

6-30-2021

Calibration-Based Estimators using Different Distance Measures under Two Auxiliary Variables: A Comparative Study

Piyush Kant Rai

Banaras Hindu University, Varanasi, India, raipiyush5@gmail.com

Alka Singh

Banasthali University, Rajasthan, India, singhalka2889@gmail.com

Muhammad Qasim

University of Veterinary and Animal Sciences, Lahore, Pakistan, muhammad.qasim@ju.se

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Rai, Piyush Kant; Singh, Alka; and Qasim, Muhammad (2021) "Calibration-Based Estimators using Different Distance Measures under Two Auxiliary Variables: A Comparative Study," *Journal of Modern Applied Statistical Methods*: Vol. 19 : Iss. 1 , Article 15.

DOI: 10.22237/jmasm/1619481600

Available at: <https://digitalcommons.wayne.edu/jmasm/vol19/iss1/15>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Calibration-Based Estimators using Different Distance Measures under Two Auxiliary Variables: A Comparative Study

Piyush Kant Rai

Banaras Hindu University
Varanasi, India

Alka Singh

Banasthali University
Rajasthan, India

Muhammad Qasim

University of Veterinary and
Animal Sciences
Lahore, Pakistan

This article introduces calibration estimators under different distance measures based on two auxiliary variables in stratified sampling. The theory of the calibration estimator is presented. The calibrated weights based on different distance functions are also derived. A simulation study has been carried out to judge the performance of the proposed estimators based on the minimum relative root mean squared error criterion. A real-life data set is also used to confirm the supremacy of the proposed method.

Keywords: Calibration estimation, stratified sampling, distance function, auxiliary information

Introduction

In survey sampling, the precision of the estimate of study variable can be increased by using the most popular stratified sampling technique. The calibration-based estimation method helps in improving the survey estimates by means of auxiliary information (e.g., known population total or mean of the auxiliary variables) through adjusting the initial design weights. A calibration estimator uses modified weights which are known as calibrated weights. These calibrated weights are determined by minimizing a given distance function to the initial design weights respecting a set of constraints associated with auxiliary information. Furthermore, the auxiliary variables are used to increase the precision of survey estimates of the population. In such situations, the ratio and regression methods of estimation are well known in sampling theory. Deville and Särndal (1992) defined calibration as an approach to estimation for finite populations consisting of (a) A computation of

weights that incorporate specified auxiliary information and are restrained by calibration equation(s), (b) The use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units, (c) An objective to obtain nearly design unbiased estimates as long as non-response and other non-sampling errors are absent.

In literature, many researchers considered the case of one auxiliary variable, but not much attention is given to estimate the character of the main variable taken into account of two or more auxiliary variables. Olkin (1958) introduced a multivariate ratio estimate by using more than one auxiliary variables. Shukla (1965) followed the work of Olkin and gave a multivariate regression estimate only for one sampling case and also extended it for the double sampling scheme. Similarly, Raj (1965) introduced a multivariate difference estimator by taking the benefit of two or more auxiliary variables over single auxiliary variable. Though, these problems become much severe in computation when two or more auxiliary variables are considered. Some calibration-based estimators were defined by Tracy et al. (2003), Kim et al. (2007), and Koyuncu (2012) in stratified sampling. Koyuncu and Kadilar (2013) considered various loss functions to define some new weights and based on these weights, calibration estimators are compared under single auxiliary variable. The main aim of this article is to introduce calibration estimator under different distance measures based on two auxiliary variables in stratified sampling scheme. Also, we have shown that all the derived estimators under two auxiliary variables perform better than the estimators under single auxiliary variable.

The rest of the article is as follows. We first discuss notations and calibration-based estimators under different distance functions. We then describe the simulation study. A real-life application on Forced Expiratory Volume (FEV) data set studied in East Boston is then illustrated. Finally, the conclusions of this study are presented.

Notations and Calibration Estimators

Consider a case where the information on two auxiliary variables is available. Suppose there is a finite population $\Omega = \{1, 2, \dots, N\}$ which is divided into H strata, with the h^{th} stratum containing N_h ($h = 1, 2, \dots, H$) units, such that

CALIBRATION ESTIMATOR USING DIFFERENT DISTANCE MEASURES

$$\sum_{h=1}^H N_h = N .$$

A simple random sample of size n_h is drawn without replacement from the h^{th} stratum such that

$$\sum_{h=1}^H n_h = n .$$

The study and two auxiliary variables are denoted as y , x_1 , and x_2 , respectively. Furthermore, y_{hi} , x_{1hi} , and x_{2hi} denote the values taken by the i^{th} ($i = 1, 2, \dots, N_h$) unit from the h^{th} ($h = 1, 2, \dots, H$) stratum in the population by y , x_1 , and x_2 . Let the population mean of the two auxiliary variables

$$\bar{X}_1 = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{1hi} \quad \text{and} \quad \bar{X}_2 = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{2hi}$$

be known. The primary aim is to estimate the population mean

$$\bar{Y} = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi} .$$

In a stratified sampling scheme, the classical unbiased estimator of the population mean is given by

$$\bar{Y}_{\text{st}} = \sum_{h=1}^H w_h \bar{y}_h , \tag{1}$$

where $w_h = N_h / N$ are the stratum weights.

Tracy et al. (2003) proposed the calibrated estimator of the population mean under stratified sampling as

$$\bar{Y}_{\text{st}}^c = \sum_{h=1}^H \Omega_h \bar{y}_h , \tag{2}$$

where Ω_h ($h = 1, 2, \dots, H$) are the calibrated weights which can be obtained by minimizing any of the following distance functions:

$$D_1(\Omega_h, w_h) = \sum_{h=1}^H \frac{(\Omega_h - w_h)^2}{w_h q_h}, \quad (3)$$

$$D_2(\Omega_h, w_h) = 2 \sum_{h=1}^H \frac{(\sqrt{\Omega_h} - \sqrt{w_h})^2}{w_h q_h}, \quad (4)$$

$$D_3(\Omega_h, w_h) = \sum_{h=1}^H \frac{1}{q_h} \left(\frac{\Omega_h}{w_h} - 1 \right)^2, \quad (5)$$

$$D_4(\Omega_h, w_h) = \sum_{h=1}^H \frac{1}{q_h} \left(\frac{\sqrt{\Omega_h}}{\sqrt{w_h}} - 1 \right)^2, \quad (6)$$

$$D_5(\Omega_h, w_h) = \sum_{h=1}^H \frac{1}{q_h} \left(-w_h \log \frac{\Omega_h}{w_h} + \Omega_h - w_h \right), \quad (7)$$

subject to the constraints

$$\sum_{h=1}^H \Omega_h \bar{x}_{1h} = \sum_{h=1}^H w_h \bar{X}_{1h}, \quad (8)$$

$$\sum_{h=1}^H \Omega_h \bar{x}_{2h} = \sum_{h=1}^H w_h \bar{X}_{2h}. \quad (9)$$

Where q_h are suitably chosen weights. The form of estimator depends on the choice of q_h .

Case 1

The Lagrange function for the weights Ω_h , which satisfy the calibration Equations (8) and (9) by minimizing the distance function $D_1(\Omega_h, w_h)$, is given by

$$L_1 = \sum_{h=1}^H 2 \frac{(\Omega_h - \bar{x}_{1h})^2}{w_h q_h} - 2\lambda_1 \left(\sum_{h=1}^H \Omega_h \bar{x}_{1h} - \sum_{h=1}^H w_h \bar{X}_{1h} \right) - 2\lambda_2 \left(\sum_{h=1}^H \Omega_h \bar{x}_{2h} - \sum_{h=1}^H w_h \bar{X}_{2h} \right) \quad (10)$$

where λ_1 and λ_2 are Lagrangian multipliers. Differentiating L_1 with respect to Ω_h and setting equal to 0, we have

$$\Omega_h = w_h (1 + \lambda_1 q_h \bar{x}_{1h} + \lambda_2 q_h \bar{x}_{2h}) \quad (11)$$

On putting the value of Ω_h from Equation (11) in Equations (8) and (9),

$$\lambda_1 \sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 + \lambda_2 \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} = \sum_{h=1}^H w_h \bar{X}_{1h} - \sum_{h=1}^H \Omega_h \bar{x}_{1h}, \quad (12)$$

$$\lambda_1 \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} + \lambda_2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 = \sum_{h=1}^H w_h \bar{X}_{2h} - \sum_{h=1}^H \Omega_h \bar{x}_{2h}. \quad (13)$$

Solving for λ_1 and λ_2 , we obtain the following:

$$\lambda_1 = \frac{\sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 \hat{X}_{1h} - \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}, \quad (14)$$

$$\lambda_2 = \frac{-\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{1h} - \sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \hat{X}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}. \quad (15)$$

where

$$\hat{X}_{1h} = \sum_{h=1}^H w_h \bar{X}_{1h} - \sum_{h=1}^H w_h \bar{x}_{1h} \quad (16)$$

$$\hat{X}_{2h} = \sum_{h=1}^H w_h \bar{X}_{2h} - \sum_{h=1}^H w_h \bar{x}_{2h} \quad (17)$$

The value of Ω_h after putting the value of λ_1 and λ_2 is given by

$$\begin{aligned} \Omega_h = w_h + w_h q_h \bar{x}_{1h} & \frac{\sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 \hat{X}_{1h} - \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2} \\ & + w_h q_h \bar{x}_{2h} \frac{-\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{1h} - \sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \hat{X}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2} \end{aligned} \quad (18)$$

Thus, the calibration estimator under stratified sampling based on two auxiliary variables is stated as:

$$\bar{Y}_{st2}^c = \sum_{h=1}^H w_h \bar{y}_h + \hat{\beta}_0 \hat{X}_{1h} + \hat{\beta}_1 \hat{X}_{2h} \quad (19)$$

where

$$\begin{aligned} \hat{\beta}_0 &= \frac{\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \sum_{h=1}^H w_h q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2} \\ \hat{\beta}_1 &= \frac{\sum_{h=1}^H w_h q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 - \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2} \end{aligned}$$

The calibration estimator under $D_1(\Omega_h, w_h)$ for a single auxiliary variable is available in literature:

$$\bar{Y}_{TR1} = \sum_{h=1}^H w_h \bar{y}_h + \frac{\sum_{h=1}^H w_h \bar{X}_h - \sum_{h=1}^H w_h \bar{x}_h}{\sum_{h=1}^H w_h q_h \bar{x}_h^2} \sum_{h=1}^H w_h q_h \bar{x}_h \bar{y}_h. \quad (20)$$

Case 2

The Lagrange function for the weights Ω_h , satisfying the calibration Equations (8) and (9) by minimizing the distance function $D_2(\Omega_h, w_h)$, is given by

$$L_2 = \sum_{h=1}^H 2 \frac{(\sqrt{\Omega_h} - \sqrt{w_h})^2}{q_h} - 2\lambda_1 \left(\sum_{h=1}^H \Omega_h \bar{x}_{1h} - \sum_{h=1}^H w_h \bar{X}_{1h} \right) - 2\lambda_2 \left(\sum_{h=1}^H \Omega_h \bar{x}_{2h} - \sum_{h=1}^H w_h \bar{X}_{2h} \right)$$

where λ_1 and λ_2 are Lagrangian multipliers. Differentiating L_2 with respect to Ω_h and setting equal to 0, we have

$$\Omega_h = \frac{w_h}{(1 - \lambda_1 q_h \bar{x}_{1h} - \lambda_2 q_h \bar{x}_{2h})^2}. \quad (21)$$

Under higher-order approximations,

$$\Omega_h = w_h (1 + 2\lambda_1 q_h \bar{x}_{1h} + 2\lambda_2 q_h \bar{x}_{2h}). \quad (22)$$

On putting the value of Ω_h from Equation (22) in Equations (8) and (9),

$$\lambda_1 \sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 + \lambda_2 \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} = \frac{1}{2} \left(\sum_{h=1}^H w_h \bar{X}_{1h} - \sum_{h=1}^H w_h \bar{x}_{1h} \right), \quad (23)$$

$$\lambda_1 \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} + \lambda_2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 = \frac{1}{2} \left(\sum_{h=1}^H w_h \bar{X}_{2h} - \sum_{h=1}^H w_h \bar{x}_{2h} \right). \quad (24)$$

Solving the system of equations for λ_1 and λ_2 and using the values of Ω_h , λ_1 , and λ_2 , the form of the estimator will be

$$\bar{Y}_{st2}^c = \sum_{h=1}^H w_h \bar{y}_h + \frac{1}{2} \left(\hat{\beta}_0 \hat{\bar{X}}_{1h} + \hat{\beta}_1 \hat{\bar{X}}_{2h} \right) \quad (25)$$

where

$$\hat{\beta}_0 = \frac{\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \sum_{h=1}^H w_h q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}$$

$$\hat{\beta}_1 = \frac{\sum_{h=1}^H w_h q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 - \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}$$

Case 3

The Lagrange function for the weights Ω_h , satisfying the calibration Equations (8) and (9) by minimizing the distance function $D_3(\Omega_h, w_h)$, is given by

$$L_3 = \sum_{h=1}^H \frac{1}{q_h} \left(\frac{\Omega_h}{w_h} - 1 \right)^2 - 2\lambda_1 \left(\sum_{h=1}^H \Omega_h \bar{x}_{1h} - \sum_{h=1}^H w_h \bar{X}_{1h} \right) - 2\lambda_2 \left(\sum_{h=1}^H \Omega_h \bar{x}_{2h} - \sum_{h=1}^H w_h \bar{X}_{2h} \right)$$

where λ_1 and λ_2 are the Lagrange multipliers. Differentiating L_3 with respect to Ω_h and setting equal to 0, we have

$$\Omega_h = w_h \left(1 + \lambda_1 w_h q_h \bar{x}_{1h} + \lambda_2 w_h q_h \bar{x}_{2h} \right). \quad (26)$$

On putting the value of Ω_h from Equation (26) in Equations (8) and (9),

$$\lambda_1 \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 + \lambda_2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} = \sum_{h=1}^H w_h \bar{X}_{1h} - \sum_{h=1}^H w_h \bar{x}_{1h}, \quad (27)$$

$$\lambda_1 \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} + \lambda_2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 = \sum_{h=1}^H w_h \bar{X}_{2h} - \sum_{h=1}^H w_h \bar{x}_{2h}. \quad (28)$$

Solving the system of equations for λ_1 and λ_2 ,

$$\lambda_1 = \frac{\sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 \hat{X}_{1h} - \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}, \quad (29)$$

$$\lambda_2 = \frac{-\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{1h} - \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \hat{X}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}. \quad (30)$$

The form of the estimator is

$$\bar{Y}_{st3}^c = \sum_{h=1}^H w_h \bar{y}_h + \hat{\beta}_0 \hat{X}_{1h} + \hat{\beta}_1 \hat{X}_{2h}, \quad (31)$$

where

$$\hat{\beta}_0 = \frac{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}$$

$$\hat{\beta}_1 = \frac{\sum_{h=1}^H w_h^2 q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 - \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}$$

The calibration estimator under $D_3(\Omega_h, w_h)$ for a single auxiliary variable is given as:

$$\bar{Y}_{TR3} = \sum_{h=1}^H w_h \bar{y}_h + \frac{\sum_{h=1}^H w_h \bar{X}_h - \sum_{h=1}^H w_h \bar{x}_h}{\sum_{h=1}^H w_h^2 q_h \bar{x}_h^2} \sum_{h=1}^H w_h^2 q_h \bar{x}_h \bar{y}_h \quad (32)$$

Case 4

The Lagrange function for the weights Ω_h , satisfying the calibration Equations (8) and (9) by minimizing the distance function $D_4(\Omega_h, w_h)$, is given by

$$L_4 = \sum_{h=1}^H \frac{1}{q_h} \left(\frac{\sqrt{\Omega_h}}{\sqrt{w_h}} - 1 \right)^2 - 2\lambda_1 \left(\sum_{h=1}^H \Omega_h \bar{x}_{1h} - \sum_{h=1}^H w_h \bar{X}_{1h} \right) - 2\lambda_2 \left(\sum_{h=1}^H \Omega_h \bar{x}_{2h} - \sum_{h=1}^H w_h \bar{X}_{2h} \right)$$

where λ_1 and λ_2 are the Lagrange multipliers. Differentiating L_4 with respect to Ω_h and setting equal to 0, we have

$$\Omega_h = w_h (1 + 2\lambda_1 w_h q_h \bar{x}_{1h} + 2\lambda_2 w_h q_h \bar{x}_{2h}). \quad (33)$$

On putting the value of Ω_h from Equation (33) in Equations (8) and (9),

$$\lambda_1 \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 + \lambda_2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} = \frac{1}{2} \left(\sum_{h=1}^H w_h \bar{X}_{1h} - \sum_{h=1}^H w_h \bar{x}_{1h} \right), \quad (34)$$

$$\lambda_1 \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} + \lambda_2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 = \frac{1}{2} \left(\sum_{h=1}^H w_h \bar{X}_{2h} - \sum_{h=1}^H w_h \bar{x}_{2h} \right). \quad (35)$$

Solving the system of equations for λ_1 and λ_2 ,

$$\lambda_1 = \frac{1}{2} \frac{\sum_{h=1}^H w_h^2 q_h \bar{x}_{2h} \hat{X}_{1h} - \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}, \quad (36)$$

$$\lambda_2 = \frac{1}{2} \frac{-\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \hat{X}_{1h} - \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \hat{X}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}. \quad (37)$$

The form of the estimator is

$$\bar{Y}_{st4}^c = \sum_{h=1}^H w_h \bar{y}_h + \frac{1}{2} \left(\hat{\beta}_0 \hat{X}_{1h} + \hat{\beta}_1 \hat{X}_{2h} \right), \quad (38)$$

where

$$\hat{\beta}_0 = \frac{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}$$

$$\hat{\beta}_1 = \frac{\sum_{h=1}^H w_h^2 q_h \bar{x}_{2h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 - \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{y}_h \sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h}}{\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h}^2 \sum_{h=1}^H w_h^2 q_h \bar{x}_{2h}^2 - \left(\sum_{h=1}^H w_h^2 q_h \bar{x}_{1h} \bar{x}_{2h} \right)^2}$$

Case 5

The Lagrange function for the weights Ω_h , satisfying the calibration Equations (8) and (9) by minimizing the distance function $D_5(\Omega_h, w_h)$, is given by

$$L_5 = \sum_{h=1}^H q_h^{-1} \left(-w_h \log \frac{\Omega_h}{w_h} + \Omega_h - w_h \right) - \lambda_1 \left(\sum_{h=1}^H \Omega_h \bar{x}_{1h} - \sum_{h=1}^H w_h \bar{X}_{1h} \right) - \lambda_2 \left(\sum_{h=1}^H \Omega_h \bar{x}_{2h} - \sum_{h=1}^H w_h \bar{X}_{2h} \right)$$

where λ_1 and λ_2 are the Lagrange multipliers. Differentiating L_5 with respect to Ω_h and setting equal to 0, we have

$$\Omega_h = w_h (1 + \lambda_1 q_h \bar{x}_{1h} + \lambda_2 q_h \bar{x}_{2h}), \quad (39)$$

which is same as the value of Ω_h in Equation (11). Thus, the calibration estimator under $D_5(\Omega_h, w_h)$ will have the same form of estimator as in Equation (19).

A Simulation Study

The main objective of this study is to introduce calibration estimators for finite population mean under two auxiliary variables by considering different distance measures using stratified sampling design. The percent Relative Root Mean Squared Error (RRMSE%) is considered as a performance criterion to judge the performance of the proposed calibration estimators with the available existing calibration estimators by means of the simulation study. The proposed calibration estimators under different distance functions with two auxiliary variables are compared with the estimators under single auxiliary variable. Thus, four different artificial populations are considered to judge the performance of the proposed estimators. These populations involve three strata, i.e., $h = 1, 2, 3$, and within each stratum the populations following a particular distribution are shown in Table 1 (for more details see, e.g., Ozgul, 2018).

Table 1. Probability distributions and their parameters value for the study and the auxiliary variables

Population 1	Population 2
$f(y_{hi}) = \frac{1}{\Gamma(2.9)} y^{2.9-1} e^{-y_{hi}}$	$f(y_{hi}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_{hi})^2}$
$f(x_{1hi}) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{1hi}}{5}\right)^2}$	$f(x_{1hi}) = \frac{1}{10-5}$ for $x_{1hi} \in [5,10]$
$f(x_{2hi}) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{2hi}}{2}\right)^2}$	$f(x_{2hi}) = \frac{1}{5-3}$ for $x_{2hi} \in [3,5]$
Population 3	Population 4
$f(y_{hi}) = \frac{1}{\Gamma(1.2)} y^{1.2-1} e^{-y_{hi}}$	$f(y_{hi}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_{hi}-5)^2}$
$f(x_{1hi}) = \frac{1}{\Gamma(0.9)} y^{0.9-1} e^{-x_{1hi}}$	$f(x_{1hi}) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{1hi}}{5}\right)^2}$
$f(x_{2hi}) = \frac{1}{\Gamma(0.2)} y^{0.2-1} e^{-x_{2hi}}$	$f(x_{2hi}) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{2hi}}{2}\right)^2}$

The auxiliary variables are normally distributed, slightly positively skewed, and strongly positively skewed in first stratum, second stratum and third stratum, respectively. The stratum sizes N_h for each population determined as $N_1 = 1000$, $N_2 = 2000$, and $N_3 = 3000$. The sample size ($n = 300$) is drawn by using simple random sampling without replacement and for each stratum the sample size n_h is defined by using proportion allocation as given below:

$$n_h = nW_h.$$

Three different levels of the correlation ($\rho = 0.50, 0.70, 0.90$) between study and auxiliary variables are considered for each stratum, respectively. Therefore, each stratum is generated based on these correlation coefficients. The Cholesky decomposition of the covariance matrix is used to generate the correlated variables in the simulation study. So, the covariance matrix is defined as:

$$\mathbf{Q} = \begin{bmatrix} \sigma_{y_h}^2 & \sigma_{y_h x_{1h}} & \sigma_{y_h x_{2h}} \\ \sigma_{y_h x_{1h}} & \sigma_{x_{1h}}^2 & \sigma_{x_{1h} x_{2h}} \\ \sigma_{y_h x_{2h}} & \sigma_{x_{1h} x_{2h}} & \sigma_{x_{2h}}^2 \end{bmatrix}, \quad (40)$$

where

$$\sigma_{y_h}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2, \sigma_{x_{1h}}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (x_{1hi} - \bar{X}_{1h})^2,$$

$$\sigma_{x_{2h}}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (x_{2hi} - \bar{X}_{2h})^2$$

are the population variances of the study variable, first auxiliary variable, and second auxiliary variable, respectively.

$$\sigma_{y_h x_{1h}} = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(x_{1hi} - \bar{X}_{1h})$$

is the population covariance between the study variable and the first auxiliary variable,

$$\sigma_{y_h x_{2h}} = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(x_{2hi} - \bar{X}_{2h})$$

is the population covariance between the study variable and the second auxiliary variable, and

$$\sigma_{x_{1h} x_{2h}} = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (x_{1hi} - \bar{X}_{1h})(x_{2hi} - \bar{X}_{2h})$$

is the population covariance between the auxiliary variables in the h^{th} stratum. The anticipated covariance matrix, denoted by \mathbf{Q}^* , is defined in Equation (41):

$$\mathbf{Q}^* = \begin{bmatrix} \sigma_{y_h}^2 & \rho_{y_h x_{1h}} \sqrt{\sigma_{y_h}^2 \sigma_{x_{1h}}^2} & \rho_{y_h x_{2h}} \sqrt{\sigma_{y_h}^2 \sigma_{x_{2h}}^2} \\ \rho_{y_h x_{1h}} \sqrt{\sigma_{y_h}^2 \sigma_{x_{1h}}^2} & \sigma_{x_{1h}}^2 & \rho_{x_{1h} x_{2h}} \sqrt{\sigma_{x_{1h}}^2 \sigma_{x_{2h}}^2} \\ \rho_{y_h x_{2h}} \sqrt{\sigma_{y_h}^2 \sigma_{x_{2h}}^2} & \rho_{x_{1h} x_{2h}} \sqrt{\sigma_{x_{1h}}^2 \sigma_{x_{2h}}^2} & \sigma_{x_{2h}}^2 \end{bmatrix}, \quad (41)$$

where $\rho_{y_h x_{1h}}$, $\rho_{y_h x_{2h}}$, and $\rho_{x_{1h} x_{2h}}$ are the population correlation coefficients between the study and first auxiliary variable, the study and second auxiliary variable, and the first and second auxiliary variables, respectively, in the h^{th} stratum.

Since, the Cholesky decomposition of the matrices \mathbf{Q} and \mathbf{Q}^* can be written as

$$\mathbf{Q} = \mathbf{C}_d \mathbf{C}_d', \quad \mathbf{Q}^* = \mathbf{C}_d^* \mathbf{C}_d^{*'}$$

where \mathbf{C}_d and \mathbf{C}_d^* are the Cholesky decompositions of the covariance matrices of \mathbf{Q} and \mathbf{Q}^* , respectively. The correlated variables are obtained as

$$\mathbf{Z} = \mathbf{H} \mathbf{C}_d^{-1} \mathbf{C}_d^*,$$

where \mathbf{H} are the uncorrelated original values and \mathbf{Z} are the correlated values.

We selected $R = 50000$ ($j = 1, 2, \dots, 50000$) samples of size $n = 300$) from four different populations under stratified sampling. To judge the performance of the proposed estimators, we defined RRMSE in percentage:

$$\text{RRMSE}(\bar{y}_{\text{st}(\theta)}^c) = \sqrt{\frac{1}{M} \sum_{i=1}^M (\bar{y}_{\text{st}(\theta)}^c - \bar{Y})^2} \times 100\%, \quad (42)$$

where θ is the estimator under D_1, D_2, D_3, D_4 under single and multi-auxiliary variables and M is the 50000 times replication of the samples. The RRMSE(%) and percent Relative Bias (RB) are shown in Tables 2 and 3. The proposed calibration estimators under different distance function with two auxiliary variables always outperformed to the available single auxiliary variable-based calibration estimators. The simulation results show that the RRMSE(%) of the proposed estimators always less than the existing methods for both symmetric and skewed population. The RB(%) is also minimum of proposed estimators when the population is symmetric.

CALIBRATION ESTIMATOR USING DIFFERENT DISTANCE MEASURES

Table 2. The RRMSE(%) of the estimators

Population	\bar{Y}_{TR1}^c	\bar{Y}_{TR3}^c	\bar{Y}_{st1}^c	\bar{Y}_{st2}^c	\bar{Y}_{st3}^c	\bar{Y}_{st4}^c
1	4.36118	14.74066	4.18992	2.90169	2.63285	2.46374
2	99.71640	99.79625	77.25652	88.53831	26.93738	52.03558
3	2.08660	8.99232	1.96409	3.13342	3.91231	4.64022
4	0.48799	1.82036	0.39502	0.59626	1.75235	0.48761

Note: $\bar{Y}_{TR1}^c, \bar{Y}_{TR3}^c$ are the calibration estimators based on single auxiliary variable and $\bar{Y}_{st1}^c - \bar{Y}_{st4}^c$ are the estimators based on two auxiliary variables

Table 3. The RB(%) of the estimators

Population	\bar{Y}_{TR1}^c	\bar{Y}_{TR3}^c	\bar{Y}_{st1}^c	\bar{Y}_{st2}^c	\bar{Y}_{st3}^c	\bar{Y}_{st4}^c
1	0.10095	0.04328	0.21260	0.02637	0.51281	0.38908
2	99.61680	97.72390	74.51697	87.25654	9.90099	45.04756
3	0.58883	0.43212	0.68120	0.69430	0.66072	0.68406
4	0.11696	0.26094	0.03394	0.02212	0.19372	0.10201

Note: $\bar{Y}_{TR1}^c, \bar{Y}_{TR3}^c$ are the calibration estimators based on single auxiliary variable and $\bar{Y}_{st1}^c - \bar{Y}_{st4}^c$ are the estimators based on two auxiliary variables

An Application

In this section, we illustrate a real-life application and its results. For this we consider the Forced Expiratory Volume (FEV) data set which was used by Singh (2013). The real data set FEV is an index of pulmonary function that measures the volume of air expelled after one second of constant effort and can be downloaded from <http://jse.amstat.org/datasets/fev.dat.txt>. The FEV data set was gathered in East Boston, Massachusetts, 1980, on 654 children aged from 3 to 19 years who were seen in the childhood respiratory disease. The study variable Y is defined as FEV, and the two auxiliary variables are X_1 , age from 3-19 years, and X_2 , height in inches. For this data set, $\bar{Y} = 2.6367$, $\bar{X}_1 = 9.5841$, $\bar{X}_2 = 60.5444$, $\rho(X_1, Y) = 0.75646$, $\rho(X_2, Y) = 0.86814$, and $\rho(X_1, X_2) = 0.79194$. Our main aim is to estimate $\bar{Y} = 2.6367$ (assumed unknown), when $\bar{X}_1 = 9.5841$ and $\bar{X}_2 = 60.5444$ are assumed to be known. A pictorial representation of the real-life dataset is shown in Figure 1.

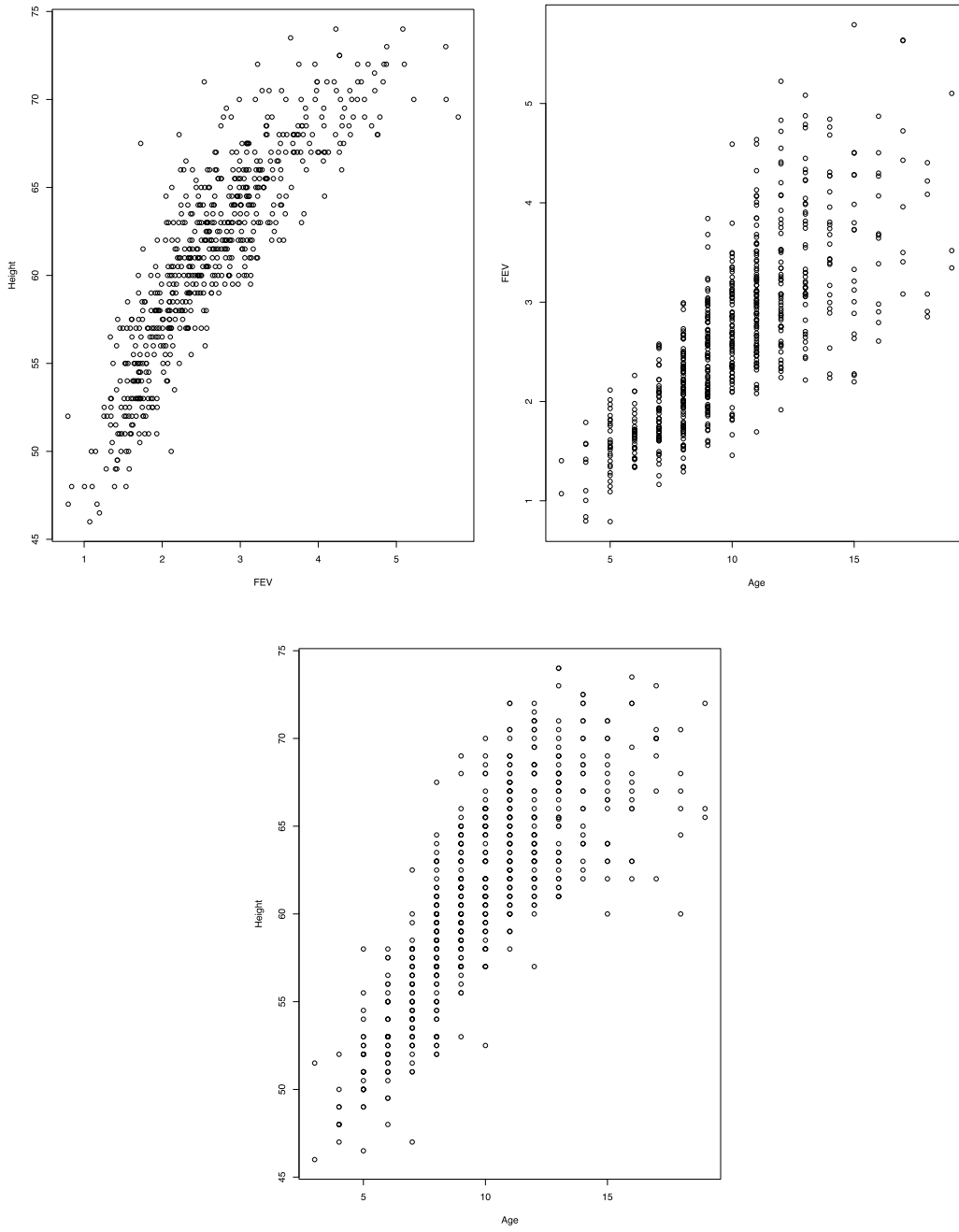


Figure 1. Scatter plots of the three variables considered in the study

CALIBRATION ESTIMATOR USING DIFFERENT DISTANCE MEASURES

Table 4. The RB(%) of the estimators

Sample proportion	\bar{Y}_{TR1}^c	\bar{Y}_{TR3}^c	\bar{Y}_{st1}^c	\bar{Y}_{st2}^c	\bar{Y}_{st3}^c	\bar{Y}_{st4}^c
0.10	0.03950	0.04615	0.01779	0.00739	0.02129	0.05326
0.20	0.02028	0.06530	0.01285	0.02056	0.02099	0.00879
0.25	0.02355	0.02155	0.00006	0.01813	0.01112	0.02021
0.30	0.01879	0.03494	0.00416	0.02333	0.02307	0.02546
0.35	0.02238	0.03277	0.00715	0.01284	0.01099	0.01383
0.40	0.00425	0.01670	0.00323	0.01319	0.00616	0.01437

Table 5. The RRMSE(%) of the estimators

Sample proportion	\bar{Y}_{TR1}^c	\bar{Y}_{TR3}^c	\bar{Y}_{st1}^c	\bar{Y}_{st2}^c	\bar{Y}_{st3}^c	\bar{Y}_{st4}^c
0.10	2.68373	5.49804	2.50300	3.01185	2.58724	3.14774
0.20	1.77938	3.63132	1.66208	1.98960	1.73115	2.07870
0.25	1.53051	3.15137	1.42790	1.71235	1.49165	1.79023
0.30	1.34280	2.77982	1.26252	1.49914	1.31864	1.56496
0.35	1.21450	2.42789	1.12322	1.36014	1.16936	1.42141
0.40	1.08345	2.17341	1.00891	1.20786	1.05171	1.26214

The estimated RB(%) and RRMSE(%) of the estimators using the real-life data set are shown in Table 4 and Table 5, respectively. It can be seen that the RB(%) and RRMSE(%) of the estimators \bar{Y}_{TR1}^c and \bar{Y}_{TR3}^c under single auxiliary variables are always greater than the estimators $\bar{Y}_{st1}^c - \bar{Y}_{st3}^c$ under two auxiliary variables. Therefore, we conclude that proposed estimators under two auxiliary variables outperforms to the usual estimator under single auxiliary variable. Moreover, on comparing the efficiency of the estimators under different distance measures, the Chi-square distance function and minimum entropy distance function performs uniformly better than any other available in literature. The proposed estimators \bar{Y}_{st1}^c and \bar{Y}_{st3}^c are always exhibit minimum RB(%) and RRMSE(%) compared to the others.

Conclusion

Calibration estimation is a method of adjusting the original design weights to improve sample survey estimates by using auxiliary information such as the known population total (or mean) of the auxiliary variables. A calibration-based estimator uses calibrated weights that are determined to minimize a given distance measure

to the original design weights while satisfying a set of constraints related to the auxiliary information. Thus, this new approach of calibration estimation has gained much attention in recent years. In this study, we proposed calibration-based estimators of finite population mean with different weights in stratified sampling based on two auxiliary variables. We define some new weights by means of different distance functions. The derived estimators are compared with the existing single auxiliary variable based calibrated estimators. For this purpose, a simulation study has been made to evaluate the performance of the proposed estimators in terms of RB(%) and RRMSE(%). The simulation results show that calibrated estimators under two auxiliary variables outperforms for single auxiliary variable calibrated estimators and always the RB(%) and RRMSE(%) are minimum for symmetric population. Based on the simulation and real-life data results, the proposed estimators are more efficient and reliable than the single auxiliary variable based calibrated estimators.

References

- Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376-382. doi: 10.1080/01621459.1992.10475217
- Kim, J.-M., Sungur, E. A., & Heo, T.-Y. (2007). Calibration approach estimators in stratified sampling. *Statistics and Probability Letters*, 77(1), 99-103. doi: 10.1016/j.spl.2006.05.015
- Koyuncu, N. (2012). *Application of calibration method to estimators in sampling theory* [Unpublished doctoral thesis]. Hacettepe University.
- Koyuncu, N., & Kadilar, C. (2013). Calibration estimator using different distance measures in stratified random sampling. *International Journal of Modern Engineering Research*, 3(1), 415-419. Retrieved from http://www.ijmer.com/papers/Vol3_Issue1/CT31415419.pdf
- Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika*, 45(1/2), 154-165. doi: 10.2307/2333053
- Ozgul, N. (2018). New calibration estimator based on two auxiliary variables in stratified sampling. *Communications in Statistics – Theory and Methods*, 48(6), 1481-1492. doi: 10.1080/03610926.2018.1433852

CALIBRATION ESTIMATOR USING DIFFERENT DISTANCE MEASURES

Raj, D. (1965). On a method of using multi-auxiliary information in sample surveys. *Journal of the American Statistical Association*, 60(309), 270-277. doi: 10.1080/01621459.1965.10480789

Shukla, G. K. (1965). Multivariate regression estimate. *Journal of the Indian Statistical Association*, 3, 202-211.

Singh, S. (2013). A dual problem of calibration of design weights. *Statistics*, 47(3), 566-574. doi: 10.1080/02331888.2011.615397

Tracy, D. S., Singh, S., & Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology*, 29(1), 99-104.