

Research Article

Nonlinear Stochastic Multiobjective Optimization Problem in Multivariate Stratified Sampling Design

Shokrya Saleh A. Alshqaq,¹ Abdullah Ali H. Ahmadini ,¹ and Irfan Ali ²

¹Department of Mathematics, College of Science, Jazan University, Jazan 45142, P.O. Box 114, Saudi Arabia

²Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh, India

Correspondence should be addressed to Irfan Ali; irfii.st@amu.ac.in

Received 24 December 2021; Accepted 28 June 2022; Published 29 August 2022

Academic Editor: Jiyuan Tao

Copyright © 2022 Shokrya Saleh A. Alshqaq 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.

Decision-making in survey sampling planning is a tricky situation; sometimes it involves multiple objectives, with various decision variables emanating from heterogeneous and homogeneous populations. Dealing with the entire population under study and its uncertain nature becomes a challenging issue for researchers and policymakers. Hence, an appropriate sampling design and optimization methodology are imperative. The study presents a useful discussion on stochastic multiobjective multivariate stratified sampling (MSS) models theoretically, and the concepts are illustrated with numerical examples. Also, it has been found that the linearization of sampling variance in survey sampling does not help determine the optimal sampling allocation problem with minimum variability. Optimal allocation problems under the weighted goal programming, stochastic goal programming, and Chebyshev goal programming methods are also discussed with numerical examples. Finally, the study discussed the linear approximation of the MSS problem with examples. The study is a conceptual and theoretical framework for MSS under a stochastic environment. The numerical data is simulated using the stratifyR package.

1. Introduction

The classical method of optimization for differential calculus is too restrictive and challenging in terms of applicability to many statistical areas in a real-life situation. The lack of numerical algorithms suitable for solving optimization problems poses some severe limitations in this regard and hence led to the utilization of some inefficient statistical procedures in choosing the objective functions and constraints. For decades, a better technique for optimization with broader applicability in statistics, with an increasing computing power able to be implemented has been forthcoming. Mathematical programming is one of such evolving methods with potential application in statistical methodologies. Several optimization techniques have various applications in statistical problems such as designing a specific experiment, extensive survey for data collection, characterizing observed data using a model, drawing inferences about a

population based on sample data, testing of hypothesis, and estimation in the decision-making process [1]. In all the applications, one has to optimize (minimize or maximize) an objective function subject to a set of constraints, such as cost or other input parameters. The sampling problem about the population characteristics remains deriving information on several populations statistically. In a sampling survey, the objectives are to minimize the sampling variance and cost, these depend on the sample size, the sampling scheme, the size of the sampling unit or the scope of the study. Alternatively, a different formulation may be to minimize survey inaccuracy, given that the survey cost is within the budgetary limits. Thus, the research aims at finding a solution for this challenging problem of optimal sample size or sampling scheme that could help in estimating the desired characteristics of a population under prescribed properties. The objective of this research is to successfully formulate the problems of sample surveys as

mathematical programming problems and develop an efficient algorithm or technique to solve them. The objective is to identify existing and future works of allocation problems in survey sampling and to investigate and suggest solutions to them, and also, study the problems in an uncertain environment, i.e., a stochastic. The uncertainty that exists in real life has motivated us to work on this aspect. The problems become more complicated when some or all of the parameters involved are uncertain; it may be either stochastic or fuzzy. The objective is to develop an efficient algorithm to solve such types of sample surveys. The formulated problems may be single-objective or multiobjective. For solving the multiobjective optimization problems, we need to develop efficient algorithms for the formulated problems. The goal programming, fuzzy goal programming, and other new modified version or extended version of these techniques will be used to solve the multiobjective optimization problems.

This study comprises modeling and optimization of different sampling design that helps in providing the efficient allocation of samples simultaneously by achieving the highest accuracy and minimizing the sampling variances. This project provides a useful insight into the decision-making for implementing strategies in different socio-economic sectors for the country based on sampling results. Our contribution to this project is proposing new models and techniques for the sampling scheme to determine the optimal allocation of samples based on which policymakers can suggest what kinds of additional efforts can simultaneously be taken in the planning of socio-economic sectors. The problems related to the case studies are usually complicated, but it has become more complicated when some or all of the input information parameters involved are uncertain. The study provides mathematical optimization problems in survey sampling, which is a powerful tool to make the best policy on national planning and industries. Therefore, the study is an integration of sample surveys, operational research, and computational modeling. Optimal sampling techniques can play an essential role in annual budgeting, income and expenditure forecasting for the preparation of five-year plans in national planning and budgeting. It can also be used in major projects scheduling of national interest, estimation of country's population, agricultural yields, employment, gross national product (GNP), and gross domestic products (GDP) amongst others. Optimal sampling techniques provide the best (optimal) solutions to the problem under study.

There is a need for Statistical Information in modern society now more than any time before, in particular, when data is to be collected periodically to satisfy the information need on a specified set of elements, known as finite populations. Surveys played a significant role in issues relating to real life, if we want to get a sense of a massive population. Sampling is the best tool that gives us a fresh idea about the whole population. A sample survey is one of the most critical data collection modes for meeting this need. Over time, an extensive literature survey sampling has developed into a vast array of

theories, processes, and operations that are used every day throughout the globe. It is appropriate to speak of a worldwide survey industry with different sectors, namely, a government sector, academic sector, a private and mass media sector, a residual sector consisting of ad-hoc and in-house surveys. Optimization is the science of selecting the best among many possible decision alternatives in a complicated real-life situation. The ultimate target of any decision is to either maximize the desired benefit or to minimize the effort (cost or time) required or incurred in a particular course of action. In recent times, more authors formulated different types of sampling problem as a nonlinear mathematical programming problem or integer programming problem and tried to find the best solution [2]. An integer compromise allocation in MSS has been determined using the goal programming approach [3]. A multiobjective all-integer nonlinear model for MSS design considering some travelling costs has been developed, and a compromising solution was obtained using the value function approach, ϵ -constraint method, and distance-based method [4]. Also, uncertainties in the MSS problem have been investigated where some cost parameters were considered as fuzzy parabolic numbers [5]. The authors formulated a fuzzy multi-objective nonlinear programming problem with a quadratic cost function and solved using fuzzy programming. A case of nonresponse in the MSS problem has been studied and modeled as an all-integer multiple objective problem [6]. The solution was sought using four different procedures. Several authors have worked in optimum allocation problems of sampling and parameter estimation, for instance a Multiobjective Integer Nonlinear Programming Problem has been formulated and converted to a single-objective problem using the value function technique [7]. Also, they used Lagrange Multipliers Technique to obtain the continuous sample sizes formula, which approximates the optimal solutions. Similarly, traveling costs within strata has been considered, and a multiple objective nonlinear stochastic programming problem was formulated for finding a compromise allocation in the sample survey [8]. The problem was solved using D_1 -distance, goal programming, and the Chebyshev approximation technique. A problem of estimating p -population means considering nonresponse and nonlinear cost functions have been investigated, and solution procedures suggested using lexicographic goal programming [9]. The dynamic programming technique has been employed in proposing an efficient methodology for optimum stratum boundaries and determining optimum sample size in survey variables under the Neyman allocation [10]. A multiple pooling of the standard deviations of the estimates in an MSS for more than three strata has been studied and formulated as a multiobjective (MO) problem, which was solved using fuzzy programming [11]. Others considered compromise allocation problems under stratified samples with two-stage randomized and multiresponse models [12, 13]. An optimum allocation problem in MSS has been considered as an integer nonlinear stochastic programming and

solved with five different techniques [14]. The authors suggested the use of coefficient of variations instead of variances. Also, the MSS problem has been studied with stochastic optimal design [15, 16], with flexible goals [17], and with integer solution [18].

Several mathematical models have been designed based on multiobjective optimization for solving different aspects of human endeavors. For instance, a mixed-integer linear programming (MILP) model has been developed for addressing a closed-looped supply chain network problem during the coronavirus pandemic. The study considered different items such as recycling, reusing, quarantine, collection, distribution, production, supply, and location within a multiperiod, multiechelon, and multiproduct supply chain [19]. A multiple criteria decision-making tool has been used in determining the supply chain performance in a petrochemical industry incorporating sustainable strategies [20]. An optimization method has been designed to optimize the distribution and allocation of scarce resources amongst individuals during a crisis, based on credibility theory and a harmony search algorithm considering random simulation [21]. A scheduling problem has been studied, and a mathematical model developed with a view to obtain near-optimal solution using meta-heuristic algorithms (MHA) [22]. Multiobjective optimization has been widely used in different sectors considering diverse applications and scenarios. For instance, robust optimization with artificial intelligence (AI) has been hybridized as multiobjective optimization applied to the product portfolio problem [23]. Location, allocation, and routing problem have been studied with the help of an improved harmony search algorithm [24]. Another important application area is that of dairy product's demand prediction, where an integrated approach based on AI and novel MHA has been used in achieving the desired future demands [25]. The MOOP has been used to formulate socio-economic and environmental issues related to sustainable development goals in several countries, such as India [26], Nigeria [27], Saudi Arabia [28], and other areas, such as municipal waste management system [29].

1.1. Organization of the Paper. The introduction of the subject matter, the background of the study, the literature review, and paper organization are presented in section 1. In Section 2, the multiobjective MSS techniques are presented. Section 3 provides single-objective stochastic MSS models. Section 4 discussed the MO stochastic MSS models. The linear approximation of MSS is discussed in Section 5. Section 7 concludes the article.

2. Multiobjective Multivariate Stratified Sampling

Let N be the size of the population partitioned into L strata each of sizes N_h , $h = 1, 2, \dots, L$. Suppose p is considered as characteristics ($p \geq 2$) that are measured on each unit of the population, and the interest is on p -population characteristics estimation. Let n_h , $h = 1, 2, \dots, L$ be the units taken randomly from the stratum without replacement.

2.1. Sampling Variance Function. The population mean (\bar{Y}_j) for the j^{th} character is

$$\begin{aligned} \bar{Y}_j &= \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} y_{jhi} \\ &= \sum_{h=1}^L W_h \bar{y}_{jh}; j = 1, 2, \dots, p, \end{aligned} \quad (1)$$

where $W_h = N_h/N$ is stratum weights and $\bar{y}_{jh} = (1/N_h) \sum_{i=1}^{N_h} y_{jhi}$ is a stratum means.

The sampling means of the j^{th} character is given as

$$\bar{y}_{j,st} = \frac{1}{n} \sum_{h=1}^L \sum_{i=1}^{n_h} y_{jhi}; j = 1, 2, \dots, p. \quad (2)$$

The sampling variance of the estimator of the mean for the j^{th} characteristic is given as follows:

$$\bar{y}_{j,st} = \sum_{h=1}^L \left(\frac{1}{n} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2; j = 1, 2, \dots, p, \quad (3)$$

where $S_{jh}^2 = (1/N_h - 1) \sum_{i=1}^{N_h} (y_{jhi} - \bar{Y}_{jh})^2$ are stratum variances and y_{jhi} is the value of the i^{th} unit in the h^{th} stratum for the j^{th} characteristics ($j = 1, 2, \dots, p$ and $h = 1, 2, \dots, L$).

2.2. Sampling Cost Function. In survey sampling, when enumeration cost, traveling cost, and labor costs are high [30, 31], the total cost function is defined as follows:

$$C = \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} + \omega \sum_{h=1}^L \frac{n_h}{\lambda}, \quad (4)$$

where c_h is the per-unit cost of measurement in the h^{th} stratum, t_h is the travel cost for enumerating for a unit of the j -th character in the h -th stratum, and ω is the cost of labor for a unit time. The labor time is available with respect to the time for a sampling unit within a stratum and follows an exponential distribution with rate λ .

If in (4) the labor expenses are not significant, then we have a quadratic cost function given in (5).

$$C = \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h}. \quad (5)$$

If in (5) the traveling cost is not significant, then we have a linear cost function with a fixed overhead cost of sampling (C_0) given in (6).

$$C = \sum_{h=1}^L c_h n_h + C_0. \quad (6)$$

In a particularly important case of (6), if $c_h = c$, that is, if the per-unit cost in all strata is assumed to be the same, then the enumeration cost terms become constant, and the fixed cost for optimum allocation reduces to fixed sample size optimum allocation and is called Neyman [32].

2.3. *Multiobjective Optimization Problem.* The multiobjective optimization problem (MOOP) using the above-given definitions can be defined as

$$\left. \begin{aligned} \min f_j(n_h) &= \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 & \text{(i)} \\ \text{subject to :} & & \\ n_h &\in X & \text{(ii)} \\ 2 \leq n_h &\leq N_h & \text{(iii)} \end{aligned} \right\} j = 1, \dots, p, \quad (7)$$

where $X = X_1$, or X_2 , or X_3 is the feasible space of the problem and it is defined as

$$\begin{aligned} X_1 &= \left\{ n_h \in R^n \mid \sum_{h=1}^L c_h n_h \leq C, 2 \leq n_h \leq N_h, \text{ for } h = 1, 2, \dots, L \right\}, \\ X_2 &= \left\{ n_h \in R^n \mid \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C, 2 \leq n_h \leq N_h, \text{ for } h = 1, 2, \dots, L \right\}, \\ X_3 &= \left\{ n_h \in R^n \mid \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} + \omega \sum_{h=1}^L \frac{n_h}{\lambda} \leq C, 2 \leq n_h \leq N_h, \text{ for } h = 1, 2, \dots, L \right\}. \end{aligned} \quad (8)$$

2.4. *Weighted Goal Programming for Optimum Allocation Problem.* In the goal programming approach, the p objective functions goals have been identified by solving the problem for individual j^{th} objective function ignoring the other ($j - 1$) objective functions with the feasible set constraints. The general form of goal programming is

$$\min_{n_h \in X} d(f(n_h), \hat{f}), \quad (9)$$

where $\hat{f} = (\hat{f}_1, \hat{f}_2, \dots, \hat{f}_p)$ is the targeted goal vector which has been obtained at individual solutions and $d(f(n_h), \hat{f})$ is the distance function between $d(f(n_h))$, and \hat{f} in some selected norm. The function (7) in the l_1 norm is given as

$$\min_{n_h \in X} d_1(f(n_h), \hat{f}) = \sum_{j=1}^p \left| \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j \right|. \quad (10)$$

The weighted l_1 norm is

$$\min_{n_h \in X} d_1^w(f(n_h), \hat{f}) = \sum_{j=1}^p w_j \left| \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j \right|, \quad (11)$$

where $w_j \geq 0$ is the weight assigned to the j th objective function.

The goal programming can be converted to a single-objective optimization problem by introducing the auxiliary variables

$$\begin{aligned} d_j^+ &= \frac{\left\{ \left| \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j \right| + \left(\sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j \right) \right\}}{2}, \\ d_j^- &= \frac{\left\{ \left| \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j \right| - \left(\sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j \right) \right\}}{2}. \end{aligned} \quad (12)$$

Finally, the weighted goal programming model is

$$\left. \begin{aligned}
 & \min \sum_{j=1}^p w_j (d_j^+ + d_j^-) && \text{(i)} \\
 & \text{subject to :} \\
 & \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - d_j^+ + d_j^- = \hat{f}_j (n_h^*), \forall j && \text{(ii)} \\
 & n_h \in X && \text{(iii)} \\
 & d_j^+ \cdot d_j^- = 0, \forall j && \text{(iv)} \\
 & d_j^+ \geq 0, d_j^- \geq 0, \forall j && \text{(v)}
 \end{aligned} \right\} \quad (13)$$

In (13) (i), the definition of d_j^+ & d_j^- is

$$d_j^+ = \begin{cases} \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 - \hat{f}_j, & \text{if } \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 \geq \hat{f}_j, \\ 0, & \text{if } \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2 < \hat{f}_j, \end{cases} \quad (14)$$

$$d_j^- = \begin{cases} \hat{f}_j - \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2, & \text{if } \hat{f}_j \geq \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2, \\ 0, & \text{if } \hat{f}_j < \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_{jh}^2, \end{cases} \quad (15)$$

where d_j^+ and d_j^- are overachievement and underachievement functions respectively for the j^{th} goal value. It is further noted from (13) (iv) that d_j^+ and d_j^- can never be achieved simultaneously. It means that when overachievement is more significant than zero, then underachievement functions will be zero and vice versa. If the objective is maximization type function and hence underachievement function is not desirable. For this situation $w_j^+ = 0$ and $w_j^- = 1$, and equation (13) (i) objective function is reduced to $\min \sum_{j=1}^p w_j^- d_j^-$, where w_j^+ and w_j^- are the weight assigned to overachievement and underachievement functions, respectively. Conversely, for minimization type objective functions, the underachievement function is not desirable, that is, $w_j^+ = 1$ and $w_j^- = 0$, and (13) (i) objective function is reduced to $\min \sum_{j=1}^p w_j^+ d_j^+$.

3. Single-Objective Stochastic Multivariate Stratified Sampling Models

Deciding under uncertainty is challenging and unavoidable in most real-life problematic situations. The problems are

mainly aim to optimize a set of function (s) under uncertain conditions by the decision-maker (s). If some or all of the constraints' parameters are unknown and are considered random, then such an optimization function becomes a stochastic programming problem.

Any modeling framework that optimized a problem under uncertainty can be viewed as a stochastic programming problem. The ultimate goal of these modeling types is to obtain a set of solution(s) that is feasible and optimal in some kind of set of data. Most of the models under this category involve parameters that follow probability distributions and can be known in advance or estimated using established procedures.

In general terms, stochastic programming can also be called probabilistic programming if some or all data of the optimization function follow probability distributions. In other words, variables that behave randomly in optimization problems can be regarded as stochastic or probabilistic as the case may be. Charnes & Cooper [33] developed and converted the chance-constrained programming technique into its equivalent deterministic nonlinear constraints. Many

authors have discussed the stochastic optimization problem. Among them are Prekopa [34], and Charnes and Cooper [35]. In the context of response surface methodology, Diaz Gracia et al. [36] had studied the problem under several stochastic optimization techniques. Diaz Gracia and Ramos-Quiroga [15, 16] formulated the problem of stratified

sampling. In this problem, the authors have considered the sampling variances as a random variable. The sampling variances s_h^2 have an asymptotically normal distribution. The given problem converts into an equivalent deterministic problem by using a modified-E model.

$$\left. \begin{aligned} \min &= k_1 \left[\sum_{h=1}^L \frac{W_h^2}{(n_h - 1)} s_h^2 - \sum_{h=1}^L \frac{W_h^2}{N} \left(\frac{n_h}{n_h - 1} \right) s_h^2 \right] \\ &+ k_2 \left[\sum_{h=1}^L \frac{W_h^4}{n_h (n_h - 1)^2} \left(C_{yh}^4 - (s_h^2)^2 \right) - \sum_{h=1}^L \frac{W_h^2}{N^2} \left(\frac{n_h}{(n_h - 1)^2} \left(C_{yh}^4 - (s_h^2)^2 \right) \right) \right]^{1/2} \quad (\text{i}) \\ \text{subject to :} \\ \sum_{h=1}^L c_h n_h + c_o &= C \quad (\text{ii}) \\ n_h &\in N, h = 1, 2, \dots, L. \quad (\text{iii}) \end{aligned} \right\} \quad (16)$$

Again, Diaz Gracia and Garay Tapia [14] formulated the problem of stratified sampling. In this problem, authors have considered stochastic programming for minimizing the cost function under the constraint to a known bound for the

estimated variance of the mean. The following problem converts into an equivalent deterministic problem by using chance constraints;

$$\left. \begin{aligned} \min &= \sum_{h=1}^L c_h n_h + c_o \quad (\text{i}) \\ \text{subject to :} \\ \left[\sum_{h=1}^L \frac{W_h^2}{(n_h - 1)} s_h^2 - \sum_{h=1}^L \frac{W_h^2}{N} \left(\frac{n_h}{n_h - 1} \right) s_h^2 \right] &+ k_\alpha \\ \left[\sum_{h=1}^L \frac{W_h^4}{n_h (n_h - 1)^2} \left(C_{yh}^4 - (s_h^2)^2 \right) - \sum_{h=1}^L \frac{W_h^2}{N^2} \left(\frac{n_h}{(n_h - 1)^2} \left(C_{yh}^4 - (s_h^2)^2 \right) \right) \right]^{1/2} &\leq V_o \quad (\text{ii}) \\ n_h &\in N, h = 1, 2, \dots, L. \quad (\text{iii}) \end{aligned} \right\} \quad (17)$$

where V_o is a known non-negative constant and K_α is the value of the standard normal variable.

4. Multiobjective Stochastic Multivariate Stratified Sampling Models

In this section, we discussed the various nonlinear optimization sampling models under stochastic approaches. For

instance, a problem of attaining several goals targets under probabilistic intervals was formulated as a linear stochastic model [37]. Problems involving stochastic MO have been analyzed considering different efficient concepts and establishing the relationships between the identified concepts [38]. A Multivariate Stratified random Sampling has been investigated where the asymptotic normality of the optimal solution was established, as well as the perturbation

effect of the stratum variance on the optimal solution [39]. Similarly, a problem of estimating several population means in an MSS design has been investigated [40]. The authors formulated an all-integer nonlinear model and proposed the solution using dynamic programming concepts with numerical illustrations. A multiobjective goal optimization in stratified sampling design was conducted by trading off between the sampling cost and its variance [41]. More than a single parameter estimation in a stratified sampling problem has been studied with a fixed budget and nonlinear cost [42]. Beale described the convex function minimization as a linear programming problem considering the coefficients as random variables [43].

Consider a multiobjective nonlinear programming problem (MNLPP) is

$$\left. \begin{aligned} \min f_j &= \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) & (i) \\ \text{subject to :} & \\ \sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} &\leq C & (ii) \\ 2 \leq n_h &\leq N_h & (iii) \end{aligned} \right\} j = 1, \dots, p. \quad (18)$$

Then, Equation (18) has been defined under the stochastic assumption given as the following stochastic nonlinear programming problem (SNLPP) for the p characteristics.

$$\left. \begin{aligned} \min &= P \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \leq f_j \right) \geq \beta & (i) \\ \text{subject to :} & \\ P \left(\sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} \leq C \right) &\geq \beta & (ii) \\ 2 \leq n_h &\leq N_h & (iii) \end{aligned} \right\} j = 1, \dots, p. \quad (19)$$

Definition 1. A point is called feasible if and only if the probability measure of the event $g_j(n_h, \xi) \leq 0$, $j = 1, \dots, p$ is at least β . Or equivalently the constraints will be violated at most $(1 - \beta)$ times. The joint chance constraint is separately defined and is referred to as a separate chance constraint. That is,

$$P\{g_j(n_h, \xi) \leq 0\} \geq \beta_j, j = 1, \dots, p. \quad (20)$$

Now, by applying minimax chance constrained programming, (18) is as follows:

$$\left. \begin{aligned} \min_{n_h} \max f_j & & (i) \\ \text{subject to :} & \\ P \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \leq f_j \right) &\geq \beta & (ii) \\ P \left(\sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} \leq C \right) &\geq \beta & (iii) \\ 2 \leq n_h &\leq N_h & (iv) \end{aligned} \right\} j = 1, \dots, p. \quad (21)$$

where β is the predetermined confidence level and $\min f_j$ is the variance term.

We can also formulate a stochastic goal programming for the problem defined in (21) with target goal values.

$$\left. \begin{aligned} \min_{n_h} \sum_{j=1}^p \delta_j^+ & & (i) \\ \text{subject to :} & \\ P \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) - \delta_j^+ \leq f_j \right) &\geq \beta & (ii) \\ P \left(\sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} \leq C \right) &\geq \beta & (iii) \\ 2 \leq n_h &\leq N_h & (iv) \end{aligned} \right\} j = 1, \dots, p. \quad (22)$$

Remark 1. (i) The stochastic objective constraints

$$P \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \leq f_j \right) \geq \beta. \quad (23)$$

coincide with the form in (20) by defining

$$g_j(n_h, \xi) = \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) - f_j. \quad (24)$$

(ii) The stochastic goal constraints

$$P \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) - f_j \geq \delta^+ \right) \geq \beta. \quad (25)$$

coincide with the form in (20) by defining

$$g_j(n_h, \xi) = f_j - \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) + \delta^+. \quad (26)$$

where δ^+ is an overachievement goal.

(iii) The stochastic problem constraints

$$P \left(\sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} \leq C \right) \geq \beta. \quad (27)$$

coincide with the form in (20) by defining

$$g_j(n_h, \xi) = \sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} - C. \quad (28)$$

(iv) The stochastic problem constraints

$$P \left(\sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} \leq C \right) \geq \beta. \quad (29)$$

coincide with the form in (20) by defining

$$g_j(n_h, \xi) = C - \sum_{h=1}^L c_n h_n + \sum_{h=1}^L t_n \sqrt{n_h} + \delta^+. \quad (30)$$

where overachievement goal.

(v) For a continuous random variable ξ , the value $Pk_\beta \leq \xi = 1 - \Phi(k_\beta)$ holds always, and we have $k_\beta = \Phi^{-1}(1 - \beta)$, where Φ^{-1} is the inverse function of Φ .

4.1. Conversion of Stochastic Inequalities to Equivalent Deterministic. In (18) (i), the term s_{jh}^2 is assumed to be a random variable. In practice, some approximations of these

parameters, which are known from some preliminary or recent survey, may be used. The concept of limiting the distribution of the sample variances in a sampling problem is used in [39], considering the random variable ξ_h defined as

$$\xi = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{jhi} - \bar{Y}_{jh})^2, \quad (31)$$

where $\bar{Y}_{jh} = (1/N_h) \sum_{i=1}^{N_h} y_{jhi}$. Note that ξ_h has an asymptotically normal distribution with mean

$$E(\xi_h) = \frac{n_h}{n_h - 1} S_{jh}^2, \quad (32)$$

and variance

$$V(\xi_h) = \frac{n_h}{(n_h - 1)^2} [C_{jh}^4 - (S_{jh}^2)^2], \quad (33)$$

respectively, where

$$C_{jh}^4 = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_{jhi} - \bar{Y}_{jh})^4, \quad (34)$$

is the fourth central moment of j^{th} character in the h^{th} stratum. The sequence of sample variances is given by

$$S_{jh}^2 = \xi_h - \frac{n_h}{N_h - 1} \sum_{i=1}^{N_h} (\bar{y}_{jhi} - \bar{Y}_{jh})^2, \quad (35)$$

where $n_h/N_h - 1 \rightarrow 1$ and $(\bar{y}_{jhi} - \bar{Y}_{jh})^2 \rightarrow 0$ in probability and hence under the asymptotically normal property $S_{jh}^2 \xrightarrow{a} N(E(\xi), \text{Var}(\xi))$, $h = 1, 2, \dots, L$ are independents.

Based on the above discussion, the multivariate stratified sampling variance function with the following expected function and variance function

$$\left. \begin{aligned} E \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right) &= \sum_{h=1}^L \left(\frac{W_h^2 E(\xi_h)}{n_h} - \frac{W_h^2 E(\xi_h)}{N_h} \right) \\ &= \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h - 1} \right) - \left(\frac{W_h^2 S_{jh}^2}{N_h} \right) \frac{n_h}{n_h - 1} \cong \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h - 1} \right) - \left(\frac{W_h^2 S_{jh}^2}{N_h} \right) \end{aligned} \right\}, \quad (36)$$

as n_h is sufficiently large,

$$\left. \begin{aligned} V \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right) &= \sum_{h=1}^L \left(\frac{W_h^4 V(\xi_h)}{n_h^2} - \frac{W_h^4 V(\xi_h)}{N_h^2} \right) \\ &= \sum_{h=1}^L \left(\frac{W_h^4 (C_{jh}^4 - (S_{jh}^2)^2)}{n_h (n_h - 1)} \right) - \left(\frac{W_h^4 (C_{jh}^4 - (S_{jh}^2)^2)}{N_h^2} \right) \frac{n_h}{(n_h - 1)^2}, \\ &\cong \sum_{h=1}^L \left(\frac{W_h^4 (C_{jh}^4 - (S_{jh}^2)^2)}{n_h (n_h - 1)} \right) - \left(\frac{W_h^4 (C_{jh}^4 - (S_{jh}^2)^2)}{N_h^2} \right) \end{aligned} \right\}, \quad (37)$$

and as n_h is sufficiently large.

Theorem 1. Assume that the stochastic vector ζ degenerates to a random variable ξ with a probability distribution Φ , and the function $g_j(n_h, \zeta)$ has the form $g_j(n_h, \zeta) = \bar{w}_j(n_h) - \xi$. Then, $P\{g_j(n_h, \zeta) \leq 0\} \geq \beta$ iff $\bar{w}_j(n_h) \leq k_\beta$, where k_β is the maximal value such that $P\{k_\beta \leq \xi\} \geq \beta$. Note that the probability $P\{k_\beta \leq \xi\}$ will be increased if k_β is replaced with a smaller number.

Theorem 2. Assume that the stochastic function and $g(n_h, \xi)$ has the form

$$g_j(n_h, \xi) = \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) - f_j. \quad (38)$$

If s_{jh}^2 are assumed to be independent normally distributed random variables, then $\{g(n_h, \xi) \leq 0\} \geq \beta$ if and only if

$$E \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right) + \Phi^{-1}(\hat{\beta}) \sqrt{V \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right)} \leq f_j, \quad (39)$$

where Φ is the standardized normal distribution function.

Proof. In the probability model of survey sampling, the probability that the sampling variance value is smaller or equal to an absolute goal value is maximized. That is,

$$P \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right) \leq f_j^*, \quad (40)$$

where f_j^* is the minimum target goal value for the j th objective function.

Recall the (18) (i), independently normally distributed random variables s_{jh}^2 . Moreover, covariance terms will be zero, and only the variance terms will be remain.

$$\begin{aligned} & \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \leq f_j^* \\ &= P \left(\frac{f_j(n_h) - E(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}{\sqrt{V(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}} \leq \frac{f_j^* - E(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}{\sqrt{V(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}} \right) \\ &= P \left(\eta \leq \frac{f_j^* - E(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}{\sqrt{V(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}} \right) = \Phi \left(\frac{f_j^* - E(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}{\sqrt{V(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}} \right), \end{aligned} \quad (41)$$

where η is the standardized normally distributed random variable.

From the fact that

$$\Phi \left(\frac{f_j^* - E(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}{\sqrt{V(\sum_{h=1}^L ((W_h^2 S_{jh}^2/n_h) - W_h^2 S_{jh}^2/N_h))}} \right) \geq \beta, \quad (42)$$

it is equivalent to

$$\Phi^{-1} \beta \sqrt{V \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right)} \leq f_j^* \quad (43)$$

$$- E \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right),$$

and from (43), the maximum of β is searched in the interval $(0, 1)$. The following function is defined for convenience:

$$f_o(n_h; \beta) = E \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right) + \Phi^{-1} \beta \sqrt{V \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h} \right) \right)} - f_j^*. \quad (44)$$

Here, we assume that for an optimal solution (n_h^*, β^*) to the problem (44), $\beta^* > 0.5$ holds. From this assumption, $\Phi^{-1}(\beta) > 0$, and then for a fixed value of β^* , it follows that $f_o(n_h^*; \beta^*)$ is convex. \square

Proposition 1. Let (n_h^*, β^*) be an optimal solution to the problem (21) with a target value f_j larger than

$$f_{oj} = E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right), \quad (45)$$

i.e., $f_j > f_{oj}$. Then, $\beta^* > 0.5$ holds.

$$E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right) + \Phi^{-1}(\beta) \sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)} = f_j. \quad (48)$$

Therefore, $(\hat{n}_h, \hat{\beta})$ is a feasible solution to the problem (44) with the target value f_j . Since the value (n_h^*, β^*) is the optimal solution of (44), it holds.

Theorem 3. Assume that the stochastic vector $\xi = (c_1, c_2, \dots, c_L, t_1, t_2, \dots, t_L, C)$ and the function $g(n_h, \xi)$ has the form $g(n_h, \xi) = \sum_{h=1}^L L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} - C$. If c_i and C are assumed to be independent normally distributed random variables, then $\{g(n_h, \xi) \leq 0\} \geq \beta$ if and only if

$$\Phi^{-1}(\beta) \leq -\frac{\left(\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} - E(C)\right)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}}, \quad (49)$$

or equivalently

$$\frac{\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} + \Phi^{-1}(\beta)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}} \leq E(C), \quad (50)$$

where Φ is the standardized normal distribution function.

Proof. Let chance constraint, that is,

$$P\left(\sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C\right) \geq \beta. \quad (51)$$

It is assumed that C, c_h , and t_h are normally distributed random variables. Moreover, assume that all are independent of each other. We note that

$$\frac{g_j(n_h, \xi) - \left(\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} - E(C)\right)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}}. \quad (52)$$

Proof. From the condition $f_j > f_{oj} > 0$, of the preposition, $f_j - f_{oj} > 0$, and then we have

$$\Phi\left(\frac{f_j^* - E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)}{\sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)}}\right) > 0.5. \quad (46)$$

Letting

$$\hat{\beta} = \Phi\left(\frac{f_j^* - E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)}{\sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)}}\right), \quad (47)$$

we have

Equation (52) is the standard normal random variable $N(0, 1)$, and it follows

$$\begin{aligned} & P\left(\sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C\right) \\ &= P\left(\frac{g_j(n_h, \xi) - \left(\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} - E(C)\right)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}}\right) \\ &\leq \frac{\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} - E(C)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}}, \end{aligned} \quad (53)$$

or equivalent to

$$P\left(\eta \leq -\frac{\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} - E(C)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}}\right) \geq \beta, \quad (54)$$

where η is the standardized, normally distributed random variable. The above-given constraint holds if and only if

$$\Phi^{-1}(\beta) \leq -\frac{\left(\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} - E(C)\right)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}}. \quad (55)$$

Hence, the chance constraint (12) (ii) can be transformed into

$$\frac{\sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} + \Phi^{-1}(\beta)}{\sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h + V(C)}} \leq E(C). \quad (56)$$

It can be further assumed that only c_h and t_h are normally distributed random variables and independent to each other where the total budget for the survey is fixed. The same procedure as discussed above will be followed and hence the cost constraint defined in (56) will be defined as follows:

$$\sum_{h=1}^L E(c_h)n_h + \sum_{h=1}^L E(t_h)\sqrt{n_h} + \Phi^{-1}(\beta) \sqrt{\sum_{h=1}^L V(c_h)n_h^2 + \sum_{h=1}^L V(t_h)n_h} \leq C. \quad (57)$$

□

$$\min_{n_h} = \sum_{j=1}^p \delta_j^+, \quad (i)$$

Subject to :

$$E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right) + \Phi^{-1}(\beta) \sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)} - \delta_j^+ \leq f_j^* \quad (ii)$$

$$\sum_{h=1}^L E(c_h)n_h + \sum_{h=1}^L E(t_h)\sqrt{n_h} + \Phi^{-1}(\beta) \sqrt{\sum_{h=1}^L V(c_h)n_h^2 + \sum_{h=1}^L V(t_h)n_h} \leq C \quad (iii)$$

$$2 \leq n_h \leq N_h \text{ and } \delta_j^+ \geq 0 \quad (iv).$$

The individual sampling variance goal value can be obtained using the following define equation:

$$\min_{j=1,2,\dots,p} f_j = E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right) + \Phi^{-1}(\beta) \sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)},$$

Subject to :

$$\sum_{h=1}^L E(c_h)n_h + \sum_{h=1}^L E(t_h)\sqrt{n_h} + \Phi^{-1}(\beta) \sqrt{\sum_{h=1}^L V(c_h)n_h^2 + \sum_{h=1}^L V(t_h)n_h} \leq C$$

$$2 \leq n_h \leq N_h.$$

(59)

4.3. Chebychev Goal Programming Sampling Model. In this method, first, we set goals for each objective that we want to attain. Let some goals $g = (g_1, g_2, \dots, g_k)'$ be identified for the objective functions $f = (f_1(x_h), f_2(x_h), \dots, f_k(x_h))'$. Let the objective functions $f = (f_1^*(x_h), f_2^*(x_h), \dots, f_k^*(x_h))'$ be defined as close as possible to goals $g = (g_1, g_2, \dots, g_k)'$.

The difference between $f = (f_1^*(x_h), f_2^*(x_h), \dots, f_k^*(x_h))'$ and $g = (g_1, g_2, \dots, g_k)'$ is defined as the deviation function $D(f(n_h), g)$. In a sampling optimization problem, the aim is to find an $n_h^* \in X$, which minimizes $D(f(n_h), g)$, that is,

4.2. Stochastic Goal Programming Sampling Variance Model. In light of the above discussion, the problem formulated in (57) is transformed equivalently as follows:

$$n_h^* = \arg \min_{n_h \in X} D(f(n_h), g), \quad (60)$$

where

$$D(f(n_h), g) = \max D\{f_1(n_h), g_1, \dots, f_k(n_h), g_k\}, \quad (61)$$

is the maximum deviation of individual goals. Finally, a preferred solution is then defined as one that minimizes the maximum deviations from the goals. In light of the above discussion, the problem formulated in (55) is transformed equivalently with an auxiliary variable δ as follows:

$$\left. \begin{aligned}
 & \min_{n_h} \delta, \quad (i) \\
 & \text{subject to :} \\
 & E\left(\sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right) + \Phi^{-1} \beta \sqrt{V\left(\sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)} \quad (ii) \\
 & \sum_{h=1}^L E(t_h) \sqrt{n_h} + \Phi^{-1} \beta \sqrt{\sum_{h=1}^L V(c_h) n_h^2 + \sum_{h=1}^L V(t_h) n_h} - C \leq \delta \quad (iii) \\
 & 2 \leq n_h \leq N_h \delta \text{ and } \delta \geq 0 \quad (iv)
 \end{aligned} \right\} \quad (62)$$

4.4. Stochastic Sampling Cost Model. Similar to the sampling variance model, in a sampling cost model, we minimize the

objective cost function subject to variance constraints. That is,

$$\left. \begin{aligned}
 & \min_{n_h} = \sum_{h=1}^L E(c_h) n_h + \sum_{h=1}^L E(t_h) \sqrt{n_h} + \Phi^{-1} (\beta) \sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)} \\
 & \text{Subject to :} \\
 & E\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right) + \Phi^{-1} (\beta) \sqrt{V\left(\sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{n_h} - \frac{W_h^2 S_{jh}^2}{N_h}\right)\right)} \leq f_j^* \\
 & 2 \leq n_h \leq N_h \text{ and } j = 1, 2, \dots, p.
 \end{aligned} \right\} \quad (63)$$

where C^* is the target goal value.

5. Linear Approximation of Multivariate Stratified Sampling Problem

The objective function in (7) f_j is linearized at the individual optimum points [44]. Thus, for $j = q$ at the point $n_q^* = (n_{q1}^*, n_{q2}^*, \dots, n_{qh}^*)$, f_q' may be approximated by the linear function with n_h as

$$f_q' \cong f_q(n_{qh}^*) + \nabla' f_q(n_{qh}^*)(n_h - n_{qh}^*), \quad (64)$$

where f_q' is the variation term and $\nabla' f_q(n_{qh}^*)$ is the value of the vector of partial derivatives of f_q with respect to n_{qh} ($h = 1, 2, \dots, L$) at the point n_{qh}^* given as follows:

$$\nabla' f_q'(n_{qh}^*) = \left[-\left(\frac{W_1^2 S_{q1}^2}{(n_{q1}^*)^2}\right), -\left(\frac{W_2^2 S_{q2}^2}{(n_{q2}^*)^2}\right), \dots, -\left(\frac{W_L^2 S_{qL}^2}{(n_{qL}^*)^2}\right) \right]. \quad (65)$$

This gives

$$\begin{aligned}
 & \nabla' f_q'(n_{qh}^*)(n_h - n_{qh}^*) \\
 & = \left\{ \left(\sum_{h=1}^L \frac{W_h^2 S_{qh}^2}{n_{qh}^*} \right) - \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{qh}^2}{n_{qh}^*} \right) n_h \right) \right\}, \\
 & f_q' \cong 2 \left\{ \sum_{h=1}^L \frac{W_h^2 S_{qh}^2}{n_{qh}^*} \right\} - \left\{ \sum_{h=1}^L \left(\frac{W_h^2 S_{qh}^2}{n_{qh}^*} \right) n_h \right\} = f_q' \text{ (say)}. \quad (66)
 \end{aligned}$$

After dropping the constant terms in the linear objective function, the NLPP (6) can be approximated, and the final problem is equivalent to maximizing $(-f_k')$. That is,

$$\left. \begin{aligned}
 & \max f_j = \sum_{h=1}^L \left(\frac{W_h^2 S_{jh}^2}{(n_{jh}^*)^2} \right) n_h, \quad (i) \\
 & \text{subject to :} \\
 & \sum_{h=1}^L c_n t_n \leq C \quad (ii) \\
 & 2 \leq n_h \leq N_h \quad (iii)
 \end{aligned} \right\} j = 1, 2, \dots, p. \quad (67)$$

TABLE 1: Data for two characteristics and four strata.

h	N_h	S_{h1}^2	S_{h2}^2	W_h	c_h
1	174	268535.9	136457.3	0.11233	9
2	196	322381.3	146465.2	0.12653	8
3	676	355325.6	127272	0.43641	12
4	503	324026.4	141460.7	0.32473	11

TABLE 2: Data for two characteristics and four strata.

h	N_h	W_h	S_{h1}^2	S_{h2}^2	C_{h1}^4	C_{h2}^4	$E(c_h)$	$V(c_h)$	$E(t_h)$	$V(t_h)$
1	174	0.11233	21.12788	6.393020	1812.391	106.4593	9	2.10	2.00	0.35
2	196	0.12653	24.56889	5.326552	1635.366	66.60126	8	1.75	2.45	0.75
3	676	0.43641	22.81545	6.220776	1517.975	112.4234	12	2.50	2.30	0.60
4	503	0.32473	23.37032	6.076825	1750.445	106.6261	11	2.25	2.10	0.55

6. Numerical Results

This section presents some numerical examples to illustrate the various theoretical concepts discussed above.

Example 1. A simulation study is used to show the computational procedure for the theoretical discussion of multiobjective MSS. The R package (stratifyR) [45] is used to simulate the data for the two different characteristics, which are divided into four strata. The information on simulation studies is given in Table 1.

The available budget for the survey is $C_0 = \$2500$.

Using (7), in Section 2.3, the best individual optimal solutions for both characteristic $j = 1, 2$ are obtained as follows:

$f_1 = 1218.183, n_1 = 26, n_2 = 33, n_3 = 99, n_4 = 74$ and $f_2 = 494.3353, n_1 = 29, n_2 = 36, n_3 = 92, n_4 = 77$. The Weighted Goal Programming discussed in 2.4 is applied to obtain the compromised allocations using (13) with the help of the LINGO optimization package [46], as follows: $f_1 = 1219.624, f_2 = 496.7824, n_1 = 27, n_2 = 35, n_3 = 96, n_4 = 75$.

Example 2. A simulation study is used to show the computational procedure of the stochastic multiobjective multivariate stratified sampling. The R package (stratifyR) [45] is used to simulate the data for the two different characteristics, which are divided into four strata. The information on simulation studies is given in Table 2.

The available budget for the survey is $C_0 = \$2000$.

The calculated parameters used in this study are presented in Table 3.

The individual solutions of numerical Example 2 for both characteristics $j = 1, 2$ are obtained using (59) of Section 4.2 as follows:

$f_1 = 0.1517723, n_1 = 20, n_2 = 26, n_3 = 64, n_4 = 52$. and $f_2 = 0.3982894, n_1 = 19, n_2 = 22, n_3 = 67, n_4 = 52$. The stochastic Goal programming discussed in Sec. 4.2 is applied to obtain the compromised allocations using (58) as follows: $f_1 = 0.03982894, f_2 = 0.03982894, n_1 = 19, n_2 = 22, n_3 = 67, n_4 = 52$. The Chebychev Goal programming discussed in Section 4.3 is applied to obtain the compromised allocations using (62) as follows: $f_1 = 0.03982894, f_2 =$

$= 0.03982894, n_1 = 19, n_2 = 22, n_3 = 67, n_4 = 52$. The Stochastic Sampling Cost Model discussed in Section 4.4 is applied to obtain the compromised allocations using (63) as follows: $C = 1997.716, n_1 = 21, n_2 = 24, n_3 = 67, n_4 = 49$.

Example 3. Here, the linearization of sampling variance are discussed numerically. Using the data of Table 1 in (67), the following sample allocations for $j = 1$ are $n_{11} = 2, n_{12} = 195, n_{13} = 75, n_{14} = 2$ are obtained with the sampling variance of $f_{\text{linear}} = \sum_{h=1}^L ((1/n_h) - (1/N_h))W_h^2 S_{jh}^2 = 1465.552$. Solving the nonlinear sampling variance problem defined in (7) with the same data, the sample allocation was perceived to be $n_{11} = 26, n_{12} = 33, n_{13} = 99, n_{14} = 74$ with sampling variance of $f_{\text{nonlinear}} = 1218.183$. It is observed that the sample allocation from the nonlinear problem is better than the linearized one. Also, the sampling variability is higher in the linearized model than in the nonlinear. Therefore, it can be concluded that linearization of the sampling problem does not give better sample allocations as well as minimum sample variance.

Example 4. Here, the linear approximation of sampling variance are numerically presented. Using the data of Table 1 in (67), the following sample allocations for $j = 2$ are $n_{21} = 2, n_{22} = 3, n_{23} = 203, n_{24} = 2$ are obtained with the sampling variance $f_{\text{linear}} = \sum_{h=1}^L (1/n_h - 1/N_h)W_h^2 S_{jh}^2 = 595.91$. Solving the nonlinear sampling variance problem with the same data, the sample allocation was perceived to be $n_{21} = 29, n_{22} = 36, n_{23} = 92, n_{24} = 77$ with sampling variance of $f_{\text{nonlinear}} = 494.3353$. It is again observed that the sample allocation from the nonlinear problem is better than the linearized one as in the case of Example 3. Also, the sampling variability is higher in the linearized model than in the nonlinear. Therefore, it can be concluded that linearization of the sampling problem does not give optimal sample allocations as well as minimum sample variance.

In general, It can be concluded that linearization of nonlinear sampling variance in a survey sampling problem does not help to determine the optimal sample allocations with minimum variability since approximation of nonlinear into a linear function will not sufficiently optimize the function value as a result of a loss of generality during the

TABLE 3: Data for two characteristics and four strata.

$W_h^2 S_{1h}^2$	$W_h^2 S_{2h}^2$	$C_{1h}^4 - (S_{h1}^2)^2$	$W_h^4 (C_{1h}^4 - (S_{h1}^2)^2)$	$C_{2h}^4 - (S_{h2}^2)^2$	$W_h^4 (C_{2h}^4 - (S_{h2}^2)^2)$
0.2665922	0.08066731	1366.004	0.2174879	65.5886	0.01044267
0.3933440	0.08527725	1031.736	0.2644493	38.2291	0.009798693
4.345287	1.184770	997.4302	36.17940	75.4956	2.738422
2.46439	0.6407986	1204.273	13.39105	69.6983	0.7750179

linear transformation of sampling variance. In such cases, it has been observed that the optimal global solution (Pareto optimal solution) of a function can suffer. Suppose if the original nonlinear function is convex and we approximate it into the linear function, we know that the linear function can be convex or concave. Indeed, the transformation process of nonlinear to a linear function could compromise several properties of the nonlinear based on the fact that a nonlinear function has a high convergence rate to a linear function. Hence, it can be verified that the linearized sampling variance case optimal allocation has high variability compared to the actual sampling problem. Hence, it can be concluded that the need for linearizing the sampling variance function for obtaining the optimal sample allocation is not an optimal decision in a sampling survey. Since the linearization of sampling design under deterministic does not give an efficient solution, there is no need to carry out the same under stochastic, and therefore, the study did not consider linearization under the uncertainty. Other interested researchers can explore the context of different sampling designs.

7. Conclusion

In sample design, allocating samples efficiently and attaining maximum accuracy in minimizing variances plays an important role. Various techniques, theorems, properties and prepositions, and stochastic models were studied, discussed, and presented for the multiobjective multivariate stratified sampling scheme. The discussion is supported with numerical examples in each case. This research is a theoretical framework and conceptual methodology for survey sampling in optimal allocation problems in a certain and stochastic environment. Based on the discussion and numerical illustrations, it can be deduced that sampling variance values resulting from the linearization have higher variability than the nonlinear sampling variance case. Therefore, the study suggests that there is no need for linearizing the original sampling variance function with the hope of getting an optimal decision regarding sampling allocation in survey sampling. The interested researchers can further demonstrate the usefulness and power of the techniques and methods presented in this study. In the future, the study could be extended to more sampling designs in optimal allocation problems for survey sampling.

Data Availability

Not applicable.

Conflicts of Interest

The authors declare that there are no known conflicts of interest regarding financial or authorship arrangement for this research.

Authors' Contributions

The authors contributed equally to the paper.

Acknowledgments

This research was funded by the Deanship of Scientific Research of Jazan University, under the Waed grant number w41-54.

References

- [1] P. V. Sukhatme, B. V. Sukhatme, S. Sukhatme, and C. Asok, *Sampling Theory of Surveys with Application*, pp. 21–25, Indian society of agricultural statistics and IOWA State University Press Ames, India and USA, 1984.
- [2] M. G. 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, p. 695, 2010.
- [3] R. Varshney, M. G. M. KhanKhan, U. Fatima, and M. J. Ahsan, "Integer compromise allocation in multivariate stratified surveys," *Annals of Operations Research*, vol. 226, no. 1, pp. 659–668, 2015.
- [4] S. Ghufuran, S. Khowaja, and M. J. Ahsan, "Optimum multivariate stratified sampling designs with travel cost: a multiobjective integer nonlinear programming approach," *Communications in Statistics - Simulation and Computation*, vol. 41, no. 5, pp. 598–610, 2012.
- [5] N. Gupta and A. Bari, "Fuzzy multi-objective optimization for optimum allocation in multivariate stratified sampling with quadratic cost and parabolic fuzzy numbers," *Journal of Statistical Computation and Simulation*, vol. 87, no. 12, pp. 2372–2383, 2017.
- [6] Y. S. Raghav, I. Ali, and A. Bari, "Multi-objective nonlinear programming problem approach in multivariate stratified sample surveys in the case of non-response," *Journal of Statistical Computation and Simulation*, vol. 84, no. 1, pp. 22–36, 2014.
- [7] S. Khowaja, S. Ghufuran, and M. J. Ahsan, "Estimation of population means in multivariate stratified random sampling," *Communications in Statistics - Simulation and Computation*, vol. 40, no. 5, pp. 710–718, 2011.
- [8] I. Ali, Y. S. Raghav, and A. Bari, "Compromise allocation in multivariate stratified surveys with stochastic quadratic cost function," *Journal of Statistical Computation and Simulation*, vol. 83, no. 5, pp. 962–976, 2013.
- [9] R. Varshney, "Optimum allocation in multivariate stratified sampling design in the presence of nonresponse with Gamma cost function," *Journal of Statistical Computation and Simulation*, vol. 89, no. 13, pp. 2454–2467, 2019.
- [10] K. G. Reddy and M. G. M. Khan, "Optimal stratification in stratified designs using weibull-distributed auxiliary information," *Communications in Statistics - Theory and Methods*, vol. 48, no. 12, pp. 3136–3152, 2019.

- [11] R. Varshney, S. Gupta, and I. Ali, "An optimum multivariate-multiobjective stratified sampling design: fuzzy programming approach," *Pakistan Journal of Statistics and Operation Research*, vol. 13, no. 4, pp. 829–855, 2017.
- [12] S. Ghufuran, S. Khowaja, and M. J. Ahsan, "Compromise allocation in multivariate stratified sample surveys under two stage randomized response model," *Optimization Letters*, vol. 8, no. 1, pp. 343–357, 2014.
- [13] S. Khowaja, S. Ghufuran, and M. J. Ahsan, "On the problem of compromise allocation in multi-response stratified sample surveys," *Communications in Statistics - Simulation and Computation*, vol. 42, no. 4, pp. 790–799, 2013.
- [14] J. A. Díaz-García and M. M. Garay-Tápia, "Optimum allocation in stratified surveys: stochastic programming," *Computational Statistics & Data Analysis*, vol. 51, no. 6, pp. 3016–3026, 2007.
- [15] J. A. Díaz-García and R. Ramos-Quiroga, "Optimum allocation in multivariate stratified random sampling: a modified prékopa's approach," *Journal of Mathematical Modelling and Algorithms in Operations Research*, vol. 13, no. 3, pp. 315–330, 2014.
- [16] J. A. Díaz-García and R. Ramos-Quiroga, "Stochastic optimal design in multivariate stratified sampling," *Optimization*, vol. 63, no. 11, pp. 1665–1688, 2014.
- [17] A. Haq, I. Ali, and R. Varshney, "Compromise allocation problem in multivariate stratified sampling with flexible fuzzy goals," *Journal of Statistical Computation and Simulation*, vol. 90, no. 9, pp. 1557–1569, 2020.
- [18] M. G. M. KhanKhan, M. J. Ahsan, and N. Jahan, "Compromise allocation in multivariate stratified sampling: an integer solution," *Naval Research Logistics*, vol. 44, no. 1, pp. 69–79, 1997.
- [19] E. B. Tirkolaei, A. Goli, P. Ghasemi, and F. Goodarziyan, "Designing a sustainable closed-loop supply chain network of face masks during the COVID-19 pandemic: pareto-based algorithms," *Journal of Cleaner Production*, vol. 333, Article ID 130056, 2022.
- [20] A. Goli and H. Mohammadi, "Developing a Sustainable Operational Management System Using Hybrid Shapley Value and Multimoor Method: Case Study Petrochemical Supply Chain," *Environment, Development and Sustainability*, pp. 1–30, 2021.
- [21] A. Goli and B. Malmir, "A covering tour approach for disaster relief locating and routing with fuzzy demand," *International Journal of Intelligent Transportation Systems Research*, vol. 18, no. 1, pp. 140–152, 2020.
- [22] A. Goli and T. Keshavarz, "Just-in-time scheduling in identical parallel machine sequence-dependent group scheduling problem," *Journal of Industrial and Management Optimization*, 2021.
- [23] A. Goli, H. Khademi ZareKhademi Zare, R. Tavakkoli-Moghaddam, and A. Sadeghieh, "Hybrid artificial intelligence and robust optimization for a multi-objective product portfolio problem Case study: the dairy products industry," *Computers & Industrial Engineering*, vol. 137, Article ID 106090, 2019.
- [24] M. Alinaghian and A. Goli, "Location, allocation and routing of temporary health centers in rural areas in crisis, solved by improved harmony search algorithm," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 894–913, 2017.
- [25] A. Goli, H. Khademi-Zare, R. Tavakkoli-Moghaddam, A. Sadeghieh, M. Sasanian, and R. Malekalipour Kordestanizadeh, "An integrated approach based on artificial intelligence and novel meta-heuristic algorithms to predict demand for dairy products: a case study," *Network: Computation in Neural Systems*, vol. 32, no. 1, pp. 1–35, 2021.
- [26] I. Ali, U. M. Modibbo, J. Chauhan, and M. Meraj, "An integrated multi-objective optimization modelling for sustainable development goals of India," *Environment, Development and Sustainability*, vol. 23, no. 3, pp. 3811–3831, 2021.
- [27] U. M. Modibbo, I. Ali, and A. Ahmed, "Multi-objective optimization modelling for analysing sustainable development goals of Nigeria: agenda 2030," *Environment, Development and Sustainability*, vol. 23, no. 6, pp. 9529–9563, 2021.
- [28] A. AlArjani, U. M. Modibbo, I. Ali, and B. Sarkar, "A new framework for the sustainable development goals of Saudi Arabia," *Journal of King Saud University Science*, vol. 33, no. 6, Article ID 101477, 2021.
- [29] M. Barma and U. M. Modibbo, "Multiobjective mathematical optimization model for municipal solid waste management with economic analysis of reuse/recycling recovered waste materials," *Journal of Computational and Cognitive Engineering*, 2022.
- [30] Y. S. Muhammad, J. Shabbir, I. Husain, and M. Abdel-Moemen, "Multi-objective compromise allocation in multivariate stratified sampling using extended lexicographic goal programming with gamma cost function," *Journal of Mathematical Modelling and Algorithms in Operations Research*, vol. 14, no. 3, pp. 255–265, 2015.
- [31] J. Beardwood, J. H. Halton, and J. M. Hammersley, "The shortest path through many points," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 55, no. 4, pp. 299–327, 1959.
- [32] J. Neyman, "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*, vol. 97, no. 4, pp. 558–625, 1934.
- [33] A. Charnes and W. W. Cooper, "Chance-constrained programming," *Management Science*, vol. 6, no. 1, pp. 73–79, 1959.
- [34] A. Prekopa, "On probabilistic constrained programming," *Mathematical Programming Study*, vol. 28, pp. 113–138, 1970.
- [35] A. Charnes and W. W. Cooper, "Deterministic equivalents for optimizing and satisficing under chance constraints," *Operations Research*, vol. 11, no. 1, pp. 18–39, 1963.
- [36] J. A. Díaz-García, R. Ramos-Quiroga, and E. Cabrera-Vicencio, "Stochastic programming methods in the response surface methodology," *Computational Statistics & Data Analysis*, vol. 49, no. 3, pp. 837–848, 2005.
- [37] B. Contini, "A stochastic approach to goal programming," *Operations Research*, vol. 16, no. 3, pp. 576–586, 1968.
- [38] A. M. Geoffrion, "Stochastic programming with aspiration or fractile criteria," *Management Science*, vol. 13, no. 9, pp. 672–679, 1967.
- [39] A. Melaku, "Asymptotic normality of the optimal allocation in multivariate stratified random sampling," *Sankhya: The Indian Journal of Statistics, Series B*, vol. 48, pp. 224–232, 1986.
- [40] M. G. Khan, E. A. Khan, and M. J. Ahsan, "An optimal multivariate stratified sampling design using dynamic programming," *Australian & New Zealand Journal of Statistics*, vol. 45, no. 1, pp. 107–113, 2003.
- [41] M. Yousaf Shad and I. Husain, "Trade-off between cost and variance for a multi-objective compromise allocation in stratified random sampling," *Communications in Statistics - Theory and Methods*, vol. 46, no. 6, pp. 2655–2666, 2017.
- [42] R. Varshney, M. J. 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.
- [43] E. M. L. Beale, “On minimizing a convex function subject to linear inequalities,” *Journal of the Royal Statistical Society: Series B*, vol. 17, no. 2, pp. 173–184, 1955.
- [44] L. Kish, “Optima and proxima in linear sample designs,” *Journal of the Royal Statistical Society: Series A*, vol. 139, no. 1, pp. 80–95, 1976.
- [45] K. G. Reddy and M. G. M. KhanKhan, “stratifyR: an R Package for optimal stratification and sample allocation for univariate populations,” *Australian & New Zealand Journal of Statistics*, vol. 62, no. 3, pp. 383–405, 2020.
- [46] Lingo, “LINGO-User’s Guide,” North Dayton Street, Chicago, Illinois, 60622, USA, 2016.