



INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY
A Generalized Class of Log-Type Estimators of Finite Population Mean
Based on Correlation Coefficient

Wasiu Babayemi ¹, Gerald Ike Onwuka ² and Aminu Bashir Isah ³

^{1,2,3} Faculty of Physical Science, Department of Mathematics, Statistics Unit
 Kebbi State University of Science and Technology Aleiro Kebbi State, Nigeria

Corresponding Author's Email: alaminbashiru@gmail.com

Received on: August, 2023

Revised and Accepted on: September, 2023

Published on: October, 2023

ABSTRACT

In sampling theory, it is a popular trend to use auxiliary information to obtain more efficient estimators for the population parameters to increase the precision of the estimator. Estimators obtained using auxiliary information are supposed to be more efficient than the estimators obtained without using auxiliary information. The ratio, regression, product and difference methods take advantage of the auxiliary information at the estimation stage. Therefore, this study considered a generalized class of log-type estimators of finite population mean based on correlation coefficient as the proposed estimator for estimating the population mean of the study variable. It has been shown that the generalized class of log-type estimators has lesser mean square errors (MSEs) under the optimum values of the characterizing scalar as compared to some of the commonly used related estimators available in the literature. Further, an extension of the proposed generalized class of log-type estimators using multiple auxiliary variables such as coefficient of variation (C_x), coefficient of kurtosis ($\beta_{2(x)}$), and correlation coefficient (ρ_{xy}) have also been initiated in this dissertation. The expressions for the properties of the proposed family of estimators, that is; Bias and Mean Square Error (MSE), were derived to the first degree of approximation. We also obtained the optimum Mean Square Error (MSE_{opt}), and theoretical comparisons were made with the related existing estimators in literature. Following theoretical comparisons, it was demonstrated that the proposed family of estimators was more efficient than various related existing estimators compared with, under the obtained conditions.

Keywords: Population, Sampling Unit, Sampling Frame, Sample Size, Statistic, Unbiased Estimator, Efficiency, Auxiliary variable.

1.0 INTRODUCTION

The use of auxiliary information to improve estimates of population parameters of the study variable is well-known in sample surveys. The ratio-type estimators are preferred when there is a highly positive linear relationship between the study and auxiliary variables described by straight line that passes through the origin. Regression-type estimators are preferred if the straight line does not pass through the origin. When there is a highly negative linear relationship between the study and auxiliary variables with the regression line passing through the origin, the product-type estimators are most preferable. Cochran (1940) introduced the customary ratio-type estimator, while Murthy (1964) proposed the customary product-type estimator. Srivastava (1967) discussed a power transformation estimator for which the customary ratio and product type estimators are special cases.

In sampling theory, it is a popular trend to use auxiliary information to obtain more efficient estimators for the population parameters to increase the precision of the estimator. Estimators obtained using auxiliary information is supposed to be more efficient than the estimators obtained without using auxiliary information. The ratio, regression, product and difference methods take advantage of the auxiliary information at the estimation stage. Many authors like, Pandey and Dubey (1988), Upadhyaya and Singh (1999), Kalidar and Cingi (2003), Singh and Tailor (2003), Singh (2003), Sisodia and Dwivedi (1981), along with many others have proposed various estimators using auxiliary information on various population parameters like coefficient of skewness, kurtosis, variation, standard deviation, correlation coefficient, etc. Sometimes, it is more economical to obtain information on more than one auxiliary

information; this would probably help in improvising the efficiency of the estimator, used to estimate the parameter under consideration. The literature deals with a wide range of ratio, product, difference and regression estimators proposed by various renowned authors using auxiliary information. Recently, Bhushan et al. (2015) had made the use of logarithmic relationship between

the study variable and auxiliary variable, extending the work, in this paper we have made the use of multiple auxiliary variables x'_s for estimating the population mean Y . The proposed estimators would work in case when the study variable is logarithmically related to the auxiliary variable.

3.0 METHODOLOGY

IMPROVED ESTIMATOR OF POPULATION MEAN

$$t_{\log} = k \bar{y} \left[1 + \log \left(\frac{a \bar{X} + b}{a \bar{x} + b} \right) \right]^{\frac{a \bar{X}}{a \bar{X} + b}} \quad i = 1, 2, 3, 4, 5, \text{ and } 6. \tag{1}$$

($a \neq 0$, b) are either constants or function of some known population parameters such as mean (\bar{X}), mean square (S_x^2), coefficient of variation (C_x^2), and coefficient of kurtosis ($\beta_2(x)$) of the auxiliary variable. It is important to mention that some new estimators have been found to be members of the proposed generalized class of estimators for different values of ϕ_i , which are given thus;

Table 1: Members of the generalized class of estimators for different values of ρ_{yx} , C_x , and $\beta_{2(x)}$ are the correlation coefficient between study and auxiliary variables, coefficient of variation and coefficient of kurtosis of the auxiliary variable respectively

New estimators	Values of ϕ_i
$t_{\log 1}^* = k \bar{y} \left[1 + \log \left(\frac{\bar{X} C_x + \rho_{yx}}{\bar{x} C_x + \rho_{yx}} \right) \right]^{\frac{\bar{X} C_x}{\bar{X} C_x + \rho_{yx}}}$	$\phi_1 = \frac{\bar{X} C_x}{\bar{X} C_x + \rho_{yx}}$
$t_{\log 2}^* = k \bar{y} \left[1 + \log \left(\frac{\bar{X} + \rho_{yx}}{\bar{x} + \rho_{yx}} \right) \right]^{\frac{\bar{X}}{\bar{X} + \rho_{yx}}}$	$\phi_2 = \frac{\bar{X}}{\bar{X} + \rho_{yx}}$
$t_{\log 3}^* = k \bar{y} \left[1 + \log \left(\frac{\bar{X} \beta_{2(x)} + \rho_{yx}}{\bar{x} \beta_{2(x)} + \rho_{yx}} \right) \right]^{\frac{\bar{X} \beta_{2(x)}}{\bar{X} \beta_{2(x)} + \rho_{yx}}}$	$\phi_3 = \frac{\bar{X} \beta_{2(x)}}{\bar{X} \beta_{2(x)} + \rho_{yx}}$
$t_{\log 4}^* = k \bar{y} \left[1 + \log \left(\frac{\bar{X} \beta_{2(x)} + \rho_{yx} C_x}{\bar{x} \beta_{2(x)} + \rho_{yx} C_x} \right) \right]^{\frac{\bar{X} \beta_{2(x)}}{\bar{X} \beta_{2(x)} + \rho_{yx} C_x}}$	$\phi_4 = \frac{\bar{X} \beta_{2(x)}}{\bar{X} \beta_{2(x)} + \rho_{yx} C_x}$
$t_{\log 5}^* = k \bar{y} \left[1 + \log \left(\frac{\bar{X} \rho_{yx} \beta_{2(x)} + C_x}{\bar{x} \rho_{yx} \beta_{2(x)} + C_x} \right) \right]^{\frac{\bar{X} \rho_{yx} \beta_{2(x)}}{\bar{X} \rho_{yx} \beta_{2(x)} + C_x}}$	$\phi_5 = \frac{\bar{X} \rho_{yx} \beta_{2(x)}}{\bar{X} \rho_{yx} \beta_{2(x)} + C_x}$

$t_{\log 6}^* = k \bar{y} \left[1 + \log \left(\frac{\bar{X} \beta_{2(x)} C_x + \rho_{yx}}{x \beta_{2(x)} C_x + \rho_{yx}} \right) \right]^{\frac{\bar{X} \beta_{2(x)} C_x}{\bar{X} \beta_{2(x)} C_x + \rho_{yx}}}$	$\phi_6 = \frac{\bar{X} \beta_{2(x)} C_x}{\bar{X} \beta_{2(x)} C_x + \rho_{yx}}$
---	--

Bias and Mean Square Error of the Proposed Estimator $t_{\log i}^*$

Expressing the above estimator in terms of e 's, we have;

$$\bar{y} = \bar{Y} (1 + e_0)$$

$$\bar{x} = \bar{X} (1 + e_1)$$

Then,

$$E(e_0) = E(e_1) = 0$$

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_Y^2$$

$$E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_X^2$$

$$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N} \right) \rho C_X C_Y$$

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log \left(\frac{a \bar{X} + b}{a \bar{X} (1 + e_1) + b} \right) \right]^{\frac{a \bar{X}}{a \bar{X} + b}} \quad (2)$$

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log \left(\frac{a \bar{X} + b}{a \bar{X} + b + a \bar{X} e_1} \right) \right]^{\frac{a \bar{X}}{a \bar{X} + b}}$$

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log \left(\frac{1}{1 + \frac{a \bar{X}}{a \bar{X} + b} e_1} \right) \right]^{\frac{a \bar{X}}{a \bar{X} + b}}$$

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log \left(\frac{1}{1 + \phi_i e_1} \right) \right]^{\phi_i}$$

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log (1 + \phi_i e_1)^{-1} \right]^{\phi_i}$$

Limiting the expansion to first order of approximation, we have;

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log (1 - \phi_i e_1 + \phi_i^2 e_1^2) \right]^{\phi_i}$$

$$t_{\log}^* = k \bar{Y} (1 + e_0) \left[1 + \log (1 + \{-\phi_i e_1 + \phi_i^2 e_1^2\}) \right]^{\phi_i}$$

By expanding logarithmically, we have;

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 + \left(\{-\phi_i e_1 + \phi_i^2 e_1^2\} - \frac{1}{2} \{-\phi_i e_1 + \phi_i^2 e_1^2\}^2 \right) \right]^{\phi_i}$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 + \left(\{-\phi_i e_1 + \phi_i^2 e_1^2\} - \frac{1}{2} \{\phi_i^2 e_1^2\} \right) \right]^{\phi_i}$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 - \phi_i e_1 + \phi_i^2 e_1^2 - \frac{1}{2} \phi_i^2 e_1^2 \right]^{\phi_i}$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 - \phi_i e_1 + \frac{1}{2} \phi_i^2 e_1^2 \right]^{\phi_i}$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 + \left\{ -\phi_i e_1 + \frac{1}{2} \phi_i^2 e_1^2 \right\} \right]^{\phi_i}$$

Limiting the power expansion to first order of approximation, we have;

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 + \phi_i \left\{ -\phi_i e_1 + \frac{1}{2} \phi_i^2 e_1^2 \right\} + \frac{\phi_i(\phi_i-1)}{2} \left\{ -\phi_i e_1 + \frac{1}{2} \phi_i^2 e_1^2 \right\}^2 + \dots \right]$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 + \phi_i \left\{ -\phi_i e_1 + \frac{1}{2} \phi_i^2 e_1^2 \right\} + \frac{\phi_i(\phi_i-1)}{2} \{\phi_i^2 e_1^2\} \right]$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 + \left\{ -\phi_i^2 e_1 + \frac{1}{2} \phi_i^3 e_1^2 \right\} + \frac{\phi_i^2 - \phi_i}{2} \{\phi_i^2 e_1^2\} \right]$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 - \phi_i^2 e_1 + \frac{\phi_i^3}{2} e_1^2 + \frac{\phi_i^4}{2} e_1^2 - \frac{\phi_i^3}{2} e_1^2 \right]$$

$$t_{\log}^* = k\bar{Y}(1+e_0) \left[1 - \phi_i^2 e_1 + \frac{\phi_i^4}{2} e_1^2 \right]_x$$

$$t_{\log}^* = k\bar{Y} \left[1 - \phi_i^2 e_1 + \frac{\phi_i^4}{2} e_1^2 + e_0 - \phi_i^2 e_0 e_1 \right]$$

$$t_{\log}^* = k\bar{Y} \left[1 + e_0 - \phi_i^2 e_1 - \phi_i^2 e_0 e_1 + \frac{\phi_i^4}{2} e_1^2 \right]$$

$$t_{\log}^* = \bar{Y} \left[k + k e_0 - k \phi_i^2 e_1 - k \phi_i^2 e_0 e_1 + k \frac{\phi_i^4}{2} e_1^2 \right]$$

$$t_{\log}^* - \bar{Y} = \bar{Y} \left[k + k e_0 - k \phi_i^2 e_1 - k \phi_i^2 e_0 e_1 + k \frac{\phi_i^4}{2} e_1^2 \right] - \bar{Y}$$

$$t_{\log}^* - \bar{Y} = \bar{Y} \left[k - 1 + k e_0 - k \phi_i^2 e_1 - k \phi_i^2 e_0 e_1 + k \frac{\phi_i^4}{2} e_1^2 \right]$$

Taking the expectation of both sides;

$$E(t_{\log}^* - \bar{Y}) = \bar{Y} E \left[k - 1 + k e_0 - k \phi_i^2 e_1 - k \phi_i^2 e_0 e_1 + k \frac{\phi_i^4}{2} e_1^2 \right]$$

$$E(e_0) = E(e_1) = 0$$

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_Y^2$$

$$E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_X^2$$

$$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho C_X C_Y$$

By substitution, we obtain the Bias of the proposed estimator as;

$$Bias(t_{\log}^*) = \bar{Y} \left[(k-1) + k \left(\frac{1}{n} - \frac{1}{N} \right) \left\{ \frac{\phi_i^4}{2} C_X^2 - \phi_i^2 \rho C_X C_Y \right\} \right]$$

(3)

Now the mean squared error (MSE) of the proposed estimator ($t_{\log i}^*$) can be derived thus;

$$(t_{\log}^* - \bar{Y})^2 = \bar{Y}^2 \left[k - 1 + k e_0 - k \phi_i^2 e_1 - k \phi_i^2 e_0 e_1 + k \frac{\phi_i^4}{2} e_1^2 \right]^2$$

Limiting the expansion to first order of approximation and taking the expectation of both sides, we have;

$$E(t_{\log}^* - \bar{Y})^2 = \bar{Y}^2 E \left[k - 1 + k e_0 - k \phi_i^2 e_1 - k \phi_i^2 e_0 e_1 + k \frac{\phi_i^4}{2} e_1^2 \right]^2$$

$$E(t_{\log}^* - \bar{Y})^2 = \bar{Y}^2 E \left[k^2 \{ e_0^2 + 2\phi_i^4 e_1^2 - 4\phi_i^2 e_0 e_1 + 1 \} - k \{ \phi_i^4 e_1^2 - 2\phi_i^2 e_0 e_1 + 2 \} \right]$$

$$E(t_{\log}^* - \bar{Y})^2 = \bar{Y}^2 E \left[(k-1)^2 + k^2 \{ e_0^2 + 2\phi_i^4 e_1^2 - 4\phi_i^2 e_0 e_1 \} - k \{ \phi_i^4 e_1^2 - 2\phi_i^2 e_0 e_1 \} \right]$$

By simplification and substitution, we have;

$$MSE(t_{\log}^*) = \bar{Y}^2 \left[\begin{array}{l} (k-1)^2 + k^2 \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \} \\ - k \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \} \end{array} \right]$$

To obtain the **optimum mean square error**, we differentiate the MSE with respect to k .

$$\frac{\partial MSE(t_{\log}^*)}{\partial k} = 0 \Rightarrow 2k - 2 + 2k \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \} - \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \} = 0$$

$$\Rightarrow k \left[2 + 2 \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \} \right] = 2 + \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \}$$

$$k = \frac{2 + \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \}}{2 + 2 \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \}}$$

$$k = \frac{1 + \frac{1}{2} \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \}}{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \}}$$

$$k = \frac{A}{B} \quad (5)$$

where;

$$A = 1 + \frac{1}{2} \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \}$$

$$B = 1 + \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \}$$

By substituting the values of k in (3.4), we have;

$$MSE_{opt.}(t_{log}^*) = \bar{Y}^2 \left[\left(\frac{A}{B} - 1 \right)^2 + \left(\frac{A}{B} \right)^2 \left(\frac{1}{n} - \frac{1}{N} \right) \{ C_Y^2 + 2\phi_i^4 C_X^2 - 4\phi_i^2 \rho C_X C_Y \} \right. \\ \left. - \left(\frac{A}{B} \right) \left(\frac{1}{n} - \frac{1}{N} \right) \{ \phi_i^4 C_X^2 - 2\phi_i^2 \rho C_X C_Y \} \right]$$

$$MSE_{opt.}(t_{log}^*) = \bar{Y}^2 \left[1 - \frac{A^2}{B} \right]$$

5.0 CONCLUSION

We have proposed a generalized class of log-type estimator for the population mean of study variable y , under the framework of simple random sampling scheme when information is available on single auxiliary variable. The properties of the proposed estimator are derived up to first order of approximation, and it is compared with other existing estimators, both theoretically and empirically. Numerical results are given in table 4.1. Results have shown that the proposed generalized class of log-type estimator t_{logi} is more efficient than the other existing estimators by smallest mean square error when empirical studies is carried-out using a real life, secondary data which is sourced from Kadilar and Cingi (2006a), that is population I, which is positively correlated in nature.

6.0 RECOMMENDATIONS

The suggested estimators are recommended for use in the estimation of population means of any variable of interest especially when the study and auxiliary variables are highly associated or correlated.

REFERENCES

- Bahl, S. & Tuteja, R. K. (1991). Ratio and Product type exponential estimator. *Information and Optimization sciences*, 12(1), 159-163.
- Bhushan, S., Gupta, R. & Pandey, S.K. (2015): Some log-type classes of estimators using auxiliary information. *International Journal of Computational and Applied Mathematics*, 12(1), 1-10.
- Cochran, W. (1940). The estimation of the yields of cereal experiments by sampling for the of grain to total produce. *The Journal of Agricultural Science*, 30(2), 262-275.
- Kadilar, C. & Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal*, 45(1), 218-225.
- Kadilar, C. & Cingi, H. (2006). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151(1), 893-902.
- Kadilar, C. & Cingi, H. (2006). "Improvement in estimating the population mean in simple random sampling". *Applied Mathematics Letters*, 19(1), 75-79.
- Murthy, M. N. (1964). Product method of estimation, *The Indian Journal of Statistics, Series A*, 26(1), 69-74.
- Pandey, B. N. & Dubey, V. (1988). Modified product estimator using coefficient of variation of auxiliary variate. *Assam Statistical Review*, 2(1), 64-66.

- Robson, D. S. (1958). Application of Multivariate Polykeys to the theory of Unbiased Ratio Type Estimators. *Journal of American Statistical Association*, 52(1), 511-522.
- Singh, H. P. (2003). On linear regression and ratio-product estimation of a finite population mean, *Journal of the Royal Statistical Society Series*. 52(1), 59-67.
- Singh, H. P. & Tailor, R. (2003). Estimation of finite population mean using known correlation coefficient between auxiliary characters, *Statistical*, 65(1), 407-418.
- Singh, H. P., Tailor, R., Tailor, R. & Kakran, M. (2004). An improved estimation of population mean using power transformation, *Journal Indian Society Agricultural Statistics*, 58(1), 223-230.
- Srivastava, S. K. (1967): "An Estimator Using Auxiliary Information in Sample Surveys," *Calcutta Statistical Association Bulletin*, 16(1), 121-32.
- Sisodia, B. V. S. & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal Indian Society Agricultural Statistics*, 33(1), 13-18.
- Tailor, R. (2009). A modified ratio-cum-product estimator of finite population mean in stratified random sampling, *Data Science Journal*, 8(1), 182-189 (on line).
- Tailor R. & Sharma, B. K. (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis, *Statistics in Transition*, 10(1), 15-24.
- Upadhyaya, L. N. & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*, 41(1), 627-636.