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, \( \Psi_1 \) can be defined as the number of hours spent on studying (low vs high) and \( \Psi_2 \) 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 \( \Psi_1 \) can be the irrigation status (irrigated vs. non-irrigated) and \( \Psi_2 \) 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 SRSSWOR 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 into two 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 by \( N_{h1} \), \( N_{h2} = N - N_{h1} \), \( n_{h1} \), \( n_{h2} = n - n_{h1} \) 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 \( \psi_{jhi} \) be the \( i^{th} \) value of the study variable (Y) and the auxiliary attribute (\( \Psi \)) respectively having the \( j^{th} \) characteristic in the \( h^{th} \) stratum. Let \( \psi_{jhi} \) is an artificial dichotomous attribute which 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 values less 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 define \( \psi_{jhi} \). Let the population and sample proportion of units of auxiliary attributes \( \psi^*_j \) are defined as:

\[
\bar{\psi}_j = \frac{\sum_{i=1}^{N_h} \psi_{jhi}}{N_h} \text{ and } \bar{\psi}_j = \frac{\sum_{i=1}^{n_h} \psi_{jhi}}{n_h}
\]

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

\[
S^2_{yjhi} = \frac{\sum_{h=1}^{L} \left( y_{jhi} - \bar{y}_{jhi} \right)^2}{N_h - 1}, \text{ where } \bar{y}_{jhi} = \frac{\sum_{i=1}^{N_h} \psi_{jhi}}{N_h}
\]
Similarly we can define the variants and covariates under non-response as:

\[
S^2_{\psi j h 2} = \sum_{h=1}^{L} \frac{\left( y_{j hi} - \bar{\psi}_{j h 2} \right)^2}{N_h - 1}, \text{ where } \bar{\psi}_{j h 2} = \sum_{i=1}^{N_h} \frac{y_{j hi}}{N_h}
\]

\[
S_{\left(\psi j h 1, \psi j h 2\right)} = \sum_{i=1}^{N_h} \frac{(y_{j hi} - \bar{\psi}_{j h 1})(\bar{\psi}_{j h 1} - \bar{\psi}_{j h 2})}{N_h - 1}
\]

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 \( \bar{\psi}_{j h} \) in the \( h^{th} \) stratum mean for the \( j^{th} \) characteristic is given by:

\[
\bar{\psi}_{j h} = \frac{n_{h1} \bar{y}_{j h 1} + n_{h2} \bar{y}_{j h 2}}{n_h}
\]

Where \( \bar{y}_{j h 1} \) is the sample mean based on \( n_{h1} \) respondent units and \( y_{j h 2} \) is the sample mean based on \( r_h \) subsample units from non-respondents.

**Case 2**

When non-response is in auxiliary attributes and study variables then we define the unbiased estimator of \( \bar{\psi}_{j h} \) for the \( j^{th} \) characteristic in the \( h^{th} \) stratum mean as:

\[
\bar{\psi}_{j h} = \frac{n_{h1} \bar{y}_{j h 1} + n_{h2} \bar{y}_{j h 2}}{n_h}
\]

where \( \bar{y}_{j h 1} \) is the sample mean based on \( n_{h1} \) respondent units and \( \bar{y}_{j h 2} \) is the sample mean based on \( r_h \) subsample units from non-respondents.

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 majority 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, \( c_{h1} = \sum_{j=1}^{p} c_{j h 1} \) is the cost per unit in taking account of responding units \( n_{h1} \) and \( c_{h2} = \sum_{j=1}^{p} c_{j h 2} \) is the cost per unit in taking account of subsamples \( r_h \) from non-responding units \( n_{h2} \) in the \( h^{th} \) stratum. Here, \( c_{j h 1} \) and \( c_{j h 2} \) 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 \( W_{h2} r_h \) respectively. The total expected cost \( \hat{C} \) of the survey may be given as:

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

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

$$\bar{y}_{t,r,s} = \sum_{h=1}^{L} W_h \bar{y}_{r,jh},$$

where $\bar{y}_{r,jh} = \bar{y}_{j} + b_{jh} (\bar{\psi}_{jh} - \bar{\psi}_{j})$

Here $b_{jh}$ is the sample regression coefficient.

The Squared CV of $y_{t,r,s}$ for $j_{th}$ characteristic is:

$$CV_j^2 = \sum_{h=1}^{L} \frac{W_h^2}{n_h Y_j^2} \left( S_{yj}^2 + B_{jh}^2 S_{\psi_j}^2 - 2 B_{jh} S_{y\psi_j} \right) + \sum_{h=1}^{L} \frac{W_h^2}{r_h^2} \left( \frac{W_{h2}}{r_h} - \frac{1}{n_h} \right) W_{h2} S_{yj}^2$$

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

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

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

Where

$$\bar{y}_{r,jh} = \bar{y}_{j} + b_{jh} (\bar{\psi}_{jh} - \bar{\psi}_{j})$$

The squared CV of $\bar{y}_{r,s}$ for $j_{th}$ characteristic is

$$CV_j^2 = \sum_{h=1}^{L} \frac{W_h^2}{n_h Y_j^2} \left( S_{yj}^2 + B_{jh}^2 S_{\psi_j}^2 - 2 B_{jh} S_{y\psi_j} \right) + \sum_{h=1}^{L} \frac{W_h^2}{r_h^2} \left( \frac{W_{h2}}{r_h} - \frac{1}{n_h} \right) W_{h2} \left( S_{yj}^2 + B_{jh}^2 S_{\psi_j}^2 - 2 B_{jh} S_{y\psi_j} \right)$$

(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 $[Z_1, Z_2, \ldots, Z_p]$

Subject to

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

$$2 \leq n_h \leq N_h$$

$$2 \leq r_h \leq R_{h2}$$

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

(6)

where $Z_j, (j = 1, 2, \ldots, p)$ 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

$$Z_j = CV_j^2 = \sum_{j=1}^{p} \sum_{h=1}^{L} \frac{W_h^2}{n_h Y_j^2} \left( S_{yj}^2 + B_{jh}^2 S_{\psi_j}^2 - 2 B_{jh} S_{y\psi_j} \right) + \sum_{j=1}^{p} \sum_{h=1}^{L} \frac{W_h^2}{r_h^2} \left( \frac{W_{h2}}{r_h} - \frac{1}{n_h} \right) W_{h2} \left( S_{yj}^2 + B_{jh}^2 S_{\psi_j}^2 - 2 B_{jh} S_{y\psi_j} \right)$$

Or

108
\[ Z_j = \hat{C}V_j^2 = \sum_{j=1}^{P} \sum_{h=1}^{L} \frac{W_h^2}{n_{h,j}} \left( S_{y,jh}^2 + B_{j,h}^2 S_{\psi,jh}^2 - 2B_{j,h} S_{(y,\psi,jh)} \right) \]

\[ + \sum_{j=1}^{P} \sum_{h=1}^{L} \frac{W_h^2}{n_{h,j}} \left( \frac{W_{h,j}^2}{n_h} - \frac{1}{n_h} \right) W_{h,j} \left( S_{y,jh,2}^2 + B_{j,h} S_{\psi,jh,2}^2 - 2B_{j,h} S_{(y,\psi,jh,2)} \right) \]

Subject to

\[ \sum_{h=1}^{L} (c_{h_0} + c_{h_1} W_{h,1}) n_h + c_{h_2} r_h \leq C_0 \]

\[ 2 \leq n_h \leq N_h \]

\[ 2 \leq r_h \leq \hat{r}_{h_2} \]

\[ n_h \text{ and } r_h \text{ are integers; } h = 1, 2, \ldots, L \]

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.

4.1 Modified 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 \( a = 1, \ldots, p \). 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:

\[ \text{Minimize } Z_j = f(n_{j,h}, r_{j,h}) \]

Subject to

\[ \sum_{h=1}^{L} (c_{h_0} + c_{h_1} W_{h,1}) n_h + c_{h_2} r_h \leq C_0 \]

\[ 2 \leq n_h \leq N_h \]

\[ 2 \leq r_h \leq \hat{r}_{h_2} \]

\[ n_{j,h} \text{ and } r_{j,h} \text{ are integers and } n_{j,h} \in F; h = 1, 2, \ldots, L \]

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:

\[ \hat{z}_j + d^-_j - d^+_j = Z_j^* \]

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

\[ \hat{r}_j \in F \]

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:

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

Subject to

$$
\bar{Z}_j + d_j^- - d_j^+ (\leq \text{ or } \geq) Z_j^*, \\
\mathbf{n} \in \mathbf{F}
$$

where $$f_1, f_2, \ldots, f_p$$ represent priority-wise functions and $$d_j^-, d_j^+$$ 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 deviations 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]:

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

Subject to

$$
\bar{Z}_j + d_j^- - d_j^+ (\leq \text{ or } \geq) Z_j^*, \\
\mathbf{n} \in \mathbf{F}, h = 1, 2, \ldots, L
$$

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:

$$
\text{Minimize D}
$$

Subject to

\[
\begin{align*}
\left[ d_j^-, d_j^+ \right] & \leq D \\
\bar{Z}_j + d_j^- - d_j^+ (\leq \text{ or } \geq) Z_j^* \\
\mathbf{n} \in \mathbf{F}
\end{align*}
\]

$$
\mathbf{n} \in \mathbf{F}, h = 1, 2, \ldots, L
$$
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:

$$
\text{Minimize} \{(1 - \rho)D + \rho \sum_{j=1}^{p} f_j \left( W_1 d_j^{-}, W_2 d_j^{+} \right) \}
$$

Subjectto

$$
\left[ \left( W_1 d_j^{-}, W_2 d_j^{+} \right) \right] \leq D \quad (12)
$$

$$
\bar{Z}_j + d_j^{-} - d_j^{+} (\leq \text{ or } \geq) Z_j^*
$$

$$
\bar{n}_j \in \mathcal{F}
$$

\( n_{jh} \) and \( r_{jh} \) are integers and \( n_{jh} \in \mathcal{F}; h = 1,2,\ldots,L \)

Parameter \( \rho \) 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 \( (1 - \rho)[W_1 d_j, W_2 d_j] \leq D \) imposed in [21] into \( [W_1 d_j, W_2 d_j] \leq D \) as \( 0 \leq \rho \leq 1 \). 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.

4.2 Other Techniques

4.2.1 Fuzzy 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

$$
\bar{Z}_j = Z_j(n_1, n_2, \ldots, n_h, \ldots, n_L)
$$

Denote the value of the CV under the compromise allocation, where \( n_h : h = 1,2,\ldots,L \) are to be worked out. Obviously

$$
\bar{Z}_j \geq Z_j^* \text{ and } \bar{Z}_j - Z_j^* \geq 0; \ j = 1,2,\ldots,p
$$

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 \( k = 1,2,\ldots,p \).

Now,

$$
L_k = \min_{j} Z_k(n_{h,j}^*) U_k = \max_{j} Z_k(n_{h,j}^*)
$$

where \( n_{h,j}^* \) denote the optimum allocation for the \( j^{th} \) characteristic in four strata.

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

$$
\text{Minimize} \delta
$$

Subjectto

$$
\bar{Z}_j - \delta d_k \leq Z_j^*
$$

or

$$
\bar{Z}_j - \delta d_k \leq \hat{Z}_j^*
$$
\[ \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + c_{h2} r_h \leq C_0 \]
\[ 2 \leq n_h \leq N_h \]
\[ 2 \leq r_h \leq \bar{n}_{h2} \]
\[ n_h \text{ and } r_h \text{ are integers; } h = 1, 2, ..., L \]

4.2.2 The Value Function Technique (VFT)

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

\[ \text{Minimize } \varphi \left( \sum_{j=1}^{p} (Z_j^2) \right) \]

Subjectto

\[ \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + c_{h2} r_h \leq C_0 \]
\[ 2 \leq n_h \leq N_h \]
\[ 2 \leq r_h \leq \bar{n}_{h2} \]
\[ n_h \text{ and } r_h \text{ are integers; } h = 1, 2, ..., L \]

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:

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

Subjectto

\[ \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + c_{h2} r_h \leq C_0 \]
\[ 2 \leq n_h \leq N_h \]
\[ 2 \leq r_h \leq \bar{n}_{h2} \]
\[ n_h \text{ and } r_h \text{ are integers; } h = 1, 2, ..., L \]

where \( \sum_{j=1}^{p} \alpha_j = 1 \), \( \alpha_j \geq 0 \) \( j = 1, 2, ..., p; \) \( \alpha_j \) are the weights according to the relative importance 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 \( \alpha_j = \alpha \sum_{h=1}^{L} S_{jh}^2 \), \( j = 1, 2, ..., p \) or \( \alpha_j = \beta \sum_{h=1}^{L} S_{jh}^2 \), where \( \beta \) 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

\[ \beta = \frac{1}{\sum_{j=1}^{p} \sum_{h=1}^{L} S_{jh}^2} \]
This provides

\[
\alpha_j = \frac{\sum_{h=1}^{L} S_{jh}^2}{\sum_{j=1}^{p} \sum_{h=1}^{L} S_{jh}^2}
\]

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

\[
\text{Minimize } Z_j = \alpha_j CV_j^2 \text{ or } Z_j = \alpha_j CV_j^2
\]

Subject to

\[
\sum_{h=1}^{H} (c_{h0} + c_{h1} W_{h1}) n_h + c_{h2} r_h \leq C_0
\]

\[
2 \leq n_h \leq N_h
\]

\[
2 \leq r_h \leq \bar{r}_{h2}
\]

\[n_h \text{ and } r_h \text{ are integers; } h = 1, 2, \ldots, L\]

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

\(\Psi_1\) : The quantity of Corn harvested in 2002

\(\Psi_2\) : The quantity of Soybean 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 \(\psi_{jhi}\) 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:

\(\psi_{11i} = 1, \text{ if quantity is greater than } 11778829.32\)

\(\psi_{11i} = 0, \text{ otherwise}\)

\(\psi_{12i} = 1, \text{ if quantity is greater than } 17339481.28\)

\(\psi_{12i} = 0, \text{ otherwise}\)

\(\psi_{13i} = 1, \text{ if quantity is greater than } 21277529.04\)

\(\psi_{13i} = 0, \text{ otherwise}\)

\(\psi_{14i} = 1, \text{ if quantity is greater than } 29384771.62\)

\(\psi_{14i} = 0, \text{ otherwise}\)

\(\psi_{21i} = 1, \text{ if quantity is greater than } 3340383.227\)

\(\psi_{21i} = 0, \text{ otherwise}\)

\(\psi_{22i} = 1, \text{ if quantity is greater than } 4727488.8\)

\(\psi_{22i} = 0, \text{ otherwise}\)

\(\psi_{23i} = 1, \text{ if quantity is greater than } 5215178.667\)

\(\psi_{23i} = 0, \text{ otherwise}\)

\(\psi_{24i} = 1, \text{ if quantity is greater than } 7740663.538\)

\(\psi_{24i} = 0, \text{ otherwise}\)
The data statistics are presented in the APPENDIX:

**Figure 1: CV under Different Weights**

![Figure 1: CV under Different Weights](image)

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. However parameter $\rho = 0.1$ 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 soyabeans 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 $\rho$. Value function technique shows higher difference among CVs with changing weights. CV of corn harvested in 2007 is higher than CV of soybean harvested in 2007 for every arbitrary selection of weights in both cases i.e. case1 and case2.
References


Appendix

Table 1. Summary statistics

<table>
<thead>
<tr>
<th>( \Psi_{1h} )</th>
<th>( \Psi_{2h} )</th>
<th>( S^2_{\Psi_{1h}} )</th>
<th>( S^2_{\Psi_{2h}} )</th>
<th>( S_{(\Psi)_{1h}} )</th>
<th>( S_{(\Psi)_{2h}} )</th>
<th>( S_{(\Psi)_{1a}} )</th>
<th>( S_{(\Psi)_{2a}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.591</td>
<td>0.591</td>
<td>0.253</td>
<td>0.267</td>
<td>2622322.766</td>
<td>2235505.444</td>
<td>550824.487</td>
<td>536741.460</td>
</tr>
<tr>
<td>0.575</td>
<td>0.475</td>
<td>0.255</td>
<td>0.242</td>
<td>4340658.415</td>
<td>5431693.843</td>
<td>590425.819</td>
<td>720235.2173</td>
</tr>
<tr>
<td>0.5</td>
<td>0.542</td>
<td>0.259</td>
<td>0.167</td>
<td>3072607.891</td>
<td>2634536.542</td>
<td>745074.7953</td>
<td>1044380.885</td>
</tr>
<tr>
<td>0.308</td>
<td>0.461</td>
<td>0.269</td>
<td>0.333</td>
<td>2664563.75</td>
<td>3459303</td>
<td>802484.31</td>
<td>203389.997</td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics for Data

<table>
<thead>
<tr>
<th>( h )</th>
<th>( N_h )</th>
<th>( W_h )</th>
<th>( W_{h1} )</th>
<th>( W_{h2} )</th>
<th>( c_{h0} )</th>
<th>( c_{h1} )</th>
<th>( c_{h2} )</th>
<th>( R_{1h} )</th>
<th>( R_{2h} )</th>
<th>( B_{1h} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>0.222</td>
<td>0.73</td>
<td>0.27</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>23389830.450</td>
<td>5077552.846</td>
<td>9833710.373</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0.404</td>
<td>0.7</td>
<td>0.3</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>36982967.878</td>
<td>8726614.632</td>
<td>17318227.95</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>0.242</td>
<td>0.73</td>
<td>0.27</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>52736301.58</td>
<td>8029735.769</td>
<td>11778330.25</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>0.131</td>
<td>0.8</td>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>115037662.884</td>
<td>15041861.56</td>
<td>11546442.92</td>
</tr>
</tbody>
</table>

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

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