FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 4 No. 4, December, 2020, pp 507 – 512
DOI: https://doi.org/10.33003/fjs-2020-0404-510

A MODIFIED CALIBRATION APPROACH FOR RATIO ESTIMATOR IN STRATIFIED SAMPLING

[1]Olayiwola. O. M., [1]Apantaku, F. S., [*1]Oyekunle, K. A., [1]Akintunde, A. A., [1]Wale-Orojo, O. A. and [2]Olawoore, S. A.

[*1]Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria
[2]Department of Statistics, Oyo State College of Agriculture and Technology, Igboora , Oyo State, Nigeria.

[*]Corresponding Author's Email: oyekunle.kazeema@pg.funaab.edu.ng

ABSTRACT

Calibration is a technique for the adjustment of the original design weight to improve the precision of the survey. There is a dearth of information on calibration approach adapted for survey such that the survey cost is put into consideration. This research work developed a modified calibration approach for improving survey precision by considering the cost function. Data set on vegetable and tobacco productions (metric tonnes) were considered for this study. The data were obtained from the website of Food and Agriculture Organization. Data used was stratified based on geographical location. The population under study was divided into subpopulation of units, these subpopulations were non-overlapping homogenous sub- group. Observations were drawn within each stratum by simple random sampling with optimum allocation procedure. The proposed estimator was derived and used to determine the linear weight estimator of population parameters. The statistical properties of the derived estimator was examined. Using Lagrange multiplier, Mean Square Error and Relative Efficiency was obtained. The proposed estimator is found to be efficient.

Keywords: Auxiliary variable, calibration approach, optimality conditions, separate ratio estimator, stratified Sampling, study variable.

INTRODUCTION

Calibration estimation in sample surveys has since its introduction by Deville and Särndal (1992) developed into an established theory and method for estimation of finite population parameter. Calibration of weights is a techniques that use population data on auxiliary variables to improve estimates in sample surveys (Deville and Särndal, 1992). If auxiliary data are available, some improvement in the precision of estimate may be achieved. Incorporation of auxiliary data in the estimation process is known as calibration. Calibration weights defined by minimizing a distance measure under calibration equations can be very large or negative. If the weights are to be used to estimate the population total, it seems reasonable that no individual weight should be less than one. Raoand Singh (1997) proposed a method of ridge shrinkage which is an iterative method of adjusting weight to meet a range restriction and to satisfy the calibration equation within given tolerances. Horn and Yu (1998), and Kim, Sungur and Heo (2006) introduced the calibration estimation in stratified sampling. They suggested the calibration estimators, respectively, for combined generalized regression estimator and combined ratio estimator using a single auxiliary information. Chen and Qin (1993) suggested a calibrated estimator that makes an efficient use of auxiliary variables for equal probability sampling by maximizing the constrained empirical likelihood.

Kott (2006) defines calibration weights as a set of weights, for units in the sample, that satisfy a calibration to known populati on totals, and such that the resulting estimator is randomization consistent (design consistent), or, more rigorously, that the d esign bias is, under mild conditions, an asymptotically insi gnificant contribution to the estimator's mean squared error Chami et al (2012) proposed a two-parameter ratio-product-ratio estimator for a finite population mean in a simple random sample without replacement following the methodology used in the studies of Ray and Sahai (1981) estimators. Malik and singh (2012) defined a multivariate-ratio type estimators using geometric and harmonic means in stratified random sampling. Subramani and Kumarapandiyan (2012) proposed two modified ratio estimators for estimating the population mean by using the linear combination of the known population values of median and co-efficient of kurtosis of the auxiliary variable.

In stratified random sampling, calibration approach is used to obtain optimum strata weights for improving the precision of survey estimates of population parameters. Kim, Sungur and Heo (2007), Koyuncu and Kadilar (2013) defined some calibration estimators in stratified random sampling for population characteristics and Clement *et al* (2014b) defined calibration estimators for domain totals in stratified random sampling. In this study, we intend to modify calibration approach for improving survey precision by considering the cost function.

METHODOLOGY

The reviews of Calibration Estimation

Consider a population consisting of N = (1,2,…,I,…,N) from which a probability sample $s(s\mathbb{C}H)$ is drawn with a given sampling design p(s). The inclusion probabilities$\pi_i = p(i\epsilon s)$and $\pi_{ij} \in s$ are assumed to be strictly positive and known. Let y_i be the value of study variable, y, for the i[th] population unit and let x_i be the value of i^{th} unit of the associated auxiliary variable. The population total $X=\sum_{i\epsilon s} x_i$ of the auxiliary variable x is assumed to be accurately known. Under a probability sampling design *P*, with probability*P*(s). Deville and Sarndal (1992) gave an unbiased estimator of the population total as $\hat{y} = \sum_{i=1}^{n} d_i y_i$, where $d_i=\frac{N}{n}$ is the design weight associated with unit i and defined as the inverse of the inclusion probability($\pi_i > 0$ *for each i*), *where* $\pi_i = \sum_{i\epsilon s} p(s)$.Followed from Etebong (2015), Calibration ratio estimator under the stratified sampling is given by

$$\hat{Y}_G = \sum_{i\epsilon s} w^*{}_i y_i \quad (1)$$

Let $\hat{R}_j = \frac{\bar{y}_i}{\bar{x}_i}$; $\bar{x}_i \neq 0$ be the estimate of the ratio $\hat{R}_i = \frac{\bar{Y}_i}{\bar{X}_i}$,$\bar{X}_i \neq 0$ of j_{th} stratum in the population. $\hat{Y}_G = \sum_{i\epsilon s} w^*{}_i y_i \hat{R}_i$ (2)

With the new weight $w^*{}_j$ called the calibration weights which are chosen such that the chi square (CS) type of distance

$$\sum_{i\epsilon s}(w^*{}_i - w^*{}_i)^2 (w^*{}_i q_i)^2 \quad (3)$$

is minimum, subject to the condition

$$\sum_{i\epsilon s} w^*{}_i x_i = \bar{X} \quad (4)$$

Note that q_i in (3), is a suitably chosen weight which determines the form of the estimator. Minimization of (3), subject to the calibration equation (4), leads to the combined generalized regression estimator (GREG) given by

$$\hat{Y}_G = \sum_{i\epsilon s} d_i y_i + \hat{\beta}_{ds}\{X - \sum_{i\epsilon s} d_i x_i\} \quad (5)$$

for the optimum choice of weights

$$w^*{}_j = d_i + (d_{iq_i}\bar{x}_i / \sum_{i\epsilon s} d_i\, q_i \bar{x}_i{}^2)(\bar{X} - \sum_{i\epsilon s} d_i \bar{x}_i)$$

Square both side

$$W_i^{*2} = (d_i + (\bar{X} - \sum_{i\epsilon s} d_i \bar{x}_i) d_i q_i \bar{x}_i / \sum_{i\epsilon s} d_i\, q_i \bar{x}_i{}^2)^2 \quad (6)$$

And setting the turning parameter $q_i = x_i{}^{-1}$, then $W_i^{*2} = (\frac{\bar{X}}{\bar{x}_{st}})^2$

Where $\bar{x}_{st} = \sum_{i\epsilon s} d_i \bar{x}_i$

Substituting (6) into (2), We obtained the calibration ratio estimator under the stratified sampling as:

$$\hat{Y}_G = (\hat{R}_j d_i \bar{x}_i + \frac{\sum_{i\epsilon s} \hat{R}_j d_i q_i \bar{y}_i \bar{x}_i{}^2}{\sum_{i\epsilon s} d_i q_i \bar{x}_i{}^2}(\bar{X} - \sum_{i\epsilon s} d_i \bar{x}_i) \quad (7)$$

Variance Estimator For the calibration Approach Separate Ratio Estimator

In other to get the calibration approach ratio estimator in stratified sampling, the general estimator of variance of the calibration ratio estimator of (7) is given b y

$$\text{Var}(\hat{Y}_G) = \sum_{i=1}^{k} W_i^{*2} y_i S_{iy}{}^{*2}$$

$$\text{Var}(\hat{Y}_G) = \sum_{i=1}^{k} (\frac{\bar{X}}{\bar{x}_{st}})^2 y_i S_{iy}{}^2$$

Where $y_i = \frac{(1-f_i)}{n_i}$; $S_{iy}{}^2 = \frac{1}{(N_j-1)} \sum_{i=1}^{k} (y_{ij} - \bar{y}_i)^2$ is the j-th stratum variance.

Let $\bar{\bar{Y}}_G = \frac{\bar{y}_i}{\bar{x}_i}\bar{X}$ be the ratio estimates of population mean $\bar{Y}$ under the simple random sampling of size $n_i(n_i\ large), then$;

$$\text{Var}(\hat{Y}_G) = \frac{1-f_i}{n} \frac{1}{(N_i-1)} \sum_{i=1}^{k} (y_i - Rx_i)^2 \quad (8)$$

Let $\bar{\bar{Y}}_{RI} = \frac{\bar{y}_i}{\bar{x}_i}$; $\bar{x}_i$ be the ratio estimate of the population mean $\bar{Y}$ under stratified sampling of size $n_i(n_i\ large), then;$

$$\text{Var}(\hat{Y}_G) = W_i^{*2} y_i \frac{1}{(N_i-1)} \sum_{i=1}^{k} (y_i - Rx_i)^2$$

By partitioning

$$\text{Var}(\hat{Y}_G) = W_i^{*2} y_i \frac{1}{(N_i-1)} \sum_{i=1}^{k} [(y_i - \bar{y}_i) - R(x_i - \bar{x}_i)]^2$$

$$\text{Var}(\hat{Y}_G) = W_i^{*2} y_i \frac{1}{(N_i-1)} \sum_{i=1}^{k} [(y_i - \bar{y}_i)^2 + R^2[(x_i - \bar{x}_i)^2 - 2R[(y_i - \bar{y}_i)(x_i - \bar{x}_i)] \quad (9)$$

$$\text{Var}(\hat{Y}_G) = W_i^{*2} y_i (S_{iy}{}^2 + R^2 S_{ix}{}^2 - 2RS_{ixy})$$

Therefore

$$\text{Var}(\hat{Y}_G) = \sum_{i=1}^{k} W_i^{*2} y_i (S_{iy}{}^2 + R^2 S_{ix}{}^2 - 2RS_{ixy})$$

Note that since $\hat{Y}_G = \sum_{i=1}^{k} \bar{\bar{Y}}_G$ and sampling is independent in each stratum, then

Var$(\hat{Y}_G) = \sum_{i=1}^{k} Var(\bar{\bar{Y}}_G)$

the variance for the calibration approach separate ratio estimator is obtained as follows:

$$\text{Var}(\hat{Y}_G) = \sum_{i=1}^{k} (\frac{\bar{X}}{\bar{x}_{st}})^2 y_i (S_{iy}{}^2 + R^2 S_{ix}{}^2 - 2RS_{ixy}) \quad (10)$$

Generalized Distances (Chi-Square Distance)

In some cases it may be necessary to conduct a sample survey with a fixed **cost**. This method is based on the cost aspect of the survey and the cost function is defined as:

$$C_1 = C_0 + \sum_{i=1}^{L} n_i C_i \quad (11)$$

Where C_0 stands for the known overhead cost.

The proposed Estimator

Motivated by Etebong (2015) a modified calibration approach for ratio estimator in stratified sampling was developed using the existing variance by Etebong where is derived variance is

$$\text{Var}(\hat{Y}_G) = \sum_{i=1}^{k} (\frac{\bar{X}}{\bar{x}_{st}})^2 y_i \left(S_{iy}{}^2 + R^2 S_{ix}{}^2 - 2RS_{ixy}\right) \quad (12)$$

To derive our variance using the cost function stated in (11)

Let $S_{iy}{}^2 + R^2 S_{ix}{}^2 - 2RS_{ixy} = A_i$ such that the variance (12), becomes

$$\text{Var}(\hat{Y}_G) = \sum_{i=1}^{L} W_i^{*2} y_i A_i \quad (13)$$

Total cost is fixed: Minimization of (12) subject to (11) leads to the Lagrange function

$$L_1 = \sum_{i=1}^{L} W_i^{*2} y_i A_i + \lambda \left[C_0 + \sum_{i=1}^{L} n_i C_i - C_1\right] \quad (14)$$

On differentiating (14) with respect to n_i and equating to zero we have

$$W_i^{*2} A_i = -\lambda C_i n_i^2$$

To make n_i the subject of the formula

$$\frac{W_i^* \sqrt{A_i}}{\sqrt{\lambda}\sqrt{C_i}} = n_i \quad (15)$$

note that $\sum_{i=1}^{L} n_i = n$ from (15) we have

$$\frac{1}{\sqrt{\lambda}} = n - \sum_{i=1}^{L} \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}} \quad (16)$$

On substituting (16) in (15) we have

$$n_i = \frac{W_i^* \sqrt{A_i}}{\sqrt{\lambda}\sqrt{C_i}} = nW_i^* \sqrt{A_i} \Big/ \left[\sqrt{C_i} \sum_{i=1}^{L} \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}}\right] = n \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}} \Big/ \sum_{i=1}^{L} \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}} \quad (17)$$

In a particular case if $C_1 = C_2 = \cdots = C_L = C$, that is the cost of sampling in each stratum is the same then (17) becomes

$$n_i = n \left[\frac{W_i^* \sqrt{A_i}}{\sum_{i=1}^{L} W_i^* \sqrt{A_i}}\right] \quad (18)$$

In other words the optimum allocation reduces to the famous Neyman (1934) allocation. On substituting (17) in (13) , the variance of the estimator $\hat{y}_G$ under optimum allocation is given by

$$\text{Var}(\hat{Y}_G)_{opt} = \sum_{i=1}^{k} W_i^{*2} A_i \left[\sum_{i=1}^{L} \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}} \Big/ n \left[\frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}}\right] - \frac{1}{N_i}\right]$$

$$= \frac{1}{n} \left[\sum_{i=1}^{L} W_i^* \sqrt{C_i} \sqrt{A_i}\right] \left[\sum_{i=1}^{L} \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}}\right] - \sum_{i=1}^{L} \frac{W_i^{*2} A_i}{N_i} \quad (19)$$

Bysubstituting , the variance of the estimator $\hat{Y}_G$ under Neyman allocation is given by

$$\text{Var}(\hat{Y}_G)_N = \frac{1}{n} (\sum_{i=1}^{L} W_i^* \sqrt{A_i})^2 - \sum_{i=1}^{L} \frac{W_i^{*2} A_i}{N_i} \quad (20)$$

it can be easily shown that if f_i is negligible then

$$\text{Var}(\hat{Y}_G)_N = \frac{1}{n} \left[\sum_{i=1}^{L} W_i^* \sqrt{C_i} \sqrt{A_i}\right] \left[\sum_{i=1}^{L} \frac{W_i^* \sqrt{A_i}}{\sqrt{C_i}}\right] \quad (21)$$

And

$$\text{Var}(\hat{Y}_G)_N = \frac{1}{n} (\sum_{i=1}^{L} W_i^* \sqrt{A_i})^2 \quad (22)$$

RESULT AND DISCUSSION

To judge the relative performances of the proposed modified calibration approach separate ratio estimator in stratified sampling, data set from food and agriculture's organization on vegetable and tobacco production where consider for analysis.

Table I: Sample Information For Vegetable Production

parm	Stratum1	Stratum2	Stratum3	Stratum4	Stratum5	Stratum6	Stratum7	Stratum8	Stratum9	Stratum10
N_i	6	6	8	10	12	4	26	15	10	3
n_i	4	3	2	6	7	2	15	14	5	2
S_{iy}^2	51.795	27.299	50	64.47	295.7	41.49	10.7	293.373	65.5	92.2
S_{ix}^2	3E+09	2E+10	4E+10	8E+10	4E+10	2E+10	1.93E+1	1.96E+15	1E+10	5E+09
S_{xy}^2	21325	361875	-3E+06	1E+06	2E+06	2E+06	274636	-1.5E+07	8E+05	4E+05
A_i	51.626	32.332	113.5	52.2	291	21.34	91.53	86366.74	60.89	84.97
W_i	0.06	0.06	0.08	0.1	0.12	0.04	0.26	0.15	0.1	0.03
V_i	32.929	20.636	54.28	19.97	92.81	20.43	13.473	22035.86	23.3	108.4
$\bar{X}_{ist}$	2989.2	20493	45955	12223	60843	14224	139587	2231893	22582	2347
$\bar{Y}$	11.965	13.688	27.32	16.57	27.83	16.19	8.975	18.631	15.14	29.02
$\bar{X}$	49821	341548	5E+05	1E+05	5E+05	4E+05	536873	1487928	2E+05	78217

Table II: Mean ,Skweness and Kutorsis of yield of Vegetable

Yield of vegetable in Metric Tonnes(Y)		
Mean	**Skweness**	**Kutorsis**
185.3346	**1.408**	**1.022**

$$\text{Var}(\hat{Y}_G)_N = \frac{1}{n}\left(\sum_{i=1}^{L} W_i^{*}\sqrt{A_i}\right)^2 = 9913620.8$$

$$\text{Var}(\hat{Y}_{RS}) = \sum_{i=1}^{k}\left(\frac{\bar{X}}{\bar{x}_{st}}\right)^2 y_i\left(S_{iy}^{\ 2} + R^2 S_{ix}^{\ 2} - 2RS_{ixy}\right) = 9459491$$

$$\text{Var}(\hat{Y}_{ST}) = \sum_{i=1}^{L} W_i^{*2} y_i\, S_{iy}^{\ 2} = 9913620.8$$

Performance of estimators from analytical study for vegetable.

The Mean square error of the usual unbiased estimator in stratified sampling ($\hat{Y}_{ST}$), the existing estimator $\widehat{(Y_{RS})}$ and the modified estimator Var$(\hat{Y}_G)_N$ were computed and present below.

Estimator	Variance	$PRE(\theta, \hat{Y}_{ST})$
$Var(\hat{Y}_{ST})$	9913620.8	100
Existing	9459491	1.04801
Proposed	2314910	0.23351

The relative efficiency for vegetable is 4.09

TABLE III: Sample Information for Tobacco Production

Parameter	statum1	statum2	statum3	statum4	statum5	statum6	statum7	statum8	statum9	Statum10
N_i	6	6	8	10	12	30	17	10	3	4
n_i	2	3	7	7	2	17	15	2	3	3
S^2_{iy}	0.0268	0.2181	0.347	0.235	0.582	0.153	0.344	0.379	2.018	0.975
S^2_{ix}	108996	584984	6E+05	2E+05	3E+05	6E+05	8E+05	1E+05	8E+05	29633
S^2_{xy}	249.83	-8156	-3667	-1327	892	-561	3984	-1107	-2841	1153
A_i	51.626	32.332	113.5	52.2	291	21.34	91.53	86367	60.89	84.97
W_i	0.0566	0.0566	0.075	0.094	0.113	0.038	0.283	0.16	0.094	0.028
V_i	0.0587	0.5676	0.551	0.332	0.213	0.141	0.517	2.143	1.89	0.251
$\bar{X}_{ist}$	2989.2	20493	45955	12223	60843	14224	1E+05	2E+06	22582	2347
$\bar{Y}$	1.9733	1.3883	2.5562	1.549	1.83166	1.47	1.115	1.38176	1.721	2.08666
$\bar{X}$	3194.5	14660	18309.4	14923.5	5987.83	3450	11682.7	68662.4	33976.1	1333.33

Table IV:Mean ,SkwenessandKutorsis of yield of Tobacco

Yield of Tobacco in Metric Tonnes(Y)		
Mean	**Skweness**	**Kutorsis**
17.0731	5.232	31.077

$$Var(\hat{Y}_G)_N = \frac{1}{n}\left(\sum_{i=1}^{L} W_i^{*}\sqrt{A_i}\right)^2 = 0.7403158$$

$$Var(\hat{Y}_{RS}) = \sum_{i=1}^{k}\left(\frac{\bar{X}}{\bar{x}_{st}}\right)^2 y_i\left(S_{iy}^{\ 2} + R^2 S_{ix}^{\ 2} - 2RS_{ixy}\right) = 1.601317$$

$$Var(\hat{Y}_{ST}) = \sum_{i=1}^{L} W_i^{*2} y_i S_{iy}^{\ 2} = 4.696582$$

Performance of estimators from analytical study for Tobacco.

The percent relative efficiency of the usual unbiased estimator in stratified sampling ($\hat{Y}_{ST}$), the existing estimator $(\widehat{Y_{RS}})$ and the modified estimator Var$(\hat{Y}_G)_N$ were computed and present in table IV.

Estimator	Variance	$PRE(\theta, \hat{Y}_{ST})$
$Var(\hat{Y}_{ST})$	4.696582	100
Existing	1.601317	0.34095
Proposed	0.7403158	0.15763

The relative efficiency for Tobacco is 2.16

From the tables above, the statistical properties of the derived estimator and the proposed estimator were examined by considering the fixed cost using lagrange multiplier to derived Mean Square Error and Relative Efficiency for each dataset. The average yield for vegetable and tobacco was 185.3346 and 17.07301 respectively. The coefficients of skewness for the vegetable and tobacco data set were 1.408 and 5.232 respectively with kurtosis 1.022 and 3.770. The MSE of existing estimator for vegetable and tobacco were 9459491 and 1.601317 while the proposed estimator were 2314910 and 0.7403158 respectively. The relative efficiency for the vegetable and tobacco data sets was 4.09 and 2.16 respectively. This showed that the relative efficiency for both data sets was greater than one which indicate that the proposed estimator is efficient.

CONCLUSION

This research work developed a modified calibration approach for improving survey precision by considering the cost function. Data set on vegetable and tobacco productions (metric tonnes) were considered for this study. The data were obtained from the website of Food and Agriculture Organization. Data used was stratified based on geographical location. The population under study was divided into subpopulation of units, these subpopulations were non-overlapping homogenous sub- group. Observations were drawn within each stratum generally by well known procedure of simple random sampling using optimum allocation. The statistical properties of the derived estimator was examined. Using Lagrange multiplier the Mean Square Error and Relative Efficiency was obtained. An application to real life situation indicate that the proposed estimator is found to be efficient than other compared estimator.

REFERENCES

Chami, P.S. and Singh, B. and Thomas, D. A (2012) two-parameter ratio-product-ratio estimator using auxiliary information, ISRN probability and Statistics. Article ID 103860. doi:10.5402/2012/103860.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. Biometrika, 80:107–116.

Clement, E. P., Udofia, G.A. and Enang, E. I. (2014b). Sample design for domain calibration estimators. International Journal of Probability and Statistics, **3** (1), 8-14.

Deville, J.C. and Sarndal, C. E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, **87**, 376-382.

Kim, J.M. Sungur, E.A. and Heo T.Y. (2007). Calibration approach estimators in stratified sampling.Statistics and Probability Letters, **77**(1), 99-103.

Kott, P. S. (2006). Using calibration weighting to adjust for nonres ponse and coverage errors. Survey Methodology, 32, 133-142.

Koyuncu, N. and Kadilar, C. (2013). Calibration estimators using different measures in stratified random sampling. International Journal of Modern Engineering Research, **3** (1), 415-419.

Malik, S. & Singh, R. (2012): Some improved multivariate-ratio-type estimators using geometric and harmonic means in stratified random sampling. ISRN Probability and Statistics, Article ID 509186.

Rao, J. N. K. and Singh, A. C. (1997). A ridge shrinkage method for range restricted weight calibration in survey sampling. Proceedings of the section on survey research methods, American Statistical Association. 57–64.

Ray, S.K. and Singh, R.K. (1981) Difference-cum-ratio type estimators. J. Ind. Statist. Assoc., 19, 147-151.

Horn. S. and Yu, F. (1998). Estimation of variance of gene ral regression estimator: Higher level calibration approach. *Survey Methodology*, 24, 41-50.

Subramani, J. and Kumarapandiyan, G. (2012). A Class of almost unbiased modified ratio estimators for population mean with known population parameters, Elixir Statistics 44, 7411-741.