International Journal of Statistical Distributions and Applications
2021; 7(1): 13-24
http://www.sciencepublishinggroup.com/j/ijsda
doi: 10.11648/j.ijsd.20210701.13
ISSN: 2472-3487 (Print); ISSN: 2472-3509 (Online)

Optimal Allocation in Small Area Mean Estimation Using Stratified Sampling in the Presence of Non-Response

Ongoma Jackson[1], Alilah David Anekeya[1], Okuto Erick[2]

[1]Department of Mathematics Masinde Muliro University of Science and Technology, Kakamega-Nairobi, Kenya

[2]Departments of Mathematics, Jaramogi Oginga Odinga University of Science and Technology, Siaya-Nairobi, Kenya

Email address:
ongomajackson@gmail.com (O. Jackson), aliladavid2010@gmail.com (A. D. Anekeya), erickokuto@gmail (O. Erick)

To cite this article:
Ongoma Jackson, Alilah David Anekeya, Okuto Erick. Optimal Allocation in Small Area Mean Estimation Using Stratified Sampling in the Presence of Non-Response. *International Journal of Statistical Distributions and Applications*. Vol. 7, No. 1, 2021, pp. 13-24.
doi: 10.11648/j.ijsd.20210701.13

Received: November 3, 2020; **Accepted**: December 1, 2020; **Published**: March 12, 2021

Abstract: Sample survey provides reliable current statistics for large areas or sub-population (domains) with large sample sizes. There is a growing demand for reliable small area statistics, however, the sample sizes are too small to provide direct (or area specific) estimators with acceptable and reliable accuracy. This study gives theoretical description of the estimation of small area mean by use of stratified sampling with a linear cost function in the presence of non-response. The estimation of small area mean is proposed using auxiliary information in which the study and auxiliary variable suffers from non-response during sampling. Optimal sample sizes have been obtained by minimizing the cost of survey for specific precision within a given cost using lagrangian function multiplier lambda and Partial Differential Equations (PDEs). Results demonstrate that as the values of the respondent sample increases sample units that supply information to study and auxiliary variable tends to small area population size, the non-response sample unit tends to sample units that supply the information as the sampling rate tends to one. From theoretic analysis it is practical that the Mean Square Error will decrease as the sub-sampling fraction and auxiliary characters increase. As the sub-sampling fraction increases and the value of beta increases then the value of large sample size is minimized with a reduction of Lagrangian multiplier value which minimizes the cost function.

Keywords: Stratified Sampling for Ratio Estimation, Small Area Mean, Auxiliary Variable, Linear Cost Function and Non-response

1. Introduction

1.1. Small Area

Small Area refers to a population for which reliable statistics of interest could not be computed using standard methods because of small or even zero sample sizes in the area. Some of the perceived small areas include geographical regions such as county, sub-county and wards, and demographic regions such as age, sex and race. In sampling the units are divided into two strata for homogeneity, the first strata represent respondents while the second strata represent non-respondents.

1.2. Small Area Estimation

According to Rahman [13] small area estimation has received much attention in recent decades due to increasing demand for reliable small area estimates for both public and private sectors. Sample data on small areas is inadequate to provide statistical estimates with high precision. This therefore makes it necessary to borrow strength from data on related auxiliary variables using appropriate models.

Small area estimation is therefore any statistical technique that involves the estimation of parameters for small sub-populations. Methods used in small area estimation are categorized as design based and model based. According to Rahman [13] design-based method reference was made for particular sampling design used whereas model-based method involves statistical method based on Bayesian approaches.

Among the models used in small area estimation and prediction is Linear Mixed Model that has found wide range of applications particularly for its ability to predict linear combination of fixed and random effects. Henderson [10] proposed the Best Linear Unbiased Prediction (BLUP) method for mixed models of the form

$$Y = X\beta + Zu + e$$

with $E(Y) = X\beta$

Where;

Y is an n x 1 vector of records to use in the predictions

X refers to known matrix

β is a fixed vector

u is a random vector with mean zero

Z being a known matrix and

e a random vector of errors with mean zero.

The Best Linear Unbiased Estimator (BLUP) method was widely used especially in fitting models for the genetic trends in animal population based on different traits measured both on continuous and categorical scale. Henderson [10] assumed that the variances associated with random effect in mixed model were known but in practice that was not the case. Such variance components are unknown and have to be estimated from the sampled data. Several researchers have proposed methods of estimating variance, among them was Harville [9] who reviewed methods suggested by Henderson of Maximum Likelihood and residual maximum likelihood. In his proposition he assumed normality which was not the case in all estimations. Therefore, in this study a design-based model is developed to solve non-linearity of the cost estimation.

Fay and Herriot model have received much attention in the previous years. Abhishek [1] applied it in estimating small area indicators. The model used was of the form;

$$Y_i = x_i^T \beta + v_i + e_i \quad i = 1, \ldots, m$$

Where;

x_i is a vector of known covariates

β is a vector of unknown regression coefficients

v_i being the specific random effect

e_i represents the sampling error

Generally, the Fay and Herriot model assumed linearity in estimation of parameters thus making it difficult to estimate costs when traveling costs was considered as a component of survey cost. Wanjoya et'al [15] carried out a study on small area estimation by incorporating a turning (index) parameter into the standard area-level (Fay-Herriot) model. In his model it was realized that the proposed model was a good alternative to the standard Fay-Herriot model though it did not consider a case of non-response. Different designs and models have been adopted in small area estimation. In this study, stratified sampling is considered in the presence of non-response during sampling and a linear cost function to cater for travel costs that are incurred during sampling. The main objective of this study is to develop linear cost model considering stratified sampling design in the presence of non-response and compute reliable estimates for a given small area.

Arnold et'al [4] estimated small area via generalized linear model of the form;

$$Y_c = x_1\beta_{1c} + x_2\beta_{2c} + \varepsilon_c$$

Which was used to solve census data and

$$Y_s = x_1\beta_{1s} + x_2\beta_{2s} + \varepsilon_s$$

Which was used to solve survey data where;

β is the (p x 1) vector of parameter

x_i being the (n x p) design matrix and

ε as the error term

According to Lohr [11], model-based estimator uses prediction approach in which the depended variable Y is predicted. The model-based estimates are only model unbiased within the structure of that specific model. It was realized that the model provides precise parameter estimates and explicit model specification. Aditya et'al [2] developed a method of estimating domain total for unknown domain size in the presence of non-response with linear cost function using two stage sampling design. The assumption was that the response mechanism was deterministic. Expression of the variance of the estimator and a suitable cost function for obtaining optimum sample size was developed. Empirical results showed that the percentage reduction in the expected cost decreased with a decrease in unit travel.

1.3. Optimal Allocation

Saini [14] proposed a method of optimum allocation for stratified two stage sampling design for multivariate surveys. The total cost of the survey was expressed as

$$\mathrm{C} = c_0 + c_1 n' + \sum_{h=1}^{L} c_{1h} n_h + c_{2h} \sum_{i=1}^{nh} m_{hi} \qquad (1)$$

Where;

c_o is the overhead cost of the survey

c_1 defined as the cost used in sampling at the preliminary stage

c_{1h} being the cost incurred in the first stage sampling

c_{2h} being cost incurred in the second stage sampling

n_h is the first stage sample size

m_{hi} is the second stage sample size

n' is a preliminary sample used for double sampling

In his method the problem of determining optimum allocations was formulated as a non-linear programming problem (NLPP). The langragian multiplier technique was used to solve the formulated NLPPs.

Cherniyak [7] proposed optimum allocation in double sampling with stratification using non-linear cost function. The proposed non-linear cost function given was of the form;

$$C = c^{/}(n^{/})^{\alpha} + \sum_{k=1}^{L} c_k n_{k,} \alpha \succ 0 \tag{2}$$

Where;

$c^{/}$ is the cost of classification per unit,

c_k is the cost of measuring per unit in stratum k

Also proposed, was logarithmic cost function of the form;

$$C = c^{/} \log n^{/} + n^{/} \sum_{k=1}^{L} c_k v_k W_k$$

Alilah et'al [3] proposed a cost function for domain mean estimation of the form

$$C_d = c_d^{/}(n_d^{/})^{\theta} + c_{d_0} n_d + c_{d_1} n_{d_1} + c_{d_2} r_{d_2} \tag{3}$$

Where;

c_d' represents the cost of measuring a unit in the first sample of size n_d'

c_{d_0} represents the cost of measuring a unit of the first attempt on y_d with second phase sample size n_d

c_{d_1} represents the unit cost for processing the responded data on the y_d at the first attempt n_{d_1}

c_{d_2} represents the unit cost associated with the subsample r_{d_2} of n_{d_2}

In his findings, it was noted that the Mean Squared Error (MSE) increased as the second sample size decreased for all values computed using linear and non-linear cost function. Also, MSE of the estimator computed using linear and non-linear cost function increased with an increase in the inverse sampling and non-response rates. Therefore, it was noted that an increase in use of auxiliary information reduced non-response error thus increasing the MSE.

1.4. Estimation of Population Parameters in the Presence of Non-Response

Hansen and Hurwitz [8] suggested a technique for handling the non-response in mail surveys. Mail survey is advantageous over the other survey since it is inexpensive. Okafor [12] extended Hansen and Hurwitz problem to the estimation of the population total in element sampling on two successive occasions. Later, Chaundhary and Kumar [6] used the Hansen and Hurwitz techniques to estimate the population product for sampling on two occasions when there was non-response on both occasions. Cochran [5] extended Hansen and Hurwitz technique for the case when the information on the characteristic under study was also available on auxiliary characteristics. Chaundhary and Kumar [6] proposed a method of estimating the mean of finite population using double sampling scheme under non-response. The proposed model was based on the fact that both the study and auxiliary variables suffered from non-response with the information of auxiliary variable X not available. The estimate of $\bar{x}$ at the first phase was given by;

$$x' = \frac{n_1'\bar{x}_{n_1} + n_2'\bar{x}_{h_2}}{n'} \tag{4}$$

With the corresponding variance of

$$V(\overline{X'^*}) = (\frac{1}{n'} - \frac{1}{N})S_x^2 + \frac{L'-1}{n'} W_2 S_{x_2}^2 \tag{5}$$

Where;

n_1' is the non-responding unit

h_2 is the responding units from the n_2' non-responding unit

$\bar{x}_{n_1}, h_2$ are the means from the non-responding units n_2

$\bar{x}_{h_2}$ are the means from the non-responding unit h_2

S_x^2 and $S_{x_2}^2$ are mean squares error of entire group and non-response group respectively with phase of

L' as the inverse sampling rate at first phase of sampling.

From the previous studies, researchers have considered linear cost function when estimating small area. In solving for non-response most of them considered double sampling.

2. Model Formulation

2.1. Proposed Small Area Concept in the Presence of Non-Response

Let U be a finite population with known population size N. The population is divided into small area groups defined by $U_1, U_2, \ldots, U_s$ with group sizes $N_1, N_2, \ldots, N_s$ respectively. Define the population total as $U = \sum_{s=1}^{S} U_s$ and overall population size as $N = \sum_{s=1}^{S} N_s$ respectively. The small area study and auxiliary variable are defined as Y_s and X_s with respective means as $\bar{Y}_s$ and $\bar{X}_s$ respectively. Let y_{s_i} and x_{s_i} be the i^{th} unit observation of the small area population for the study and auxiliary variable respectively with $i = 1, 2, \ldots, N_s$. Stratified sampling is used to estimate the small area auxiliary population mean $\bar{X}_s$ of the variable X_s from a large sample size n_s. Attributes from the auxiliary variable observation x_{s_i} are obtained and a sample small area mean $\bar{x}_{s_1}$ are obtained and a sample small area mean $\bar{x}_s$ computed. Define n_{s_1} as sample units that supply the information on y_{s_1} and x_{s_1} respondents while n_{s_2} be the non-respondents for both the study and auxiliary variables respectively. The sum total small area sample size is given by

$$n_s = n_{s_1} + n_{s_2}$$

By considering the n_{s_2} as the non-respondent subgroup using SRSWOR of m_{s_2} units are drawn with an inverse sampling rate k_{s_2} defined by

$$m_{s_2} = \frac{n_{s_2}}{k_{s_2}} \text{ With } k_{s_2} \succ 1 \tag{6}$$

Such that $n_{s_2} = k_{s_2} m_{s_2}$

All the m_{s_2} units respond after sub-sampling n_{s_2} non-responding units.

In developing the non-response theory let N_{s_1} be the stratum containing small area population units that respond in the first attempt and N_{s_2} be the stratum with units that do not respond such that; $N_{s_2} = N_s - N_{s_1}$. Both the small area stratum units N_{s_1} and N_{s_2} are not known in advance. Further let W_{s_1} and W_{s_2} be the small area stratum weights defined by $W_{s_1} = \frac{N_{s_1}}{N_s}$ and $W_{s_2} = \frac{N_{s_2}}{N_s}$ With the corresponding estimates defined as $\hat{W}_{s_1} = \frac{n_{s_1}}{n_s}$ and $\hat{W}_{s_2} = \frac{n_{s_2}}{n_s}$ respectively.

Also define the total small area sample size as $n_{s_1} + m_{s_2}$ then the small area estimate of population mean for the study variable will be defined by;

$$\bar{y}_s = \frac{n_{s_1}}{n_s}\bar{y}_{s_1} + \frac{n_{s_2}}{n_s}\bar{y}_{m_{s_2}} \tag{7}$$

While for auxiliary variable;

$$\bar{x}_s = \frac{n_{s_1}}{n_s}\bar{x}_{s_1} + \frac{n_{s_2}}{n_s}\bar{x}_{m_{s_2}} \tag{8}$$

This can then be written as

$$\bar{y}_s = w_{s_1}\bar{y}_{s_1} + w_{s_2}\bar{y}_{m_{s_2}} \tag{9}$$

and

$$\bar{x}_s = w_{s_1}\bar{x}_{s_1} + w_{s_2}\bar{x}_{m_{s_2}} \tag{10}$$

Respectively where $\bar{y}_s$ and $\bar{x}_s$ are the sample small area means for the observation y_{s_i} and x_{s_i} respectively. The following sample characteristics are defined when estimating small area mean,

i) $\bar{y}_{s_1} = \frac{1}{n_{s_1}}\sum_{i=1}^{n_{s_1}} y_{s_i}$ small area mean of the study character from the response group based on n_{s_1} units

ii) $\bar{y}_{m_{s_2}} = \frac{1}{m_{s_2}}\sum_{i=1}^{m_{s_2}} y_{s_i}$ small area mean of the study character for the non-responding group of m_{s_2} respondent units

iii) $\bar{x}_{s_1} = \frac{1}{n_{s_1}}\sum_{i=1}^{n_{s_1}} x_{s_i}$ small area mean of the auxiliary character from the response group based on n_{s_1} units

iv) $\bar{x}_{m_{s_2}} = \frac{1}{m_{s_2}}\sum_{i=1}^{m_{s_2}} x_{s_i}$ small area mean of the auxiliary character for the non-responding group of m_{s_2} respondent units

In estimating the overall small area population mean in the presence of non-response, stratified sampling ratio estimation of the small area mean is used. Define

$$\hat{\bar{Y}}_s = \frac{\bar{y}_s}{\bar{x}_s} * \bar{X} = \hat{R}_s\bar{X}_s = r_s\bar{X}_s$$

With assumption that;

$$E[\bar{y}_s] = \bar{Y}_s \text{ and } E[\bar{x}_s] = \bar{X}_s \tag{11}$$

2.2. Bias of the Ratio Estimator

Define

$$\varepsilon_{s_0} = \frac{\bar{y}_s - \bar{Y}_s}{\bar{Y}_s} \tag{12}$$

$$\varepsilon_{s_1} = \frac{\bar{x}_s - \bar{X}_s}{\bar{X}_s} \tag{13}$$

From (12)

$$\varepsilon_{s_0}\bar{Y}_s + \bar{Y}_s = \bar{y}_s$$

$\bar{Y}_s\left(\varepsilon_{s_0} + 1\right) = \bar{y}_s$ and from (13)

$$\varepsilon_{s_1}\bar{X}_s + \bar{X}_s = \bar{x}_s$$

$$\bar{X}_s\left(\varepsilon_{s_1} + 1\right) = \bar{x}_s$$

The assumption is that

$$E\left(\varepsilon_{s_0}\right) = E\left(\varepsilon_{s_1}\right) = 0$$

Define

$$E\left[\varepsilon_{s_0}^2\right] = E\left[\frac{\bar{y}_s - \bar{Y}_s}{\bar{Y}_s}\right]^2$$

$$= \frac{1}{\bar{Y}_s^2} E\left[\bar{y}_s - \bar{Y}_s\right]^2$$

$$= \frac{1}{\bar{Y}_s^2} Var\left(\bar{y}_s\right) \text{ but}$$

$$\bar{y}_s = \frac{n_{s_1}\bar{y}_{s_1} + n_{s_2}\bar{y}_{m_{s_2}}}{n_s}$$

$$V\left(\bar{y}_s\right) = V_1 E_2\left[\bar{y}_s \middle| n_s\right] + E_1 V_2\left[\bar{y}_s \middle| m_s\right]$$

Considering the first term in equation above

$$V_1 E_2\left[\frac{n_{s_1}\bar{y}_{s_1} + n_{s_2}\bar{y}_{m_{s_2}}}{n_s}\right]$$

$$V_1\left[\frac{n_{s_1}\bar{y}_{s_1} + n_{s_2}\bar{y}_{s_2}}{n_s}\right]$$

$V_1\left[\bar{y}_s\right] = \frac{N_s - n_s}{n_s}\frac{S_{y_s}^2}{N_s}$ this is equal to

$$= \left(\frac{1}{n_s} - \frac{1}{N_s}\right) S_{y_s}^2 \qquad (14)$$

Where,

$S_{y_s}^2$ = Variance of the whole small area population mean of the study variable Y_s

$S_{y_{s_2}}^2$ =Variance of the small area population mean for the stratum of non-respondents for the study variable Y_d

Consider also

$$E_1 V_2\left(\bar{y} \middle| m_{s_2}\right)$$

$$= E_1 V_2\left(\frac{n_{s_1}\bar{y}_{s_1} + n_{s_2}\bar{y}_{m_{s_2}}}{n_s}\right)$$

$$= E_1\left(\frac{n_{s_1}^2}{n_s^2} V_2\left(\bar{y}_{s_1}\right) + \frac{n_{s_2}^2}{n_s^2} V\left(\bar{y}_{m_{s_2}}\right)\right)$$

$$= E_1\left\{\frac{n_{s_1}^2}{n_s^2}\left(\frac{n_s - n_{s_1}}{n_s}\right)\frac{s_{y_{s_1}}^2}{n_{s_1}} + \frac{n_{s_2}^2}{n_s^2}\left(\frac{n_{s_2} - m_{s_2}}{n_{s_2}}\right)\frac{s_{y_{m_{s2}}}^2}{m_{s_2}}\right\}$$

$$= E_1\left\{\frac{n_{s_1}}{n_s^2}\left(\frac{n_s - n_{s_1}}{n_s}\right) s_{y_{s_1}}^2 + \frac{n_{s_2}^2}{n_s^2}\left(\frac{n_{s_2} - m_{s_2}}{n_{s_2}}\right)\frac{s_{y_{m_{s2}}}^2}{m_{s_2}}\right\}$$

$$= E_1\left\{\frac{n_{s_1}}{n_s^2}\left(1 - \frac{n_{s_1}}{n_s}\right) + \frac{n_{s_2}^2}{n_s^2}\left(\frac{n_{s_2} - m_{s_2}}{n_{s_2}}\right)\frac{s_{y_{m_{s2}}}^2}{m_{s_2}}\right\}$$

But

$n_{s_2} = k_{s_2} m_{s_2}$

$$= E_1\left\{\frac{n_{s_1}}{n_s^2}\left(1 - \frac{n_{s_1}}{n_s^2}\right) + \frac{n_{s_2}^2}{n_s^2}\left(\frac{k_{s_2} m_{s_2} - m_{s_2}}{n_{s_2}}\right)\frac{s_{y_{m_{s2}}}^2}{m_{s_2}}\right\}$$

$$= E_1\left\{\frac{n_{s_2}^2}{n_s^2}\left(\frac{k_{s_2} - 1}{n_{s_2}}\right) m_{s_2} \frac{s_{y_{m_{s2}}}^2}{m_{s_2}}\right\}$$

$$= E_1\left\{\frac{n_{s_2}}{n_s}\left(\frac{k_{s_2} - 1}{n_s}\right) s_{y_{m_{s2}}}^2\right\}$$

$$= E_1\left\{w_{s_2}\left(\frac{k_{s_2} - 1}{n_s}\right) s_{y_{m_{s2}}}^2\right\}$$

$$= W_{s_2}(k_{s_2} - 1)\frac{s_{y_{s2}}^2}{n_s}$$

$$E(\varepsilon_{s_0}^2) = \frac{1}{\bar{Y}_s^2}\left[\left(\frac{1}{n_s} - \frac{1}{N_s}\right) S_{y_s}^2 + W_{s_2}\left(\frac{k_{s_2} - 1}{n_s}\right) S_{y_{s2}}^2\right]$$

$$= \left(\frac{1}{n_s} - \frac{1}{N_s}\right)\frac{S_{y_s}^2}{\bar{Y}_s^2} + W_{s_2}\left(\frac{k_{s_2} - 1}{n_s}\right)\frac{S_{y_{s2}}^2}{\bar{Y}_s^2}$$

$$E\left[\varepsilon_{s_1}^2\right] = E\left[\frac{\bar{x}_s - \bar{X}_s}{\bar{X}_s}\right]^2$$

$$= \frac{1}{\bar{X}_s^2} E\left[\bar{x}_s - \bar{X}_s\right]^2$$

$$= \frac{1}{\bar{X}_s^2} Var\left(\bar{x}_s\right)$$

$$= \frac{1}{\bar{X}_s^2}\left[V_1\left[E_2\left(\bar{x}_s \middle| n_s\right)\right] + E_1\left[V_2\left(\bar{x}_s \middle| m_{s_2}\right)\right]\right]$$

$$= \frac{1}{\bar{X}_s^2}\left[\left(\frac{1}{n_s} - \frac{1}{N_s}\right) S_{x_s}^2 + W_{s_2}\left(\frac{k_{s_2} - 1}{n_s}\right) S_{x_{s2}}^2\right]$$

$$= \left(\frac{1}{n_s} - \frac{1}{N_s}\right)\frac{S_{x_s}^2}{\bar{X}_s^2} + W_{s_2}\left(\frac{k_{s_2} - 1}{n_s}\right)\frac{S_{x_{s2}}^2}{\bar{X}_s^2} \qquad (15)$$

Where;

$S_{x_s}^2$ =Variance of the whole small area population mean of the auxiliary variable X_s

k_{s_2} =The inverse sampling rate

$S_{x_{s_2}}^2$ =Variance of the domain population mean for the stratum of non-respondents for

Stratum of non-respondents for the auxiliary variable X_s

Next consider $E\left(\varepsilon_{s_0}\varepsilon_{s_1}\right)=E\left[\left(\frac{\bar{y}_s-\bar{Y}_s}{\bar{Y}_s}\right)\left(\frac{\bar{x}_s-\bar{X}_s}{\bar{X}_s}\right)\right]$

$$=\frac{1}{\bar{X}_s\bar{Y}_s}E\left[\left(\bar{y}_s-\bar{Y}_s\right)\left(\bar{x}_s-\bar{X}_s\right)\right]$$

$$=\frac{1}{\bar{X}_s\bar{Y}_s}Cov\left(\bar{x}_s\bar{y}_s\right) \tag{16}$$

Consider

$$\frac{1}{\bar{X}_s\bar{Y}_s}Cov\left[E\left(\bar{y}_s|n_s\right)E\left(\bar{x}_s|n_s\right)\right]+\frac{1}{\bar{X}_s\bar{Y}_s}E\left[Cov\left(\bar{y}_s\bar{x}_s|n_s\right)\right]+\frac{1}{\bar{X}_s\bar{Y}_s}E\left[Cov\left(\bar{y}_s\bar{x}_s|m_{s_2}\right)\right]$$

$$=\frac{1}{\bar{X}_s\bar{Y}_s}E\left[Cov\left(\bar{y}_s\bar{x}_s|n_s\right)\right]+\frac{1}{\bar{X}_s\bar{Y}_s}E\left[Cov\left(\bar{y}_s\bar{x}_s|m_{s_2}\right)\right]$$

$$=\frac{1}{\bar{X}_s\bar{Y}_s}E\left[\left(\frac{N_s-n_s}{N_s}\right)\frac{s_{x_sy_s}}{n_s}\right]+\frac{1}{\bar{X}_s\bar{Y}_s}E\left[\frac{n_{s_2}^2}{n_s^2}\left(\frac{n_{s_2}-m_{s_2}}{n_{s_2}}\right)\frac{s_{x_{m_{s2}}y_{m_{s2}}}}{m_{s_2}}\right]$$

$$=\frac{1}{\bar{X}_s\bar{Y}_s}E\left[\left(\frac{1}{n_s}-\frac{1}{N_s}\right)s_{x_sy_s}\right]+\frac{1}{\bar{X}_s\bar{Y}_s}E\left[\frac{n_{s_2}^2}{n_s^2}\left(\frac{k_{s_2}m_{s_2}-m_{s_2}}{n_{s_2}}\right)s_{x_{m_{s2}}y_{m_{s2}}m_{s2}}\right]$$

$$=\frac{1}{\bar{X}_s\bar{Y}_s}E\left[\left(\frac{1}{n_s}-\frac{1}{N_s}\right)s_{x_sy_s}\right]+\frac{1}{\bar{X}_s\bar{Y}_s}E\left[w_2\left(\frac{k_{s_2}-1}{n_s}\right)s_{x_{m_{s2}}y_{m_{s2}}}\right]$$

$$=\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{S_{x_sy_s}}{\bar{X}_s\bar{Y}_s}+w_2\left(\frac{k_{s_2}-1}{n_s}\right)\frac{S_{x_{s2}y_{s2}}}{\bar{X}_s\bar{Y}_s} \tag{17}$$

3. Small Area Ratio Estimator of the Mean

$$\hat{\bar{Y}}_s=\frac{\bar{y}_s}{\bar{x}_s}*\bar{X}_s$$

$$\bar{y}_s=\bar{Y}_s\left(1-\varepsilon_{s_0}\right)$$

$$\bar{x}_s=\bar{X}_s\left(1+\varepsilon_{s_1}\right)$$

$$\hat{\bar{Y}}_s=\frac{\bar{Y}_s\left(1+\varepsilon_{s0}\right)}{\bar{X}_s\left(1+\varepsilon_{s_1}\right)}*\bar{X}_s$$

$$\hat{\bar{Y}}_s=\bar{Y}_s\left(1+\varepsilon_{s_0}\right)\left(1+\varepsilon_{s_1}\right)^{-1}$$

$$\hat{\bar{Y}}_s=\bar{Y}_s\left(1+\varepsilon_{s_0}\right)\left(1-\varepsilon_{s_1}+\varepsilon_{s_1}^2+,...\right)$$

$$\bar{Y}_s\left(1-\varepsilon_{s_1}+\varepsilon_{s_1}^2+\varepsilon_{s_0}-\varepsilon_{s_0}\varepsilon_{s_1}+\varepsilon_{s_0}\varepsilon_{s_1}^2,...\right)$$

$$\hat{\bar{Y}}_s=\bar{Y}\left(1+\varepsilon_{s_0}-\varepsilon_{s_1}+\varepsilon_{s_0}\varepsilon_{s_1}+\varepsilon_{s_1}^2\right) \tag{18}$$

3.1. Bias of Ratio Estimator $\hat{\bar{Y}}_s$

Proposition 1

The Bias of the ratio estimator ?$\hat{\bar{Y}}_s$?is given by

$$\text{Bias}\left(\bar{Y}_s\right)=\bar{Y}_s\left[\left(\frac{1}{n_s}-\frac{1}{N_s}\right)C_{x_s}^2+W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)C_{x_{s2}}^2-\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\rho_{x_s y_s}C_{x_s}C_{y_s}-W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\rho_{x_{s2}y_{s2}}C_{x_{s2}}C_{y_{s2}}\right]\ C_{x_s}=\frac{S_{x_s}}{\bar{X}_s},$$

$$C_{y_s}=\frac{S_{y_s}}{\bar{Y}_s},\ C_{x_{s2}}=\frac{S_{x_{s2}}}{\bar{X}_s}\ \text{and}\ C_{y_{s2}}=\frac{S_{y_{s2}}}{\bar{Y}_s}$$

$S_{y_s}^2$ = Variance of the whole small area population mean of the study variable Y_s

$S_{y_{s2}}^2$ =Variance of the small area population mean for the stratum of non-respondents for the study variable Y_s

$S_{x_s}^2$ =Variance of the whole small area population mean of the auxiliary variable X_s

k_{s_2} =The inverse sampling rate.

$S_{x_{s2}}^2$ =Variance of the small area population mean for the stratum of non-respondents for the auxiliary variable X_s

Proof

$$\text{Define } \bar{Y}_s=\bar{Y}_s\left(1+\varepsilon_{s_0}-\varepsilon_{s_1}-\varepsilon_{s_0}\varepsilon_{s_1}+\varepsilon_{s_1}^2\right)$$

$$E\left(\hat{\bar{Y}}_s\right)=\bar{Y}_s\left\{1+E\left(\varepsilon_{s_0}\right)-E\left(\varepsilon_{s_1}\right)-E\left(\varepsilon_{s_0}\varepsilon_{s_1}\right)+E\left(\varepsilon_{s_1}^2\right)\right\}$$

Since

$$E\left(\varepsilon_{s_1}\right)=E\left(\varepsilon_{s_0}\right)=0$$

$$E\left(\hat{\bar{Y}}_s\right)=\bar{Y}_s\left[1-E\left(\varepsilon_{s_0}\varepsilon_{s_1}\right)+E\left(\varepsilon_{s_1}^2\right)\right]$$

$$=\bar{Y}_s\left[1-\left\{\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{S_{x_s y_s}}{\bar{X}_s\bar{Y}_s}+W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\frac{S_{x_{s2}y_{s2}}}{\bar{X}_s\bar{Y}_s}\right\}+\left\{\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{S_{x_s}^2}{\bar{X}_s^2}+W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\frac{S_{x_{s2}}^2}{\bar{X}_s^2}\right\}\right]$$

$$E\left[\hat{\bar{Y}}_s-\bar{Y}_s\right]=\bar{Y}_s\left[\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{S_{x_s}^2}{\bar{X}_s^2}+W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\frac{S_{x_{s2}}^2}{\bar{X}_s^2}-\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\rho_{x_s y_s}\frac{S_{x_s}}{\bar{X}_s}\frac{S_{y_s}}{\bar{Y}_s}-W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\frac{\rho_{x_{s2}y_{s2}}S_{x_{s2}y_{s2}}}{\bar{X}_s\bar{Y}_s}\right]$$

$$\text{Bias}\left(\bar{Y}_s\right)=\bar{Y}_s\left[\left(\frac{1}{n_s}-\frac{1}{N_s}\right)C_{x_s}^2+W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)C_{x_{s2}}^2-\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\rho_{x_s y_s}C_{x_s}C_{y_s}-W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\rho_{x_{s2}y_{s2}}C_{x_{s2}}C_{y_{s2}}\right]$$

3.2. Mean Square Error (MSE) of the Ratio Estimator $\hat{\bar{Y}}_s$

Proposition 2

The Mean Squared Error (MSE) of the ratio estimator $\hat{\bar{Y}}_s$ is given by

$$MSE\left(\hat{\bar{Y}}_s\right)=\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\phi_{s_1}^2+\left(\frac{k_{s_2}-1}{n_s}\right)W_{s_2}\phi_{s_2}^2$$

Proof

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

$$E\left[\hat{\bar{Y}}_s - \bar{Y}_s\right]^2 = E\left[\frac{\bar{y}_s}{\bar{x}_s} * \bar{X}_s - \bar{Y}_s\right]^2$$

$$= E\left[\frac{\bar{Y}_s\left(1+\varepsilon_{s_0}\right)}{\bar{X}_s\left(1+\varepsilon_{s_1}\right)} * \bar{X}_s - \bar{Y}_s\right]^2$$

$$= \bar{Y}_s^2 E\left[\left(\frac{1+\varepsilon_{s_0}}{1+\varepsilon_{s_1}}\right) - 1\right]^2$$

$$= \bar{Y}_s^2 E\left[\frac{\left(1+\varepsilon_{s_0}\right)-\left(1+\varepsilon_{s_1}\right)}{\left(1+\varepsilon_{s_1}\right)}\right]^2$$

$$= \bar{Y}_s^2 E\left[\frac{\varepsilon_{s_0} - \varepsilon_{s_1}}{1+\varepsilon_s}\right]^2$$

$$= \bar{Y}_s^2 E\left[\left(\varepsilon_{s_0} - \varepsilon_{s_1}\right)\left(1+\varepsilon_{s_1}\right)^{-1}\right]^2$$

$$= \bar{Y}_s^2 E\left[\left(\varepsilon_{s_0} - \varepsilon_{s_1}\right)\left(1-\varepsilon_{s_1} + \varepsilon_{s_1}^2 +,\ldots\right)\right]^2$$

$$= \bar{Y}_s^2 E\left[\varepsilon_{s_0} - \varepsilon_{s_0}\varepsilon_{s_1} + \varepsilon_{s_0}\varepsilon_{s_1}^2 - \varepsilon_{s_1} + \varepsilon_{s_1}^2 - \varepsilon_{s_1}^3\right]$$

$$= \bar{Y}_s^2 E\left[\varepsilon_{s_0} - \varepsilon_{s_1}\right]^2$$

$$= \bar{Y}_s^2 E\left[\varepsilon_{s_0}^2 + \varepsilon_{s_1}^2 - 2E\left(\varepsilon_{s_0}\varepsilon_{s_1}\right)\right]$$

$$= \bar{Y}_s^2\left\{\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{s_{y_s}^2}{\bar{Y}_s^2} + W_{s_2}\frac{s_{y_{s_2}}^2}{\bar{Y}_s^2}\left(\frac{k_{s_2}-1}{n_s}\right) + \left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{s_{x_s}^2}{\bar{X}_s^2} + W_{s_2}\frac{s_{x_s}^2}{\bar{X}_s^2}\left(\frac{k_{s_2}-1}{n_s}\right) - 2\left[\left(\frac{1}{n_s}-\frac{1}{N_s}\right)\frac{s_{x_s}s_{y_s}}{\bar{X}_s\bar{Y}_s}\rho_{x_sy_s} + \rho_{x_{s_2}y_{s_2}}W_{s_2}\frac{s_{x_{s_2}}s_{y_{s_2}}}{\bar{X}_s\bar{Y}_s}\left(\frac{k_{s_2}-1}{n_s}\right)\right]\right\}$$

$$= \left(\frac{1}{n_s}-\frac{1}{N_s}\right)S_{y_s}^2 + W_{s_2}S_{y_{s_2}}\left(\frac{k_{s_2}-1}{n_s}\right) + \left(\frac{1}{n_s}-\frac{1}{N_s}\right)S_{x_s}^2\frac{\bar{Y}_s^2}{\bar{X}_s^2} + W_{s_2}S_{x_{s_2}}^2\frac{\bar{Y}_s^2}{\bar{X}_s^2}\left(\frac{k_{s_2}-1}{n_s}\right) - 2\left\{\left(\frac{1}{n_s}-\frac{1}{N_s}\right)R_sS_{x_s}S_{y_s}\rho_{x_sy_s} + W_{s_2}R_s\rho_{x_{s_2}y_{s_2}}S_{x_{s_2}}S_{y_{s_2}}\left(\frac{k_{s_2}-1}{n_s}\right)\right\}$$

$$= \left\{\left(\frac{1}{n_s}-\frac{1}{N_s}\right)S_{y_s}^2 + \left(\frac{1}{n_s}-\frac{1}{N_s}\right)R_s^2S_{x_s}^2 - 2\left(\frac{1}{n_s}-\frac{1}{N_s}\right)R_s\rho_{x_sy_s}R_sS_{x_s}S_{y_s}\right\} + \left\{W_{s_2}S_{y_2}^2\left(\frac{k_{s_2}-1}{n_s}\right) + W_{s_2}R_s^2S_{x_{s_2}}^2\left(\frac{k_{s_2}-1}{n_s}\right) - 2W_{s_2}R_s\rho_{x_{s_2}y_{s_2}}S_{x_{s_2}}S_{y_{s_2}}\left(\frac{k_{s_2}-1}{n_s}\right)\right\}$$

$$= \left(\frac{1}{n_s}-\frac{1}{N_s}\right)\left\{S_{y_s}^2 + R_s^2S_{x_s}^2 - 2R_s\rho_{x_sy_s}S_{x_s}S_{y_s}\right\} + W_{s_2}\left(\frac{k_{s_2}-1}{n_s}\right)\left\{S_{y_s}^2 + R_s^2S_{x_{s_2}}^2 - 2R_s\rho_{x_{s_2}y_{s_2}}S_{x_{s_2}}S_{y_{s_2}}\right\}$$

$$\phi_{s_1}^2 = \left\{S_{y_s}^2 + R_s^2S_{x_s}^2 - 2R_s\rho_{x_sy_s}S_{x_s}S_{y_s}\right\}$$

$$MSE\left(\hat{\bar{Y}}_s\right) = \left(\frac{1}{n_s}-\frac{1}{N_s}\right)\phi_{s_1}^2 + \left(\frac{k_{s_2}-1}{n_s}\right)W_{s_2}\phi_{s_2}^2$$

3.3. Optimal Allocation

An optimum size of a sample is required so as to balance the precision and cost involved in the survey. The optimum allocation of a sample size is attained either by minimizing the precision against a given cost or minimizing cost against a given precision. In this study, a linear cost function has been considered.

Denote the cost function for the ratio estimation by

$$C_s = c_s\left(n_s\right)^{\beta} + c_{s_1}n_{s_1} + c_{s_2}m_{s_2} \tag{19}$$

c_s = Represents the cost of identifying sampling target population n_s

c_{s_1} = Represents the cost of measuring a unit in the response sample n_{s_1} .

c_{s_2} = Represents the cost of measuring unit ascertained with sub-sample m_{s_2} from n_{s_2}

Further more n_{s_1} and m_{s_2} are not known. Let

$$n_{s_1} = W_{s_1}n_s$$

$$m_{s_2} = W_{s_2}\frac{n_s}{k_{s_2}} = \frac{n_{s_2}}{k_{s_2}} \tag{20}$$

$$C_s = c_s\left(n_s\right)^{\beta} + c_{s_1}n_{s_1} + c_{s_2}m_{s_2}$$

$$C_s = c_s\left(n_s\right)^{\beta} + c_{s_1}W_{s_1}n_s + c_{s_2}W_{s_2}\frac{n_s}{k_{s_2}}$$

$$C_s = c_s\left(n_s\right)^{\beta} + n_s\left(c_{s_1}W_{s_1} + c_{s_2}\frac{W_{s_2}}{k_{s_2}}\right) \tag{21}$$

Proposition 3

The optimal values of n_s and k_{s_2} are given by

$$n_s = \left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s}\right)^{\frac{1}{\beta+1}}\left(\frac{1}{\lambda}\right)^{\frac{1}{\beta+1}}$$

and

$$k_{s_2} = \frac{\sqrt{c_{s_2}}}{\phi_{s_2}}\left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s}\right)^{\frac{1}{\beta+1}}\lambda^{-\frac{1}{2}}$$

To determine the optimum values of n_s and k_{s_2} that minimizes variance at a fixed cost

Proof

Define

$$\psi(W) = \left(\frac{1}{n_s} - \frac{1}{N_s}\right)\phi_{s_1}^2 + \left(\frac{k_{s_2}-1}{n_s}\right)W_s\phi_{s_2}^2 + \lambda\left\{c_s\left(n_s\right)^{\beta} + n_s\left(c_{s_1}W_{s_1} + \frac{c_{s_2}W_{s_2}}{k_{s_2}}\right) - C_s'\right\} \tag{22}$$

To obtain the normal equations, the expression of Equation (22) is differentiated partially with respect to k_{s_2} and n_s and the partial derivatives are equated to zero

$$\frac{\partial\psi(W)}{\partial k_{s_2}} = \frac{W_{s_2}\phi_{s_2}^2}{n_s} - \frac{\lambda n_s c_{s_2} W_{s_2}}{k_{s_2}^2} = 0$$

$$k_{s_2}^2 W_{s_2}\phi_{s_2}^2 = \lambda n_s^2 c_{s_2}$$

$$\frac{k_{s_2}^2}{n_s^2} = \frac{\lambda c_{s_2}}{\phi_{s_2}^2}$$

$$\frac{k_{s_2}}{n_s} = \frac{\sqrt{\lambda c_{s_2}}}{\phi_{s_2}}$$

$$\frac{n_s^2}{k_{s_2}^2} = \frac{\phi_{s_2}^2}{\lambda c_{s_2}}$$

$$\frac{n_s}{k_{s_2}} = \frac{\phi_{s_2}}{\sqrt{\lambda c_{s_2}}}$$

$$n_s = \frac{k_{s_2}\phi_{s_2}}{\sqrt{\lambda c_{s_2}}}$$

But

$$\psi(W) = \left(\frac{1}{n_s} - \frac{1}{N_s}\right)\phi_{s_1}^2 + \left(\frac{k_{s_2}}{n_s} - \frac{1}{n_s}\right)W_s\phi_{s_2}^2 + \lambda\left\{c_s\left(n_s\right)^{\beta} + n_s c_{s_1} W_{s_1} + c_{s_2} W_{s_2}\frac{n_s}{k_{s_2}} - C_s'\right\}$$

Satisfying for

$$\frac{k_{s_2}}{n_s}, \frac{n_s}{k_{s_2}}$$

We obtain

$$\psi(W) = \left(\frac{1}{n_s} - \frac{1}{N_s}\right)\phi_{s_1}^2 + \left(\frac{\sqrt{\lambda c_{s_2}}}{\phi_{s_2}} - \frac{1}{n_s}\right)W_{s_2}\phi_{s_2}^2 + \lambda\left\{c_s\left(n_s\right)^{\beta} + \frac{k_{s_2}\phi_{s_2}}{\sqrt{\lambda c_s}} c_{s_1} W_{s_1} + c_{s_2} W_{s_2}\left(\frac{\phi_{s_2}}{\sqrt{\lambda c_{s_2}}}\right) - C_s'\right\} \tag{23}$$

Next the partial derivative with respect to n_s obtained as;

$$\frac{\partial\psi(W)}{\partial n_s} = -\frac{1}{n_s^2}\phi_{s_1}^2 + \frac{1}{n_s^2}W_{s_2}\phi_{s_2}^2 + \lambda\beta c_s\left(n_s\right)^{\beta-1} = 0$$

$$W_{s_2}\phi_{s_2}^2 - \phi S_1^2 + \lambda\beta c_s\left(n_s\right)^{\beta+1} = 0$$

$$\lambda\beta c_s\left(n_s\right)^{\beta+1} = \phi_{s_1}^2 - W_{s_2}\phi_{s_2}^2$$

$$\left(n_s\right)^{\beta+1} = \frac{\phi_{s_1}^2 - W_{s_2}\phi_{s_2}^2}{\lambda\beta c_s}$$

$$n_s = \left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\lambda \beta c_s} \right)^{\frac{1}{\beta+1}}$$

$$n_s = \left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right)^{\frac{1}{\beta+1}} \left(\frac{1}{\lambda} \right)^{\frac{1}{\beta+1}} \tag{24}$$

$$k_{s_2} = \frac{n_s \sqrt{\lambda c_{s_2}}}{\phi_{s_2}}$$

Substituting n_s

$$k_{s_2} = \frac{\left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right)^{\frac{1}{\beta+1}} \left(\frac{1}{\lambda} \right)^{\frac{1}{\beta+1}} \lambda^{\frac{1}{2}} c_{s_2}^{\frac{1}{2}}}{\phi_{s_2}}$$

$$k_{s_2} = \left(\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2} \right)^{\frac{1}{\beta+1}} \left(\beta c_s \right)^{-\frac{1}{\beta+1}} \lambda^{-\frac{1}{\beta+1}+\frac{1}{2}} c_{s_2}^{\frac{1}{2}} \phi_{s_2}^{-1} \tag{25}$$

But the overall cost function is defined as

$$C_s' = c_s \left(n_s \right)^{\beta} + n_s \left(c_{s_1} W_{s_1} + \frac{c_{s_1} W_{s_2}}{k_{s_2}} \right)$$

Substituting in the values of n_s and k_{s2} we obtain

$$C_s' = c_s \left\{ \frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right\}^{\frac{\beta}{\beta+1}} \left(\frac{1}{\lambda} \right)^{\frac{\beta}{\beta+1}} + \left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right)^{\frac{1}{\beta+1}} \left(\frac{1}{\lambda} \right)^{\frac{1}{\beta+1}} \left\{ c_{s_1} W_{s_1} + c_{s_2} W_{s_2} \frac{\phi_{s_2}}{\sqrt{c_{s_2}}} \left(\frac{\beta c_s}{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}} \right)^{\frac{1}{\beta+1}} \bullet \lambda^{\frac{1}{2}} \right\}$$

$$C_s' = c_s \left\{ \frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right\}^{\frac{\beta}{\beta+1}} \left(\frac{1}{\lambda} \right)^{\frac{\beta}{\beta+1}} + \left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right)^{\frac{1}{\beta+1}} c_{s_1} W_{s_1} \left(\frac{1}{\lambda} \right)^{\frac{1}{\beta+1}} + c_{s_2} W_{s_2} \phi_{s_2} \lambda^{\frac{1}{2}} \left(\frac{1}{\lambda} \right)^{\frac{1}{\beta+1}} \tag{26}$$

Let

$$c_s \left\{ \frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right\}^{\frac{\beta}{\beta+1}} = A$$

$$\left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right) c_{s_1} W_{s_1} = B$$

And

$$c_{s_2} W_{s_2} \phi_{s_2} = D$$

Substituting A, B and D in (26)

$$C_s' = A \lambda^{-\left(\frac{\beta}{\beta+1} \right)} + B \lambda^{-\left(\frac{1}{\beta+1} \right)} + D \lambda^{\frac{\beta-1}{2(\beta+1)}} - C_s \tag{27}$$

$$\beta = 1$$

$$A \lambda^{-\frac{1}{2}} + B \lambda^{-\frac{1}{2}} - C_s = 0$$

$$\lambda^{-\frac{1}{2}} = \frac{C_s}{A+B}$$

$$\frac{1}{\lambda^{\frac{1}{2}}} = \frac{C_s}{A+B}$$

Applying the reciprocals to all terms

$$\lambda^{\frac{1}{2}} = \frac{A+B}{C_s}$$

Squaring both sides

$$\lambda = \left(\frac{A+B}{C_s} \right)^2$$

When

$$\beta = 0$$

Substituting from equation (27) we obtain a linear equation of the form

$$B\lambda^{-1} + D\lambda^{-\frac{1}{2}} - C_s \tag{28}$$

With the values of B and D defined as

$$B = \left(\frac{\phi_{s_1}^2 - \phi_{s_2}^2 W_{s_2}}{\beta c_s} \right) c_{s_1} W_{s_1},\ D = c_{s_2} W_{s_2} \phi_{s_2}$$

$$\text{let } \lambda^{-\frac{1}{2}} = K$$

$$\lambda^{-1} = \lambda^{-\frac{1}{2}} \lambda^{-\frac{1}{2}} = K^2$$

$$BK^2 + DK - C_s = 0$$

$$K = \frac{-D \pm \sqrt{D^2 + 4BC_s}}{2B}$$

but

$$K = \lambda^{-\frac{1}{2}}$$

$$\lambda^{-\frac{1}{2}} = \frac{-D \pm \sqrt{D^2 + 4BC_s}}{2B}$$

$$\lambda^{\frac{1}{2}} = \frac{2B}{-D \pm \sqrt{D^2 + 4BC_s}}$$

Solving for λ in equation (28) the solution becomes,

$$\lambda = \frac{4B^2}{\left(-D \pm \sqrt{D^2 + 4BC_s} \right)^2}$$

4. Conclusion

From the results it is noted that as values the respondent sample n_{s_1} tends to small area population size N_s the non-response m_{s_2} tends to n_{s_1} and the sampling rate k_{s_2} tends to 1. From theoretical analysis it is observed that the Mean Square Error of the proposed estimator will decrease as the sub-sampling fraction together with the number of auxiliary characters is increased. As the sub-sampling fraction also increases and the value of β increases then the value of n_s is minimized with the reduction in the value of Lagrangian multiplier (λ) which minimizes the cost function.

References

[1] Abhishek. N (2013). An overview of Fay Herriot model with our package in small area.

[2] Aditya K. (2014). Estimation of Domain Mean Using Two-Stage Sampling with Sub-Sampling Non-response. Journal of the Indian Society of Agricultural Statistics 68 (1) pp. 39-54.

[3] Alilah D. A and Ouma C. O (2018). Domain Mean Estimation Using Double Sampling with Non-Linear Cost functions in the presence of Non-response. Science Journal of the Applied Mathematics and Statistics Vol 6, No 1, pp. 28-42.

[4] Arnold G. S. Harslet and Noble N. (2002). Small Area Estimation via generalized linear model. Journal of official Statistics vol. 18 (1) pp 45-60.

[5] Cochran W. G (1977). Sampling techniques, New York John Wiley and sons.

[6] Chaundhary M. K, Kumar A. (2016). Estimation of Mean of Finite Population using Double sampling Scheme under Non-response, Journal of Mathematical Sciences 5 (2), pp 287-297.

[7] Cherniyak O. I (2001). Optimal allocation in stratified sampling and double sampling with non-linear cost function, Journal of Mathematical Sciences 103, 4 pp. 525-528.

[8] Hansen M. H and Hurwitz W. N (1946). The problem of non-response in sampling surveys, Journal of American Statistical Association.

[9] Harville D. A (1977). Maximum Likelihood approaches to variance component estimation and related problems. Journal of American Statistical ass. 72 (320-340).

[10] Henderson, C. R. (1975). Best linear unbiased estimation and prediction under selection model. Biometrics. 31 (423-447).

[11] Lohr, S. L. (2010) Sampling Design and Analysis. Boston, MA 02210, USA: Brooks/Col, Cengage Learning.

[12] Okafor F. C (2001) Treatment of non response in successive sampling, Statistica, 61 (2) pp. 195-204.

[13] Rahman (2008). A review of small area estimation problems and methodological developments. Discussion paper NATSEM-University of Canberra.

[14] Saini M. and Kumar A. (2015). Method of optimum allocation for multivariate stratified two stage sampling design using double sampling. Journal of probability and statistic forum, 8 pp 19-23.

[15] Wanjoya A. K, Torelli N and Datta G. (2012). Small Area Estimation JAGST Vol 14 (1).