

## Research Article

# Estimation of Finite Population Mean in Multivariate Stratified Sampling under Cost Function Using Goal Programming

Atta Ullah,<sup>1</sup> Javid Shabbir,<sup>2</sup> Zawar Hussain,<sup>2</sup> and Bander Al-Zahrani<sup>3</sup>

<sup>1</sup> Department of Mathematics, Comsats Institute of Information Technology, Attock 43600, Pakistan

<sup>2</sup> Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

<sup>3</sup> Department of Statistics, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Correspondence should be addressed to Atta Ullah; [attaqaul3@gmail.com](mailto:attaqaul3@gmail.com)

Received 17 March 2014; Revised 26 May 2014; Accepted 16 July 2014; Published 14 August 2014

Academic Editor: Dongqing Wang

Copyright © 2014 Atta Ullah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In practical utilization of stratified random sampling scheme, the investigator meets a problem to select a sample that maximizes the precision of a finite population mean under cost constraint. An allocation of sample size becomes complicated when more than one characteristic is observed from each selected unit in a sample. In many real life situations, a linear cost function of a sample size  $n_h$  is not a good approximation to actual cost of sample survey when traveling cost between selected units in a stratum is significant. In this paper, sample allocation problem in multivariate stratified random sampling with proposed cost function is formulated in integer nonlinear multiobjective mathematical programming. A solution procedure is proposed using extended lexicographic goal programming approach. A numerical example is presented to illustrate the computational details and to compare the efficiency of proposed compromise allocation.

## 1. Introduction

It is common practice in sample survey related to agriculture, market, industries, and social research, and so forth that usually more than one characteristic is observed from each sampled unit of population. Stratified random sampling is more suitable than other survey designs used for obtaining information from heterogeneous population for reasons of economy and efficiency. The theory of stratified random sampling deals with the properties of estimator constructed from stratified random sample and with the best (optimum) choice of sample size to be selected from various strata either to maximize the precision of constructed estimator for a fixed cost or to minimize the cost of survey for fixed precision of estimator. The sample sizes selected according to above criteria are known as “optimum allocation.” In general, variance of study variate varies from stratum to stratum that provides basis for selecting optimum sample size.

Tschuprow [1] and Neyman [2] independently proposed an allocation procedure that minimizes variance of sample

mean under a linear cost function of sample size  $n_h$  in stratified random sampling scheme. Neyman [2] used Lagrange multiplier optimization technique to get optimum sample size for single variable under study. In stratified sampling, sample allocation problem becomes complicated when more than one characteristic is observed from each selected unit of a finite population. An allocation which is optimum for single characteristic may not be optimum for others unless the characteristics are highly correlated. There is need to use some compromise allocation criteria which produce an optimum allocation for all characteristics in some sense, for example, an allocation that minimizes the trace of variance-covariance matrix of the estimator of population mean or an allocation that minimizes the weighted average of variances or an allocation that maximizes the total relative efficiency of the estimators as compared to corresponding individual optimum allocation (Varshney et al. [3]). Many authors such as Dalenius [4, 5], Ghosh [6], Folks and Antle [7], Chromy [8], Bethel [9], Jahan et al. [10, 11], Khan et al. [12], Khan et al. [13, 14], Ansari et al. [15], Khan et al. [16], and Varshney et al.

[17] used different compromise criterion to solve allocation problem in stratified random sampling scheme.

The cost of survey is an important factor of sample allocation to various strata. The linear cost function used in stratified sampling is given as

$$C = c_0 + \sum_{h=1}^L c_h n_h, \quad (1)$$

where  $C$  denotes total budget available for survey,  $c_h$  for  $h = 1, 2, \dots, L$  represents measurement per unit cost in the  $h$ th stratum,  $c_0$  represents fixed cost of survey, and  $n_h$  is number of sample units selected in the  $h$ th stratum. In many practical situations, measurement unit cost and travel cost within strata are important factors of survey cost. The nonlinear cost function including measurement unit cost and traveling cost within strata is good approximation to actual cost of survey. Beardwood et al. [18] suggested that the shortest rout among  $k$  randomly disperse destination within a region is asymptotically proportional to  $\sqrt{k}$  for large  $k$ . Varshney et al. [17] used nonlinear cost function for large sample size given in (2). Consider

$$C = c_0 + \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h}, \quad (2)$$

where  $t_h$  is travel cost within  $h$ th stratum. The problem of finding the shortest rout among  $n_h$  selected units in  $h$ th stratum is often called the "shortest rout problem" in the operation research literature. If rout map and its length is given for each strata, we find shortest rout among  $n_h$  units within strata that  $n_h$  is either small or large. This shortest rout is used for practical purpose with confidence (Beardwood et al. [18]).

Consider following proposed nonlinear cost function:

$$\dot{C} = \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h n_h^\delta, \quad (3)$$

where  $\dot{C} = C - c_0$  and  $\delta$  represents the effect of travel within strata to cost function. The value of  $\delta$  is determined by solving shortest rout problem using methods discussed by Hiller and Lieberman [19]. The cost function in (2) becomes particular case of our proposed cost function given in (3) if  $\delta = 0.5$ .

Generally, Lagrange multiplier technique (LMT) is used to determine sample size. However, the constraint  $2 \leq n_h \leq N_h$ , where  $n_h$  ( $h = 1, 2, 3, \dots, L$ ) is an integer neglected in using LMT. For integer value of sample size  $n_h$ , rounding rule is used which may lead to violating the optimality or feasibility conditions (or both). We need integer value of sample size  $n_h$  for practical implementation. Therefore, the authors did not try to use LMT and used integer programming for integer value of strata sample sizes  $n_h$ .

In this paper, we discuss compromise allocation based on minimization of coefficients of variation of regression estimators of population mean in multivariate stratified random sampling design under proposed nonlinear cost function (3). The problem is formulated in multiobjective integer

nonlinear programming. The extended lexicographic goal programming technique is applied to solve formulated allocation problem. The GAMS—AlphaECP Rosenthal [20] optimization software is used to solve numerical example which illustrates the computational detail of allocation procedure.

## 2. Formulation of the Problem

Consider a population of  $N$  units divided in to  $L$  mutually exclusive strata of size  $N_h$  ( $h = 1, 2, \dots, L$ ) such that  $\sum_{h=1}^L N_h = N$ . The simple random sample of size  $n_h$  is drawn from each stratum independently. Suppose we observe  $Y_{ji}$  ( $i = 1, 2, \dots, N_h, j = 1, 2, \dots, p$ ),  $p \geq 2$ , characteristics from each unit in  $h$ th stratum and estimate population mean of  $p \geq 2$  characteristics. Let  $\bar{y}_{jh}$  and  $\bar{x}_{jh}$  be the sample means and  $\bar{Y}_{jh}$  and  $\bar{X}_{jh}$  the population means of study variable  $Y_{jh}$  and auxiliary variable  $X_{jh}$ , respectively, of  $j$ th characteristics in the  $h$ th stratum.  $S_{yjh}^2$  and  $S_{xjh}^2$  are population variance and  $S_{yxjh}$  is population covariance between the  $j$ th study and auxiliary variable in the  $h$ th stratum.  $b_{jh} = s_{yxjh}/s_{xjh}^2$  and  $\beta_{jh} = S_{yxjh}/S_{xjh}^2$  are sample and population regression coefficients and  $W_h = N_h/N$  is stratum weight.

Consider an estimator,

$$\bar{y}_{j,lrs} = \sum_{h=1}^L W_h \bar{y}_{j,lrh}, \quad (4)$$

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

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

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

$$\text{MSE}(\bar{y}_{j,lrs}) = \sum_{h=1}^L \frac{W_h^2 \dot{U}_{jh}}{n_h} - \sum_{h=1}^L \frac{W_h^2 \dot{U}_{jh}}{N_h}, \quad (6)$$

where

$$\dot{U}_{jh} = S_{yjh}^2 - 2\beta_{jh} S_{yxjh} + \beta_{jh}^2 S_{xjh}^2. \quad (7)$$

If we ignore the second term in RHS of (6) because it is independent of sample size  $n_h$ , then

$$\text{MSE}(\bar{y}_{j,lrs}) = \sum_{h=1}^L \frac{W_h^2 \dot{U}_{jh}}{n_h}. \quad (8)$$

Since different characteristics are measured with different units, we need to use an estimate which should be independent of measurement unit. Therefore, coefficient of variation is used instead of mean square error; that is,

$$\text{C.V}(\bar{y}_{j,lrs}) = \sqrt{\frac{\text{MSE}(\bar{y}_{j,lrs})}{\bar{Y}_j^2}} \quad (9)$$

or

$$C.V(\bar{y}_{j,lrs}) = Z_j = \sqrt{\sum_{h=1}^L \frac{\hat{u}_{jh}}{n_h}} \tag{10}$$

where

$$\hat{u}_{jh} = \frac{W_h^2 \hat{U}_{jh}}{\bar{Y}_j^2} \tag{11}$$

A sample size  $n_h$  is determined under proposed nonlinear cost function in (3) that minimizes coefficients of variation of the estimator of population mean for each characteristics  $Y_j$  ( $j = 1, 2, \dots, p$ ). This problem may be formulated in multiobjective integer nonlinear programming as in (12). Consider

$$\begin{aligned} & \text{Minimize} \quad (Z_1, Z_2, \dots, Z_p) \\ & \text{Subject to} \quad \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h n_h^\delta \leq \hat{C} \\ & \qquad \qquad \qquad 2 \leq n_h \leq N_h \\ & \qquad \qquad \qquad n_h \text{ are integers} \\ & \qquad \qquad \qquad n_h \in F; \quad h = 1, 2, \dots, L, \end{aligned} \tag{12}$$

where  $F$  represents the feasible region that fulfills all constraints and sign restrictions. Any solution that exists within feasible region is implementable in practice.

### 3. Extended Lexicographic Goal Programming

Romero [21] proposed extended lexicographic goal programming method that provides a general framework which covers and allows the mixture of most common method of solving multiobjective decision making problems. It is also encompasses distance based multicriteria decision making technique. Romero [22] extended this work to make more general form of objective function. It is a technique used by decision makers for optimizing more than one objective under some constraints. In goal programming, all specified objectives are included in the model. The decision maker tries to minimize the potential deviations from specified objectives.

Consider the following individual optimum problem:

$$\begin{aligned} & \text{Minimize} \quad Z_j \\ & \text{Subject to} \quad \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h n_{jh}^\delta \leq \hat{C} \\ & \qquad \qquad \qquad 2 \leq n_{jh} \leq N_h \\ & \qquad \qquad \qquad n_{jh} \text{ are integers, } n_{jh} \in F. \\ & \qquad \qquad \qquad h = 1, 2, \dots, L, \quad j = 1, 2, \dots, p. \end{aligned} \tag{13}$$

Let  $Z_j^*$  be the individual optimum values of  $Z_j$  obtained by solving above problem. These optimum values  $Z_j^*$  specify

objectives and try to achieve these objectives using multi-objective mathematical programming. Let  $\widehat{Z}_j$  be values of objectives obtained by applying multiobjective optimization method. It is obvious that  $\widehat{Z}_j \geq Z_j^*$  or  $\widehat{Z}_j - Z_j^* \geq 0$  is the increase in  $Z_j$  due to compromise among objectives using compromise criterion. Suppose this increase is  $d_j^+ \geq 0$ . To achieve these specified objectives, we must have

$$\widehat{Z}_j - Z_j^* \leq d_j^+ \tag{14}$$

or

$$\widehat{Z}_j - d_j^+ \leq Z_j^* \tag{15}$$

In goal programming method, we minimize the deviations  $d_j^+$  using additional constraint equation (15). To solve multi-objective allocation problem (12), the extended lexicographic goal programming has following mathematical model:

$$\begin{aligned} & \text{Minimize} \quad \alpha(\lambda) + (1 - \alpha) \left( \sum_{j=1}^p d_j^+ \right) \\ & \text{Subject to} \quad d_j^+ \leq \lambda \\ & \qquad \qquad \qquad Z_j - d_j^+ \leq Z_j^* \\ & \qquad \qquad \qquad \sum_{h=1}^L c_h n_{hc} + \sum_{h=1}^L t_h n_{hc}^\delta \leq \hat{C} \\ & \qquad \qquad \qquad 2 \leq n_{hc} \leq N_h \\ & \qquad \qquad \qquad n_{hc} \text{ are integers} \\ & \qquad \qquad \qquad n_{hc} \in F, \quad n_{hc} \geq 0, \quad d_j^+ \geq 0 \\ & \qquad \qquad \qquad h = 1, 2, \dots, L, \quad j = 1, 2, \dots, p, \end{aligned} \tag{16}$$

where  $\alpha$  is a constant that can assume minimum value zero and maximum value one.  $d_j^+$  is positive deviational variable.

### 4. Some Other Compromise Allocations

In this section, some other compromise allocations are discussed for the sake of comparison with the proposed allocation.

**4.1. Cochran Compromise Allocation.** Cochran [23] proposed a compromise allocation criteria by averaging the individual optimum allocation  $n_{jh}^*$  ( $j = 1, 2, 3$ ) that is solution to integer nonlinear programming problem (INLPP) (13) over the characteristics.

Cochran's compromise allocation is given by

$$n_h = \frac{1}{p} \sum_{j=1}^p n_{jh}^* \tag{17}$$

**4.2. Khan et al. Compromise Allocation.** Khan et al. [13] compromise allocation is obtained by minimizing the weighted

sum of variances. The mathematical model of Khan et al. [13] compromise allocation is given by

$$\begin{aligned}
 & \text{Minimize } \sum_{j=1}^p \alpha_j Z_j \\
 & \text{Subject to } \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h n_{jh}^\delta \leq \dot{C} \\
 & \qquad \qquad \qquad 2 \leq n_{jh} \leq N_h \\
 & \qquad \qquad \qquad n_{jh} \text{ are integers, } n_{jh} \in F. \\
 & \qquad \qquad \qquad h = 1, 2, \dots, L, \quad j = 1, 2, \dots, p,
 \end{aligned} \tag{18}$$

where  $\alpha_j = \sum_{h=1}^L S_{jh}^2 / \sum_{j=1}^p \sum_{h=1}^L S_{jh}^2$  is the relative weights proposed by Khan et al.

### 5. Numerical Example

The data are taken from agricultural census in Iowa state conducted by National Agricultural Statistics Service, USDA, Washington DC as reported by Khan et al. [16]. We assume that  $c_1 = 12, c_2 = 8, c_3 = 6, c_4 = 10, t_1 = 6, t_2 = 4, t_3 = 3,$  and  $t_4 = 5$ .

Let

- $Y_1$  denote the quantity of corn harvested in 2002;
- $Y_2$  denote the quantity of oats harvested in 2002;
- $X_1$  denote the quantity of corn harvested in 1997;
- $X_2$  denote the quantity of oats harvested in 1997.

The data summary is given as  $\bar{Y}_1 = 474973.90, \bar{X}_1 = 405654.19, \bar{Y}_2 = 1576.25,$  and  $\bar{X}_2 = 2116.70$ . The detailed summary of data is given in Tables 1 and 2.

The allocation problem formulated in multiobjective integer nonlinear programming is

$$\text{Minimize } \begin{pmatrix} Z_1 = \sqrt{\frac{0.000066}{n_1} + \frac{0.000809}{n_2} + \frac{0.001212}{n_3} + \frac{0.000332}{n_4}} \\ Z_2 = \sqrt{\frac{0.000181}{n_1} + \frac{0.009411}{n_2} + \frac{0.023390}{n_3} + \frac{0.000610}{n_4}} \end{pmatrix} \tag{19}$$

Subject to

$$\begin{aligned}
 & 12n_1 + 8n_2 + 6n_3 + 10n_4 + 6n_1^\delta \\
 & \quad + 4n_2^\delta + 3n_3^\delta + 5n_4^\delta \leq \dot{C} \\
 & \qquad \qquad \qquad 2 \leq n_1 \leq 8 \\
 & \qquad \qquad \qquad 2 \leq n_2 \leq 34 \\
 & \qquad \qquad \qquad 2 \leq n_3 \leq 45 \\
 & \qquad \qquad \qquad 2 \leq n_4 \leq 12
 \end{aligned} \tag{20}$$

$n_1, n_2, n_3,$  and  $n_4$  are integers.

#### 5.1. (a) Individual Optimum Allocation Method

5.1.1. Individual Optimum Allocation for Characteristic  $Y_1$ . Consider

Minimize  $Z_1 =$

$$\sqrt{\frac{0.000066}{n_{11}} + \frac{0.000809}{n_{12}} + \frac{0.001212}{n_{13}} + \frac{0.000332}{n_{14}}} \tag{21}$$

subject to

$$\begin{aligned}
 & 12n_{11} + 8n_{12} + 6n_{13} + 10n_{14} + 6n_{11}^\delta \\
 & \quad + 4n_{12}^\delta + 3n_{13}^\delta + 5n_{14}^\delta \leq \dot{C}
 \end{aligned}$$

$$2 \leq n_{11} \leq 8$$

$$2 \leq n_{12} \leq 34$$

$$2 \leq n_{13} \leq 45$$

$$2 \leq n_{14} \leq 12$$

$n_{11}, n_{12}, n_{13},$  and  $n_{14}$  are integers.

$$\tag{22}$$

5.1.2. Individual Optimum Allocation for Characteristic  $Y_2$ . Consider

Minimize  $Z_2 =$

$$\sqrt{\frac{0.000181}{n_{21}} + \frac{0.009411}{n_{22}} + \frac{0.023390}{n_{23}} + \frac{0.000610}{n_{24}}} \tag{23}$$

subject to

$$\begin{aligned}
 & 12n_{21} + 8n_{22} + 6n_{23} + 10n_{24} + 6n_{21}^\delta \\
 & \quad + 4n_{22}^\delta + 3n_{23}^\delta + 5n_{24}^\delta \leq \dot{C}
 \end{aligned}$$

$$2 \leq n_{21} \leq 8$$

$$2 \leq n_{22} \leq 34 \tag{24}$$

$$2 \leq n_{23} \leq 45$$

$$2 \leq n_{24} \leq 12$$

$n_{21}, n_{22}, n_{23},$  and  $n_{24}$  are integers.

TABLE 1: Data summary.

$h$	$N_h$	$W_h$	$S_{y1h}^2$	$S_{x1h}^2$	$S_{y2h}^2$	$S_{x2h}^2$
1	8	0.0808	29267524195.5	21601503189.8	777174.1	1154134.2
2	34	0.3434	26079256582.8	19734615816.7	4987812.9	7056074.8
3	45	0.4545	42362842460.8	27129658750.0	1074510.6	2082871.3
4	12	0.1212	30728265336.9	17258237358.5	388378.5	732004.9

TABLE 2: Data summary.

$h$	$S_{x1y1h}$	$S_{x2y2h}$	$\beta_{1h}$	$\beta_{2h}$	$u'_{1h}$	$u'_{2h}$
1	24360422802.3	902170.6	1.1249	0.7834	0.000066	0.000181
2	22003466630.3	5813439.5	1.1150	0.8239	0.000809	0.009411
3	33367597192.0	1285355.6	1.2300	0.6171	0.001212	0.023390
4	21033769867.3	456991.5	1.2188	0.4243	0.000332	0.000610

$Z_1^*$  and  $Z_2^*$  are coefficients of variation under individual allocation at different values of  $\delta$  and  $\hat{C}$  given in Table 3.

5.2. *Proposed Compromise Allocation.* We used extended lexicographic goal programming model (16) for sample allocation to different strata taking into account two characteristics  $Y_1$  and  $Y_2$ . Consider

$$\text{Minimize } 0.2(\lambda) + (1 - 0.2)(d_1^+ + d_2^+) \quad (25)$$

subject to

$$d_1^+ \leq \lambda$$

$$d_2^+ \leq \lambda$$

$$\sqrt{\frac{0.000066}{n_{1c}} + \frac{0.000809}{n_{2c}} + \frac{0.001212}{n_{3c}} + \frac{0.000332}{n_{4c}}} - d_1^+ \leq Z_1^*$$

$$\sqrt{\frac{0.000181}{n_{1c}} + \frac{0.009411}{n_{2c}} + \frac{0.023390}{n_{3c}} + \frac{0.000610}{n_{4c}}} - d_2^+ \leq Z_2^*$$

$$12n_{1c} + 8n_{2c} + 6n_{3c} + 10n_{4c} + 6n_{1c}^\delta$$

$$+ 4n_{2c}^\delta + 3n_{3c}^\delta + 5n_{4c}^\delta \leq \hat{C}$$

$$2 \leq n_{1c} \leq 8$$

$$2 \leq n_{2c} \leq 34$$

$$2 \leq n_{3c} \leq 45$$

$$2 \leq n_{4c} \leq 12$$

$n_{hc}$  ( $h = 1, 2, 3, 4$ )  $\in F$  are integer

$$d_1^+, d_2^+ \geq 0.$$

(26)

Let  $\widehat{Z}_1$  and  $\widehat{Z}_2$  be the coefficients of variation at various values of constants  $\delta$  and  $\hat{C}$  under proposed allocation given in Table 6.

5.3. *Khan et al. Compromise Allocation.* We have applied model (18) to find compromise allocation proposed by Khan et al. Consider

$$\text{Minimize} \left( \begin{array}{l} 0.999944 \sqrt{\frac{0.000066}{n_1} + \frac{0.000809}{n_2} + \frac{0.001212}{n_3} + \frac{0.000332}{n_4}} \\ + 0.000056 \sqrt{\frac{0.000181}{n_1} + \frac{0.009411}{n_2} + \frac{0.023390}{n_3} + \frac{0.000610}{n_4}} \end{array} \right) \quad (27)$$

subject to

$$12n_1 + 8n_2 + 6n_3 + 10n_4 + 6n_1^\delta$$

$$+ 4n_2^\delta + 3n_3^\delta + 5n_4^\delta \leq \hat{C}$$

$$2 \leq n_1 \leq 8$$

$$2 \leq n_2 \leq 34$$

$$2 \leq n_3 \leq 45$$

$$2 \leq n_4 \leq 12$$

$n_1, n_2, n_3,$  and  $n_4$  are integers.

(28)

The values  $\widehat{Z}_1$  and  $\widehat{Z}_2$  are the coefficients of variation under Khan et al. compromise allocation obtained by solving above model at different values of constants  $\delta$  and  $\hat{C}$  given in Table 5.

TABLE 3: Individual optimum allocation.

$\delta$	$\hat{C}$	Allocation	$n_1$	$n_2$	$n_3$	$n_4$	Used $\hat{C}$	$Z_1^*$	$Z_2^*$	Trace = $Z_1^* + Z_2^*$
0.5	300	$Y_1$	2	9	18	5	298.00	0.01602	0.05057	0.06659
		$Y_2$	2	10	22	2	298.27	0.01830	0.04899	0.06729
1	500	$Y_1$	4	11	21	7	498.00	0.01398	0.04584	0.05982
		$Y_2$	2	13	29	3	498.00	0.01601	0.04271	0.05872
1.5	850	$Y_1$	4	12	15	9	847.56	0.01420	0.04950	0.06376
		$Y_2$	2	14	21	3	833.18	0.01610	0.04560	0.06170
2	1500	$Y_1$	3	10	13	7	1470.00	0.01560	0.05374	0.06934
		$Y_2$	2	10	16	3	1467.00	0.01733	0.05193	0.06926

TABLE 4: Cochran's compromise allocation.

$\delta$	$\hat{C}$	$n_{1c}$	$n_{2c}$	$n_{3c}$	$n_{4c}$	Used $\hat{C}$	$\hat{Z}_1$	$\hat{Z}_2$	Trace = $\hat{Z}_1 + \hat{Z}_2$
0.5	300	2	10	20	4	308.55	0.01605	0.04851	0.06456
1.0	500	3	12	25	5	498.00	0.01429	0.04361	0.05790
1.5	850	3	13	18	6	829.25	0.01438	0.04675	0.06113
2.0	1500	3	10	15	5	1510.00	0.01581	0.05180	0.06761

TABLE 5: Khan et al. compromise allocation.

$\delta$	$\hat{C}$	$n_{1c}$	$n_{2c}$	$n_{3c}$	$n_{4c}$	Used $\hat{C}$	$\hat{Z}_1$	$\hat{Z}_2$	Trace = $\hat{Z}_1 + \hat{Z}_2$
0.5	300	2	9	18	5	298.00	0.01602	0.05057	0.06659
1	500	4	11	21	7	498.00	0.01398	0.04584	0.05982
1.5	850	4	13	15	8	844.91	0.01418	0.04903	0.06321
2.0	1500	4	12	12	5	1495.00	0.01585	0.05386	0.06971

TABLE 6: Proposed compromise allocation.

$\delta$	$\hat{C}$	$n_{1c}$	$n_{2c}$	$n_{3c}$	$n_{4c}$	Used $\hat{C}$	$\hat{Z}_1$	$\hat{Z}_2$	Trace = $\hat{Z}_1 + \hat{Z}_2$
0.5	300	2	11	19	3	299.49	0.01676	0.04879	0.06557
1.0	500	2	16	23	4	495.00	0.01480	0.04299	0.05779
1.5	850	2	11	22	5	848.58	0.01473	0.04583	0.06056
2.0	1500	3	9	15	6	1491.00	0.01575	0.05260	0.06835

### 6. Discussion

In this section, a comparative study of proposed compromise allocation with Cochran compromise allocation, Khan et al. compromise allocation, and individual optimum allocation has been made. The comparison is based on trace of variance-covariance matrix of the estimates of finite population means under compromise allocations. We assume that characteristics are independent; therefore, covariances are zero. Table 3 gives a individual optimum allocation. Tables 4 and 5 give Cochran compromise allocation and Khan compromise allocation as discussed in Section 5. The proposed compromise allocation is given in Table 6.

Table 4 shows that Cochran compromise allocation gives high trace values for  $\delta = 1, 1.5$  as compared to proposed compromise allocation given in Table 6. For  $\delta = 0.5, 2.0$ , Cochran compromise allocation gives slightly low value of trace but is infeasible because corresponding cost exceeds the available cost. Table 5 shows that Khan et al. compromise allocation gives higher trace values than proposed compromise

TABLE 7: PRE of proposed compromise allocation to individual optimum allocation.

$\delta$	$\hat{C}$	$Y_1$	$Y_2$	$\delta$	$\hat{C}$	$Y_1$	$Y_2$
0.5	300	101.59	102.65	1.5	850	105.28	101.88
1.0	500	103.51	101.61	2.0	1500	101.45	101.33

allocation. The performance comparison of proposed compromise allocation relative to individual optimum allocation of one characteristic is used for both characteristics given in Table 7 based on percentage relative efficiency (PRE) expression given as

$$PRE = \frac{T_I}{T_C} \times 100, \tag{29}$$

where  $T_I$  is the value of trace using individual optimum allocation and  $T_C$  is the value of trace using proposed compromise allocation. Table 7 shows that proposed compromise



allocation provides more efficient estimates of population means as compared to individual optimum allocation.

## 7. Conclusion

On the basis of the comparison made in Section 6, we can conclude that the extended lexicographic goal programming approach always secures a feasible solution which is not granted Cochran's compromise method and it provides better results comparative to Khan et al. compromise approach and individual optimum allocation approach from the point of view of efficiency.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] A. A. Tschuprow, "On the mathematical expectation of the moments of frequency distributions in the case of correlated observations," *Metron*, vol. 2, pp. 461–493, 1923.
- [2] J. Neyman, "On the two different aspect of representative method: the method of stratified sampling and method of purposive selection," *Journal of the Royal Statistical Society*, vol. 97, no. 4, pp. 558–625, 1934.
- [3] R. Varshney, M. J. Ahsan, and M. G. M. Khan, "An optimum multivariate stratified sampling design with nonresponse: a lexicographic goal programming approach," *Journal of Mathematical Modelling and Algorithms*, vol. 10, no. 4, pp. 393–405, 2011.
- [4] T. Dalenius, "The problem of optimum stratification-II," *Scandinavian Actuarial*, vol. 33, pp. 203–213, 1950.
- [5] T. Dalenius, *Sampling in Sweden: Contributions to the Methods and Theories of Sample Survey Practice*, Almqvist and Wiksell, Stockholm, Sweden, 1957.
- [6] S. P. Ghosh, "A note on stratified random sampling with multiple characters," *Calcutta Statistical Association Bulletin*, vol. 8, pp. 81–90, 1958.
- [7] J. L. Folks and C. E. Antle, "Optimum allocation of sampling units to strata when there are R responses of interest," *Journal of the American Statistical Association*, vol. 60, pp. 225–233, 1965.
- [8] J. R. Chromy, "Design optimization with multiple objectives," in *Proceeding of the Survey Research Section*, pp. 194–199, American Statistical Association, Washington, DC, USA, 1987.
- [9] J. Bethel, "An optimum allocation algorithm for multivariate surveys," *Proceedings of the Survey Research Section*, pp. 209–212, 1985.
- [10] N. Jahan, M. G. M. Khan, and M. J. Ahsan, "A generalized compromise allocation," *Journal of Indian Statistical Association*, vol. 32, pp. 95–101, 1994.
- [11] N. Jahan, M. G. M. Khan, and M. J. Ahsan, "Optimum compromise allocation using dynamic programming," *Dhaka University Journal of Science*, vol. 49, no. 2, pp. 197–202, 2001.
- [12] E. A. Khan, M. G. M. Khan, and M. J. Ahsan, "Optimum stratification: a mathematical programming approach," *Calcutta Statistical Association Bulletin*, vol. 52, pp. 323–333, 2002.
- [13] M. G. M. Khan, E. A. Khan, and M. J. Ahsan, "An optimal multivariate stratified sampling design using dynamical programming," *Australian & New Zealand Journal of Statistics*, vol. 45, no. 1, pp. 107–113, 2003.
- [14] M. G. M. Khan, E. A. Khan, and M. J. Ahsan, "Optimum allocation in multivariate stratified sampling in presence of non-response," *Journal of the Indian Society of Agricultural Statistics*, vol. 62, no. 1, pp. 42–48, 2008.
- [15] A. H. Ansari, Najmussehar, and M. J. Ahsan, "On multiple response stratified random sampling design," *International Journal of Statistical Sciences, Kolkata, India*, vol. 1, no. 1, pp. 45–54, 2009.
- [16] M. G. M. Khan, T. Maiti, and M. J. Ahsan, "An optimal multivariate stratified sampling design using auxiliary information: an integer solution using goal programming approach," *Journal of Official Statistics*, vol. 26, no. 4, pp. 695–708, 2010.
- [17] R. Varshney, Najmussehar, and M. J. Ahsan, "Estimation of more than one parameters in stratified sampling with fixed budget," *Mathematical Methods of Operations Research*, vol. 75, no. 2, pp. 185–197, 2012.
- [18] J. Beardwood, J. H. Halton, and J. M. Hammersley, "The shortest path through many points," vol. 55, pp. 299–327, 1959, *Mathematical Proceedings of the Cambridge Philosophical Society*.
- [19] F. S. Hiller and G. J. Lieberman, *Introduction to Operation Research*, McGRAW-Hill, New York, NY, USA, 1995.
- [20] R. E. Rosenthal, *A User's Guide Tutorial*, Gams Development Corporation, Washington, DC, USA, 2008.
- [21] C. Romero, "Extended lexicographic goal programming: a unifying approach," *Omega*, vol. 29, no. 1, pp. 63–71, 2001.
- [22] C. Romero, "A general structure of achievement function for a goal programming model," *European Journal of Operational Research*, vol. 153, no. 3, pp. 675–686, 2004.
- [23] W. G. Cochran, *Sampling Techniques*, John Wiley & Sons, New York, NY, USA, 1977.