

A MULTI-CONSTRAINED NONLINEAR PROGRAMMING APPROACH FOR OPTIMUM ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING

SHUAIBU IDRIS ADAM^{1*}, QUAZZAFI RABBANI¹

Manuscript received: 24.11.2025; Accepted paper: 02.09.2026;

Published online: 30.09.2026.

Abstract. This study aimed to establish a multi-constrained nonlinear programming framework for optimal allocation in multivariate stratified sampling. The study incorporates time and cost as constraints with the objective of minimizing the sum of the weighted variance. The concept of the Lagrange multiplier was adopted, where two multipliers (λ and γ) were used to derive the optimum sample size, with time explicitly regarded as a constraint. A comparative analysis between linear and quadratic time functions was carried out using LINGO optimization software, and the results reveal that quadratic time functions reduce sample sizes and resource consumption but lead to slightly higher weighted variances compared to linear time functions. Sensitivity analysis further confirms this finding and demonstrates a trade-off between time, cost, and precision, showing how changes in resource budgets affect precision and optimal allocations. The study concluded that, based on the analyses demonstrated, the Linear and Quadratic time functions lead to distinct allocation strategies, reflecting unavoidable trade-offs among cost, time, and precision. It is recommended that this framework should be extended to stochastic or fuzzy environments, where uncertain time and cost are explicitly modelled.

Keywords: multivariate stratified sampling; optimum allocation; linear time function; quadratic time function; framework.

Mathematics Subject Classification (2020): 62D05.

1. INTRODUCTION

It is widely acknowledged that the implementation of stratified sampling is imperative for ensuring maximal efficacy and precision in the inference of populations within the overarching framework of survey design and analysis. By segmenting a heterogeneous population into a composite of relatively uniform strata, the evaluation of population parameters is subjected to enhanced control, thereby augmenting our precision [1]. This technique proves particularly advantageous in intricate survey environments where uncontrolled variability within the population may obscure critical findings. However, ascertaining the optimal distribution of sample sizes across strata can pose significant challenges, particularly in multivariate contexts where multiple attributes are assessed simultaneously, necessitating sophisticated, adaptive, and computationally advanced methodologies [2]. Study [3] posits that [4] was the pioneering scholar to articulate the ideal allocation for a fixed sample size. Nonetheless, his findings remained largely unrecognized

¹ Integral University, Department of Mathematics and Statistics, 226026 Lucknow, India.

E-mail: quazzafi@gmail.com.

* Corresponding author: shuaibuidrisadam3@gmail.com.

until [5] rediscovered them. Subsequently, [6] employed the Cauchy-Schwarz inequality to advance the discourse on optimal allocation in broader conditions. However, in multivariate surveys where multiple characteristics of each selected unit of the population are to be measured, the optimal allocation for one characteristic may significantly diverge from that of others. In such instances, it becomes nearly infeasible to identify an allocation that is optimal for all characteristics concurrently. Consequently, one must resort to an allocation that is optimal in some capacity for all characteristics; such an allocation is referred to as a compromise allocation within the domain of sampling literature.

The traditional paradigm of stratified sampling operates under the premise that the cost function is linear, suggesting that the time or expense associated with measurement per unit remains constant across samples. Acknowledging that this assumption seldom reflects real-world conditions, numerous scholars have expanded the allocation problem by simulating nonlinear cost functions. Integrated travel expenses into multivariate stratified sampling by conceptualizing the issue as a multi-objective integer nonlinear programming problem according to [7]. Complementary research by [8] investigated a generalized model accommodating both linear and nonlinear cost functions with respect to alternative allocations.

Furthermore, research [9] proposed a Gamma cost function that incorporates fixed, variable, and nonlinear escalations of survey costs, particularly in contexts characterized by nonresponse. Lexical goal programming was employed to resolve this model, providing a flexible framework for managing diverse nonlinear cost structures. Additionally, a scrutinized integer compromise allocation within multivariate stratified surveys emphasizes the necessity of reconciling competing survey objectives through goal programming methodologies [10]. Optimal sample size and variance in multivariate stratified sampling with nonlinear time functions by integrating a power-law function are emphasized in [11]. There exists a body of research dedicated to stochastic extensions of nonlinear allocation models. Another study [12] developed a stochastic multi-objective nonlinear programming model for optimal allocation that treated both cost and time as nonlinear functions exhibiting randomness. This solution was addressed using goal programming techniques and concentrated on practical applications within uncertain survey environments.

While prior studies have largely focused on cost nonlinearity, the temporal aspect of survey operations has remained underexplored in the existing academic literature. Yet, time is a binding resource in survey design-governing fieldwork, logistics, personnel deployment, and data quality, and the temporal dimension is often as critical as the cost aspect. In this research, the existing gap is addressed by incorporating both linear and nonlinear time functions into the framework of compromise allocation. By conceptualizing the allocation as a multi-constrained nonlinear programming problem (MCNLPP) and employing multiple Lagrange multipliers, we derive an optimal allocation that minimizes the weighted sum of variances while concurrently adhering to constraints related to cost, precision, and temporal factors.

2. PROBLEM FORMULATION

Let a multivariate stratified population have L number of strata and p characteristics on each population unit. Let N_h denotes the size of the h^{th} stratum and n_h units to be drawn without replacement. Then:

N_h : population size of h^{th} stratum

n_h : sample size to be determine in h^{th} stratum

S_{ih}^2 : population variance of variable i th characteristics in h th stratum

w_h : weight of h^{th} stratum

α_i : variance weight of i^{th} characteristics

$var(\hat{Y}_i)$: variance of the estimate

Objective function:

The objective is to minimize the total weighted variance of the i^{th} characteristics

$$z = \min \left(\sum_{i=1}^p \alpha_i \text{var}(\hat{Y}_i) \right) = \min \left(\sum_{i=1}^p \alpha_i \left(\sum_{h=1}^L w_h^2 S_{ih}^2 / n_h \right) \right). \quad (1)$$

Let

$$k_h \equiv \sum_{i=1}^p \alpha_i w_h^2 S_{ih}^2 > 0.$$

Now,

$$z = \sum_{h=1}^L \frac{k_h}{n_h}.$$

Ideally, $\text{var}(\hat{Y}_i) = \sum_{h=1}^L w_h^2 S_{ih}^2 / n_h - \sum_{h=1}^L w_h^2 S_{ih}^2 / N_h$ for large N_h , the second term

will be negligible, or it could be ignored as it is independent of n_h .

Subject to Constraints:

– Nonlinear Time: $T(n) = \sum_{h=1}^L t_h(n_h) \leq T_{max}$

t_h : sampling time per unit in stratum h

T : total available time.

– Cost: $C(n) = \sum_{h=1}^L c_h n_h \leq C_{max}$

c_h : monetary cost per unit sample in stratum h .

C : Total available cost budget.

where $2 \leq n_h \leq N_h$, for all h are added to take care of oversampling and to provide an estimate of strata variances S_{ih}^2 .

Min $\sum_{i=1}^3 \alpha_i \left(\sum_{h=1}^5 w_h^2 S_{ih}^2 / n_h \right)$

Subject to

$$\begin{aligned} \sum_{h=1}^5 c_h n_h &\leq C_{max} \\ \sum_{h=1}^L t_h n_h &\leq T_{max} \\ 2 &\leq n_h \leq N_h \end{aligned} \quad (2)$$

Lagrangian function:

To simultaneously handle these constraints, the Lagrange multiplier technique provides an elegant analytical framework for deriving optimal stratum allocations under complex nonlinear dependencies.

Introducing the Lagrangian multipliers.

- $\lambda \geq 0$ for time.
- $\gamma \geq 0$ for cost.

$$L(n_h, \lambda, \gamma) = \sum_{h=1}^L \frac{k_h}{n_h} + \lambda (\sum_{h=1}^L t_h(n_h) - T) + \gamma (\sum_{h=1}^L c_h n_h - C)$$

Using Karush-Kuhn-Tucker (KKT) conditions: The KKT conditions are necessary for optimality in non-linear programming, especially when dealing with inequality constraints.

They involve:

1. Stationarity conditions (partial derivatives of the Lagrangian with respect to n_h set to zero).
2. Primal feasibility (original constraints must be satisfied).
3. Dual feasibility (Lagrange Multipliers must be non-negative).
4. Complementary slackness (product of multiplier and slack in constraint is zero).

Let's start with the stationary condition. Take the partial derivative of L w.r.t n_h and set to zero.

$$\frac{\delta L}{\delta n_h} = \frac{\delta}{\delta n_h} \left[\sum_{h=1}^L \frac{k_h}{n_h} + \lambda \left(\sum_{h=1}^L t_h(n_h) - T \right) + \gamma \left(\sum_{h=1}^L c_h n_h - C \right) \right] = 0$$

$$\frac{\delta L}{\delta n_h} = \frac{-k_h}{n_h^2} + \lambda t_h^1(n_h) + \gamma c_h = 0$$

$$\frac{\delta L}{\delta n_h} = \frac{-k_h}{n_h^2} + \lambda t_h^1(n_h) + \gamma c_h = 0$$

Rearranging

$$\frac{k_h}{n_h^2} = \lambda t_h^1(n_h) + \gamma c_h$$

Solving for n_h :

$$n_h^2 = \frac{k_h}{\lambda t_h^1(n_h) + \gamma c_h}$$

By taking the square roots of both sides, we have:

$$n_h = \left(\frac{k_h}{\lambda t_h^1(n_h) + \gamma c_h} \right)^{\frac{1}{2}} \quad (3)$$

Above is a derived framework for optimum sample size where different time functions can be integrated.

2.1. INTEGRATING THE TIME FUNCTIONS INTO THE FRAMEWORK

Let's write $t_h(n_h)$ and its derivatives $\lambda t_h^1(n_h)$ under different time functions.

- Linear time: $t_h(n_h) = \beta_h n_h$

Then $t_h^1(n_h) = \beta_h$,

substituting $t_h^1(n_h) = \beta_h$ in equation (2) above, we have

$$n_h = \left(\frac{k_h}{\lambda\beta_h + \gamma c_h} \right)^{\frac{1}{2}} \quad (4)$$

– Quadratic time: $t_h(n_h) = a_h n_h + b_h n_h^2$

Then $t_h^1(n_h) = a_h + 2b_h n_h$

Again substituting $t_h^1(n_h) = a_h + 2b_h n_h$ in equation (2) above, we have

$$n_h = \left(\frac{k_h}{\lambda(a_h + 2b_h n_h) + \gamma c_h} \right)^{\frac{1}{2}} \quad (5)$$

The linear time function assumes that every observation takes the same amount of time no matter the size or complexity of the stratum. The quadratic time function, on the other hand, accounts for rising marginal time costs per observation, such as logistical delays, interviewer fatigue, and longer travel distance. By including this nonlinear time structure, the framework is better able to represent actual survey conditions, where sampling slows down as the survey grows.

3. RESULTS AND DISCUSSION

3.1. NUMERICAL ILLUSTRATIONS

In research [13], the following example is discussed, having Population ($N = 2050$) split into 5 strata with 3 characteristics. It is assumed that the total available survey cost budget (C_{max}) = 9500 units and Total Time budget (T_{max}) = 4500 units. The values $N_h, S_{1h}^2, S_{2h}^2, S_{3h}^2, t_h, c_h$ and w_h are shown in Table1.

Table 1. Data set.

h	N_h	S_{1h}^2	S_{2h}^2	S_{3h}^2	t_h	c_h	w_h
1	395	35	1005	20	4.0	7	0.19
2	412	107	67	35	4.5	9	0.20
3	438	53	533	19	5.0	10	0.21
4	421	204	805	33	5.5	12	0.21
5	384	91	101	26	6.0	13	0.19

The Multi-constraint Nonlinear Programming Problem (MNLPP) in (2) may be expressed as:

Since all variances have equal importance $\alpha_1 = \alpha_2 = \alpha_3 = 0.333$

Min $0.333 \text{ var } (\hat{Y}_1) + 0.333 \text{ var } (\hat{Y}_2) + 0.333 \text{ var } (\hat{Y}_3)$

Subject to $7n_1 + 9n_2 + 10n_3 + 12n_4 + 13n_5 \leq 9500$

$4n_1 + 4.5n_2 + 5n_3 + 5.5n_4 + 6n_5 \leq 4500$

$2 \leq n_1 \leq 395$

$2 \leq n_2 \leq 412$

$2 \leq n_3 \leq 438$

$2 \leq n_4 \leq 421$

$2 \leq n_5 \leq 384$
 where n_h are integers.

Using the LINGO program, the optimum compromise allocation for both linear and quadratic time functions is obtained by solving equations (4) and (5), respectively, and the results are in Table 2.

Table 2. Results for linear and quadratic time functions.

Time functions	Strata					Total sample size	Total cost	Total time	Weighted variance
	1	2	3	4	5				
Linear	267	117	200	241	96	921	9062	4496	0.206
quadratic	241	109	187	223	92	852	8410	4168	0.222

3.2. SENSITIVITY ANALYSIS

Here, we used different cost and time budgets while observing their effect on total sampling and weighted variance (Tables 3 and 4); a target variance of 0.5 was maintained throughout the analysis.

Table 3. Sensitivity results for linear time function.

Cost budget	Time budget	Weighted variance	Total cost	Total time	Total sample size
10000	5000	0.187	9998	4978.5	826
10500	5500	0.178	10492	5225.0	1077
11000	6000	0.169	10995	5474.5	1113

Table 4. Sensitivity results for Quadratic time functions.

Cost budget	Time budget	Weighted variance	Total cost	Total time	Total sample size
10000	5000	0.201	9299	4606.5	941
10500	5500	0.184	10150	5028.0	1027
11000	6000	0.170	11005	5450.5	1113

3.3. DISCUSSION

The findings presented in Table 2 compared the allocations of sample sizes, total survey costs, cumulative time requirements, and weighted variances under both Linear and Quadratic time functions. The results revealed that the Quadratic time function requires fewer samples, lower costs, and less time, albeit with a slightly higher variance compared to the linear function. Importantly, in both models, the weighted variances remain consistently below the target precision threshold, confirming that the available financial and temporal resources are sufficient to achieve the desired level of precision.

Tables 3 and 4 further highlight the effects of varying cost and time budgets on sample allocations and weighted variances. Sensitivity analysis reveals how increasing budget allocations (both cost and time) influence sample allocations and survey precision, demonstrating that improved precision is directly linked to more efficient resource utilization. This clearly illustrates the inherent trade-off between cost, time, and precision.

Unlike some earlier studies (e.g., (14), (15), (16), (17), (18), (19)) that assume linear measurement of time and/or costs, the present framework explicitly incorporates nonlinear (quadratic) resource functions into the allocation process. By directly contrasting linear and

nonlinear time formulations within the same cost framework, this study provides empirical evidence of how nonlinear time significantly alters allocation strategies, resource demands, and achievable level of precision. These findings underscore the practical importance of accurately modeling realistic nonlinear constraints when designing surveys under complex logistical and financial conditions.

4. CONCLUSIONS

This study established a rigorous framework for optimal sample allocation in multivariate stratified sampling by simultaneously addressing time, cost, and precision constraints using the Lagrange multiplier technique. By formulating the problem as a multi-constrained nonlinear programming problem (MCNLPP) and solving it through (20), the analysis demonstrated that Linear and Quadratic time functions lead to distinct allocation strategies, reflecting unavoidable trade-offs among cost, time, and precision. The explicit integration of nonlinear time constraints within a Lagrangian optimization framework constitutes a novel contribution to the stratified sampling literature. The results clearly illustrated that overlooking nonlinear time effects can misestimate resource requirements, leading to suboptimal budget planning and reduced precision. Hence, this work emphasized the decisive role of time in survey sampling and provides a robust foundation for extending the framework to stochastic or fuzzy environments, where uncertain time and cost are explicitly modelled.

Acknowledgement: The authors are thankful to Integral University for providing the Manuscript Communication Number (IUL/R&D/2025-MCN0004111). Also, the author SHUAIBU IDRIS ADAM is grateful to the Borno State Government for sponsoring his Ph.D. studies.

REFERENCES

- [1] Lohr, S. L., *Sampling: Design and Analysis*, Chapman & Hall/CRC, Boca Raton, New York, USA, 2021. <https://doi.org/10.1201/9780429298899>
- [2] Khan, M. G. M., Ahsan, M. J., A note on optimum allocation in multivariate stratified sampling, *The South Pacific Journal of Natural and Applied Sciences*, 21, 91, 2003. <https://doi.org/10.1071/SP03017>
- [3] Cochran, W. G., *Sampling Techniques*, 3rd Edition, John Wiley & Sons, New York, 1977.
- [4] Tschuprow, A. A., On the mathematical expectation of the moments of frequency distributions, *Metron*, 2, 461, 1923.
- [5] Neyman, J., On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, *Journal of the Royal Statistical Society*, **97**(4), 558, 1934. <https://doi.org/10.1111/j.2397-2335.1934.tb04184.x>
- [6] Stuart, A., A simple presentation of optimum sampling results, *Journal of the Royal Statistical Society*, **B16**, 239, 1954.

- [7] Ghufuran, S., Khawaja, S., Ahsan, M. J., Optimum Multivariate Stratified Sampling Designs with Travel Cost: A Multiobjective Integer Nonlinear Programming Approach, *Communications in Statistics –Simulation and Computation*, **41**, 598, 2012.
- [8] Fajana, O. B., Fadiji, F. A., Generalised Model of Optimum Stratification Allocation Size and Variance for Linear and Non-Linear Cost Function Considering Other Allocation Procedures, *International Journal of Science and Research*, **8**(11), 1406, 2019.
- [9] Varshney, R., Mradula, Optimum allocation in multivariate stratified sampling design in the presence of nonresponse with Gamma cost function, *Journal of Statistical Computation and Simulation*, **89**, 2454, 2019.
- [10] Varshney, R., Khan, M. G. M., Fatima, U., Ahsan, M. J., Integer compromise allocation in multivariate stratified surveys, *Annals of Operations Research*, **226**(1), 659, 2015.
- [11] Ahmad, A., Rabbani, Q., Ansari, A. H., Determination of optimum sample size and variance in multivariate stratified sampling with non-linear time function, *Journal of Science and Arts*, **23**(2), 513, 2023. <https://doi.org/10.46939/J.Sci.Arts-23.2-a18>
- [12] Alshqaq, S. S. A., Ahmadini, A. A. H., Ali, I., Nonlinear Stochastic Multiobjective Optimization Problem in Multivariate Stratified Sampling Design, *Mathematical Problems in Engineering*, **2022**, 2502346, 2022.
- [13] Raghav, Y. S., Haq, A., Ali, I., Multiobjective intuitionistic fuzzy programming under pessimistic and optimistic applications in multivariate stratified sample allocation problems, *PLoS One*, **18**, e0284784, 2023. <https://doi.org/10.1371/journal.pone.0284784>
- [14] Ahmad A., Ansari, A. H., Rabbani, Q., Compromise allocation for two-stage sampling with quadratic travel cost using dynamic programming technique, *International Journal of Applied Mathematics*, **35**, 173, 2022. <https://doi.org/10.12732/ijam.v35i1.13>
- [15] Ahsan A. H., Khan S. U., Mixed allocation in stratified sampling, *Aligarh Journal of Statistics*, **25**, 87, 1982.
- [16] Ahsan M. J., Khan S. U., Optimum allocation in multivariate stratified random sampling with overhead cost, *Metrika*, **29**(1), 71, 1982. <https://doi.org/10.1007/BF01893366>
- [17] Khan M. G.M., Ahsan M. J., Jahan N., Compromise allocation in multivariate stratified sampling: An integer solution, *Naval Research Logistics*, **44**, 69, 1997. [https://doi.org/10.1002/\(SICI\)1520-6750\(199702\)44:1%3C69::AID-NAV4%3E3.0.CO;2-K](https://doi.org/10.1002/(SICI)1520-6750(199702)44:1%3C69::AID-NAV4%3E3.0.CO;2-K)
- [18] Varshney R., Khan M. G. M., Fatima U., Ahsan M. J., Integer compromise allocation in multivariate stratified surveys, *Annals of Operations Research*, **226**(1), 659, 2015. <https://doi.org/10.1007/s10479-014-1734-z>
- [19] Fatima U., Ghufuran S., Ahsan M. J., The problem of optimum allocation in multivariate stratified sampling, *The South Pacific Journal of Natural and Applied Sciences*, **30**, 65, 2012. <https://doi.org/10.1071/SP12007>
- [20] *LINGO 20 User's Manual*, Lindo Systems, Chicago, Illinois, USA, 2020. https://lindo.com/doc/online_help/lingo20_0/index.html?binary_search_the_model.htm