

# A Goal Programming Approach for Multivariate Calibration Weights Estimation in Stratified Random Sampling

Siham Rabee\*, Ramadan Hamed, Ragaa Kassem, Mahmoud Rashwaan

Department of Statistics, Faculty of Economic and Political Science, Cairo University, Egypt

Received February 4, 2021; Revised April 9, 2021; Accepted May 10, 2021

## Cite This Paper in the following Citation Styles

(a): [1] Siham Rabee, Ramadan Hamed, Ragaa Kassem, Mahmoud Rashwaan, "A Goal Programming Approach for Multivariate Calibration Weights Estimation in Stratified Random Sampling," *Mathematics and Statistics*, Vol. 9, No. 3, pp. 326 - 334, 2021. DOI: 10.13189/ms.2021.090314.

(b): Siham Rabee, Ramadan Hamed, Ragaa Kassem, Mahmoud Rashwaan (2021). A Goal Programming Approach for Multivariate Calibration Weights Estimation in Stratified Random Sampling. *Mathematics and Statistics*, 9(3), 326 - 334. DOI: 10.13189/ms.2021.090314.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** Calibration estimation is one of the most important ways to improve the precision of the survey estimates. It is a method in which the designs weights are modified as little as possible by minimizing a given distance measure to the calibrated weights respecting a set of constraints related to suitable auxiliary information. This paper proposes a new approach for Multivariate Calibration Estimation (MCE) of the population mean of a study variable under stratified random sampling scheme using two auxiliary variables. Almost all literature on calibration estimation used Lagrange multiplier technique in order to estimate the calibrated weights. While Lagrange multiplier technique requires all equations included in the model to be differentiable functions, some undifferentiable functions may be faced in some cases. Hence, it is essential to look for using another technique that can provide more flexibility in dealing with the problem. Accordingly, in this paper, using goal programming approach is newly suggested as a different approach for MCE. The theory of the proposed calibration estimation is presented and the calibrated weights are estimated. A comparison study is conducted using actual and generated data to evaluate the performance of the proposed approach for multivariate calibration estimator with other existing calibration estimators. The results of this study prove that using the proposed GP approach for MCE is more flexible and efficient compared to other calibration estimation methods of the population mean.

**Keywords** Multivariate Calibration Estimation, Auxiliary Variables, Stratified Sampling, Goal Programming Technique

---

## 1. Introduction

Statisticians in survey sampling are often interested in the precision of the estimates of the population parameters. This paper focuses on MCE approach, which is considered to improve the efficiency of the estimates of the population parameters. MCE is a systematic way to incorporate auxiliary information in the estimation procedure. With respect to the auxiliary information, it is the additional information correlated in some way with the study variable. It can be available for the target population, either in an aggregated form or in a detailed form for each individual population unit.

In General, any estimation approach includes attaching weights to sample data and then the weighted averages are computed. These attached weights require a little modification making use of the available auxiliary information through calibration estimation. Hence, the main goal of calibration estimation is to get new weights that modify as little as possible the design weights. These design weights have the desirable property of yielding unbiased estimates. Accordingly, the survey statistician

wants to stay as close as possible to these design weights. There is no guarantee that the new weights still give unbiased estimates, but it is expected to be close to the unbiasedness [1].

Deville and Sarndal [1] first introduced calibration estimation in survey sampling. Later, many researchers have contributed to calibration estimation approach which provides calibrated estimators for population parameters using various calibration constraints under different sampling designs. Singh [2] is considered the first study that extended the calibration approach to Stratified Random Sampling (SRS) design. According to [2], many researchers have contributed to the theory of calibration estimation under SRS [3-5].

This study proposes, mainly, new MCE through developing a new set of calibrated weights that can be used to increase the precision of the estimates under SRS by incorporating two auxiliary variables. Therefore, the estimation of the calibrated weights is considered as a Mathematical Programming Problem (MPP) in which the Manhattan distance is minimized respecting a set of calibration constraints. In addition, the constraints that improve the precision of the calibrated estimator should be taken into consideration. The considered optimization problem will be solved using Goal Programming technique; the computational details of the proposed procedure will be illustrated in the existence of two auxiliary variables.

The remaining part of this paper is organized as follows. Section 2 presents notations about calibration approach in SRS, while the literature of the calibration estimation approach is reviewed in section 3. Section 4 presents the suggested mathematical programming model used in estimating the calibrated weights. Section 5 illustrates the performance of the proposed mathematical programming model. This is done through comparing the suggested model to the previously introduced methods using a numerical example and through a simulation study that is demonstrated in section 6. Finally, section 7 summarizes the main conclusions and contribution of this paper.

## 2. Basic Notations of Calibration Estimation Approach in Stratified Random Sampling

Consider finite population sized  $N$  units with  $L$  homogeneous subgroups called strata. Each size of stratum is  $N_h$  units  $\ni \sum_{h=1}^L N_h = N; h = 1, 2, \dots, L$  and  $W_h = N_h/N$  is the weight of stratum. For this population, a sample with size  $n_h$  is selected by Simple Random Sampling Without Replacement (SRSWOR) from each stratum  $\ni \sum_{h=1}^L n_h = n$ , where  $n$  is the total sample size.

Suppose the  $i^{\text{th}}$  unit value of the study variable selected from the  $h^{\text{th}}$  stratum is denoted by  $y_{hi}; i=1, 2, \dots, n_h$ . It must be noted that the unbiased estimator of the population mean under SRS design  $\bar{Y} = \sum_{h=1}^L W_h \bar{Y}_h$  is given by [6]

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h \tag{2.1}$$

where  $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$  denotes the  $h^{\text{th}}$  stratum sample mean. Additionally, the estimated variance of  $\bar{y}_{st}$  under the SRSWOR scheme is given by [6]

$$\hat{v}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \left(\frac{1-f_h}{n_h}\right) s_{hy}^2 \tag{2.2}$$

where  $s_{hy}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$  is the sample variance of  $Y$  in the stratum  $h$ , and  $f_h = n_h/N_h$  is the sampling fraction for the  $h^{\text{th}}$  stratum;  $h=1, 2, \dots, L$ .

Assume that  $X_{hji}$  denotes the  $i^{\text{th}}$  unit value of the  $j^{\text{th}}$  auxiliary variable in the stratum  $h; h=1, 2, \dots, L, i=1, 2, \dots, N_h$  and  $j=1, 2$ . The stratum means  $\bar{X}_{hj} = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hij}$  and the stratum variances  $S_{hj}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (x_{hij} - \bar{X}_{hj})^2$  of each auxiliary variable are assumed to be known.

The main purpose of any calibration estimation approach is to provide an improved estimator of the parameter of the population by incorporating information from number of auxiliary variables. Hence, the calibration estimator of the population mean, under SRS, can be given by [3]

$$\bar{y}_{st}^C = \sum_{h=1}^L \Omega_h \bar{y}_h, \tag{2.3}$$

where  $\Omega_h; h=1, 2, \dots, L$  denote the estimated calibration weights that are chosen to minimize a given distance function, subject to some specific constraints. In addition, the variance of the calibrated estimator can be computed by substituting the design weights,  $W_h$  with the calibration weights  $\Omega_h; h=1, 2, \dots, L$  in (2.2). Accordingly, the variance of the calibrated estimator can be given by [3]

$$\hat{v}(\bar{y}_{st}^C) = \sum_{h=1}^L \Omega_h^2 \left(\frac{1-f_h}{n_h}\right) s_{hy}^2, \tag{2.4}$$

## 3. Review for the Previous Calibration Estimators in Stratified Random Sampling

Mathematical programming tools offer researchers the advantages of optimizing an objective function respecting a set of system constraints that identify the structure of the main problem. This is one of the main features that have been utilized by many authors in the field of calibration estimation. In this section, some of the mathematical programming models, which were suggested in the literature to determine the optimal calibrated weights, are presented.

Most of the literature considered the calibration weights estimation as an optimization problem, in which the chi-square distance function is minimized subject to some important constraints. Finally, they used Lagrange multiplier technique to estimate the optimum calibrated weights.

The literature can be classified according to the number

of incorporated auxiliary variables into two groups. The first group represents the studies that contributed to univariate calibration estimation by incorporating one auxiliary variable, while the second group introduces some multivariate calibration estimators using two auxiliary variables.

With respect to the first group, the calibration estimator introduced by Singh [3] is considered one of the main estimators for the population mean in SRS using one auxiliary variable. Singh [3] minimized the chi-square distance function subject to two main constraints. However, the first constraint assumes that the weighted sum of mean of the auxiliary variable has to be equal to the known parameter of that auxiliary variable. The second constraint indicates that the sum of the calibrated weights is equal to 1. Therefore, Singh [3] optimization problem was formulated as follows

Find  $\Omega_h$  that

$$\text{Minimize } \sum_{h=1}^L \frac{(W_h - \Omega_h)^2}{W_h Q_h} \quad (3.1)$$

Subject to

$$\sum_{h=1}^L \Omega_h \bar{x}_h = \bar{X} \quad , \quad (3.2)$$

$$\sum_{h=1}^L \Omega_h = 1 \quad (3.3)$$

where  $\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$  and  $\bar{X} = \sum_{h=1}^L W_h \bar{x}_h$  are the sample and population means of auxiliary variable in  $h^{\text{th}}$  stratum respectively.  $Q_h$  is appropriately set constant that determine the final form of the calibrated estimators [9].

Moreover, Tracy [4] introduced their calibration estimator by minimizing the objective function in (3.1) subject to two constraints. The first one is the same as the constraint in (3.2), while the second constraint used the second order moments of the auxiliary variable. Hence, the second constraint can be expressed as follows:

$$\sum_{h=1}^L \Omega_h s_h^2 = S^2, \quad (3.4)$$

where  $s_h^2$  and  $S^2 = \sum_{h=1}^L W_h s_h^2$  are the auxiliary variable's sample and population variance in the  $h^{\text{th}}$  stratum respectively.

Motivated by Singh [3] and Tracy [4], Koyuncu and Khadilar [7] proposed their calibration estimator. They also minimized the same objective function in (3.1) subject to three constraints which are the same as the constraints proposed in [3] and [4]. However, they considered using them simultaneously at one optimization problem. Hence, it can be stated that they minimized (3.1) subject to the three constraints expressed in (3.2), (3.3) and (3.4) together simultaneously [7].

Regarding the second group that focused on MCE, Rao [5] can be considered the first study that proposed MCE for the population mean in SRS through incorporating two auxiliary variables. They considered minimizing (3.1) subject to one constraint in which the weighted sum of the sample means of  $j^{\text{th}}$  auxiliary variable is equal to the known parameter for that auxiliary variable. Their optimization

problem is expressed by

Find  $\Omega_h$  that

$$\text{Minimize } \sum_{h=1}^L \frac{(W_h - \Omega_h)^2}{W_h Q_h} \quad (3.1)$$

Subject to

$$\sum_{h=1}^L \Omega_h \bar{x}_{hj} = \bar{X}_j \quad , \quad (3.5)$$

where  $\bar{x}_{hj} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hji}$  and  $\bar{X}_j = \sum_{h=1}^L W_h \bar{x}_{hj}$  are the sample of  $j^{\text{th}}$  auxiliary variable and population means in  $h^{\text{th}}$  stratum respectively.

Further, Rao [8] proposed MCE for the population mean in SRS using different constraints related to two auxiliary variables. They also suggested minimizing (3.1) as their objective function subject to the same three constraints used in [7] but with respect to multi-auxiliary variables.

Almost all previously mentioned literature used the Lagrange multiplier technique in order to derive the values of their calibrated weights. While Lagrange multiplier technique requires all equations included in the model to be differentiable functions, some un-differentiable functions may be faced in some cases. Hence, it is essential to look for another technique that can provide more flexibility in dealing with different optimization problems. In the following section, MPP will be proposed along with its suitable solving method.

## 4 The Suggested Multivariate Mathematical Programming Model

In this section a Multivariate Mathematical Programming (MMP) for MCE in SRS using two auxiliary variables is presented in order to estimate the values of the calibrated weights of strata. The objective function in the proposed model is presented in subsection 4.1 while the constraints included in the model and the suggested MMP will be introduced in the subsections 4.2 and 4.3, respectively.

### 4.1. The Objective Function

This model is concerned with minimizing an objective function that represents a distance measure between the design weights ( $W_h$ ;  $h=1 \dots L$ ) and the calibrated weights ( $\Omega_h$ ;  $h=1 \dots L$ ). It must be noted that most of the literature on calibration estimation in SRS considered minimizing the Chi-squared distance function as their objective function. Contrary to most of this literature, this paper considers the Manhattan distance (L1 Norm) from the design weights  $W_h$  as the main objective function. The main advantage behind using the Manhattan distance measure is decreasing the effect of the existence of the outlier. If the design weights  $W_h$  contain outliers, the variance of the estimator may be large and inference on the

population parameter may be inefficient.

Moreover, the decision variables in the suggested model are the L calibrated weights  $\Omega_h$ ;  $h=1, 2, \dots, L$  assuming that the population under the study consists of L strata. Accordingly, the considered objective function is presented as follows:

$$\text{Minimize } Z = \sum_{h=1}^L |W_h - \Omega_h|, \quad (4.1)$$

where  $W_h$  and  $\Omega_h$ ; ( $h=1, 2, \dots, L$ ) are the design weight and calibrated weight for the  $h^{\text{th}}$  stratum respectively.

### 4.2. The Constraints

This model considers two sets of constraints. The first set represents the calibration constraints while the second set contains the constraints that concentrate on improving the precision of the calibrated estimator. The following three constraints belong to the first set that represents the main key of the calibration procedure. In these calibration constraints, both of the 1<sup>st</sup> and 2<sup>nd</sup> order moments of each auxiliary variable are used. Hence, the first constraint, in which the weighted sum of the  $j^{\text{th}}$  auxiliary variable equals the known parameter for that auxiliary variable, can be expressed as follows

$$\sum_{h=1}^L \Omega_h \bar{x}_{hj} = \bar{X}_j; j=1, 2 \quad (4.2)$$

The second constraint, in which the weighted sum of the sample variances of  $j^{\text{th}}$  auxiliary variable equals population variance for that auxiliary variable, can be represented as follows

$$\sum_{h=1}^L \Omega_h s_{hj}^2 = S_j^2, j=1, 2 \quad (4.3)$$

The third constraint, which guarantees that the sum of the estimated calibration weights doesn't exceed 1, can be given by

$$\sum_{h=1}^L \Omega_h = 1 \quad (4.4)$$

Regarding to the second set of the constraints, two constraints will be included in the MMP model. The first constraint is designed to guarantee that the estimated variance of the calibrated estimator is less than or equal to a specified positive constant  $\zeta$ , and it can be expressed as follows:

$$\sum_{h=1}^L \Omega_h^2 \left(\frac{1-f_h}{n_h}\right) s_{hy}^2 \leq \zeta \quad (4.5)$$

On the other hand, the second constraint is designed to guarantee that the improved calibrated estimator is still close to unbiasedness. It can be expressed as follows

$$\left| \sum_{h=1}^L \Omega_h \bar{y}_h - \bar{Y} \right| \leq \epsilon \quad (4.6)$$

where  $\sum_{h=1}^L \Omega_h \bar{y}_h$  represents the calibration estimator of the population mean, and  $\bar{Y}$  denotes its parameter,  $\epsilon$  is a specified positive constant. Accordingly,  $\left| \sum_{h=1}^L \Omega_h \bar{y}_h - \bar{Y} \right|$  denotes the estimated biasness of the calibrated estimator.

### 4.3. The MMP Model

Based on the elements of the model presented in the previous subsections, the main structure for the MMP model can be expressed as follows:

Find  $\Omega_h$  that

$$\text{Min } Z = \sum_{h=1}^L |W_h - \Omega_h|, \quad (4.7)$$

Subject to

$$\sum_{h=1}^L \Omega_h \bar{x}_{hj} = \bar{X}_j, j=1, 2 \quad (4.8)$$

$$\sum_{h=1}^L \Omega_h s_{hj}^2 = S_j^2, j=1, 2 \quad (4.9)$$

$$\sum_{h=1}^L \Omega_h = 1, \quad (4.10)$$

$$\sum_{h=1}^L \Omega_h^2 \left(\frac{1-f_h}{n_h}\right) s_{hy}^2 \leq \zeta, \quad (4.11)$$

$$\left| \sum_{h=1}^L \Omega_h \bar{y}_h - \bar{Y} \right| \leq \epsilon, \quad (4.12)$$

$$\Omega_h \geq 0; h=1, 2, \dots, L \quad (4.13)$$

where;

- $W_h$  &  $\Omega_h$  is the design and calibrated weights, respectively;  $h=1, 2, \dots, L$ .
- $\bar{x}_{hj}$  is sample mean for the  $j^{\text{th}}$  auxiliary variable in  $h^{\text{th}}$  stratum;  $h=1, 2, \dots, L, j=1, 2$ .
- $\bar{X}_j$  is the population mean for the  $j^{\text{th}}$  auxiliary variable.
- $s_{hj}^2$  is sample variance for the  $j^{\text{th}}$  auxiliary variable in  $h^{\text{th}}$  stratum;  $h=1, 2, \dots, L$ .
- $S_j^2$  is the population variance for the  $j^{\text{th}}$  auxiliary variables.
- $\sum_{h=1}^L \Omega_h^2 \left(\frac{1-f_h}{n_h}\right) s_{hy}^2$  denotes the estimated variance of the calibrated estimator of the population parameter.
- $\zeta$  is a positive upper bound for the estimated variance of the calibrated estimator.
- $\left| \sum_{h=1}^L \Omega_h \bar{y}_h - \bar{Y} \right|$  denotes the estimated biasness of the population parameter's calibrated estimator.
- $\epsilon$  is a specified positive constant as an upper bound for the estimated bias of the calibrated estimator.

On the other hand, Goal programming approach (GP) has been accepted as a basic mathematical programming method for solving different optimization problems. Practically, in some mathematical programming problems that contain too restrictive hard constraints, it is still available to find a solution for that problem using GP technique. This can be obtained by modifying these hard constraints to be soft constraints, and then these soft constraints will be treated as goals to be achieved by adding penalty term in the objective function. Finally, the overall objective is to get as close as possible to these goals [9]. Nevertheless, up to researcher's knowledge, there is not any study that considered using GP Technique for calibration estimation in stratified sampling. Accordingly, based on the wide flexibility of GP technique, it is suggested to convert the original MMP model (4.7- 4.13) into a GP model for estimating the calibrated weights of strata. In the suggested GP model, the constraints of the

entire problem are considered as goals to be attained and satisfied.

#### 4.4. The GP Model

In order to solve any mathematical programming model using Goal programming technique, positive and/or negative deviational variables have to be added to the main objective function in addition to each goal has to be attained. Further, it can be proved that minimizing the absolute deviation between any parameter and its estimate is equivalent to minimizing the sum of the positive and negative deviational variables added to this function [10].

In another words, minimizing the sum of positive and negative deviational variables for any objective (goal constraint) is equivalent to minimizing the absolute deviation between the right-hand side and the left-hand side of this objective (goal constraint). Here is an example that can be used to clarify the previous statement. Assuming that the objective (goal constraint) expressed in (4.14) is considered

$$aX_1 + bX_2 + n_1 - p_1 = A, \quad (4.14)$$

where A is the target value for this objective (goal constraint),  $n_1$  &  $p_1$  are called positive and negative deviational variable respectively.

The positive and negative deviational variables can be defined by the following [10]:

$$n_1 = \frac{1}{2}[A - aX_1 + bX_2 + |A - aX_1 + bX_2|] \quad (4.15)$$

$$p_1 = \frac{1}{2}[aX_1 + bX_2 - A + |A - aX_1 + bX_2|] \quad (4.16)$$

Accordingly, from (4.15) & (4.16), we get:

$$n_1 + p_1 = \frac{1}{2}A - \frac{1}{2}(aX_1 + bX_2) + \frac{1}{2}|A - aX_1 + bX_2| + \frac{1}{2}(aX_1 + bX_2) - \frac{1}{2}A + \frac{1}{2}|A - aX_1 + bX_2|.$$

Hence  $n_1 + p_1 = |A - aX_1 + bX_2|$ .

Further, to construct the Goal programming model for the previous MMP [4.7- 4.13], each objective (goal constraint) will be divided by its right-hand side value in order to avoid the effect of different measurement units between different constraints assuming that the right-hand side value doesn't equal zero.

Accordingly, the suggested GP model can be expressed as follows:

**Find**  $\Omega_h, dn_{1h}, dp_{1h}, dn_k, dp_k; h=1,2,..L, k=2, 3,..7$ , that:  
Minimize  $Z = \sum_{h=1}^L (dn_{1h} + dp_{1h}) + dp_2 + \sum_{k=3}^7 (dn_k + dp_k)$ . (4.17)

**subject to;**

$$[\Omega_h/W_h] + dn_{1h} - dp_{1h} = 1; h=1, 2, .. L \quad (4.18)$$

$$[\sum_{h=1}^L \Omega_h^2 (\frac{1-f_h}{n_h}) s_{hy}^2] / \zeta + dn_2 - dp_2 = 1 \quad (4.19)$$

$$[\sum_{h=1}^L \Omega_h \bar{x}_{h1}] / \bar{X}_1 + dn_3 - dp_3 = 1; \quad (4.20)$$

$$[\sum_{h=1}^L \Omega_h \bar{x}_{h2}] / \bar{X}_2 + dn_4 - dp_4 = 1; \quad (4.21)$$

$$[\sum_{h=1}^L \Omega_h s_{1h}^2] / S_{1+}^2 + dn_5 - dp_5 = 1; \quad (4.22)$$

$$[\sum_{h=1}^L \Omega_h s_{2h}^2] / S_{2+}^2 + dn_6 - dp_6 = 1; \quad (4.23)$$

$$[\sum_{h=1}^L \Omega_h \bar{y}_h] / \bar{Y} + dn_7 - dp_7 = 1; \quad (4.24)$$

$$\sum_{h=1}^L \Omega_h = 1; \quad (4.25)$$

$$\Omega_h, dn_{1h}, dp_{1h}, h=1,2,..L \quad dn_k, dp_k; \geq 0; k=2, 3, ..7.$$

Theoretically, there is a tradeoff between getting a calibrated estimator with a minimum variance and a minimum bias simultaneously. Minimizing the variance of the calibrated estimator by adding the goal constraint (4.19) without adding the goal constraint (4.24) may cause getting a biased calibrated estimator. Hence, it is important to add both of the two constraints in order to control the overall mean squared error of the calibrated estimator.

Therefore, according to this theoretical basis, three cases for the proposed GP approach will be considered. The main difference between these cases depends on the relative importance of each goal (variance and bias). This relative importance can be determined by the priority weights given to the deviational variables added to each goal constraint, specifically;  $dp_2$  and  $(dn_7 + dp_7)$  deviational variables.

The three cases can be further explained as follows:

**Case 1:** Both of the variance and the bias of the calibrated estimator has the same relative importance so that equal priority weights are given to both  $dp_2$  and  $(dn_7 + dp_7)$  deviational variables of the variance and the bias goal constraints respectively.

**Case 2:** The variance of the calibrated estimator is relatively more important than its unbiasedness. Hence, the priority weight given to  $dp_2$  is equal to 1, while the priority weight for  $(dn_7 + dp_7)$  is zero.

**Case 3:** This case is considered the inverse of case 2, where having unbiased calibrated estimator is more important than decreasing its variance. Hence, the priority weight given to  $dp_2$  is equal to zero, while the priority weight for  $(dn_7 + dp_7)$  is one.

In the following two sections, the performance of the proposed approach with its three cases will be demonstrated and compared with some literature using a numerical example and a simulation study.

## 5. Numerical Illustrations and Comparison Study

For explaining the computational details of the proposed multivariate calibration estimator of the population mean, a tobacco data containing 106 units for three variables: area, yield, and production is used. The data is obtained from the Agriculture Statistics 1999 reported in [3] and commonly used in the literature [3, 5, and 8]. The population in this data is divided into  $L = 10$

subgroups (strata). Using proportional allocation, a sample of 40 observations is selected.

**5.1. Numerical Illustration**

Assuming that estimating the average production of tobacco  $\bar{Y}$  using both of area and yield as the auxiliary variables is considered. To estimate the multivariate

calibration weights and then compute the calibration estimator of  $\bar{Y}$  in SRS, the same sample units taken in [3] is used. The information needed for solving the model and computational details is summarized in Tables 1 & 2.

GAMS software is used to solve the suggested GP. It solves the nonlinear programming problems using CONOPT 3.15L solver (by default) and the estimated calibration weights are obtained and given in table 3.

**Table 1.** Population's Information on Tobacco Data

$\bar{X}_1$	$\bar{X}_2$	$S_1^2$	$S_2^2$	$\zeta$
34438.61	1.5507	22340.229388	0.6323	512333228

**Table 2.** Needed Information on Tobacco Data's Sample

H	$n_h$	$f_h$	$\bar{x}_1$	$\bar{x}_2$	$s_{h1}^2$	$s_{h2}^2$	$s_{hy}^2$
1	3	0.5	1304.7	1.9400	0.722185	0.0171	3205948
2	3	0.5	29075	1.3767	839.008125	0.4757	451299457
3	3	0.375	5191.7	2.7933	74.3878583	0.8010	5692179003
4	3	0.3	21700	1.4433	6.070000	0.7362	220530000
5	4	0.333	6808	1.7875	63.5729813	1.0698	302045625
6	2	0.500	6808	1.7850	1.620000	0.0612	7411250
7	11	0.367	24481.5	1.3209	1801.653230	0.4824	5002786596
8	6	0.353	294809	1.3200	322774.1010042	0.2462	1013000000000
9	3	0.300	6303.7	1.3267	59.9398903	0.0306	76515960
10	2	0.667	350	1.7650	0.125000	1.3285	1058512.5

**Table 3.** The Estimated Calibrated Weights Using the Proposed GP approach

h	$(W_h)$	$\Omega_h^{case 1}$	$\Omega_h^{case 2}$	$\Omega_h^{case 3}$
1	0.0566	0.0566	0.0566	0.0566
2	0.0566	0.0566	0.0566	0.0566
3	0.0755	0.0755	0.0755	0.0755
4	0.0943	0.0943	0.0943	0.0943
5	0.1132	0.1132	0.1132	0.1132
6	0.0377	0.0377	0.0377	0.0377
7	0.283	0.4435	0.377	0.4435
8	0.160	0	0.066	0
9	0.0943	0.0943	0.0943	0.0943
10	0.0283	0.0283	0.0283	0.0283

**5.2. Comparison Study**

In this section, the performance of the proposed multivariate calibration estimators and other existing calibration estimators will be compared using the Relative Efficiency RE. The proposed GP approach will be compared with following existing calibration estimators:

- Tracy [4] estimator:

$$\bar{y}_{st}^{(TR)} = \sum_{h=1}^L \Omega_h^{(1)} \bar{y}_h.$$

- Singh [3] estimator:

$$\bar{y}_{st}^{(S)} = \sum_{h=1}^L \Omega_h^{(2)} \bar{y}_h.$$

- Rao [5] estimator:

$$\bar{y}_{st}^{(R12)} = \sum_{h=1}^L \Omega_h^{(3)} \bar{y}_h.$$

- Rao [8] estimator:

$$\bar{y}_{st}^{(R18)} = \sum_{h=1}^L \Omega_h^{(4)} \bar{y}_h.$$

To assess the performance of different calibration estimators compared to the unbiased estimator  $\bar{y}_{st}$ , the measure of Relative Efficiency (RE) is computed as follows [8]:

$$RE = \frac{\hat{v}(\bar{y}_{st})}{\hat{v}(\bar{Y})} \times 100 \tag{5.1}$$

where

$$\hat{v}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hy}^2 \tag{5.2}$$

It is the estimated variance of the unbiased estimator  $\bar{y}_{st}$ ,  $\hat{v}(\bar{Y})$  is the estimated variance of the calibrated estimators, that can be computed through substituting by the calibrated weights instead of the design weights in (5.2) [8].

Moreover, table 4 presents the variance  $\hat{v}(\bar{Y})$  of calibration estimator and its relative efficiency (RE) using tobacco data.

**Table 4.** Comparison Results for Numerical Example using GP approach.

Estimator	$\hat{v}(\bar{Y})$	RE%
Stratified: $\bar{y}_{st}$	2822731121	100.0
Tracy (2003): $\bar{y}_{st}^{(TR)}$	609889505.3	462.8
Singh (2003): $\bar{y}_{st}^{(S)}$	569781129.3	495.4
Rao (2012): $\bar{y}_{st}^{(R12)}$	605381935.5	466.3
Rao (2018): $\bar{y}_{st}^{(R18)}$	512333228.2	551
Proposed: $\bar{y}_{st}^{(GP)}$ Case 1	58808436.7	4799
Proposed: $\bar{y}_{st}^{(GP)}$ Case 2	526358402.6	536.3
Proposed: $\bar{y}_{st}^{(GP)}$ Case 3	58808436.7	4799.7

The results in table 4 show that the proposed GP approach is more efficient than other calibration estimation methods. This can be observed in the increase in the efficiency of the proposed multivariate calibration estimator over the unbiased stratified estimator which is 4799% compared to 462.8%, 495.4% and 551% for the estimators of [3, 4, and 8] respectively. This emphasizes the importance of using GP approach for multivariate calibration estimation that provides more flexibility in dealing with the constraints of the model.

**6. Simulation Study**

In this section, a simulation study is conducted and demonstrated to assess the performance of the proposed estimator compared to the same methods used in the numerical example. A population with three strata is generated; the size of the stratum is taken as 3000, 2000, 5000 respectively. The study variable and two auxiliary variables are generated using R software from Multivariate Normal Distribution with parameters  $\mu = (0, 0, 0)$  and  $\Sigma$ . The covariance matrix  $\Sigma$  is determined based on the correlation levels  $\rho$  between the study and auxiliary variables. Therefore, the covariance matrix  $\Sigma$  can be given by

$$\Sigma = \begin{bmatrix} S_{y_h}^2 & S_{y_h \times x_{1h}} & S_{y_h \times x_{2h}} \\ S_{y_h \times x_{1h}} & S_{x_{1h}}^2 & S_{x_{1h} \times x_{2h}} \\ S_{y_h \times x_{2h}} & S_{x_{1h} \times x_{2h}} & S_{x_{2h}}^2 \end{bmatrix} \tag{6.1}$$

where the main diagonal's elements represent the variances of the three variables and the elements of the off diagonal represent the covariance between each two variables.

Using the proportional allocation, a sample with size  $n_h=1000$  is selected by SRSWR, so  $n_h$  can be given by

$$n_h = nW_h \tag{6.2}$$

Detailed information about the simulated data is given in table (5).

In order to compare the proposed GP approach for multivariate calibration estimation with the other existing calibration methods through the simulation study, a total of 1000 datasets are generated where each dataset represents a population of size 10,000 observations. Moreover, a sample with size 1000 observations is proportionally selected from each population under stratified sampling scheme. For each population and its selected sample, the means and variances for the three generated variables are calculated within the three strata. Since the values of the population parameters are unknown, they can be calculated from a pilot survey or obtained from a previous study. These calculated values for the means and the variances represent the input for the models under comparison.

**Table 5.** Properties of the Generated Population Data

h	$N_h$	$n_h$	$\rho_{yx_h}$	Variables distribution
1	3000	300	0.5	$N(0,10) = f(y) = \frac{1}{10\sqrt{2\pi}}$ $exp \left[ -\frac{\left(\frac{y}{10}\right)^2}{2} \right]$
2	2000	200	0.7	
3	5000	500	0.9	$N(0,5) = f(x_1) = \frac{1}{5\sqrt{2\pi}}$ $exp \left[ -\frac{\left(\frac{x_1}{5}\right)^2}{2} \right]$ $N(0,3) = f(x_2) = \frac{1}{3\sqrt{2\pi}}$ $exp \left[ -\frac{\left(\frac{x_2}{3}\right)^2}{2} \right]$

As mentioned above, five models are considered in the comparison. Four models are from the literature and the fifth model is the proposed one. It must be noted that all of the models under comparison are solved for a total number of 1000 runs using GAMS software.

The performance of the different methods for calibration estimators under comparison are evaluated by the Relative Root Mean Squared Error "RRMSE" that can be defined by [11].

$$RRMSE = \sqrt{\frac{\sum_{k=1}^{1000} \left( \frac{\bar{y}_{st}^{(\alpha)} - \bar{Y}}{\bar{Y}} \right)^2}{1000}} \tag{6.3}$$

where  $\bar{Y}$  denotes the population parameter and  $\bar{y}^{(\alpha)}$  denotes the calibration estimators; specifically,  $\alpha =$  Singh [3], Tracy [4], Rao [5], Rao [8] & the proposed GP estimator. Table (6) summarizes the simulation results for the three cases.

Table 6 shows that in case 1, where equal weights are given for both of the variance and the bias of the calibrated estimator, the suggested multivariate calibration estimator is better than other existing calibration estimators. The simulated data's results showed that its RRMSE is lower than the RRMSE of [3], [4], [5] and [8] calibration estimators. Moreover, the proposed calibrated estimator in case 2 has a lower RRMSE than [8] calibration estimator but higher than the other estimators. Finally, the proposed multivariate calibration estimator in case 3 is considered the best among all estimators even the estimator in case 1.

### 7. Conclusions

This paper proposed a new approach for multivariate calibration estimation for the population mean of the study variable under SRS. New calibration weights are developed by minimizing the Manhattan distance function subject to new calibration constraints using Goal Programming approach. A comparison study using both of numerical example and generated data is performed to assess the precision of the proposed estimator. The results showed that the proposed estimator provides better estimates compared to other existing calibration methods. Specifically, using GP approach provides more efficient and flexible calibration estimator than other existing calibration estimation in the stratified sampling design.

For future research, the Euclidean distance (L2 Norm) could be used rather than the Manhattan distance (L1 Norm).

**Table 6.** Different Calibration Estimator's RRMSE

Cases	$\bar{y}^{(TR)}$	$\bar{y}^{(SI)}$	$\bar{y}^{(RA12)}$	$\bar{y}^{(RA18)}$	$\bar{y}^{(GP)}$
Case 1:	68.13468	97.38585	91.10178	125.6647	15.35545
Case 2:	68.13468	97.38585	91.10178	125.6647	99.38826
Case 3:	68.13468	97.38585	91.10178	125.6647	15.3041



---

## REFERENCES

- [1] J. C. Deville, C. E. Sarndal, Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, Vol. 87, No.418, 376-382, 1992.
- [2] S. Singh, S. Horn, F. Yu, Estimation of variance of the general regression estimator: Higher level calibration approach, *Survey Methodology*, 24, 41-50, 1998.
- [3] S. Singh *Advanced Sampling Theory with Applications*, Kluwer Academic Publishers, 2003.
- [4] D.S. Tracy, S. Singh, R. Arnab. Note on calibration in stratified and double sampling, *Survey Methodology*, Vol. 29, No. 1, pp. 99-104, 2003.
- [5] D. K. Rao, M. G. Khan, S. Khan. Mathematical Programming on Multivariate Calibration Estimation in Stratified Sampling, *International Journal of Mathematical, Computational, Physical, electrical and Computer Engineering*, Vol 6, No.12, pp 58-62, 2012.
- [6] W.G. Cochran *Sampling Techniques*, John Wiley and Sons, New York, 1977.
- [7] N., Koyuncu, C. Kadilar. A New Calibration Estimator in Stratified Double Sampling, *Hacettepe Journal of Mathematics and Statistics*, Vol. 43, No.2, pp. 1-9, 2016.
- [8] D. K. Rao, M. G. Khan, G. Singh On Calibrated Weights in Stratified Sampling, *ANZIAM Journal*, vol. 59, pp.190-204, 2018.
- [9] I. Berrada, J. A. Ferland, P. Michelon. A Multi-objective Approach to Nurse Scheduling with both Hard and Soft Constraints, *Elsevier Science Ltd*, vol. 30, no.3, pp. 183-193, 1996.
- [10] M. J. Schniederjans. *Goal Programming: Methodology and Applications*, Kluwer Academic Publishers, 1995.
- [11] N. Özgül New Calibration Estimator Based on Two Auxiliary Variables in Stratified Sampling. *Communications in Statistics - Theory and Methods*, 1-12, 2018.