Comparison of Modified Extended Lexicographic Technique with Fuzzy and Value Function Techniques Using the Auxiliary Information as Attributes

Saima Khan

Yousaf Shad Muhammad

Javid Shabbir

Department of Statistics
Quaid-I-Azam University
Islamabad

Abstract

In this article, we propose compromise allocations for multivariate stratified random sampling using the auxiliary attributes under non-response. We modified extended lexicographic goal programming technique and compared it with fuzzy goal programming and value function technique. We addressed the problem of compromise allocation when the auxiliary information is in the form of an auxiliary attributes. A comparative study is carried out to find the best compromise allocation by the numerical example.

Keyword: Multivariate stratified sampling; Compromise allocation; Multi-objective programming techniques; Nonlinear cost function; Lexicographic goal programming

1. Introduction

A well established sampling plan plays a significant role to make the results useful, obtained from statistical studies and provides close approximation to the population estimates. A suitably selected sampling plan and samples, representing population, produce more reliable estimates. Stratified random sampling is the most representative of population when the data is heterogeneous.

The important consideration in stratified random sampling is the sample size allocation in each stratum with the criteria either to minimize variance of stratified sample mean for a fixed cost or to minimize cost for the specified variance.

Sampling efficiency depends largely on how the sample size is allocated. Holmberg (2002) addressed the problem of compromise allocation in multivariate stratified sampling by taking into consideration the minimization of sum of variances or coefficients of variation of population parameters and minimization of sum of efficiency loss which may result due to increase in variance because of using the compromise allocation.

It is more logical to consider the minimization of coefficient of variation (CV) instead of sampling variance as an objective because coefficient of variation is a dimensionless number. For practical implementation, constraint for integer values of the sample sizes are implemented because rounding off the non integer sample sizes to their nearest integer values may produce infeasible solutions.

Non-response is the phenomenon in which required information is not obtained from persons selected in the sample. When information on the auxiliary variable is known in the presence of non response, the problem of estimation of population mean $\bar{Y}$ of the study variable Y has been discussed by Cochran (1977), Rao (1986), Khare and Srivastava (1997), Singh and Kumar (2008) and Khan et al.(2010). Methods for solving the problem of optimum allocation in multivariate stratified sampling under non response are discussed by Varshney et al. (2011), Varshney et al. (2012), Gupta et al.(2012), Haseen et al. (2012) and Raghav et al. (2012)

It is well known that use of the auxiliary information increases the precision of estimators by taking the advantages of correlation between the study variable and the auxiliary variable. Ratio, product and regression estimators are good examples in this context. In many situations, the auxiliary information is available in the form of qualitative information.

For illustration, consider the following examples: When the study variable (Y) is the student's grade point average, Ψ1 can be defined as the number of hours spent on studying (low vs high) and Ψ2 can be defined as the use of library facilities (yes or no). The study variable (Y) can be the product of wheat crop in a certain region, the auxiliary attribute Ψ1 can be the irrigation status (irrigated vs. non-irrigated) and Ψ2 can be the status of land ownership (rich vs. poor).

When information on the auxiliary attributes is known in the presence of non response, the problem of estimation of population mean $\bar{Y}$of the study variable Y has been discussed by Saghir and Shabbir (2012). To best of our knowledge the problem of compromise allocation in multivariate stratified sampling using the auxiliary attributes is not found in literature so the proposed study aims to discuss methods for compromise allocation under non-response using the auxiliary attributes by use of existing criterions. Two cases of non-response are considered. By using various multi-objective optimization techniques we find out the best which gives minimized coefficients of variation (CVs) for both cases of non-response in presence of the auxiliary attributes. The application of these mathematical models on our proposed study is illustrated by numerical example using GAMS.

The paper is organized as: First sampling strategy, required for selection of sample and sub-sample under non response in presence of the auxiliary attributes is described and then mathematical form of the multi-objective problem under certain constraints is presented. A modified extended lexicographic technique is described along with other optimization methods used. Finally, it is shown by the numerical illustration which optimization method is best and results are discussed.

2. Sampling Strategies

Consider a population of size N divided into L strata such that $\sum_{h=1}^{L} N_h =$ N. We draw a sample of size n_h from each stratum by SRSWOR such that $\sum_{h=1}^{L} n_h =$ n. Let $\bar{Y}_h$, $S_h$and$W_h = N_h / N$ be the population mean, population variance and known stratum weight of the h^{th} stratum respectively. It is convenient to think of stratum as a single group divided intotwo disjoint groups, one group containing the responding units and the other containing the non-responding units. The number of responding units, number of non-responding units, size of sample in responding units, the size of sample in non-responding units in the h^{th} stratum is denoted byN_{h1}, $N_{h2} = N - N_{h1}$, n_{h1} , $n_{h2} = n - n_{h1}$respectively. Let $r_h = n_{h2} / k_h$ is a sub sample taken from n_{h2} units in the h^{th}stratum where $k_h > 1$ and $1/k_h$ denotes sampling fraction among nonrespondents assuming that all units in the subsample will respond. Let y_{jhi} and ψ_{jhi} be the i^{th}value of the study variable (Y) and the auxiliary attribute (Ψ) respectively having the j^{th} characteristic in the h^{th} stratum. Let Ψ_{jhi} is an artificial dichotomous attributewhich is useful for simplifying the analysis and presentation of data in many situations. An artificial dichotomy in values of quantitative variable is created out by assigning a designated limit due to which two classes will be formed, one possessing the values greater than that cutt off and the other possessing valuesless than that cut off. It can be defined as:

$$\Psi_{jhi} = \begin{cases} 1, \text{ if } i^{th} \text{unit possess attribute} \Psi_j, \text{ for } (j = 1,2,\ldots,p) \text{ in the } h^{th} \text{ stratum} \\ 0, \text{ otherwise} \end{cases}$$

Similarly we can de ineΨ_{jhi} . Let the population and sample proportion of units of auxiliary attributesΨ_j are defined as:

$$\bar{\Psi}_{jh} = \sum_{i=1}^{N_h} \frac{\Psi jhi}{N_h} \text{ and } \bar{\Psi}_{jh} = \sum_{i=1}^{n_h} \frac{\Psi jhi}{n_h}$$

For$(j = 1,2,\ldots,p)$ respectively. Let the variance and covariance of the study variable Y_j and the auxiliary attribute Ψ_j in the h^{th} stratum for the j^{th} (j = 1, 2, ..., p) characteristic is defined as

$$S_{yjh}^2 = \sum_{h=1}^{L} \frac{\left(y_{jhi} - \bar{Y}_{jh}\right)^2}{N_h - 1}, \text{ where } \bar{Y}_{jh} = \sum_{i=1}^{N_h} \frac{y_{jhi}}{N_h}$$

$$S^2_{\Psi_{jh}} = \sum_{h=1}^{L} \frac{\left(\Psi_{jhi} - \bar{\Psi}_{jh}\right)^2}{N_h - 1}, \text{ and } S_{(y\Psi)jh} = \sum_{i=1}^{N_h} \frac{\left(y_{jhi} - \bar{Y}_{jh}\right)\left(\Psi_{jhi} - \bar{\Psi}_{jh}\right)}{N_h - 1}$$

Similarly we can define the variants and covariates under non-response as:

$$S^2_{yjh2} = \sum_{h=1}^{L} \frac{\left(y_{jhi} - \bar{Y}_{jh2}\right)^2}{N_h - 1}, \text{ where } \bar{Y}_{jh2} = \sum_{i=1}^{N_{h2}} \frac{y_{jhi}}{N_{h2}}$$

$$S_{(y\Psi)jh2} = \sum_{i=1}^{N_h} \frac{\left(y_{jhi} - \bar{Y}_{jh2}\right)\left(\Psi_{jhi} - \bar{\Psi}_{jh2}\right)}{N_h - 1}$$

In our study we have considered two situations of non-response.

Case 1

When non-response is in the study variable Y_j . Using Hansen and Hurwitz (1946) methodology, the unbiased estimator of $\bar{Y}_{jh}$in the h^{th}stratum mean for the j^{th} characteristic is given by:

$$\bar{y}'_{jh} = \frac{n_{h1}\bar{y}_{jh1} + n_{h2}\bar{y}_{jh2}}{n_h} \quad (1)$$

Wherey_{jh1} is the sample mean based on n_{h1} respondent units and y_{jh2} is the sample mean based on r_h sub sample units from non respondents.

Case 2

When non-response is in auxiliary attributes and study variables then we define the unbiasedestimator of $\bar{\Psi}_{jh}$for the j^{th} characteristic in the h^{th} stratum mean as:

$$\bar{\Psi}'_{jh} = \frac{n_{h1}\bar{\Psi}_{jh1} + n_{h2}\bar{\Psi}_{jh2}}{n_h} \quad (2)$$

where$\bar{\Psi}_{jh1}$ is the sample mean based on n_{h1} respondent units and $\bar{\Psi}_{jh2}$ is the sample mean based on r_h subsample units from nonrespondents.

We propose optimum allocation using the auxiliary attributes in presence of non response with objective to minimize coefficients of variation under simple cost constraint. Usually the total cost of a sample survey is represented by the function of sample allocations n_h , $h = 1, 2, \ldots, L$. The simple cost function is suitable when the major part of cost is that of taking the measurements on each unit. It can be represented as:

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

where C is the total cost of the survey, c_{h0} denotes the cost in selection of per unit in the h^{th} stratum, $c_{h1} = \sum_{j=1}^{p} c_{jh1}$ is the cost per unit in taking account of responding units n_{h1} and$c_{h2} = \sum_{j=1}^{p} c_{jh2}$ is the cost per unit in taking account of subsamplesr_h from non responding units n_{h2} in the h^{th} stratum. Here, $c_{jh1} and c_{jh2}$ are the costs per unit for obtaining the j^{th} characteristics in first and second attempts respectively in the h^{th}stratum.As$n_{h1} and n_{h2}$ are unknown before the first and second attempt is made therefore their expected values are used as $W_{h1}n_h and W_{h2}n_h$ respectively. The total expected cost $\hat{C}$ of the survey may be given as:

$$\hat{C} = \sum_{h=1}^{L} (c_{h0} + c_{h1}W_{h1})n_h + \sum_{h=1}^{L} c_{h2}r_h \quad (3)$$

Now we discuss the estimator used in our study.

When nonresponse is in the study variable Y_j , the separate regression estimator for multivariate stratified sampling is

$$\bar{y}_{lrs} = \sum_{h=1}^{L} W_h \bar{y}lr_{jh},$$

Where $\bar{y}lr_{jh} = \bar{y}'_{jh} + b_{jh}\left(\bar{\Psi}_{jh} - \bar{\Psi}_{jh}\right)$

Here b_{jh} is the sample regression coefficient.

The Squared CVofy_{lrs} for j_{th} characteristic is:

$$CV_j^2 = \sum_{h=1}^{L} \frac{W_h^2}{n_h \bar{Y}_j^2}\left(S_{yjh}^2 + B_{jh}^2 S_{\Psi jh}^2 - 2B_{jh} S_{(y\Psi)jh}\right) + \sum_{h=1}^{L} \frac{W_h^2}{\bar{Y}_j^2}\left(\frac{W_{h2}}{r_h} - \frac{1}{n_h}\right) W_{h2} S_{yjh2}^2 \qquad (4)$$

where $\beta_{jh} = S_{(y\Psi)jh}/S_{\Psi jh}^2$ is the population regression coefficient.

When nonresponse is in both the study variable Y_j and the auxiliary attribute Ψ_j , the separate regression estimator is given by

$$\bar{y}'_{lrs} = \sum_{h=1}^{L} W_h \bar{y}l\acute{r}_{jh},$$

Where $$\bar{y}l\acute{r}_{jh} = \bar{y}_{jh} + b_{jh}\left(\bar{\Psi}_{jh} - \bar{\Psi}'_{jh}\right)$$

The squared CV of$\bar{y}'_{lrs}$ forj^{th} characteristic is

$$CV_j'^2 = \sum_{h=1}^{L} \frac{W_h^2}{n_h \bar{Y}_j^2}\left(S_{yjh}^2 + B_{jh}^2 S_{\Psi jh}^2 - 2B_{jh} S_{(y\Psi)jh}\right) + \sum_{h=1}^{L} \frac{W_h^2}{\bar{Y}_j^2}\left(\frac{W_{h2}}{r_h} - \frac{1}{n_h}\right) W_{h2}\left(S_{yjh2}^2 + B_{jh}^2 S_{\Psi jh2}^2 - 2B_{jh} S_{(y\Psi)jh2}\right) \qquad (5)$$

3. Formulation of the Problem

The formulation of Multi-objective integer nonlinear programming problem MOINLPP with a simple cost constraint to found the optimum sample and subsample sizes may be expressed as:

Minimize $[Z_1, Z_2, \dots, Zp]$
Subject to
$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \le C_0$
$2 \le n_h \le N_h$ (6)
$2 \le r_h \le \hat{n}_{h2}$
$n_h and r_h are integers; h = 1,2, \dots, L$

where Z_j , $(j = 1,2,\dots,p)$ denotes the CV for j^{th} characteristic which are to be minimized for fixed cost. Using the expression defined in Eq. (4) and Eq. (5), Eq. (6) can be expressed as:

Minimize

$$Z_j = CV_j^2 = \sum_{j=1}^{p}\sum_{h=1}^{L} \frac{W_h^2}{n_h \bar{Y}_j^2}\left(S_{yjh}^2 + B_{jh}^2 S_{\Psi jh}^2 - 2B_{jh} S_{(y\Psi)jh}\right) + \sum_{j=1}^{p}\sum_{h=1}^{L} \frac{W_h^2}{\bar{Y}_j^2}\left(\frac{W_{h2}}{r_h} - \frac{1}{n_h}\right) W_{h2} S_{yjh2}^2$$

Or

$$\acute{Z}_J = \acute{C}V_j^2 = \sum_{j=1}^{p}\sum_{h=1}^{L}\frac{W_h^2}{n_h\bar{Y}_j^2}\left(S_{yjh}^2 + B_{jh}^2 S_{\Psi jh}^2 - 2B_{jh}S_{(y\Psi)jh}\right)$$
$$+\sum_{j=1}^{p}\sum_{h=1}^{L}\frac{W_h^2}{\bar{Y}_j^2}\left(\frac{W_{h2}}{r_h}-\frac{1}{n_h}\right)W_{h2}\left(S_{yjh2}^2 + B_{jh}^2 S_{\Psi jh2}^2 - 2B_{jh}S_{(y\Psi)jh2}\right)$$

$Subjectto$

$$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \le C_0 \quad (7)$$
$$2\le n_h \le N_h$$
$$2\le r_h \le \hat{n}_{h2}$$
$$n_h and r_h are integers; h = 1,2,\ldots,L$$

4. Optimization Methods for Solving the Multi-objective Programming Problem

The various methods proposed to solve the multi-objective programming problem of multivariate stratified sample surveys in the case of non-respondents can be classified according to the available information about the population. We proposed modified extended lexicographic goal programming (MELGP) technique to find compromise allocation in presence of the auxiliary attributes under non-response.

4.1Modified Extended Lexicographic Goal Programming (MELGP)

Goal programming (GP) is the multiple criteria decision making approach. Let us consider our goal program to have$p$ goals, which may be$j = 1,\ldots,p$. We also define $n_h and r_h$ decision variables. These are the factors over which the decision maker(s) have control and define the decision to be made. Each goal has an achieved value, Z_j , on its underlying criterion. $Z_j$is a function of the decision variables. The whole situation may expressed as below:

$$Minimize Z_j = f(n_{jh}, r_{jh})$$
$$Subjectto$$
$$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \le C_0 \quad (8)$$
$$2\le n_h \le N_h$$
$$2\le r_h \le \hat{n}_{h2}$$
$$n_{jh} and r_{jh} are integers and n_{jh} \in F; h = 1,2,\ldots,L$$

Note that in this generic form no assumptions have yet been made about the nature of the decision variables of goals. The decision maker(s) sets a real target level for each goal denoted by Z_j^*(generally an individual optimal of the j^{th} objective). This then leads to the basic formulation of the j^{th} goal:

$$\widehat{Z}_J + d_j^- - d_j^+ = Z_j^*,$$

where$d_j^- and d_j^+$ are negative and positive deviational variables. They are also called goal variables. Sometimes the set of goals are termed as soft constraints. That is, the decision maker(s) desires to optimize each goal but if the goal is not achieved then this does not imply that the solution is infeasible. Goal programming also allows for an addition of a set of linear programming style hard constraints whose violation will make the solution infeasible. These are modeled by adding the condition

$$\underline{\widehat{n}_J} \in F$$

where$F$ is feasible region established by points in decision space.

Finally, the unwanted deviational variables are put into an achievement function whose purpose is to minimize them and ensure that solution is as close as possible to the set of desired goals.

Lexicographic goal programming is termed as pre-emptive goal programming. The lexicographic ordering philosophy is available via the priority structure of the achievement function. All unwanted deviations are minimized at each priority level. The generic form program of compromise allocation can be written as:

$$Minimize\, a = \left[f_1\left(\underline{d_j^-}, \underline{d_j^+}\right), f_2\left(\underline{d_j^-}, \underline{d_j^+}\right), \ldots, f_p\left(\underline{d_j^-}, \underline{d_j^+}\right)\right]$$

$$\begin{aligned} & Subject\, to \\ & \hat{Z}_J + d_j^- - d_j^+ (\leq or \geq) Z_j^*, \\ & \underline{\hat{n}} \in F \\ & n_{jh}\, and\, r_{jh}\, are\, integers\, and\, n_{jh} \in F; h = 1,2, \ldots, L \end{aligned} \quad (9)$$

where $f_1, f_2, \cdots, f_p$ represent priority-wise functions and $\underline{d_j^-}$, $\underline{d_j^+}$ are vectors of unwanted deviations in the respective priority. The other variant of goal programming is Weighted Goal Programming (WGP), which formulate to minimize a composite objective function formed by a weighted sum of unwanted deviational variables. The third is MINMAX (Chebyshev) Goal Programming, which attempts to minimize the maximum deviation from the desired goals.

In most of the cases, the goal programming variant is chosen without justifying the reason for the selection. It then appears as the choice of the goal programming variant is related to the analyst's taste or to the capability of getting solution. However, the selection of the right goal programming variant or mix of variants is a crucial matter if we want the goal programming model to capture the essential features of the reality modeled [21]. Goal programming can be analyzed in terms of utility theory which always maximizes the utility. The utility function described from the given situation may be of any form i.e. linear, non-linear, etc and a certain satisfaction level of aspiration for a particular goal can be set within a feasible space [21]. A goal program becomes equivalent to minimize the weighted discrepancy for a certain aspiration level $\forall j = 1, 2, \ldots, p$ goals within a feasible space. Now, if we consider that negative deviational variable and positive deviational variable have different impact on achievement function in a particular preference sequence.

Let $W_{1j}\, and\, W_{2j}$ represent the weights of normalizing parameter and preferential of negative deviation variable and positive deviation variable on the j^{th} goal respectively, then following formulation is discussed in [21]:

$$Minimize\, a = \sum_{j=1}^{p}\left[f_j\left(\underline{W_{1j} d_j^-}, \underline{W_{2j} d_j^+}\right)\right]$$

$$\begin{aligned} & Subject\, to \\ & \hat{Z}_J + d_j^- - d_j^+ (\leq or \geq) Z_j^*, \\ & \underline{\hat{n}_J} \in F, h = 1,2, \ldots, L \end{aligned} \quad (10)$$

The maximum utility function may subject to deviate from its desired aspiration level. An Archimedean goal programming model has a clear utility interpretation; it implies the maximization of a separable and additive utility function in the p attributes considered [21]. The MINMAX (Chebyshev) structure corresponds to a utility function where the maximum deviation is minimized. This structure is discussed in [21] as:

$$Minimize\, D$$

$$\begin{aligned} & Subject\, to \\ & \left[d_j^-, d_j^+\right] \leq D \\ & \hat{Z}_J + d_j^- - d_j^+ (\leq or \geq) Z_j^* \\ & \underline{\hat{n}_J} \in F \\ & n_{jh}\, and\, r_{jh}\, are\, integers\, and\, n_{jh} \in F; h = 1,2, \ldots, L \end{aligned} \quad (11)$$

where D is maximum deviation from utility. The concept of extended goal programming, the utility maximization of the Archimedean and MINMAX (Chebyshev) goal programming models, can be generalized as:

$$Minimize(1-\rho)D + \rho\sum_{j=1}^{p}\left[f_j\left(\underline{W_{1j}d_j^-}, \underline{W_{2j}d_j^+}\right)\right]$$

$$Subjectto$$

$$\left[\left(\underline{W_{1j}d_j^-}, \underline{W_{2j}d_j^+}\right)\right] \le D \qquad (12)$$

$$\widehat{Z_J} + d_j^- - d_j^+ (\le or \ge) Z_j^*$$
$$\underline{\widehat{n_J}} \in F$$
$$n_{jh} and r_{jh} are integers and n_{jh} \in F; h = 1,2,\dots,L$$

Parameter ρ assigns the importance attached to the minimization of the weighted sum of unwanted deviation variables. Above formulation increase the feasible region by relaxing the constraint $(1-\rho)[W_{1j}d_j, W_{2j}d_j] \le D$ imposed in [21] into $[W_{1j}d_j, W_{2j}d_j] \le D$ as $0 \le \rho \le 1$. Integer nonlinear programming problems have a small feasible solution grid and we are already compromising on allocating sample size. This will help us to find feasible and optimal solution considering larger grid using this relaxation.

4.2Other Techniques

4.2.1Fuzzy Programming (FP)

When the optimal solution is not a firmly decisive solution, instead a compromise solution is required for the problem. The problem is required to be formulated into a fuzzy programming problem [10].

Let (Z_j^*) be the optimal value of (Z_j) obtained by solving the MOINLPP (7).

Further let

$$\widetilde{Z_J} = \widetilde{Z_J}(n_1, n_2, \dots, n_h, \dots, n_L)$$

Denote the value of the CV under the compromise allocation, where n_h ; $h = 1, 2, \dots, L$ are to be worked out. Obviously

$$\widetilde{Z_J} \ge Z_j^* and \widetilde{Z_J} - Z_j^* \ge 0;\ j = 1,2,\dots,p$$

Will give the increase in variance due to not using the individual optimum allocation for j^{th} characteristic.

To obtain a fuzzy solution, we first compute the maximum value U_k and the minimum value L_k , for each $k = 1, 2, \dots, p$.

Now,

$$L_k = \min_j Z_k\left(n_{h,j}^*\right) U_k = \max_j Z_k\left(n_{h,j}^*\right)$$

where $n_{h,j}^*$ denote the optimum allocation for the j^{th} charcteristic in four strata.

The differences of the maximum and minimum values of the Z_k are denoted by $d_k = U_k - L_k, k = 1, 2, \dots, p$. The fuzzy programming formulation of the MOINLPP in (7) is given by the following INLPP:

$$Minimize\delta$$
$$Subjectto$$
$$\widehat{Z_J} - \delta d_k \le Z_j^*$$
$$or$$
$$\acute{\widehat{Z_J}} - \delta d_k \le \acute{Z}_j^*$$

$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \leq C_0$ (13)
$2 \leq n_h \leq N_h$
$2 \leq r_h \leq \hat{n}_{h2}$
$n_h and r_h are integers; h = 1,2, \dots, L$

4.2.2The value Function Technique (VFT)

Khan et al.and Diaz-Garcia and Ulloa expressed problem under the value function technique as [29]:

$$Minimize \varphi\left(\sum_{j=1}^{p}(Z_j^2)\right)$$
$Subject to$

$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \leq C_0$ (14)
$2 \leq n_h \leq N_h$
$2 \leq r_h \leq \hat{n}_{h2}$
$n_h and r_h are integers; h = 1,2, \dots, L$

where $\varphi(.)$ is a scalar function that summarizes the importance of each of the coefficients of variance of the p characteristics. Usually, $\varphi(.)$ is taken as the weighted sum of the squares of p coefficient of variances. Under this property Eq. (14) becomes:

$$Minimize\left(\sum_{j=1}^{p}\alpha_j Z_j^2\right)$$

$Subject to$

$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \leq C_0$ (15)
$2 \leq n_h \leq N_h$
$2 \leq r_h \leq \hat{n}_{h2}$
$n_h and r_h are integers; h = 1,2, \dots, L$

where$\sum_{j=1}^{p}\alpha_j = 1$, $\alpha_j \geq 0$ j = 1,2, ...,p; α_j are the weights according to the relativeimportance of the characteristics. When complete information is available, the weights may be decided according to some measures of the relative importance of the characteristics. For Example, weights $\alpha_j$may be taken as $\alpha_j \alpha \sum_{h=1}^{L} S_{jh}^2$, $j = 1,2, \dots, por \alpha_j = \beta\sum_{h=1}^{L} S_{jh}^2$,where β is the constant of proportionality. Without loss of generality, we can assume that$\sum_{j=1}^{p}\alpha_j = 1$. Thus,

$$\sum_{j=1}^{P}\alpha_j = \beta\sum_{j=1}^{P}\sum_{h=1}^{L} S_{jh}^2$$
$$or$$
$$\beta = 1\Big/\sum_{j=1}^{P}\sum_{h=1}^{L} S_{jh}^2$$

This provides

$$\alpha_j = \sum_{h=1}^{L} S_{jh}^2 \Big/ \sum_{j=1}^{P}\sum_{h=1}^{L} S_{jh}^2$$

Using Eq. (4) and Eq. (5), MOINLPP (7) can be rewritten as:

$Minimize Z_j = \alpha_j CV_j^2 or \acute{Z}_j = \alpha_j C\acute{V}_j^2$

$Subject to$

$\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1})n_h + c_{h2}r_h \le C_0$ (16)

$2 \le n_h \le N_h$

$2 \le r_h \le \hat{n}_{h2}$

n_h and r_h are integers; $h = 1,2,\dots,L$

5. Application

[Data Source: www.agcensus.usda.gov]

Y_1 : The quantity of Corn harvested in 2007

Y_2 : The quantity of Soybean harvested in 2007

Ψ_1 : The quantity of Corn harvested in 2002

Ψ_2 : The quantity of Soyabean harvested in 2002

Here, $\bar{Y}_1$= 22698622.75 and $\bar{Y}_2$= 4306561.045

It is assumed that total cost of survey in Case 1 is C_0 =331 and in Case 2 is C_0 =346 units.

We considered last 27%, 30%, 27% and 20% values in each stratum as non response respectively.

The area of counties is used to stratify the population into 4 strata.

Let Ψ_{jhi} is artificial dichotomous variable, the cut off for quantitative variable to be transformed into attribute is set as respective stratum mean for each characteristic in the h^{th} stratum defined below:

Ψ_{11i} =1, if quantity is greater than 11778829.32
0, otherwise

Ψ_{12i} =1, if quantity is greater than 17339481.28
0, otherwise

Ψ_{13i} =1, if quantity is greater than 21277529.04
0, otherwise

Ψ_{14i} =1, if quantity is greater than 29384771.62
0, otherwise

Ψ_{21i} =1, if quantity is greater than 3340383.227
0, otherwise

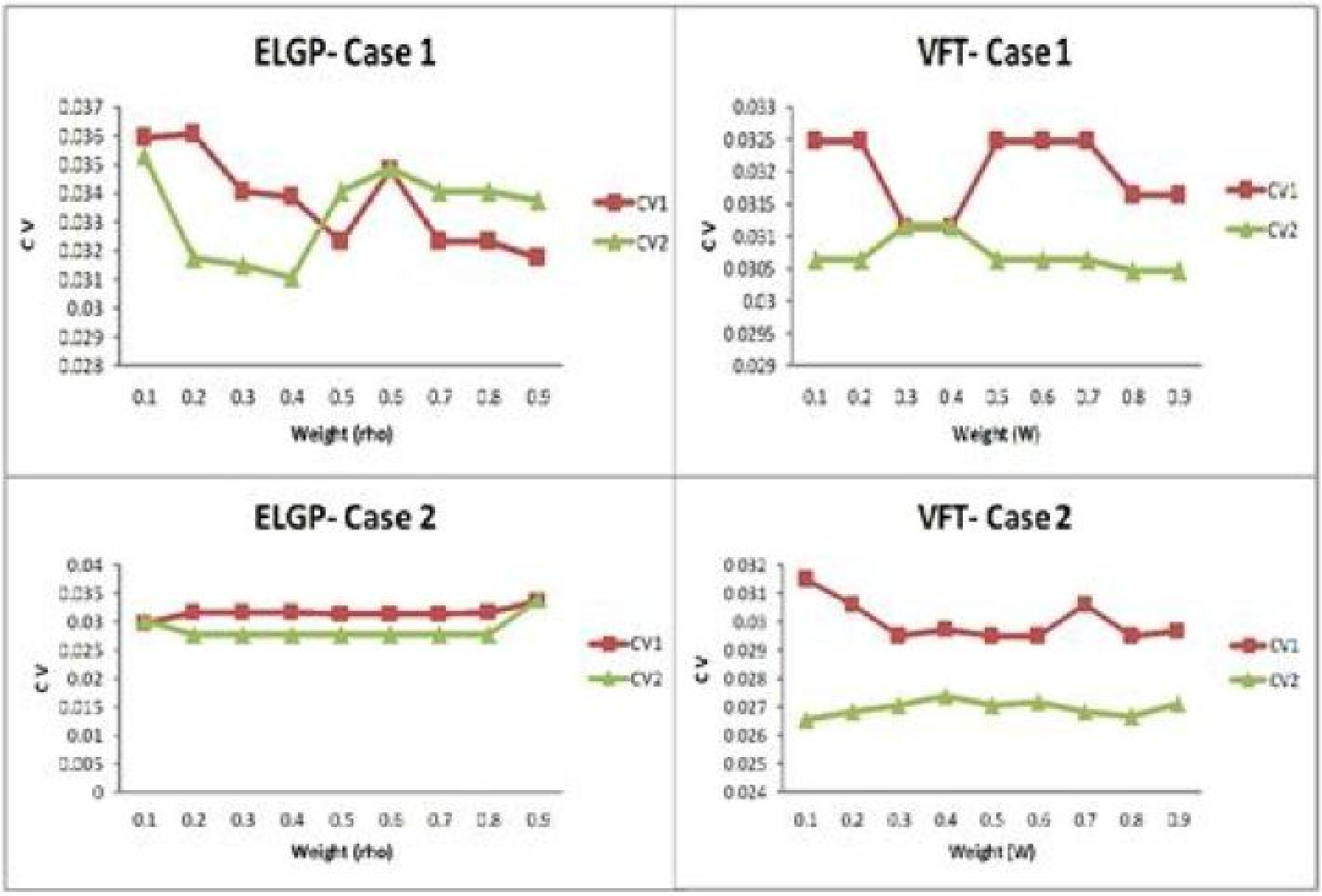
Ψ_{22i} =1, if quantity is greater than 4727488.8
0, otherwise

Ψ_{23i} =1, if quantity is greater than 5215178.667
0, otherwise

Ψ_{24i} = 1, if quantity is greater than 7740663.538
0, otherwise

The data statistics are presented in the APPENDIX:

Figure 1: CV under Different Weights



We assumed equal weights for both the characteristics in value function technique and extended lexicographic goal programming. The minimized CVs sample and subsample optimal allocations using the three Multiobjective optimization techniques for two cases of nonresponse using Data set are presented as:

Table 2 gives the optimum allocations for both cases of non-response using auxiliary attributes by different optimization methods. Results show that value function technique gives minimum value of objective function (i.e. CV) for compromised allocations. Howeverparameter $\rho = 0.1$ and$W = 0.5$ are randomly selected. The results by changing these values are discussed in the Fig 1.

Fig 1 shows the relation of changing weights with CVs in both data sets. In extended lexicographic programming technique for first data set, CV of corn harvested in 2007 is higher when we use $\rho \leq 0.4$ as compared to soyabean harvested in 2007. Both CVs are equal when we use $\rho=0.6$ but in case2, CV of corn harvested in 2007 is higher for all ρ. Value function technique shows higher difference among CVs with changing weights. CV of corn harvested in 2007 is higher than CV of soya bean harvested in 2007 for every arbitrary selection of weights in both cases i.e. case1 and case2.

References

Bretthauer, K. M., Ross, A., & Shetty, B. (1999). Nonlinear integer programming for optimal allocation in stratified sampling. European Journal of Operational Research, 116 (3),667-680.

Cochran, W. G.(1977). Sampling techniques. New York, Wiley and Sons, 98, 259-261.

Csenki, A. (1997). Optimum allocation in stratified random sampling via Holders inequality. Journalof the Royal Statistical Society: Series D (The Statistician), 46(3), 439-441.

Diaz-Garcia, J. A., & Cortez, L. U. (2006). Optimum allocation in multivariate strat-ified sampling: multi-objective programming. ComunicacionTecnica No. I-06-07/28-03-206 (PE/CIMAT). Guanajuato, Mexico: Centro deInvestigacionenMatematicas, AC.

El-Badry, M. A. (1956). A sampling procedure for mailed questionnaires.Journal of the American Statistical Association, 51(274), 209-227.

Gupta, N., Iftekhar, S. S., & Bari, A. Fuzzy goal programmming approach to solvenon-linear bi-level programming problem in stratified double sampling design in pres-ence of nonresponse. International Journal of Scientific & Engineering Research, 3(10)

Hansen, M. H., & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. Journal of the American Statistical Association, 41(236), 517-529.

Holmberg, A. (2002).A multiparameter perspective on the choice of sampling design in surveys.Statistics in Transition, 5(6), 969-994.

Haseen, S., Iftekhar, S., Ahsan, M., & Bari, A. (2012). A fuzzy approach for solving double sampling design in presence of non-response. International Journal of Engi-neering Science, 4.

Iftekhar, S., Haseen, S., Ali, Q. M., & Bari, A. A compromise solution in multivariate surveys with stochastic random cost function. International Journal of Scientific & Engineering Research, 4(6).

Khare, B. B. (1987). Allocation in stratified sampling in presence of nonresponse.Metron, 45(1-2), 213-221.

Khare, B. B., & Srivastava, S. (1997). Transformed ratio type estimators for the pop-ulation mean in the presence of nonresponse. Communications in Statistics-Theory and Methods,26(7), 1779-1791.

Khan, M. G., Khan, E. A., & Ahsan, M. J. (2008). Optimum allocation in multivariate stratified sampling in presence of non-response. Journal of Indian Society Agricultural Statistics, 62(1), 42-48.

Khan, M. G. M., Maiti, T., & Ahsan, M. J. (2010). An optimal multivariate stratified sampling design using auxiliary information: an integer solution using goal programming approach. Journal of Official Statistics, 26(4), 695.

Kumar, S. (2011). Estimation of the Population Mean in Presence of Non-Response.Communication of the Korean Statistical Society, 18(4), 537-548.

Khowaja, S., Ghufran, S., & Ahsan, M. J. (2011). Estimation of population means in multivariate stratified random sampling. Communications in Statistics-Simulation and Computation, 40(5), 710-718.

Khowaja, S., Guerin, S., & Ahsan, M. J. (2013). On the problem of compromise allocation in multi-response stratified sample surveys.Communications in Statistics−Simulation and Computation, 42(4), 790-799.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, 97(4), 558-625.

Naik, V. D., & Gupta, P. C. (1996). A note on estimation of mean with known population proportion of an auxiliary character. Journal of Indian Society Agriculture Statistics, 48(2), 151-158.

Rao, P. S. R. S. (1986). Ratio estimation with subsampling the nonrespondents.Survey Methodology, 12(2), 217-230.

Romero, C. (2001). Extended lexicographic goal programming: a unifying approach.Omega, 29(1), 63-71.

Singh Raghav, Y., Ali, I., & Bari, A. (2012). Multi-objective nonlinear program-ming problem approach in multivariate stratified sample surveys in the case of non-response. Journal of Statistical Computation and Simulation, (ahead-of-print), 1-15.

Srikantan, K. (1963). A problem in optimum allocation.Operations Research, 11(2),265-273.

Sukhatme, P. V., &Sukhatme, B. V. (1970). Sampling theory of surveys with appli-cations, (pp. 27-29). London: Asia Publishing House.

Shabbir, J., & Gupta, S. (2007). On estimating the finite population mean with known population proportion of an auxiliary variable. Pakistan Journal of Statistics-All Series, 23(1), 1.

Singh, H. P., & Kumar, S. (2008). A regression approach to the estimation of the finite population mean in the presence of nonresponse. Australian & New ZealandJournal of Statistics, 50(4), 395-408.

Sahin, S. T. (2011). Determination of Sample Size Selecting from Strata under Non-linear Cost Constraint by Using Goal Programming and Kuhn-Tucker Methods. Gazi University Journal of Science, 24(2), 249-262.

Shabbir, J., &Sagir, A. (2012). Estimation of Finite Population Mean in Stratified Random Sampling Using Auxiliary Attribute (s) under Non-Response. Pakistan Journal of Statistics and Operation Research, 8(1).

Singh Raghav, Y., Ali, I., & Bari, A. (2012). Multi-objective nonlinear programming problem approach in multivariate stratified sample surveys in the case of nonresponse. Journal of Statistical Computation and Simulation, 1-15.

Shabbir, J., &Sagir, A. (2012). Estimation of Finite Population Mean in Stratified Random Sampling Using Auxiliary Attribute(s) under Non-Response. Pakistan Journal of Statistics and Operation Research, 8(1).

Singh, H. P., & Solanki, R. S. (2012). Improved estimation of population mean in simple random sampling using information on auxiliary attribute. Applied Mathematics and Computation, 218(15), 7798-7812.

Singh, R. V. K., &Audu, A. (2013). Efficiency of Ratio Estimators in Stratified Random Sampling Using Information on Auxiliary Attribute. International Journal of Engineering Science and Innovative Technology(IJESIT), 2(1).

Sharma, M. K., & Ismail, M. V. (2013). Some Multi Convex Programming Problems Arising in Multivariate Sampling.International Journal of Computer Applications,61 (21).

Varshney, R., Ahsan, M. J., & Khan, M. G. (2011). An optimum multivariate strati-fied sampling design with nonresponse: A lexicographic goal programming approach. Journal of Mathematical Modelling and Algorithms, 10(4), 393-405.

Varshney, R., & Ahsan, M. J. (2012). An optimum multivariate stratified double sampling design in presence of non-response. Optimization Letters, 6(5), 993-1008.

Appendix

Table 1.Summary statistics

Ψ_{1h}	Ψ_{2h}	$S^2_{\psi_{2h}}$	$S^2_{\psi_{2h2}}$	$S_{(y\psi)_{1h}}$	$S_{(y\psi)_{1h2}}$	$S_{(y\psi)_{2h}}$	$S_{(y\psi)_{2h2}}$
0.591	0.591	0.253	0.267	2622322.766	2233805.44	550824.487	536741.460
0.575	0.475	0.255	0.242	4340658.415	5431693.843	590425.8192	720235.2173
0.5	0.542	0.259	0.167	3072607.891	2634536.542	745074.7953	1044380.885
0.308	0.461	0.269	0.333	2664563.75	3489303	802484.31	203389.997

Table 1: Summary Statistics for Data

h	N_h	W_h	W_{h1}	W_{h2}	c_{h0}	c_{h1}	c_{h2}	R_{1h}	R_{2h}	B_{1h}
1	22	0.222	0.73	0.27	1	2	4	23389830.450	5077552.846	9833710.373
2	40	0.404	0.7	0.3	1	3	5	36982957.878	8726614.632	17318227.95
3	24	0.242	0.73	0.27	1	4	6	52736301.58	8029735.769	11778330.25
4	13	0.131	0.8	0.2	1	5	7	115037662.884	15041861	11546442.92

B_{2h}	$S^2_{y_{1h}}$	$S^2_{y_{1h2}}$	$S^2_{y_{2h}}$	$S_{y^2_{2h2}}$	$S^2_{\psi_{1h}}$	$S^2_{\psi_{1h2}}$
2175050.538	5.76×10^{13}	7.80×10^{13}	1.67×10^{12}	1.48×10^{12}	0.253	0.267
2308431.774	1.21×10^{14}	1.22×10^{14}	2.50×10^{12}	2.807×10^{12}	0.251	0.265
2876092.916	5.57×10^{13}	2.63×10^{13}	3.58×10^{12}	3.02×10^{12}	0.261	0.267
2980656	7.08×10^{13}	4.01×10^{13}	4.44×10^{12}	6.28×10^{11}	0.231	0.333

Table 2: Compromise Allocations and Corresponding Values of the Objective Functions Obtained by Different Methods

Compromise allocation	Extended Lexicographic goal programming	Fuzzy goal programming	Value function technique
Case 1			
(n_1, r_1)	(12,3)	(12,3)	(15,4)
(n_2, r_2)	(39,11)	(40,11)	(24,7)
(n_3, r_3)	(8,2)	(8,2)	(19,5)
(n_2, r_2)	(10,2)	(10,2)	(10,2)
CV_1	0.0323	0.0322	0.0324
CV_2	0.0340	0.0340	0.0306
Case 2			
(n_1, r_1)	(8,2)	(15,4)	(15,4)
(n_2, r_2)	(40,10)	(34,8)	(34,7)
(n_3, r_3)	(17,2)	(17,2)	(17,2)
(n_4, r_4)	(10,2)	(11,2)	(12,2)
CV_1	0.0315	0.0295	0.0295
CV_2	0.0277	0.0270	0.0270