Modified Ratio Estimator Using Multiple Auxiliary Variables.

O.M. Olayiwola, Ph.D.[*1]; F.S. Apantaku, Ph.D.[1]; A.A. Adewara, Ph.D.[2]; and A.F. Gboyega, M.Sc.[1]

[1]Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria.
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria.

[*]E-mail: laniyimathew@yahoo.com

ABSTRACT

This study developed a modified ratio estimator using one variable of interest, multi-auxiliary variables and the product of Kurtosis and Median of the auxiliary variables. The questionnaire contained detail about workers income and expenditures. Boxplot was used to check for outliers. The biases and mean squared errors (MSE) were computed for both existing and modified ratio estimator. Principal Components Analysis (PCA) was used to determine the linear combination of auxiliary variables that best explain the variable of interest. The scree and bar plots were used to arrange eigenvalues in descending order. The expenditure on food stands as an outlier. The bias and MSE of the existing ratio estimator were 74.39586 and 8,382,470 while that of modified ratio estimator were 30.97307 and 6,128,431 respectively. PCA generated two principal components, which accounted for 76.2% of the total variance. The modified ratio estimator was efficient.

(Keywords: ratio estimator, mean square error, MSE, principal component, auxiliary variable)

INTRODUCTION

In surveys, samples are used instead of population and most of these samples are prepared by statisticians. One of the areas of statistics that is most commonly used in all fields of scientific investigation is that of probabilistic sampling. Surveys used by social scientists are based on complex sampling designs (Lumley, 2004; Winship and Radbill, 1994).

Sampling is not mere substitution of a partial coverage for a total coverage. Sampling is the science and art of controlling and measuring the reliability of useful statistical information through the theory of probability (Deming, 1950). The simplest and the most common method of sampling is simple random sampling in which a sample is drawn unit by unit, with equal probability of selection for each unit at each draw, where there is no additional information available.

In sample surveys, along with study variable Y, information on auxiliary variable X, which is correlated with is also collected. This information on auxiliary variable may be utilized to obtain a more efficient estimator of the population mean. Ratio method of estimation using auxiliary information is an attempt in this direction. The auxiliary information is frequently used to increase precision of the population estimates by taking advantage of the correlation between the study variable and the auxiliary variable. Several authors including Kadilar and Cingi (2004, 2006a, b, c) and Gupta and Shabbir (2008) have proposed different estimators by utilizing information on the auxiliary variable for estimation of population mean.

For ratio estimators in sampling theory, population information of the auxiliary variable, such as Coefficient of variation, Coefficient of Correlation, Coefficient of Skewness, is often used to increase the efficiency of the estimation for a population mean. Murthy (1967), Cochran (1977), Prasad (1989), Sen (1993), Upadhyaya and Singh (1999), Singh and Tailor (2003, 2005), Singh *et al.* (2004), Kadilar and Cingi (2004, 2006), Koyuncu and Kadilar (2009), and Yan and Tian (2010) among others used the population information of the auxiliary variable to increase precision. Though, attempts have been made by Subramani, and Kumarapandiyan, (2012) using the linear combination of known population value of coefficient of Kurtosis and Median of the auxiliary variable, however this study is combining the interactive effect of the coefficient of Kurtosis and Median of the multiple auxiliary variables to improve the ratio estimators.

The main objective of this study is to modify the existing ratio estimator by Subramani and Kumarapandiyan, (2012), using multiple auxiliary variables to facilitate bias reduction and improve the precision of ratio estimator.

MATERIALS AND METHODS

Basic: Khoshnevisan *et al* (2007) defined their family of estimators as:

$$t = \bar{y}\left[\frac{a\bar{X}+b}{\alpha(a\bar{x}+b)+(1-\alpha)(a\bar{X}+b)}\right]^g$$

Where $\alpha \neq 0$, a and b are either real numbers or a functions of the known parameters of the auxiliary variable x such as standard derivation σ_x, coefficient of variation C_x, Skewness $\beta_{1(x)}$, kurtosis $\beta_{2(x)}$ and correlation coefficient ρ

When α=1 a=1, b= 0, g=0, we have the usual ratio estimator, $\theta_0 = \bar{y}$ with MSE $\theta_0 = \frac{N-n}{Nn}\bar{Y}^2 C_y^2$

When α=0 a=1, b= 0, g=1, we have the usual ratio estimator, $\theta_1 = \bar{y}\frac{\bar{X}}{\bar{x}}$ with MSE $\theta_1 = \frac{N-n}{Nn}\bar{Y}^2(C_y^2 + C_x^2 - 2\rho C_x C_y)$

When α=1 a=1, b= 0, g=-1, we have the usual product estimator, $\theta_2 = \bar{y}\frac{\bar{x}}{\bar{X}}$ with MSE $\theta_2 = \frac{N-n}{Nn}\bar{Y}^2(C_x^2 + C_y^2 + 2\rho C_x C_y)$

When α=1 a=1, b= C_x, g=1, Sisodia and Dwivedi (1981) ratio estimator $\theta_3 = \bar{y}(\frac{\bar{x}+C_x}{\bar{X}+C_x})$ with MSE

$$\theta_3 = \frac{N-n}{Nn}\bar{Y}^2\left[C_y^2 + \left(\frac{\bar{X}}{\bar{x}+C_x}\right)^2 C_x^2 - 2\frac{\bar{x}}{\bar{x}+C_x}\rho C_x C_y\right]$$

When α=1 a=1, b=C_x, g=1, we have Pandey and Dubey [1988] product estimator with $\theta_4 = \bar{y}(\frac{\bar{X}+C_x}{\bar{x}+C_x})$ with MSE

$$\theta_4 = \frac{N-n}{Nn}\bar{Y}^2\left[C_y^2 + \left(\frac{\bar{x}}{\bar{x}+C_x}\right)^2 C_x^2 + 2\frac{\bar{x}}{\bar{x}+C_x}\rho C_x C_y\right]$$

When α=1 a=1, b= ρ, g=1, we have Singh, Taylor [2003] ratio estimator as:

$$\theta_5 = \bar{y}(\frac{\bar{X}+\rho}{\bar{x}+\rho}) \text{ with MSE}$$

$$\theta_5 = \frac{N-n}{Nn}\bar{Y}^2\left[C_x^2 + \left(\frac{\bar{x}}{\bar{x}+\rho}\right)^2 C_x^2 - 2\frac{\bar{x}}{\bar{x}+C_x}\rho C_x C_y\right]$$

When α=1 a=-1, b=ρ5, g=1, we have Singh, Taylor [2003] product estimator $\theta_6 = \bar{y}(\frac{\bar{x}+\rho}{\bar{X}+\rho})$ with MSE

$$\theta_6 = \frac{N-n}{Nn}\bar{Y}^2\left[C_y^2 + \left(\frac{\bar{X}}{\bar{X}+\rho}\right)^2 C_x^2 + 2\frac{\bar{X}}{\bar{X}+\rho}\rho C_x C_y\right]$$

When $\alpha = 1$, a=β_2 , b=M_d, g=1, we have Subramani, J. and Kumarapandiyan, G. (2012) ratio estimator. $\theta_7 = \left[\frac{\bar{y}(\bar{X}\beta_2+Md)}{\bar{x}\beta_2+Md}\right]$ with MSE

$$\theta_7 = \frac{N-n}{Nn}\bar{Y}^2\left[C_y^2 + \left(\frac{\bar{x}\beta_2}{\bar{x}\beta_2+Md}\right)^2 C_x^2 - 2\theta_3 C_y C_x \rho\right]$$

Notations used are:

N- Population Size
n- Sample Size
$f = {}^{n}/_{N}$ Sampling Fraction
Y- Variable of Interest
X- Auxiliary Variable
$\bar{X}, \bar{Y}$- Population Means
$\bar{x}, \bar{y}$- Sample Means
S_x, S_y- Population Standard Deviations
C_x, C_y- Co-Efficient of Variations
ρ - Co-Efficient of Correlation
$\beta_1 = \frac{N\sum_{i=1}^{N}(X_i-\bar{X})^3}{(N-1)(N-2)S^3}$ Co-Efficient of Skewness of the Auxiliary Variable
$\beta_2 = \frac{N(N+1)\sum_{i=1}^{N}(X_i-\bar{X})^4}{(N-1)(N-2)(N-3)S^4} - \frac{3(N-1)^2}{(N-2)(N-3)}$ Co-Efficient of Kurtosis of the Auxiliary Variable
Md - Median of the Auxiliary Variable

The new ratio estimator will be derived from the family estimator explained above. Where $\alpha = 1$, a = β_2, g = 1, and b = $\beta_2 Md$. We have:

Estimator $(\hat{\bar{\bar{Y}}})$ $= \left[\frac{\bar{y}(\bar{X}\beta_2+\beta_2 Md)}{\bar{x}\beta_2+\beta_2 Md}\right]$

Multiple auxiliary variables with one variable of interest were used and the new estimator new estimator becomes:

$$\left[\frac{\bar{y}\left(\begin{pmatrix}\bar{X}_1\\ \bar{X}_2\\ \vdots\\ \bar{X}_p\end{pmatrix}(\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P) + (\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P)\begin{pmatrix}Md_1\\ Md_2\\ \vdots\\ Md_p\end{pmatrix}\right)}{\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_p\end{pmatrix}(\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P) + (\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P)\begin{pmatrix}Md_1\\ Md_2\\ \vdots\\ Md_p\end{pmatrix}}\right]$$

Mean Square Error (MSE)

$$MSE \quad = \frac{N-n}{Nn}\bar{Y}^2[C_y^2 + \theta_p^2 C_x^2 - 2\theta_p C_y C_x \rho]$$

$$= \frac{(1-f)\bar{Y}^2(C_y^2+\theta_p^2 C_x^2-2\theta_p C_y C_x \rho)}{n}$$

Bias $= \frac{(1-f)\bar{Y}(\theta_p^2 C_x^2-\theta_p C_y C_x \rho)}{n}$

Where $C_{xx} = \frac{S_x^2}{\bar{X}^2}$, $C_{yy} = \frac{S_y^2}{\bar{Y}^2}$ and $C_{yx} = \frac{S_{yx}}{\bar{Y}\bar{X}}$

The constant (θ_p) $= \left[\frac{\bar{x}\beta_2}{\bar{x}\beta_2+\beta_2 Md}\right]$

$$= \left[\frac{\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_p\end{pmatrix}(\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P)}{\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_p\end{pmatrix}(\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P) + (\beta_1, \ \beta_2 \ \dots \ \dots \ \beta_P)\begin{pmatrix}Md_1\\ Md_2\\ \vdots\\ Md_p\end{pmatrix}}\right]$$

The constant (θ_p) $= \left[\frac{\bar{x}\beta_2}{\bar{x}\beta_2+\beta_2 Md}\right]$

Principal Components

Given a sample of n observation on a vector of P variables X = $(x_1, x_2 \dots, x_p)$, define the first principal component (PC) of the sample by linear transformation:

$$z_1 = a_1^T X = \sum_{i=1}^{p} a_{i1} x_i$$

where the vector $a_1 = (a_{11}, a_{21} \dots, a_{pi})$ is chosen such that var (z_1) is maximum. Likewise, defined the k^{th} PC of the sample by linear transformation:

$$z_k = a_k^T X \quad k= (1\dots p)$$

Where vector $a_k = (a_{1k}, a_{2k}, \dots, a_{pk})$ is chosen such that var $([z_k]$ is maximum, subject to cov $[z_k, z_1]$ = 0 for k $\geq l \geq 1$ and $a_k^T a_k = 1$

Scree Graph

The term scree refers to the geological term for the debris at the bottom of a rocky cliff. We could plot the eigenvalues in an attempt to find a visual break between the large eigenvalues and the small eigenvalues. This plot is called scree graph.

RESULTS AND DISCUSSION

Four hundred and sixty-two (462) questionnaires were administered to staff of Federal University of Agriculture, Abeokuta, (FUNAAB) Ogun state, Nigeria to determine their total monthly expenditure (Y) , house rent (X_1), Transportation fare (X_2), feeding cost (X_3), Electricity bill (X_4), Toiletries cost(X_5), Private coaching fee(X_6), Fueling of car (X_7), Fueling of generator (X_8) and Airtime cost (X_9).

The boxplot revealed that X_3 stands as an outlier, as shown in Figure 1, which indicates that the respondent spent more on food items. The variable of interest (Y) and each of the auxiliary variables (X) were used to test the efficiency of the new ratio estimator and the existing ratio estimator. The biases and MSE of the new estimator is lesser than the existing ratio estimator. Combining all the auxiliary variables we observed that the bias and MSE of the new ratio estimator are smaller compared with the existing estimator.

The parameters estimates were computed for variable of interest (Y) and all the auxiliary variables (X).

The proposed estimator is more efficient than the existing estimator, as the biases and mean squared error are less than that of existing estimators.

The graph of the biases and mean squared error of the proposed (blue line) and existing (red line) were shown in Figures 3 and 4, respectively. The graph also shows that the proposed estimator is more efficient than the existing estimator.

The principal component on the basis of the sample covariance matrix for the merged sample data sets for the first two components of the auxiliary variables which accounted for 76.2% of the total components are:

$$\begin{aligned} Y_1 = -0.1453589X_1 - 0.1755014X_2 \\ - 0.8521833X_3 \\ - 0.1398828X_4 \\ - 0.1290020X_5 \\ - 0.1481015X_6 \\ - 0.3121008X_7 \\ - 0.1863985X_8 \\ - 0.1776501X_9 \end{aligned}$$

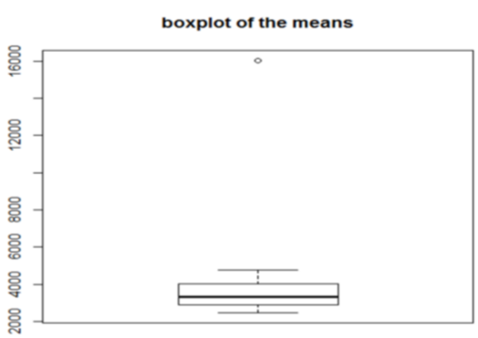

Figure 1: Box Plot for the Means of Auxiliary Variables.

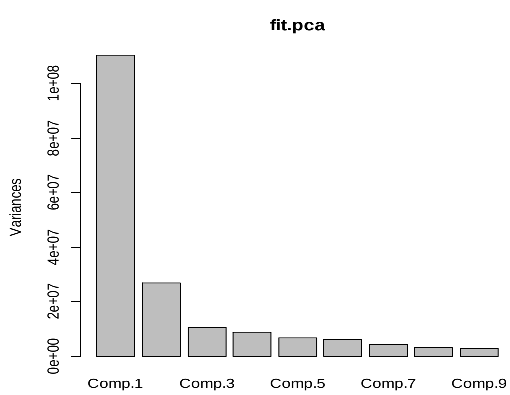

Figure 2: The Bar Plot.

$$\begin{aligned} Y_2 = -0.2390244X_1 - 0.3463498X_2 \\ + 0.5017204X_3 \\ - 0.2928875X_4 \\ - 0.3308010X_5 \\ - 0.2515492X_6 \\ - 0.2944923X_7 \\ - 0.4094889X_8 \\ - 0.2414322X_9 \end{aligned}$$

with corresponding sample variance of 0.6126 and 0.1492 respectively, the total variance is 0.7618. Using the two components obtained from the PCA in the proposed ratio estimator, the bias equals zero and MSE also equals zero, which also establish that the proposed estimator is efficient than existing estimator.

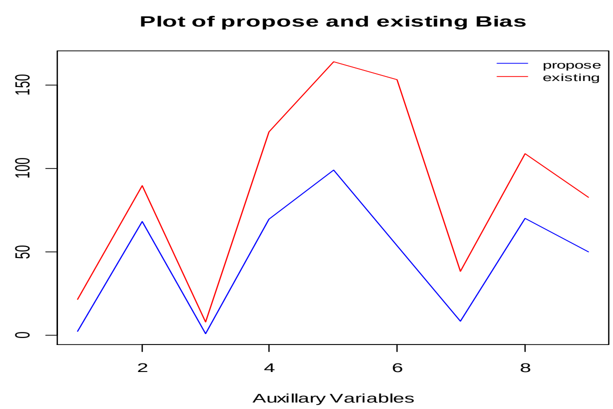

Figure 3: Graph of Bias of Proposed and Existing Estimator.

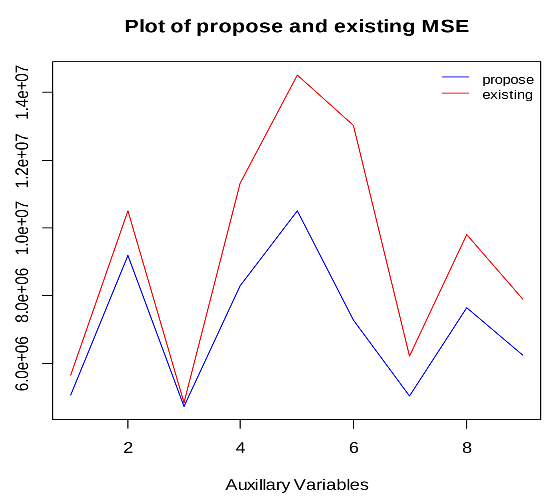

Figure 4: Graph of Mean Squared Error of Proposed and Existing Estimator

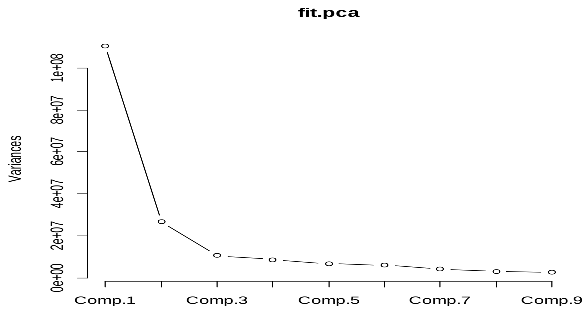

Figure 5: The Scree Plot.

CONCLUSION

The biases and the mean squared errors of the proposed estimator is lesser than the biases and the mean squared errors of the existing estimator by Subramani and Kumarapandiyan (2012). Hence, the proposed estimator is efficient

REFERENCES

1. Cochran, W. G. 1977. *Sampling Techniques (3rd Edition)*. John Wiley & Son, Inc.: New York, NY.

2. Deming, W.E. 1950. Sampling Techniques. John Wiley & Son, Inc.: New York, NY.

3. Kadilar, C. and H. Cingi. 2004. "Ratio Estimators in Simple Random Sampling". *Appl. Math. and Comput*. 151:893-902.

4. Kadilar, C. and H. Cingi.2006. "Ratio Estimators for the Population Variance in Sample and Stratified Random Sampling". *Appl. Math. and Comput*. 173:1043-1059.

5. Kadilar, C. and H. Cingi. 2006a. "An Improvement in Estimating the Population Mean by using the Correlation Co-efficient". *Hacettepe Journal of Mathematics and Statistics*. 35(1):103-109.

6. Kadilar, C. and H. Cingi. 2006b. "Improvement in Estimating the Population Mean in Simple Random Sampling". *Applied Mathematics Letters*. 19:75-79.

7. Koyuncu, N. and C. Kadilar. 2009. "Efficient Estimators for the Population Mean". *Hacettepe Journal of Mathematics and Statistics*. 38(2):217-225.

8. Lumley, T. 2004. *Analysis of Complex Survey Samples*. Department of Biostatistics. University of Washington Press.

9. Murthy, M.N. 1967. *Sampling Theory and Methods*. Statistical Publishing Society: Calcutta, India.

10. Prasad, B. 1989. "Some Improved Ratio Type Estimators of Population Mean and Ratio in Finite Population Sample Surveys". *Communications in Statistics: Theory and Methods*. 18:379–392.

11. Sen, A.R. 1993. "Some Early Developments in Ratio Estimation". *Biometric Journal*. 35(1):313

12. Singh, H.P. and M.R. Espejo. 2003. "On Linear Regression and Ratio-Product Estimation of a Finite Population Mean". *The Statistician*. 52:59-67.

13. Singh, H.P., R. Tailor, and M.S. Kakran. 2004. "An Improved Estimator of Population Mean using Power Transformation". *Journal of the Indian Society of Agricultural Statistics*. 58(2):223-230.

14. Singh, H.P. and R. Tailor. 2005. "Estimation of Finite Population Mean with known Co-efficient of Variation of an Auxiliary". *Statistica*. anno LXV(3):301-313.

15. Subramani, J. and G. Kumarapandiyan. 2012. "A Class of almost Unbiased Modified Ratio Estimators for Population Mean with Known Population Parameters". *Elixir Statistics*. 44:7411-7415.

16. Winship, C. and L. Radbill. 1994. "Sampling Weights and Regression Analysis". *Sociological Methods and Research*. 23(2):230 – 257.

17. Yan, Z. and B. Tian. 2010. "Ratio Method to the Mean Estimation Using Co-efficient of Skewness of Auxiliary Variable". *ICICA 2010, Part II*. CCIS 106:103–110

ABOUT THE AUTHORS

Dr. Olaniyi Mathew Olayiwola is a Lecturer in the Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria. He is a Chartered Statistician and Chartered Scientist. He is a member of Royal Statistical Society, London. He holds Doctor of Philosophy in Statistics from the University of Ibadan, Nigeria. His research interests are in sample surveys.

Dr. Fadeke Sola Apantaku is a Lecturer in the Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria. She is a member of Royal

Statistical Society, London. She holds Doctor of Philosophy in Statistics from the Federal University of Agriculture, Abeokuta, Nigeria. Her research interests are in sample surveys.

Dr. Adedayo Amos Adewara is a Lecturer in the Department of Statistics, Faculty of Science, University of Ilorin, Nigeria. He is a member of Nigerian Statistical Society. He holds Doctor of Philosophy in Statistics from the University of Ilorin, Nigeria. His research interests are in sample surveys.

Ms. Ayobami Fadilat Gboyega is a Postgraduate student in the Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria. Her research interests are in sample surveys.

SUGGESTED CITATION

Olayiwola, O.M., F.S. Apantaku, A.A. Adewara, and A.F. Gboyega. 2015. "Modified Ratio Estimator Using Multiple Auxiliary Variables". *Pacific Journal of Science and Technology*. 16(2):108-113.

Pacific Journal of Science and Technology