



Modified Estimator of Finite Population Variance in Simple Random Sampling

J.O. Muili^{a}, A. Audu^b and R. V. K. Singh^b*

^{a,b}Department of Mathematics, Kebbi State University of Science and Technology, Aliero, Nigeria

^bDepartment of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria

Corresponding Author*: jamiunice@yahoo.com

ARTICLE INFORMATION

Article history:

Received 17 February 2019

Revised 2 March 2019

Accepted 14 March 2019

Available online 25 March 2019

Keywords:

unknown-weight, coefficient of kurtosis, median, auxiliary variable, median, mean square error

ABSTRACT

This paper deals with modification of Milton's estimator of population variance using coefficient of quartile deviation and unknown weight of auxiliary variable. The Bias and Mean Square Error (MSE) of proposed estimator are obtained up to the first order of approximation by Taylor's Series Method. The expression for the unknown weight is obtained and an empirical study was conducted to assess the performance of proposed estimator over some selected existing estimators. A numerical study is carried out to assess the efficiency of proposed estimator over the existing estimator with the aid of some known natural populations. Numerical study results shown that the proposed estimator performs better than the existing estimators.

1. Introduction

The use of auxiliary information, being constant with unit (population variance, population mean, population standard deviation, etc.) or unit free constant (Coefficients of variation, Kurtosis, Skewness etc.), can enhance the efficiency at the estimation stage. [1, 2,3], and [4] utilized this concept to improve the efficiency of ratio and product type estimators for estimating the population variance as well as population mean of study variable. To effectively estimate the population parameter of the variable of interest, there is need for the population values of the auxiliary variables. When auxiliary information is available researchers are able to utilize it in methods of estimation to obtain the most efficient estimator [5]. In many situations, information on the auxiliary is required either at the designing stage or estimation stage or both stages, to increase precision of the estimators. Ratio, Product and regression estimators are often used when advance knowledge of population variance of the auxiliary variable is readily available. In a class of estimators, an estimator with minimum variance or mean square error is regarded as the most efficient estimator. Estimation of the population variance of the study variable Y has received a considerable attention from experts engaged in survey statistics. For example, in agriculture the variation in production of crop is required for further planning or in manufacturing industries and pharmaceutical laboratories, the variation life of their products is a necessity for their quality control. Although, in literature, a great variety of techniques have been used mentioning the use of auxiliary information by means of ratio, product and regression methods for estimating population variance and other parameters [6].

Let $\Omega = (1, 2, 3, \dots, N)$ be a population of size N and Y, X be two real valued functions having values $(Y_i, X_i) \in \mathbb{R}^+ > 0$ on the i^{th} unit of $U(1 \leq i \leq N)$. Let S_y^2 and S_x^2 be the finite population

variance of Y and X respectively and S_y^2 and S_x^2 be respective sample variances based on the random sample of size n drawn without replacement.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i, \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad f = \frac{n}{N}, \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2, \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

The sample variance estimator of the finite population variance is defined as

$$t_1 = s_y^2 \tag{1}$$

which is an unbiased estimator of finite population variance S_y^2 and its variance is

$$Var(t) = \frac{1-f}{n} S_y^4 (\lambda_{40} - 1) \tag{2}$$

[7] proposed a ratio type variance estimator for the finite population variance S_y^2 when the finite population variance S_x^2 of auxiliary variable X is known. The bias and its mean squared error are given below:

$$t_2 = s_y^2 \frac{S_x^2}{s_x^2} \tag{3}$$

$$Bias(t_2) = \frac{1-f}{n} S_y^4 [(\lambda_{04} - 1) - (\lambda_{22} - 1)] \tag{4}$$

$$MSE(t_2) = \frac{1-f}{n} S_y^4 [(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)] \tag{5}$$

[8] proposed a ratio estimator for finite population variance by imposing Coefficient of kurtosis on the work of [4] as:

$$t_3 = s_y^2 \left(\frac{S_x^2 + \beta_{x(2)}}{s_x^2 + \beta_{x(2)}} \right) \tag{6}$$

$$Bias(t_2) = \frac{1-f}{n} A_2 S_y^2 [A_2 (\lambda_{04} - 1) - (\lambda_{22} - 1)] \tag{7}$$

$$MSE(t_3) = \frac{1-f}{n} S_y^4 [(\lambda_{40} - 1) + A_3^2 (\lambda_{04} - 1) - 2A_3 (\lambda_{22} - 1)] \tag{8}$$

Where $A_3 = \frac{S_x^2}{S_x^2 + \beta_{2(x)}}$

[9] proposed a class of ratio type estimators for finite population variance by imposing Coefficient of variation and Coefficient of kurtosis on the work of [7] as:

$$t_4 = s_y^2 \left(\frac{S_x^2 - C_x}{s_x^2 - C_x} \right) \tag{9}$$

$$t_5 = s_y^2 \left(\frac{S_x^2 - \beta_{x(2)}}{s_x^2 - \beta_{x(2)}} \right) \tag{10}$$

$$t_6 = s_y^2 \left(\frac{S_x^2 \beta_{x(2)} - C_x}{s_x^2 \beta_{x(2)} - C_x} \right) \tag{11}$$

$$t_7 = s_y^2 \left(\frac{S_x^2 C_x - \beta_{x(2)}}{s_x^2 C_x - \beta_{x(2)}} \right) \quad (12)$$

$$Bias(t_i) = \frac{1-f}{n} A_i S_y^2 [A_i (\lambda_{04} - 1) - (\lambda_{22} - 1)], \text{ where } i=4,5,6,7 \quad (13)$$

$$MSE(t_i) = \frac{1-f}{n} S_y^4 [(\lambda_{40} - 1) + A_i^2 (\lambda_{04} - 1) - 2A_i (\lambda_{22} - 1)], \text{ where } i=4,5,6,7 \quad (14)$$

$$\text{Where } A_4 = \frac{S_x^2}{S_x^2 - C_x}, A_5 = \frac{S_x^2}{S_x^2 - \beta_{2(x)}}, A_6 = \frac{S_x^2 \beta_{2(x)}}{S_x^2 \beta_{2(x)} - C_x}, A_7 = \frac{S_x^2 C_x}{S_x^2 C_x - \beta_{2(x)}}$$

[10] proposed ratio type estimators for finite population variance using quartiles and median of auxiliary variable as:

$$t_8 = s_y^2 \left(\frac{S_x^2 + M_d}{s_x^2 + M_d} \right) \quad (15)$$

$$t_9 = s_y^2 \left(\frac{S_x^2 + Q_1}{s_x^2 + Q_1} \right) \quad (16)$$

$$t_{10} = s_y^2 \left(\frac{S_x^2 + Q_3}{s_x^2 + Q_3} \right) \quad (17)$$

$$Bias(t_i) = \frac{1-f}{n} S_y^2 A_i [A_i (\lambda_{04} - 1) - (\lambda_{22} - 1)] \quad \text{where } i=8,9,10 \quad (18)$$

$$MSE(t_i) = \frac{1-f}{n} S_y^4 [(\lambda_{40} - 1) + A_i^2 (\lambda_{04} - 1) - 2A_i (\lambda_{22} - 1)] \quad (19)$$

$$A_i = \frac{S_x^2}{S_x^2 + M_d}, A_i = \frac{S_x^2}{S_x^2 + Q_1}, A_i = \frac{S_x^2}{S_x^2 + Q_3} \quad \text{where } i=8,9,10$$

$$t_{11} = s_y^2 \left(\frac{S_x^2 C_x + M_d}{s_x^2 C_x + M_d} \right) \quad (20)$$

$$Bias(t_{11}) = \frac{1-f}{n} S_y^2 A_{11} [A_{11} (\lambda_{04} - 1) - (\lambda_{22} - 1)] \quad (21)$$

$$MSE(t_{11}) = \frac{1-f}{n} S_y^4 [(\lambda_{40} - 1) + A_{11}^2 (\lambda_{04} - 1) - 2A_{11} (\lambda_{22} - 1)] \quad (22)$$

$$\text{Where } A_{11} = \frac{S_x^2}{S_x^2 + M_d}$$

[11] proposed a class of ratio type variance estimators utilizing different known parameters of auxiliary variable as:

$$t_{12} = s_y^2 \left(\frac{S_x^2 \beta_x + M_d^2}{s_x^2 \beta_x + M_d^2} \right) \quad (23)$$

$$Bias(t_{12}) = \frac{1-f}{n} A_{12} S_y^2 [A_{12} (\lambda_{04} - 1) - (\lambda_{22} - 1)] \quad (24)$$

$$MSE(t_{12}) = \frac{1-f}{n} S_y^4 [(\beta_{2(y)} - 1) + A_{12}^2 (\beta_{2(x)} - 1) - 2A_{12} (\lambda_{22} - 1)] \quad (25)$$

$$\text{Where } A_{12} = \frac{S_x^2}{S_x^2 \beta_x + M_d^2}$$

In this paper, modified estimator of population variance in simple random sampling has been proposed with objective to produce efficient estimator and its properties have been established.

2. Materials and Method

2.1 Proposed Estimator

Motivated by the work of [11], we proposed a ratio estimator of finite population variance by imposing unknown weight (k) and quartile deviation (Q_c) of auxiliary variable as:

$$\hat{S}_{MJ}^2 = kS_y^2 \left(\frac{S_x^2 Q_c^2 + M_d}{S_x^2 Q_c^2 + M_d} \right) \quad (26)$$

Where k is unknown weight to be determined such that the MSE of the proposed estimator \hat{S}_{MJ}^2 is minimized.

2.2 Properties of Proposed Estimator

In order to determine the bias and MSE of \hat{S}_{MJ}^2 , we define $e_0 = \frac{s_y^2 - S_y^2}{S_y^2}$ and $e_1 = \frac{s_x^2 - S_x^2}{S_x^2}$ such that

$s_y^2 = S_y^2(1 + e_0)$ and $s_x^2 = S_x^2(1 + e_1)$, from the definition of e_0 and e_1 , obtaining

$$\left. \begin{aligned} E(e_0) = E(e_1) = 0, E(e_0^2) &= \frac{1-f}{n} (\beta_{2(y)} - 1) \\ E(e_1^2) &= \frac{1-f}{n} (\beta_{2(x)} - 1), E(e_0 e_1) = \frac{1-f}{n} (\lambda_{22} - 1) \end{aligned} \right\} \quad (27)$$

Now expressing (26) in error terms as:

$$\hat{S}_{MJ}^2 = S_y^2 (1 + e_0) k \left(\frac{S_x^2 Q_c^2 + M_d}{S_x^2 Q_c^2 (1 + e_1) + M_d} \right) \quad (28)$$

Simplifying (28) up to first order approximation, it reduces to (28) as:

$$\hat{S}_{MJ}^2 = S_y^2 (k + ke_0 - kte_1 - kte_0 e_1 + kt^2 e_1^2) \quad (29)$$

Where $t = \frac{S_x^2 Q_c^2}{S_x^2 Q_c^2 + M_d}$

Subtracting S_y^2 from both sides of (29)

$$\hat{S}_{MJ}^2 - S_y^2 = S_y^2 (k + ke_0 - kte_1 - kte_0 e_1 + kt^2 e_1^2) - S_y^2 \quad (30)$$

Taking Expectation of both sides of (30)

$$E \left(\hat{S}_{MJ}^2 - S_y^2 \right) = S_y^2 E \left((k-1) + ke_0 - kte_1 - kte_0 e_1 + kt^2 e_1^2 \right) \quad (31)$$

Applying the results of (27), obtaining the *Bias* $\left(\hat{S}_{MJ}^2 \right)$ as:

$$\text{Bias} \left(\hat{S}_{MJ}^2 \right) = S_y^2 \left[(k-1) + \frac{1-f}{n} kt^2 (\lambda_{04} - 1) - \frac{1-f}{n} kt (\lambda_{22} - 1) \right] \quad (32)$$

Squaring, taking expectation (31) using the results in equation (27), and differentiating partially with respect to k_i and equate to zero, we obtain the $MSE\left(\hat{S}_{MJ}^2\right)_{\min}$ of the proposed estimator as:

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} = S_y^4 \left(1 - \frac{P^2}{Q}\right) \quad (33)$$

$$\text{Where } k = \frac{\frac{1-f}{n} \left[t^2 (\beta_{2(x)} - 1) - t (\lambda_{22} - 1) \right] + 1}{\frac{1-f}{n} \left[3t^2 (\beta_{2(x)} - 1) - 4t (\lambda_{22} - 1) + (\beta_{2(y)} - 1) \right] + 1} = \frac{P}{Q} \quad (34)$$

2.3 Efficiency Comparison

In this section efficiency of the proposed estimator is compared with efficiencies of some estimators in the literature.

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than sample variance if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < Var(t_1)$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} (\lambda_{40} - 1) \quad (35)$$

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than t_2 if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < MSE\left(\hat{S}_R^2\right)$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} \left[(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1) \right] \quad (36)$$

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than t_3 if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < MSE(t_3)$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} \left[(\lambda_{40} - 1) + A_3^2 (\lambda_{04} - 1) - 2A_3 (\lambda_{22} - 1) \right] \quad (37)$$

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than t_i if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < MSE(t_i) \quad i=4,5,6,7$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} \left[(\lambda_{40} - 1) + A_i^2 (\lambda_{04} - 1) - 2A_i (\lambda_{22} - 1) \right] \quad (38)$$

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than t_i if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < MSE(t_i) \quad i = 8,9,10$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} [(\lambda_{40} - 1) + A_i^2 (\lambda_{04} - 1) - 2A_i (\lambda_{22} - 1)] \quad (39)$$

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than t_{11} if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < MSE(t_{11})$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} [(\lambda_{40} - 1) + A_{11}^2 (\lambda_{04} - 1) - 2A_{11} (\lambda_{22} - 1)] \quad (40)$$

The \hat{S}_{MJ}^2 of estimator of the finite population variance is more efficient than t_{12} if,

$$MSE\left(\hat{S}_{MJ}^2\right)_{\min} < MSE(t_{12})$$

$$\left(1 - \frac{P^2}{Q}\right) < \frac{1-f}{n} [(\lambda_{40} - 1) + A_{12}^2 (\lambda_{04} - 1) - 2A_{12} (\lambda_{22} - 1)] \quad (41)$$

When conditions (35), (36), (37), (38), (39), (40), and (41) are satisfied, we can conclude that the proposed estimator is more efficient than some selected existing estimators.

3. Results and Discussion

3.1 Numerical Illustration

In order to investigate the merits of the proposed estimator of finite population variance over sample variance and some selected existing estimators under simple random sampling without replacement (SRSWOR), we have considered the following real populations.

Table 1: Populations I & II: [11]

Characteristics	Population I	Population II
N	80	70
n	20	25
\bar{Y}	51.826	96.700
\bar{X}	11.265	175.2671
ρ	0.9413	0.7293
S_y	18.357	60.7140
C_y	0.354	0.6254
S_x	8.456	140.8572
C_x	0.751	0.8037
$\beta_{2(y)}$	2.267	4.7596
$\beta_{2(x)}$	2.8664	7.0952
λ_{22}	2.221	4.6038
Md	10.300	72.4375
Q_1	5.150	80.1500
Q_3	16.975	225.0250

Table 2: Bias and MSE of the Reviewed and Proposed Estimators

Estimator	Population I BIAS (.)	MSE (.)	BIAS (.)	Population II MSE (.)
$t_1 = S_y^2$	0	5395.289	0	1313625.261
$t_2 = s_y^2 \frac{S_x^2}{s_x^2}$	8.151	3276.421	236.154	924946.481
$t_3 = s_y^2 \left(\frac{S_x^2 + \beta_{x(2)}}{s_x^2 + \beta_{x(2)}} \right)$	6.956	2740.349	235.656	924324.375
$t_4 = s_y^2 \left(\frac{S_x^2 - C_x}{s_x^2 - C_x} \right)$	8.512	3006.373	236.656	925017.011
$t_5 = s_y^2 \left(\frac{S_x^2 - \beta_{x(2)}}{s_x^2 - \beta_{x(2)}} \right)$	9.518	3186.399	236.445	925569.577
$t_6 = s_y^2 \left(\frac{S_x^2 \beta_{x(2)} - C_x}{s_x^2 \beta_{x(2)} - C_x} \right)$	8.279	2965.067	236.159	924956.421
$t_7 = s_y^2 \left(\frac{S_x^2 C_x - \beta_{x(2)}}{s_x^2 C_x - \beta_{x(2)}} \right)$	10.002	3275.722	236.517	925721.916
$t_8 = s_y^2 \left(\frac{S_x^2 + M_d}{s_x^2 + M_d} \right)$	4.530	2377.418	233.201	918641.426
$t_9 = s_y^2 \left(\frac{S_x^2 + Q_1}{s_x^2 + Q_1} \right)$	6.126	2609.91	232.889	917976.121
$t_{10} = s_y^2 \left(\frac{S_x^2 + Q_3}{s_x^2 + Q_3} \right)$	2.934	2181.488	227.099	905689.896
$t_{11} = s_y^2 \left(\frac{S_x^2 C_x + M_d}{s_x^2 C_x + M_d} \right)$	3.656	2314.033	232.485	917116.922
$t_{12} = s_y^2 \left(\frac{S_x^2 \beta_x + M_d^2}{s_x^2 \beta_x + M_d^2} \right)$	0.708	1993.270	207.653	865134.030
$\hat{S}_{MJ}^2 = ks_y^2 \left(\frac{S_x^2 Q_c^2 + M_d}{s_x^2 Q_c^2 + M_d} \right)$	-5.809	1957.662	-193.735	714144.085

Table 2 shows the results of the Bias and Mean Square Error (MSE) of the proposed and some existing estimators considered in the study for populations I and II.

Table 3: PRE of Estimators with respect to s_y^2

Estimator	Percentage Relative Efficiency (PRE)	
	Population I	Population II
t_1	100	100
t_2	164.67	142.02
t_3	196.88	142.12
t_4	179.46	142.01
t_5	169.32	141.93
t_6	181.96	142.02
t_7	164.71	141.90
t_8	226.94	143.00
t_9	206.72	143.10
t_{10}	247.32	145.04
t_{11}	233.16	143.23
t_{12}	270.68	151.84
Proposed Estimator \hat{S}_{MJ}^2	275.60	183.94

Table 3 shows the results of Percentage Relative Efficiency (PRE) of the proposed and some existing estimators considered in the study for populations I and II. The results revealed that the proposed estimator has the highest PRE among the estimators considered in the study.

We proposed an estimator of population mean. The performance of the proposed estimator over the usual ratio estimator and some selected existing estimators whereby two real populations were examined. The Bias and Mean Square Error (MSE) of the proposed estimator were derived. Table 2 shows the biases values of the existing and proposed estimators. The results revealed that the proposed estimator has minimum values of MSE in the populations considered in the study. Table 3 shows the results of Percentage Relative Efficiency (PRE) of the proposed and some related estimators considered in the study for populations I and II with difference ranging from 4.92%, 110.93%, and 32.10%, 42.04% respectively.

4. Conclusion

The results of this work have revealed that the proposed estimator, having minimum mean square error (MSE) (i.e. 1957.662 and 714144.085 for populations I and II respectively) and highest percentage relative efficiency (PRE) ranging from 4.92% to 110.93% efficient in population I and 32.10% to 42.04% efficient in population II among the selected existing estimators. With this fact, we conclude that the proposed estimator is proficient than some selected existing estimators and should be used in practical applications in estimating finite population variance.

Nomenclature

- N : Population size
 n : Sample size
 Y : Study variable
 X : Auxiliary variable
 \bar{y}, \bar{x} : Sample means of study and auxiliary variables
 \bar{Y}, \bar{X} : Population means of study and auxiliary variables
 ρ : Coefficient of correlation
 C_y, C_x : Coefficient of variations of study and auxiliary variables
 Q_3 : The upper quartile
 Q_1 : The lower quartile
 Q_c : Coefficient of quartile deviation
 $\beta_{2(y)}$: Coefficient of kurtosis of study variable
 $\beta_{2(x)}$: Coefficient of kurtosis of auxiliary variable
 M_d : Median of the auxiliary variable

5. Acknowledgement

The authors wish to acknowledge the anonymous reviewers for their valuable and the Editor in Chief for their recommendations.

6. Conflict of Interest

There is no conflict of interest associated with this work.

References

- [1] Gupta, S. and Shabbir, J. (2008). Variance Estimation in Simple Random Sampling using Auxiliary Information, Hacettepe *Journal of Mathematics and Statistics*. 37:57-67.
- [2] Subramani, J., and Kumarapandiyam, G. (2012). Variance Estimation Using Quartiles and their Functions of an Auxiliary Variable. *International Journal of Statistics and Applications*, 2(5):67-72.
- [3] Tailor, R. and Sharma, B. (2012). Modified Estimators of Population Variance in presence of Auxiliary Information. *Statistics in Transition-New series*, 13(1), 37-46.
- [4] Yadav, S. K. and Kadilar, C. (2013). Improved Class of Ratio and Product Estimators, *Applied Mathematics and Computation*, 219, 10726–10731.
- [5] Cochran, W. G. (1977). *Sampling Techniques*: Third Edition. John Wiley & Sons Limited.
- [6] Kazeem, A. A., and Olanrewaju, I. S (2013). On the Efficiency of Ratio Estimator Based on Linear Combination of Median, Coefficients of Skewness and Kurtosis. *American Journal of Mathematics and Statistics*. 3(3): 130-134.
- [7] Isaki, C.T. (1983). Variance Estimation using Auxiliary Information. *Journal of the American Statistical Association* 78:117–123.

- [8] Upadhyaya, L.N. and Singh, H. P. (1999). An Estimator for Population Variance that utilizes the Kurtosis of an Auxiliary Variable in Sample Surveys, *Vikram Mathematical Journal*, 19: 14-17.
- [9] Kadilar, C. and Cingi, H. (2007). Improvement in Variance Estimation in Simple Random Sampling, *Communications in Statistics – Theory and Methods* 36: 2075–2081.
- [10] Subramani, J., and Kumarapandiyam, G. (2012). Variance Estimation Using Median of the Auxiliary Variable. *International Journal of Probability and Statistics*, 1: 6-40.
- [11] Milton T.K., Odhiambo, R.O. and Orwa, G.O. (2017). Estimation of Population Variance Using the Coefficient of Kurtosis and Median of an Auxiliary Variable under Simple Random Sampling. *Open Journal of Statistics*, 7, 944-955.