bar{X}_{dh}	529	4385	485.4	1816	493	1694	354.1	2205
S_{Ydh}^2	40.5	136775663	71715.5	4652460	135721	5643626	81863.7	10236271
S_{Xdh}^2	101250	81259476	29239.7	2530489	162	2354475	18128.3	2902966
S_{XYdh}	-2025	104701660	34761.1	3144594	-4689	2999017	13306	3557068
ρ_{XYdh}	-1.000	0.993	0.759	0.916	-1.000	0.823	0.345	0.653

Source: Statistical computation from original data 2023

Table 2 The parameter values of strata (1 and 2) for domains (1, 2, 3 and 4) in case 1

Domain	Strata	S_{Ydh2}^2	S_{Xdh2}^2	S_{XYdh2}	K_{dh}	n_{dh2}	w_{dh2}	n_{dh1}	n_{dh}
1	1	0	0	0	3	0	0.3	0	0
	2	223415888	132328227	171255299	2	4	0.3	10	14
2	1	120583	9862.3	28095.5	2	1	0.3	2	3
	2	3977507	2517165	2669307	2	3	0.3	8	11
3	1	0	0	0	2	0	0.3	0	0
	2	987699	1771955	1137277	3	1	0.3	3	4
4	1	79129	20311.8	8114.3	2	3	0.3	6	9
	2	141512	5618	-28196	2	1	0.3	1	2

Source: Statistical computation from original data 2023

Table 3 The parameter values of Strata (1 and 2) for domain (1,2,3 and 4) in case 2

Domain	Strata	S_{Ydh2}^2	S_{Xdh2}^2	S_{XYdh2}	k_{dh}	n_{dh2}	W_{dh2}	n_{dh1}	n_{dh}
1	1	0	0	0	2	0	0.2	0	0
	2	256872328	152987958	197463945	3	5	0.4	7	12
2	1	120583	9862.3	28095.5	2	1	0.2	2	3
	2	3977507	2517165	2669307	3	4	0.4	7	11
3	1	0	0	0	2	0	0.2	0	0
	2	987699	1771955	1137277	4	2	0.4	2	4
4	1	79129	20311.8	8114.3	2	2	0.2	7	9
	2	141512	5618	-28196	3	1	0.4	1	2

Source: Statistical computation from original data 2023

Table 4 The parameter value of the strata for the domains (1, 2, 3 and 4)

Domain Parameter	Domain							
	25	32	15	29	25	32	15	29
Stratum	1	2	1	2	1	2	1	2
N_{dh}	2	23	14	18	7	8	20	9
W_{dh}	0.08	0.92	0.438	0.563	0.467	0.533	0.69	0.31
\bar{Y}_{dh}	75.5	594	67.5	260.6	73	315	44.55	345.2
\bar{X}_{dh}	9.00	67.1	10.643	34.5	10.714	40.63	6.55	41.89
S_{Ydh}^2	840.5	1551426	275.96	41200.8	369.67	51631.7	187.21	54848.9
S_{Xdh}^2	18.00	16649.6	4.555	544.97	6.905	731.13	5.103	681.61

Table 4 Continued...

Domain Parameter	Domain							
Domain Size	25	32		15		29		
S_{XYdh}	123	160633.5	32.038	4559.147	49.5	6116.143	29.839	6076.778
ρ_{XYdh}	1.00	0.999	0.904	0.962	0.98	0.995	0.965	0.994
λ_{12}	0.003414	0.0000163	0.0004078	0.0000537	0.005061	0.0001168	0.0003394	0.0000987
λ_{03}	0.001047	0.0000066	0.000086	0.0000156	0.000813	0.0000207	0.0000813	0.0000208
λ_{04}	0.250000	0.5941080	0.407718	0.543903	0.001194	0.417192	0.221445	0.283596

Table 5 The parameter values of Strata (1 and 2) for domain (1,2,3 and 4) in case 1

Domain	Strata	S_{Ydh2}^2	S_{Xdh2}^2	S_{XYdh2}	k_{dh}	n_{dh2}	W_{dh2}	n_{dh1}	n_{dh}
1	1	0	0	0	2	0	0.3	0	0
	2	2478255	26533	256331	2	4	0.3	10	14
2	1	373.7	4.200	35.05	2	2	0.3	3	5
	2	64541.6	875.25	7146.75	2	3	0.3	6	9
3	1	0	0	0	2	0	0.3	0	0
	2	0	0	0	2	0	0.3	0	0
4	1	168.16	5.018	27.945	3	3	0.3	8	11
	2	0	0	0	2	0	0.3	0	0

Source: Statistical computation from original data 2023

Table 6 Parameter values of strata (1 and 2) for each domain in the case 2

Domain	Strata	S_{Ydh2}^2	S_{Xdh2}^2	S_{XYdh2}	k_{dh}	n_{dh2}	W_{dh2}	n_{dh1}	n_{dh}
1	1	0	0	0	2	0	0.2	0	0
	2	2654913	28395.1	274463.7	3	5	0.4	8	13
2	1	176.25	1.333	9.667	2	1	0.2	3	4
	2	64184.8	874.3	7069.214	2	3	0.4	5	8
3	1	0	0	0	2	0	0.2	0	0
	2	0	0	0	3	0	0.4	0	0
4	1	169.778	5.511	30	2	2	0.2	8	10
	2	0	0	0	3	0	0.4	0	0

Table 7 MSE of Estimators for domain mean in both cases 1 and 2(Population 1&2)

Estimator	1	2	3	4	AMSE
Case 1 (Population 1)					
t_{2j}	1950061	74298	817651	367172	802295.5
$T_{DG.st.-1.d}^*$	969486	1524915	8225052	7635549	4588751
t_{exp1}	3388887	2662270	2108968	1580771	2435224
t_{cal}^*	389212	68914.02	613475	41078.01	278169.8
Case2 (Population 1)					
t_{2j}	1724942	69614.41	771605.6	36752.6	650728.7
$T_{(DG.st.-1.d)}^*$	531387	334565	501732.17	1024511	598048.8
t_{exp1}	1190074	28456.45	270455	927116.3	604025.4
t_{cal}^*	31623	2178.416	35028.05	32543.11	25343.14

Table 7 Continued...

Estimator	1	2	3	4	AMSE
Case 1 (Population 2)					
t_{2j}	716146	115492	-	721	208089.8
$T_{DG.st.-1.d}^*$	279503	68413	-	409	87081.25
t_{exp1}	30790869	615419.7	-	246	7851634
t_{cal}^*	222071	21236	-	169	60869
Case 2 (Population 2)					
t_{2j}	11048.4	246.8	-	9	2826.05
$T_{DG.st.-1.d}^*$	11989.8	639.6	-	313	3235.6
t_{exp1}	97184	942	-	206	24583
t_{cal}^*	10431	127.3	-	4.8	2640.775

Note: AMSE, average mean square error

Source: Statistical computation from original data 2023

Table 8 PRE of the estimators for domain mean in both cases 1 and 2(Population 1 and 2)

	D_1	D_2	D_3	D_4
Case 1 (Population 1)				
Estimator				
t_{2j}	19.95897	92.75353	75.02895	11.18767
$T_{DG.st.-1.d}^*$	40.14622	4.519204	7.458615	0.537984
t_{exp1}	11.48495	2.588544	29.08887	2.598606
t_{cal}^*	100	100	100	100
Case 2(Population 1)				
t_{2j}	1.833279	3.12926	4.539631	88.54642
$T_{DG.st.-1.d}^*$	5.95103	0.651119	6.981424	3.176453
t_{exp1}	2.65723	7.655263	12.95153	3.510143
t_{cal}^*	100	100	100	100
Case 1 (Population 2)				
t_{2j}	31.00918	18.38742	0	23.43967
$T_{DG.st.-1.d}^*$	79.4521	31.04088	0	41.32029
t_{exp1}	0.721224	3.450653	0	68.69919
t_{cal}^*	100	100	0	100
Case 2(Population 2)				
t_{2j}	94.41186	51.58023	0	53.33333
$T_{DG.st.-1.d}^*$	86.99895	19.90306	0	1.533546
t_{exp1}	10.73325	13.5138	0	2.330097
t_{cal}^*	100	100	0	100

Conclusion

This study develops the concept of calibration estimator for ratio estimation and proposes calibration ratio estimators of population mean in single stage sampling. The study contributes to the theory of domain estimation in stratified random sampling of the population mean of the study variable with sub-sampling the non-respondents when there is non-response in the study variable and auxiliary variable is free from non-response.

The proposed class of estimators provide opportunity for different known values of the domain population parameters of the auxiliary variable to be incorporated in constructing estimators in the presence of non-response using the concept of calibration. The study revealed that the first constraint is just the sum of the calibration weight equals to one and the third constraint which has to do with the stratum variance also contributes immensely to the efficiency of the proposed estimator. Furthermore, with the adoption of the procedure of sub-sampling the non-respondents even with ratio estimator, the study has reveal that subjecting an estimator to conditions where the study variable is affected by non-response while the auxiliary variable is free of non-response has no effect in the mean estimate.

From the efficiency comparison and empirical work, it becomes pertinent that the use of calibration technique has really paid off in providing estimates of the population mean with sub-sampling the non-respondents that provides greater gains in efficiency better than the existing estimators. This will proffer useful results to users of statistics and researchers when working on economic data that requires the use of auxiliary data either from the records or from previous survey.

However, it could be seen clearly from Table 7 that it was impossible to compute estimates for domain 3 in both cases of population II and hence, the mean square error was not computed. As a result, the PRE was accorded zero value. This is as a result of no sample size for both the respondents and the non-respondents as indicated in Table 6. Future research is encouraged in the light of this through the use of synthetic estimation technique.

Acknowledgments

None.

Conflicts of interest

The authors declare there is not any conflict of interest.

Funding

None.

References

1. Deville JC, Särndal C. E. Calibration estimators in survey sampling. *JASA*. 1992;87:376–382.
2. Koyuncu N, Kadilar, C. (2013). Calibration estimators using different measures in stratified random sampling. *International Journal of Modern Engineering Research*. 2013;3(1):415–419.
3. Clement EP, Udofia GA, Enang EI. Sample design for domain calibration estimators. *International Journal of Probability and Statistics*. 2014;3(1):8–14.
4. Clement EP, Enang E I. Calibration approach alternative ratio estimator for population mean in stratified sampling. *International Journal of Statistics and Economics*. 2015;16(1):83–93.
5. Godwin A Udofia. Ratio estimation for small domains with subsampling the non-respondents:an application of Rao strategy. *Statistics in Transition*. 2004;6(5):713–724.
6. Rao Poduri SRS. Ratio estimation with sub-sampling the non-respondents. *Survey Methodology*. 1986;12:217–230.
7. Iseh, MJ, Bassey MO. Calibration estimators for population mean with subsampling the nonrespondents under stratified sampling. *Science Journal of Applied Mathematics and Statistics*. 2022;10(4):45–56.
8. Iseh Matthew, Bassey Mbuotidem. Smoothing of estimators of population mean using calibration technique with sample errors. *Journal of Modern Applied Statistical Methods*. 2024; 23(1):.
9. Cochran WG. *Sampling Techniques*, 3rd edition, New York: Wiley. 1977.
10. Ashutosh (2021) Estimator of domain mean using stratified sampling in the presence on non-response.,*Sri Lankan Journal of Applied Statistics*. 2021;22(1):13–29.
11. Clement EP, Inyang EJ. Improving the efficiency of ratio estimators by calibration weightings. *International Journal of Statistics and Mathematics*. 2021;8(1):164–172.
12. Iseh, M J, Bassey KJ. A New calibration estimator of population mean for small area with nonresponse. *Asian Journal of Probability and Statistics*. 2021(a);12(2):14–51.
13. Iseh, M. J, Bassey, KJ. Calibration estimator for population mean in small sample size with non-response. *European Journal of Statistics and Probability*. 2021(b);9(1):32–42.
14. Iseh, M.J, Enang EI. A calibration synthetic estimator of population mean in small area under stratified sampling design. *Transition in Statistics new series*. 2021;22(3):15–30.
15. Pal SK, Singh HP. A class of ratio-cum-ratio-type exponential estimators for population mean with subsampling the non-respondents. *Jordan Journal of Mathematics and Statistics*. 2017;10(1):73–94.
16. Sarndal CE, Swensson B, Wretman J. *Model-assisted surveys*. New York: Springer-Verlag. 1992.