International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

Estimation of Finite Population Mean under Model Based Approach using Auxiliary Variables

Damaris Felistus Mulwa[1], Mutua Kilai[2]
[1,2]*Student, Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya*

***Abstract*: This paper proposes an estimator for estimating finite population mean in the presence of two auxiliary variables. For an estimator, two important properties are estimated which are the variance ad the unbiased property and in this paper the properties were investigated. The proposed estimator has been compared to the ratio-cum product estimator developed by [1]. The ratio regression estimator which has been proposed was found to performs well compared to the ratio-cum product estimator developed by [1] in datasets which have high correlation between the variables. Two datasets from the finance and motor manufacturing sections were used so as to fully explore the properties of the estimators. To compare the performance of the proposed estimator, mean squared error was used and the proposed estimator registered minimum squared error values in many of the datasets. The findings in this paper will benefit in understanding finance and motor manufacturing sector especially where the variables have a high correlation.**

***Keywords*: Auxiliary variable, Correlation, Predictive approach**

1. Introduction

The precision of an estimator in survey sampling is increased by the use of auxiliary information and it takes advantage of the correlation that is there between the auxiliary variables and study variables. Many authors have proposed estimators that take into consideration one and two auxiliary variables. [2] studied a class of estimator that are ratio based in the presence of two or more auxiliary variables. The presence of multivariate auxiliary variables enables the formation of more robust estimators by combining different estimators, [3]. The use of auxiliary information in a parametric super population model in the estimation stage has been done by several researchers including, [4]-[6].

In double sampling, auxiliary information has been used. A study by [7] found out that the estimator he had proposed performed better than the mean per unit estimator. He also compared the estimator to other estimators that did not take into consideration auxiliary variables and estimators that are not optimum asymptotically with the two auxiliary variables. Estimation of finite population mean and totals using local polynomial regression with two auxiliary variables has been investigated by [8]. In their study, they used a super population approach and a simulation was done. In both models, when the model was specified incorrectly, the local linear regression dominated the linear regression.

In this paper, a ratio regression type estimator is proposed which is used in estimating the finite population mean and derived its variance, Mean Squared Error and we compare the proposed estimator with the ratio-cum product estimator developed by [1]. The proposed estimator has been developed in relation to the motivation behind Cem Kadilar, Hulya Cingi [9], [10]. Two datasets from the finance and motor manufacturing sectors were used in order to explore the properties of the proposed estimator and they gave satisfactory results.

2. The proposed estimator

A. Some useful information

Auxiliary information can be used in two stages which include the sample stage and the estimation stage. The regression estimate of the population mean when there are two auxiliary variables W_1 and W_2, will be:

$$\overline{zp} = \bar{z} + t1(\overline{W_1} - \overline{w_1}) + t2(\overline{W_2} - \overline{w_2}) \quad (1)$$

Where $t_1 = \frac{s_{zw1}}{s_{w1}^2}$ and $t_2 = \frac{s_{zw2}}{s_{w2}^2}$, s_{zw1} and s_{zw2} are the sample covariance between the study variable and the auxiliary information.

The MSE of the estimator is given by:

$$MSE(\overline{zp}) = \frac{1-f}{n} s_z^2(1 - \rho_{zw1}^2 - \rho_{zw2}^2) + 2\rho_{zw1}\rho_{zw2}\rho_{w1w2} \quad (2)$$

Survey variables are often estimated by the auxiliary variables, a super population approach is used whereby a model which is working relating the two auxiliary variables is used.

B. Existing estimators using two auxiliary variables

An estimator for the population mean that relies on the assumption that the means of the two auxiliary variables are known was proposed by Abu-Dayyeh [11]. The proposed estimator is given by:

$$\bar{y} = \bar{y}\left(\frac{\overline{X_1}}{\overline{x_1}}\right)^{\alpha_1}\left(\frac{\overline{X_2}}{\overline{x_2}}\right)^{\alpha_2} \quad (3)$$

Where α_1 and α_2 are real numbers.
[9] proposed the estimator given in equation 4:

$$\bar{y} = \bar{y}\left(\frac{\overline{X_1}}{\overline{x_1}}\right)^{\alpha_1}\left(\frac{\overline{X_2}}{\overline{x_2}}\right)^{\alpha_2} + b_1(\overline{X_1} - \overline{x_1}) + b_2(\overline{X_2} - \overline{x_2}) \quad (4)$$

An exponential-ratio estimator proposed by [12] for estimating finite population mean is given by:

$$\overline{yBT} = \bar{y}_{exp}\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \quad (5)$$

Jinglu Lu [13] proposed an exponential ratio type estimator given by:

$$\overline{yclr} = \bar{y}_{exp}\left(\frac{\overline{Xlc}-\overline{xlc}}{\overline{Xlc}+\overline{xlc}}\right) \quad (6)$$

C. Proposed estimator

On the lines of [9] we propose an estimator which belongs to the exponential ratio regression class using two auxiliary variables for estimating population totals given by:

$$\hat{Y}_{pr} = \left[\bar{y}_{exp}\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right) + b_1(\overline{X_1} - \overline{x_1}) + b_2(\overline{X_2} - \overline{x_2})\right] \quad (7)$$

We now show that this estimator is unbiased

$$E\left[\hat{Y}_{pr}\right] = E\left[\left[\bar{y}_{exp}\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right) + b_1(\overline{X_1} - \overline{x_1}) + b_2(\overline{X_2} - \overline{x_2})\right]\right]$$

$$= E\left[\bar{y}_{exp}\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right) + b_1(\overline{X_1} - \overline{x_1}) + b_2(\overline{X_2} - \overline{x_2})\right]$$

$$= \bar{Y}_{exp}E\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right) + b_1E(\overline{X_1} - \overline{x_1}) + b_2E(\overline{X_2} - \overline{x_2}) = N\bar{Y}_{exp}(0) = \bar{Y} \quad (8)$$

In a similar manner, the variance of the estimator is derived as follows;

$$Var\left[\hat{Y}_{pr}\right] = E\left[\left[\bar{y}_{exp}\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right) + b_1(\overline{X_1} - \overline{x_1}) + b_2(\overline{X_2} - \overline{x_2})\right]\right]^2 - E\left[\hat{Y}_{pr}\right]^2$$

But,

$$E\left[\hat{Y}_{pr}\right] = \bar{Y} = E\left[\bar{y}^2 exp\left\{2\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right)\right\} + \left(b_1(\overline{X_1} - \overline{x_1})\right)^2 + \left(b_2(\overline{X_2} - \overline{x_2})\right)^2\right]$$

$$+ b_1(\overline{X_1} - \overline{x_1})b_2(\overline{X_2} - \overline{x_2}) = \bar{Y}_{exp}\left(E\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right)\right) + b_1(\overline{X_1} - \overline{x_1}) + b_2(\overline{X_2} - \overline{x_2})$$

$$= \bar{Y}_{exp}(0) = Y = E\left(\bar{y}^2 exp\left(2\left(\frac{\overline{X_1}}{\overline{x_1}} - \frac{\overline{x_1}}{\overline{X_1}}\right)\left(\frac{\overline{X_2}}{\overline{x_2}} - \frac{\overline{x_2}}{\overline{X_2}}\right)\right)\right) + \left(b_1^2\left(\frac{\sigma_1^2}{n} + \bar{X}_1^2\right) + b_1^2\left(\frac{\sigma_1^2}{n} + \bar{X}_1^2\right)\right)$$

$$\left(b_1^2\left(\frac{\sigma_1^2}{n} + \bar{X}_1^2\right) + b_1^2\left(\frac{\sigma_1^2}{n} + \bar{X}_1^2\right) - 2b_1^2\bar{X}_1^2 + b_2^2\left(\frac{\sigma_2^2}{n} + \bar{X}_2^2\right) + b_2^2\left(\frac{\sigma_1^2}{n} + \bar{X}_2^2\right)\right)$$

$$= \left(\frac{\sigma_y^2}{n} + \bar{Y}^2 + 2b_1^2\frac{\sigma_1^2}{n} + 2b_2^2\frac{\sigma_2^2}{n} - \bar{Y}^2\right) = \frac{\sigma_y^2}{n} + 2b_1^2\frac{\sigma_1^2}{n} + 2b_2^2\frac{\sigma_2^2}{n}$$

Thus

$$Var\left[\hat{Y}_{pr}\right] = \frac{1}{n}\left(\sigma_y^2 + 2b_1^2\sigma_{X1}^2 + 2b_2^2\sigma_{X2}^2\right) \quad (9)$$

3. Description of the datasets and studied variables

A. Datasets

In order to illustrate how the estimator given in equation 9 performs in relation to the ratio-cum product estimator proposed by Singh for finite population mean, two datasets were used and empirical results presented.

B. Description of variables

1) Dataset I: mtcars.

Y: Miles per gallon (mpg)
X_1: Rear axle ratio (drat)
X_2: Mile time (qsec)
$N = 32$
$n = 20$
$\bar{X}_1 = 3.597$
$\bar{X}_2 = 17.849$
$\bar{Y} = 20.091$
$S_{x1}^2 = 0.3158$
$S_{x2}^2 = 2.5412$
$S_y^2 = 41.724$
$\rho_{yx1} = 0.7814$
$\rho_{yx2} = 0.4498$
$\rho_{x1x2} = 0.2427$

2) Dataset II: The freeny data

Y: y
X_1: Lag quarterly revenue
X_2: Income level
$N = 39$
$n = 25$
$\bar{X}_1 = 9.281$

$\bar{X}_2 = 6.03$
$\bar{Y} = 9.306$
$S_{x1}^2 = 0.0424$
$S_{x2}^2 = 0.0096$
$S_y^2 = 0.0435$
$\rho_{yx1} = 0.9941$
$\rho_{yx2} = 0.9878$
$\rho_{x1x2} = 0.9864$

3) Graphical relationships between variables

In investigating the linearity assumption between the dependent variable and the auxiliary variables, dot plots were used. Figure 1 and 2, depicts a positive relationship between the auxiliary variables and the study variable. In literature reviewed, the positive relationship between auxiliary variables and study variable is of importance.

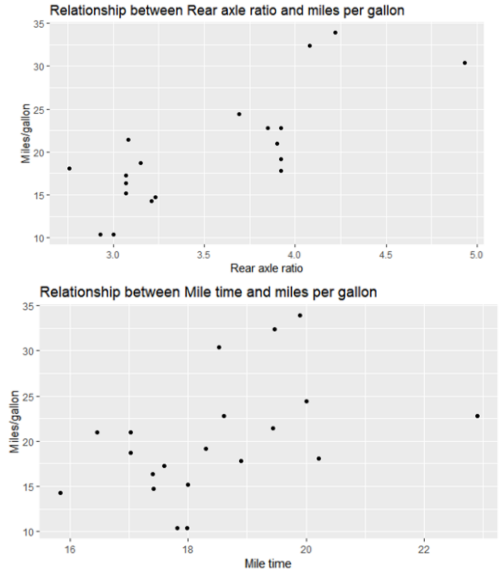

Fig. 1. Scatter plots for dataset 1

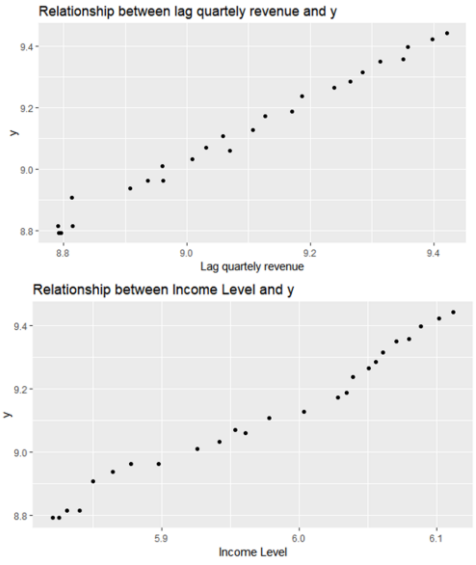

Fig. 2. Scatter plots for dataset 2

4. Results and discussions

For a study to utilize two auxiliary variables, they should be positively correlated. Figure 1 and 2 supports this claim sufficiently. The proposed estimator was used to compute the population mean for the two datasets and the population mean was also calculated using the ratio-cum estimator proposed by Singh. The performance of the proposed estimator and the ratio-cum product estimator proposed by Singh was compared using the Mean Squared Error of the two estimators. The estimator that produced the least MSE was deemed the best performing in the presence of two auxiliary variables.

Table 1 presents the Mean Squared Error of the proposed estimator and the estimator by [1]. The estimators were applied to the two datasets and results presented. In population 2 the variables were highly correlated and the proposed estimator produced the least mean squared estimator compared to the ratio-cum estimator proposed by Singh.

Table 1
MSE of the existing estimator and proposed estimator

Estimator	Population 1	Population 2
$\overline{Y_{singh}}$	2.688	1.44685
$\overline{Y_{pr}}$	4.535	0.00385

Table 2
Finite Population Mean estimates

Population	Estimator	Estimate
1	Singh's Estimator	20.7976
	Proposed estimator	19.8551
2	Singh's Estimator	9.02465
	Proposed Estimator	9.29730

In Table 2 the finite population mean estimates were calculated. The estimates from the proposed estimator were close to the Population mean.

5. Conclusion

In this paper, a ratio-regression type estimator using two auxiliary variables has been proposed and compared to the ratio-cum product estimator developed by Singh. In the presence of two auxiliary variables that are highly correlated the proposed estimator performs better than the ratio-cum product estimator developed by Singh. In data sets which have moderate correlation the ratio-cum product estimator by Singh performs better. The Mean Squared Error was used to compare the performance of the two estimators. In cases where the variables are highly correlated, we recommend the use of the proposed estimator as it gives more accurate estimates.

References

[1] S. M. P, "Ratio Cum Product Method of Estimation," Metrika, vol. 12, no. 1, pp. 34-42, 1967.
[2] L. Chand, "Some ratio type estimators based on two or more auxiliary variables," Unpublished PHD disertation, 1975.
[3] D. Robson, "Applications of multivariate polykays to the theory of unbiased ratio-type estimation," Journal of the American Statistical Association, vol. 52, no. 280, pp. 511-522, 1957.
[4] R. L. a. D. R. Chambers, "Estimating distribution functions from survey data," Biometrika, vol. 73, no. 3, pp. 597-604, 1986.

[5] A. H. a. H. P. Dorfman, Estimators of the finite population distribution function using nonparametric regression., The Annals of Statistics, 1993.
[6] J. K. J. a. M. H. Rao, "On estimating distribution functions and quantiles from survey data using auxiliary information.," Biometrika, pp. 365-375.
[7] K. M, A. O and I. A, "Use of auxiliary variables and asymptotically optimum estimators in double sampling," International Journal of Statistics and Probability, vol. 5, no. 3, 2016.
[8] R. E-H and Z. D, "Estimation of population tota; using local polynomial regression with two auxiliary variables," Journal of Statistics Application and Probability, vol. 3, no. 2, 2014.
[9] K. Cem and C. Hulya, "A new estimator using two auxiliary variables," Applied Mathematics and Computation, vol. 162, no. 2, pp. 901-908, 2005.
[10] M. Sachin and J. Singh, "An improved estimator using two auxiliary attributes," Applied Mathematical and Computation, vol. 219, no. 23, pp. 10983-10986, 2013.
[11] W. A. A. M. A. R. a. M. H. A. Abu-Dayyeh, "Some estimators of a finite population mean using auxiliary information.," Applied Mathematics and computation, vol. 139, no. 2, pp. 287-298, 2003.
[12] S. Bahl and R. Tuteja, "Ratio and Product Type exponential estimators," Journal of Information and Optimization Sciences, vol. 12, no. 1, pp. 159-164, 1991.
[13] J. Lu, Efficient estimator of a finite population mean using two auxiliary variables, biomedical, and power engineering, Mathematical problems in engineering, 2017.
[14] L. P. S, "Theorie analtytique des probabilities," Courcier, 1820.
[15] S. C and L. S, Estimation in surveys with nonresponse, John Wiley & Sons, 2005.
[16] M. H. e. a. Hansen, "Some history and reminiscences on survey sampling," Statistical science, vol. 2, no. 2, pp. 180-190, 1987.
[17] H. P. a. E. M. R. Singh, "Double sampling ratio-product estimator of a finite population mean in sample surveys.," Journal of Applied Statistics, vol. 34, no. 1, pp. 71-85, 2007.