Journal of the Nigerian Association of Mathematical Physics
***Volume 48* (*Sept. & Nov., 2018 Issue*), pp93 –102**
© J. of NAMP

OPTIMUM COMPROMISE ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING DESIGN WITH GAMMA COST FUNCTION: A CASE OF NON-RESPONSE

Abubakar Yahaya and Okon Ekom-Obong Jackson*

Department of Statistics, Ahmadu Bello University, Zaria

Abstract

In this paper, we proposed a new gamma cost function which approximates the linear and quadratic cost functions in the case of non-response when more than one characteristic is considered. Since different characteristics are measured in different units, it is ideal to consider the minimization of the squared coefficient of variation as an objective for a given fixed cost. The problem of optimum allocation in multivariate stratified sampling is described as a multi-objective integer non-linear programming problem based on the separate regression estimator. A solution procedure is developed using two different compromise allocations namely extended lexicographic goal programming approach and value function technique which were compared to the corresponding individual optimum allocations. A numerical example is presented to illustrate the computational applicability of such approaches.

Keywords: Multivariate Stratified Sampling, Compromise Allocation, Non-response, Quadratic Cost Function, Lexicographic Goal programming.

1. Introduction

Stratified random sampling is more suitable compared to other survey designs based on efficiency when obtaining information from heterogeneous population. In sampling surveys, non-response is a situation that occurs when the desired information is not obtained from a sampled person involved in the survey. The extent of non-response depends on various factors such as type of target population, type of survey and the time of survey [1].

The problem of optimum allocation was considered in [2,3] for stratified random sampling using a univariate population. When more than one characteristics are defined on each and every unit of the population, optimum allocation tends to be cumbersome in the sense that; univariate allocation methods are not optimum when planning multivariate surveys, in such situations the problem may be resolved using an allocation that is optimum in some sense for all the characteristics; the resulting allocation is called a "compromise allocation". Various optimum compromise allocation techniques for multipurpose survey have been formulated either by minimizing the coefficient of variation (CV) for a fixed cost or minimizing the fixed cost for a specified variance. Furthermore, sampling efficiency largely depends on how the sample sizes are allocated to different strata. Several authors [4 - 12] and many others worked out different compromise allocation criterion or examined the existing approaches under different situations such as the presence of non-response, use of auxiliary information, use of double sampling techniques and so forth.

This paper proposes a gamma cost function which approximates the linear and quadratic cost function in the presence of non-response and the problem of determining compromise allocation in multivariate stratified random sampling is formulated as a Multi-objective integer non-linear programming problem (MOINLPP). LINGO optimization software is utilized to solve numerical example which illustrates the computational details of the allocation procedure.

The article is organized in such a way that, section 2 discusses some sampling strategies and estimation procedures. Section 3 discusses the proposed gamma cost function given in the presence non-response. In section 4, we discuss the adopted optimization technique (MOINLPP). Some numerical illustrations were given in section 5 and the results obtained were discussed in section 6. Some concluding remarks were presented in section 7.

Corresponding Author**:** Abubakar Y., Email: ensiliyu2@yahoo.co.uk, Tel: +2348069334466, +2348182486302 (OEJ)

2. Sampling Strategies and Estimation Procedures

Consider a finite population of size N which is divided into L mutually exclusive and exhaustive strata with size $N_h (h = 1,2,...L)$ such that $\sum_{h=1}^{L} N_h = N$. An independent simple random sample without replacement (SRSWOR) of size n_h is drawn from each stratum such that $\sum_{h=1}^{L} n_h = n$. It is assumed that every stratum is initially divided into two mutually exclusive and exhaustive groups, one comprising of respondents and the other for non-respondents. Let N_{h1} and $N_{h2} = N_h - N_{h1}$ be the sizes of the respondents and non-respondents group respectively in the h^{th} stratum. The true values of N_{h1} and N_{h2} or their estimate are not known prior to when the sample observations are obtained. Let $n_h, h = 1,2,..., L$ units be drawn from the h^{th} stratum and let n_{h1} units belong to the respondents and the remaining $n_{h2} = n_h - n_{h1}$ units belong to the non-respondents group.

To have some representation from the non-respondent group of the sample, a more careful second attempt is made to obtain information on a random subsample of size r_h out of the n_{h2} non-respondents. The desired information is now collected from the r_h units by personal interview. It is assumed that this time around all the r_h units respond.

At the second attempt, let the subsample sizes

$$r_h = \frac{n_{h2}}{k_h}, \; h = 1,2,...L \tag{1}$$

be drawn from n_{h2} non-respondents group of the h^{th} stratum, where $k_h \geq 1$, and $1/k_h$ denote the sampling fraction among non-respondents.

Suppose we observe Y_{jhi} for $j = 1,2,..., p$ characteristics, $h = 1,2,..., L$ strata with $i = 1,2,..., N_h$ the sampling units in the h^{th} stratum. Let $\bar{y}_{jh}$ and $\bar{x}_{jh}$ be the sample means, $\bar{Y}_{jh}$ and $\bar{X}_{jh}$ be the strata means of the study variable Y_{jhi} and the auxiliary variable X_{jhi} respectively of the j^{th} characteristics in the h^{th} stratum, thus, the population mean is estimated for $p \geq 2$ characteristics ineach stratum. Let S_{yjh}^2 and S_{xjh}^2 be strata variances and S_{yxjh} is the strata covariance between the study and auxiliary variables for the j^{th} characteristics in the h^{th} stratum. $b_{jh} = \frac{s_{yxjh}}{s_{xjh}^2}$ and $\beta_{jh} = \frac{S_{yxjh}}{S_{xjh}^2}$ are sample and population regression coefficients and $W_h = N_h / N$ be strata weights.

Using estimator for non-response in the study variable by [13], the stratum mean $\bar{Y}_{jh}$ for the j^{th} characteristics in the h^{th} stratum is estimated using

$$\bar{y}_{jh} = \frac{n_{h1}\bar{y}_{jh1} + n_{h2}\bar{y}_{jh2(r_h)}}{n_h} \tag{2}$$

where $\bar{y}_{jh1}$ is the sample mean of the n_{h1} respondents and $\bar{y}_{jh2(r_h)}$ is the sample mean of the r_h unit of non-respondents measured at the second attempt.

Now, considering the separate regression estimator,

$$\bar{y}_{j,lrs} = \sum_{h=1}^{L} W_h \bar{y}_{j,lrh} \tag{3}$$

where $\bar{y}_{j,lrh} = \bar{y}_{jh} + b_{jh}\left(\bar{X}_{jh} - \bar{x}_{jh}\right)$ (4)

the mean square error (MSE) of $\bar{y}_{j,lrs}$ is given by

$$MSE(\bar{y}_{j,lrs}) = \sum_{h=1}^{L} W_h^2 \left[\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\left[S_{yjh}^2 + \beta_{jh}^2 S_{xjh}^2 - 2\beta_{jh} S_{yxjh}\right] + \frac{W_{h2}^2 S_{yjh2}^2}{r_h} - \frac{W_{h2} S_{yjh2}^2}{n_h}\right] \tag{5}$$

Ignoring finite population correction (FPC) in equation 5, we obtain,

$$MSE(\bar{y}_{j,lrs}) = \sum_{h=1}^{L} \frac{W_h^2\left(\xi'_{jh} - W_{h2}S_{yjh2}^2\right)}{n_h} + \sum_{h=1}^{L} \frac{W_h^2 W_{h2}^2 S_{yjh2}^2}{r_h} \tag{6}$$

where $\xi'_{jh} = S_{yjh}^2 + \beta_{jh}^2 S_{xjh}^2 - 2\beta_{jh} S_{yxjh}$ and $W_{h2} = N_{h2}/N_h$ is the proportion of non-respondent in the h^{th} stratum;

$\bar{Y}_{jh} = N_h^{-1}\sum_{i=1}^{N_h} y_{jhi}$: Population mean of the study variable for the h^{th} stratum;

$\bar{X}_{jh} = N_h^{-1}\sum_{i=1}^{N_h} x_{jhi}$: Population mean of the auxiliary variable for the h^{th} stratum; $S_{yjh}^2 = (N_h - 1)^{-1}\sum_{i=1}^{N_h}\left(y_{jhi} - \bar{Y}_{jh}\right)^2$: Population variance of the study variable for the h^{th} stratum; $S_{xjh}^2 = (N_h - 1)^{-1}\sum_{i=1}^{N_h}\left(x_{jhi} - \bar{X}_{jh}\right)^2$: Population variance of the auxiliary variable for the h^{th} stratum.

$S_{yxjh} = (N_h - 1)^{-1}\sum_{i=1}^{N_h}\left(y_{jhi} - \bar{Y}_{jh}\right)\left(x_{jhi} - \bar{X}_{jh}\right)$: Population covariance between the study and auxiliary variables for the h^{th} stratum.

$S_{yjh2}^2 = (N_{h2} - 1)^{-1}\sum_{i=1}^{N_{h2}}\left(y_{jhi} - \bar{Y}_{jh2}\right)^2$: Population variance of the study variable for non-response group in the h^{th} stratum.

$\bar{y}_{jh1} = n_{h1}^{-1}\sum_{i=1}^{n_{h1}} y_{jhi}$: Sample mean of the study variable of units respond on first call in the h^{th} stratum.

$\bar{y}_{jh2(r_h)} = r_h^{-1}\sum_{i=1}^{r_h} y_{jhi}$: Sample mean of the study variable of units respond on second call in the h^{th} stratum.

Since different characteristics are measured in different units, we need an estimate free of unit measurement. Hence the squared coefficient of variation is used instead of the MSE.

$$CV(\bar{y}_{j,lrs}) = Z_j = \sqrt{\frac{MSE(\bar{y}_{j,lrs})}{\bar{Y}_j^2}} = \sqrt{\sum_{h=1}^{L}\frac{\Psi'_{jh}}{n_h} + \sum_{h=1}^{L}\frac{\Phi'_{jh}}{r_h}} \tag{7}$$

$$\left.\begin{aligned} \Psi'_{jh} &= \bar{Y}_j^{-2} W_h^2\left(\xi'_{jh} - W_{h2}S_{yjh2}^2\right) \\ \Phi'_{jh} &= \bar{Y}_j^{-2} W_h^2 W_{h2}^2 S_{yjh2}^2 \end{aligned}\right\} \tag{8}$$

where Z_j is the CV of the j^{th} characteristics and $\bar{Y}_j = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{jhi}$ is the population mean of the j^{th} characteristics.

3. Gamma Cost Function with Non-Response

If there is non-response the linear cost function used in stratified random sampling is given by:

$$C = \sum_{h=1}^{L} c_h = \sum_{h=1}^{L} c_{h0} n_h + \sum_{h=1}^{L} c_{h1} n_{h1} + \sum_{h=1}^{L} c_{h2} n_{h2} \tag{9}$$

where c_h is the total cost in the h^{th} stratum, c_{h0} is the per unit cost of selecting the n_h units or making the first attempt, $c_{h1} = \sum_{j=1}^{p} c_{jh1}$ is the per unit cost for measuring and processing (enumerating) the results of all the p characteristics on the n_{h1} selected units from respondent group in the h^{th} stratum in the first attempt and $c_{h2} = \sum_{j=1}^{p} c_{jh2}$ is the per unit cost for measuring and processing (enumerating) the results of all the p characteristics on the r_h units selected from the non-respondents group in the h^{th} stratum. As n_{h1} is not known until the first attempt is made, the quantity $W_{h1}n_h$ may be used as its expected value. So the expected cost will be

$$\hat{C} = \sum_{h=1}^{L}\left(c_{h0} + c_{h1}W_{h1}\right)n_h + \sum_{h=1}^{L} c_{h2} r_h \tag{10}$$

In practical situations, measurement unit cost and travel cost within strata are important factors in survey cost. If the travel cost is significant then the cost function remains no more linear. The non-linear cost function including measurement unit cost and travel cost within strata to approach the units selected in the sample is a good approximation to the actual budget of the survey. The distance between k randomly scattered (dispersed) destinations within a region is asymptotically proportional to $\sqrt{k}$ for large k [14]. Hence, the cost of visiting the n_h selected units in the h^{th} stratum scattered all over the stratum can be taken as $t_h\sqrt{n_h}, h = 1,2,..., L$, approximately, where t_h is the travel cost per unit in the h^{th} stratum. As the travel cost is proportional to the distance traveled it will also be proportional to $\sqrt{n_h}$. The quadratic cost function considered by [1] is:

$$\hat{C} = \sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + \sum_{h=1}^{L} c_{h2}r_h + \sum_{h=1}^{L} t_{h0}\sqrt{n_h} + \sum_{h=1}^{L} t_{h2}\sqrt{r_h} \tag{11}$$

where t_{h0} is the per unit traveling cost at the first attempt, t_{h2} is the travel cost for visiting the non-respondent unit within the h^{th} stratum and r_h is the sub-sample from the non-respondent.

3.1. Proposed Gamma Cost Function

Reward given to a respondent may reflect the preciousness of the respondent stand point, availability, approachability and time; labour cost is the multiple of time units consumed to obtain data from the respondents [15].

Now, if time taken to obtain data from sampling unit's that is, the selected units in the first attempt and non-respondent in a more careful second attempt follows exponential distributions [16, 17] with rates λ and λ^*. The sum of independent identically distributed exponential random variables will follow gamma distributions with parameters (n_h, λ) and (r_h, λ^*) respectively. If we extend this idea to the whole sample from all strata, then the gamma functions will have parameters $\left(\sum_{h=1}^{L} n_h, \lambda\right)$ and $\left(\sum_{h=1}^{L} r_h, \lambda^*\right)$ respectively where the probability distribution functions for time used is approximately exponentially distributed.

Consider the proposed gamma cost function in the presence of non-response given by:

$$C = \sum_{h=1}^{L}(c_{h0} + c'_{h1}W_{h1})n_h + \sum_{h=1}^{L} c'_{h2}r_h + \sum_{h=1}^{L} t_{h0}\sqrt{n_h} + \sum_{h=1}^{L} t_{h2}\sqrt{r_h} + \alpha\int_0^\infty \lambda e^{-\lambda t}\frac{(\lambda t)^{\sum_{h=1}^{L} n_h - 1}}{(\sum_{h=1}^{L} n_h - 1)!}dt \tag{12}$$

$$+ \omega\int_0^\infty \lambda^* e^{-\lambda^* t^*}\frac{(\lambda^* t^*)^{\sum_{h=1}^{L} r_h - 1}}{(\sum_{h=1}^{L} r_h - 1)!}dt^*$$

here c_{h0} is the per unit cost of selecting the n_h units or making the first attempt, $c'_{h1} = c_{h1} + u_{h1}$ is the measurement and processing unit cost with reward paid to respondent in strata h ; $c'_{h2} = c_{h2} + u_{h2}$ is the measurement and processing unit cost with reward paid to non-respondent in the second attempt, t and t^* are the times taken to obtain data, t_{h0} is the per unit traveling cost at the first attempt, t_{h2} is the travel cost for visiting the non-respondent unit within the h^{th} stratum and r_h are the sub-samples from non-respondent and n_h are the sample sizes from each stratum. $W_{h1} = N_{h1}/N_h$ is the weight of the respondents in the h^{th} stratum. If equation 12 is considered over all characteristics that is, $j = 1,2,..., p$, let

$$\varphi = \int_0^\infty \lambda e^{-\lambda t}\frac{(\lambda t)^{\sum_{h=1}^{L} n_{jh} - 1}}{(\sum_{h=1}^{L} n_{jh} - 1)!}dt; \quad \varphi^* = \int_0^\infty \lambda^* e^{-\lambda^* t^*}\frac{(\lambda^* t^*)^{\sum_{h=1}^{L} r_{jh} - 1}}{(\sum_{h=1}^{L} r_{jh} - 1)!}dt^*$$

then equation 12 becomes,

$$C = \sum_{h=1}^{L}(c_{h0} + c'_{h1}W_{h1})n_{jh} + \sum_{h=1}^{L} c'_{h2}r_{jh} + \sum_{h=1}^{L} t_{h0}\sqrt{n_{jh}} + \sum_{h=1}^{L} t_{h2}\sqrt{r_{jh}} + \alpha\varphi + \omega\varphi^* \tag{13}$$

let α be the cost of unit time in the first attempt and ω be the cost of unit time on non-respondent in the second attempt. The two gamma functions are effects on labour cost. Estimates of the gamma functions φ and φ^* are replaced with aggregate expected time used in obtaining information about the j^{th} characteristic, that is

$$\left.\begin{aligned}\alpha\sum_{h=1}^{L}E(T_h)=\alpha\sum_{h=1}^{L}\left[\int_0^\infty \lambda e^{-\lambda t}\,\frac{(\lambda t)^{n_{jh}-1}}{(n_{jh}-1)!}\,dt\right]=\alpha\sum_{h=1}^{L}\frac{n_{jh}}{\lambda}\\ \omega\sum_{h=1}^{L}E(T_h^*)=\omega\sum_{h=1}^{L}\left[\int_0^\infty \lambda^* e^{-\lambda^* t^*}\,\frac{(\lambda^* t^*)^{r_{jh}-1}}{(r_{jh}-1)!}\,dt^*\right]=\omega\sum_{h=1}^{L}\frac{r_{jh}}{\lambda^*}\end{aligned}\right\} \qquad (14)$$

from the above formulations the proposed expected gamma cost constraint is given as:

$$\sum_{h=1}^{L}(c_{h0}+c'_{h1}W_{h1})n_{jh}+\sum_{h=1}^{L}c'_{h2}r_{jh}+\sum_{h=1}^{L}t_{h0}\sqrt{n_{jh}}+\sum_{h=1}^{L}t_{h2}\sqrt{r_{jh}}+\alpha\sum_{h=1}^{L}\frac{n_{jh}}{\lambda}+\omega\sum_{h=1}^{L}\frac{r_{jh}}{\lambda^*}\le\hat{C} \qquad (15)$$

4. Optimization Techniques with Gamma Cost

Let Z_j^* be the objective function values of Z_j under the individual optimum allocation for the j^{th} characteristics obtained by solving the formulated MOINLPP separately for each $j=1,2,...,p$.

$$\left.\begin{aligned}&Minimize\ \ Z_j\,(j=1,2,...,p)\\ &subject\ \ to\\ &\sum_{h=1}^{L}(c_{h0}+c'_{h1}W_{h1})n_{jh}+\sum_{h=1}^{L}c'_{h2}r_{jh}+\sum_{h=1}^{L}t_{h0}\sqrt{n_{jh}}+\sum_{h=1}^{L}t_{h2}\sqrt{r_{jh}}+\alpha\sum_{h=1}^{L}\frac{n_{jh}}{\lambda}+\omega\sum_{h=1}^{L}\frac{r_{jh}}{\lambda^*}\le C_0\\ &2\le n_{jh}\le N_h\\ &2\le r_{jh}\le \hat{n}_{h2}\\ &n_{jh}\ and\ r_{jh}\ are\ \text{integers and } n_{jh}\in\mathbf{F},\ \forall\ h=1,2,...,L,\ \text{and}\ j=1,2,...,p\end{aligned}\right\} \qquad (16)$$

where Z_j is as defined in equation 7 subject to the cost constraints given in equation 15 and $\hat{n}_{h2}$ is the estimated value of the n_{h2}, obviously $\hat{n}_{h2}=W_{h2}n_{jh}$ and the optimum allocation of one characteristic may be good for all characteristics, the solution of the above formulation for each $j=1,2,...,p$ gives the individual optimum allocations n_{jh}^* and r_{jh}^* for the j^{th} characteristics.

4.1.Value Function Technique (VFT)

The problem given in equation 16 expressed by [8, 18] using the VFT isgiven as

$$\left.\begin{aligned}&Minimize\ \ \left(\sum_{j=1}^{p}\theta_j Z_j\right)\\ &subject\ \ to\\ &\sum_{h=1}^{L}(c_{h0}+c'_{h1}W_{h1})n_{hc}+\sum_{h=1}^{L}c'_{h2}r_{hc}+\sum_{h=1}^{L}t_{h0}\sqrt{n_{hc}}+\sum_{h=1}^{L}t_{h2}\sqrt{r_{hc}}+\alpha\sum_{h=1}^{L}\frac{n_{hc}}{\lambda}+\omega\sum_{h=1}^{L}\frac{r_{hc}}{\lambda^*}\le\hat{C}_0\\ &2\le n_{hc}\le N_h\\ &2\le r_{hc}\le\hat{n}_{h2}\\ &\underline{\hat{n}}_c\in\mathbf{F}\\ &n_{hc}\ and\ r_{hc}\ are\ \text{integers}\ \forall\ h=1,2,...,L,\ \text{and}\ j=1,2,...,p\end{aligned}\right\} \qquad (17)$$

$$\theta_j=\frac{\sum_{h=1}^{L}S_{jh}^2}{\sum_{j=1}^{p}\sum_{h=1}^{L}S_{jh}^2},\ j=1,2,...,p \quad \text{where} \quad \sum_{j=1}^{p}\theta_j=1,\ \ \theta_j\ge 0 \qquad (18)$$

θ_j are the weights according to the relative importance of the characteristics when complete information is available.

4.2. Extended Lexicographic Goal Programming (ELGP)

The distinct feature of preemptive or lexicographic goal programming is that the objectives can be divided into different priority classes, and it is assumed that no two goals have equal priorities. This is known as the non-Archimedean goal programming (GP).

If $\hat{Z}_j$ denote the values of Z_j for the compromise allocation n_{hc} and r_{hc} then obviously $\hat{Z}_j\ge Z_j$ and $\hat{Z}_j-Z_j\ge 0;\ j=1,2,...,p$ will give an increase in Z_j due to not using the individual optimum allocation of the j^{th}

characteristics. We find n_{hc} and r_{hc} such that for the j^{th} characteristics, the increase in the value of Z_j for each j due to the use of compromise allocation is less or equal to d_j^+, where $d_j^+ \geq 0$, $j = 1,2,...,p$ are the goal variables or positive deviational variables. To achieve this we must have $\hat{Z}_j - Z_j^* \leq d_j^+$ or $\hat{Z}_j - d_j^+ \leq Z_j^*$. The total increase in the value of Z_j for not using the individual optimum allocation is given by $\sum_{j=1}^{p} d_j^+$. Using the above formulations to solve multi-objective allocation problem given in 16, the ELGP will have the following model

$$\left.\begin{aligned}
&\textit{Minimize}\ \ (1-\rho)D + \rho\sum_{j=1}^{p} d_j^+ \\
&\textit{subject to} \\
&d_j^+ \leq D \\
&\hat{Z}_j - d_j^+ \leq Z_j^* \\
&\sum_{h=1}^{L}(c_{h0} + c'_{h1}W_{h1})n_{hc} + \sum_{h=1}^{L} c'_{h2} r_{hc} + \sum_{h=1}^{L} t_{h0}\sqrt{n_{hc}} + \sum_{h=1}^{L} t_{h2}\sqrt{r_{hc}} + \alpha\sum_{h=1}^{L}\frac{n_{hc}}{\lambda} + \omega\sum_{h=1}^{L}\frac{r_{hc}}{\lambda^*} \leq C_0 \\
&2 \leq n_{hc} \leq N_h \\
&2 \leq r_{hc} \leq \hat{n}_{h2} \\
&d_j^+ \geq 0 \\
&\rho \in [0,1] \\
&n_{hc}\ \textit{and}\ r_{hc}\ \textit{are integers and}\ n_{hc} \in \mathbf{F},\ \ \forall\ h = 1,2,...,L,\ \text{and}\ j = 1,2,...,p
\end{aligned}\right\} \quad (19)$$

where D is the maximum deviation from utility and ρ is the parameter that weights the value attached to the minimization of the weighted sum of unwanted deviation variables and uses arbitrary value(s) on the interval $0 \leq \rho \leq 1$. For $\rho = 0$, we have the MINMAX GP achievement function, for $\rho = 1$ the GP achievement function and for other values of parameter ρ belonging to the interval $[0,1]$ are intermediate solutions provided by the weighted combination of these two GP options [19].

5. Numerical Illustration

The data used in the study was obtained from the agricultural census conducted by National Agricultural Statistics Services, USDA, Washington DC (http://www.agcensus.usda.gov). The census results are for 2002 and 2007 as presented in[20] for linear cost function. For the purpose of illustration, the data of $N = 99$ counties in Iowa State are divided into four strata, with respect to two characteristics, the quantity of Corn harvested and the quantity of Soybean harvested. $Y_1 =$ Corn harvested in 2007; $Y_2 =$ Soybean harvested in 2007 $X_1 =$ Corn harvested in 2002; $X_2 =$ Soybean harvested in 2002, where Y_1, Y_2 are study variables and X_1, X_2 are auxiliary variables respectively. The estimate of the population means for the study variable Y_1 and Y_2 is given as $\bar{Y}_1 = 22698622.750$ and $\bar{Y}_2 = 4306561.045$ respectively, with 27, 30, 27, and 20 percentage of non-response in each stratum. Let the reward u_{h1} and u_{h2} given for a unit time on the respondent and non-respondent be 2 units. Let α and ω which are the cost for a unit time of labour be (say 20, 30, 40, etc per hour per individual) obviously the results will not be the same if different cost is chosen in another study. Let also λ and $\lambda^* = \dfrac{1}{avg.\ time}$ taken to obtain data from the selected units in the first attempt and non-respondent on the second call be (say 15min, 20min, 25min etc on the average from an individual) with $j = 1, 2$ and $h = 1, 2, 3, 4$ and the total budget for the survey is assumed to be $C_0 = 800$ units.

Table 1: Summary Statistics of Data

h	N_h	W_h	W_{h1}	W_{h2}	c_{h0}	c_{h1}	c_{h2}	S^2_{x1h}	S^2_{x2h}	S^2_{y1h}
1	22	0.2222	0.73	0.27	1	2	4	0.253	0.253	5.76×10^{13}
2	40	0.4040	0.70	0.30	1	3	5	0.251	0.255	1.21×10^{14}
3	24	0.2424	0.73	0.27	1	4	6	0.261	0.259	5.57×10^{13}
4	13	0.1313	0.80	0.20	1	5	7	0.231	0.269	7.08×10^{13}

S_{y2h}^2	S_{yx1h}	S_{yx2h}	S_{y1h2}^2	S_{y2h2}^2	t_{h0}	t_{h2}
1.67×10^{12}	2622322.766	550824.4870	7.80×10^{13}	1.48×10^{12}	0.5	2
2.50×10^{12}	4340658.415	590425.8192	1.22×10^{14}	2.807×10^{12}	0.5	2.5
3.58×10^{12}	3072607.891	745074.7953	2.63×10^{13}	3.02×10^{12}	0.5	3
4.44×10^{12}	2664563.750	802484.3100	4.01×10^{13}	6.28×10^{11}	0.5	3.5

ξ'_{1h}	ξ'_{2h}	Ψ'_{1h}	Ψ'_{2h}	Φ'_{1h}	Φ'_{2h}
3.041985498×10^{13}	4.707604131×10^{11}	0.000897	0.000189	0.000545	0.000287
4.59349981×10^{13}	1.132943158×10^{12}	0.002957	0.002560	0.003478	0.002223
1.952789559×10^{13}	1.436616021×10^{12}	0.001417	0.001968	0.000219	0.000697
4.006450226×10^{13}	2.046020094×10^{12}	0.001072	0.001785	0.000054	0.000023

5.1. Individual Optimum Allocation with Gamma Cost

The individual optimum allocation for characteristics Y_1 formulated as a MOINLPP is given as

$$Min\ \ Z_1=\left(\frac{0.000897}{n_{11}}+\frac{0.002957}{n_{12}}+\frac{0.001417}{n_{13}}+\frac{0.001072}{n_{14}}+\frac{0.000545}{r_{11}}+\frac{0.003478}{r_{12}}+\frac{0.000219}{r_{13}}+\frac{0.000054}{r_{14}}\right)^{\frac{1}{2}}$$

subject to

$$[8.92\quad 9.50\quad 10.38\quad 11.60][n_{11}\quad n_{12}\quad n_{13}\quad n_{14}]^T+[11\quad 12\quad 13\quad 14][r_{11}\quad r_{12}\quad r_{13}\quad r_{14}]^T$$

$$+[0.5\quad 0.5\quad 0.5\quad 0.5]\left[\sqrt{n_{11}}\quad \sqrt{n_{12}}\quad \sqrt{n_{13}}\quad \sqrt{n_{14}}\right]^T+[2\quad 2.5\quad 3\quad 3.5]\left[\sqrt{r_{11}}\quad \sqrt{r_{12}}\quad \sqrt{r_{13}}\quad \sqrt{r_{14}}\right]^T\le 800$$

$$2\le n_{1h}\le N_h$$

$$2\le r_{1h}\le \hat{n}_{h2}$$

$$\text{where}\quad n_{1h}\ and\ \ r_{1h}\ \text{are integers},\ \ h=1,2,3,4$$

The individual optimum allocation for characteristics Y_2 formulated as a MOINLPP is given as

$$Min\ \ Z_2=\left(\frac{0.000189}{n_{21}}+\frac{0.002560}{n_{22}}+\frac{0.001968}{n_{23}}+\frac{0.001785}{n_{24}}+\frac{0.000287}{r_{21}}+\frac{0.002223}{r_{22}}+\frac{0.000697}{r_{23}}+\frac{0.000023}{r_{24}}\right)^{\frac{1}{2}}$$

subject to

$$[8.92\quad 9.50\quad 10.38\quad 11.60][n_{21}\quad n_{22}\quad n_{23}\quad n_{24}]^T+[11\quad 12\quad 13\quad 14][r_{21}\quad r_{22}\quad r_{23}\quad r_{24}]^T$$

$$+[0.5\quad 0.5\quad 0.5\quad 0.5]\left[\sqrt{n_{21}}\quad \sqrt{n_{22}}\quad \sqrt{n_{23}}\quad \sqrt{n_{24}}\right]^T+[2\quad 2.5\quad 3\quad 3.5]\left[\sqrt{r_{21}}\quad \sqrt{r_{22}}\quad \sqrt{r_{23}}\quad \sqrt{r_{24}}\right]^T\le 800$$

$$2\le n_{2h}\le 22$$

$$2\le r_{2h}\le \hat{n}_{h2}$$

$$\text{where}\quad n_{2h}\ and\ \ r_{2h}\ \text{are integers},\ \ h=1,2,3,4$$

From the above models, we obtain individual optimal values Z_1^* and Z_2^* as coefficient of variation for two characteristics*j* = 1 and *j* = 2.

5.2 VFT with Gamma Cost

Optimum compromise allocation of samples and subsamples using VFT is given as

$$Min \left[0.9616\left(\frac{0.000897}{n_{1c}}+\frac{0.002957}{n_{2c}}+\frac{0.001417}{n_{3c}}+\frac{0.001072}{n_{4c}}+\frac{0.000545}{r_{1c}}+\frac{0.003478}{r_{2c}}+\frac{0.000219}{r_{3c}}+\frac{0.000054}{r_{4c}}\right)^{\frac{1}{2}} +0.0384\left(\frac{0.000189}{n_{1c}}+\frac{0.002560}{n_{2c}}+\frac{0.001968}{n_{3c}}+\frac{0.001785}{n_{4c}}+\frac{0.000287}{r_{1c}}+\frac{0.002223}{r_{2c}}+\frac{0.000697}{r_{3c}}+\frac{0.000023}{r_{4c}}\right)^{\frac{1}{2}}\right]$$

subject to

$$[8.92 \quad 9.50 \quad 10.38 \quad 11.60]\,[n_{1c} \quad n_{2c} \quad n_{3c} \quad n_{4c}]^T + [11 \quad 12 \quad 13 \quad 14]\,[r_{1c} \quad r_{2c} \quad r_{3c} \quad r_{4c}]^T$$

$$+[0.5 \quad 0.5 \quad 0.5 \quad 0.5]\,[\sqrt{n_{1c}} \quad \sqrt{n_{2c}} \quad \sqrt{n_{3c}} \quad \sqrt{n_{4c}}]^T + [2 \quad 2.5 \quad 3 \quad 3.5]\,[\sqrt{r_{1c}} \quad \sqrt{r_{2c}} \quad \sqrt{r_{3c}} \quad \sqrt{r_{4c}}]^T \le 800$$

$$2 \le n_{hc} \le N_h$$

$$2 \le r_{hc} \le \hat{n}_{h2}$$

where n_{hc} *and* r_{hc} are integers, $h = 1,2,3,4$

5.3. ELGP with Gamma Cost

The optimum compromise allocation of the samples and subsamples using ELGP while selecting arbitrary value(s) say $\rho = 0.7$ and changing ρ different results are expected, hence the following model is established

$$Min \;\; 0.3D + 0.7(d_1^+ + d_2^+)$$

subject to

$$d_1^+ \le D$$

$$d_2^+ \le D$$

$$\left(\frac{0.000897}{n_{1c}}+\frac{0.002957}{n_{2c}}+\frac{0.001417}{n_{3c}}+\frac{0.001072}{n_{4c}}+\frac{0.000545}{r_{1c}}+\frac{0.003478}{r_{2c}}+\frac{0.000219}{r_{3c}}+\frac{0.000054}{r_{4c}}\right)^{\frac{1}{2}} - d_1^+ \le Z_1^*$$

$$\left(\frac{0.000189}{n_{1c}}+\frac{0.002560}{n_{2c}}+\frac{0.001968}{n_{3c}}+\frac{0.001785}{n_{4c}}+\frac{0.000287}{r_{1c}}+\frac{0.002223}{r_{2c}}+\frac{0.000697}{r_{3c}}+\frac{0.000023}{r_{4c}}\right)^{\frac{1}{2}} - d_2^+ \le Z_2^*$$

$$[8.92 \quad 9.50 \quad 10.38 \quad 11.60]\,[n_{1c} \quad n_{2c} \quad n_{3c} \quad n_{4c}]^T + [11 \quad 12 \quad 13 \quad 14]\,[r_{1c} \quad r_{2c} \quad r_{3c} \quad r_{4c}]^T$$

$$+[0.5 \quad 0.5 \quad 0.5 \quad 0.5]\,[\sqrt{n_{1c}} \quad \sqrt{n_{2c}} \quad \sqrt{n_{3c}} \quad \sqrt{n_{4c}}]^T + [2 \quad 2.5 \quad 3 \quad 3.5]\,[\sqrt{r_{1c}} \quad \sqrt{r_{2c}} \quad \sqrt{r_{3c}} \quad \sqrt{r_{4c}}]^T \le 800$$

$$2 \le n_{hc} \le N_h$$

$$2 \le r_{hc} \le \hat{n}_{h2}$$

$$d_1^+, d_2^+ \ge 0$$

where n_{hc} *and* r_{hc} are integers, $h = 1,2,3,4$

6. Results and Discussion

Table 2: Individual Optimum Allocations

	n_{j1}^*	n_{j2}^*	n_{j3}^*	n_{j4}^*	r_{j1}^*	r_{j2}^*	r_{j3}^*	r_{j4}^*	Used Cost	Z_1^*	Z_2^*	Trace $= Z_1^* + Z_2^*$
Y_1	15	27	8	10	4	8	2	2	795.1334	0.03407556	0.03564043	0.06971599
Y_2	8	24	15	12	2	7	4	2	793.1848	0.03523549	0.03251251	0.06894350

Table 3: Compromise Allocations

Allocations	n_{1c}^*	n_{2c}^*	n_{3c}^*	n_{4c}^*	r_{1c}^*	r_{2c}^*	r_{3c}^*	r_{4c}^*	Used Cost	$\hat{Z}_1$	$\hat{Z}_2$	Trace $=\hat{Z}_1+\hat{Z}_2$
VFT	12	27	12	10	3	8	3	2	799.4709	0.03409499	0.03444715	0.06854214
ELGP (0.7)	12	24	14	10	3	7	3	2	793.2179	0.03442232	0.03314452	0.06756684

The percentage relative efficiency (PRE) of a compromise allocation with respect to the individual optimum allocation is given in [12] as

$$PRE = \frac{T_I}{T_C} \times 100$$

where T_I is the value of the trace using individual optimum allocation of one characteristic for both characteristics and T_C is the value of the trace using the compromise allocations.

Table 4: PREs as Compared to Individual Optimum Allocations

Allocations	PREs	
Ind. Opt. Allocation	$Y_1(100)$	$Y_2(100)$
VFT	101.71	100.59
ELGP (0.7)	103.18	102.04

The data for four strata and two characteristics is presented in Table 1. The individual optimum allocation of one characteristic used for both characteristics is given in Table 2 and compromise allocations using VFT and ELGP is given in Table 3. The traces are sum of diagonal element of the variance-covariance matrices, which are the variances of estimate of finite population means of different characteristics. Since the characteristics under study are assumed to be independent the covariances are zero which provides the basis for performance comparison using PRE. Table 2 and Table 3 also show that the corresponding cost does not exceed the available cost. Table 4 gives the percentage relative efficiencies of the VFT and ELGP $(\rho = 0.7)$ as compared to the individual optimum allocation of both characteristics and shows that the compromise allocations outperform the individual optimum allocations of each characteristic. With the results obtained we deduce that there is no high correlation between the characteristics of interest since the compromise allocations provides a better allocation for both characteristics.

7. Conclusion

The proposed gamma function approximates the linear and quadratic cost functions in the presence of non-response. The problem of optimum allocation in multivariate stratified sampling has been solved using some compromise allocations based on the data used for a fixed cost. Furthermore, we can attest that the VFT and ELGP approach based on the comparison made in section 6 always secure a feasible solution and provides better results as compared to the individual optimum allocation approach from the perspective of efficiency.

References

[1] Shafiullah, Khan, M.F. and Ali, I. (2014) Fuzzy Geometric Programming in Multivariate Stratified Sample Surveys in Presence of Non-Response with Quadratic Cost Function. *American Journal of Operations Research*, 4; 173-188.

[2] Cochran, W.G. (1977). *Sampling Techniques*. 3rd Edition, John Wiley and Sons, NewYork.

[3] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984) Sampling Theory of Surveys with Applications. Iowa State University Press, Ames and Indian Society of Agricultural Statistics, New Delhi.

[4] Dalenius, T. (1957). *Sampling in Sweden: Contributions to the Methods and Theories of Sample Survey Practice*, Almqvist and Wiksell, Stockholm.

[5] Aoyama, H. (1963). Stratified Random Sampling with Optimum Allocation for Multivariate Populations. *Annals of the Institute of Statistical Mathematics*, 14; 251-258.

[6] Chatterjee, S. (1968). Multivariate Stratified Surveys. *Journal American Statistical Association*, 63; 530-534.

[7] Khan, M.G.M., Ahsan, M.J., and Jahan, N. (1997). Compromise Allocation in Multivariate Stratified Sampling: An Integer Solution. *Naval Research Logistics*44, 69-79.

[8] Khan, M.G.M., Khan, E.A. and Ahsan, M.J. (2003). An Optimal Multivariate Stratified Sampling Design Using Dynamic Programming. *Australia and New Zealand Journal of Statistics*, 45(1); 107-113.

[9] Díaz-García, J. A. and Cortez L.U. (2006). Optimum Allocation in Multivariate Stratified Random Sampling: Multi-Objective Programming. *Comunicación Técnica*, No. I-06-07/28-03-206 (PE/CIMAT).

[10] Díaz-García, J.A. and Cortez, L.U. (2008). Multi-Objective Optimization for Optimum Allocation in Multivariate StratifiedSampling. *Survey Methodology,* 34(2); 215–222.
[11] Varshney, R., Ahsan, M.J. and Khan, M.G.M. (2011). An Optimum Multivariate Stratified Sampling Design with Non-Response: A Lexicographic Goal Programming Approach. *Journal of Mathematical Modeling and Algorithms*, 65; 291-296.
[12] Ullah, A., Shabbir, J., Hussain, Z. and Al-Zahrani B. (2014). Estimation of Finite Population Mean in Multivariate Stratified Sampling under Cost Function Using Goal Programming. *Journal of Applied Mathematics*, 2014; 1-7.
13] Hansen, M. H., and Hurwitz, W. N. (1946). The Problem of Non-Response in Sample Surveys. *Journal of American Statistical Association,* 41; 517–529.
[14] Beardwood, J., Halton, J. H. and Hammersley, J. M. (1959). The Shortest Path through Many Points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55; 299-327.
[15] Muhammad, Y. S., Shabbir, J., Husain, I. and Abd-el. Moemen M. (2014). Multi-objective Compromise Allocation in Multivariate Stratified Sampling Using Extended Lexicographic Goal Programming with Gamma Cost Function. *Journal of Mathematical Modelling and Algorithms*, 14(3); 225-265.
[16] Hogg, R. V. and Craig, A. T. (1978). *Introduction to Mathematical Statistics,* 4th edition, Macmillan, New York.
[17] Ross, S. M. (2009). *Introduction to Probability and Statistics for Engineers and Scientists,* 4th edition. Elsevier Academic Press, Burlington, USA.
[18] Raghav, Y.S., Ali, I., and Bari, A. (2014). Multi-Objective Nonlinear Programming Problem Approach in Multivariate Stratified Sample Surveys in Case of Non-Response. *Journal of Statistical Computation and Simulation*, 84(1); 22-36.
[19] Romero, C. (2001). Extended Lexicographic Goal Programming: A Unifying Approach. *Omega*, 29(1); 63-71.
[20] Khan, S., Muhammad, Y. S. and Shabbir, J. (2014). Comparison of Modified Extended Lexicographic Technique with Fuzzy and Value Function Techniques Using the Auxiliary Information as Attributes. *International Journal of Business and Social Science,* 5(13);105-117.