



Calibration Approach Product Type Estimators of Population Mean in Stratified Sampling with Single Constraint: A Comparison of Three Distance Measures

Enang, Ekaette Inyang¹, Ojua, Doris Nkan^{1*} and T. T. Ojewale¹

¹Department of Statistics, University of Calabar, Calabar- Cross River State, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author EEI designed the study and wrote the protocol. Author ODN performed the statistical analysis, wrote the first draft of the manuscript and managed the analyses of the study. Author TTO managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v15i230350

Editor(s):

(1) Dr. Dariusz Jacek Jakóbczak, Koszalin University of Technology, Poland.

Reviewers:

(1) Mansour Sattam Aldosari, King Abdulaziz University, Saudi Arabia.

(2) Simón Sandoval Rocha, Universidad de Concepción, Chile.

Complete Peer review History: <https://www.sdiarticle4.com/review-history/67505>

Received 25 February 2021

Accepted 30 April 2021

Published 21 October 2021

Original Research Article

Abstract

This study employed the method of calibration on product type estimator to propose calibration product type estimators using three distance measures namely; chi-square distance measure, the minimum entropy distance measure and the modified chi-square distance measure for single constraint. The estimators of variances of the proposed estimators were also obtained. An empirical study to ascertain the performance of these estimators was carried out using real life and stimulated data set. The result with the real life data showed that the proposed calibration product type estimator \bar{y}_{pcp1} produced better estimates of the population mean \bar{Y} compared to \bar{y}_{pcp2} and \bar{y}_{pcp3} . Results from the simulation study showed that the proposed calibration product type estimators had a high gain in efficiency as compared to the product type estimator. The simulation result also showed that the proposed estimators were more consistent and reliable under the Gamma and Exponential distributions with the exponential distribution taking the lead. The conventional product type estimator however was found to be better if the underlying distributional assumption is normal in nature.

*Corresponding author: Email: ujuadoris@yahoo.com, dorisojua@gmail.com;

Keywords: Finite population; auxiliary variable; stratified sampling; calibration estimators; population mean.

1 Introduction

Over the years information on auxiliary variable has been used to improve on the precision of an estimate of the population mean or total, as in ratio, product, regression and difference method of estimation. In each case the advance knowledge of the population mean, \bar{X} of the auxiliary variate x is required. The ratio estimator is used when the variables are positively correlated, while the product type estimator is preferred when the variates are negatively correlated. Robson [1], Murthy [2] and Perri [3] had established that both the ratio and product type estimator are good estimators of the population parameters if the regression line is a straight line and passes through the origin. However in many practical situations the regression line does not pass through the origin and in such situations the ratio and product estimators do not perform as well as the regression estimator [4].

It is an established fact that in stratified sampling design, the use of auxiliary variable increases the precision of estimates of a population characteristic. In stratified sampling design, the population under investigation is divided into different strata so as to obtain the homogeneity within each stratum, and sample observations are drawn within each stratum generally by the procedure of simple random sampling. Members of the sample are assigned a sampling weight, which represents the fraction of the population that is accounted for by the sample members.

Calibration technique which was introduced by Deville and Sarndal [5] seeks to adjust sampling weights in stratified sampling design with the aim of improving the precision of the estimates of the population parameters. To adjust the sample weights, information on auxiliary variable is used based on existing data, or other large surveys, it is often possible to know the population total, mean or proportion for other variables measured in the survey, as well as the values recorded for the members of the sample.

Mathematically, the problem of calibration can be defined informally as follows. Suppose there are some initial weights assigned to n objects of a survey. Suppose further that there are m auxiliary variables and that for these auxiliary variables the sample values are known, either exactly or approximately. The calibration problem seeks to improve on the initial weights by finding new weights that incorporate the auxiliary information. Several authors including Singh [6], Singh [7], Singh [8], Farrell and Singh [9], Farrell and Singh [10], Wu and Sitter [11], Estevao and Särndal [12], Kott [13], Montanari and Ronalli [14], Clement and Enang [15] amongst others considered the Deville and Särndal [5] method and derived important calibration estimators.

In sampling literature, many calibration estimators have been proposed using auxiliary information, but this work seeks to extend the calibration technique to adjust the weight of the conventional product type estimator in stratified sampling using one auxiliary variable under one constraint.

1.1 Definition of terms

\bar{X}_h is the population mean of the auxiliary variable
 \bar{x}_h is the sample mean of the auxiliary variable
 \bar{Y}_h is the population mean of the variable of interest
 \bar{y}_h is the sample mean of the variable of interest
 S_{hy}^2 is the population variance of the variable of interest
 s_{hy}^2 is the sample variance of the variable of interest
 S_{hx}^2 is the population variance of the auxiliary variable
 s_{hx}^2 is the sample variance of the auxiliary variable
 S_{hxy} is the covariance between the auxiliary variable and variable of interest
 ρ_{xy} is the correlation between the variable of interest and the auxiliary variable
 N is the population size
 n is the sample size
 N_h is the stratum population size
 n_h is the stratum sample size
 Q_h is a positive constant

$MSE(\hat{y}_p)$ is the estimated mean square error of the conventional product type estimator
 $MSE(\hat{y}_{pcp})$ is the mean square error of the proposed estimators

1.2 Percentage average relative efficiency (% \overline{RE})

The relative efficiency of two procedures is given by the ratio of their efficiencies and is often defined using variance or mean square error. This shall be used to measure the average efficiency of each proposed estimator. It can be computed as:

$$\% \overline{RE}(\hat{y}_{pcp}) = \left\{ \sqrt{\frac{MSE(\hat{y}_p)}{MSE(\hat{y}_{pcp})}} \right\} \times 100 \tag{1}$$

Where

$$\overline{MSE}(\hat{y}_{pcp}) = \frac{1}{h} \sum_{h=1}^H MSE(\hat{y}_{pcp}) \tag{2}$$

It should be noted that a $\% \overline{RE}(\hat{y}_{pcp})$ of value greater than 100 predicts a relative increase in efficiency of the proposed estimator, while a $\% \overline{RE}(\hat{y}_{pcp})$ of value less than 100 indicates a loss in efficiency of the proposed estimator.

1.3 Percentage average absolute relative bias $\%(\overline{ARB})$

If \hat{y}_{pcp} , then, for each stratum $h = 1, 2, \dots, L$, the relative bias is given by:

$$RB(\hat{y}_{pcp}) = \frac{1}{R} \sum_{r=1}^R \left(\frac{\hat{y}_{pcp}}{\bar{y}_p} - 1 \right) \tag{3}$$

and the percentage average absolute relative bias $\%(\overline{ARB})$ is computed as

$$\% \overline{ARB}(\hat{y}_{pcp}) = \left\{ \frac{1}{L} \sum_{h=1}^L ARB(\hat{y}_{pcp}) \right\} \times 100 \tag{4}$$

where

$$ARB(\hat{y}_{pcp}) = \left| \frac{1}{R} \sum_{r=1}^R \left(\frac{\hat{y}_{pcp}}{\bar{y}_p} - 1 \right) \right| \tag{5}$$

and R is the number of runs

1.4 Average coefficient of variation (\overline{CV})

This measure shall be used to measure the reliability of the proposed estimators compared to the conventional product type estimator in stratified sampling. The percentage average coefficient of variation of \hat{y}_{pcp} is given as:

$$\% \overline{CV}(\hat{y}_{pcp}) = \left\{ \frac{1}{L} \sum_{h=1}^L CV(\hat{y}_{pcp}) \right\} \times 100 \tag{6}$$

Where

$$CV(\hat{y}_{pcp}) = \frac{\sqrt{MSE(\hat{y}_{pcp})}}{\bar{y}_p}$$

The interpretation is that, high values of $\%CV(\hat{y}_{pcp})$ indicates unreliable estimates while low value predicts reliable estimates.

2 Proposed Estimators

Theorem 2.1: Given the product type estimator

$$\bar{y}_{ps} = \sum_{h=1}^L W_h \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h}$$

a calibration product type estimator \bar{y}_{pcp1} for population mean \bar{Y} given as

$$\bar{y}_{pcp1} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{X}_h} + \frac{\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \bar{y}_h}{\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \bar{X}_h} \left(\bar{X} - \sum_{h=1}^L W_h \bar{x}_h \right)$$

can be obtained by

$$MinD = \sum_{h=1}^L \frac{(\gamma_{h1} - W_h)^2}{W_h Q_h}$$

s.t.

$$\sum_{h=1}^L \gamma_{h1} \bar{x}_h = \bar{X}$$

Proof:

Given

$$\bar{y}_{ps} = \sum_{h=1}^L W_h \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h}$$

A calibration product type estimator

$$\bar{y}_{pcp1} = \sum_{h=1}^L \gamma_{h1} \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h} \tag{7}$$

where the weight γ_{h1} are chosen such that the distance measure

$$\sum_{h=1}^L \frac{(\gamma_{h1} - W_h)^2}{W_h Q_h} \tag{8}$$

is minimized subject to the constraint

$$\sum_{h=1}^L \gamma_{h1} \bar{x}_h = \bar{X} \tag{9}$$

Combing (8) and (9) gives the optimization function

$$\varphi(\gamma_{h1}, \lambda_1) = \sum_{h=1}^L \frac{(\gamma_{h1} - W_h)^2}{W_h Q_h} - 2\lambda_1 (\sum_{h=1}^L \gamma_{h1} \bar{x}_h - \bar{X}) \tag{10}$$

where λ_1 is a Lagrange multiplier.

Differentiating equation (10) partially with respect to γ_{h1} and λ_1 , and equating to zero gives

$$\gamma_{h1} = W_h + \lambda_1 Q_h \bar{x}_h \tag{11}$$

and

$$\lambda_1 = \frac{\bar{X} - \sum_{h=1}^L W_h \bar{x}_h}{\sum_{h=1}^L W_h Q_h \bar{x}_h^2} \tag{12}$$

substituting (12) into (11) yields

$$\gamma_{h1} = W_h + \frac{W_h Q_h \bar{x}_h}{\sum_{h=1}^L W_h Q_h \bar{x}_h^2} (\bar{X} - \sum_{h=1}^L W_h \bar{x}_h) \tag{13}$$

substituting equation (13) into equation (7) we obtain

$$\bar{y}_{pcp1} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{x}_h} + \frac{\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \bar{y}_h}{\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \bar{x}_h} (\bar{X} - \sum_{h=1}^L W_h \bar{x}_h) \tag{14}$$

which is the proposed calibration product type estimator for population mean \bar{Y} in stratified random sampling as required to prove. This estimator is in form of a regression estimator with $\sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{x}_h}$ as the intercept and $\frac{\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \bar{y}_h}{\sum_{h=1}^L W_h Q_h \bar{x}_h^2 \bar{x}_h}$ as slope.

By letting $Q_h = 1$ then equation (14) becomes

$$\bar{y}_{pcp11} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{x}_h} + \frac{\sum_{h=1}^L W_h \bar{x}_h^2 \bar{y}_h}{\sum_{h=1}^L W_h \bar{x}_h^2 \bar{x}_h} (\bar{X} - \sum_{h=1}^L W_h \bar{x}_h) \tag{15}$$

Which is the proposed regression calibration product type estimator for population mean \bar{Y} in stratified sampling.

Also by letting $Q_h = \frac{1}{\bar{x}_h}$, then equation (14) becomes

$$\bar{y}_{pcp12} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{X}_h} + \frac{\sum_{h=1}^L W_h \bar{x}_h \bar{y}_h}{\sum_{h=1}^L W_h \bar{x}_h \bar{X}_h} (\bar{X} - \sum_{h=1}^L W_h \bar{x}_h) \tag{16}$$

Which is the proposed ratio calibration product type estimator for population mean \bar{Y} in stratified sampling.

Theorem 2.2: Given the product type estimator, a calibration product type estimator \bar{y}_{pcp2} for population mean \bar{Y} given as

$$\bar{y}_{pcp2} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{X}_h} \exp\left(\sum_{h=1}^L \ln\left[\frac{\bar{X}}{W_h \bar{x}_h}\right]\right) = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{X}_h} \prod_{h=1}^L \left(\frac{\bar{X}}{W_h \bar{x}_h}\right)$$

can be obtained by

$$MinD = \sum_{h=1}^L \frac{1}{Q_h} \left\{ \gamma_{h2} \log\left(\frac{\gamma_{h2}}{W_h}\right) - \gamma_{h2} - W_h \right\}$$

s.t.

$$\sum_{h=1}^L \gamma_{h2} \bar{x}_h = \bar{X}$$

Proof:

Given

$$\bar{y}_{ps} = \sum_{h=1}^L W_h \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h}$$

We define a calibration estimator

$$\bar{y}_{pcp2} = \sum_{h=1}^L \gamma_{h2} \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h} \tag{17}$$

where the weights γ_{h2} are chosen such that the distance measure

$$\sum_{h=1}^L \frac{1}{Q_h} \left\{ \gamma_{h2} \log\left(\frac{\gamma_{h2}}{W_h}\right) - \gamma_{h2} - W_h \right\} \tag{18}$$

is minimized subject to the constraint

$$\sum_{h=1}^L \gamma_{h2} \bar{x}_h = \bar{X}$$

by combing (18) and this constraint gives

$$\varphi(\gamma_{h2}, \lambda_2) = \sum_{h=1}^L \frac{1}{Q_h} \left\{ \gamma_{h2} \log\left(\frac{\gamma_{h2}}{W_h}\right) - \gamma_{h2} - W_h \right\} - \lambda_2 (\sum_{h=1}^L \gamma_{h2} \bar{x}_h - \bar{X}) \tag{19}$$

Differentiating equation (19) partially with respect to γ_{h2} and λ_2 , and equating to zero gives

$$\gamma_{h2} = W_h \exp[\lambda_2 Q_h \bar{x}_h] \tag{20}$$

and

$$\lambda_2 = \frac{1}{\sum_{h=1}^L Q_h \bar{x}_h} \sum_{h=1}^L \ln \left(\frac{\bar{X}}{W_h \bar{x}_h} \right) \tag{21}$$

substituting (21) into (20) gives

$$\gamma_{h2} = W_h \exp \left[\frac{1}{\sum_{h=1}^L Q_h \bar{x}_h} \sum_{h=1}^L \ln \left(\frac{\bar{X}}{W_h \bar{x}_h} \right) Q_h \bar{x}_h \right] = W_h \prod_{h=1}^L \left(\frac{\bar{X}}{W_h \bar{x}_h} \right)^{Q_h \bar{x}_h / \sum_{h=1}^L Q_h \bar{x}_h} \tag{22}$$

substituting equation (22) into equation (17) we obtain

$$\bar{y}_{pcp2} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{x}_h} \exp \left(\sum_{h=1}^L \ln \left[\frac{\bar{X}}{W_h \bar{x}_h} \right] \right) = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{x}_h} \prod_{h=1}^L \left(\frac{\bar{X}}{W_h \bar{x}_h} \right)$$

Which is the proposed calibration product type estimator for population mean \bar{Y} in stratified random sampling.

In this case, after substitution we observed that there is no need for a tuning parameter.

Theorem 2.3: Calibration product type estimator \bar{y}_{pcp3} for population mean \bar{Y} can be obtained from the product type estimator by

$$MinD = \sum_{h=1}^L \frac{(\gamma_{h3} - W_h)^2}{\gamma_{h3} Q_h}$$

s.t.

$$\sum_{h=1}^L \gamma_{h3} \bar{x}_h = \bar{X}$$

given as

$$\bar{y}_{pcp3} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{x}_h} \left(L + \frac{(\sum_{h=1}^L W_h^2 \bar{x}_h^2 - \bar{X}^2)}{\bar{X}^2} \right)^{-\frac{1}{2}}$$

Proof:

Given the product estimator

$$\bar{y}_{ps} = \sum_{h=1}^L W_h \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h}$$

an estimator

$$\bar{y}_{pcp3} = \sum_{h=1}^L \gamma_{h3} \frac{\bar{x}_h \bar{y}_h}{\bar{X}_h} \tag{23}$$

where the weights γ_{h3} are chosen such that the distance measure

$$\sum_{h=1}^L \frac{(\gamma_{h3} - W_h)^2}{\gamma_{h3} Q_h} \tag{24}$$

is minimized subject to the constraint

$$\sum_{h=1}^L \gamma_{h3} \bar{x}_h = \bar{X}$$

combining (24) and the constraint gives

$$\varphi(\gamma_{h3}, \lambda_3) = \sum_{h=1}^L \frac{(\gamma_{h3} - W_h)^2}{\gamma_{h3} Q_h} - 2\lambda_3 (\sum_{h=1}^L \gamma_{h3} \bar{x}_h - \bar{X}) \tag{25}$$

Differentiating equation (25) partially with respect to γ_{h3} and λ_3 , and equating to zero gives

$$\gamma_{h3} = \frac{W_h}{[1 - 2\lambda_3 Q_h \bar{x}_h]^{\frac{1}{2}}} \tag{26}$$

and

$$\lambda_3 = \frac{\bar{X}^2 - \sum_{h=1}^L W_h^2 \bar{x}_h^2}{2\bar{X}^2 \sum_{h=1}^L Q_h \bar{x}_h} \tag{27}$$

substituting (27) into (26) gives

$$\gamma_{h3} = W_h \left[1 + \frac{(Q_h \bar{x}_h)(\sum_{h=1}^L W_h^2 \bar{x}_h^2 - \bar{X}^2)}{\bar{X}^2 \sum_{h=1}^L Q_h \bar{x}_h} \right]^{\frac{1}{2}} \tag{28}$$

substituting equation (28) into equation (23) yields

$$\bar{y}_{pcp3} = \sum_{h=1}^L \frac{W_h \bar{x}_h \bar{y}_h}{\bar{X}_h} \left(L + \frac{(\sum_{h=1}^L W_h^2 \bar{x}_h^2 - \bar{X}^2)}{\bar{X}^2} \right)^{\frac{1}{2}} \tag{29}$$

Which is the proposed calibration product type estimator for population mean \bar{Y} in stratified random sampling. Also in this case, after substitution of the adjusted weight into the calibration equation there is no need of a tuning parameter.

3 Variance Estimators of the Proposed Estimators

Theorem 3.1: Given the variance estimator of the product type estimator of population mean in stratified sampling as

$$\hat{V}(\bar{y}_p) = \sum_{h=1}^L W_h^2 \gamma (s_{hy}^2 + R^2 s_{hx}^2 + 2R s_{hxy}) \tag{30}$$

where

$$\gamma = \frac{1 - f_h}{n_h}$$

$$s_{hxy} = \rho_{xy} s_{hx} s_{hy}$$

and

$$R = \frac{\bar{Y}}{\bar{X}}$$

a variance estimator of the calibrated product type estimator $\hat{V}(\bar{y}_{pcp1})$ for population mean \bar{Y} given as

$$\hat{V}(\bar{y}_{pcp1}) = \sum_{h=1}^L \frac{D_h \gamma_{h1}^2}{W_h^2} s_p + \frac{\left(\sum_{h=1}^L \frac{D_h Q_h \gamma_{h1}^2 s_p s_{hx}^2}{W_h^2} \right)}{\sum_{h=1}^L D_h Q_h S_{hx}^4} (V(\bar{x}_{st}) - \hat{v}(\bar{x}_{st}))$$

can be obtained by

$$MinD = \sum_{h=1}^L \frac{(\omega_{h1}^0 - D_h)^2}{D_h Q_h}$$

s.t.

$$\sum_{h=1}^L \omega_{h1}^0 S_{hx}^2 = V(\bar{x}_{st})$$

Proof: Let (30) be rewritten in the form of

$$\hat{V}(\bar{y}_p) = \sum_{h=1}^L \frac{D_h \gamma_{h1}^2}{W_h^2} s_p \tag{31}$$

where

$$D_h = \frac{W_h^2 (1 - f_h)}{n_h}$$

and

$$s_p = (s_{hy}^2 + R^2 s_{hx}^2 + 2R s_{hxy})$$

with

$$s_{hxy} = \rho_{hxy} s_{hx} s_{hy}$$

and γ_{h1} is the calibrated weights

Now consider a calibration variance as estimator of the form

$$\hat{V}(\bar{y}_{pcp1}) = \sum_{h=1}^L \frac{\omega_{h1}^o \gamma_{h1}^2}{W_h^2} S_p \tag{32}$$

where the weights ω_{h1}^o , are chosen such that the distance measure

$$\sum_{h=1}^L \frac{(\omega_{h1}^o - D_h)^2}{D_h Q_h} \tag{33}$$

is minimized subject to the constraint

$$\sum_{h=1}^L \omega_{h1}^o S_{hx}^2 = V(\bar{x}_{st}) \tag{34}$$

Combing (33) and (34) gives the optimization function

$$\varphi(\omega_{h1}^o, \lambda_{11}) = \sum_{h=1}^L \frac{(\omega_{h1}^o - D_h)^2}{D_h Q_h} - 2\lambda_{11} \left(\sum_{h=1}^L \omega_{h1}^o S_{hx}^2 - V(\bar{x}_{st}) \right) \tag{35}$$

Differentiating equation (35) partially with respect to ω_{h1}^o and λ_{11} , and equating to zero gives

$$\omega_{h1}^o = D_h [1 + \lambda_{11} Q_h S_{hx}^2] \tag{36}$$

and

$$\lambda_{11} = \frac{(V(\bar{x}_{st}) - \hat{v}(\bar{x}_{st}))}{\sum_{h=1}^L D_h Q_h S_{hx}^4} \tag{37}$$

where

$$\hat{v}(\bar{x}_{st}) = \sum_{h=1}^L D_h S_{hx}^2$$

substituting (37) into (36) gives

$$\omega_{h1}^o = D_h + \frac{D_h Q_h S_{hx}^2}{\sum_{h=1}^L D_h Q_h S_{hx}^4} (V(\bar{x}_{st}) - \hat{v}(\bar{x}_{st})) \tag{38}$$

substituting (38) into (31) yields

$$\hat{V}(\bar{y}_{pcp1}) = \sum_{h=1}^L \frac{D_h \gamma_{h1}^2}{W_h^2} S_p + \frac{\left(\sum_{h=1}^L \frac{D_h Q_h \gamma_{h1}^2 S_p S_{hx}^2}{W_h^2} \right)}{\sum_{h=1}^L D_h Q_h S_{hx}^4} (V(\bar{x}_{st}) - \hat{v}(\bar{x}_{st})) \tag{39}$$

which is the proposed calibration product type variance estimator for population mean \bar{Y} in stratified random sampling.

Substituting $Q_h = 1$ and $Q_h = \frac{1}{\bar{x}_h}$ in (39) gives

$$\hat{V}(\bar{y}_{pcp11}) = \sum_{h=1}^L \frac{D_h \gamma_{h11}^2}{W_h^2} S_p + \frac{\left(\sum_{h=1}^L \frac{D_h \gamma_{h11}^2 S_p S_{hx}^2}{W_h^2} \right)}{\sum_{h=1}^L D_h S_{hx}^4} (V(\bar{x}_{st}) - \hat{v}(\bar{x}_{st})) \tag{40}$$

and

$$\hat{V}(\bar{y}_{pcp12}) = \sum_{h=1}^L \frac{D_h \gamma_{h12}^2}{W_h^2} S_p + \frac{\left(\sum_{h=1}^L \frac{D_h \frac{1}{\bar{x}_h} S_p \gamma_{h12}^2 S_{hx}^2}{W_h^2} \right)}{\sum_{h=1}^L D_h \frac{1}{\bar{x}_h} S_{hx}^4} (V(\bar{x}_{st}) - \hat{v}(\bar{x}_{st})) \tag{41}$$

This is the regression and ratio type calibration product variance estimator for population mean for stratified sampling respectively.

Theorem 3.2: Given the product type variance estimator in (30), its weight can be adjusted by

$$MinD = \sum_{h=1}^L \frac{1}{Q_h} \left\{ \omega_{h2}^o \log \left(\frac{\omega_{h2}^o}{D_h} \right) - \omega_{h2}^o - D_h \right\}$$

s.t.

$$\sum_{h=1}^L \omega_{h2}^o S_{hx}^2 = V(\bar{x}_{st})$$

to obtain the calibration product type variance estimator $\hat{V}(\bar{y}_{pcp2})$ for population mean \bar{Y} given as

$$\hat{V}(\bar{y}_{pcp2}) = \sum_{h=1}^L \frac{D_h \gamma_{h2}^2}{W_h^2} S_p \exp \left(\sum_{h=1}^L \ln \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right) \right) = \sum_{h=1}^L \frac{D_h \gamma_{h2}^2}{W_h^2} S_p \prod_{h=1}^L \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right)$$

Proof: Rewriting the estimator (30) as

$$\hat{V}(\bar{y}_{pcp2}) = \sum_{h=1}^L \frac{\omega_{h2}^o \gamma_{h2}^2}{W_h^2} S_p \tag{42}$$

where the weights ω_{h2}^o , are chosen such that the distance measure

$$\sum_{h=1}^L \frac{1}{Q_h} \left\{ \omega_{h2}^o \log \left(\frac{\omega_{h2}^o}{D_h} \right) - \omega_{h2}^o - D_h \right\} \tag{43}$$

is minimized subject to the constraint

$$\sum_{h=1}^L \omega_{h2}^o S_{hx}^2 = V(\bar{x}_{st})$$

Then by combining (43) and the constraint gives the optimization function

$$\varphi(\gamma_{h2}, \lambda_{22}) = \sum_{h=1}^L \frac{1}{Q_h} \left\{ \omega_{h2}^o \log \left(\frac{\omega_{h2}^o}{D_h} \right) - \omega_{h2}^o - D_h \right\} - \lambda_{22} (\sum_{h=1}^L \omega_{h2}^o S_{hx}^2 - V(\bar{x}_{st})) \tag{44}$$

Differentiating equation (44) partially with respect to ω_{h2}^o and λ_{22} , and equating to zero gives

$$\omega_{h2}^o = D_h \exp[\lambda_{22} Q_h S_{hx}^2] \tag{45}$$

and

$$\lambda_{22} = \frac{1}{\sum_{h=1}^L Q_h S_{hx}^2} \sum_{h=1}^L \ln \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right) \tag{46}$$

Substituting (46) into (45) we obtain

$$\omega_{h2}^o = D_h \exp \left[\frac{Q_h S_{hx}^2}{\sum_{h=1}^L Q_h S_{hx}^2} \sum_{h=1}^L \ln \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right) \right] = D_h \prod_{h=1}^L \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right)^{Q_h S_{hx}^2 / \sum_{h=1}^L Q_h S_{hx}^2} \tag{47}$$

Substituting (47) into (42) gives

$$\hat{V}(\bar{y}_{pcp2}) = \sum_{h=1}^L \frac{D_h \gamma_{h2}^2}{W_h^2} s_p \exp \left(\sum_{h=1}^L \ln \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right) \right) = \sum_{h=1}^L \frac{D_h \gamma_{h2}^2}{W_h^2} s_p \prod_{h=1}^L \left(\frac{V(\bar{x}_{st})}{D_h S_{hx}^2} \right) \tag{48}$$

which is the proposed calibration product type variance estimator for population mean \bar{Y} in stratified random sampling.

Theorem 3.3: Given the product type variance estimator, a calibration product type variance estimator $\hat{V}(\bar{y}_{pcp3})$ for population mean \bar{Y} can be obtained by

$$\text{Min}D = \sum_{h=1}^L \frac{(\omega_{h3}^o - D_h)^2}{\omega_{h3}^o Q_h}$$

s.t.

$$\sum_{h=1}^L \omega_{h3}^o S_{hx}^2 = V(\bar{x}_{st})$$

given as

$$\hat{V}(\bar{y}_{pcp3}) = \sum_{h=1}^L \frac{D_h \gamma_{h3}^2}{W_h^2} s_p \left(L + \frac{1}{(V(\bar{x}_{st}))^2} (V(\bar{x}_{st}))^2 - \sum_{h=1}^L (D_h S_{hx}^2)^2 \right)^{-1/2}$$

Proof: Given the product type variance estimator, we define a calibration variance estimator as

$$\hat{V}(\bar{y}_{pcp3}) = \sum_{h=1}^L \frac{\omega_{h3}^o \gamma_{h3}^2}{W_h^2} s_p \tag{49}$$

where the weights ω_{h3}^o , are chosen such that the distance measure

$$\sum_{h=1}^L \frac{(\omega_{h3}^o - D_h)^2}{\omega_{h3}^o Q_h} \tag{50}$$

is minimized subject to the constraint

$$\sum_{h=1}^L \omega_{h3}^o S_{hx}^2 = V(\bar{x}_{st})$$

By combining (50) and the constraint gives an optimization function

$$\varphi(\omega_h^o, \lambda_{22}) = \sum_{h=1}^L \frac{(\omega_{h3}^o - D_h)^2}{\omega_{h3}^o Q_h} - 2\lambda_{22} (\sum_{h=1}^L \omega_{h3}^o S_{hx}^2 - V(\bar{x}_{st})) \tag{51}$$

Differentiating equation (51) partially with respect to ω_{h3}^o and λ_{22} , and equating to zero gives

$$\omega_{h3}^o = \frac{D_h}{[1 - 2\lambda_{22} Q_h S_{hx}^2]^{\frac{1}{2}}} \tag{52}$$

and

$$\lambda_{22} = \frac{(V(\bar{x}_{st}))^2 - \sum_{h=1}^L (D_h S_{hx}^2)^2}{2(V(\bar{x}_{st}))^2 \sum_{h=1}^L Q_h S_{hx}^2} \tag{53}$$

Substituting (52) into (53) we obtain

$$\omega_{h3}^o = D_h \left(1 + \left(\frac{(V(\bar{x}_{st}))^2 - \sum_{h=1}^L (D_h S_{hx}^2)^2}{(V(\bar{x}_{st}))^2 \sum_{h=1}^L Q_h S_{hx}^2} \right) Q_h S_{hx}^2 \right)^{-\frac{1}{2}} \tag{54}$$

Substituting (54) into (49) we obtain

$$\hat{V}(\bar{y}_{pcp3}) = \sum_{h=1}^L \frac{D_h \gamma_{h3}^2}{W_h^2} s_p \left(L + \frac{1}{(V(\bar{x}_{st}))^2} (V(\bar{x}_{st}))^2 - \sum_{h=1}^L (D_h S_{hx}^2)^2 \right)^{-\frac{1}{2}} \tag{55}$$

Which is the proposed calibration product type variance estimator for population mean \bar{Y} in stratified random sampling.

4 Empirical Studies

In this section empirical evaluation of the proposed calibration estimators are done using stimulated data set with underlying distributional assumption of Normal, Gamma and Exponential and real – life data set from a secondary source by Ojua et al. [16] was used to authenticate the result of our study.

4.1 Empirical evaluation of estimators using real-life data

In this section estimate of the mean fat content in pepper is obtained using the proposed calibration product type estimator and the conventional product type estimator. This will help to compare the precision of the proposed estimators. The data summary is presented:

$$N = 84, n = 43, \bar{X} = 5.002, \bar{Y} = 1.8042, L = 2 \quad \rho = -0.892 \quad R = 0.3607 \quad s_x^2 = 15.1722, [16]$$

The results of the analysis using excel work sheet is presented in Tables.

Table 1. Mean fat estimates for the proposed calibration product type estimators

Estimators	Estimates
\bar{y}_p	1.7271
\bar{y}_{pcp11}	1.7729
\bar{y}_{pcp12}	1.7735
\bar{y}_{pcp2}	30.0007
\bar{y}_{pcp3}	1.4807

Table 1 above shows the estimate for the mean fat, of the proposed calibration product type estimators and the conventional product type estimator with real-life data from Ojua et al. [16]. It was observed that the ratio type calibration estimator \bar{y}_{pcp12} obtained from the chi-square distance measure under one constraint gave a more precised estimate of the population mean than the other estimators in under consideration. It was also observed that the estimator \bar{y}_{pcp2} over estimated the population mean than the other estimators.

Table 2. Estimate of variance estimators for the proposed calibration product type estimator

Variance Estimators	Estimates
$\hat{V}(\bar{y}_p)$	0.002579
$\hat{V}(\bar{y}_{pcp11})$	0.002563
$\hat{V}(\bar{y}_{pcp12})$	0.002563
$\hat{V}(\bar{y}_{pcp21})$	0.04112
$\hat{V}(\bar{y}_{pcp22})$	0.04112
$\hat{V}(\bar{y}_{pcp31})$	0.00264
$\hat{V}(\bar{y}_{pcp32})$	0.00269

Table 2 shows the variance estimates for the proposed calibration product type estimators and the conventional product type variance estimator. It was observed that the regression type calibration variance estimator $\hat{V}(\bar{y}_{pcp11})$ and $\hat{V}(\bar{y}_{pcp12})$ obtained from the chi-square distance gave minimum variance.

4.2 Simulation study

To further examined the performance of the proposed calibration product type estimators for population mean, simulation was done for R = 10,000 runs using different sample sizes using R software with seed of (1113329). The result of the simulation is presented below:

Table 3. Percent average relative efficiency for gamma, normal and exponential distribution

Sample size	Distributions	\bar{y}_p	\bar{y}_{pcp11}	\bar{y}_{pcp12}	\bar{y}_{pcp2}	\bar{y}_{pcp3}
10%	Gamma	100	155.11	399.25	267.49	291.75
	Normal	100	21.01	55.01	47.70	72.60
	Exponential	100	13108.27	124843.64	1030.56	634.53
15%	Gamma	100	155.23	399.55	267.65	1715.49
	Normal	100	66.48	54.91	47.61	72.47
	Exponential	100	12283.90	203270.74	1029.61	633.60
20%	Gamma	100	154.7809	400.0738	267.9135	1725.133
	Normal	100	21.08	55.11	47.77	72.77
	Exponential	100	3838.82	57498.90	324.70	200.21
25%	Gamma	100	155.20	400.40	268.11	1733.87
	Normal	100	66.50	55.02	47.69	72.64
	Exponential	100	12104.99	139458.70	1025.78	633.55

Table 3 shows the percent average relative efficiency ($\% \overline{RE}$) of the proposed estimators and the conventional product type estimator. It was observed that under the exponential distributional assumption the proposed estimators had a high gain in efficiency with the estimator \bar{y}_{pcp11} taking the lead. When the distributional assumption was Gamma, the estimator \bar{y}_{pcp3} took the lead and when the distribution was Normal in nature, the proposed estimators had decrease efficiency.

Table 4. Percentage average absolute relative bias for gamma, normal and exponential distribution

	Distributions	\bar{y}_p	\bar{y}_{pcp11}	\bar{y}_{pcp12}	\bar{y}_{pcp2}	\bar{y}_{pcp3}
10%	Gamma	266.893	172.068	66.849	99.777	15.629
	Normal	47.623	226.642	86.568	99.841	65.597
	Exponential	1018.364	7.769	0.816	98.817	160.491
15%	Gamma	267.119	172.077	66.854	99.799	15.571
	Normal	47.538	226.116	86.568	99.859	65.594
	Exponential	1018.813	8.294	0.501	98.951	160.798
20%	Gamma	267.400	172.760	66.838	99.808	15.500
	Normal	47.709	226.286	86.567	99.865	65.565
	Exponential	1016.788	8.376	0.559	99.027	160.601
25%	Gamma	269.618	172.431	66.837	99.816	15.435
	Normal	47.628	226.498	86.563	99.869	65.566
	Exponential	1016.097	8.394	0.728	99.056	160.382

Table 4 above shows the percent average absolute bias ($\% \overline{ARB}$) for Normal distribution, Gamma distribution and Exponential distribution respectively using different sample sizes of 10%, 15%, 20% and 25%. It was observed that the proposed calibration product type estimators were more consistent under the gamma and exponential distribution, with exponential distribution taking the lead.

Table 5. Average coefficient of variation for gamma, normal and exponential distribution

Sample size	Distributions	\bar{y}_p	\bar{y}_{pcp11}	\bar{y}_{pcp12}	\bar{y}_{pcp2}	\bar{y}_{pcp3}
10%	Gamma	266.893	172.068	66.849	99.777	15.629
	Normal	47.623	226.642	86.568	99.841	65.597
	Exponential	1018.364	7.769	0.816	98.817	160.491
15%	Gamma	267.119	172.077	66.854	99.799	15.571
	Normal	47.538	226.116	86.568	99.859	65.594
	Exponential	1018.813	8.294	0.501	98.951	160.798
20%	Gamma	267.400	172.760	66.838	99.808	15.500
	Normal	47.709	226.286	86.567	99.865	65.565
	Exponential	1016.788	8.376	0.559	99.027	160.601
25%	Gamma	269.618	172.431	66.837	99.816	15.435
	Normal	47.628	226.498	86.563	99.869	65.566
	Exponential	1016.097	8.394	0.728	99.056	160.382

Table 5 shows the average coefficient of variation (\overline{CV}) for the proposed estimators and the conventional product type estimator. The result shows that the calibration product type estimator \bar{y}_{pcp11} and \bar{y}_{pcp12} are more reliable estimators of the population mean \bar{Y} . Also under the normal distribution the conventional product type estimator is more reliable than the proposed calibration product type estimators. It was also observed that as the sample size increases there was no significant difference in the reliability for the proposed estimators.

5 Discussion of Findings

Using the real life data, the population mean fat of pepper fruits \bar{Y} was calculated to be 1.8042. The result in Table 1 shows that this value was best estimated by the estimators \bar{y}_{pcp12} and \bar{y}_{pcp11} obtained using the chi-

square distance measure, with estimated value of mean fats of 1.7735 and 1.7729 respectively. This result agrees with Koyuncu and Kadilar [17] which say that in the presence of other distance measures, the chi-square distance measure gives the best estimator. This result could be because the chi-square distance measure satisfies the set constraint which says that the sum of the product of the calibrated weight and the mean of the auxiliary variable equals the population mean of the auxiliary variable.

Within the two estimators obtained using the chi-square distance measure, the ratio calibration product type estimator \bar{y}_{pcp12} , provided a better estimate of the population mean than the regression calibration product type estimator \bar{y}_{pcp11} . This result agrees with Clement and Enang [15] results which said that the ratio estimator estimate the population mean better then the regression estimator when the regression line passes through the origin. It was also observed that the conventional product type estimator was more precised in estimating the mean fat than estimators obtained using the minimum entropy and modify chi-square distance measures \bar{y}_{pcp2} and \bar{y}_{pcp3} . It is worthy of note that the estimator \bar{y}_{pcp2} obtained with the minimum entropy distance measure grossly overestimated the mean fat indicating that this estimator is highly bias than the other estimators. This could be as a result of the large weight associated with using this distance measure.

For the real life data, the result in Table 2 shows that the estimates of the variance estimators obtained using the chi-square distance gave smaller variance estimates and hence are more efficient than those obtained using other distance measures.

From the simulation study carried out under the distributional assumption of Normal, Gamma and Exponential, it was observed that the proposed calibration product type estimators had a higher gain in efficiency than the conventional product type estimator with a higher gain in efficiency recorded when the distributional assumption is exponential in nature and a loss in efficiency when the distributional assumption is normal in nature which agrees to the fact that the variate are negatively correlated. The proposed estimators are consistent estimators since as the sample size increases the performance of the estimators did not vary and the estimators obtained from the chi-square distance measure had a smaller relative bias as compared to the conventional product type estimator when the distribution is exponential. Therefore, proposed calibration product type estimators are more reliable estimators as compared to the conventional product type estimator and reaffirms that the estimators perform better when the distribution is exponential.

6 Conclusion

In this paper, we proposed calibration product type estimators of population mean in stratified sampling to be used in survey when the variables of interest are negatively correlated. The performance of these proposed estimators was compared using real – life data obtained from Ojua et al (2018) and simulated data set under the distributional assumption of Normal, Gamma and Exponential. It was shown that the calibration product type estimators obtained by minimizing the chi-square distance measure gave a better estimator with minimum variance than the other estimators obtained from the minimum entropy and modified chi-square distance measures. Also when the underlying distribution is exponential in nature, the proposed estimators outperform the conventional product type estimator.

7 Recommendation

This study recommends the proposed ratio-product type calibration estimator \bar{y}_{pcp12} for use in estimating population mean when the variable of interest is negatively correlated with the auxiliary variable and the data set is exponential in nature.

The use the constraint $\sum_{h=1}^L \gamma_h \bar{x}_h = 2\bar{X}$ to minimize the minimum entropy distance measure is recommended for further researchers. Secondly other distance measures can also be use to adjust the weight on the product type estimator in stratified sampling and compare with the once proposed in this study.

Competing Interests

Authors have declared that no competing interests exist.

References

- [1] Robson DS. Application of multivariate polykeys to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*. 1957;52:511-522.
- [2] Murthy MN. Product method of estimation. *The Indian Journal of Statistics, Series A*. 1964;26(1):69-74.
- [3] Perri PF. Combining two auxiliary variables in ratio-cum-product type estimators. *Proceedings of Italian Statistical Society. International Meeting on Statistics and Environment, Messina, 21-23 September, 2005*. 2005;193-196.
- [4] Solanki RS, Singh HP, Rathour A. An alternative estimator for estimating the finite population mean using auxiliary information in sampling surveys. *International Scholarly Research Network, ISRN Probability and Statistics*. 2012;12.
- [5] Deville JC, Sarndal CE. Calibration estimators in survey sampling. *Journal of the American Statistical Association*. 1992;87(418):376-382.
- [6] Singh S. *Advanced sampling theory with applications: How Michael 'Selected' Amy*. Kluwer Academic Publisher. 2003;1-1247.
- [7] Singh S. Golden and Silver Jubilee Year–2003 of the linear regression estimators. Presented at the Joint Statistical Meeting, Toronto. 2004;4382-4389.
- [8] Singh S. Survey statisticians celebrate Golden Jubilee Year-2003 of the linear regression estimator. *Metrika*. 2006;1-18.
- [9] Farrell PJ, Singh S. Penalized chi square distance function in survey sampling. *Proceedings of JSM, New York*; 2002.
- [10] Farrell PJ, Singh S. Model-assisted higher order calibration of estimators of variance. *Australia. & New Zealand Journal of Statistics*. 2005;47(3):375-383.
- [11] Wu C, Sitter RR. A model-calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc*. 2001;96:185-193.
- [12] Estevao VM, Sarndal CE. A new perspective on calibration estimators. *JSM-Section on Survey Research Methods*. 2003;1346-1356.
- [13] Kott PS. An overview of calibration weighting. *Joint Statistical Meeting-Section of Survey Methods*. 2003;2241-2252.
- [14] Montanari GE, Ranalli G. Nonparametric model calibration estimation in survey sampling. *Jour. Amer. Statist. Assoc*. 2005;100(472):1429-1442.
- [15] Clement EP, Enang EI. On the efficiency of ratio estimators over regression estimators. *Communication in Statistics-Theory and Methods*. 2017;46(11):5357-5367.

- [16] Ojua E, Abu N, Omeke J. Data for sensitivity of some micronutrient composition in two Nigerian pepper to treatment with different mutagens. Mendeley Data. 2018;1.
Available:<http://dx.doi.org/10.17632/Kyv29yxcj>
- [17] Koyuncu N, Kadilar C. On improvement in estimating population mean in stratified random sampling. J. Appl. Statist. 2010;37(6):999–1013.

© 2021 Enang et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<https://www.sdiarticle4.com/review-history/67505>