Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 4(4) July 2026
DOI: https://dx.doi.org/10.4314/jobasr.v4i4.8

Calibration Technique for Compensation of Non-Response in Domains of Study with Probability Proportional to Size

Iseh, Matthew Joshua[1*], Ikot, Ekemini E[2] & Bassey, Kufre John[3]
[1]Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria
[2]Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria
[3]Statistics Department, Central Bank of Nigeria
[*]Corresponding Author Email: matthewiseh@aksu.edu.ng

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 using probability proportional to size. 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. The bias and mean square error (MSE) for the proposed estimators are derived. The use an auxiliary variable to estimate the population mean assuming that the non-response is observed only on the study variable is adopted. The proposed estimator and the existing estimators were compared empirically in the domains with small sampling units and two populations were considered in terms of the MSE and Percentage Relative Efficiency (PRE). Also, two cases were considered, where non-responses are uniform in the two strata at approximately (30%) and a case where the nonresponse rates are different with 20% and 40% in strata 1 and 2 respectively. The proposed estimator gained more efficiency than the existing estimators. However, the proposed estimator performed better in case II than in case I in some domains but the aggregate result suggests otherwise in both populations.

Keywords:
Auxiliary variable,
Calibration,
Sub-sampling,
Population Mean,
Domain,
Probability
Proportional to size

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. In recent times, 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. 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). 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.

Calibration estimation in sample surveys has since its introduction by Deville and Sarndal (1992) developed into 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 (Iseh & Enang 2021, and Iseh 2023). The efficiencies of some proposed estimators to existing unbiased estimators, which do not utilize the auxiliary information have been overwhelming with the use of calibration estimation technique as shown under simple random sampling, stratified sampling and double sampling estimation by several authors. This technique has been profitable with higher precision and bias reduction through weights adjustments which maybe as a result of nonresponse and small sample size in domains of interest

How to cite this article: Iseh, M. J., Ikot, E. E. & Bassey, K. J. (2026). Calibration Technique for Compensation of Non-Response in Domains of Study with Probability Proportional to Size. *Journal of Basics and Applied Sciences Research*, 4(4), 71-81. https://dx.doi.org/10.4314/jobasr.v4i4.8

The challenge of the presence of possible non-response in the estimation of population mean was first considered by Hansen and Hurwitz (1946) and several works have been done in this area including Singh and Kumar (2008), Iseh and Bassey (2021), and Iseh and Bassey (2022). Since Hansen and Hurwitz first introduced the use of probability proportional to size (PPS) sampling, a number of procedures for selecting samples without replacement have been developed. Many of them are reviewed and compared in Brewer and Hanif (1983). Survey statisticians have found probability proportional to size sampling scheme useful for selecting units from the population as well as estimating parameters of interest especially when it is clear that the survey is large in size and involves multiple characteristics. Studies on inferences in finite population sampling including the works of Godambe (1955), Agrawal and Mrida (2007), and Chaudhuri (2010) have postulated the non-existence of an unbiased estimator of population characteristics with the uniformly least value of its variance. With this development, lots of alternative estimators have been suggested in PPS sampling scheme following the pioneering work by Hansen and Hurwitz (1943).
Selection with PPS means that, on the average, a primary sampling unit (psu), which is, for example, five times as large as another will be in the sample five times as frequently as the other psu. It might appear, at first, that this would introduce a bias in the sample result with some psu's overrepresented and others under-represented - and it would, in fact, if no special attention were given to the varying probabilities in the estimating or subsampling procedures (Hansen, Hurwitz, and Madow, 1953a). One of the advantages gained by the use of PPS sampling is that: for many common populations, and with a fixed number of primary units in the sample, a smaller variance will be obtained by sampling with probability proportionate to size than by sampling with equal probability, (Hansen, et al. l953b) .Rao (1966) suggested an alternative estimator in PPS sampling by assuming that the correlation between study variables and measure of size variable is zero. Pathak (1966) proved this theory correct while Bansal and Singh (1985) argued that population correlation can never be zero and provided a non-linear transformation in the selection probabilities. Amahia et al. (1989) developed a transformation that is linear in pi and possesses the properties of arithmetic mean. This development brought about numerous contributions including the works of Grewal et al. (1997) which suggested that a subsample should be drawn from the non-responding units selected in the sample and an extra effort should be made for the collection of information from the units of the subsample.

Enang and Amahia (2012) worked on an alternative class of estimators in probability proportional to size with replacement sampling scheme for multi- character surveys in which the study variables are poorly correlated with selection probabilities is developed. This is achieved by redefining the selection probabilities. Some existing estimators have been shown to be special cases of the proposed class. Recently, Ikughur and Amahia (2011) developed a generalized transformation for a class of alternative linear estimators in PPS sampling scheme within which optimum estimator for any target population is located. An interesting feature of these estimators is that it is defined by the kth moment in correlation coefficient as are related with the statistical properties of target population namely, coefficients of variation, determination, skewness and kurtosis (Michael et al 2019, 2021, 2025 & Okereke et. al.2026). Tillé (1996) presents a new PPS, without replacement sampling procedure. The general idea of this procedure is to begin with the universe and eliminate one unit at a time until the number of remaining units is equal to the desired sample size. The set of remaining units then becomes the original sample. Although Tillé does not discuss applying his procedure to the sample expansion problem, the method of doing so is essentially immediate provided a record had been kept for the order of elimination of the units not in the original sample. Proposed class are more efficient than existing estimators under a super-population model.

Considering the area of interest, Udofia (2002a & 2002b) expanded Yates (1953) results to double sampling for probability proportional to size in estimation for domains with sub sampling the non-respondent when information on the size X of each sampling unit is known. However, it was observed that the estimators could not adjust for weights to help counter the gaps from missing survey responses, and as a result performs poorly in terms of consistency, bias and gains in efficiency in some domains of interest with higher rate of nonresponse or small sample units. Consequently, this paper seeks to extend Udofia (2002a) and (2002b) on double sampling with subsampling the nonrespondents for probability proportional to size, to a new improved estimator for population mean in stratified random sampling through the use of the theory of calibration to compensate for nonresponse and small sample size in domains of interest.

MATERIALS AND METHODS

Sample Design and Theoretical Underpinnings

Let the population, $\pi = \{U_1, U_2, \dots, U_N\}$, consist of H strata with Nh elements in the hth stratum, h = 1, 2, …, H. Let Ahd denote the part of the domain d (d= 1, 2, …, M), in stratum h. The number of elements, Nhd in domain d is not known. We assume that $0 < N_{hd} < N_h$, and information on X, the size of each element in the population is known before the start of the survey. An initial sample, $S(n_h')$, of size n_h' is drawn from stratum h, with probabilities $p'_h = \frac{x'_{hi}}{\bar{x}_h}$, $\bar{X}_h = \frac{1}{n'_h}\sum_{i=1}^{n'_h} x'_{hi}$, proportional to X and with replacement. We denote by n'_{hd} the number of units in $S(n'_h)$ that fall in A_{hd}. For a fixed n'_{hd}, the n'_{hd}units constitute an initial random sample from A_{hd}, and we assume that $0 < n'_{hd} < n'_h$. A second sample, $S(n_h)$, of $n_h$units, $n_h < n'_h$, is drawn from $S(n'_h)$ by simple random sampling without replacement (SRSWOR) and the study variable, Y, is measured on it. Let n_{hd} denote the number of units in $n_h$that belong to A_{hd}. These n_{hd} units constitute a second random sample from A_{hd} and we assume that $0 < n_{hd} < n_h$. Again, considering the theory of subsampling the nonrespondents by Hansen and Hurwitcz (1946), let n_{hd} denote the number of units in s_{n_h} that fall in domain A_{hd}, n_{1h} the number of respondents in s_{n_h} out of which n_{1hd} belong to A_{dh}, and $n_{2h} = n_h - n_{1h}$ the number of non-respondents in s_{n_h} out of which n_{2hd} belong to A_{hd}. A sub-sample of $m_h = n_{2h}/k_h$, $k_h \geq 1$, is drawn from the n_{2h} non-respondents. Let m_{hd} of the m_h non-respondents belong to A_{hd}. Since it is a well-known fact that some elements may fall outside the domains of study, let $x'_{hi} = x_{hdi}$ for all $i \in A_{hd}$and zero otherwise and $y_{hi} = y_{hdi}$ if $i \in A_{hd}$ and zero otherwise.

Some existing estimators for domains in Double Sampling with Probabilities Proportional to Known size

Udofia (2002a) Estimator:

Udofia (2002a) proposed an estimator of jth domain total for a small domain of study in double sampling with probabilities proportional to size as:

$$\hat{Y}_j = \sum_{h=1}^{H} \frac{N_h}{n'_h}\frac{1}{n_h}\sum_{i=1}^{n_h}\frac{y_{hij}}{P_{hij}}$$

with variance

$$V(\hat{Y}_j) = \sum_{h=1}^{H} \frac{N_h^2}{n_h'^2}\frac{X_{1h}^2}{n_h(n_h-1)}\sum_{i=1}^{n_{hj}}\left(\frac{y_{hij}}{x_{hij}} - \frac{1}{n_{hj}}\sum_{i=1}^{n_{hj}}\frac{y_{hij}}{x_{hij}}\right)^2 + \sum_{h=1}^{H}\frac{N_h(N_h-n'_h)}{(N_h-1)n_h n'_h}\left\{\left[X_{1hj}\sum_{i=1}^{n_{hj}}\frac{y_{hij}}{x_{hij}} - \frac{X_{1hj}^2}{n'_h(n_h-1)}\left(\left(\sum_{i=1}^{n_{hj}}\frac{y_{hij}}{x_{hij}}\right)^2 - \sum_{i=1}^{n_{hj}}\frac{y_{hij}^2}{x_{hij}^2}\right)\right] + \frac{n'_{hj}}{n'_h}N_h\left(\mathbf{1} - \frac{n'_{hj}}{n'_h}\right)X_{1hj}^2\left(\frac{\mathbf{1}}{\boldsymbol{n_{hj}}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n_{hj}}}\frac{\boldsymbol{y_{hij}}}{\boldsymbol{x_{hij}}} - \frac{\mathbf{1}}{n_h}\sum_{\boldsymbol{h=1}}^{\boldsymbol{H}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n_{hj}}}\frac{\boldsymbol{y_{hij}}}{\boldsymbol{x_{hij}}}\right)^{\mathbf{2}}\right\} \quad (1)$$

Udofia (2002b) Estimator:

Udofia (2002b) proposed an estimator of the domain total for a small domain of study in double sampling with probabilities proportional to Size as:

$$\hat{Y}_j = \sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{n_h}\frac{y_{hij}}{P_{hij}}$$

with variance estimator

$$\hat{V}(\hat{Y}_j) = \sum_{h=1}^{H}\frac{1}{n_h(n_{hj}-1)}\sum_{i=1}^{n_{hj}}\left(\frac{y_{hij}}{P_{hij}} - \frac{1}{n_{hj}}\sum_{i=1}^{n_{hj}}\frac{Y_{hij}}{P_{hij}}\right)^2 \quad (2)$$

Des Raj (1964) Estimator:

Des Raj (1964) proposed an unbiased estimator for the hth domain of the population total:

$$\hat{Y}_h = (N_h/n'_h)(x'_h/n_h)\sum_{i=1}^{n_h}(y_{hi}/x_{hi})$$

with variance

$$V(\hat{Y}_h) = \frac{{N_h}^2}{n_h(n_h-1)n'_h}\left[\sum_{i=1}^{n_h}\frac{y_i^2}{x_i^2} - \frac{1}{n_h}\left(\sum_{i=1}^{n}\frac{y_{hi}}{x_{hi}}\right)^2\right]\left[\frac{1}{n'_h}x'^2 - \left(1-\frac{n'_h}{N_h}\right)S_x^2 + \frac{N_h(N_h-n'_h)}{n'_h}\left(\frac{1}{n_h}\sum_{i=1}^{n}\frac{y_{hi}}{x_{hi}}\right)^2 S_x^2\right] \quad (3)$$

The proposed Calibration estimator

Following Udofia (2002), and by adopting the theory of subsampling the nonrespondents in section 2.1, the population mean when there is nonresponse in the domains of study is obtained as

Let

$y'_{hi} \quad = y_{hdi}$ if the ith element is in Ahdi

$= 0$ if the ith element is not in Ahdi

$$\hat{\hat{Y}}_{dPPS} = \sum_{h=1}^{H}\frac{1}{n'_h}\frac{1}{P'_h}\frac{1}{n_h}\sum_{i=1}^{n_h} y'_{hi}$$

$$= \sum_{h=1}^{H}\alpha'_h\bar{y}_h^{*\prime} \quad (4)$$

where $\bar{y}_h^{*\prime} = w_{1h}\bar{y}_{n_{1dh}} + w_{2h}\bar{y}_{n_{2dh}}$ and $\alpha'_h = \frac{1}{n'_h P'_h}$

$w_{1dh} = \frac{n_{1dh}}{n_{dh}}$ response rate of h^{th} stratum in the d^{th} domain

$w_{2dh} = \frac{n_{2dh}}{n_{dh}}$ non-response rate of h^{th} stratum in the d^{th} domain

$\bar{y}_{n_{1dh}} = \frac{\sum_{i=1}^{n_{1dh}} y_{dhi}}{n_{1dh}}$ is the mean of y_{dhi} for the n_{1dh} respondents and

$\bar{y}_{n_{2dh}} = \frac{\sum_{i=1}^{n_{2dh}} y_{dhi}}{n_{2dh}}$ is the mean of y_{dhi} for the n_{2dh} non-respondents.

Then the proposed calibration estimator is given as

$$\hat{\hat{Y}}_{dpps} = \sum_{h=1}^{H} W_h\bar{y}_h^{*\prime} \quad (5)$$

where W_h is the calibration weights aimed at adjusting the existing weight in the estimator of the population mean version of Udofia (2002) through a chi-square distance measure

$$\varphi = \frac{\sum_{h=1}^{H}(W_h-\alpha'_h)^2}{Q_h\alpha'_h}$$

subject to the following constraints

$$\sum_{h=1}^{H} W_h\bar{x}_h = \bar{X}_h \quad (6)$$

$$\sum_{h=1}^{H} W_h s_h^2 = S_h^2 \quad (7)$$

which is aimed at minimizing the distance measure in the weights as well as obtaining an optimal gain in efficiency through the optimization equation

$$\varphi(\alpha_h', W_h) = \frac{\sum_{h=1}^{H}(W_h - \alpha_h')^2}{Q_h\alpha_h'} - 2\lambda_1\left(\sum_{h=1}^{H} W_h\bar{x}_h - \bar{X}_h\right) - 2\lambda_2\left(\sum_{h=1}^{H} W_h s_h^2 - S_h^2\right)$$

where λ_1 *and* λ_2 are the Lagrange multipliers

$$\frac{\partial\varphi(\alpha_h', W_h)}{\partial W_h} = \frac{2(W_h - \alpha_h')}{Q_h\alpha_h'} - 2\lambda_1\bar{x}_h - 2\lambda_2 s_h^2 = 0$$

$$\Rightarrow W_h = \alpha_h' + Q_h\alpha_h'(\lambda_1\bar{x}_h + \lambda_2 s_h^2) \quad (8)$$

Substituting (8) into (6) and (7) give:

$$\lambda_1\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2 + \lambda_2\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2 = \bar{X}_h - \sum_{h=1}^{H}\alpha_h'\bar{x}_h \quad (9)$$

$$\lambda_1\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2 + \lambda_2\sum_{h=1}^{H} Q_h\alpha_h' s_h^4 = S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2 \quad (10)$$

From (9) and (10) we have:

$$\begin{pmatrix} \sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2 & \sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2 \\ \sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2 & \sum_{h=1}^{H} Q_h\alpha_h' s_h^4 \end{pmatrix}\begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} = \begin{pmatrix} \bar{X}_h - \sum_{h=1}^{H}\alpha_h'\bar{x}_h \\ S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2 \end{pmatrix}$$

$$\lambda_1 = \frac{(\bar{X} - \sum_{h=1}^{H}\alpha_h'\bar{x}_h)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (S_{hi}^2 - \sum_{h=1}^{H}\alpha_h' s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}$$

$$\lambda_2 = \frac{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_{hi}^2)(S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_{hi} s_h^2)(\bar{X} - \sum_{h=1}^{H}\alpha_h'\bar{x}_h)}{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}$$

Thus substituting λ_1 *and* λ_2 in (8) gives:

$$W_h = \alpha_h' + Q_{hi}\alpha_h' x_h\left(\frac{(\bar{X} - \sum_{h=1}^{H}\alpha_h'\bar{x}_h)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}\right) + Q_{hi}\alpha_h' s_h^2\left(\frac{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\bar{X} - \sum_{h=1}^{H}\alpha_h'\bar{x}_h)}{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}\right)$$

Therefore, the proposed estimator in (5) is written as:

$$\hat{\hat{\bar{Y}}}_{dpps} = \sum_{h=1}^{H}\alpha_h'\bar{y}_h^{*\prime} + \beta_{1(h)}(\bar{X}_h - \sum_{h=1}^{H}\alpha_h'\bar{x}_h) + \beta_{2(h)}(S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2) \quad (11)$$

$$\beta_{1(h)} = \frac{(\sum_{h=1}^{H}\alpha_h'\bar{x}_h\bar{y}_h^{*\prime})(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H}\alpha_h'\bar{y}_h^{*\prime} s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}$$

$$\beta_{2(h)} = \frac{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H}\alpha_h'\bar{y}_h^{*\prime} s_h^2) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H}\alpha_h'\bar{x}_h\bar{y}_h^{*\prime})}{(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h^2)(\sum_{h=1}^{H} Q_h\alpha_h' s_h^4) - (\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)(\sum_{h=1}^{H} Q_h\alpha_h'\bar{x}_h s_h^2)}$$

Bias and Mean square error of the Proposed Estimator

Recall that Udofia (2002a) estimator was shown to be unbiased with variance derived under the conditional variance technique. It should be noted that the intractability and complexity of the conditional approach, though technically exact, it is difficult to use in the case of complex survey designs. Again, the distributional dependency of the conditional approach which requires knowing the precise conditional distribution, which may be unknown, poses challenges whereas the Taylor series (Linearization) method which often only requires moments (mean/variance) and more robust to distributional assumptions. Consequently, in the proposed estimator, the Taylors series linearization method is adopted which is assumed to favor complex survey statistics, or when only local behavior around the mean is required

Recall the large number approximation for the sample means and variance

$\bar{y}_h^{*\prime} = \bar{Y}_h(1+e_0)$, $\bar{x}_h = \bar{X}_h(1+e_1)$, $\bar{x}_h' = \bar{X}_h(1+e_1')$, $s_h^2 = S_h^2(1+e_2)$

$E(e_0^2) = \frac{1}{\bar{Y}_h^2}\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_{\bar{y}h}^2 + \frac{(k_h-1)}{n_h\bar{Y}_h^2}W_{h2}S_{\bar{y}h2}^2$,

$E(e_1^2) = \frac{1}{\bar{X}^2}\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_{\bar{x}h}^2$ $E(e_2^2) = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)(\lambda_{04} - 1)$, $E(e_1'^2) = \frac{1}{\bar{X}_h^2}\left(\frac{1}{n_h'} - \frac{1}{N_h}\right)S_{\bar{x}h}^2$

$E(e_1e_0) = \frac{1}{\bar{X}_h\bar{Y}_h}\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}$,

$E(e_1'e_0) = \frac{1}{\bar{X}_h\bar{Y}_h}\left(\frac{1}{n_h'} - \frac{1}{N_h}\right)\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}$

$E[e_0e_2] = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\frac{S_{yh}}{\bar{Y}_h}\lambda_{12}$,

$E[e_1e_2] = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\frac{S_{xh}}{\bar{X}_h}\lambda_{03}$

$E[e_1'e_2] = \left(\frac{1}{n_h'} - \frac{1}{N_h}\right)\frac{S_{xh}}{\bar{X}_h}\lambda_{03}$, $E[e_0] = E[e_1] = E[e_2] = 0$ and

$S_{yh}^2 = \sum_{i=1}^{N} P_{hi}\left(\frac{y_{hi}}{P_{hi}} - \bar{Y}_h\right)^2$, $S_{x_h}^2 = \sum_{i=1}^{N} P_{hi}\left(\frac{x_{hi}}{P_{hi}} - \bar{X}_h\right)^2$,

$S_{xhi}S_{yhi} = \sum_{i=1}^{N} P_{hi}\left(\frac{y_{hi}}{P_{hi}} - \bar{Y}_h\right)\left(\frac{x_{hi}}{P_{hi}} - \bar{X}\right)$, where

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

and $\mu_{rs} = \frac{1}{N_h - 1}\sum_{i=1}^{N}(Y_{hi} - \bar{Y}_h)^r (X_{hi} - \bar{X}_h)^s$, hence

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

and $\lambda_{12} = \frac{\mu_{12}}{\mu_{20}^{1/2}\mu_{02}}$

Bias of the Proposed Estimator

From (11)

$$E\left[\hat{\bar{Y}}_{dpps}\right] = E\left[\sum_{h=1}^{H}\alpha_h'\bar{y}_h^{*\prime} + \beta_{1(h)}\left(\bar{X} - \sum_{h=1}^{H}\alpha_h'\bar{x}_h\right) + \beta_{2(hi)}\left(S_h^2 - \sum_{h=1}^{H}\alpha_h' s_h^2\right)\right]$$

$$= E[\textstyle\sum_{h=1}^{H}\alpha_h'\bar{y}_h^{*\prime}] + \beta_{1(h)}\big(\bar{X} - E(\sum_{h=1}^{H}\alpha_h'\bar{x}_h)\big) + \beta_{2(hi)}\big(S_h^2 - E(\sum_{h=1}^{H}\alpha_h' s_h^2)\big)$$

$$= \bar{Y} + \beta_{1(h)}(\bar{X} - \bar{X}) + \beta_{2(h)}(S_h^2 - S_h^2)$$

$$= \bar{Y}$$

This shows that the proposed estimator is asymptotically unbiased.

Mean Square Error of the Proposed Estimator

$$MSE\left(\hat{\bar{Y}}_{dpps}\right) = V\left(\hat{\bar{Y}}_{dpps}\right) = E\left[\hat{\bar{Y}}_{dpps} - \bar{Y}\right]^2$$

$$= E\left[\textstyle\sum_{h=1}^{H}\alpha_h'\bar{Y}_h(1+e_0) + \beta_{1(h)}(\bar{X}_h(1+e_1') - \sum_{h=1}^{H}\alpha_h'\bar{X}_h(1+e_1)) + \beta_{2(h)}(S_h^2 - \sum_{h=1}^{H}\alpha_h' S_h^2(1+e_2)) - \bar{Y}\right]^2$$

$$= E\left[\textstyle\sum_{h=1}^{H}\alpha_h'\bar{Y}_h + \sum_{h=1}^{H}\alpha_h'\bar{Y}_h e_0 + \beta_{1(h)}(\bar{X}_h + \bar{X}_h e_1' - \sum_{h=1}^{H}\alpha_h'\bar{X}_h - \sum_{h=1}^{H}\alpha_h'\bar{X}_h e_1) + \beta_{2(h)}(S_h^2 - \sum_{h=1}^{H}\alpha_h' S_h^2 - \sum_{h=1}^{H}\alpha_h' S_h^2 e_2) - \bar{Y}\right]^2$$

Since $\sum_{h=1}^{H}\alpha_h'\bar{Y}_h = \bar{Y}$, $\sum_{h=1}^{H}\alpha_h'\bar{X}_h = \bar{X}$ and $\sum_{h=1}^{H}\alpha_h' S_h^2 = S^2$

$$\Rightarrow E\left[\bar{Y} + \textstyle\sum_{h=1}^{H}\alpha_h'\bar{Y}e_0 + \beta_{1(h)}(\bar{X} - \bar{X} + \sum_{h=1}^{H}\alpha_h'\bar{X}e_1' - \sum_{h=1}^{H}\alpha_h'\bar{X}e_1) + \beta_{2(h)}(S_h^2 - S_h^2 - \sum_{h=1}^{H}\alpha_h' S_h^2 e_2) - \bar{Y}\right]^2$$

$$= E\left[\sum_{h=1}^{H}\alpha_h'\bar{Y}e_0 + \beta_{1(h)}\sum_{h=1}^{H}\alpha_h'\bar{X}e_1' - \beta_{1(h)}\sum_{h=1}^{H}\alpha_h'\bar{X}e_1 - \beta_{2(h)}\sum_{h=1}^{H}\alpha_h' S_h^2 e_2\right]^2$$

By expansion;

$$= \textstyle\sum_{h=1}^{H}\alpha_h'^2\bar{Y}^2E[e_0^2] + 2\beta_{1(h)}(\sum_{h=1}^{H}\alpha_h'^2\bar{Y}\bar{X}E[e_0e_1']) - 2\beta_{1(h)}(\sum_{h=1}^{H}\alpha_h^2\bar{Y}\bar{X}E[e_0e_1]) - 2\beta_{2(h)}(\sum_{h=1}^{H}\alpha_{hi}'^2\bar{Y}S_h^2E[e_0e_2]) + \beta_{1(h)}^2(\sum_{h=1}^{H}\alpha_h'^2\bar{X}^2E[e_1'^2]) - 2\beta_{1(h)}^2(\sum_{h=1}^{H}\alpha_{hi}'^2\bar{X}\bar{X}E[e_1'e_1]) - 2\beta_{1(h)}\beta_{2(h)}(\sum_{h=1}^{H}\alpha_h'^2\bar{X}S_h^2E[e_1'e_2]) + \beta_{1(h)}^2(\sum_{h=1}^{H}\alpha_h'^2\bar{X}^2E[e_1^2]) + 2\beta_{1(h)}\beta_{2(h)}(\sum_{h=1}^{H}\alpha_h'^2\bar{X}S_h^2E[e_1e_2]) + \beta_{2(h)}^2(\sum_{h=1}^{H}\alpha_h'^2S_h^4E[e_2^2])$$

$$= \textstyle\sum_{h=1}^{H}\alpha_h'^2\left[\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_{\bar{y}h}^2 + \frac{(k_h-1)}{n_h}W_{h2}S_{\bar{y}h2}^2\right] + 2\beta_{1(h)}\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_h'} - \frac{1}{N_h}\right)\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}\right) - 2\beta_{1(h)}\left(\sum_{h=1}^{H}\alpha_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}\right) - 2\beta_{2(h)}\left(\sum_{h=1}^{H}\alpha_h'^2S_h^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{ydh}\lambda_{12}\right) + \beta_{1(h)}^2\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_h'} - \frac{1}{N_h}\right)S_{\bar{x}h}^2\right) - 2\beta_{1(h)}^2\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_{\bar{x}h}^2\right) - 2\beta_{1(h)}\beta_{2(h)}\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_{dh}'} - \frac{1}{N_{dh}}\right)S_{\bar{x}h}^3\lambda_{03}\right) + \beta_{1(h)}^2\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_{\bar{x}h}^2\right) + 2\beta_{1(h)}\beta_{2(h)}\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{\bar{x}h}^3\lambda_{03}\right) + \beta_{2(h)}^2\left(\sum_{h=1}^{H}\alpha_h'^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{\bar{x}h}^4(\lambda_{04} - 1)\right) \quad (12)$$

To obtain minimum variance, (12) is differentiated partially with respect to $\boldsymbol{\beta_{1(h)}}$ and $\boldsymbol{\beta_{2(h)}}$

In (12) let $\left(\frac{1}{n_h} - \frac{1}{N_h}\right) = f_1$, $\left(\frac{1}{n_h'} - \frac{1}{N_h}\right) = f_2$, and

$f_3 = f_1 - f_2 = \left[\left(\frac{1}{n_h} - \frac{1}{N_h}\right) - \left(\frac{1}{n_h'} - \frac{1}{N_h}\right) = \left(\frac{1}{n_h} - \frac{1}{n_h'}\right)\right]$;

$f_4 = \frac{(K_h - 1)}{n_h}$

$$\Rightarrow \textstyle\sum_{h=1}^{H}\alpha_h'^2[f_1S_{\bar{y}h}^2 + f_4W_{h2}S_{\bar{y}h2}^2] + 2\beta_{1(h)}(\sum_{h=1}^{H}\alpha_h'^2f_2\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}) - 2\beta_{1(h)}(\sum_{h=1}^{H}\alpha_h^2f_1\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}) - 2\beta_{2(h)}(\sum_{h=1}^{H}\alpha_h'^2S_h^2f_1S_{ydh}\lambda_{12}) + \beta_{1(h)}^2(\sum_{h=1}^{H}\alpha_h'^2f_2S_{\bar{x}h}^2) - 2\beta_{1(h)}^2(\sum_{h=1}^{H}\alpha_h'^2f_1S_{\bar{x}h}^2) - 2\beta_{1(h)}\beta_{2(h)}(\sum_{h=1}^{H}\alpha_h'^2f_2S_{\bar{x}h}^3\lambda_{03}) + \beta_{1(h)}^2(\sum_{h=1}^{H}\alpha_h'^2f_1S_{\bar{x}h}^2) + 2\beta_{1(h)}\beta_{2(h)}(\sum_{h=1}^{H}\alpha_h'^2f_1S_{\bar{x}h}^3\lambda_{03}) + \beta_{2(h)}^2(\sum_{h=1}^{H}\alpha_h'^2f_1S_{\bar{x}h}^4(\lambda_{04} - 1))$$

$$= \textstyle\sum_{h=1}^{H}\alpha_h'^2[f_1S_{\bar{y}h}^2 + f_4W_{h2}S_{\bar{y}h2}^2] - 2\beta_{1(h)}(\sum_{h=1}^{H}\alpha_h'^2f_3\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}) -$$

$$2\beta_{2(h)}\left(\sum_{h=1}^{H}\alpha_h'^2 S_h^2 f_1 S_{yh}\lambda_{12}\right)-\beta_{1(h)}^2\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^2\right)+2\beta_{1(h)}\beta_{2(h)}\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^3\lambda_{03}\right)+\beta_{2(h)}^2\left(\sum_{h=1}^{H}\alpha_h'^2 f_1 S_{\bar{x}h}^4(\lambda_{04}-1)\right) \quad (13)$$

Differentiating and solving with respect to $\beta_{1(h)}$ and $\beta_{2(h)}$ respectively give

$$\beta_{1(hopt)}=\frac{\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^3\lambda_{03}\right)\left(\sum_{h=1}^{H}\alpha_h'^2 S_h^2 f_1 S_{yh}\lambda_{12}\right)-\left(\sum_{h=1}^{H}\alpha_h'^2 f_3\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}\right)\left(\sum_{h=1}^{H}\alpha_h'^2 f_1 S_{\bar{x}h}^4(\lambda_{04}-1)\right)}{\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^2\right)\left(\sum_{h=1}^{H}\alpha_h'^2 f_1 S_{\bar{x}h}^4(\lambda_{04}-1)\right)+\left(\sum_{h=1}^{H}\alpha_h'^2 f_2 S_{\bar{x}h}^3\lambda_{03}\right)^2}$$

$$\beta_{2(hopt)}=\frac{\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^2\right)\left(\sum_{h=1}^{H}\alpha_h'^2 S_h^2 f_1 S_{ydh}\lambda_{12}\right)+\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^3\lambda_{03}\right)\left(\sum_{h=1}^{H}\alpha_h'^2 f_3\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}\right)}{\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^2\right)\left(\sum_{h=1}^{H}\alpha_h'^2 f_1 S_{\bar{x}h}^4(\lambda_{04}-1)\right)+\left(\sum_{h=1}^{H}\alpha_h'^2 f_2 S_{\bar{x}h}^3\lambda_{03}\right)^2}$$

Substituting for $\beta_{1(hopt)}$ and $\beta_{2(hopt)}$ into (13) we have

$$V\left(\hat{\bar{Y}}_{dpps}\right)_{min}=\sum_{h=1}^{H}\alpha_h'^2\left[f_1 S_{\bar{y}h}^2+f_4 W_{h2}S_{\bar{y}h2}^2\right]-2\beta_{1(hopt)}\left(\sum_{h=1}^{H}\alpha_h'^2 f_3\rho_{\bar{x}\bar{y}}S_{\bar{x}h}S_{\bar{y}h}\right)-2\beta_{2(hopt)}\left(\sum_{h=1}^{H}\alpha_h'^2 S_h^2 f_1 S_{yh}\lambda_{12}\right)-\beta_{1(hopt)}^2\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^2\right)+2\beta_{1(hopt)}\beta_{2(hopt)}\left(\sum_{h=1}^{H}\alpha_h'^2 f_3 S_{\bar{x}h}^3\lambda_{03}\right)+\beta_{2(hopt)}^2\left(\sum_{h=1}^{H}\alpha_h'^2 f_1 S_{\bar{x}h}^4(\lambda_{04}-1)\right) \quad (14)$$

Equation (14) is the minimum variance of the proposed estimator in probability proportional to size.

RESULTS AND DISCUSSION

EMPIRICAL STUDY

To validate the theoretical claims, this study adopts the data from the Sweden municipalities MU284 (Sarndal et al. 1992, appendix B). The population is geographically sub-divided into eight different parts (domains) 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 estimators are calibration estimators. The required variables were computed based on existing information from the populations. In the case of double sampling, we computed the first phase sample size (n_h') and the second phase sample size (n_h). Hence, we assumed n_h' to be 70% of the stratum size (N_h) in both populations. Then each of the domains is classified into homogeneous according to our convenient into two strata: value below 1500 (millions of kronor) and above 1500 (millions of kronor). We consider two cases I and II of non-response (in both Population 1 and Population 2).

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

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

Population 1

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

X: Total number of municipal employees in 1984.

Table 1: Statistics of the Parameter for the Domains across the Strata for population 1

DOMAIN PARAMETER	DOMAIN							
DOMAIN SIZE	25		32		15		29	
STRATUM	1	2	1	2	1	2	1	2
N_h	2	23	12	20	2	13	18	11
W_h	0.080	0.920	0.375	0.625	0.133	0.867	0.620	0.379
$\bar{Y}_h$	955.50	6888	1056.9	3364	1231	4020	723.2	4799
$\bar{X}_h$	529	4385	485	1816	493	1694	354	2205
$S_{\bar{Y}h}^2$	40.5	136775663	71715.5	4652460	135721	5643626	81863.7	10236271
$S_{\bar{X}h}^2$	101250	81259476	29239.7	2530489	162	2354475	18128.3	2902966
S_{XYh}	-2025	104701660	34761.1	3144594	-4689	2999017	13306	3557068
ρ_{XYh}	-1.000	0.993	0.759	0.916	-1.000	0.823	0.345	0.653
$\frac{1}{p_{hi}}$	14.081	75773.23	2204.417	29643.79	8.008	8511.403	8898.614	8520.82
$\frac{y_{hi}}{p_{hi}}$	146089.5	4540415	157002.1	1599412	4942.67	932014	254689.7	1041294
$\frac{x_{hi}}{p_{hi}}$	2116	2319573	69900	726400	1972	286286	114714	266838
X_h	1058	100851	5825	36320	986	22022	6373	24258
Y_h	1911	158422	12683	67277	2461	52266	13017	52787
S_{Yhi}^2	4263166	2.64E+10	1.41E+08	4.615E+09	1833280	3.263E+09	1.81E+08	6.428E+09
S_{Xhi}^2	279841	930568916	28511328.2	11905260	243049	41322758	36227157	48633481
S_{XYhi}	-4651.9	-210191	-13917.9	-72922.6	-4900.61	-59187.1	-254316	-60368.8
λ_{12}	1.089745	1.013467	1.805651	1.635687	1.675385	1.385685	1.099383	1.300187
λ_{03}	1.097657	1.380915	1.284943	1.385578	1.097381	1.003987	1.328766	1.087474
λ_{04}	1.078602	1.097466	1.200367	1.316240	1.073478	1.010431	1.408913	1.063587

Source: Statistical computation from original data 2026

Table 2: The parameter estimates for the Domains across the Strata in case I

DOMAIN	STRATA	S^2_{Yh2}	S^2_{Xh2}	S_{XYh2}	k_h	n_{2h}	w_{h2}	n_{1h}	n_h	n'_h
1	1	0	0	0	3	0	0.3	0	0	1
	2	223415888	132328227	171255299	2	4	0.3	10	14	16
2	1	120583	9862.3	28095.5	2	1	0.3	2	3	8
	2	3977507	2517165	2669307	2	3	0.3	8	11	14
3	1	0	0	0	2	0	0.3	0	0	1
	2	987699	1771955	1137277	3	1	0.3	3	4	9
4	1	79129	20311.8	8114.3	2	3	0.3	6	9	13
	2	141512	5618	-28196	2	1	0.3	1	2	8

Source: Statistical computation from original data 2026

Table 3: The parameter estimates for the Domains across the Strata in case II

DOMAIN	STRATA	S^2_{Yh2}	S^2_{Xh2}	S_{XYh2}	k_h	n_{2h}	W_{h2}	n_{1h}	n_h	n'_h
1	1	0	0	0	2	0	0.2	0	0	1
	2	256872328	152987958	197463945	3	5	0.4	7	12	16
2	1	120583	9862.3	28095.5	2	1	0.2	2	3	8
	2	3977507	2517165	2669307	3	4	0.4	7	11	14
3	1	0	0	0	2	0	0.2	0	0	1
	2	987699	1771955	1137277	4	2	0.4	2	4	9
4	1	79129	20311.8	8114.3	2	2	0.2	7	9	13
	2	141512	5618	-28196	3	1	0.4	1	2	8

Source: Statistical computation from original data 2026

Population 2

Another population is considered (Sarndal et al. 1992, 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).

Table 4: Statistics of the Parameter for the Domains across the Strata for population 2

DOMAIN PARAMETER	DOMAIN							
DOMAIN SIZE	25		32		15		29	
STRATUM	1	2	1	2	1	2	1	2
N_h	2	23	14	18	7	8	20	9
W_h	0.080	0.920	0.438	0.563	0.467	0.533	0.690	0.310
$\bar{Y}_h$	75.5	594	67.5	260.6	73.0	315.0	44.6	345.2
$\bar{X}_h$	9.00	67.1	10.6	34.5	10.7	40.6	6.6	41.9
S^2_{Yh}	840.50	1551426	275.96	41200.80	369.67	51631.70	187.21	54848.90
S^2_{Xh}	18.00	16649.60	4.56	544.97	6.91	731.13	5.103	681.61
S_{XYh}	123.00	160633.5	32.038	4559.147	49.500	6116.143	29.839	6076.778
ρ_{XYh}	1.000	0.999	0.904	0.962	0.980	0.995	0.965	0.994

$\frac{1}{p_{hi}}$	11.25	60403.37	3113.304	12019.2	405.202	1213.689	11927.27	1900.777
$\frac{y_{hi}}{p_{hi}}$	309	295862.5	13155.13	80503.97	26350.49	214526.9	18136.48	27169.94
$\frac{x_{hi}}{p_{hi}}$	18	1543	149	621	75	325	131	377
X_h	578	983	541	229	306	94	283	225
Y_h	4719	9083	3670	1966	2249	782	2080	1918
S^2_{Yhi}	81	2178281	19142.66	343982.3	4132.69	80866.3	15487.8	112298.7
S^2_{Xhi}	17.16667	191.745	88.28943	129.636	47.5801	60.63821	138.4464	72.06882
λ_{12}	1.54877	1.08961	1.36774	1.93846	1.30965	1.08653	1.04573	1.36574
λ_{03}	1.34801	1.97464	1.34856	1.38447	1.67248	1.02357	1.28351	1.23734
λ_{04}	1.27324	1.90231	1.29362	1.45232	1.25561	1.20721	1.15342	1.20825

Table 5: The parameter estimates for the Domains across the Strata in case I

DOMAIN	STRATA	S^2_{Yh2}	S^2_{Xh2}	S_{XYh2}	k_h	n_{2h}	W_{h2}	n_{1h}	n_h	n'_h
1	1	0	0	0	2	0	0.3	0	0	1
	2	2478255	26533.0	256331.0	2	4	0.3	10	14	16
2	1	373.70	4.200	35.050	2	2	0.3	3	5	10
	2	64541.6	875.250	7146.750	2	3	0.3	6	9	13
3	1	0	0	0	2	0	0.3	0	0	5
	2	0	0	0	2	0	0.3	0	0	6
4	1	168.16	5.018	27.945	3	3	0.3	8	11	14
	2	0	0	0	2	0	0.3	0	0	6

Source: Statistical computation from original data 2026

Table 6: The parameter estimates for the Domains across the Strata in case II
Source: Statistical computation from original data 2026

DOMAIN	STRATA	S^2_{Yh2}	S^2_{Xh2}	S_{XYh2}	k_h	n_{2h}	W_{h2}	n_{h1}	n_h	n'_h
1	1	0	0	0	2	0	0.2	0	0	1
	2	2654913	28395.1	274463.7	3	5	0.4	8	13	16
2	1	176.25	1.333	9.667	2	1	0.2	3	4	10
	2	64184.8	874.3	7069.214	2	3	0.4	5	8	13
3	1	0	0	0	2	0	0.2	0	0	5
	2	0	0	0	3	0	0.4	0	0	6
4	1	169.778	5.511	30.000	2	2	0.2	8	10	14
	2	0	0	0	3	0	0.4	0	0	6

Table 7: Variance of the Existing and Proposed Estimators

ESTIMATOR	Case 1(Population1)				AMSE
	1	2	3	4	
$(\hat{Y}_j)2002a$	2.14096E+14	1.66927E+13	2.89814E+11	1.10157E+12	5.8045E+13
$(\hat{Y}_j)2002b$	1.65058E+13	1.92926E+12	3.66462E+11	1.81095E+11	4.7457E+12
$(\hat{Y}_h)$	1.73802E+13	1.09775E+16	1.31411E+17	6.05391E+12	3.5603E+16
$(\bar{Y}_{dPPS})min$	3.46515E+09	4.34417E+09	2.16504E+09	8.44616E+09	4.61630E+09
		Case II (Population1)			
$(\hat{Y}_j)2002a$	2.14096E+14	1.66927E+13	2.89814E+11	1.10157E+12	5.8045E+13

$(\hat{Y}_j)2002b$	1.65058E+13	1.92926E+12	3.66462E+11	1.81095E+11	4.7457E+12
$(\hat{Y}_h)$	1.73802E+13	1.09775E+16	1.31411E+17	6.05391E+12	3.5603E+16
$(\bar{Y}_{dPPS})min$	3.41279E+08	7.10620E+08	2.20223E+08	2.12329E+10	5.63041E+09
	Case 1 (Population 2)				
$(\hat{Y}_j)2002a$	1.70763E+14	1.41753E+12	0	1.63695E+11	5.74481E+13
$(\hat{Y}_j)2002b$	7.00851E+10	4.64034E+09	0	2.47131E+08	2.49909E+10
$(\hat{Y}_h)$	5.79685E+14	1.02461E+12	0	1.39587E+12	1.94035E+14
$(\bar{Y}_{dPPS})min$	2.36553E+09	4.36163E+08	0	6.35160E+07	7.16933E+08
	Case II (Population2)				
$(\hat{Y}_j)2002a$	1.70763E+14	1.41753E+12	0	1.63695E+11	5.74481E+13
$(\hat{Y}_j)2002b$	7.00851E+10	4.64034E+09	0	2.47131E+08	2.49909E+10
$(\hat{Y}_h)$	5.79685E+14	1.02461E+12	0	1.39587E+12	1.94035E+14
$(\bar{Y}_{dPPS})min$	2.07653E+10	1.01611E+08	0	2.01313E+07	5.22260E+09

Note: Average Mean Square Error (AMSE)

This discussion is based on the empirical analysis carried out and results presented in Tables. From the result in Table 7 (Populations I and 2), it is observed that the mean squared error for the proposed estimator $(\bar{Y}_{dPPS})min$ is less than that of the existing estimators $(\hat{Y}_j)\,2002a$, $(\hat{Y}_j)\,2002b$ and $(\hat{Y}_h)$ in all the domains. This is seen in both cases of nonresponse where the nonresponse rate was uniform across the strata and where it was non-uniform as specified in the data. The Average Mean Square Errors (AMSE) also confirms the behavior and performance of the proposed and existing estimators, with the former being superior to the latter. From the result, it is obvious that the use of calibration approach has been fruitful in the estimation of domain means through the adjustment of survey weights and reduction in sampling variance, which conforms to the literature by Iseh and Bassey (2021).

Again, from the empirical results of Table 7, it was noticed that the proposed estimator performed better in case II with non-uniform nonresponse rate of population 1 in domains 1, 2 and 3 than in domain 4, which made the average performance of the estimators to be in favor of case I under uniform nonresponse rate. However, in population 2, the suggested estimator performed better in case II under domains 2 and 4 but failed in domain 1. Again, the average performance of the estimator was noticeably better in case I. This signifies that even with different rates of nonresponse in the domains of study, the estimators ensure consistency in performance except in areas of inadequate sample size. It is worth noting that the domain estimators have gained efficiency in the domains of study than as seen in the aggregate performance. This is in line with Udofia (2004) that there would be no need for additional estimation if domains of study are the same with strata. Besides, it was impossible to compute the mean square error of the estimators in domain 3 of population 2 because there were zero sample unit in the two strata. This collaborates with the literature on the challenges of domain estimation when there are zero sample size in domains of interest (see Sarndal et. al. 1992, Iseh & Bassey 2021).

CONCLUSION

This study develops the concept of calibration estimator for double sampling with probability proportional to size (PPS). 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 nonrespondents when there is nonresponse in the study variable and auxiliary variable is free from nonresponse. Apart from the gains in efficiency which the proposed estimator has over the existing estimators, it was noticed that this study has identified the individual performance of estimators within domains of interest instead of relying on the aggregate performance. This is principally as a result of the weights adjustment which has compensated for unit nonresponse as well as aligning survey results with known population results.

Again, 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 nonresponse using the concept of calibration, and by so doing encourages performance within the respective domains of interest. Futthermore, this concept has identified that the proposed estimators performed better in some domain of interest when the nonresponse rate is not uniform, which is practically a real-life application though the aggregate result suggested otherwise. In fact, this attribute of the proposed estimator can be useful in estimation involving highly skewed survey-data like income and expenditure.

From the empirical work and efficiency comparison, it becomes pertinent that the use of calibration technique has really paid off in providing estimates of the population mean with sub-sampling the nonrespondents that provides greater gains in efficiency better than the existing estimators. It suffices to say that, this study is made possible through the theory of stratified random sampling which has made the computations much easier. However, future study involving other sampling designs are encouraged to broaden this concept in domains of study. Again, it is also pertinent to suggest future research to enable researchers and users of statistics to compute estimates of the population mean when there are small/zero sampling units in domains of interest.

Acknowledgement

The authors gratefully acknowledge the insightful feedback provided by the reviewers

REFERENCE

Agrawal, M. C., and Mrida, C. K. (2007). Some efficient estimators of the domain parameters. Statistics and Probability Letters, 77: 704-709.

Amahia, G. N., Chaubey Y. P., and Rao T. J. (1989). Efficiency of a new PPS sampling for multiple characteristics, J. Stat. Plan. Inference, 21:75-84.

Bansal M. L. and Singh R., (1985). An alternative estimator for multiple characteristics in PPS sampling, J. Stat. Plan. Inference, 11(3): 313-320.

Brewer, K. R. W., and Hanif, M. (1983). Sampling with unequal probabilities. Springer, New York.

Chaudhuri, A. (2010). Essentials of survey sampling, PHI Learning Private Limited, New Delhi.

Des Raj (1964). On double sampling for PPS estimation, Ann. Math.Statist., 35(2): 900-902. difference cum ratio estimator with full response on the auxiliary character. Res. J. App. Sci., 2, 769-772.

Deville, J. C. and Sarndal, C. E., (1992). Calibration estimators in survey sampling, Journal of the American Statistical Association, 87: 376‒382.

Enang, E. I and Amahia, G. N. (2012). A Class of alternative estimators in probability proportional to size sampling with replacement for multiple characteristics. Journal of Mathematics Research; 4(3). 66-77

Godambe V. P. (1955). A unified theory of sampling from finite population, J. Roy. Stat. Soc. B., 17: 269-278.

Grewal, I. S, Bansal, M. L, and Singh (1997). An alternative estimator for multiple characteristics using randomized response technique in PPS sampling. Aligarh J. Stat., 19:51-65.

Hansen, M. H., and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. Journal of the American Statistical Association, 41:517-529.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953a). Sample survey methods and theory, Volume 1-Methods and Applications, New York, John Wiley and Sons.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (l953b). Sample survey methods and theory, Volume 2-Theory, New York. John Wiley and Sons.

Ikughur A. J. and Amahia, G. N, (2011). A generalized transformation for selection probabilities in unequal probability sampling scheme, J. Sci. Ind. Stud., 9(1): 58-62.

Iseh, M. J. and Enang, E. I. (2021). A calibrated synthetic estimator for small area estimation. Statistics in Transition New Series, 22 (3):15-30. DOI 10.21307/stattrans-2021-025.

Iseh, M. J. and Bassey, M. O. (2022). Calibration estimators for population mean with subsampling the nonrespondents under stratified sampling. Science Journal of Applied Mathematics and Statistics, 10(4): 45-56. doi: 10.11648/j.sjams.20221004.11.

Iseh, M. J, and Bassey, K. J. (2021). A new calibration estimator of population mean for small area with non-response. Asian Journal of Probability and Statistics 12(2):41-51. DOI: 10.9734/AJPAS/2021/v12i230286

Iseh, M. J. (2023). Model formulation on efficiency for median estimation under fixed cost in survey sampling, Model Assisted Statistica and Applications. 18(4): 373-385. DOI 10.3233/MAS-231437

Michael, I. T., Usoro, A. E., Ikpang, N. I., George, E. U. and Bassey, D. I. (2019). Fitting a gamma distribution using a chi-squared approach to heights of students of Akwa Ibom State University, Nigeria. International Journal of Advanced Statistics and Probability, 7(2): 42-46.

Michael, I. T., Iseh, M. J., Ikpang, N. I., George, E. U. and Bassey, D. I. (2021). A goodness of fit test: using chi-squared approach to fit the lognormal probability model to the weights of students of Akwa Ibom State University. International Journal of Applied Science and Research, 4(6): 146-152.

Michael, I. T., Isaac, A. A. and Etim, A. E. (2025). Fitting a normal distribution to the heights of Akwa Ibom State University students using chi-square. Asian Journal of Probability and Statistics, 27(2): 42-49.

Okereke, C. U., Ekpenyong, E. J., Usoro, A. E., Iseh, M. J. and Iwuji, A. C. (2026). Comparative analysis of conditional error distributions in asymmetric GARCH modelling. Journal of Basic and Applied Sciences Research (JOBASR), 4(3):120-129. https://dx.doi.org/10.4314/jobasr.v4i3.13

Pathak, P. K. (1966). An estimator in PPS sampling for multiple characteristics. Sankhya, A. 28(1):35-40.

Rao, J. N. K. (1966). Alternative estimators in PPS sampling for multiple characteristics, Sankhya: The Indian Journal of Statistics, Series A, 28: 47–60.

Sarndal, C. E., Swensson, B., and Wretman, J. (1992). Model assisted survey sampling, Springer-Verlag, New York. 24(2):167-191.

Tillé, Y. (1996). An elimination procedure for unequal probability sampling without replacement, Biometrika, 83: 238-241.

Udofia, G. A. (2004). Ratio estimation for small domains with subsampling the non-respondents: An application of Rao strategy; Statistics in Transition, 6(5): 713-724.

Udofia, G. (2002a). Estimation for domains in double sampling for probabilities proportional to known size. Global Journal of Pure and Applied Science, 1(1): 67-74

Udofia G. (2002b). On the precision of an estimator of mean for domains in double sampling for inclusion probabilities, Global Journal of Pure and Applied Science. 1(2): 49-58.

Yates, F. (1953). Sampling methods for censuses and surveys. London: Charles W. Griffin.