Journal of Modeling and Simulation of Materials
ISSN: 2582-2365
Volume 5, Issue 1, pp. 21-30, 2022
DOI: https://doi.org/10.21467/jmsm.5.1.21-30

RESEARCH ARTICLE

Calibration Ratio Estimators of Population Mean Using Median of Auxiliary Variable

Effiong Eyo Eyo* and I. E. Enang

University of Calabar

*Corresponding Author's e-mail: effiongeyo52@gmail.com
Received: 01 July 2022 / Revised: 24 December 2022 / Accepted: 26 December 2022 / Published: 29 December 2022

ABSTRACT

In this study we propose a calibration ratio estimator and a calibration separate ratio-product estimator of population mean of study variable under stratified sampling using the median of auxiliary variable. The calibration estimator used calibrated weight determined to minimize a chi-square distance measure subject to a set of constraint related to the auxiliary variable in other to increase precision of the estimators. The median of the auxiliary variable was used in defining the calibration constraints. The variances of the proposed estimators were also obtained. An empirical study to ascertain the performance of these estimators using simulated data under underlying distribution assumption of Student-T distribution, Cauchy distribution, Lognormal distribution, and Standard normal distribution with varying sample sizes of 10%, 20%, and 25% were carried out. The result of simulation reveals that when the underlying distribution is Student-T, at 10% sample size, the efficiency performance of the proposed calibration separate ratio-product estimator is better than other competing estimators. As the sample size is increased to 20% and 25%, the efficiency performance of the existing stratified ratio estimator and existing calibration ratio estimator respectively become better than the other estimators. Under the skewed distributions (Cauchy and Lognormal) and the standard normal distribution, it is observed that the proposed calibration ratio estimator is better than other competing estimators in terms of efficiency, consistency and reliability. The result also reveals that under the lognormal distribution, the conventional stratified ratio estimator and the conventional calibration ratio estimator give the same result.

Keywords: Calibration Estimation, Stratified Sampling, Ratio Estimators

1 Introduction

The simplest estimator for estimating population mean of a study variable is the sample mean, obtained by using simple random sampling without replacement. If the population parameters are not known, supplementary information may be obtained from space (related area to the study variable) or from time (from previous research) and used to estimate parameters of the study variable. In survey sampling, using auxiliary information is observed to yield extensive gain in performance (better efficiency, precision, less bias etc.) over the estimators lacking such information. Auxiliary information is obtained from an auxiliary variable which is a variable having high correlation with the study. Noor-ul-Amin *et al.* [1]used auxiliary information in the estimation of population mean. Auxiliary variables can be either positively correlated or negatively correlated with the study variable. When the parameters of the auxiliary variable X such as Population Mean, Co-efficient of Variation, Co-efficient of Kurtosis, Co-efficient of Skewness, Median etc., are known, a number of estimators such as linear regression, ratio, and product estimators and their modifications like product-ratio estimators, exponential estimators etc., can be used for improved estimation of the population parameters of the study variable. When the auxiliary variable is positively correlated with the study variable, a ratio estimation technique is used to improve the estimators' performance. However, when the correlation is negative, a product estimation technique is employed to improve estimators' performance. Over the years, different researchers have used these two forms of estimators (ratio and product estimation) to improve the quality of estimation with respect to the type of correlation existing between the auxiliary variable and the study variable. Singh *et al.* [2] proposed a two-parameter ratio-product estimator in post stratification, and derived conditions under which the proposed estimators have smaller mean squared error than some conventional estimators. Zaman *et al.* [3] proposed exponential ratio estimators in the stratified two-phase sampling utilizing an auxiliary attribute.

Recently, other parameters of the auxiliary variable such as the median, coefficient of skewness, coefficient of kurtosis, coefficient of correlation has been used to estimate the population parameters of the study variable. Subramani [4] suggested a median ratio-based estimator of the population mean, $\bar{Y}$.

Calibration technique can also be used boost precision of an estimator. Calibration is commonly used when auxiliary information is available to increase the precision of estimators of population parameters. This is done by modifying the original design weights using the known population parameters, in practice population totals or population means, of the auxiliary variables. Garg *et al.* [5] proposed a calibration estimator of the finite population mean in stratified sampling using the median of auxiliary variable. Rai *et al.*[6] proposed calibration-based estimators using different distance measures under two auxiliary variables. Singh *et al.* [7] suggested new technique to calibrate estimators of the variance of simple mean, ratio and regression estimators under different sampling schemes.

2 Research and method

2.1 Notations and some existing estimators

Suppose the finite population $\boldsymbol{U}$ of N elements $\boldsymbol{U} = (\boldsymbol{U_1}, \boldsymbol{U_2}, \dots, \boldsymbol{U_N})$ and consist of $\boldsymbol{L}$ strata with $\boldsymbol{N_h}$ units in the h^{th} stratum from which a simple random sample of size $\boldsymbol{n_h}$ is obtained without replacement. Given that total population size $\boldsymbol{N} = \sum_{h=1}^{L} \boldsymbol{N_h}$ and the sample size $\boldsymbol{n} = \sum_{h=1}^{L} \boldsymbol{n_h}$, respectively. Associated with the ith element of the hth stratum are $\boldsymbol{y_{hi}}$ and $\boldsymbol{x_{hi}}$ with $\boldsymbol{x_{hi}} > \boldsymbol{0}$, being the covariate; where $\boldsymbol{y_{hi}}$ is the y value of the i^{th} element in stratum $\boldsymbol{h}$, and $\boldsymbol{x_{hi}}$ is the $\boldsymbol{x}$ value of the i^{th} element in $\boldsymbol{h}$, $\boldsymbol{h} = \boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{L}$ and $\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{N_h}$ where y and x are the study and auxiliary variables respectively. For the hth stratum, let $\boldsymbol{W_h} = \frac{\boldsymbol{N_h}}{\boldsymbol{N}}$ be the stratum weights and $\boldsymbol{f_h} = \frac{\boldsymbol{n_h}}{\boldsymbol{N_h}}$, the sample fraction.

Let the h^{th} stratum means of the study variable y and the auxiliary variable x $\left(\bar{\boldsymbol{y}}_{\boldsymbol{h}} = \sum_{h=1}^{L} \frac{\boldsymbol{y_{hi}}}{\boldsymbol{n_{hi}}};\ \bar{\boldsymbol{x}}_{\boldsymbol{h}} = \sum_{h=1}^{L} \frac{\boldsymbol{x_{hi}}}{\boldsymbol{n_{hi}}}\right)$ be the unbiased estimator of the population mean $\left(\bar{\boldsymbol{Y}}_{\boldsymbol{h}} = \sum_{h=1}^{L} \frac{\boldsymbol{Y_{hi}}}{\boldsymbol{N_h}};\ \bar{\boldsymbol{X}}_{\boldsymbol{h}} = \sum_{h=1}^{L} \frac{\boldsymbol{X_{hi}}}{\boldsymbol{N_h}}\right)$ of $\boldsymbol{y}$ *and* $\boldsymbol{x}$ respectively, based on n_h observations.

The Horvitz Thompson stratified sampling estimator is given as:

$$\bar{\boldsymbol{y}}_{\boldsymbol{st}}(\boldsymbol{\alpha}) = \sum_{h=1}^{L} \boldsymbol{W_h} \bar{\boldsymbol{y}}_{\boldsymbol{h}} \quad (1)$$

Where, $\boldsymbol{W_h} = \frac{\boldsymbol{n_h}}{\boldsymbol{N_h}}$, is the stratum weight, $\bar{\boldsymbol{y}}_{\boldsymbol{h}} = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{h=1}^{L} \boldsymbol{y_h}$, and the variance of $\bar{y}_{st}(\alpha)$ is given as

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

Where, $S_{hy}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$, $f_h = \frac{n_h}{N_h}$

The conventional ratio type estimator in stratified sampling is given as:

$$\bar{y}_{st}(\alpha) = \sum_{h=1}^{L} W_h \bar{y}_h R_h \quad (2)$$

Where, $R_h = \frac{\bar{X}_{hx}}{\bar{x}_{hx}}$

And the variance is

$$V(\bar{y}_{rs}) = \sum_{h=1}^{L} W_{h=1}^2 \frac{1-f_h}{n_h} \left(S_{hy}^2 + R_h^2 S_{hx}^2 - 2 R_h S_{hxy}\right) \quad (3)$$

The calibration ratio estimator under stratified sampling is given as:

$$\bar{y}_{st}{}^{*}(\alpha_h) = \sum_{h=1}^{L} W_h^{*}\, \bar{y}_h R_h \quad (4)$$

Garg *et al.* [5] proposed a calibration estimator of the finite population mean in stratified sampling using the median of auxiliary variable as follow:

$$\bar{y}_{md} = \sum_{h=1}^{L} \Omega_h \bar{y}_h \qquad (5)$$

Where Ω_h , $h = 1, 2, \dots, L$ are the calibration weight obtained by minimizing the chi-square distance measure $\sum_{h=1}^{L}\left(\frac{\Omega_h - W_h}{Q_h W_h}\right)^2$, subject to the two calibration constraints:

$$\sum_{h=1}^{L} \Omega_h m_h = \sum_{h=1}^{L} W_h M_h \qquad (6)$$

Where m_h and M_h are the sample and population median of auxiliary variable, respectively.

The Lagrange function is defined as:

$$L = \sum_{h=1}^{l} \frac{(\Omega_h - W_h)^2}{W_h Q_h} - 2\lambda\left(\sum_{h=1}^{L} \Omega_h m_h - \sum_{h=1}^{L} W_h M_h\right) \qquad (7)$$

Where λ is the Lagrange multipliers. To determine the optimum value of Ω_h , differentiate the Lagrange function in (7) with respect to Ω_h and equate to zero. Thus, the calibration weight can be obtained as:

$$\Omega_h = W_h + \lambda(W_h Q_h m_h) \qquad (8)$$

Here λ is determined by substituting the value of Ω_h from equation (8) to equation (6), so this leads to a calibrated weight given as:

$$\Omega_h = W_h + W_h Q_h m_h \left[\frac{\sum_{h=1}^{L} W_h (M_h - m_h)}{\sum_{h=1}^{L} W_h Q_h {m_h}^2}\right] \qquad (9)$$

After substituting the value of Ω_h from equation (9) to (5) ,we obtain the proposed calibrated estimator as:

$$\bar{y}_{md} = \sum_{h=1}^{L} W_h \bar{y}_h + \hat{\beta}_{md}\left[\sum_{h=1}^{L} W_h (M_h - m_h)\right] \qquad (10)$$

Where $\hat{\beta}_{md} = \frac{\sum_{h=1}^{L} W_h Q_h m_h \bar{y}_h}{\sum_{h=1}^{L} W_h Q_h {m_h}^2}$

Vishwakarma *et al.* [8] proposed a separate ratio-product estimator for population mean in stratified random sampling using auxiliary information as:

$$\hat{\bar{Y}}_{RP}^{(s)} = \sum_{h=1}^{L} W_h \bar{y}_h \left\{\alpha_h \frac{\bar{X}_h}{\bar{x}_h} + (1 - \alpha_h)\frac{\bar{x}_h}{\bar{X}_h}\right\} \qquad (11)$$

With the variance of $\hat{\bar{Y}}_{RP}^{(s)}$ to the first order of approximation given as

$$Var\left(\hat{\bar{Y}}_{RP}^{(s)}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left[C_{hy}^2 + (1 - 2\alpha_h)\{(1 - 2\alpha_h) + 2K_h\}C_{hx}^2\right] \qquad (12)$$

Motivated by [8], [9] who proposed a separate ratio-product estimator for population mean in stratified random sampling using calibration estimation theory as follow:

$$\bar{y}_{st}{}^*(\alpha_h) = \sum_{h=1}^{L} W_h^* \, \bar{y}_h \lambda \qquad (13)$$

Where the coefficient $\lambda = \left\{\alpha_h \frac{\bar{X}_h}{\bar{x}_h} + (1 - \alpha_h)\frac{\bar{x}_h}{\bar{X}_h}\right\}$ and W_h^* is the new weights called the calibration weight and are chosen such that a chi-square-type loss function of the form

$$\sum_{h=1}^{L}\left(\frac{W_h^* - W_h}{Q_h W_h}\right)^2 \qquad (14)$$

Is minimized subject to the calibration constrain of the form

$$\sum_{h=1}^{L} W_h^* S_{hx}^2 = V(\bar{X}_{st}) \qquad (15)$$

Minimizing the loss function (14) subject to the calibration constraint (15) leads to the calibration weight for stratified sampling given by

$$W_h^* = W_h + \left(V(\bar{X}_{st}) - \sum_{h=1}^{L} W_h^* S_{hx}^2\right)\frac{Q_h W_h S_{hx}^2}{\sum_{h=1}^{L} Q_h W_h (S_{hx}^2)^2} \qquad (16)$$

Let ${W_h^*}^2 = \left[W_h + \left(V(\bar{X}_{st}) - \sum_{h=1}^{L} W_h^* S_{hx}^2\right)\frac{Q_h W_h S_{hx}^2}{\sum_{h=1}^{L} Q_h W_h (S_{hx}^2)^2}\right]^2$

And setting the turning parameter $Q_h = S_{hx}^{-2}$, then

$${W_h^*}^2 = W_h^2 \left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \qquad (17)$$

Where $V(\bar{X}_{st}) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{hx}^2$ and $\hat{V}(\bar{X}_{st}) = \sum_{h=1}^{L} W_h S_{hx}^2$

Substituting (17) into (13) gives the proposed estimator of the mean as

$$\bar{y}_{st}{}^{*}(\alpha_h) = \left[\sum_{h=1}^{L} W_h + \left(V(\bar{X}_{st}) - \sum_{h=1}^{L} W_h^* S_{hx}^2\right) \frac{Q_h W_h S_{hx}^2}{\sum_{h=1}^{L} Q_h W_h (S_{hx}^2)^2}\right] \bar{y}_h \lambda \qquad (18)$$

The MSE of the estimator $\bar{y}_{st}{}^{*}(\alpha_h)$ is given as

$$MSE\left(\bar{y}_{st}{}^{*}(\alpha_h)\right) = \lambda^2 \sum_{h=1}^{L} {W_h^*}^2 \gamma_h S_{hy}^2 + \bar{Y}^2(\lambda-1)^2 \qquad (19)$$

Writing (19) with respect to (17) gives

$$MSE\left(\bar{y}_{st}{}^{*}(\alpha_h)\right) = \lambda^2 \left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2 + \bar{Y}^2(\lambda-1)^2 \qquad (20)$$

2.2 The proposed calibration ratio type estimators under one constraint:

In this section we present some calibration ratio type estimators under one constraint using the chi-square distance measures

Theorem 1

Given the ratio estimator as

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

Where $R_h = \frac{M_{hx}}{m_{hx}}$

A calibration ratio type estimator $\bar{y}_{stm}^*$ for population mean $\bar{Y}$ given as

$$\bar{y}_{stm}^* = \sum_{h=1}^{L} \left[W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2}\right] \bar{y}_h R_h$$

Is proposed with variance

$$V(\bar{y}_{stm}^*) = (R_h)^2 \sum_{h=1}^{L} \left[W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2}\right]^2 \theta_{nh} S_{yh}^2 + \bar{Y}^2\left(R_h^{-1} - 1\right)^2$$

Proof:

Given the ratio estimator

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

Define a calibration estimator of the form

$$\bar{y}_{stm}^* = \sum_{h=1}^{L} W_h^* \bar{y}_h R_h \qquad (21)$$

Where the coefficient $R_h = \frac{M_{hx}}{m_{hx}}$ and W_h^* is the new weight chosen such that a chi-square type loss function

$$\sum_{h=1}^{L} \left(\frac{W_h^* - W_h}{Q_h W_h}\right)^2 \qquad (22)$$

is minimized subject to a calibration constraint

$$\sum_{h=1}^{L} W_h^* m_{hx} = \sum_{h=1}^{L} W_h M_{hx} \qquad (23)$$

The Lagrange's function, using Calibration constraints and chi-square distance measure is

$$\Delta = \sum_{h=1}^{L} \left(\frac{W_h^* - W_h}{Q_h W_h}\right)^2 - 2\lambda_1\left(\sum_{h=1}^{L} W_h^* m_{hx} - \sum_{h=1}^{L} W_h M_{hx}\right) \qquad (24)$$

Differentiating (24) with respect to W_h^*, setting result equal to zero and solving for W_h^*

$$W_h^* = W_h + \lambda_1 m_{hx} Q_h W_h \qquad (25)$$

Putting equation (25) into (23) and solving for λ_1 gives

$$\lambda_1 = \frac{\sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2} \qquad (26)$$

Substituting for λ_1 from equation (26) in (25) gives the calibration weight for stratified sampling:

$$W_h^* = W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2} \qquad 2 \quad (27)$$

Substituting for W_h^* from equation (27) into (21) gives the population estimator

$$\bar{y}_{stm}^* = \sum_{h=1}^{L} \left[W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2}\right] \bar{y}_h R_h \qquad (28)$$

By substituting the turning parameter $Q_h = M_{hx}^{-1}$ we have the calibration ratio type estimator as

$$\bar{y}^*_{stm} = \sum_{h=1}^{L}\left[W_h + \frac{(M_{hx}^{-1}W_h m_{hx})\sum_{h=1}^{L}W_h(M_{hx}-m_{hx})}{\sum_{h=1}^{L}W_h M_{hx}^{-1}m_{hx}^2}\right]\bar{y}_h R_h \quad (29)$$

2.3 The Variance of the proposed estimator calibration estimator

The variance of the calibration ratio type estimator $\bar{y}^*_{stm}$ for population mean $\bar{Y}$ using median as auxiliary variable is defined as

$$V(\bar{y}^*_{stm}) = (\bar{y}^*_{stm} - \bar{Y})^2$$

$$\bar{y}^*_{stm} - \bar{Y} = \sum_{h=1}^{L} W_h^*\bar{y}_h R_h - \bar{Y} \quad (30)$$

Squaring both sides of equation (30) and taking expectation gives

$$E(\bar{y}^*_{stm} - \bar{Y})^2 = \left[\sum_{h=1}^{L} W_h^*\bar{y}_h R_h - \bar{Y}\right]^2 \quad (31)$$

$$E(\bar{y}^*_{stm} - \bar{Y})^2 = E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h R_h\right)^2 - 2R_h\bar{Y}E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right) + \bar{Y}^2 \quad (32)$$

$$= var\left(\sum_{h=1}^{L} W_h^*\bar{y}_h R_h\right) + \left[E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h R_h\right)\right]^2 - 2\lambda R_h E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right) + \bar{Y}^2$$

Where $E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right)^2 = var\left(\sum_{h=1}^{L} W_h^*\bar{y}_h R_h\right) + \left[E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h R_h\right)\right]^2$

$$= (R_h)^2 var\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right) + (R_h)^2 E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right)^2 - 2R_h\bar{Y}E\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right) + \bar{Y}^2$$

$$= (R_h)^2 var\left(\sum_{h=1}^{L} W_h^*\bar{y}_h\right) + (R_h)^2\left(\sum_{h=1}^{L} W_h^*\bar{Y}_h\right)^2 - 2R_h\bar{Y}\left(\sum_{h=1}^{L} W_h^*\bar{Y}_h\right) + \bar{Y}^2$$

$$= (R_h)^2\sum_{h=1}^{L} W_h^{*2} var(\bar{y}_h) + (R_h)^2\bar{Y}^2 - 2R_h\bar{Y}^2 + \bar{Y}^2$$

$$E(\bar{y}^*_{stm} - \bar{Y})^2 = (R_h)^2\sum_{h=1}^{L} W_h^{*2} var(\bar{y}_h) + \bar{Y}^2((R_h)^2 - 2R_h + 1) \quad (33)$$

$$E(\bar{y}^*_{stm} - \bar{Y})^2 = (R_h)^2\sum_{h=1}^{L} W_h^{*2} var(\bar{y}_h) + \bar{Y}^2(R_h - 1)^2$$

$$V(\bar{y}^*_{stm}) = (R_h)^2\sum_{h=1}^{L} W_h^{*2}\theta_h S_{yh}^2 + \bar{Y}^2(R_h - 1)^2$$

$$V(\bar{y}^*_{stm}) = (R_h)^2\sum_{h=1}^{L}\left[W_h + \frac{(Q_h W_h m_{hx})\sum_{h=1}^{L}W_h(M_{hx}-m_{hx})}{\sum_{h=1}^{L}W_h Q_h m_{hx}^2}\right]^2\theta_h S_{yh}^2 + \bar{Y}^2(R_h - 1)^2 \quad (34)$$

Where $\theta_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$

2.4 The proposed calibration separate ratio-product type estimators

Theorem 2: Given the separate ratio-product estimator of population mean

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

Where $\lambda = \left\{\alpha_h\frac{M_{hx}}{m_h} + (1-\alpha_h)\frac{m_h}{M_{hx}}\right\}$

A calibration separate ratio-product type estimator $\bar{y}^{*\,RP}_{stm}$ for population mean $\bar{Y}$ given as

$$\boldsymbol{\bar{y}^{*\,RP}_{stm}} = \sum_{h=1}^{L} W_h\bar{y}_h\lambda + \frac{\sum_{h=1}^{L}W_h^2\bar{y}_h\lambda(Q_h W_h m_{hx})(M_{hx}-m_{hx})}{\sum_{h=1}^{L}W_h Q_h m_{hx}^2}$$

Is proposed with variance

$$V(\bar{y}^{*RP}_{stm}) = \lambda^2\sum_{h=1}^{L}\left[W_h + \frac{(Q_h W_h m_{hx})\sum_{h=1}^{L}W_h(M_{hx}-m_{hx})}{\sum_{h=1}^{L}W_h Q_h m_{hx}^2}\right]^2\theta_h S_{hy}^2 + \bar{Y}^2(\lambda - 1)^2$$

Proof:

For the separate ratio-product estimator of population mean $\bar{Y}$

$$\bar{y}^{RP}_{stm} = \sum_{h=1}^{L} W_h\bar{y}_h\lambda$$

Let a calibration estimator be of the form

$$\bar{y}^{*\,RP}_{stm} = \sum_{h=1}^{L} W_h^*\bar{y}_h\lambda \quad (35)$$

Where $\lambda = \left\{\alpha_h\frac{M_{hx}}{m_h} + (1-\alpha_h)\frac{m_h}{M_{hx}}\right\}$ and W_h^* is the new weight and chosen such that a chi-square type loss function

$$\sum_{h=1}^{L}\left(\frac{W_h^* - W_h}{Q_h W_h}\right)^2 \quad (36)$$

is minimized subject to a calibration constraint

$$\sum_{h=1}^{L} W_h^{*} m_{hx} = \sum_{h=1}^{L} W_h M_{hx} \quad (37)$$

The Lagrange's function, using Calibration constraints and chi-square distribution measure is

$$\Delta = \sum_{h=1}^{L} \left(\frac{W_h^{*} - W_h}{Q_h W_h}\right)^2 - 2\lambda_1\left(\sum_{h=1}^{L} W_h^{*} m_{hx} - \sum_{h=1}^{L} W_h M_{hx}\right) \quad (38)$$

Differentiating (38) with respect to W_h^{*}, setting result equal to zero and solving for W_h^{*}

$$W_h^{*} = W_h + \lambda_1 m_{hx} Q_h W_h \quad (39)$$

Putting equation (39) into (37) and solving for λ_1 gives

$$\lambda_1 = \frac{\sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2} \quad (40)$$

Substituting for λ_1 from (40) in (39) gives the calibration weight for stratified sampling:

$$W_h^{*} = W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2} \quad (41)$$

Substituting equation (41) into (35) gives the required calibration estimator of population mean as

$$\bar{y}_{stm}^{*\,RP} = \sum_{h=1}^{L} W_h \bar{y}_h \lambda + \frac{\sum_{h=1}^{L} W_h^2 \bar{y}_h \lambda (Q_h W_h m_{hx})(M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2} \quad (42)$$

By putting the turning parameter $Q_h = \frac{1}{M_{hx}}$

$$\bar{y}_{stm}^{*\,RP} = \sum_{h=1}^{L} W_h \bar{y}_h \lambda + \frac{\sum_{h=1}^{L} W_h^2 \bar{y}_h \lambda (M_{hx}^{-1} W_h m_{hx})(M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h M_{hx}^{-1} m_{hx}^2} \quad (43)$$

2.5 The variance of the proposed calibration separate ratio-product estimator

The variance of the calibration separate ratio-product estimator $\bar{y}_{stm}^{*RP}$ for population mean $\bar{Y}$ using median as auxiliary variable is defined as

$$V(\bar{y}_{stm}^{*\,RP}) = \lambda^2 \sum_{h=1}^{L} \left[W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2}\right]^2 \theta_h S_{hy}^2 + \bar{Y}^2 (\lambda - 1)^2$$

Proof:

$$V(\bar{y}_{stm}^{*RP}) = (\bar{y}_{stm}^{*\,RP} - \bar{Y})^2$$

$$\bar{y}_{stm}^{*\,RP} - \bar{Y} = \sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda - \bar{Y} \quad (44)$$

Squaring equation (44) and taking expectation gives

$$E(\bar{y}_{stm}^{*\,RP} - \bar{Y})^2 = E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda - \bar{Y}\right)^2$$

$$E(\bar{y}_{stm}^{*\,RP} - \bar{Y})^2 = E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda\right)^2 - 2\lambda \bar{Y} E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h\right) + \bar{Y}^2$$

$$= var\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda\right) + \left[E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda\right)\right]^2 - 2\lambda \bar{Y} E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h\right) + \bar{Y}^2$$

Where $E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda\right)^2 = var\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda\right) + \left[E\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h \lambda\right)\right]^2$

$$= \lambda^2 var\left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h\right) + \lambda^2 \left(\sum_{h=1}^{L} W_h^{*} \bar{y}_h\right)^2 - 2\lambda \bar{Y}\left(\sum_{h=1}^{L} W_h^{*} \bar{Y}_h\right) + \bar{Y}^2$$

$$= \lambda^2 \sum_{h=1}^{L} W_h^{*2} \, var(\bar{y}_h) + \lambda^2 \bar{Y}^2 - 2\lambda \bar{Y}^2 + \bar{Y}^2$$

$$= \lambda^2 \sum_{h=1}^{L} W_h^{*2} \, var(\bar{y}_h) + \bar{Y}^2 (\lambda^2 - 2\lambda + 1)$$

$$E(\bar{y}_{stm}^{*RP} - \bar{Y})^2 = \lambda^2 \sum_{h=1}^{L} W_h^{*2} var(\bar{y}_h) + \bar{Y}^2 (\lambda - 1)^2$$

$$V(\bar{y}_{stm}^{*RP}) = \lambda^2 \sum_{h=1}^{L} W_h^{*2} \theta_h S_{hy}^2 + \bar{Y}^2 (\lambda - 1)^2 \quad (45)$$

Where $\theta_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$

Substituting (41) into (45) gives

$$V(\bar{y}_{stm}^{*RP}) = \lambda^2 \sum_{h=1}^{L} \left[W_h + \frac{(Q_h W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h Q_h m_{hx}^2}\right]^2 \theta_h S_{hy}^2 + \bar{Y}^2 (\lambda - 1)^2 \quad (46)$$

Setting the turning parameter $Q_h = S_{hx}^{-2}$

$$V(\bar{y}_{stm}^{*RP}) = \lambda^2 \sum_{h=1}^{L} \left[W_h + \frac{(S_{hx}^{-2} W_h m_{hx}) \sum_{h=1}^{L} W_h (M_{hx} - m_{hx})}{\sum_{h=1}^{L} W_h S_{hx}^{-2} m_{hx}^2}\right]^2 \theta_h S_{hy}^2 + \bar{Y}^2 (\lambda - 1)^2 \quad (47)$$

2.6 Empirical Results

In this section result of empirical evaluation of the proposed calibration estimators are done using simulated data set with underlying distributional assumption of Student-T, Cauchy, Lognormal, and standard normal. The result of the simulation study for percent average relative efficiency, percent average coefficient of variation, and percent average absolute relative bias of the existing stratified ratio estimator $\bar{y}_{st}$, existing calibration ratio estimator $\bar{y}_{st}^{*}$, existing calibration separate ratio-product estimator $\bar{y}_{st}^{*RP}$, proposed calibration ratio estimation $\bar{y}_{stm}^{*}$, proposed calibration separate ratio-product estimator $\bar{y}_{stm}^{*RP}$ for different underlying distributions, and sample sizes are presented in Table 1, 2 and 3.

3 Discussion of Finding

From the result of percent average relative efficiency in Table 1, it is observed that when the underlying distribution is student-t in nature, the efficiency performance of the separate ratio-product estimator $\bar{y}_{st}^{*RP}$ is better than other competing estimators at a sample size of 10%. As sample size is increased to 20%, the existing stratified ratio estimator $\bar{y}_{st}$ is more efficient than the other estimators under study. As the sample size is further increased to 25%, the efficiency performance of the existing calibration separate ratio-product estimator is better than those of the other competing estimators.

Table 1: *Percent Average Relative Efficiency for T-distribution, Cauchy distribution, Lognormal distribution, Standard normal distribution*

Distribution	Sample size	$\hat{\bar{y}}_{st}$	$\hat{\bar{y}}_{st}^{*}$	$\hat{\bar{y}}_{st}^{*RP}$	$\hat{\bar{y}}_{stm}^{*}$	$\hat{\bar{y}}_{stm}^{*RP}$
Student t	10%	100	1.59	78.66	53.30	103.04
	20%	100	4.97	86.23	30.03	14.49
	25%	100	10.70	244.39	74.18	74.18
Cauchy	10%	100	151.233	0	4440.25	0.04
	20%	100	107.15	0	639.21	0.19
	25%	100	99.24	0.03	988.37	0.90
Lognormal	10%	100	100	0	5187397.31	0
	20%	100	100	0	2004581.91	0
	25%	100	100	0	1694912.17	0
Standard normal	10%	100	1.23	95.47	208.081	70.61
	20%	100	5.52	232.25	115.15	90.24
	25%	100	63.54	1591.89	774.80	438.55

For the Cauchy distribution, it is observed that efficiency performance of the proposed calibration ratio estimator $\bar{y}_{st}^{*}$ is better (for all sample sizes considered) than those of the other estimators compared. However, there is no defined trend as the sample size increases. The result also shows a very poor efficiency performance for the proposed calibration separate ratio-product estimator $\bar{y}_{stm}^{*RP}$ under the Cauchy distribution. When the underlying distribution is lognormal in nature, it is again observed that the proposed calibration ratio estimator $\bar{y}_{stm}^{*}$ using the median of the auxiliary variable is highly efficient when compared to existing estimators $\bar{y}_{st}$ and $\bar{y}_{st}^{*}$ that uses mean of auxiliary variable.

For the standard normal distribution, there is a notable gain in efficiency for the proposed calibration ratio estimator $\bar{y}_{stm}^{*}$ across all sample sizes. There is still no observable trend as the sample size increases. However, the proposed calibration separate ratio-product estimator $\bar{y}_{stm}^{*RP}$ show a considerable gain in efficiency at 25%.

The result of the average coefficient of variation in Table 2 shows that when the underlying distribution is student t-distribution, the calibration separate ratio-product estimator $\bar{y}_{stm}^{*RP}$ is a more reliable estimator compared to the other estimators understudy at sample size of 10%. As sample size is increased to 20%, the existing stratified ratio estimator $\bar{y}_{st}$ becomes a more reliable estimator when compared to the other estimators. At sample size of 25%, the existing separate ratio-product estimator $\bar{y}_{st}^{*RP}$ is a more reliable estimator than the competing estimators.

Table 2: *Average coefficient of variation for T-distribution, Cauchy distribution, Lognormal distribution, Standard normal distribution*

Distribution	Sample size	$\hat{\bar{y}}_{st}$	$\hat{\bar{y}}^{*}_{st}$	$\hat{\bar{y}}^{*RP}_{st}$	$\hat{\bar{y}}^{*}_{stm}$	$\hat{\bar{y}}^{*RP}_{stm}$
Student-t	10%	917.98	57618.69	1167.02	1722.29	890.90
	20%	330.50	6655.77	383.29	1100.39	2280.58
	25%	963.81	28487.24	429.51	1299.29	1299.32
Cauchy	10%	3060	2383.09	1.36e+09	256.68	94436
	20%	786012612	7332	102164238	-12.299	-41589
	25%	11623	11709.09	42321142	1175.65	12865
Lognormal	10%	89577234	89577234	2.0e+117	8.25e+22	6.03e+93
	20%	10862134	10862134	1.79e+50	543	2.13e+38
	25%	8012143	8012143	7.3e+49	475	1.27e+38
Standard normal	10%	1.46	119.51	1.53	0.70	2.07
	20%	1.81	32.75	0.78	1.56	2.00
	25%	9.53	15.01	0.60	1.23	2.17

For the Cauchy distribution, it is observed that the proposed calibration ratio estimator $\bar{y}^{*}_{stm}$ is a more reliable estimator of the estimators under consideration at sample size 10%, 20% and 25%. However, there is no defined trend as the sample size increases. The result also reveals that the estimate obtain from the proposed calibration separate ratio-product estimator is highly unreliable. When the underlying distribution is lognormal in nature, it is observed that the proposed calibration ratio estimator $\bar{y}^{*}_{stm}$ is more reliable than the other estimators under consideration. It is also notable that the existing stratified ratio estimator $\bar{y}_{st}$ and the existing calibration ratio-product estimators are the same at sample size 10%, 20% and 25%.
For the standard normal distribution, it is observed that the proposed calibration ratio estimator $\bar{y}^{*}_{stm}$ is more reliable than the other estimators under consideration at sample size of 10%. At a sample size of 20% and 25%, the existing separate ratio-product estimators is observed to be more reliable than the other estimators under study.
From the result of the simulation study in Table 3, when the underlying distribution is student-t in nature, it is observed that the proposed calibration separate ratio-product estimator $\bar{y}^{*RP}_{stm}$ has minimum bias compared to the other estimators under consideration at sample size of 10%. As the sample size is increased to 20%, the existing stratified ratio estimator $\bar{y}_{st}$ show a minimum biasness. At 25% sample size, the existing calibration separate ratio estimator $\bar{y}^{*RP}_{st}$ show minimum biasness compared to the other estimators under consideration.

Table 3: *Percentage Average Absolute Relative Bias for T-distribution, Cauchy distribution, Lognormal distribution, Standard normal distribution*

Distribution	Sample size	$\hat{\bar{y}}_{st}$	$\hat{\bar{y}}^{*}_{st}$	$\hat{\bar{y}}^{*RP}_{st}$	$\hat{\bar{y}}^{*}_{stm}$	$\hat{\bar{y}}^{*RP}_{stm}$
Student-t	10%	9.179	576.19	11.67	17.22	8.90
	20%	3.305	66.56	3.83	11.00	22.81
	25%	9,64	284.87	4.29	12.99	12.99
Cauchy	10%	36.1	23.83	1.36e+07	2.57	94421
	20%	78.6	73.3	1.0e+06	12.30	41523
	25%	116	117.09	423421	11.76	12861
Lognormal	10%	8.9e+05	8.9e+05	2.89e+49	17.2	1.98e+37
	20%	1.08e+05	1.08e+05	1.79e+48	5.43	2.13e+36
	25%	8.01e+04	8.01e+04	7.3e+47	4.72	1.27e+36
Standard normal	10%	146.449	11950.8	153.399	70.38	207.41
	20%	180.75	3275.09	77.8	156.96	200.315
	25%	953.51	1500.74	59.90	123.07	217.43

For the Cauchy distribution, it is observed that the proposed calibration ratio estimator $\bar{y}^{*}_{stm}$ show minimum biasness at sample size of 10%, 20% and 25% compared to the other estimators under study.

However, there is no trend as the sample size increases. The result also indicates that the proposed calibration separate ratio-product estimator is highly bias when the underlying distribution is Cauchy.
When the underlying distribution is lognormal in nature, it is observed that the proposed calibration ratio estimator $\bar{y}^{*}_{stm}$ has minimum biasness compared to the other competing estimators at 10%, 20% and 25% sample size. The result also reveals that the proposed calibration separate ratio-product estimator $\bar{y}^{*RP}_{stm}$ is highly bias under the lognormal distribution.
For the standard normal distribution, it is observed that at a sample size of 10%, the proposed calibration ratio estimator $\bar{y}^{*}_{stm}$ has minimum biasness compared to the other estimators under study. As sample size increase to 20% and 25%, the existing calibration separate ratio-product estimator show minimum biasness than the other estimators under consideration.
In summary, under the skewed distributions (Cauchy distribution and Lognormal distribution) and the standard normal distribution, it is observed that the proposed calibration ratio estimator is a more precise and efficient estimator of the population mean than the competing estimators in this study. This estimator is found to be consistently better than the other estimators in efficiency and minimum bias as the sample size increases. This result agrees with [4], [10], [11] result which suggest the use of median of auxiliary variable as an alternative to the use of auxiliary mean to give a more efficient and less bias estimator of the population mean. It is also necessary to note that under the lognormal distribution, the conventional ratio estimator and the calibration ratio estimator give the same results.

4 Conclusion

Calibration estimation technique is a known method used to modify the design weights in other to improve sample survey estimates by minimizing a distance function subject to one or more constraints when external information related to the population otherwise known as the auxiliary variable is available. In this work, calibration ratio estimator and calibration separate ratio-product estimator has been proposed in the presence of auxiliary information using the median of the auxiliary variable, stratified sampling scheme and chi-square distance measure. A simulation study was conducted to evaluate the performance of the proposed estimators in terms of percent average relative efficiency, percent average coefficient of variation, and percent average absolute relative bias. The result of simulation study shows that under the stratified sampling, proposed calibration ratio type estimator $\hat{\bar{y}}^{*}_{stm}$ gives better estimate of population mean when the auxiliary variable is highly positively correlated with the study variable and the underlying distribution is Cauchy distribution, Lognormal distribution or Standard normal distribution. For lognormal distribution, it appears that the efficiency performance of the proposed calibration ratio estimator $\hat{\bar{y}}^{*}_{stm}$ increases as the sample size is increased. This suggest that the proposed calibration ratio estimator is the most efficient and list bias estimator of the population mean when the underlying distribution is Cauchy, Lognormal or Standard normal. However, this is not the case under the other distributions (Student-t, Cauchy and Standard normal). When the distribution is T-distribution, the efficiency performance of the proposed calibration separate ratio-product type estimator $\hat{\bar{y}}^{*RP}_{stm}$ is better than the other estimators when the sample size is 10%, but as the sample size increases, the Horvitz Thompson estimator $\hat{\bar{y}}_{st}$ and the calibration separate ratio-product estimator $\hat{\bar{y}}^{*RP}_{st}$ give better precision than the other estimators. The Horvitz Thompson type estimator $\hat{\bar{y}}_{st}$ and the Sarndal Calibration estimator $\hat{\bar{y}}^{*}_{stm}$ give equal precision when the underlying distribution is lognormal.

5 Declarations

5.1 Study Limitations:

The simulation studies in this work are limited to only standard normal distribution and skewed distributions.

5.2 Competing Interests

The authors declared that no conflict of interest exist in this publication.

5.3 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published institutional affiliations.

How to Cite this Article:

Will be updated in the final version.

References

[1] M. Noor-ul-Amin, A. Javaid, and M. Hanif, "Estimation of population mean in systematic random sampling using auxiliary information," *Journal of Statistics and Management Systems*, vol. 20, no. 6, pp. 1095–1106, Nov. 2018, doi: 10.1080/09720510.2017.1364500.

[2] H. P. Singh and P. Nigam, "A Two Parameter Ratio-Product-Ratio Estimator in Post Stratification," *Pakistan Journal of Statistics and Operation Research*, vol. 18, no. 1, pp. 273–296, Mar. 2022, doi: 10.18187/pjsor.v18i1.3507.

[3] T. Zaman and C. Kadilar, "Exponential ratio and product type estimators of the mean in stratified two-phase sampling," *AIMS MATHEMATICS*, vol. 6, no. 5, pp. 4265–4279, 2021, doi: 10.3934/MATH.2021252.

[4] J. Subramani, *A NEW MEDIAN BASED RATIO ESTIMATOR FOR ESTIMATION OF THE FINITE POPULATION MEAN*, vol. 17, no. 4. Główny Urząd Statystyczny, 2016, pp. 591–604.

[5] N. Garg and M. Pachori, "Use of Median in Calibration Estimation of the Finite Population Mean in Stratified Sampling," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 5, pp. 667–671, May 2019, doi: 10.26438/IJCSE/V7I5.667671.

[6] P. K. Rai, A. Singh, and M. Qasim, "Calibration-Based Estimators using Different Distance Measures under Two Auxiliary Variables: A Comparative Study," *Journal of Modern Applied Statistical Methods*, vol. 19, no. 1, p. 15, Jun. 2021, doi: 10.22237/jmasm/1619481600.

[7] S. Singh, S. Horn, S. Chowdhury, and F. Yu, "Theory & Methods: Calibration of the estimators of variance," *Aust N Z J Stat*, vol. 41, no. 2, pp. 199–212, Jun. 1999, doi: 10.1111/1467-842X.00074.

[8] G. K. Vishwakarma, H. P. Singh, G. K. Vishwakarma, and H. P. Singh, "Separate ratio-product estimator for estimating population mean using auxiliary information," *Journal of Statistical Theory and Applications*, vol. 10, no. 4, pp. 653–664, 2011.

[9] E. P. Clement, "A New Ratio Estimator of Mean in Survey Sampling by Calibration Estimation," *Elixir International Journal: Statistics*, vol. 106, pp. 46461–46465, 2017.

[10] J. Subramani and G. Kumarapandiyan, "Estimation of Population Mean Using Known Median and Co-Efficent of Skewness," *American Journal of Mathematics and Statistics*, vol. 2, no. 5, pp. 101–107, Dec. 2012, doi: 10.5923/J.AJMS.20120205.01.

[11] S. K. Yadav, S. Misra, A. K. Shukla, and V. Tiwari, "Improved Estimation of Population Mean Using Median and Coefficient of Variation of Auxiliary Variable," *Int. Journal of Engineering Research and Applications*, vol. 4, no. 1, pp. 206–211, 2014, Accessed: Jan. 19, 2023. [Online]. Available: https://www.ijera.com/papers/Vol4_issue1/Version%203/AJ4103206211.pdf