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

Alilah David Anekeya^{1,*}, Ouma Christopher Onyango², Nyongesa Kennedy¹

¹Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

²Departments 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:

Alilah David Anekeya, Ouma Christopher Onyango, Nyongesa Kennedy. Optimal Allocation in Domains Mean Estimation Using Double Sampling with Non-Linear Cost Function in the Presence of Non-Response. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 2, 2018, pp. 45-57. doi: 10.11648/j.ajtas.20180702.11

Received: December 13, 2017; Accepted: January 5, 2018; Published: February 12, 2018

Abstract: Studies have been carried out on domain mean estimation using non-linear cost function. However little has been done on domain stratum estimation using non-linear cost function using ratio estimation in the presence of non-response. This study develops a method of optimal stratum sample size allocation in domain mean estimation using double sampling with non-linear cost function in the presence of non-response. To obtain an optimum sample size, Lagrangian multiplier technique is employed by minimizing precision at a specified cost. In the estimation of the domain mean, auxiliary variable information in which the study and auxiliary variables both suffers from non-response in the second phase sampling is used. The expressions of the biases and mean square errors of proposed estimator has also been obtained.

Keywords: Optimal Allocation, Double Sampling, Non-Linear Cost Function, Non-Response

1. Introduction

1.1. Domains

In sampling, estimates are made in each of the class into which the population is subdivided. Such subgroups or classes are known as the domain of study. Units of domains may sometimes be identified prior to sampling. Such domains are called planned domains. For unplanned domain the units cannot be identified prior to sampling and hence the estimates of certain domains is often evident only after the sampling design has been identified or after the

Sampling and field work have been completed. Hence the size of unplanned domain cannot be controlled. The sample sizes for sub-populations are random variables since formation of these sub-populations is unrelated to sampling.

According to Eurostat [4] the precision threshold and or minimum effective sample sizes are set up for effective planned domains. The minimum sample sizes required to achieve a relative margin error of 100.k% for the total Y_d

(Domain total) of a study variable y_d over domain U_d of size N_d given by

$$n_d(\min) = \frac{Z_{1-\alpha/2}^2 \cdot N_d^2 S_{y_d}^2}{K^2 Y_d^2 + Z_{1-\alpha/2}^2 N_d^2 S_{y_d}^2} \quad (1)$$

Where $S_{y_d}^2$ is the variance of y over the domain and $Z_{1-\alpha/2}$ is the percentile value at $100(1-\alpha/2)\%$ of normal distribution with mean 0 and variance 1, K is the relative margin of error expressed as a proportion while 100.k% is the relative margin error expressed as a percentage. The population value Y_d and $S_{y_d}^2$ is unknowns and have to be estimated using data from auxiliary sources.

1.2. Optimal Allocation with Non-Linear Cost Function

Optimal sample allocation involves determining the

sample size n_1, n_2, \dots, n_H that minimizes the various cost characters under a given sampling budget C (where C is the upper limit of the total cost of the survey). Linear cost function is appropriate when the cost involved is associated with non-travel activities of survey e.g drawing sample, preparing survey methods, locating, identifying, interviewing respondents and coding data. Such a linear cost function can

be of the form, $C = c' n' + \sum_{h=1}^H c_h n_h$

where

c' = cost of classification per unit

n' =size of the first sample

c_h = Cost of measuring a unit in stratum h

n_h = number of units in stratum h .

Generally the above linear cost function is mostly applicable when the major cost item is that of taking the measurements on each unit without considering the cost of the distance between the sample units.

Chernyak [1] proposed optimum allocation in double sampling for stratification with a non-linear cost function. The proposed non-linear cost function is of the form,

$C = c' (n')^\alpha + \sum_{k=1}^L c_k n_k$, $\alpha > 0$ where, c' is the cost of

classification per unit and c_k is the cost of measuring a unit in stratum k . Another non-linear cost function proposed was logarithmic in nature of the form,

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

Okafor and Lee [9] employed the double sampling method to estimate the mean of the auxiliary variable and proceeded to estimate the mean of the study variable in a similar way as Cochran [3]. In this method double sampling ratio and regression estimation was considered. The distribution of the auxiliary information was not known and hence 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 was derived for the estimators at a fixed cost. Performance of the proposed estimators was 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. Tschuprow [11] and Neyman [8] proposed the allocation procedure that minimizes variance of sample mean under a linear cost function of sample size $n = \sum_{h=1}^H n_h$, where n_h is the size of the stratum.

Neyman [8] used Lagrange multiplier optimization technique to get optimum sample sizes for a single variable under study. Holmberg [6] addressed the problem of compromised allocation in multivariate Stratified sampling by taking into consideration minimization of some of the

variances or coefficient of variation of the population parameters and of some of the efficiency losses which may be as a result of increase in the variance due to the use of compromise allocation. Saini [10] developed a method of optimum allocation for multivariate stratified two stage sampling design by using double sampling. In this method the problem of determining optimum allocations was formulated as non-linear programming problem (NLPP) in which each NLPP has a convex objective function under a single linear constraints. The Lagrange multiplier technique was used to solve the formulated NLPPs. Khan et al. [7] proposed a quadratic cost function for allocating sample size in multivariate stratified random sampling in the presence of non-response in which a separate linear regression estimator is used. In this multi-objective

Non-linear integer programming problem, an extended lexicographic goal programming was used for solution purpose and comparison made with individual optimum techniques. It is observed that in the allocation techniques, the extended lexicographic goal programming gives minimum values of coefficient of variation than the individual optimum and goal programming technique. Choudhry [2] considered sample allocation issues in the context of estimating Sub-populations (stratum and domain) means as well as the aggregate population means under stratified simple random sampling. In this method non-linear programming was used to obtain the optimal sample allocation to the strata that minimizes the total sample sizes subject to a specified tolerance on the coefficient of variation of the estimators of strata and population means.

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 paper 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 component is inclusive and it is not linear.

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

2.1. Introduction

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 \bar{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, \dots, U_D of sizes $N_1, N_2, \dots, N_d, \dots, N_D$ 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, $U = \bigcup_{d=1}^D U_d$ and $N = \sum_{d=1}^D N_d$ respectively.

Let Y_d and X_d be the domain study and auxiliary variables respectively. Further, let \bar{Y}_d and \bar{X}_d be their respective domain population means and auxiliary means with $y_{d_i} (i=1,2,3,\dots,N_d)$ and $x_{d_i} (i=1,2,\dots,N_d)$ observations on the i^{th} unit. In estimating the domain auxiliary population mean \bar{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 \bar{X}_d of the auxiliary variables X_d from a large first phase sample of size n'_d , values of the observations $x'_{d_i} (i=1,2,3,\dots,n'_d)$ are obtained and a sample auxiliary domain mean \bar{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 $(i=1,2,3,\dots,n_d)$. Let n_{d_1} 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 $(n_{d_1} + r_{d_2})$ observations on y_{d_i} domain study character is given by;

$$\begin{aligned} \bar{y}_d &= \frac{n_{d_1}}{n_d} \bar{y}_{d_1} + \frac{n_{d_2}}{n_d} \bar{y}_{r_{d_2}} \\ &= w_{d_1} \bar{y}_{d_1} + w_{d_2} \bar{y}_{r_{d_2}} \end{aligned} \tag{2}$$

Similarly the estimate for domain auxiliary variable is given by;

$$\begin{aligned} \bar{x}_d &= \frac{n_{d_1}}{n_d} \bar{x}_{d_1} + \frac{n_{d_2}}{n_d} \bar{x}_{r_{d_2}} \\ &= w_{d_1} \bar{x}_{d_1} + w_{d_2} \bar{x}_{r_{d_2}} \end{aligned} \tag{3}$$

Where \bar{y}_d and \bar{x}_d are the sample domain means for the

observation y_{d_i} and x_{d_i} respectively.

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

$$\hat{Y}_{d_{R_1}} = \frac{\bar{y}_d}{\bar{x}_d} \cdot \bar{x}'_d = r_d \bar{x}'_d \text{ and } \hat{Y}_{d_2} = \frac{\bar{y}_d}{\bar{x}'_d} \cdot \bar{x}_d = r_d \bar{x}_d$$

With the assumption that,

$$E[\bar{x}'_d] = E[\bar{x}_d] = \bar{X}_d, \quad E(\bar{y}_d) = \bar{Y}_d \quad (4)$$

3. Mean Square Error of the Ratio Estimator

The expression for the Mean square error (MSE) of $\bar{Y}_{d_{R_1}}$

$$\begin{aligned} E(\epsilon_{d_0}^2) &= E\left[\left(\frac{\bar{y}_d - \bar{Y}_d}{\bar{Y}_d}\right)^2\right] \\ &= E\left[\frac{(\bar{y}_d - \bar{Y}_d)^2}{\bar{Y}_d^2}\right] = \frac{1}{\bar{Y}_d^2} E[\bar{y}_d - \bar{Y}_d]^2 \\ &= \frac{1}{\bar{Y}_d^2} \text{Var}(\bar{y}_d) \\ &= \frac{1}{\bar{Y}_d^2} \left[V_1 E_2 E_3 (\bar{y}_d / n'_d) + E_1 V_2 E_3 (\bar{y}_d / n_d) + E_1 E_2 V_3 (\bar{y}_d / r_{d_2}) \right] \\ &= \frac{1}{\bar{Y}_d^2} \left[\left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{y_d}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{y_{d_2}}^2 \right] \\ &= \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) C_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) C_{y_d}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} C_{y_{d_2}}^2 \end{aligned} \quad (6)$$

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

$$\begin{aligned} E[\epsilon_1^2] &= E\left[\frac{\bar{x}_d - \bar{X}_d}{\bar{X}_d}\right]^2 = \frac{1}{\bar{X}_d^2} E[\bar{x}_d - \bar{X}_d]^2 \\ &= \frac{1}{\bar{X}_d^2} \text{Var}(\bar{x}_d) \end{aligned}$$

and $\bar{Y}_{d_{R_2}}$ are derived by the use of the Taylor's series approximation.

Let

$$\begin{aligned} \epsilon_{d_0} &= \frac{\bar{y}_d - \bar{Y}_d}{\bar{Y}_d} \Rightarrow \bar{y}_d = \bar{Y}_d (\epsilon_{d_0} + 1) \\ \epsilon_{d_1} &= \frac{\bar{x}_d - \bar{X}_d}{\bar{X}_d} \Rightarrow \bar{x}_d = \bar{X}_d (\epsilon_{d_1} + 1) \\ \epsilon_{d_2} &= \frac{\bar{x}'_d - \bar{X}_d}{\bar{X}_d} \Rightarrow \bar{x}'_d = \bar{X}_d (\epsilon_{d_0} + 1) \end{aligned} \quad (5)$$

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

Further define;

$$\begin{aligned} &= \frac{1}{\bar{X}_d^2} \left[V_1 \left[E_2 (\bar{x}_d / n_d) \right] + E_1 \left[V_2 (\bar{x}_d / r_{d_2}) \right] \right] \\ &= \left(\frac{1}{n_d} - \frac{1}{N_d} \right) \frac{S_{x_d}^2}{\bar{X}_d^2} + W_{d_2} \left(\frac{v_{d_2} - 1}{n_d} \right) \frac{S_{x_{d_2}}^2}{\bar{X}_d^2} \end{aligned} \quad (7)$$

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

$$\begin{aligned}
 E(\varepsilon_{d_2}^2) &= E\left[\frac{\bar{x}'_d - \bar{X}_d}{\bar{X}_d}\right]^2 = \frac{1}{\bar{X}_d^2} E\left[\bar{x}'_d - \bar{X}_d\right]^2 \\
 &= \frac{1}{\bar{X}_d^2} Var(\bar{x}'_d) \\
 &= \frac{1}{\bar{X}_d^2} \left(\frac{N_d - n'_d}{N_d}\right) \frac{S_{x_d}^2}{n'_d} \\
 &= \left(\frac{1}{n'_d} - \frac{1}{N_d}\right) \frac{S_{x_d}^2}{\bar{X}_d^2} \tag{8}
 \end{aligned}$$

Consider,

$$\begin{aligned}
 E[\varepsilon_{d_0} \varepsilon_{d_1}] &= E\left[\left(\frac{\bar{y}_d - \bar{Y}_d}{\bar{Y}_d}\right)\left(\frac{\bar{x}_d - \bar{X}_d}{\bar{X}_d}\right)\right] \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} E\left[(\bar{y}_d - \bar{Y}_d)(\bar{x}_d - \bar{X}_d)\right] \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} Cov(\bar{x}_d \bar{y}_d) \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} Cov\left[E(\bar{y}_d / n_d), E(\bar{x}_d / n_d)\right] + \frac{1}{\bar{X}_d \bar{Y}_d} E\left[Cov(\bar{y}_d \bar{x}_d) / n_d\right] + \frac{1}{\bar{X}_d \bar{Y}_d} E\left[Cov(\bar{y}_d \bar{x}_d) / r_{d_2}\right] \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} E\left[Cov(\bar{y}_d \bar{x}_d) / n_d\right] + \frac{1}{\bar{X}_d \bar{Y}_d} E\left[Cov(\bar{y}_d \bar{x}_d) / r_{d_2}\right] \\
 &= \left(\frac{1}{n_d} - \frac{1}{N_d}\right) \rho_{x_d y_d} \frac{S_{x_d}}{\bar{X}_d} \frac{S_{y_d}}{\bar{Y}_d} + \left(\frac{v_{d_2} - 1}{n_d}\right) \rho_{x_{d_2} y_{d_2}} W_{d_2} \frac{S_{x_{d_2}}}{\bar{X}_d} \frac{S_{y_{d_2}}}{\bar{Y}_d} \tag{9}
 \end{aligned}$$

Next,

$$\begin{aligned}
 E[\varepsilon_{d_0} \varepsilon_{d_2}] &= E\left[\left(\frac{\bar{y}_d - \bar{Y}_d}{\bar{Y}_d}\right)\left(\frac{\bar{x}'_d - \bar{X}_d}{\bar{X}_d}\right)\right] \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} E\left[(\bar{y}_d - \bar{Y}_d)(\bar{x}'_d - \bar{X}_d)\right] \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} Cov(\bar{y}_d \bar{x}'_d) \\
 &= \frac{1}{\bar{X}_d \bar{Y}_d} Cov\left[E(\bar{y}_d / n'_d), E(\bar{x}'_d / n'_d)\right] + \frac{1}{\bar{X}_d \bar{Y}_d} E\left[Cov(\bar{y}_d \bar{x}'_d) / n'_d\right] + \frac{1}{\bar{X}_d \bar{Y}_d} E\left[Cov(\bar{y}_{r_{d_2}} \bar{x}'_{r_{d_2}}) / n'_d\right] \\
 &= \left(\frac{1}{n'_d} - \frac{1}{N_d}\right) \rho_{x_d y_d} \frac{S_{x_d}}{\bar{X}_d} \frac{S_{y_d}}{\bar{Y}_d} \tag{10}
 \end{aligned}$$

Consider,

$$\begin{aligned}
 E[\varepsilon_{d_1} \varepsilon_{d_2}] &= E\left[\left(\frac{\bar{x}_d - \bar{X}_d}{\bar{X}_d}\right)\left(\frac{\bar{x}'_d - \bar{X}_d}{\bar{X}_d}\right)\right] \\
 &= \frac{1}{\bar{X}_d^2} E_1 E_2 \left[(\bar{x}_d - \bar{X}_d)(\bar{x}'_d - \bar{X}_d) / n'_d\right] \\
 &= \frac{1}{\bar{X}_d^2} E_1 \left(\bar{x}'_d - \bar{X}_d\right)^2 = \left(\frac{1}{n'_d} - \frac{1}{N_d}\right) \frac{S_{x_d}^2}{\bar{X}_d^2} \tag{11}
 \end{aligned}$$

3.1. Mean Square Error (MSE) of the Ratio Estimator $\hat{Y}_{d_{R_1}}$ and $\hat{Y}_{d_{R_2}}$ with the Sample Size Allocation

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

$$\hat{Y}_{d_{R_1}} = \frac{\bar{y}_d}{\bar{x}_d} \cdot \bar{x}'_d = r_d \bar{x}'_d \text{ and } \hat{Y}_{d_2} = \frac{\bar{y}_d}{\bar{x}_d} \cdot \bar{x}_d = r_d \bar{x}_d \text{ respectively}$$

Proposition 1

The mean square error (MSE) of the estimator defined by

$$\hat{Y}_{d_{R_1}} = \frac{\bar{y}_d}{\bar{x}_d} \cdot \bar{x}'_d = r_d \bar{x}'_d \text{ is given by:}$$

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

Where,

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

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

With the notations defined as:

$S_{y_d}^2$ = 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

R_d = Population ratio of \bar{Y}_d to \bar{X}_d

Proof

By definition,

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

Substituting the values of equations (5) we obtain

$$\begin{aligned} &= E\left[\frac{\bar{Y}_d (1 + \varepsilon_{d_0}) \bar{X}_d (1 + \varepsilon_{d_2})}{\bar{X}_d (1 + \varepsilon_{d_1})} - \bar{Y}_d\right]^2 \\ &= \bar{Y}_d^2 E\left[\frac{(1 + \varepsilon_{d_0})(1 + \varepsilon_{d_2})}{(1 + \varepsilon_{d_1})} - 1\right]^2 \\ &= \bar{Y}_d^2 E\left[(\varepsilon_{d_2} + \varepsilon_{d_0} + \varepsilon_{d_0} \varepsilon_{d_2} - \varepsilon_{d_1})(1 - \varepsilon_{d_1} + \varepsilon_{d_1}^2 + \dots)\right]^2 \\ &= \bar{Y}_d^2 E(\varepsilon_{d_0} + \varepsilon_{d_2} - \varepsilon_{d_1})^2 \\ &= \bar{Y}_d^2 \left[E(\varepsilon_{d_0}^2) + E(\varepsilon_{d_1}^2) + E(\varepsilon_{d_2}^2) + 2E(\varepsilon_{d_0} \varepsilon_{d_2}) - 2E(\varepsilon_{d_0} \varepsilon_{d_1}) - 2E(\varepsilon_{d_1} \varepsilon_{d_2}) \right] \\ &= \bar{Y}_d^2 \left[\left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S_{y_d}^2}{\bar{Y}_d^2} + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) \frac{S_{y_d}^2}{\bar{Y}_d^2} + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} \frac{S_{y_{d_2}}^2}{\bar{Y}_d^2} \right. \\ &\quad + \left(\frac{1}{n_d} - \frac{1}{N_d} \right) \frac{S_{x_d}^2}{\bar{X}_d^2} + \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S_{x_d}^2}{\bar{X}_d^2} + W_{d_2} \left(\frac{v_{d_2} - 1}{n_d} \right) \frac{S_{x_{d_2}}^2}{\bar{X}_d^2} \\ &\quad + 2 \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \rho_{x_d y_d} \frac{S_{x_d} S_{y_d}}{\bar{X}_d \bar{Y}_d} - 2 \left(\frac{1}{n_d} - \frac{1}{N_d} \right) \rho_{x_d y_d} \frac{S_{x_d} S_{y_d}}{\bar{X}_d \bar{Y}_d} \\ &\quad \left. - 2 \left(\frac{v_{d_2} - 1}{n_d} \right) \rho_{x_{d_2} y_{d_2}} W_{d_2} \frac{S_{x_{d_2}} S_{y_{d_2}}}{\bar{X}_d \bar{Y}_d} - 2 \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S_{x_d}^2}{\bar{X}_d^2} \right] \\ &= \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{y_d}^2 + \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{x_d}^2 R_d^2 \end{aligned}$$

$$\begin{aligned}
 & - \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{x_d}^2 R_d^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} \left[S_{y_d}^2 + R_d^2 S_{x_d}^2 - 2\rho_{x_{d_2} y_{d_2}} R_d S_{x_{d_2}} S_{y_{d_2}} \right] \\
 & + 2 \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \rho_{x_d y_d} R_d S_{x_d} S_{y_d} - 2 \left(\frac{1}{n_d} - \frac{1}{N_d} \right) \rho_{x_d y_d} R_d S_{x_d} S_{y_d} \\
 = & \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) \left\{ S_{y_d}^2 + S_{x_d}^2 R_d^2 - 2\rho_{x_d y_d} R_d S_{x_d} S_{y_d} \right\} \\
 & + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} \left[S_{y_d}^2 + R_d^2 S_{x_d}^2 - 2\rho_{x_{d_2} y_{d_2}} R_d S_{x_{d_2}} S_{y_{d_2}} \right] \\
 = & \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 \quad \square
 \end{aligned}$$

Proposition 2

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

$$\left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2$$

Where,

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

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

With the notations as defined in proposition 1 above

Proof

$$\begin{aligned}
 \text{MSE of } \hat{Y}_{d_{R_2}} & = \text{MSE} \left(\hat{Y}_{d_{R_2}} \right) = E \left[\hat{Y}_{d_{R_2}} - \bar{Y}_d \right]^2 \\
 & = E \left[\frac{\bar{y}_d}{\bar{x}_d} \bar{x}'_d - \bar{Y}_d \right]^2 \\
 & = E \left[\bar{Y}_d \left(\frac{(1 + \varepsilon_{d_0})(1 + \varepsilon_{d_1})}{(1 + \varepsilon_{d_2})} - 1 \right) \right]^2 \\
 & = \bar{Y}_d^2 E \left[(\varepsilon_{d_1} + \varepsilon_{d_0} + \varepsilon_{d_1} \varepsilon_{d_0} - \varepsilon_{d_2}) (1 - \varepsilon_{d_2} + \varepsilon_{d_2}^2 + \dots) \right]^2 \\
 & = \bar{Y}_d^2 E \left(\varepsilon_{d_0} + \varepsilon_{d_1} - \varepsilon_{d_2} \right)^2 \\
 & = \bar{Y}_d^2 \left[E \left(\varepsilon_{d_0}^2 \right) + E \left(\varepsilon_{d_1}^2 \right) + E \left(\varepsilon_{d_2}^2 \right) + 2E \left(\varepsilon_{d_0} \varepsilon_{d_1} \right) - 2E \left(\varepsilon_{d_0} \varepsilon_{d_2} \right) - 2 \left(\varepsilon_{d_1} \varepsilon_{d_2} \right) \right] \\
 & = \bar{Y}_d^2 \left[\left(\frac{1}{n'_d} - \frac{1}{N_d} \right) \frac{S_{y_d}^2}{\bar{Y}_d^2} + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) \frac{S_{y_d}^2}{\bar{Y}_d^2} + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} \frac{S_{y_{d_2}}^2}{\bar{Y}_d^2} \right]
 \end{aligned}$$

$$\begin{aligned}
& + \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^2 S_{x_d}^2 + \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) R_d^2 S_{x_d}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} R_{d_2}^2 S_{x_{d_2}}^2 \\
& + 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{v_{d_2} - 1}{n_d} \right) W_{d_2} \rho_{x_{d_2} y_{d_2}} R_{d_2} S_{x_{d_2}} S_{y_{d_2}} \\
& - 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) R_d^2 S_{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}^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}}^2 + R_{d_2}^2 S_{x_{d_2}}^2 + 2 \rho_{x_{d_2} y_{d_2}} R_{d_2} 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}
\end{aligned}$$

3.2. Optimal Allocation in Double Sampling for Domain Estimation

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

$$C_d = c'_d (n'_d)^\theta + c_{d_0} n_d + c_{d_1} n_{d_1} + c_{d_2} r_{d_2} \quad (12)$$

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 $r_{d_2} = W_{d_2} \frac{n_d}{v_{d_2}}$. Hence the expected cost function is;

$$E[C_d] = C_d^* = c'_d (n'_d)^\theta + c_{d_0} n_d + c_{d_1} n_{d_1} + c_{d_2} W_{d_2} \frac{n_d}{v_{d_2}}$$

$$C_d^* = c'_d (n'_d)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} + c_{d_2} \cdot \frac{W_{d_2}}{v_{d_2}} \right) \quad (13)$$

$$= -S_{y_d}^2 + S_{d_{R_1}}^2 + \lambda c'_d (n'_d)^{\theta+1} = 0$$

$$\lambda c'_d (n'_d)^{\theta+1} = S_{y_d}^2 - S_{d_{R_1}}^2$$

$$(n'_d)^{\theta+1} = \frac{S_{y_d}^2 - S_{d_{R_1}}^2}{\lambda c'_d \theta}$$

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

Proposition 3

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

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

$$(n'_d)^{\theta+1} = \frac{S_{\omega_d}^2}{\lambda c'_d \theta}$$

$$n'_d = \left(\frac{S_{\omega_d}^2}{\lambda c'_d \theta} \right)^{\frac{1}{\theta+1}} = \left(\frac{S_{\omega_d}^2}{c'_d \theta} \right)^{\frac{1}{\theta+1}}$$

$$n'_d = \left(\frac{S_{\omega_d}^2}{\theta c'_d} \right)^{\frac{1}{\theta+1}}$$

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

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

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

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

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

$$v_{d_2} = \frac{n_d \sqrt{\lambda c_{d_2}}}{S_{d_{R_2}}} \quad (15)$$

Where,

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

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

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

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

Proof

To determine the optimum values of v_{d_2} , n_d and n'_d that minimizes variance at a fixed cost, define

Consider the equation

$$G(W_d) = \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2}}{n_d} - \frac{1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 + \lambda \left[c'_d (n'_d)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} + c_{d_2} \cdot \frac{W_{d_2}}{v_{d_2}} \right) - C_d^* \right] \quad (16)$$

But from (15),

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

$$G(W_d) = \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 + \lambda \left[c'_d (n'_d)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} + c_{d_2} \cdot \frac{W_{d_2}}{v_{d_2}} \right) - C_d^* \right] \quad (14)$$

To obtain the normal equations, the expression of Equation (14) is differentiated partially with respect to v_{d_2} , n_d and n'_d , and the partial derivatives are equated to zero

and

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

$$\frac{\partial G(W_d)}{\partial n'_d} = \frac{-S_{y_d}^2}{n_d^2} + \frac{S_{d_{R_1}}^2}{n_d^2} + \lambda c'_d \theta (n'_d)^{\theta-1} = 0$$

Substituting this in Equations (17) we obtain into (16) we obtain

$$G(W_d) = \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{d_{R_1}}^2 + \left(\frac{\sqrt{\lambda c_{d_2}}}{S_{d_{R_2}}} - \frac{1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 + \lambda \left[c'_d (n'_d)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} \right) + c_{d_2} W_{d_2} \frac{S_{d_{R_2}}}{\sqrt{\lambda c_{d_2}}} \right] - C_d^* \quad (18)$$

The partial derivative of the equation (18) with respect to n_d is obtained as

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

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

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

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

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

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

$$C_d^* = c'_d \left(\frac{1}{c'_d} \right)^{\frac{\theta}{\theta+1}} \left(\frac{S_{d_{R_1}}^2}{\theta} \right)^{\frac{\theta}{\theta+1}} \left(\frac{1}{\lambda} \right)^{\frac{\theta}{\theta+1}} + \frac{1}{\sqrt{\lambda}} \left[W_{d_2} S_{d_{R_2}} + (c_{d_0} + c_{d_1} W_{d_1}) \sqrt{\left(\frac{S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2}{c_{d_0} + c_{d_1} W_{d_1}} \right)} \right] \quad (19)$$

Let,

$$A = (c'_d)^{\frac{1}{\theta+1}} \left(\frac{S_{d_{R_1}}^2}{\theta} \right)^{\frac{\theta}{\theta+1}}$$

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

$$C = C_d^*$$

The equation (19) becomes;

$$A \lambda^{-\frac{\theta}{\theta+1}} + B \lambda^{-\frac{1}{2}} - C = 0 \quad (20)$$

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

Where $\varnothing = \frac{1}{\lambda}$ But $v_{d_2} = \frac{\sqrt{\lambda c_{d_2}}}{S_{d_{R_2}}} n_d$ from equation (17)

Thus,

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

To obtain λ the values of v_{d_2} , n_d and n'_d are substituted in the cost function equation (13) and then solve for the value of λ . Suppose the cost function is given by

$$C_d^* = c'_d (n'_d)^\theta + n_d \left(c_{d_0} + c_{d_1} W_{d_1} + c_{d_2} \cdot \frac{W_{d_2}}{v_{d_2}} \right)$$

Then,

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

With the values of A and B defined as;

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

Solving the linear equation solution obtained is,

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

estimated domain mean \bar{y}_d is minimum for a specified cost C_d^* if;

$$n'_d = \left(\frac{S^2_{\omega_d}}{\theta c'_d} \right) \varnothing$$

When $\theta = \frac{1}{3}$ and substituting this value in the equation

(20) we obtain a linear equation of the form,

$$A\lambda^{\frac{1}{4}} + B\lambda^{\frac{1}{2}} - C = 0 \tag{21}$$

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

With the values of A defined as;

$$A = (c'_d)^{\frac{1}{2}} \left(\frac{S^2_{\omega_d}}{\theta} \right)^{\frac{1}{2}}$$

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

Where,

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

While B and C remains as earlier defined

Solving the equation (21) solution obtained is,

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

Proof

The proof for n_d and v_{d_2} is the same as the one in proposition 3 above. For n'_d the Lagrangian multiplier technique is used.

Let,

Proposition 4

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

$$G(W_d) = \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S^2_{y_d} + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S^2_{d_{R_1}} + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S^2_{d_{R_2}} + \lambda \left[c'_d \log n'_d + n_d \left(c_{d_0} + c_{d_1}W_{d_1} + c_{d_2} \cdot \frac{W_{d_2}}{v_{d_2}} \right) - C_d^* \right] \tag{22}$$

To obtain the normal equations for the expression (22) the equation is differentiated partially with respect to n'_d and the partial derivatives are equated to zero

$$\frac{\partial G(W_d)}{\partial n'_d} = \frac{-S^2_{y_d}}{n'^2_d} + \frac{S^2_{d_{R_1}}}{n'^2_d} + \frac{\lambda c'_d}{n'_d} = 0$$

$$= -S^2_{y_d} + S^2_{d_{R_1}} + \lambda c'_d n'_d = 0$$

$$\lambda c'_d n'_d = S^2_{y_d} - S^2_{d_{R_1}}$$

$$n'_d = \frac{S_{y_d}^2 - S_{d_{R_1}}^2}{\lambda c'_d}$$

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,

$$\begin{aligned} G(W_d) &= \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 = V_0 \\ &= \left(\frac{1}{n'_d} - \frac{1}{N_d} \right) S_{y_d}^2 + \left(\frac{1}{n_d} - \frac{1}{n'_d} \right) S_{d_{R_1}}^2 + \left(\frac{v_{d_2} - 1}{n_d} \right) W_{d_2} S_{d_{R_2}}^2 - V_0 = 0 \\ &= \frac{1}{n'_d} (S_{y_d}^2 - S_{d_{R_1}}^2) + \frac{1}{n_d} \left\{ (S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2) + v_{d_2} W_{d_2} S_{d_{R_2}}^2 \right\} - \left(\frac{S_{y_d}^2}{N_d} + V_0 \right) = 0 \end{aligned} \quad (23)$$

Substitute the values of v_{d_2} , n_d and n'_d into the equation (23) and simplify to obtain,

$$\lambda c'_d + \sqrt{\lambda} \left[\sqrt{(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2)(c_{d_0} + c_{d_1} W_{d_1})} \right] + W_{d_2} S_{d_{R_2}}^2 \sqrt{c_{d_2}} - V_d^* = 0 \quad (24)$$

Let,

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

Thus equation (24) becomes,

$$A\lambda + B\lambda^2 - V_d^* = 0 \quad (25)$$

Solving for λ in equation (21) the solution becomes,

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

4. Conclusion

It is noted that value of inverse sampling rate (v_{d_2}) do not depend on the value of the Lagrangian multiplier (λ). Further the value of sampling rate, $v_{d_2} < 1$, if c_{d_2} (the unit cost associated with the sub-sample of size r_{d_2} from non-

respondents of size n_{d_2}) is less than both c_{d_1} (the unit cost for processing the responded data of y_d at the first attempt of size n_{d_1}) and c_{d_0} (the cost of measuring a unit of the first attempt on y_d with second phase sample size n_d) and also when the value of $S_{d_{R_1}}^2$ is not too large relative to $S_{d_{R_2}}^2$. The second phase sample size (n_d) will be minimum if the value of $(S_{d_{R_1}}^2 - W_{d_2} S_{d_{R_2}}^2) > 0$ but less than 1. If the value of $S_{\omega_d}^2 = S_{y_d}^2 - S_{d_{R_1}}^2 < 1$ and $S_{y_d}^2 > S_{d_{R_1}}^2$ with the values of c'_d (the cost of measuring a unit in the first sample of size n'_d) not being too large to the relative $S_{\omega_d}^2$ then the value of n'_d (size of the first sample) will be minimum. These minimum values therefore make the theoretic cost survey of the proposed estimator as minimal as possible.

References

- [1] 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.
- [2] Choudhry H. G., Rao, J. N. K, and Michael A., Hidiroglou, (2012). On sample allocation for efficient domain estimation, *Survey methodology*, 38 (1) pp. 23-29.
- [3] Cochran W. G., (1977) *Sampling techniques*. New York: John Wiley and Sons, (1977).
- [4] Eurostat., (2008). Introduction to Sample Design and Estimation Techniques, *Survey Sampling Reference Guidelines*. Luxembourg; Office for Publication of the European Communities pp. 36.
- [5] Hansen M. H. and Hurwitz W. W, (1946). The problem of non-response in sample surveys. *The Journal of the American Statistical Association*, 41 517-529.
- [6] Holmberg A., (2002). A multi-parameter perspective on the choice of sampling designs in surveys. *Journal of statistics in transition*. 5 (6) pp. 969-994.
- [7] Khan S. U., Muhammad Y. S., and Afgan N., (2009). Multi-objective compromise allocation stratified sampling in the presence of non-response using quadratic cost function. *International Journal of Business and social science*. 5 (13). pp. 162-169.
- [8] Neyman, C. and Jerzy D. (1934).; On the Two Different Aspects of the Representative methods of stratified sampling and the method of purposive selection, *Journal of royal statistical society*. 97 (4) pp. 558-625.
- [9] Okafor F. C, (2001). Treatment of non-response in successive sampling, *Statistica*, 61 (2) 195-204.
- [10] Saini M., and Kumar A, (2015). Method of Optimum allocation for Multivariate Stratified two stage Sampling design Using double Sampling. *Journal of Probability and Statistics forum*, 8, pp. 19-23.
- [11] Tschuprow and Al A., (1923). On mathematical expectation of the moments of frequency distribution in the case of correlated observation (chapters 4-6) *Metron* 2 (1) pp. 646-683, (1923).