Science Journal of Applied Mathematics and Statistics
2022; 10(4): 45-56
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20221004.11
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

Calibration Estimators for Population Mean with Subsampling the Nonrespondents Under Stratified Sampling

Iseh Matthew Joshua, Bassey Mbuotidem Okon

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

Email address:
matthewiseh@aksu.edu.ng (Iseh Matthew Joshua), okonmbuotidem@gmail.com (Bassey Mbuotidem Okon)

To cite this article:
Iseh Matthew Joshua, Bassey Mbuotidem Okon. Calibration Estimators for Population Mean with Subsampling the Nonrespondents Under Stratified Sampling. *Science Journal of Applied Mathematics and Statistics*. Vol. 10, No. 4, 2022, pp. 45-56.
doi: 10.11648/j.sjams.20221004.11

Received: August 7, 2022; **Accepted**: August 27, 2022; **Published**: September 16, 2022

Abstract: The existence of nonresponse in survey sampling has engendered inconsistencies in the estimation of population parameter. Such estimation, being characterized by nonresponse bias has become a rule rather than the exception in survey sampling, and this has been long acknowledged in the literature. Several authors have come up with different techniques such as subsampling the nonresponse, imputation, and calibration to curb this menace. An attempt to overcome the challenges faced in existing works, this study considered the estimation of finite population mean using calibration approach with subsampling the nonrespondents. Owning to the fact that calibration estimation has been found to reduce bias and improve efficiency of estimators. The classical estimator by Hansen and Hurwitz for estimating the population mean with subsampling the nonrespondents is calibrated upon using the chi square distance function, and different choices of the tunning parameter result in the calibration estimators of combined regression and ratio. Expressions for the bias, variance and mean square error (MSE) of the proposed estimators are derived and their properties studied. Again, the optimum conditions under which the suggested estimators have minimum variance and MSE are equally provided. Both efficiency and empirical comparisons are in favor of the proposed estimators, and suggest that the proposed estimators are more efficient and reliable with high precision than the existing estimators even in double sampling. In addition, expressions for optimal sample sizes with respect to the cost of the survey have been derived to validate the superiority of the proposed estimators, and the empirical investigation confirms the proposed estimators as highly preferable.

Keywords: Auxiliary Variable, Calibration, Nonresponse, Stratified Sampling, Sub-sampling

1. Introduction

In sample survey, the term non-response refers to the failure to collect information from one or more respondents on one or more variables. This becomes challenging in estimation, as the estimates of the population characteristics may be highly biased, and consequently may lead to wrong conclusions. Overtime, the usual estimation methods where the nonresponse is ignored assumed that the estimates based on the respondents are representatives of the combined population of respondents and non-respondents, and this assumption leads to an unknown bias. In order to utilize the advantages, M. H. Hansen and N. W. Hurwitz [1] developed an estimator which was the weighted mean of two estimators: the sample mean based on the units responded, and the sample mean based on the units of the sub-sample selected from the non-respondents.

M. H. Hansen and N. W. Hurwitz [1] presented a classical estimator for estimation of mean of a characteristic of interest with subsampling the non-respondents suitable for different practical situations. Authors like K. M. Chaudhary et al. [2], K. M. Chaudhary, V. K. Singh and R. K. Shukla [3], S. Kumar [4], K. M. Chaudhary and A. Kumar [5], and A. Sanaullah, I. Elisan and M. Noor-UI-Amin [6] etc. are among the several authors who have adopted or extended the M. H. Hansen and N. W. Hurwitz [1] estimator by utilizing supplementary information and some known population parameters of the auxiliary variable using the conventional methods to obtain improved estimates of the population total/mean of the study variable. To further enhance the performance of estimators with subsampling the nonrespondents using two auxiliary variables, G. N. Singh and M. Usman [7] suggested some regression-cum-ratio estimators to estimate the population mean while N. Garib and U. Mahamood [8] proposed both difference and ratio estimators for the estimation of population distribution function of the study variable. The authors concluded that estimators with two auxiliary variables performed better than those with single auxiliary variable.

In the progression to improve on the estimator of the population mean, A. E. Anieting and E. I. Enang [9], A. E. Anieting, E. I. Enang and C. E. Onwukwe [10], S. Guha and H. Chandra [11], and M. K Chaudhary, A. Kumar, and G. K. Vishwarkarma [12] extended the pioneer work on subsampling the nonrespondents by Hansen and Hurwitz to double sampling. However, M. J. Iseh and K. J. Bassey [13-14] suggested an estimator in small area estimation in the presence of nonresponse using calibration technique to improve on both sample size as well as gain in efficiency, and their estimators had an overwhelming less bias and gain in efficiency.

The question now becomes; what difference will calibration techniques make, since it is the process of incorporating auxiliary variable(s) into the existing estimator?

This becomes testable following the argument by J. C. Deville and C. E. Sarndal [15] in support of the calibration approach. They argued that if the calibration constraint is satisfied for unbiasedness of the auxiliary variable that it will be reasonably profitable to combine the calibration weight with the sample mean/total of the study variable in practice, if the auxiliary variable and the study variable are strongly correlated either positively or negatively. This concept of calibration was also proven to be fruitful in M. J. Iseh, and E. I. Enang [16], where the use of auxiliary variable was seen to greatly reduced the bias of the synthetic estimator.

Believing we make a good choice of the calibration constraints, this approach will be more beneficial to researchers in practical sense than the conventional approach of formulating estimators of the study variable.

As an attempt in this direction, to bridge the gap in choosing from a list of available population parameters of the auxiliary variable, and methods of combining them, this study employs the calibration approach to achieve a reliable and preferable result. Two estimators are proposed from the existing classical estimator of M. H. Hansen and N. W. Hurwitz [1] through the conceptualization of calibration technique.

2. Sampling Design

Consider a population of size N divided into k strata. Let N_i be the size of i^{th} stratum $(i = 1,2,3..........,k)$ and a sample of size n_i is drawn from the i^{th} stratum using SRSWOR scheme such that $\sum_{i=1}^{k} n_i = n$. It is assumed that nonresponse is detected on the study variable Y only and auxiliary variable X is free from nonresponse. Also, assume that of the n_i units, n_{i1} respond and $n_{i2} = n_i - n_{i1}$ fail to respond. Adopting M. H. Hansen and N. W. Hurwitz [1] techniques of subsampling the non-respondents, a sub sample of $u_{i2} = \frac{n_{i2}}{k_i}$, $k_i \geq 1$ unit is selected from the sample of n_{i2} nonrepondents and information is collected from all of them. Due to the detection of nonresponse in the study variable, the classical estimator is being calibrated upon in order to compensate for nonresponse bias and high variance of the estimate.

2.1. Hansen and Hurwitz Classical Estimator

A classical estimator proposed by M. H. Hansen and N. W. Hurwitz [1] for estimating the population mean with subsampling the non-respondents is given as follows:

$$\bar{y}_{st}^{*} = \sum_{i=1}^{k} P_i \bar{y}_i^{*}$$

where $\bar{y}_i^{*} = \frac{n_{i1}\bar{y}_{ni1} + n_{i2}\bar{y}_{ui2}}{n_i}$

$$V(\bar{y}_{st}^{*}) = \sum_{i=1}^{k} f_i P_i^2 S_{yi}^2 + \sum_{i=1}^{k} \frac{(K_i - 1)}{n_i} P_i^2 W_{i2} S_{iy2}^2 \quad (1)$$

where S_{yi1}^2 and S_{yi2}^2 are respectively the mean squares of entire group and non-response group of study variables in the population for the i^{th} stratum. $f_i = \left(\frac{1}{n_i} - \frac{1}{N_i}\right)$, $k_i = \frac{n_{i1}}{u_{i2}}$, $P_i = \frac{N_i}{N}$, and $W_{i2} = \frac{N_{i2}}{N_i}$ is the non-response rate of the i^{th} stratum in the population.

2.2. Some Existing Estimators with Auxiliary Variable

Some existing estimators for population mean in stratified sampling with single auxiliary variable with subsampling the non-respondents are presented as follows:

K. M. Chaudhary, et al. [2] Ratio Estimator.

A separate ratio estimator of population mean $\bar{Y}$ proposed by K. M. Chaudhary, et al. [2] is given as

$$T_s = \sum_{i=1}^{k} P_i T_i^{*}$$

where $T_i^{*} = \bar{y}_i^{*}\left[\frac{a\bar{X}_i + b}{\sigma(a\bar{x}_{st} + b) + (1-\sigma)(a\bar{X}_i + b)}\right]^g$

If $\sigma = 1, a = 1, b = 0$ and $g = -1$,

The estimator gives

$$T_s = \sum_{i=1}^{k} P_i \bar{y}_i^{*} \frac{\bar{x}_i}{\bar{X}_i}$$

And the bias and mean square error given as

$$B(T_s) = \sum_{i=1}^{k} = P_i f_i \bar{Y}_i \left[\frac{g(g+1)}{2}\sigma^2\lambda_i^2 C_{xi}^2 - \sigma\lambda_i g\rho_i C_{yi} C_{xi}\right]$$

$$\mathrm{MSE}\ (\mathrm{T_s}) = \sum_{i=1}^{k} \mathrm{P_i^2} \left[\mathrm{f_i}\, \overline{\mathrm{Y}}_i^2 \left(\mathrm{C_{yi}^2} + \mathrm{C_{xi}^2}\, \sigma^2\lambda_i^2 - 2\lambda\sigma g\, \rho_i\, \mathrm{C_{yi}C_{xi}}\right) + \frac{(\mathrm{k_i}-1)}{\mathrm{n_i}}\, \mathrm{W_{i2}\, S_{Yi2}^2}\right] \quad (2)$$

$$\mathrm{MSE}\ (\mathrm{T_s})_{min} = \sum_{i=1}^{k} \mathrm{P_i^2} \left[\mathrm{f_i}\, \overline{\mathrm{Y}}_i^2 \left(\mathrm{C_{yi}^2} + \mathrm{C_{xi}^2}\, \sigma^2\lambda_i^2 - 2\lambda\sigma g\, \rho_i\, \mathrm{C_{yi}C_{xi}}\right) + \frac{(\mathrm{k_{i(opt)}}-1)}{\mathrm{n_{i(opt)}}}\, \mathrm{W_{i2}\, S_{Yi2}^2}\right]$$

K. M. Chaudhary, V. K. Singh and R. K. Shukla [3] Combined Ratio Estimator.

A combined ratio type estimator proposed by K. M. Chaudhary, V. K. Singh and R. K. Shukla [3] with subsampling the non-respondents is given as

$$\mathrm{T_{FC}}\ (\sigma) = \bar{y}_{st}^{*} \left[\frac{(\mathrm{A+C})\overline{\mathrm{X}}\, \mathrm{fB}\bar{\mathrm{x}}_{st}}{(\mathrm{A+B})\overline{\mathrm{X}} + \mathrm{C}\bar{\mathrm{x}}_{st}}\right]$$

Where $\mathrm{A} = (\sigma-1)(\sigma-2), \mathrm{B} = (\sigma-1)(\sigma-4), \mathrm{C} = (\sigma-3)(\sigma-2)(\sigma-4)$,

$$\sigma > 0, f = \frac{\mathrm{n_i}}{\mathrm{N_i}}, \bar{y}_{st}^{*} = \sum_{i=1}^{k} \mathrm{P_i}\, \bar{y}_i^{*} \text{ and } \bar{\mathrm{x}}_{st} = \sum_{i=1}^{k} \mathrm{P_i}\, \bar{\mathrm{x}}_i$$

$$\mathrm{MSE}\big(\mathrm{T_{FC}}(\sigma)\big) = \sum_{i=1}^{k} \mathrm{f_i}\, \mathrm{P_i^2}\ \left[\mathrm{S_{yi}^2} + \phi(\sigma)^2\mathrm{RS}^2 - 2\phi(\sigma)\mathrm{R}\rho_i\mathrm{S_{xi}S_{yi}}\right] \quad (3)$$

where $\phi(\sigma) = \frac{\mathrm{C-fB}}{\mathrm{A+fB+C}}$

K. M. Chaudhary and A. Kumar [5] Generalized Combined Ratio Estimator.

A family of combined ratio-type estimators proposed by K. M. Chaudhary and A. Kumar [5] with subsampling the nonrespondents is given as follows:

$$T_c' = \bar{y}_{st}^{*} \left[\frac{a\bar{x}_{st}' + b}{\propto(a\bar{\mathrm{x}}_{st}+b) + (1-\propto)\left(a\bar{x}_{st}'+b\right)}\right]^g$$

With mean square error given as

$$MSE(T_c') = \sum_{i=1}^{k} f_i'\, p_i^2 S_{yi}^2 + \sum_{i=1}^{k} f_i^{*}\, p_i^2\left(S_{yi}^2 + g^2\lambda^2R^2 \propto^2 S_{xi}^2 - 2g\lambda R \propto p_i S_{xi} S_{yi}\right) + \sum_{i=1}^{k} \frac{(k_i-1)}{n_i} W_{i2}\, p_i^2 S_{yi}^2 \quad (4)$$

Where $f_i' = \left(\frac{1}{n_i'} - \frac{1}{\mathrm{N_i}}\right), f_i^{*} = \left(\frac{1}{n_i} - \frac{1}{n_i'}\right), \lambda = \frac{a\bar{X}}{a\bar{X}+b}$, and $\bar{X}_i = \sum_{i=1}^{k} \bar{x}_i$

2.3. Some Existing Estimators Under Double Sampling

Estimator from A. E. Anieting and E. I. Enang [9]

A product-ratio estimator under double sampling proposed by A. E. Anieting and E. I. Enang [9] is given as follows

$$Tae = \bar{y}_{st}^{*} \left(\frac{\bar{x}_{st}' + \Phi}{\bar{\mathrm{x}}_{st} - \Phi}\right)$$

Where $\Phi = \sum_{i=1}^{k} C_{xi}$

$$\bar{y}_{st}^{*} = \sum_{i=1}^{k} P_i\, \bar{y}_i^{*};$$

$$\bar{y}_i^{*} = \frac{n_{i1}\bar{y}_{ni1} + n_{i2}\bar{y}_{ni2}}{n_i}$$

where $\bar{x}_{st}' = \sum_{i=1}^{k} P_i\, \bar{x}_i'$ and mean square error given as

$$MSE(Tae) = \left\{b^2R^2 \sum_{i=1}^{k}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)P_i^2 S_{yi}^2\right] - 2abR^2\right\} \sum_{i=1}^{k}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)P_i^2 S_{xi}^2\right] + a^2R^2 \sum_{i=1}^{k}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)P_i^2 S_{xi}^2\right] - 2bR \sum_{i=1}^{k}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)P_i^2 \rho_i S_{xi} S_{yi} + 2aR\right] + \sum_{i=1}^{k}\left\{\left(\frac{1}{n_i} - \frac{1}{N_i}\right)P_i^2 \rho_i S_{xi} S_{yi} + \sum_{i=1}^{k}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)P_i^2 S_{yi}^2 + \frac{(k_i-1)}{n_i} P_i^2 W_{i2} S_{yi}^2\right]\right\} \quad (5)$$

Where $R = \frac{\bar{Y}}{\bar{X}}, a = \frac{\bar{X}}{\bar{X}+\Phi}$, and $b = \frac{\bar{X}}{\bar{X}-\Phi}$,

Estimator from A. E. Anieting, E. I. Enang and C. E. Onwukwe [10].

An estimator of population mean proposed by A. E. Anieting, E. I. Enang and C. E. Onwukwe [10] in double sampling with subsampling the nonrespondents using single auxiliary variable as follows:

$$Tae1 = \bar{y}_{st}^{*} \left(\frac{\bar{x}_{st}' - \rho_i}{\bar{\mathrm{x}}_{st} - \rho_i}\right)$$

with mean square error

$$MSE(Tae1) = \left[\left(\frac{\bar{X}}{\bar{X}-\rho_i}\right)^2 R^2 \sum_{i=1}^{K}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\rho_i^2 S_{xi}^2\right] - \left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)^2 R^2 \sum_{i=1}^{K}\left[\left(\frac{1}{n_i'}-\frac{1}{N_i}\right)P_i^2 S_{xi}^2\right] - 2R\left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)\sum_{i=1}^{K}\left(\frac{1}{n_i}-\frac{1}{N_i}\right)P_i^2\rho_i\rho_i S_{xi}S_{yi} + 2\left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)R\sum_{i=1}^{k}\left(\frac{1}{n_i'}-\frac{1}{N_i}\right)p_i^2\rho_i S_{xi}S_{yi} + \sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\frac{(k_i-1)}{n_i}P_i^2 W_{i2}S_{yi2}^2\right]\right] \quad (6)$$

$$MSE(Tae1)_{min} = \left\{\left(\frac{\bar{X}}{\bar{X}-\rho_i}\right)^2 R^2 \sum_{i=1}^{K}\left[\left(\frac{1}{n_{i(opt)}}-\frac{1}{N_i}\right)\rho_i^2 S_{xi}^2\right] - \left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)^2 R^2 \sum_{i=1}^{K}\left[\left(\frac{1}{n_{i(opt)}^1}-\frac{1}{N_i}\right)P_i^2 S_{xi}^2\right] - 2R\left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)\sum_{i=1}^{K}\left(\frac{1}{n_{i(opt)}}-\frac{1}{N_i}\right)P_i^2\rho_i\rho_i S_{xi}S_{yi} + 2\left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)R\sum_{i=1}^{k}\left(\frac{1}{n_{i(opt)}'}-\frac{1}{N_i}\right)p_i^2\rho_i S_{xi}S_{yi} + \sum_{i=1}^{k}\left[\left(\frac{1}{n_{i(opt)}}-\frac{1}{N_i}\right)\frac{(k_{i(opt)}-1)}{n_{i(opt)}}P_i^2 W_{i2}S_{yi2}^2\right]\right\}$$

3. Proposed Estimators

Motivated by M. H. Hansen and N. W. Hurwitz [1], we proposed the following estimator

$$\bar{y}_c^* = \sum_{i=1}^{k}\Omega_i\,\bar{y}_i^* \quad (7)$$

where Ω_i is the calibration weight chosen such that the distance function.
$\sum_{i=1}^{k}\frac{(\Omega_i-P_i)^2}{q_i p_i}$ is minimize subject to the calibration constraint

$$\sum_{i=1}^{k}\Omega_i\,\bar{x}_i = \bar{X} \quad (8)$$

And the optimization problem becomes

$$\Phi = \sum_{i=1}^{k}\frac{(\Omega_i-P_i)^2}{q_i p_i} - 2\lambda\left[\sum_{i=1}^{k}\Omega_i\,\bar{x}_i - \bar{X}\right]$$

which is minimized with respect to the calibration weights such that

$$\Omega_i = P_i[1 + \lambda\bar{x}_i q_i] \quad (9)$$

Substituting Eq. (9) in Eq. (8), and solving for λ yields.
$\lambda = \left[\frac{\bar{X}-\sum_{i=1}^{k}P_i\bar{x}_i}{\sum_{i=1}^{k}P_i\bar{x}_i^2}\right]$ Using the value of λ in Eq. (7) gives the required calibration estimator

$$\bar{y}_c^* = \sum_{i=1}^{k}P_i\,\bar{y}_i^* + \sum_{i=1}^{k}P_i q_i\bar{x}_i\,\bar{y}_i^*\left[\frac{\bar{X}-\sum_{i=1}^{k}P_i\,\bar{x}_i}{\sum_{i=1}^{k}P_i q_i\bar{x}_i^2}\right] \quad (10)$$

By Setting $q_i = 1$ in Eq. (10) gives

$$\bar{y}_{creg}^* = \sum_{i=1}^{k}P_i\,\bar{y}_i^* + \frac{\sum_{i=1}^{k}P_i\bar{x}_i\,\bar{y}_i^*}{\sum_{i=1}^{k}P_i\bar{x}_i^2}\left[\bar{X} - \sum_{i=1}^{k}P_i\,\bar{x}_i\right]$$

which is in the form of the combined regression, and can be written as

$$\bar{y}_{creg}^* = \sum_{i=1}^{k}P_i\,\bar{y}_i^* + \hat{\beta}\left[\bar{X} - \sum_{i=1}^{k}P_i\,\bar{x}_i\right] \quad (11)$$

where $\hat{\beta} = \frac{\sum_{i=1}^{k}p_i\,\bar{x}_i\bar{y}_i^*}{\sum_{i=1}^{k}p_i\bar{x}_i^2}$
using Taylor's series approximation method, we proceed as follows in deriving the variance of the suggested estimator in Eq. (11).

Let,

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

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

$$E(e_0) = E(e_1) = 0$$

$$E(e_0^2) = \frac{V(\bar{y}_i^*)}{\bar{Y}_I^2} = f_i C_{xi}^2 + \frac{(k_i-1)}{n_i\bar{Y}_i^2}W_{i2}\,S_{yi2}^2$$

$$E(e_1^2) = \frac{V(\bar{x}_i)}{\bar{X}_I^2} = f_i C_{xi}^2$$

$$E(e_0 e_1) = \frac{Cov(\bar{y}_i^*, \bar{x}_i)}{\bar{Y}_i \bar{X}_i} = f_i \rho_i C_{yi} C_{xi}$$

where $f_i = \left(\frac{1}{n_i} - \frac{1}{N_i}\right)$, $C_{xi}^2 = \frac{S_{xi2}^2}{\bar{X}_i^2}$, $C_{yi}^2 = \frac{S_{yi2}^2}{\bar{Y}_i^2}$

$$E\left(\bar{y}_{creg}^*\right) = \sum_{i=1}^{k} P_i \bar{Y}_i (1 + E(e_0)) + \beta\left[\bar{X} - \sum_{i=1}^{k} P_i \bar{X}_i (1 + E(e_1))\right]$$

$$= \bar{Y} + \beta[\bar{X} - \bar{X}]$$

$$= \bar{Y}$$

$$V\left(\bar{y}_{creg}^*\right) = E\left[\bar{y}_{creg}^* - E\left(\bar{y}_{creg}^*\right)\right]^2$$

$$= E\left[\sum_{i=1}^{k} P_i \bar{y}_i^* + \beta\left(\bar{X} - \sum_{i=1}^{k} P_i \bar{x}_i\right) - \bar{Y}\right]^2$$

$$= E\left[\sum_{i=1}^{k} P_i \bar{Y}_i (1 + e_0) + \beta\left(\bar{X} - \sum_{i=1}^{k} P_i \bar{X}_i (1 + e_1)\right) - \bar{Y}\right]^2$$

$$= \sum_{i=1}^{k} P_i^2 \left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi}^2 + \frac{(K_i - 1)}{n_i} W_{i2} S_{yi2}^2\right] + \beta^2 \sum_{i=1}^{k} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{xi}^2 - 2\beta \sum_{i=1}^{k} P_i^2 \rho_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi} S_{xi} \qquad (12)$$

Optimum Choice of β

From Eq. (12)

$$\frac{\partial V\left(\bar{y}_{creg}^*\right)}{\partial \beta} = 2\beta \sum_{i=1}^{k} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{xi}^2 - 2 \sum_{i=1}^{k} P_i^2 \rho_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi} S_{xi} = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{k} P_i^2 \rho_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi} S_{xi}}{\sum_{i=1}^{k} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{xi}^2}$$

Substituting the value of $\hat{\beta}$ in Eq. (12)

$$V\left(\bar{y}_{creg}^*\right) = \sum_{i=1}^{k} P_i^2 \left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi}^2 + \frac{(K_i - 1)}{n_i} W_{i2} S_{yi2}^2\right] + \left[\frac{\sum_{i=1}^{k} P_i^2 \rho_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi} S_{xi}}{\sum_{i=1}^{k} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{xi}^2}\right]^2 \sum_{i=1}^{k} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{xi}^2 - 2\left[\frac{\sum_{i=1}^{k} P_i^2 \rho_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi} S_{xi}}{\sum_{i=1}^{k} P_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{xi}^2}\right] \sum_{i=1}^{k} P_i^2 \rho_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi} S_{xi}$$

$$V\left(\bar{y}_{creg}^*\right) = \sum_{i=1}^{k} P_i^2 \left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi}^2 - \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \rho_i^2 S_{yi}^2 + \frac{(K_i - 1)}{n_i} W_{i2} S_{yi2}^2\right] \qquad (13)$$

Optimum Sample Size Determination for n_i and k_i

Let C_{i0} be he cost per unit of selecting n_h units, C_{i1} be the cost per unit in enumerating n_{i1} units and C_{i2} be the cost per unit of enumerating n_{i2} units, then the total cost for the i^{th} stratum is given by $C_i = C_{i0} n_i + C_{i1} n_{i1} + C_{i2} n_{i2} \; \forall \, i = 1, 2, \ldots, k$.

Now, we consider the average cost per stratum

$$E(C_i) = n_i \left[C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{i2}}{K_i}\right]$$

Thus the total cost over all the strata becomes

$$C_0 = \sum_{i=1}^{k} E(C_i)$$

$$C_0 = \sum_{i=1}^{k} n_i \left[C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{i2}}{K_i}\right] \qquad (14)$$

Consider the Lagrange function

$$\Phi = V\left(\bar{y}_{creg}^*\right) + \mu C_0$$

$$\Phi = \sum_{i=1}^{k} P_i^2 \left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_{yi}^2 - \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \rho_i^2 S_{yi}^2 + \frac{(K_i - 1)}{n_i} W_{i2} S_{yi2}^2 + \mu \sum_{i=1}^{k} n_i \left[C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{i2}}{K_i}\right]\right] \qquad (15)$$

$$\frac{\partial\Phi}{\partial n_i} = -\frac{1}{n_i^2}P_i^2S_{yi}^2 - \frac{(K_i-1)}{n_i^2}P_i^2W_{i2}S_{yi2}^2 + \frac{1}{n_i^2}\rho_i^2S_{yi}^2P_i^2 + \mu\left(C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{K_i}\right) = 0$$

$$n_i = \frac{\sqrt{P_i^2\left[(1-\rho_i^2)S_{yi}^2+(K_i-1)W_{i2}S_{yi2}^2\right]}}{\sqrt{\mu\left(C_{i0}+C_{i1}W_{i1}+C_{i2}\frac{W_{i2}}{K_i}\right)}} \tag{16}$$

Again, by taking partial derivative of Eq. (15) with respect to K_i

$$\frac{\partial\Phi}{\partial K_i} = \frac{P_i^2W_{i2}S_{yi2}^2}{n_i} - \mu n_i C_{i2}\frac{W_{i2}}{K_i^2} = 0$$

$$K_i = \sqrt{\frac{\mu n_i^2 C_{i2}}{P_i^2 S_{yi2}}} \tag{17}$$

$$\mu = \frac{P_i^2S_{yi2}^2K_i^2}{n_i^2C_{i2}}$$

$$\sqrt{\mu} = \frac{P_iS_{yi2}K_i}{n_i\sqrt{C_{i2}}} \tag{18}$$

Substitute Eq. (18) in Eq. (16) gives

$$n_i = \frac{P_i\sqrt{\left[(1-\rho_i^2)S_{yi}^2+(K_i-1)W_{i2}S_{yi2}^2\right]}}{\frac{P_iS_{yi2}K_i}{n_i\sqrt{C_{i2}}}\sqrt{\left(C_{i0}+C_{i1}W_{i1}+C_{i2}\frac{W_{i2}}{K_i}\right)}} \tag{19}$$

Then solving for K_i in Eq. (17) becomes

$$K_{i(opt)} = \frac{\sqrt{\left[(1-\rho_i^2)S_{yi}^2-W_{i2}S_{yi2}^2\right]C_{i2}}}{S_{yi2}\sqrt{(C_{i0}+C_{i1}W_{i1})}}$$

$$K_{i(opt)} = \frac{B_i\sqrt{C_{i2}}}{S_{yi2}A_i} \tag{20}$$

$$A_i = \sqrt{(C_{i0} + C_{i1}W_{i1})},\ \text{and}\ B_i = \sqrt{\left[(1-\rho_i^2)S_{yi}^2 - W_{i2}S_{yi2}^2\right]}$$

substituting Eq. (20) in Eq. (16), we have

$$n_i = \frac{P_i\sqrt{\left[B_i^2+\frac{W_{i2}S_{yi2}^2\sqrt{C_{i2}}B_i}{A_h}\right]}}{\sqrt{\mu}\sqrt{A_i^2+\frac{W_{i2}S_{yi2}^2B_i\sqrt{C_{i2}}}{B_i}}} \tag{21}$$

To obtain n_{iopt}, we substitute Eq. (21) and Eq. (20) in Eq. (14)

$$C_o = \sum_{i=1}^{k}\left[\frac{P_i\sqrt{B_i^2+\frac{\sqrt{C_{i2}}B_iW_{i2}S_{yi2}}{A_i}}}{\sqrt{\mu}\sqrt{A_i^2+\frac{\sqrt{C_{i2}}A_iW_{i2}S_{yi2}}{B_i}}}\left(C_{i0} + C_{i1}W_{i1} + C_{i2}\frac{W_{i2}}{\frac{\sqrt{C_{i2}}B_i}{S_{yi2}A_i}}\right)\right]$$

$$C_o\sqrt{\mu} = \sum_{i=1}^{k}\left[P_i\left(B_iA_i + \sqrt{C}_{i2}W_{i2}S_{yi2}\right)\right]$$

$$\sqrt{\mu} = \frac{1}{C_o}\sum_{i=1}^{k}\left[P_i\left(B_iA_i + \sqrt{C}_{i2}W_{i2}S_{yi2}\right)\right]$$

$$n_{i(opt)} = \frac{P_iC_o\sqrt{B_{i+}^2\frac{W_{i2}S_{yi2}\sqrt{C_{i\,2}}B_i}{A_i}}}{\sqrt{A_i^2+\frac{\sqrt{C_{i\,2}}W_{i2}S_{yi2}A_i}{B_i}}\sum_{i=1}^{k}\left[P_i\left(B_iA_i+\sqrt{C}_{i2}W_{i2}S_{yi2}\right)\right]} \tag{22}$$

Substituting Eq. (22) in Eq. (13) gives the minimum variance of the proposed estimator.

$$V_{min}\left(\bar{y}^*_{creg}\right) = \sum_{i=1}^{k} P_i^2 \left\{\left(\frac{1}{n_{i(opt)}} - \frac{1}{N_i}\right)\left[(1-\rho_i^2)S_{yi}^2\right]\frac{(K_{i(opt)}-1)}{n_{i(opt)}} + W_{i2}S_{yi2}^2\right\} \tag{23}$$

3.1. The Proposed Combined Ratio Estimator

Setting $q_i = \frac{1}{\bar{x}_i}$ in Eq. (10) gives

$$\bar{y}^*_{cr} = \frac{\sum_{i=1}^{k} P_i \bar{y}_i^*}{\sum_{i=1}^{k} P_i \bar{x}_i}\bar{X} \tag{24}$$

which is in the form of combined ratio estimator with subsampling the non-respondents. Since it is a ratio estimator, it is assumed biased, and the bias is derived as follows

$$\bar{y}^*_{cr} = \frac{\sum_{i=1}^{k} P_i \bar{Y}_i(1+e_0)}{\sum_{i=1}^{k} P_i \bar{X}_i(1+e_1)}\bar{X}$$

$$\bar{y}^*_{cr} = \sum_{i=1}^{k} P_i \bar{Y}_i (1+e_0)(1+e_1)^{-1}$$

By Taylor's approximation method, we have

$$\bar{y}^*_{cr} = \sum_{i=1}^{k} P_i \bar{Y}_i (1+e_0)\left[1 - e_1 + \frac{1}{2}(1+1)e_1^2\right]$$

$$\bar{y}^*_{cr} - \bar{Y} = \sum_{i=1}^{k} P_i \bar{Y}_i [e_0 - e_1 + e_1^2 - e_0 e_1] \tag{25}$$

Taking expectation of Eq. (25) to the first order of approximation, we obtain

$$E(\bar{y}^*_{cr} - \bar{Y}) = B\left(\hat{\bar{y}}_c\right) = \sum_{i=1}^{k} P_i \bar{Y}_i E(e_0 - e_1 + e_1^2 - e_0 e_1)$$

$$B(\bar{y}^*_{cr}) = \sum_{i=1}^{k} P_i \bar{Y}_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right)\left[C_{yi}^2 - \rho_i C_{yi} C_{xi}\right] \tag{26}$$

Squaring both sides of Eq. (25), and taking expectation of both sides gives

$$E(\bar{y}^*_{cr} - \bar{Y}_i)^2 = \sum_{i=1}^{k} P_i^2 \bar{Y}_i^2 \left[E(e_0^2) - 2E(e_0 e_1) + E(e_1^2)\right]$$

$$MSE\left(\bar{y}^*_{cr}\right) = \sum_{i=1}^{k} P_i^2 \bar{Y}_i^2 \left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + \frac{(K_i - 1)}{n_i} W_{i2} S_{yi2}^2\right] \tag{27}$$

3.2. Optimum Sample Size Determination for n_i and k_i

Recall the total cost over all strata in Eq. (14), then the optimization problem becomes

$$\Delta = V(\bar{y}^*_{cr}) + \lambda C_0 \tag{28}$$

$$\Delta = \sum_{i=1}^{k} P_i^2 \bar{Y}_i^2 \left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + \frac{(K_i - 1)}{n_i} W_{i2} S_{yi2}^2\right] + \lambda \sum_{i=1}^{k} n_i \left[C_{io} + C_{i1} W_{i1} + C_{i2}\frac{W_{i2}}{K_i}\right] \tag{29}$$

$$\frac{\partial \Delta}{\partial n_i} = -\frac{P_i^2}{n_i^2}\left[\bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + (K_i - 1) W_{i2} S_{yi2}^2\right] + \lambda\left[C_{io} + C_{i1} W_{i1} + C_{i2}\frac{W_{i2}}{K_i}\right] = 0$$

$$n_i = p_i \frac{\sqrt{\left[\bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + (k_i - 1) W_{i2} S_{yi2}^2\right]}}{\sqrt{\lambda}\sqrt{C_{io} + C_{i1} W_{i1} + C_{i2}\frac{W_{i2}}{K_i^2}}} \tag{30}$$

Again, taking the partial derivative of Eq. (29) with respect to K_i

$$\frac{\partial \Delta}{\partial K_i} = \frac{P_i^2 W_{i2} S_{yi2}^2}{n_i} - \lambda n_i C_{i2}\frac{W_{i2}}{K_i^2} = 0$$

$K_i = \sqrt{\frac{\lambda n_i^2 C_{i2}}{P_i^2 S_{yi2}^2}}$ Squaring both sides, and solving for λ, we have

$$\lambda = \frac{P_i^2 S_{yi2}^2 K_i^2}{n_i^2 C_{i2}}$$

$$\sqrt{\lambda} = \frac{P_i S_{yi2} K_i}{n_i \sqrt{C_{i2}}} \tag{31}$$

Substitute Eq. (31) in Eq. (30), and solve for K_{iopt}

$$n_i = P_i \frac{\sqrt{\bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + (K_i - 1) W_{i2} S_{yi2}^2}}{\frac{K_i P_i s_{yi2}}{n_i \sqrt{C_{i2}}} \sqrt{\left(C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{hi2}}{K_i}\right)}}$$

$$K_{i(opt)} = \frac{\sqrt{[\bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{Xi}\right) - W_{i2} S_{yi2}^2]\, C_{i2}}}{S_{yi2} \sqrt{C_{i0} + C_{i1} W_{i1}}}$$

$$K_{iopt} = \frac{D_i \sqrt{C_{i2}}}{S_{yi2} A_i} \tag{32}$$

where $A_i = \sqrt{C_{i0} + C_{i1} W_{i1}}$, and

$D_i = \sqrt{[\bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{Xi}\right) - W_{i2} S_{yi2}^2]}$ Substituting Eq. (32) in Eq. (30), we express n_i as follows

$$n_i = \frac{P_i \sqrt{D_{i+}^2 \frac{W_{i2} S_{yi2} \sqrt{C_{i2}} D_i}{A_i}}}{\sqrt{\lambda} \sqrt{A_i^2 + \frac{\sqrt{C_{i2}} W_{i2} S_{yi2} D_i}{D_i}}} \tag{33}$$

As before, to obtain the value of $\sqrt{\lambda}$ in terms of the total cost, we substitute Eq. (33) and Eq. (32) in Eq. (14)

$$C_o = \sum_{i=1}^{k} \left\{ \frac{P_i \sqrt{D_i^2 + \frac{\sqrt{C_{i2}} D_i W_{i2} S_{yi2}}{A_i}}}{\sqrt{\lambda} \sqrt{A_i^2 + \frac{\sqrt{C_{i2}} A_i W_{i2} S_{yi2}}{D_i}}} \left(C_{i0} + C_{i1} W_{i1} + C_{i2} \frac{W_{i2}}{\frac{\sqrt{C_{i2}} D_i}{S_{yi2} A_i}} \right) \right\}$$

$$C_o \sqrt{\lambda} = \sum_{i=1}^{k} \left\{ \frac{P_i \sqrt{D_i^2 + \frac{\sqrt{C_{i2}} D_i W_{i2} S_{yi2}}{A_i}}}{\sqrt{A_i^2 + \frac{\sqrt{C_{i2}} A_i W_{i2} S_{yi2}}{D_i}}} \left(A_i^2 + \frac{\sqrt{C_{i2}} A_i W_{i2} S_{yi2}}{D_i} \right) \right\}$$

$$C_o \sqrt{\lambda} = \sum_{i=1}^{k} \left[P_i \sqrt{\left(D_i^2 A_i^2 + 2 D_i A_i \sqrt{C}_{i2} W_{i2} S_{yi2} + C_{i2} W_{i2}^2 S_{yi2}^2 \right)} \right]$$

$$\sqrt{\lambda} = \frac{1}{C_o} \sum_{i=1}^{k} \left[P_i \left(B_i A_i + \sqrt{C}_{i2} W_{i2} S_{yi2} \right) \right] \tag{34}$$

substituting Eq. (34) in Eq. (33), we obtain the optimum sample size as

$$n_{iopt} = \frac{P_i C_o \sqrt{D_{i+}^2 \frac{W_{i2} S_{yi2} \sqrt{C_{i2}} D_i}{A_i}}}{\sqrt{A_i^2 + \frac{\sqrt{C_{i2}} W_{i2} S_{yi2} A_i}{D_i}} \sum_{i=1}^{k} [P_i (D_i A_i + \sqrt{C}_{i2} W_{i2} S_{yi2})]} \tag{35}$$

Therefore, by substituting Eq. (35) in Eq. (27), the minimum mean square error of the proposed calibration ratio estimator with subsampling the non-respondents is obtain as

$$MSE(\bar{y}_{cr}^*)_{min} = \sum_{i=1}^{k} P_i^2 \left[\bar{Y}_i^2 \left(\frac{1}{n_{i(opt)}} - \frac{1}{N_i} \right) \left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi} \right) + \frac{(K_{i(opt)} - 1)}{n_{i(opt)}} W_{i2} S_{yi2}^2 \right] \tag{36}$$

3.3. Efficiency Comparison of the Proposed Estimators with the Existing Estimators Under Single Stage Sampling

Condition I: Comparing Eq. (27) with Eq. (1)

$$\mathrm{MSE}\ (\bar{y}^*_{cr}) < V(\bar{y}^*_{st})\ \text{if}$$

$$\sum_{i=1}^{k} P_i^2 \bar{Y}_i^2 f_i\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + \frac{(K_i-1)}{n_i} W_{i2}\, S_{Yi2}^2 < \sum_{i=1}^{k} P_i^2 f_i S_{yi}^2 + \sum_{i=1}^{k} P_i^2 \frac{(K_i-1)}{n_i} W_{i2}\, S_{iy2}^2$$

$$\Rightarrow C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi} < C_{yi}^2$$

$$\Rightarrow \rho_i < \frac{C_{xi}^2}{2C_{yi}C_{xi}}$$

If this condition is satisfied, then MSE $(\bar{y}^*_{cr})$ will be better than $V(\bar{y}^*_{st})$
Condition II: Comparing Eq. (13) with Eq. (1)

$$V\left(\bar{y}^*_{creg}\right) < V(\bar{y}^*_{st})\ \text{if}$$

$$\sum_{i=1}^{k} P_i^2 \left[f_i S_{yi}^2 - \rho_i^2 f_i S_{yi}^2 + \left(\frac{K_i-1}{n_i}\right) W_{i2} S_{yi2}^2\right] < \sum_{i=1}^{k} P_i^2\, f_i S_{yi}^2 + \sum_{i=1}^{k} P_i^2 \left(\frac{K_i-1}{n_i}\right) W_{i2}\, S_{yi2}^2$$

$$\Rightarrow (1-\rho_i^2) f_i S_{yi}^2 < f_i S_{yi}^2$$

$$\Rightarrow 1-\rho_i^2 < 0$$

$$\Rightarrow \rho_i^2 < 1$$

If this condition is satisfied, then $V\left(\bar{y}^*_{creg}\right)$ will be better than $V(\bar{y}^*_{st})$
Condition III: Comparing Eq. (27) with Eq. (2)

$$MSE(\bar{y}^*_{cr}) < MSE(T_s)\ \text{if}$$

$$\sum_{i=1}^{k} P_i^2 \left[\bar{Y}_i^2 f_i\left(C_{yi}^2 + C_{xi}^2 - 2\rho_i C_{yi} C_{xi}\right) + \frac{(k_i-1)}{n_i} W_{i2} S_{iy2}^2\right] < \sum_{i=1}^{k} P_i^2 \left[\bar{Y}_i^2 f_i\left(C_{yi}^2 + C_{xi}^2\sigma^2\lambda_i^2 - 2\lambda\sigma g \rho_i C_{yi} C_{xi}\right) + \frac{(K_i-1)}{n_i} W_{i2} S_{iy2}^2\right]$$

$$\Rightarrow C_{xi}^2 - 2\rho_i C_{xi} C_{yi} < C_{xi}^2\sigma^2\lambda_i^2 - 2\lambda\sigma g \rho_i C_{xi} C_{yi}$$

By setting $\sigma = 1, \lambda = 1$, and $g = -1$

$$\Rightarrow C_{xi}^2 - 2\rho_i C_{xi} C_{yi} < C_{xi}^2 + 2\rho_i C_{xi} C_{yi}$$

$$\Rightarrow 4\rho_i C_{yi} C_{xi} > 0$$

$$\Rightarrow \rho_i > 0$$

$MSE(\bar{y}^*_{cr})$ will be more efficient if the following conditions are satisfied; $\rho_i > 0, \sigma = 1, \lambda = 1$, and $g = -1$.
Condition IV: Comparing Eq. (13) with Eq. (2)

$$V\left(\bar{y}^*_{creg}\right) < MSE(T_s)\ \text{if}$$

$$\sum_{i=1}^{k} P_i^2 \left[f_i S_{yi}^2 - \rho_i^2 f_i S_{yi}^2 + \left(\frac{K_i-1}{n_i}\right) W_{i2} S_{yi2}^2\right] < \sum_{i=1}^{k} P_i^2 \left[f_i \bar{Y}_i^2\left(C_{yi}^2 + \sigma^2\lambda_i^2 g^2 C_{xi}^2 - 2\lambda\sigma_i g \rho_i C_{yi} C_{xi}\right) + \frac{(K_i-1)}{n_i} W_{i2} S_{iy2}^2\right]$$

$$S_{yi}^2(1-\rho_i^2) < \bar{Y}_i^2\left(C_{yi}^2 + \sigma^2\lambda_i^2 g^2 C_{xi}^2 - 2\lambda\sigma_i g \rho_i C_{xi} C_{yi}\right)$$

Again, by setting $\sigma = 1, \lambda = 1$, and $g = -1$,
$V\left(\bar{y}^*_{creg}\right)$ will be more efficient if and only if;
For $\rho_i > 0$

$$\Rightarrow S_{yi}^2(1-\rho_i^2) - \bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 + 2\rho_i C_{yi} C_{xi}\right) < 0$$

For $\rho_i = 0$

$$\Rightarrow C_{xi}^2 > 0$$

For $\rho_i = 1$

$$\Rightarrow \bar{Y}_i^2\left(C_{yi}^2 + C_{xi}^2 + 2C_{yi} C_{xi}\right) > 1$$

4. Empirical Study

This section will deal extensively on numerical analysis to validate the theoretical results earlier derived. Data considered is from N. Koyuncu, and C. Kadilar [17] used for the single stage sampling.

Table 1. *Data statistics from koyuncu and kadilar (2009).*

Stratum no.	N_i	n_i	$\bar{Y}_i$	$\bar{X}_i$	S_{Yi}	S_{Xi}	S_{XYi}	ρ_i	S_{Yi2}
1	127	31	703.74	20804.59	883.835	30486.751	25237153.52	0.936	440
2	117	21	413.00	9211.79	644.922	15180.769	9747942.85	0.996	200
3	103	29	573.71	14309.30	1033.467	27549.697	28294397.04	0.994	400
4	170	39	424.66	9478.85	810.585	18218.931	14523885.53	0.983	405
5	205	22	527.03	5569.95	403.654	8497.776	3393591.75	0.989	180
6	201	39	393.84	12997.59	711.723	23094.141	1586473.97	0.965	300

Table 2. *MSE and PRE of the proposed and existing estimators in single stage design.*

W_{i2}	K_i	$V(\bar{y}_{st})$	$MSE(T_s)$	$V(\bar{y}^*_{creg})$	$MSE(\bar{y}^*_{cr})$	$PRE(T_s)$	$PRE(\bar{y}^*_{creg})$	$PRE(\bar{y}^*_{cr})$
0.1	2.0	1443.1	1203.337	102.46	116.43	119.2	1408.452	1239.457
	2.5	1461.3	1221.576	120.66	134.62	119.6	1211.089	1085.5
	3.0	1479.5	1239.774	138.87	152.82	119.3	1065.385	968.1324
	3.5	1497.7	1257.973	157.06	171.02	119.9	954.5846	875.7455
0.2	2.0	1500.3	1239.774	161.23	175.77	119.9	930.534	853.5586
	2.5	1515.9	1236.172	175.86	189.22	120.1	861.9925	801.131
	3.0	1552.3	1312.569	211.63	225.61	120.9	733.4971	688.0457
	3.5	1588.7	1348.967	248.05	262.02	121.5	640.4757	606.3278
0.3	2.0	1612.5	1276.171	254.67	278.55	120.9	633.1723	578.8907
	2.5	1620.4	1330.768	263.54	288.11	122.7	614.8592	562.4241
	3.0	1625.9	1385.364	274.51	298.41	138.8	592.f2917	544.8544
	3.5	1679.7	1430.690	290.34	353.01	157.6	578.5286	475.8222
0.4	2.0	1681.1	1312.569	320.11	359.87	161.3	525.1632	467.1409
	2.5	1690.4	1385.364	340.45	368.45	175.6	496.5193	458.7868
	3.0	1738.3	1458.195	383.87	410.43	211.3	452.8356	423.5314
	3.5	11770.7	1530.954	430.03	444.00	248.5	411.762	398.8063

Table 3. *Optimal sample sizes, costs of study and minimum variances.*

C_o = \$ 17,357 C_{i1} = \$10,097 C_{i2} = \$6230								
W_{i2}	K_i	$n_{i(cr\ opt)}$	$K_{i(cr\ opt)}$	$n_{i(creg\ opt)}$	$K_{i(creg\ opt)}$	$MSE(T_s)_{min}$	$V_{min}(\bar{y}^*_{creg})$	$MSE(\bar{y}^*_{cr})_{min}$
0.1	2.0	10	0.5	8	0.3	79.1	67.1	70.3
	2.5	11	0.6	4	0.4	100.2	70.5	79.4
	3.0	11	0.7	6	0.5	107.1	73.2	79.9
	3.5	9	1.0	5	0.1	109.4	75.7	80.5
0.2	2.0	5	0.5	3	0.2	107.2	80.3	85.3
	2.5	5	0.6	3	0.3	120.1	83.1	88.4
	3.0	6	0.7	3	0.4	120.9	87.3	90.1
	3.5	5	1.0	3	0.1	123.1	89.5	95.7
0.3	2.0	10	0.5	5	0.4	120.8	90.4	95.9
	2.5	8	0.6	4	0.5	131.7	90.7	99.8
	3.0	7	0.7	3	0.3	134.2	92.9	100.3
	3.5	7	1.0	3	0.1	137.1	95.4	107.1
0.4	2.0	9	0.5	6	0.2	140.4	97.6	111.3
	2.5	5	0.6	3	0.4	140.7	99.7	125.1
	3.0	10	0.7	5	0.2	142.1	100.1	139.9
	3.5	11	1.0	6	0.1	147.5	105.3	141.7

For the double sampling, data from K. M. Chaudhary and A. Kumar [5] is used to validate the theoretical claims as shown in Table 4.

Table 4. *Data statistics from [5] for double sampling design.*

Stratum No.	N_i	n'_i	n_i	$\bar{Y}_i$	$\bar{X}_i$	S^2_{yi}	S^2_{xi}	P_i	S^2_{yi2}
1	73	65	26	40.85	39.56	6389.1	6624.44	0.999	618.88
2	70	25	10	27.57	27.57	1051.07	1147.01	0.998	240.91
3	97	48	19	25.44	25.44	2014.97	2205.4	0.999	265.52
4	44	11	5	20.36	20.36	538.47	485.27	0.997	83.69

Table 5. *Variances and mses of the proposed and existing estimators under single/double sampling for different choices of W_{i2} and K_i.*

W_{i2}	K_i	$V(\bar{y}^*_{st})$	$MSE(T'_c)$	$MSE(Tae)$	$MSE(T_{ae1})$	$V(\bar{y}^*_{creg})$	$MSE(\bar{y}^*_{cr})$
0.1	2	34.42	6.28	4.66	4.38	0.19	0.21
	2.5	34.67	6.54	4.92	4.64	0.29	0.30
	3	34.92	6.79	5.17	4.89	0.38	0.39

W_{i2}	K_i	$V(\bar{y}^*_{st})$	$MSE(T'_c)$	$MSE(Tae)$	$MSE(T_{ae1})$	$V(\bar{y}^*_{creg})$	$MSE(\bar{y}^*_{cr})$
	2	34.92	6.79	5.26	4.89	0.38	0.39
0.2	2.5	35.43	7.30	5.66	5.64	0.56	0.58
	3	34.94	7.80	6.18	6.12	0.75	0.76
	2	35.43	7.30	5.66	5.40	0.56	0.58
0.3	2.5	36.19	8.06	6.44	6.43	0.48	0.85
	3	36.95	8.86	7.22	7.12	1.11	1.12

Table 6. *PRE of the proposed estimators under single phase with respect to $\bar{y}^*_{st}$ and the existing double sampling estimators for the different choices of W_{i2} and K_i.*

W_{i2}	K_i	$PRE(\bar{y}^*_{st})$	$PRE(T'_c)$	$PRE(T_{ae})$	$PRE(T_{ae1})$	$PRE(\bar{y}^*_{creg})$	$PRE(\bar{y}^*_{cr})$
	2	100	540.1	738.6	785.8	18115.8	16390.5
0.1	2.5	100	530.1	704.7	747.2	11955.2	11556.7
	3	100	514.3	675.4	714.1	1294.8	1261.5
	2	100	514.3	663.9	714.1	9189.4	8953.8
0.2	2.5	100	485.3	626.0	628.2	6326.8	6108.6
	3	100	447.9	581.6	587.3	4658.7	4597.4
	2	100	485.3	625.97	656.1	6326.786	6108.621
0.3	2.5	100	449.0	561.96	562.8	7539.583	4257.647
	3	100	417.0	511.8	518.9	3328.829	3299.107

Table 7. *Costs of study and optimal sample sizes.*

$c_0 = \$16327$ $c_{i1} = \$10,097$ $c_{i2} = \$6,230$							
W_{i2}	K_i	$n_{i(ae1\ opt)}$	$k_{i(ae1\ opt)}$	$K_{i(cr\ opt)}$	$n_{i(cr\ opt)}$	$n_{i(creg\ opt)}$	$K_{i(creg\ opt)}$
	2	9.7	0.51	0.13	5.1	4.0	0.10
0.1	2.5	9.9	0.60	0.15	6.7	4.2	0.12
	3	10.1	0.93	0.19	9.2	4.9	0.18
	2	10.1	0.93	0.21	9.2	4.9	0.20
0.2	2.5	10.3	1.10	0.22	9.5	5.7	0.23
	3	10.5	1.12	0.34	8.1	7.8	0.29
	2	10.3	1.10	0.37	9.8	5.7	0.20
0.3	2.5	11.1	2.3	0.41	10.1	7.9	0.30
	3	11.7	2.8	0.47	11.3	8.1	0.37

Table 8. *Minimum mses and variance for optimal sample sizes of the proposed estimators and the existing double sampling estimator.*

W_{i2}	K_i	$Min\ MSE(T_{ae1})$	$V_{min}(\bar{y}^*_{creg})$	$MSE(\bar{y}^*_{cr})_{min}$
	2	3.01	0.013	0.35
0.1	2.5	3.52	0.017	0.36
	3	3.57	0.018	0.41
	2	3.57	0.018	0.41
0.2	2.5	4.31	0.019	0.47
	3	4.99	0.02	0.51
	2	4.25	0.019	0.47
0.3	2.5	5.33	0.030	0.52
	3	5.49	0.039	0.55

5. Discussion

From Table 2, it is observed that the estimator of the proposed family certainly provides better estimates with gains in efficiency as compared to the classical estimator $\bar{y}^*_{st}$, and the generalized ratio estimator by K. M. Chaudhary, et al.[2] T_s. It is also seen that the efficiency of the existing estimators decrease rapidly with an increase in non-response rate W_{i2} as well as with an increase in inverse sampling rate K_i. However, the loss in efficiency was not remarkable in the proposed estimator. This outstanding performance can clearly be attributed to the concept of calibration as mentioned in the literature. Similarly, the performance of the existing and proposed estimators were examined for minimum variance and MSEs in Table 3, and the result reveals that the suggested estimators outperformed the existing estimator T_s in terms of gains in efficiency with optimal sample sizes.

Again, considering also the data-set used by K. M. Chaudhary and A. Kumar [5], it is observed from Tables 5 and 6 that the estimates of the population means for the proposed estimators $\bar{y}^*_{creg}$ and $\bar{y}^*_{cr}$ than the existing estimators $\bar{y}^*_{st}$, T'_c, T_{ae}, and T_{ae1} at different choices of inverse sampling rates K_i. In real life application, it suffices to say that the proposed estimators are highly preferred in terms of performance and simplicity, compared to the rigorous nature and complexity in handling the bias and MSEs of the existing generalized combined ratio estimator T'_c and the product-ratio estimators in double sampling T_{ae}, and T_{ae1}.

As shown in Table 8, minimizing the variance and mean square error for optimal sample sizes in respect to the cost of the survey on the existing estimator T_{ae1}, as well as the proposed estimators $\bar{y}^*_{creg}$ and $\bar{y}^*_{cr}$, were computed. The result shows that the proposed estimators give minimum variance and means square error than the existing estimator under double sampling.

6. Conclusion

This work was aimed at reducing the effect of biasness and variance or mean square error as the case may be because of nonresponse from the proposed estimator through the concept of calibration. As seen from both the efficiency comparison, and the numerical evaluations, the proposed estimators outperformed the existing estimators, even that of the double sampling design. It is evident to say that the calibration technique suggests the best fit of auxiliary variable into an existing estimator through the formulation of constraint equation(s), and this conceptualization guarantees stability in gains of efficiency even with increase in the nonresponse and inverse sampling rates, whereas the existing estimators considered in this study, both single and two phase that adopt the conventional approach of combining supplementary information with the study variable cannot compete favorably. The proposed estimators are also shown to be preferred in terms of cost effectiveness and minimum variance and mean square error. Consequently, if nonresponse is only measured in the study variable and the auxiliary variable is free from nonresponse, it will be fruitful to adopt the concept of calibration for optimal result. However, in a situation where the population mean is not known, the authors have considered a case of double sampling using calibration approach in future work.

References

[1] M. H. Hansen and N. W. Hurwitz, "The problem of non-response in sample surveys", *J. Amer. Statist. Assoc*., 41, 517-529, (1946).

[2] K. M. Chaudhary, R. Singh, R. K. Shukla and M. Kumar, "A family of estimators for estimating population mean in Stratified sampling under nonresponse", *Pakistan Journal of Statistics and Operations Research,* 5 (1), 47-54, (2009).

[3] K. M. Chaudhary, V. K. Singh and R. K. Shukla, "Combined-type family of estimators of population mean in stratified random sampling under nonresponse", *Journal of Reliability and Statistical studies (JRSS)*, 5 (2), 133-142, (2012).

[4] S. Kumar, "Efficient use of auxiliary information in estimating the population ratio, product and mean in the presence of non-response", *Journal for Advanced Computing*. 4, 68-87, (2015).

[5] K. M. Chaudhary and A. Kumar, "Use of double sample scheme in estimating the mean of stratified population under non-response", *STATISTICA*, anno, LXXV, n. 4 (2015).

[6] A. Sanaullah, I. Elisan and M. Noor-UI-Amin, "Estimation of mean for a finite population using subsampling of nonrespondents", Journal of Statistics and Management Systems, 22 (6), 1015-1035, (2019).

[7] G. N. Singh and M. Usman, "Improved regression cum ratio estimators using information on two auxiliary variables dealing with subsampling technique of nonresponse", Journal of Statistical Theory and Practice, 14 (1), 1-28, (2020).

[8] N. Garib and U. Mahamood, "Enhanced estimation of population distribution function in the presence of non-response", http://dio.org/10.1016/,asej. 2021.02.002.

[9] A. E. Anieting and E. I. Enang, "Two-Phase stratified sampling Estimator for population mean in the presence of nonresponse using one auxiliary variable", *Math. J. Interdiscip. Sci*. 8 (2), 49-56, (2020).

[10] A. E. Anieting, E. I. Enang and C. E. Onwukwe, "Efficient estimator for Population mean in stratified double sampling in the presence of nonresponse using one auxiliary variable", *African Journal of Mathematics and Statistical Studies* 4 (2), 40-50, (2021). DOI: 10525589/AJMSS-YF4VIIQV.

[11] S. Guha and H. Chandra, "Improved estimation of finite population mean in two-phase sampling with subsampling of the nonrespondents", *Mathematical Population Studies* 28 (1), 24-44, (2020).

[12] M. K Chaudhary, A. Kumar, and G. K. Vishwarkarma, "Some improved estimators of population mean using two-phase sampling scheme in the presence of nonresponse", Pakistan Journal of Statistics and Operations Research, 17 (4), 911-919, (2021). DOI: 10.18187/Pjsor.v17i4.2505.

[13] M. J. Iseh and K. J. Bassey, "Calibration estimator for population mean in small sample size with nonresponse", *European Journal of Statistics and Probability,* 9 (1), 32-42, (2021a).

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

[15] J. C. Deville and C. E. Sarndal, "Calibration estimators in survey sampling', Journal of the American Statistical Association, 87, 376-382. (1992).

[16] M. J. Iseh, and E. I. Enang, "A Calibration Synthetic Estimator of population Mean in Small Area under Stratified Sampling Design", *Transition in Statistics new series,* 22 (3), 15-30, (2021). DOI: *10.21307/stattrans-2021-025.*

[17] N. Koyuncu, and C. Kadilar, "Family of estimators of population mean using two auxiliary variables in stratified random sampling", *Communication in Statistics Theory and Methods,* 38 (14), 2398-2417.