Science Journal of Applied Mathematics and Statistics
2018; 6(1): 28-42
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20180601.14
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

Domain Mean Estimation Using Double Sampling with Non-Linear Cost Function in the Presence of Non Response

Alila David Anekeya[1, *], Ouma Christopher Onyango[2], Nyongesa Kennedy[1]

[1]Department of Mathematics Masinde Muliro University of Science and Technology, Kakamega, Kenya
[2]Department of Statistics and Actuarial Science Kenyatta University, Nairobi, Kenya

Email address:
aliladavid2010@gmail.com (A. D. Anekeya), chrisouma2004@yahoo.com (O. C. Onyango), knyongesa@mmust.ac.ke (N. Kennedy)
[*]Corresponding author

To cite this article:
Alila David Anekeya, Ouma Christopher Onyango, Nyongesa Kennedy. Domain Mean Estimation Using Double Sampling with Non-Linear Cost Function in the Presence of Non Response. *Science Journal of Applied Mathematics and Statistics*. Vol. 6, No. 1, 2018, pp. 28-42. doi: 10.11648/j.sjams.20180601.14

Received: December 17, 2017; **Accepted**: January 5, 2018; **Published**: February 15, 2018

Abstract: This paper describes theoretical estimation of domains mean using double sampling with a non-linear cost function in the presence of non-response. The estimation of domain mean is proposed using auxiliary information in which the study and auxiliary variable suffers from non-response in the second phase sampling. The expression of the biases and mean square errors of the proposed estimators are obtained. The optimal stratum sample sizes for given set of non-linear cost function are developed.

Keywords: Double Sampling for Ratio Estimation, Domain Mean, Auxiliary Variable, Non-Linear Cost Function and Non-Response

1. Introduction

1.1. Domains

Domain is a subgroup of the whole target population of the survey for which specific estimates are needed. In sampling, estimates are made in each of the class into which the population is subdivided; for instance, the focus may not only be the unemployment rate of the entire population but also the break-down by age, gender and education level. Units of domains may sometimes be identified prior to sampling. In such cases, the domains can be treated as separate stratum from a specific sample taken. Stratification ensures a satisfactory level of representativeness of the domains in the final sample. These domains are called planned domains.

1.2. Domain Estimation

Consider a finite population under study U of size N divided into D domains; $U_1, U_2, ..., U_D$ respectively. Domain membership of any population unit is unknown before sampling. Its assumed that the domains are quite large and for a typical d^{th} domain U_d several characteristics maybe defined as described by Gamrot [4]. This includes;

Domain total; $Y_{U_d} = \sum_{u_d} y_{d_k}$

Domain mean; $\bar{Y}_{U_d} = \frac{1}{N_d}\sum_{u_d} y_{d_k}$

Domain variance; $S^2_{U_d}(Y) = \frac{1}{N_d - 1}\sum_{k\in u_d}\left(Y_d - \bar{Y}_{u_d}\right)^2$

Domain Covariance between two characters X and Y is given by;

$$Cov_{U_d} = \frac{1}{N_d - 1}\sum_{k\in u_d}\left(x_{d_k} - \bar{X}_{u_d}\right)\left(y_{d_k} - \bar{Y}_{u_d}\right)$$

According to Meeden [7] domain can be estimated by use of a non-informative Bayesian approach where a polya posterior is used on finite population that has little or no prior information about the population. Although a prior distribution is not specified there is a posterior distribution which may be used to make inferences.

Udofia [12] proposed estimate of domains using double sampling for probabilities proportional to size (PPS) with known constituent domain. The assumptions proposed by Udofia [12] are;

(i) The size of auxiliary variable X is not known.

(ii) The distribution of the variable (Z) that defines the domain is the not known prior and therefore the population size (N_{h_j}) of the domain is also known.

(iii) The cost of measuring the variable X and Z in each stratum is much lower than that of measuring of the study variable Y.

Aditya *et al.* [1] developed a method of estimating domain total for unknown domain size in the presence of non-response with a linear cost function using two-stage sampling design. In this method the response mechanism is assumed to be deterministic.

1.3. Double Sampling in the Presence of Auxiliary Information

In many sampling procedures the prior knowledge about the population mean of the auxiliary variable is required. If there is no such information, it's easier and cheaper to take on the large initial sample from which the auxiliary variable is measured and from which the estimation of the population parameters like the total, mean or the frequency distribution of the auxiliary variable X is made. Srivastava [11] proposed a large class of ratio and product estimators in double sampling. It was found that the asymptotic minimum variance for any estimator of this class is equal to that which is generally believed to be linear regression estimators. According to Sahoo and Panda [10] if an experimenter knows the population mean of an additional auxiliary variable, say, Z whereas the population mean of an auxiliary variable X is unknown and can be estimated using double sampling scheme, it is possible to come up with a class of estimators for the finite population mean μ_Y.

1.4. Double Sampling for the Ratio Estimator in the Presence of Non-Response

Hansen and Hurwitz [5] proposed a way of dealing with non-response to address the bias problem. In this case, when dealing with non-response, a sub-sample is taken from the non-respondents to get an estimate of the sub-populations represented by the non-respondents. Cochran [2] employed Hansen and Hurwitz [5] technique and proposed ratio and regression estimation of the population mean of the study variables where the auxiliary variable information is obtained from all the sample units with some of the sample units failing to supply information on the study variable. According to Oh and Scheuren [8] and Kalton and Karsprzyk [6], non-response is often compensated by weighting adjustment and imputation respectively. In these methods it was argued that the procedure used in weighting adjustment and imputation aimed at eliminating the bias due to non-response. Okafor and Lee [9] employed the double sampling method to estimate the mean of the auxiliary variable and went ahead to estimate the mean of the study variable in a similar way as Cochran [2]. In this method double sampling for ratio and regression estimation was considered. The distribution of the auxiliary information was not known and hence the the first phase sample was used to estimate the population distribution of the auxiliary variable while the second phase was used to obtain the required information on the variable of the interest. The optimum sampling fraction for the estimators for a fixed cost was derived. Performances of the proposed estimators were computed and compared with those of Hansen and Hurwitz [5] estimators without considering the cost. It was noted that for the results for which cost component was not considered, regression estimator functions were consistent than the Hansen and Hurwitz [5] estimator. Chaudhary and Kumar [3] proposed a method of estimating mean of a finite population using double sampling scheme under non-response. The proposed model was based on the fact that both the study and auxiliary variable suffered from the non-response with the information of $\bar{X}$ not available. Hence the estimate of $\bar{X}$ at first phase is given by,

$$x' = \frac{n_1'\bar{x}_{n_1}' + n_2'\bar{x}_{h_2}'}{n'} \tag{1}$$

With the corresponding variance of,

$$V\left(\bar{x}'^{*}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_x^2 + \left(\frac{L'-1}{n'}\right)W_2 S_{x_2}^2 \tag{2}$$

Where $\bar{x}_{n_1}'$ and $\bar{x}_{h_2}'$ are means from the n_1' responding units and n_2' non-responding units respectively. S_x^2 and $S_{x_2}^2$ are mean square errors of the entire group and non-responding respectively with L' as the inverse sampling rate at first phase of the sampling.

From the previous studies, a number of researchers have considered a linear cost function when estimating domains. In dealing with non-response most of them have considered subsampling while holding to the idea that the response mechanism is deterministic. This study therefore focuses on the estimation of domain mean using double sampling for ratio estimation with non-linear cost function with a random response mechanism. In this study we therefore establish an efficient and cost effective method of estimating domains when the travel cost component is inclusive and it is not linear. The problem of minimum variance and cost is addressed while considering non-linear cost function and optimal sample size.

2. Estimation of Domain Mean and Variance in the Presence of Non-Response

2.1. Developing Domain Concept Theory with Non-Response

The problem of non-response is inherent in many surveys. It always persists even after call-backs. The estimates obtained from incomplete data will be biased especially when the respondents are different from the non-respondents. The non-response error is not so important if the characteristics of the non-responding units are similar to those of the responding units. However, such similarity of characteristics between two types of units (responding and non-responding) is not always attainable in practice. In double sampling when the problem of non-response is present, the strata are virtually divided into two disjoint and exhaustive groups of respondents and non-respondents. A sub-sample from non-responding group is then selected and a second more extensive attempt is made to the group so as to obtain the required information. Hansen and Hurwitz [5] proposed a technique of adjusting the non-response to address the problem of bias. The technique consists of selecting a sub-sample of the non-respondents through specialized efforts so as to obtain an estimate of non-responding units in the population. This sub-sampling procedure albeit costly, it's free from any assumption hence, one does not have to go for a hundred percent response which can be substantially more expensive.

In developing the concept of domain theory with non-response the following assumptions are made;

i. Both the domain study and auxiliary variables suffers from non-response.
ii. The responding and non-responding units are the same for the study and auxiliary characters.
iii. The information on the domain auxiliary variable X_d is not known and hence $\overline{X}_d$ is not available.
iv. The domain auxiliary variables do not suffer from non-response in the first phase sampling but suffers from non-response in the second phase of sampling.

2.2. Proposed Domain Estimators

Let U be a finite population with N known first stage units. The finite population is divided into D domains; $U_1, U_2, ..., U_D$ of sizes $N_1, N_2, ..., N_d, ..., N_D$ respectively. Further, let U_d be the domain constituents of any population size N_d which is assumed to be large and known. Let U and N be defined as,

$$U = \bigcup_{d=1}^{D} U_d \text{ and } N = \sum_{d=1}^{D} N_d \text{ respectively.}$$

Let Y_d and X_d be the domain study and auxiliary variables respectively. Further, let $\overline{Y}_d$ and $\overline{X}_d$ be their respective domain population means and auxiliary means with $y_{d_i}\left(i=1,2,3,...,N_d\right)$ and $x_{d_i}\left(i=1,2,...,N_d\right)$ observations on the i^{th} unit. In estimating the domain auxiliary population mean $\overline{X}_d$ double sampling design is used.

A large first phase sample of size $n^{/}$ is selected from N units of the population by simple random sampling without replacement (SRSWOR) design from which $n_d^{/}$ out $n^{/}$ first sample units falling in the d^{th} domain. The assumption here is that all the $n^{/}$ units supply information of the auxiliary variable X_d at first phase. A smaller second phase sample of size n is selected from $n^{/}$ by SRSWOR from which n_d out of n second phase sample units fall in the d^{th} domain.

For estimating the domain population mean $\overline{X}_d$ of the auxiliary variables X_d from a large first phase sample of size $n_d^{/}$, values of the observations $x_{d_i}^{/}\left(i=1,2,3,...,n_d^{/}\right)$ are obtained and a sample auxiliary domain mean $\overline{x}_d^{/}$ is computed. From the second sample of size n_d, let y_{d_i} and x_{d_i} be the domain study and auxiliary observations with $\left(i=1,2,3,...,n_d\right)$. Let n_{d_i} units supply the information on y_{d_i} and x_{d_i} respondents while n_{d_2} be the non-respondents for both the study and the auxiliary domain variables respectively such that,

$$n_d = n_{d_1} + n_{d_2}.$$

For the n_{d_2} non-respondent group at the second phase sampling, an SRSWOR of r_{d_2} units is selected with an inverse sampling rate of v_{d_2} such that,

$$r_{d_2} = \frac{n_{d_2}}{v_{d_2}}, \text{ With } v_{d_2} > 1$$

All the r_{d_2} units respond after making extra efforts of subsampling n_{d_2} non-responding units. In developing the framework of double sampling there are two strata that are non-overlapping and disjoint. Stratum one consist of those units that will respond in the first attempt of the second phase population made up of N_{d_1} units and stratum two consist of those units that would not respond in the first attempt of phase two with domain population units $N_{d_2} = N_d - N_{d_1}$. Both N_{d_1} and N_{d_2} units are not known in advance. The stratum weights of the responding and non- responding groups are defined by $W_{d_1} = \frac{N_{d_1}}{N_d}$ and $W_{d_2} = \frac{N_{d_2}}{N_d}$ respectively with their estimators defined by $\hat{W}_{d_1} = w_{d_1} = \frac{n_{d_1}}{n_d}$ and $\hat{W}_{d_2} = w_{d_2} = \frac{n_{d_2}}{n_d}$ respectively.

Following the Hansen and Hurwitz [5] techniques, the unbiased estimator for estimating the domain population mean using $\left(n_{d_1}+r_{d_2}\right)$ observations on y_{d_i} domain study character is given by;

$$\bar{y}_d=\frac{n_{d_1}}{n_d}\bar{y}_{d_1}+\frac{n_{d_2}}{n_d}\bar{y}_{r_{d_2}}$$

$$=w_{d_1}\bar{y}_{d_1}+w_{d_2}\bar{y}_{r_{d_2}} \tag{3}$$

Similarly the estimate for domain auxiliary variable is given by;

$$\bar{x}_d=\frac{n_{d_1}}{n_d}\bar{x}_{d_1}+\frac{n_{d_2}}{n_d}\bar{x}_{r_{d_2}}$$

$$=w_{d_1}\bar{x}_{d_1}+w_{d_2}\bar{x}_{r_{d_2}} \tag{4}$$

Where $\bar{y}_d$ and $\bar{x}_d$ are the sample domain means for the observation y_{d_i} and x_{d_i} respectively.

The following sample characteristics are defined when estimating domain mean,

i) $y_{d_1}=\frac{1}{n_{d_1}}\sum_{i=1}^{n_{d_1}}y_{d_i}$ -domain mean of the study character from the response group based on n_{d_1} units

ii) $\bar{y}_{r_{d_2}}=\frac{1}{r_{d_2}}\sum_{j=1}^{r_{d_2}}y_{d_j}$ -domain mean of the study character for the non-responding group of r_{d_2} respondent units

iii) $x_{d_1}=\frac{1}{n_{d_1}}\sum_{i=1}^{n_{d_1}}x_{d_i}$ -domain mean of the auxiliary character from the response group based on n_{d_1} units

iv) $\bar{x}_{r_{d_2}}=\frac{1}{r_{d_2}}\sum_{j=1}^{r_{d_2}}x_{d_j}$ -domain mean of the auxiliary character for the non-responding group of r_{d_2} respondent units

In estimating the overall domain population mean in the presence of non-response, double sampling ratio estimation of the domain mean is used. Define;

$$\hat{\bar{Y}}_{d_{R_1}}=\frac{\bar{y}_d}{\bar{x}_d}.\bar{x}_d^{/}=r_d\bar{x}_d^{/} \text{ and } \hat{\bar{Y}}_{d2}=\frac{\bar{y}_d}{\bar{x}_d^{/}}.\bar{x}_d=r_d\bar{x}_d$$

With the assumption that,

$$E\left[x_d^{/}\right]=E\left[\bar{x}_d\right]=\bar{X}_d,\ E\left(\bar{y}_d\right)=\bar{Y}_d \tag{5}$$

3. Bias and Mean Square Error of the Ratio Estimator

The expression for the Mean square error (MSE) of $\bar{Y}_{d_{R_1}}$ and $\bar{Y}_{d_{R_2}}$ are derived by the use of the Taylor's series approximation.

Let

$$\varepsilon_{d_0}=\frac{\bar{y}_d-\bar{Y}_d}{\bar{Y}_d}\Rightarrow\bar{y}_d=\bar{Y}_d\left(\varepsilon_{d_0}+1\right)$$

$$\varepsilon_{d1}=\frac{\bar{x}_d-\bar{X}_d}{\bar{X}_d}\Rightarrow\bar{x}_d=\bar{X}_d\left(\varepsilon_{d_1}+1\right)$$

$$\varepsilon_{d_2}=\frac{\bar{x}_d^{/}-\bar{X}_d}{\bar{X}_d}\Rightarrow\bar{x}_d=\bar{X}_d\left(\varepsilon_{d_0}+1\right) \tag{6}$$

With the assumption that $E\left(\varepsilon_{d_0}\right)=E\left(\varepsilon_{d_1}\right)=E\left(\varepsilon_{d_2}\right)=0$

Further define;

$$E\left(\varepsilon_{d_0}^2\right)=E\left[\left(\frac{\bar{y}_d-\bar{Y}_d}{\bar{Y}_d}\right)^2\right]$$

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

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

$$=\frac{1}{\bar{Y}_d^2}\left[V_1E_2E_3\left(\bar{y}_d/n_d^{/}\right)+E_1V_2E_3\left(\bar{y}_d/n_d\right)+E_1E_2V_3\left(\bar{y}_d/r_{d_2}\right)\right]$$

$$=\frac{1}{\bar{Y}_d^2}\left[\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)S_{y_d}^2+\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)S_{y_d}^2+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}S_{y_{d_2}}^2\right] \tag{7}$$

$$=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)C_{y_d}^2+\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)C_{y_d}^2+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}C_{y_{d_2}}^2$$

Where,

$S_{y_d}^2$ = Variance of the whole domain population mean of the study variable Y_d

$S_{y_{d_2}}^2$ =Variance of the domain population mean for the stratum of non-respondents for Stratum of non-respondents for the study variable Y_d

Consider also

$$E\left[\varepsilon_1^2\right]=E\left[\frac{\bar{x}_d-\bar{X}_d}{\bar{X}_d}\right]^2=\frac{1}{\bar{X}_d^2}E\left[\bar{x}_d-\bar{X}_d\right]^2$$

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

$$=\frac{1}{\bar{X}_d^2}\left[V_1\left[E_2\left(\bar{x}_d/n_d\right)\right]+E_1\left[V_2\left(\bar{x}_d/r_{d_2}\right)\right]\right]$$

$$=\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\frac{S_{x_d}^2}{\bar{X}_d^2}+W_{d_2}\left(\frac{v_{d_2}-1}{n_d}\right)\frac{S_{x_{d_2}}^2}{\bar{X}_d^2} \tag{8}$$

Where,

$S_{x_d}^2$ = Variance of the whole domain population mean of the auxiliary variable X_d

v_{d_2} = The inverse sampling rate

$S_{x_{d_2}}^2$ = Variance of the domain population mean for the stratum of non-respondents for Stratum of non-respondents for the auxiliary variable X_d

Next consider $E\left(\varepsilon_{d_2}^2\right)=E\left[\frac{\bar{x}_d^{/}-\bar{X}_d}{\bar{X}_d}\right]^2=\frac{1}{\bar{X}_d^2}E\left[\bar{x}_d-\bar{X}_d\right]^2$

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

$$=\frac{1}{\bar{X}_d^2}\left(\frac{N_d-n_d^{/}}{N_d}\right)\frac{S_{x_d}^2}{n_d^{/}}$$

$$=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)\frac{S_{x_d}^2}{\bar{X}_d^2} \tag{9}$$

Consider,

$$E\left[\varepsilon_{d_0}\varepsilon_{d_1}\right]=E\left[\left(\frac{\bar{y}_d-\bar{Y}_d}{\bar{Y}_d}\right)\left(\frac{\bar{x}_d-\bar{X}_d}{\bar{X}_d}\right)\right]$$

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

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

$$=\frac{1}{\bar{X}_d\bar{Y}_d}Cov\left[E\left(\bar{y}_d/n_d\right)E\left(\bar{x}_d/n_d\right)\right]+\frac{1}{\bar{X}_d\bar{Y}_d}E\left[Cov\left(\bar{y}_d\bar{x}_d\right)/n_d\right]+\frac{1}{\bar{X}_d\bar{Y}_d}E\left[Cov\left(\bar{y}_d\bar{x}_d\right)/r_{d_2}\right]$$

$$=\frac{1}{\bar{X}_d\bar{Y}_d}E\left[Cov\left(\bar{y}_d\bar{x}_d\right)/n_d\right]+\frac{1}{\bar{X}_d\bar{Y}_d}E\left[Cov\left(\bar{y}_d\bar{x}_d\right)/r_{d_2}\right]$$

$$=\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_d y_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d}+\left(\frac{v_{d_2}-1}{n_d}\right)\rho_{x_{d_2}y_{d_2}}W_{d_2}\frac{S_{x_{d_2}}}{\bar{X}_d}\frac{S_{y_{d_2}}}{\bar{Y}_d} \tag{10}$$

Next,

$$E\left[\varepsilon_{d_0}\varepsilon_{d_2}\right]=\left[\left(\frac{\bar{y}_d-\bar{Y}_d}{\bar{Y}_d}\right)\left(\frac{\bar{x}_d^{/}-\bar{X}_d}{\bar{X}_d}\right)\right]$$

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

$$=\frac{1}{\bar{X}_d\bar{Y}_d}Cov\left(\bar{y}_d\bar{x}_d^{/}\right)$$

$$=\frac{1}{\bar{X}_d\bar{Y}_d}Cov\left[E\left(\bar{y}_d/n_d^{/}\right),E\left(\bar{x}_d/n_d^{/}\right)\right]+\frac{1}{\bar{X}_d\bar{Y}_d}E\left[Cov\left(\bar{y}_d\bar{x}_d^{/}\right)/n_d^{/}\right]+\frac{1}{\bar{X}_d\bar{Y}_d}E\left[Cov\left(\bar{y}_{r_{d_2}}\bar{x}_{r_{d_2}}\right)/n_d^{/}\right]$$

$$=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)\rho_{x_d y_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d} \quad (11)$$

Consider,

$$E\left[\varepsilon_{d_1}\varepsilon_{d_2}\right]=E\left[\left(\frac{\bar{x}_d-\bar{X}_d}{\bar{X}_d}\right)\left(\frac{\bar{x}_d^{/}-\bar{X}_d}{\bar{X}_d}\right)\right]$$

$$=\frac{1}{\bar{X}_d^2}E_1E_2\left[\left(\bar{x}_d-\bar{X}_d\right)\left(\bar{x}_d^{/}-\bar{X}_d\right)/n_d^{/}\right]$$

$$=\frac{1}{\bar{X}_d^2}E_1\left(\bar{x}_d^{/}-\bar{X}_d\right)^2=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)\frac{S_{x_d}^2}{\bar{X}_d^2} \quad (12)$$

3.1. The Bias of the Ratio Estimator $\hat{\bar{Y}}_{d_{R_1}}$ and $\hat{\bar{Y}}_{d_{R_2}}$

The ratio estimator of $\hat{\bar{Y}}_{d_{R_1}}$ and $\hat{\bar{Y}}_{d_{R_2}}$ can be defined as; $\hat{\bar{Y}}_{d_{R_1}}=\frac{\bar{y}_d}{\bar{x}_d}.\bar{x}_d^{/}=r_d\bar{x}_d^{/}$ and $\hat{\bar{Y}}_{d2}=\frac{\bar{y}_d}{\bar{x}_d^{/}}.\bar{x}_d=r_d\bar{x}_d$ respectively

Define $\hat{\bar{Y}}_{d_{R_1}}$ as

$$\hat{\bar{Y}}_{d_{R_1}}=\frac{\bar{y}_d}{\bar{x}_d}.\bar{x}_d^{/}=r_d\bar{x}_d^{/}$$

$$=\left[\frac{\bar{Y}_d\left(1+\varepsilon_{d_0}\right)\bar{X}_d\left(1+\varepsilon_{d_2}\right)}{\bar{X}_d\left(1+\varepsilon_{d_1}\right)}\right]$$

$$=\bar{Y}_d\left[\frac{\left(1+\varepsilon_{d_0}\right)\left(1+\varepsilon_{d_2}\right)}{\left(1+\varepsilon_{d_1}\right)}\right]$$

$$=\bar{Y}_d\left[\left(1+\varepsilon_{d_0}+\varepsilon_{d_2}+\varepsilon_{d_0}\varepsilon_{d_2}-\varepsilon_{d_1}-\varepsilon_{d_0}\varepsilon_{d_1}-\varepsilon_{d_1}\varepsilon_{d_2}+\varepsilon_{d_1}^2\right)\right] \quad (13)$$

3.1.1. Bias of Ratio Estimator $\hat{\bar{Y}}_{d_{R_1}}$

Proposition 1

The bias of the ratio estimator $\hat{\bar{Y}}_{d_{R_1}}$ is given by,

$$\bar{Y}_d\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)C_{x_d}\left(C_{x_d}-\rho_{x_d y_d}c_{y_d}\right)+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}C_{x_{d_2}}\left(C_{x_{d_2}}-\rho_{x_{d_2}y_{d_2}}C_{x_{d_2}}\right)$$

Where,

$$C_{x_d}=\frac{S_{x_d}}{\bar{X}_d},\ C_{y_d}=\frac{S_{y_d}}{\bar{Y}_d},\ C_{x_{d_2}}=\frac{S_{x_{d_2}}}{\bar{X}_d}\ \textit{and}\ C_{y_{d_2}}=\frac{S_{y_{d_2}}}{\bar{Y}_d}$$

$S_{y_d}^2$ = *Variance of the whole domain population mean of the study variable* Y_d

$S_{y_{d_2}}^2$ =*Variance of the domain population mean for the stratum of non-respondents for Stratum of non-respondents for the study variable* Y_d

$S_{x_d}^2$ = *Variance of the whole domain population mean of the auxiliary variable* X_d

v_{d_2} = *The inverse sampling rate*

$S_{x_{d_2}}^2$ = *Variance of the domain population mean for the stratum of non-respondents for Stratum of non-respondents for the auxiliary variable* X_d

Proof

$$\hat{\bar{Y}}_{d_{R_1}}=\bar{Y}_d\left[\left(1+\varepsilon_{d_0}+\varepsilon_{d_2}+\varepsilon_{d_0}\varepsilon_{d_2}-\varepsilon_{d_1}-\varepsilon_{d_0}\varepsilon_{d_1}-\varepsilon_{d_1}\varepsilon_{d_2}+\varepsilon_{d_1}^2\right)\right]$$

$$E\left(\hat{\bar{Y}}_{d_{R_1}}\right) = \bar{Y}_d\left[\left(1+E\left(\varepsilon_{d_0}\right)+E\left(\varepsilon_{d_2}\right)+E\left(\varepsilon_{d_0}\varepsilon_{d_2}\right)-E\left(\varepsilon_{d_1}\right)-E\left(\varepsilon_{d_0}\varepsilon_{d_1}\right)-E\left(\varepsilon_{d_1}\varepsilon_{d_2}\right)+E\left(\varepsilon_{d_1}^2\right)\right)\right]$$

$$= \bar{Y}_d\left[\left(1+E\left(\varepsilon_{d_0}\varepsilon_{d_2}\right)-E\left(\varepsilon_{d_0}\varepsilon_{d_1}\right)-E\left(\varepsilon_{d_1}\varepsilon_{d_2}\right)+E\left(\varepsilon_{d_1}^2\right)\right)\right]$$

$$= \bar{Y}_d\left[1+\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)\rho_{x_dy_d}C_{x_d}C_{y_d}-\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_dy_d}C_{x_d}C_{y_d}-\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\rho_{x_{d_2}y_{d_2}}C_{x_{d_2}}C_{y_{d_2}}-\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)C_{x_d}^2\right.$$
$$\left.+\left(\frac{1}{n_d}-\frac{1}{N_d}\right)C_{x_d}^2+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}C_{x_{d_2}}^2\right]$$

$$= \bar{Y}_d\left[1+\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)\left(C_{x_d}^2-\rho_{x_dy_d}C_{x_d}C_{y_d}\right)+\left(\frac{v_{d_2}-1}{n_d}\right)\left(W_{d_2}C_{x_d}^2-\rho_{x_{d_2}y_{d_2}}C_{x_{d_2}}C_{y_{d_2}}\right)\right]$$

$$E\left(\hat{\bar{Y}}_{d_{R_1}}\right)-\bar{Y}_d = \bar{Y}_d\left[\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)\left(C_{x_d}^2-\rho_{x_dy_d}C_{x_d}C_{y_d}\right)+\left(\frac{v_{d_2}-1}{n_d}\right)\left(W_{d_2}C_{x_d}^2-\rho_{x_{d_2}y_{d_2}}C_{x_{d_2}}C_{y_{d_2}}\right)\right]$$

Hence Bias of $\hat{\bar{Y}}_{d_{R_1}} = \bar{Y}_d\left[\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)\left(C_{x_d}^2-\rho_{x_dy_d}C_{x_d}C_{y_d}\right)+\left(\frac{v_{d_2}-1}{n_d}\right)\left(W_{d_2}C_{x_d}^2-\rho_{x_{d_2}y_{d_2}}C_{x_{d_2}}C_{y_{d_2}}\right)\right]$

3.1.2. Bias of Ratio Estimator $\hat{\bar{Y}}_{d_{R_2}}$

Proposition 2

The bias of the ratio estimator $\hat{\bar{Y}}_{d_{R_2}}$ *is given by,* $\bar{Y}_d\left[\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)\rho_{x_dy_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d}+W_{d_2}\rho_{x_{d_2}y_{d_2}}\frac{S_{x_{d_2}}}{\bar{X}_d}\frac{S_{y_{d_2}}}{\bar{Y}_d}\right]$

Proof

Define $\hat{\bar{Y}}_{d_{R_2}}$ as $\hat{\bar{Y}}_{d_{R_2}} = \frac{\bar{y}_d}{\bar{x}_d^{/}}.\bar{x}_d = r_d\bar{x}_d$

$$= \left[\frac{\bar{Y}_d\left(1+\varepsilon_{d_0}\right)\bar{X}_d\left(1+\varepsilon_{d_1}\right)}{\bar{X}_d\left(1+\varepsilon_{d_2}\right)}\right]$$

$$= \bar{Y}_d\left[\frac{\left(1+\varepsilon_{d_0}\right)\left(1+\varepsilon_{d_1}\right)}{\left(1+\varepsilon_{d_2}\right)}\right]$$

$$= \bar{Y}_d\left[\left(1+\varepsilon_{d_0}+\varepsilon_{d_1}+\varepsilon_{d_0}\varepsilon_{d_1}-\varepsilon_{d_2}-\varepsilon_{d_0}\varepsilon_{d_2}-\varepsilon_{d_1}\varepsilon_{d_2}+\varepsilon_{d_2}^2\right)\right]$$

$$\hat{\bar{Y}}_{d_{R_2}} = \bar{Y}_d\left[\left(1+\varepsilon_{d_0}+\varepsilon_{d_1}+\varepsilon_{d_0}\varepsilon_{d_1}-\varepsilon_{d_2}-\varepsilon_{d_0}\varepsilon_{d_2}-\varepsilon_{d_1}\varepsilon_{d_2}+\varepsilon_{d_2}^2\right)\right]$$

$$E\left(\hat{\bar{Y}}_{d_{R_1}}\right) = \bar{Y}_d\left[\left(1+E\left(\varepsilon_{d_0}\right)+E\left(\varepsilon_{d_1}\right)+E\left(\varepsilon_{d_0}\varepsilon_{d_1}\right)-E\left(\varepsilon_{d_2}\right)-E\left(\varepsilon_{d_0}\varepsilon_{d_2}\right)-E\left(\varepsilon_{d_1}\varepsilon_{d_2}\right)+E\left(\varepsilon_{d_2}^2\right)\right)\right]$$

$$= \bar{Y}_d\left[\left(1+E\left(\varepsilon_{d_0}\varepsilon_{d_1}\right)-E\left(\varepsilon_{d_0}\varepsilon_{d_2}\right)-E\left(\varepsilon_{d_1}\varepsilon_{d_2}\right)+E\left(\varepsilon_{d_2}^2\right)\right)\right]$$

$$= \bar{Y}_d\left[1+\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_dy_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d}-\left(\frac{1}{n^{/}_d}-\frac{1}{N_d}\right)\rho_{x_dy_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d}+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\rho_{x_{d_2}y_{d_2}}\frac{S_{x_{d_2}}}{\bar{X}_d}\frac{S_{y_{d_2}}}{\bar{Y}_d}-\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\frac{S^2{}_{x_{d_2}}}{\bar{X}_d^2}\right.$$
$$\left.+\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\frac{S^2{}_{x_{d_2}}}{\bar{X}_d^2}\right]$$

Hence Bias of $\hat{\bar{Y}}_{d_{R_2}} = \bar{Y}_d \left[\left(\frac{1}{n_d} - \frac{1}{n'_d} \right) \rho_{x_d y_d} \frac{S_{x_d}}{\bar{X}_d} \frac{S_{y_d}}{\bar{Y}_d} + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} \rho_{x_{d_2} y_{d_2}} \frac{S_{x_{d_2}}}{\bar{X}_d} \frac{S_{y_{d_2}}}{\bar{Y}_d} \right]$

3.2. Mean Square Error (MSE) of the Ratio Estimator $\hat{\bar{Y}}_{d_{R_1}}$ and $\hat{\bar{Y}}_{d_{R_2}}$

The ratio estimator of $\hat{\bar{Y}}_{d_{R_1}}$ and $\hat{\bar{Y}}_{d_{R_2}}$ can be defined as;

$\hat{\bar{Y}}_{d_{R_1}} = \frac{\bar{y}_d}{\bar{x}_d} . \bar{x}'_d = r_d \bar{x}'_d$ *and* $\hat{\bar{Y}}_{d2} = \frac{\bar{y}_d}{\bar{x}'_d} . \bar{x}_d = r_d \bar{x}_d$ *respectively*

Proposition 3

The mean square error (MSE) of the estimator defined by $\hat{\bar{Y}}_{d_{R_1}} = \frac{\bar{y}_d}{\bar{x}_d} . \bar{x}'_d = r_d \bar{x}'_d$ *is given by;*

$$\left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S^2_{y_d} + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S^2_{d_{R_1}} + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S'_{d_{R_2}}$$

Where,

$$S^2_{d_{R_1}} = S^2_{y_d} + S^2_{x_d} R^2_d - 2\rho_{x_d y_d} R_d S_{x_d} S_{y_d}$$

$$S^2_{d_{R_2}} = S^2_{y_{d_2}} + R^2_d S^2_{x_{d_2}} - 2\rho_{x_{d_2} y_{d_2}} R_d S_{x_{d_2}} S_{y_{d_2}}$$

With the notations defined as in preposition 1 above and R_d =Population ratio of $\bar{Y}_d$ to $\bar{X}_d$

Proof

By definition,

$$MSE\left(\hat{\bar{Y}}_{d_{R_1}} \right) = E\left[\hat{\bar{Y}}_{d_{R_1}} - \bar{Y}_d \right]^2 = E\left[\frac{\bar{y}_d}{\bar{x}_d} . \bar{x}'_d - \bar{Y}_d \right]^2$$

Substituting the values of equations (5) we obtain

$$= E\left[\frac{\bar{Y}_d \left(1 + \varepsilon_{d_0}\right) \bar{X}_d \left(1 + \varepsilon_{d_2}\right)}{\bar{X}_d \left(1 + \varepsilon_{d_1}\right)} - \bar{Y}_d \right]^2$$

$$= \bar{Y}^2_d E\left[\frac{\left(1 + \varepsilon_{d_0}\right)\left(1 + \varepsilon_{d_2}\right)}{\left(1 + \varepsilon_{d_1}\right)} - 1 \right]^2$$

$$= \bar{Y}^2_d E\left[\left(\varepsilon_{d_2} + \varepsilon_{d_0} + \varepsilon_{d_0}\varepsilon_{d_2} - \varepsilon_{d_1} \right)\left(1 - \varepsilon_{d_1} + \varepsilon^2_{d_1} + \ldots \right) \right]^2$$

$$= \bar{Y}^2_d E\left(\varepsilon_{d_0} + \varepsilon_{d_2} - \varepsilon_{d_1} \right)^2$$

$$= \bar{Y}^2_d \left[E\left(\varepsilon^2_{d_0}\right) + E\left(\varepsilon^2_{d_1}\right) + E\left(\varepsilon^2_{d_2}\right) + 2E\left(\varepsilon_{d_0}\varepsilon_{d_2}\right) - 2E\left(\varepsilon_{d_0}\varepsilon_{d_1}\right) - 2E\left(\varepsilon_{d_1}\varepsilon_{d_2}\right) \right]$$

$$= \bar{Y}^2_d \left[\left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S^2_{y_d}}{\bar{Y}^2_d} + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) \frac{S^2_{y_d}}{\bar{Y}^2_d} + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} \frac{S^2_{y_{d_2}}}{\bar{Y}^2_d} \right.$$

$$+ \left(\frac{1}{n_d} - \frac{1}{N_d} \right) \frac{S^2_{x_d}}{\bar{X}^2_d} + \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S^2_{x_d}}{\bar{X}^2_d} + W_{d_2} \left(\frac{v_{d_2} - 1}{n_d} \right) \frac{S^2_{x_{d_2}}}{\bar{X}^2_d}$$

$$+2 \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \rho_{x_d y_d} \frac{S_{x_d}}{\bar{X}_d} \frac{S_{y_d}}{\bar{Y}_d} - 2\left(\frac{1}{n_d} - \frac{1}{N_d} \right) \rho_{x_d y_d} \frac{S_{x_d}}{\bar{X}_d} \frac{S_{y_d}}{\bar{Y}_d}$$

$$\left. -2\left(\frac{v_{d_2} - 1}{n_d} \right) \rho_{x_{d_2} y_{d_2}} W_{d_2} \frac{S_{x_{d_2}}}{\bar{X}_d} \frac{S_{y_{d_2}}}{\bar{Y}_d} - 2\left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S^2_{x_d}}{\bar{X}^2_d} \right]$$

$$= \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S^2_{y_d} + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S^2_{y_d} + \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S^2_{x_d} R^2_d$$

$$- \left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)S_{x_d}^2 R_d^2 + \left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\left[S_{y_d}^2 + R_d^2 S_{x_d}^2 - 2\rho_{x_{d_2}y_{d_2}} R_d S_{x_{d_2}} S_{y_{d_2}}\right]$$

$$+2\left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)\rho_{x_d y_d} R_d S_{x_d} S_{y_d} - 2\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_d y_d} R_d S_{x_d} S_{y_d}$$

$$=\left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)S_{y_d}^2 + \left(\frac{1}{n_d}-\frac{1}{n_d^/}\right)\left\{S_{y_d}^2 + S_{x_d}^2 R_d^2 - 2\rho_{x_d y_d} R_d S_{x_d} S_{y_d}\right\}$$

$$+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\left[S_{y_d}^2 + R_d^2 S_{x_d}^2 - 2\rho_{x_{d_2}y_{d_2}} R_d S_{x_{d_2}} S_{y_{d_2}}\right]$$

$$=\left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)S_{y_d}^2 + \left(\frac{1}{n_d}-\frac{1}{n_d^/}\right)S_{d_{R_1}}^2 + \left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2} S_{d_{R_2}}^2$$

Proposition 4

The mean square error (MSE) of the ratio estimator $\hat{\bar{Y}}_{d_{R_2}} = \frac{\bar{y}}{\bar{x}_d^/}.\bar{x}_d$ *is given by;* $\left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)S_{y_d}^2 + \left(\frac{1}{n_d}-\frac{1}{n_d^/}\right)S'^2_{d_{R_1}} + \left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2} S'^2_{d_{R_2}}$

Where,

$$S'^2_{d_{R_1}} = S_{y_d}^2 + S_{x_d}^2 R_d^2 + 2\rho_{x_d y_d} R_d S_{x_d} S_{y_d}$$

$$S'^2_{d_{R_2}} = S_{y_{d_2}}^2 + R_d^2 S_{x_{d_2}}^2 + 2\rho_{x_{d_2}y_{d_2}} R_d S_{x_{d_2}} S_{y_{d_2}}$$

With the notations as defined in proposition 1 above
Proof

$$\text{MSE of } \hat{\bar{Y}}_{d_{R_2}} = MSE\left(\hat{\bar{Y}}_{d_{R_2}}\right) = E\left[\hat{\bar{Y}}_{d_{R_2}} - \bar{Y}_d\right]^2$$

$$= E\left[\frac{\bar{y}_d}{\bar{x}_d}.\bar{x}_d^/ - \bar{Y}_d\right]^2$$

$$= E\left[\bar{Y}_d\left(\frac{\left(1+\varepsilon_{d0}\right)\left(1+\varepsilon_{d_1}\right)}{\left(1+\varepsilon_{d_2}\right)}-1\right)\right]^2$$

$$=\bar{Y}_d^2 E\left[\left(\varepsilon_{d_1}+\varepsilon_{d_0}+\varepsilon_{d_1}\varepsilon_{d_0}-\varepsilon_{d_2}\right)\left(1-\varepsilon_{d_2}+\varepsilon_{d_2}^2+\ldots\right)\right]^2$$

$$= \bar{Y}_d^2 E\left(\varepsilon_{d_0}+\varepsilon_{d_1}-\varepsilon_{d_2}\right)^2$$

$$= \bar{Y}_d^2\left[E\left(\varepsilon_{d_0}^2\right)+E\left(\varepsilon_{d_1}^2\right)+E\left(\varepsilon_{d_2}^2\right)+2E\left(\varepsilon_{d_0}\varepsilon_{d_1}\right)-2E\left(\varepsilon_{d_0}\varepsilon_{d_2}\right)-2\left(\varepsilon_{d_1}\varepsilon_{d_2}\right)\right]$$

$$= \bar{Y}_d^2\left[\left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)\frac{S_{y_d}^2}{\bar{Y}_d^2} + \left(\frac{1}{n_d}-\frac{1}{n_d^/}\right)\frac{S_{y_d}^2}{\bar{Y}_d^2} + \left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\frac{S_{y_{d_2}}^2}{\bar{Y}_d^2}\right.$$

$$+\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\frac{S_{x_d}^2}{\bar{X}_d^2} + \left(\frac{1}{n_d^/}-\frac{1}{N_d}\right)\frac{S_{x_d}^2}{\bar{X}_d^2} + \left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\frac{S_{x_{d_2}}^2}{\bar{X}_d^2}$$

$$+2\left\{\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_d y_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d}+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\rho_{x_{d_2}y_{d_2}}\frac{S_{x_{d_2}}}{\bar{X}_d}\frac{S_{y_{d_2}}}{\bar{Y}_d}\right\}$$

$$-2\left\{\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_d y_d}\frac{S_{x_d}}{\bar{X}_d}\frac{S_{y_d}}{\bar{Y}_d}+\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)\frac{S_{x_d}^2}{\bar{X}_d^2}\right\}$$

$$=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)S_{y_d}^2+\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)S_{y_d}^2+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}S_{y_{d_2}}^2$$

$$+\left(\frac{1}{n_d}-\frac{1}{N_d}\right)R_d^2S_{x_d}^2+\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)R_d^2S_{x_d}^2+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}R_d^2S_{x_{d_2}}^2$$

$$+2\left(\frac{1}{n_d}-\frac{1}{N_d}\right)\rho_{x_d y_d}R_dS_{x_d}S_{y_d}+2\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\rho_{x_{d_2}y_{d_2}}R_dS_{x_{d_2}}S_{y_{d_2}}$$

$$-2\left(\frac{1}{n'_d}-\frac{1}{N_d}\right)\rho_{x_d y_d}R_dS_{x_d}S_{y_d}-2\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)R_d^2S_{x_d}^2$$

$$=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)S_{y_d}^2+\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)\left(S_{y_d}^2+S_{x_d}^2R_d^2+2\rho_{x_d y_d}R_dS_{x_d}S_{y_d}\right)$$

$$+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}\left(S_{y_d}^2+R_d^2S_{x_{d_2}}^2+2\rho_{x_{d_2}y_{d_2}}R_dS_{x_{d_2}}S_{y_{d_2}}\right)$$

$$=\left(\frac{1}{n_d^{/}}-\frac{1}{N_d}\right)S_{y_d}^2+\left(\frac{1}{n_d}-\frac{1}{n_d^{/}}\right)S^{/2}{}_{d_{R_1}}+\left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}S^{/2}{}_{d_{R_2}}$$

4. Estimation of Sample Size in the Presence of Non-Response

Estimation of domain mean is developed using double sampling design based on the technique of sub-sampling of both the study and auxiliary variable of the non-response with unknown domain size. A study of cost surveys is therefore considered where a non-linear cost function is employed in obtaining the optimal sample sizes by minimizing variance for a fixed cost

4.1. Optimal Allocation in Double Sampling for the Estimation of Domain

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 non-linear cost function has been considered.

Denote the cost function for the ratio estimation by

$$C_d=c_d^{/}\left(n_d^{/}\right)^{\theta}+c_{d_0}n_d+c_{d_1}n_{d_1}+c_{d_2}r_{d_2} \tag{14}$$

Where,

$c_d^{/}$ = The cost of measuring a unit in the first sample of size $n_d^{/}$

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

c_{d_1} = The unit cost for processing the responded data of y_d at the first attempt of size n_{d_1}.

c_{d_2} = The unit cost associated with the sub-sample of size r_{d_2} from non-respondents of size n_{d_2}

However the first sample of size n_{d_1} and sub-samples of size r_{d_2} are not known until the first attempt is carried out. The cost will therefore be used in the planning for the survey. Hence the expected cost values of sizes n_{d_1} and r_{d_2} will be given by; $n_{d_1}=W_{d_1}n_d$ and $r_{d_2}=W_{d_2}\cdot\frac{n_d}{v_{d_2}}$. Hence the expected cost function is;

$$E\left[C_d\right]=C_d^{*}=c_d^{/}\left(n_d^{/}\right)^{\theta}+c_{d_0}n_d+c_{d_1}n_{d_1}+c_{d_2}W_{d_2}\frac{n_d}{v_{d_2}}$$

$$C_d^* = c_d^{/}\left(n_d^{/}\right)^{\theta} + n_d\left(c_{d_0} + c_{d_1}W_{d_1} + c_{d_2}.\frac{W_{d_2}}{v_{d_2}} \right) \quad (15)$$

4.2. Results for Double Sampling for Domain Estimation in the Presence of Non-Response

Proposition 5

The variance for the estimated domain mean for the estimated domain mean $\hat{\bar{Y}}_{d_{R_1}} = \frac{\bar{y}_d}{\bar{x}_d}.\bar{x}_d^{/}$ *is minimum for a specified cost* C_d^* *when,*

$$n_d^{/} = \left(\frac{S_{\omega_d}^2}{\theta c_d^{/}}\varnothing\right)^{\frac{1}{\theta+1}}$$

$$n_d = \sqrt{\varnothing\frac{\left(S_{d_{R_1}}^2 - W_{d_2}S_{d_{R_2}}^2\right)}{\left(c_{d_0} + c_{d_1}W_{d_1}\right)}}$$

$$v_{d_2} = \sqrt{\frac{c_{d_2}}{S_{d_{R_2}}^2}.\frac{\left(S_{d_{R_1}}^2 - W_{d_2}S_{d_{R_2}}^2\right)}{\left(c_{d_0} + c_{d_1}W_{d_1}\right)}}$$

Where,

$$S_{\omega_d}^2 = S_{y_d}^2 - S_{d_{R_1}}^2 > 0$$

$$\varnothing = \frac{1}{\lambda}$$

$$S_{d_{R_1}}^2 = S_{y_d}^2 + S_{x_d}^2R_d^2 \pm 2\rho_{x_d y_d}R_dS_{x_d}S_{y_d}$$

$$S_{d_{R_2}}^2 = S_{y_{d_2}}^2 + R_d^2S_{x_{d_2}}^2 \pm 2\rho_{x_{d_2}y_{d_2}}R_dS_{x_{d_2}}S_{y_{d_2}}$$

Proof

To determine the optimum values of v_{d_2}, n_d and $n_d^{/}$ that minimizes variance at a fixed cost, define

$$G(W_d) = \left(\frac{1}{n_d^{/}} - \frac{1}{N_d}\right)S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n_d^{/}}\right)S_{d_{R_1}}^2 + \left(\frac{v_{d_2}-1}{n_d}\right)W_{d_2}S_{d_{R_2}}^2$$

$$+\lambda\left[c_d^{/}\left(n_d^{/}\right)^{\theta} + n_d\left(c_{d_0} + c_{d_1}W_{d_1} + c_{d_2}.\frac{W_{d_2}}{v_{d_2}} \right) - C_d^* \right] \quad (16)$$

To obtain the normal equations, the expression of Equation (16) is differentiated partially with respect to v_{d_2}, n_d and $n_d^{/}$, and the partial derivatives are equated to zero

$$\frac{\partial G(W_d)}{\partial n_d^{/}} = \frac{-S_{y_d}^2}{n_d^{/2}} + \frac{S_{d_{R_1}}^2}{n_d^{/2}} + \lambda c_d^{/}\theta\left(n_d^{/}\right)^{\theta-1} = 0$$

$$= -S_{y_d}^2 + S_{d_{R_1}}^2 + \lambda c_d^{/}\left(n_d^{/}\right)^{\theta+1} = 0$$

$$\lambda c_d^{/}\left(n_d^{/}\right)^{\theta+1} = S_{y_d}^2 - S_{d_{R_1}}^2$$

$$\left(n_d^{/}\right)^{\theta+1} = \frac{S_{y_d}^2 - S_{d_{R_1}}^2}{\lambda c_d^{/}\theta}$$

Let $S_{\omega_d}^2 = S_{y_d}^2 - S_{d_{R_1}}^2 > 0$, thus,

$$\left(n_d^{/}\right)^{\theta+1} = \frac{S_{\omega_d}^2}{\lambda c_d^{/}\theta}$$

$$n_d^{/} = \left(\frac{S_{\omega_d}^2}{\lambda c_d^{/}\theta}\right)^{\frac{1}{\theta+1}} = \left(\frac{S_{\omega_d}^2}{c_d^{/}\theta}\varnothing\right)^{\frac{1}{\theta+1}}$$

Next the partial derivative with respect to v_{d_2} obtained as;

$$\frac{\partial G(W_d)}{\partial v_{d_2}} = \frac{W_{d_2}S_{d_{R_2}}^2}{n_d} - \frac{\lambda c_{d_2}W_{d_2}n_d}{v_{d_2}^2} = 0$$

$$\frac{W_{d_2}S_{d_{R_2}}^2}{n_d} = \frac{\lambda c_{d_2}W_{d_2}n_d}{v_{d_2}^2}$$

$$\lambda c_{d_2}W_{d_2}{n_d}^2 = v_{d_2}^2W_{d_2}S_{d_{R_2}}^2$$

$$v_{d_2} = \frac{n_d\sqrt{\lambda c_{d_2}}}{S_{d_{R_2}}} \quad (17)$$

Consider the equation

$$G(W_d) = \left(\frac{1}{n_d^{/}} - \frac{1}{N_d}\right)S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n_d^{/}}\right)S_{d_{R_1}}^2 + \left(\frac{v_{d_2}}{n_d} - \frac{1}{n_d}\right)W_{d_2}S_{d_{R_2}}^2$$

$$+\lambda\left[c_d^{/}\left(n_d^{/}\right)^{\theta} + n_d\left(c_{d_0} + c_{d_1}W_{d_1} + c_{d_2}.\frac{W_{d_2}}{v_{d_2}} \right) - C_d^* \right] \quad (18)$$

But from (17),

$$\frac{v_{d_2}}{n_d} = \frac{\sqrt{\lambda c_{d_2}}}{S_{d_{R_2}}}$$

and

$$\frac{n_d}{v_{d_2}} = \frac{S_{d_{R_2}}}{\sqrt{\lambda c_{d_2}}} \quad (19)$$

Substituting this in Equations (18) we obtain

$$G(W_d) = \left(\frac{1}{n_d^/} - \frac{1}{N_d}\right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n_d^/}\right) S_{d_{R_1}}^2 + \left(\frac{\sqrt{\lambda c_{d_2}}}{S_{d_{R2}}} - \frac{1}{n_d}\right) W_{d_2} S_{d_{R_2}}^2$$

$$+\lambda\left[c_d^/ \left(n_d^/\right)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} \right) + c_{d_2} W_{d_2} \frac{S_{d_{R_2}}}{\sqrt{\lambda c_{d_2}}} \right) - C_d^* \right] \tag{20}$$

The partial derivative of the equation (20) with respect to n_d is obtained as

$$\frac{\partial G(W_d)}{\partial n_d} = -\frac{S_{d_{R_1}}^2}{n_d^2} + \frac{W_{d_2} S_{d_{R_2}}^2}{n_d^2} + \lambda\left(c_{d_0} + c_{d_1} W_{d_1}\right) = 0$$

$$= -S_{d_{R_1}}^2 + W_{d_2} S_{d_{R_2}}^2 + \lambda n_d^2 \left(c_{d_0} + c_{d_1} W_{d_1}\right) = 0$$

$$\lambda n_d^2 \left(c_{d_0} + c_{d_1} W_{d_1}\right) = S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2$$

$$n_d^2 = \frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{\lambda\left(c_{d_0} + c_{d_1} W_{d_1}\right)}$$

$$n_d = \sqrt{\varnothing\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_1}}^2}{c_{d_0} + c_{d_1} W_{d_1}}\right)}$$

Where $\varnothing = \frac{1}{\lambda}$ But $v_{d_2} = \frac{\sqrt{\lambda c_{d_2}}}{S_{d_{R_2}}}.n_d$ from equation (19)

Thus,

$$v_{d_2} = \sqrt{\frac{c_{d_2}}{S_{d_{R_2}}^2} \cdot \frac{\left(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2\right)}{\left(c_{d_0} + c_{d_1} W_{d_1}\right)}}$$

To obtain λ the values of v_{d_2}, n_d and $n_d^/$ are substituted in the cost function equation (16) and then solve for the value of λ. Suppose the cost function is given by

$$C_d^* = c_d^/ \left(n_d^/\right)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} + c_{d_2} . \frac{W_{d_2}}{v_{d_2}} \right)$$

Then,

$$C_d^* = c_d^/ \left\{ \left(\frac{S_{\omega_d}^2}{\theta\lambda c_d^/}\right)^{\frac{1}{\theta+1}} \right\}^\theta + \frac{1}{\sqrt{\lambda}} \sqrt{\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{c_{d_0} + c_{d_1} W_{d_1}}\right)} \left(c_{d_0} + c_{d_1} W_{d_1} + \frac{c_{d_2} W_{d_2}}{\sqrt{c_{d_2}}} \sqrt{\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{c_{d_0} + c_{d_1} W_{d_1}}\right)} \right)$$

$$C_d^* = c_d^/ \left(\frac{1}{c_d^/}\right)^{\frac{\theta}{\theta+1}} \left(\frac{S_{\omega_d}^2}{\theta}\right)^{\frac{\theta}{\theta+1}} \left(\frac{1}{\lambda}\right)^{\frac{\theta}{\theta+1}} + \frac{1}{\sqrt{\lambda}} \left[W_{d_2} S_{d_{R_2}} + \left(c_{d_0} + c_{d_1} W_{d_1}\right) \sqrt{\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{c_{d_0} + c_{d_1} W_{d_1}}\right)} \right] \tag{21}$$

Let,

$$A = \left(c_d^{/}\right)^{\frac{1}{\theta+1}} \left(\frac{S_{\omega_d}^2}{\theta}\right)^{\frac{\theta}{\theta+1}}$$

$$B = \left(c_{d_0} + c_{d_1} W_{d_1}\right) \sqrt{\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{c_{d_0} + c_{d_1} W_{d_1}}\right)} + W_{d_2} S_{d_{R_2}} \sqrt{c_{d_2}}$$

$$C = C_d^*$$

The equation (20) becomes;

$$A\lambda^{-\frac{\theta}{\theta+1}} + B\lambda^{-\frac{1}{2}} - C = 0 \tag{22}$$

If $\theta = 1$ and substituting this value in the equation (20) we obtain a linear equation of the form

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

With the values of A and B defined as;

$$A = \left(c_d^{/}\right)^{\frac{1}{2}} \left(\frac{S_{\omega_d}^2}{\theta}\right)^{\frac{1}{2}},\ B = \left(c_{d_0} + c_{d_1} W_{d_1}\right) \sqrt{\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{c_{d_0} + c_{d_1} W_{d_1}}\right)} + W_{d_2} S_{d_{R_2}} \sqrt{c_{d_2}} \text{ and } C = C_d^*$$

Solving the linear equation solution obtained is,

$$\lambda = \sqrt{\frac{A+B}{C}}$$

When $\theta = \frac{1}{3}$ and substituting this value in the equation (22) we obtain a linear equation of the form,

$$A\lambda^{-\frac{1}{4}} + B\lambda^{-\frac{1}{2}} - C = 0 \tag{23}$$

With the values of A defined as;

$$A = \left(c_d^{/}\right)^{\frac{1}{2}} \left(\frac{S_{\omega_d}^2}{\theta}\right)^{\frac{1}{2}}$$

While B and C remains as earlier defined
Solving the equation (23) solution obtained is,

$$\lambda = \left[\frac{2A}{\left(\sqrt{B^2 + 4AC}\right) - B}\right]^4$$

Proposition 6

If the expected cost function is of the form $C_d^* = c_d^{/} \log n_d^{/} + n_d \left(c_{d_0} + c_{d_1} W_{d_1} + c_{d_2} \cdot \frac{W_{d_2}}{v_{d_2}} \right)$ *then the variance of the estimated domain mean* $\overline{y}_d$ *is minimum for a specified cost* C_d^* *if;*

$$n_d^{/} = \left(\frac{S_{\omega_d}^2}{\theta c_d^{/}} \varnothing\right)$$

$$n_d = \sqrt{\varnothing \frac{\left(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2\right)}{\left(c_{d_0} + c_{d_1} W_{d_1}\right)}}$$

$$v_{d_2} = \sqrt{\frac{c_{d_2}}{S_{d_{R_2}}^2} \cdot \frac{\left(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2\right)}{\left(c_{d_0} + c_{d_1} W_{d_1}\right)}}$$

Where,

$$\varnothing = \frac{1}{\lambda}$$

Proof

The proof for n_d and v_{d_2} is the same as the one in proposition 5 above. For $n_d^{/}$ the Lagrangian multiplier technique is used.

Let,

$$G(W_d) = \left(\frac{1}{n_d^{/}} - \frac{1}{N_d}\right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n_d^{/}}\right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d}\right) W_{d_2} S_{d_{R_2}}^2$$

$$+\lambda\left[c_d^{/}\log n_d^{/} + n_d\left(c_{d_0} + c_{d_1}W_{d_1} + c_{d_2}.\frac{W_{d_2}}{v_{d_2}} \right) - C_d^{*} \right] \quad (24)$$

To obtain the normal equations for the expression (24) the equation is differentiated partially with respect to $n_d^{/}$ and the partial derivatives are equated to zero

$$\frac{\partial G(W_d)}{\partial n_d^{/}} = \frac{-S_{y_d}^2}{n_d^{/2}} + \frac{S_{d_{R_1}}^2}{n_d^{/2}} + \frac{\lambda c_d^{/}}{n_d^{/}} = 0$$

$$= -S_{y_d}^2 + S_{d_{R_1}}^2 + \lambda c_d^{/} n_d^{/} = 0$$

$$\lambda c_d^{/} n_d^{/} = S_{y_d}^2 - S_{d_{R_1}}^2$$

$$n_d^{/} = \frac{S_{y_d}^2 - S_{d_{R_1}}^2}{\lambda c_d^{/}}$$

But, $S_{\omega_d}^2 = S_{y_d}^2 - S_{d_{R_1}}^2 > 0$, thus,

$$n_d^{/} = \frac{S_{\omega_d}^2}{\lambda c_d^{/}} = \left(\frac{S_{\omega_d}^2}{c_d^{/}} \varnothing \right)$$

$$\varnothing = \frac{1}{\lambda}$$

To solve for λ, let the variance be given as V_0 then substitute the values of v_{d_2}, n_d and $n_d^{/}$ into the equation,

$$G(W_d) = \left(\frac{1}{n_d^{/}} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n_d^{/}} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 = V_0$$

$$= \left(\frac{1}{n_d^{/}} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n_d^{/}} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 - V_0 = 0$$

$$= \frac{1}{n_d^{/}}\left(S_{y_d}^2 - S_{d_{R_1}}^2 \right) + \frac{1}{n_d}\left\{ \left(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2 \right) + v_{d_2} W_{d_2} S_{d_{R_2}}^2 \right\} - \left(\frac{S_{y_d}^2}{N_d} + V_0 \right) = 0 \quad (25)$$

Substitute the values of v_{d_2}, n_d and $n_d^{/}$ into the equation (25) and simplify to obtain,

$$\lambda c_d^{/} + \sqrt{\lambda}\left[\sqrt{\left(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2 \right)\left(c_{d_0} + c_{d_1} W_{d_1} \right)} \right] + W_{d_2} S_{d_{R_2}}^2 \sqrt{c_{d_2}} - V_d^{*} = 0 \quad (26)$$

Let,

$$V_d^{*} = \left(\frac{S_{y_d}^2}{N_d} + V_0 \right),\ A = c_d^{/} \text{ and } B = \sqrt{\left(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2 \right)\left(c_{d_0} + c_{d_1} W_{d_1} \right)} + W_{d_2} S_{d_{R_2}}^2 \sqrt{c_{d_2}}$$

Thus equation (26) becomes,

$$A\lambda + B\lambda^{\frac{1}{2}} - V_d^{*} = 0 \quad (27)$$

Solving for λ in equation (27) the solution becomes,

$$\lambda = \left[\frac{\left(\sqrt{B^2 + 4AC} \right) - B}{2A} \right]^{\frac{1}{2}}$$

5. Conclusion

From the results it is noted that as values of first sample domain size ($n_d^{/}$) tends to $(\rightarrow)$ domain population size (N_d), second sample size (n_d) tends to $(\rightarrow)$ $n_d^{/}$ and inverse sampling rate (v_{d_2}) tends to $(\rightarrow)$ 1 then the MSE tends asymptotically to 0. 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 θ increases then the values of $n_d^{/}$ and n_d are minimized with the reduction in the value of Lagrangian multiplier (λ) which minimizes the cost function.

References

[1] Aditya, K., Sud U., and Chandra H., (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.

[2] Cochran W. G., (1977) *Sampling techniques*. New York: John Wiley and Sons, (1977).

[3] Chaudhary M. K, and Kumar A., (2016). Estimation of Mean of Finite Population using Double sampling Scheme under Non-response, *Journal of Mathematical Sciences* 5 (2), 4 pp 287 297.

[4] Gamrot, W., (2006). Estimation of Domain Total under Non-response using Double Sampling, *Statistics in Transition*, 7 (4) pp. 831-840.

[5] Hansen M. H. and Hurwirtz W. W, (1946). The problem of Non-response in Sample Surveys, The Journal of the American Statistical Association, 41 517-529.

[6] Kalton, G., and Kasprzyk, D (1986). The treatment of Missing Survey Data, *Survey Methodology,* 12 pp. 1-16.

[7] Meeden, G., (2005). A Non-information Bayesian Approach to Domain Estimation. *Journal of Statistical Planning and Inference*, 129 (2) pp. 85-92.

[8] Oh, H. L, and Scheuren F. J., (1983). Weighting adjustments for unit non-response, In W. G Madow, I, Olkin And B Rudin (Eds), Incomplete data in sample surveys New, *York Academic press*, 2 pp 143-184.

[9] Okafor F. C, (2001). Treatment of Non-response in Successive Sampling, *Statistica*, 61 (2) 195 204.

[10] Sahoo, L. N and Panda, P., (199). Estimation Using Auxiliary Information in Two-Stage Sampling, *Austrialian and New Zealand Journal of Statistics* 41 (4) pp. 405-410.

[11] Srivastava, S. K., Jhajj, H. S., (1983). A Class of Estimators of Population Means Using Multi-Auxiliary Information. *Cal. Stat. Assoc. Bull.* 10 (2) pp. 47-56.

[12] Udofia G. A., (2002). Estimation for Domains in Double Sampling for Probabilities Proportional to Size, *the Indian Journal of Statistics* 63 pp. 82-89.