International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 339-344 www.ijaem.net ISSN: 2395-5252

On Efficiency of A New Ratio Type Estimator for Double Sampling with Two Auxiliary Variables

Faweya Olanrewaju, Ajayi Esther, Akinyemi Oluwadare
Department of Statistics, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria

Date of Submission: 01-03-2023 Date of Acceptance: 10-03-2023

ABSTRACT
Among the well-known existing estimators and Simple random sampling without replacement (SRSWOR) type estimators, this study focused on the efficiency of a newly developed ratio estimator for double sampling with two auxiliary variables. The most efficient estimator was determined through an empirical comparison of the minimum variances, relative efficiency, and coefficient of variations.
Keywords: Double Sampling; Ratio-Type Estimator; Minimum Variance; Coefficient of Variation; Relative Efficiency;

I. INTRODUCTION

Obtaining more precise estimators of parameters of interest is the primary concern of statisticians in any statistical estimation procedure. It is also well understood that including more information in the estimation procedure results in a better estimator, provided the information is valid and correct.A variety of sampling procedures are used to obtain such auxiliary information, while the ratio, product, and regression methods are used to obtain an improved estimator of population parameters such as the mean, total, or ratio.Auxiliary information on a variable that is linearly related to the variable of interest is available in the ratio estimation technique and is used to estimate the population mean. The way auxiliary information is used is determined by the format in which it is available.In the absence of prior information about an auxiliary variable, double sampling is the appropriate technique for estimating those variables based on samples. In double sampling, the auxiliary variable is collected from a large preliminary sample using simple random sampling without replacement, while the variable of interest is collected from a second sample that is smaller in size than the preliminary sample size using simple random sampling without replacement.

In literature, many authors introduced ratio, product and regression-type estimators using a general linear transformation of the auxiliary variable. Srivastava et al. (1990) considered the problem of estimating the population of the study variable using double sampling with a generalized chain estimator. Other authors such as Bahl and Tuteja (1991), Armstrong and St-Jean (1993), Singh and Choudury (2012), Singh and Majhi (2014), Singh and Ahmed (2015), Singh and Espejo (2003), Samiuddin and Hanif (2007), Khare et al (2013), Khan and Shabbir (2013), Khare and Rehman (2015), Khan (2016) have studied the problem of estimating population mean using two auxiliary variables under double sampling scheme.

Cochran (1940) appears to be the first to use auxiliary information in ratio estimator when there is highly positive correlation between variable of interest and auxiliary variables. Hansen and Hurwitz (1943) were first to suggest the use of auxiliary information in selecting the population with varyingprobabilities. Robson (1957) gave the idea of product estimator when there is highly negative correlation.

The Simple Random Sampling (SRS) estimator for a sample of size n drawn from a population of size N is defined as:

$$\bar{y}_0 = \frac{1}{n}\sum y_i = \bar{y} \qquad (1.11)$$

The mean square error is written as:

$$MSE(\bar{y}_0) = \theta S_y^2 \qquad (1.12)$$

Cochran (1977) gave a classical ratio estimator based on a single auxiliary variable as given as:

$$\bar{y}_c = \hat{R}\overline{x'} \qquad (1.13)$$

The mean square error of $\bar{y}_C$ is given as:

$$MSE(\bar{y}_c) = [\theta_1 S_y^2 + (\theta_2 - \theta_1)(S_y^2 + \hat{R}^2 S_x^2 - 2RSxy (1.14)$$

Singh and Choudhury (2012) proposed an exponential chain ratio estimator in double sampling and defined it as:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 339-344 www.ijaem.net ISSN: 2395-5252

$$\bar{y}_{SC} = \bar{y}\exp\left[\frac{\bar{x}'\frac{\bar{Z}}{\bar{z}'} - \bar{x}}{\bar{x}'\frac{\bar{Z}}{\bar{z}'} + \bar{x}}\right] \quad (1.15)$$

$$MSE(\bar{y}_{SC}) = \bar{Y}^2\left[\theta_2 C_y^2 + \frac{1}{4}((\theta_2 - \theta_1)C_x^2 + \theta_1 C_z^2) - \right.$$
θ2−θ1CyxCx2−θ1CyzCz2 (1.16)

Singh and Majhi (2014) developed a chain type exponential estimator for population mean and the estimator is given as:

$$\bar{y}_{SM} = \bar{y}\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right)\left(\frac{\bar{Z}}{\bar{z}'}\right) \quad (1.17)$$

$$MSE(\bar{y}_{SM}) = \bar{Y}^2\left[\theta_1 C_z^2 + \frac{(\theta_2-\theta_1)C_x^2}{4} + \theta_2 C_y^2 - \right.$$
2θ1ρyzCyCZ−θ2−θ1ρyxCyCx (1.18)

Singh and Ahmed (2015) proposed a chain ratio-type exponential estimator as:

$$\bar{y}_{SA} = \bar{y}\exp\left(\frac{\sqrt{\bar{x}'\frac{\bar{Z}}{\bar{z}'}} - \sqrt{\bar{x}'}}{\sqrt{\bar{x}'\frac{\bar{Z}}{\bar{z}'}} + \sqrt{\bar{x}'}}\right) \quad (1.19)$$

Its corresponding Mean square error is given as:

$$MSE(\bar{y}_{SA}) = \bar{Y}^2\left[\theta_2 C_y^2 + \frac{1}{16}((\theta_2 - \theta_1)C_x^2 + \right.$$
θ1Cz2+12θ2−θ1ρyxCyCx+ θ1ρyzCyCz (1.21)

The proposed estimator proposedby Jide-Ashaolu et al. (2020) is given as :

$$\hat{\bar{Y}}_p = \bar{y}\left[\alpha\frac{\bar{x}'}{\bar{x}} + (1-\alpha)\frac{\bar{z}'}{\bar{z}}\right] (1.22)$$
$$= \bar{y}[\alpha\bar{x}'\bar{x}^{-1} + (1-\alpha)\bar{z}'\bar{z}^{-1}]$$

Where α is a real constant

$$\alpha = \frac{(\rho_{yz} C_y C_z - C_z^2)}{(C_x^2 + \rho_{yz} C_y C_z - \rho_{xy} C_x C_y - C_z^2)} \quad (1.23)$$

The Mean square error of the estimate is:

$$MSE\left(\hat{\bar{Y}}_p\right) = \bar{Y}^2\left[\theta_2 C_y^2 + (\theta_2 - \theta_1)\left(C_z^2 - \right.\right.$$
2ρyzCyCz −
(Cz2−ρxzCxCz+ρxyCxCy−ρyzCyCz
)2Cx2−2ρxzCxCz +Cz2 (1.24)

II. EMPIRICAL COMPARISON OF THE PROPOSED ESTIMATOR OVER WELL-KNOWN EXISTING ESTIMATORS

In this section, we compare the efficiency of the proposed estimator with well-known existing estimators of double sampling using empirical data. The secondary data which includes the systolic measure, diastolic measure and rate pressure product of some students. The double sampling used rate pressure product as the study variable (y), systolic measure as the first auxiliary variable (x) and diastolic measure as the second auxiliary variable (z). The double sampling is done with the first and second sample sizes drawn at three different levels.

The mathematical efficiency of the proposed ratio type estimator is given as:
The data obtained was analyzed using Microsoft Excel and Scilab 5.33 scientific packages.
The values of the population parameters are given in Table 3.11 below.

Table 3.11: Population Parameters

Term	Value
N	1500.00
$\bar{Y}$	206.65
$\bar{Y}^2$	42705.27
$\bar{X}$	104.43
$\bar{Z}$	61.20

Table 3.12: Summary of the first and second phase sample sizes at each level

Level	1	2	3
n'	1000	700	500
n	700	400	200
$k = n/n'$	0.70	0.57	0.40
A	0.9726	1.0119	1.0067

(A) Level one with n' =1000 and n=700
The values of the estimates are given below:

Table 3.13: Sample estimates at level one

Term	Value	Term	Value
$\bar{y}'$	205.77	C_y	0.1575
$\bar{x}'$	104.03	C_x	0.1542
$\bar{z}'$	60.70	C_z	0.1941
$\bar{y}$	211.94	ρ_{yx}	0.9991
$\bar{x}$	107.23	ρ_{yz}	0.66455
$\bar{z}$	61.97	ρ_{xz}	0.676652
S_y	33.3931	θ_2	0.0007619
S_x	16.54212	S_{yz}	266.5865
S_z	12.03026	β_{yx}	1.9565
θ_1	0.0003333	S_{xz}	134.4883
S_{yx}	535.3837		

(B) Level two with n' =700 and n=400
The values of the estimates are given below:

Table 3.14: Sample estimates at level two

Term	Value	Term	Value
$\bar{y}'$	212.16	C_y	0.1557
$\bar{x}'$	107.19	C_x	0.1569
$\bar{z}'$	61.97	C_z	0.1948
$\bar{y}$	210.11	ρ_{yx}	0.9991
$\bar{x}$	106.08	ρ_{yz}	0.675407
$\bar{z}$	61.78	ρ_{xz}	0.677183
S_y	32.72945	θ_2	0.0018333
S_x	16.64961	S_{yz}	265.4321
S_z	12.03749	β_{yx}	1.959193
θ_1	0.0007619	S_{xz}	135.3815
S_{yx}	543.107		

(C) Level three with n' =500 and n=200
The values of the estimates are given below:

Table 3.15: Sample estimates at level three

Term	Value	Term	Value
$\bar{y}'$	212.55	C_y	0.148755
$\bar{x}'$	107.42	C_x	0.149399
$\bar{z}'$	62.10	C_z	0.192265
$\bar{y}$	202.51	ρ_{yx}	0.9991
$\bar{x}$	102.07	ρ_{yz}	0.6798
$\bar{z}$	60.14	ρ_{xz}	0.6812
S_y	30.1238	θ_2	0.00433
S_x	15.248	S_{yz}	235.604
S_z	11.562	β_{yx}	1.9638
θ_1	0.0013	S_{xz}	119.519
S_{yx}	456.638		

The Estimates of the population mean for different estimators are shown below.

Table 3.16: Sample Mean of the estimators at all levels

Estimator	Mean at all levels		
	1 (k = 0.7)	2 (k = 0.57)	3 (k=0.4)
$\bar{y}_0$	211.94	210.11	202.51
$\bar{y}_C$	205.6152	212.30855	213.12456
$\bar{y}_{SC}$	209.61176	209.89015	206.45429
$\bar{y}_{SM}$	208.72785	211.19275	207.68244
$\bar{y}_{SA}$	212.37511	209.45427	201.87798
$\hat{\bar{y}}_p$	205.66936	212.32707	213.15146

The alternative estimator $\bar{y}_p$ of the population mean increases as the sampling fraction decreases.

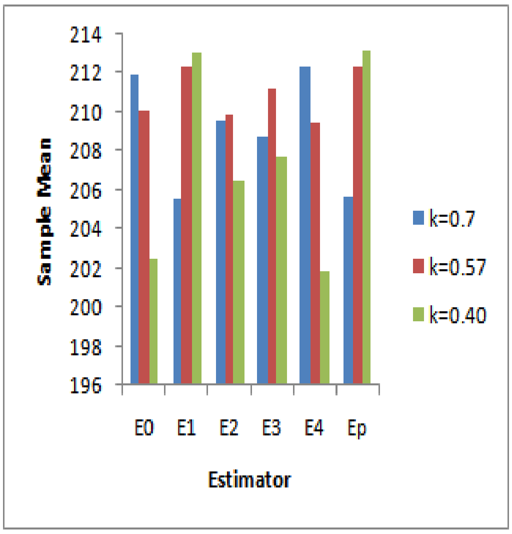

Figure 3.1: Sample Mean with respect to each estimator

Figure 3.1 above reveals that as k decreases, the sample mean obtained from estimators $\bar{y}_0$ and $\bar{y}_{SA}$ decreases
As k decreases, estimators $\bar{y}_C$ gets higher.
Estimators $\bar{y}_{SC}$ and $\bar{y}_{SM}$ are high at k=0.57 and minimum at k=0.4
The proposed estimator $\bar{y}_p$ gets higher as k decreases.
At level one, the newly developed estimator of the population mean is closest to the actual value of the population mean. This shows that the newly developed estimator is approximately unbiased.

The Estimates of the variance for different estimators are shown below:

Table 3.17: Sample Variance of the estimators at all levels

Estimator	Mean square error at all levels		
	1 (k = 0.7)	2 (k = 0.57)	3 (k=0.4)
$\bar{y}_0$	0.8495994	1.9639033	1.9639033
$\bar{y}_C$	0.400688	0.8239874	0.8239874
$\bar{y}_{SC}$	0.3169038	0.7056697	0.7056697
$\bar{y}_{SM}$	0.4298927	0.9650895	0.9650895
$\bar{y}_{SA}$	1.235367	2.939783	2.939783
$\hat{\bar{y}}_p$	0.3916279	0.7914921	0.7914921

At level one, the proposed estimator is more efficient over $\bar{y}_0$, $\bar{y}_C$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 339-344 www.ijaem.net ISSN: 2395-5252

At level two, the proposed estimator is more efficient over $\bar{y}_0$, $\bar{y}_C$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$
At level three, the proposed estimator is more efficient over $\bar{y}_0$, $\bar{y}_{SC}$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$

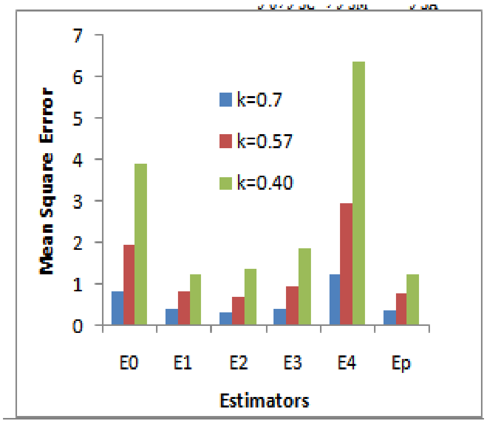

Figure 3.2: Mean Square error of the estimators at all levels

Figure 3.2 above reveals that as the sampling fraction decreases, the mean square error of each estimator increases.
All the estimators have a minimum variance when the sampling fraction is high.

The mean square error of the proposed estimator is higher than that of estimator $\bar{y}_{SC}$ at levels one and two. The mean square error of the proposed estimator is however lower than that of estimator $\bar{y}_{SC}$ at level three.

The mean square error of the proposed estimator is lower than that of estimator $\bar{y}_C$ at levels one and two. The mean square error of the proposed estimator is however higher than that of estimator $\bar{y}_C$ at level three.

The sample variances are minimum at level one when k=0.7. This implies that as the sampling fraction increase, minimum variances are obtained. At level one which is best, the mean square error of the proposed estimator $\hat{\bar{y}}_p$ is lower than that of estimators $\bar{y}_0$, $\bar{y}_C$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$

Table 3.18: Coefficient of Variation of the estimators at all levels

	Coefficient of variation at all levels		
Estimator	**1 (k = 0.7)**	**2 (k = 0.57)**	**3 (k=0.4)**
$\bar{y}_0$	0.4349	0.6670	0.9792
$\bar{y}_C$	0.3079	0.4276	0.5230
$\bar{y}_{SC}$	0.2686	0.4002	0.5699
$\bar{y}_{SM}$	0.3141	0.4652	0.6559
$\bar{y}_{SA}$	0.5234	0.8186	1.2513
$\hat{\bar{y}}_p$	0.3043	0.4190	0.5277

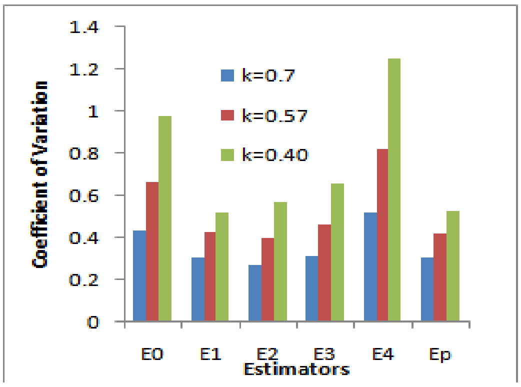

Figure 3.3: Coefficient of Variation of the estimators at all levels

Figure 3.3 above reveals that the coefficients of variation of the estimators are minimum at level one (k=0.7)
This implies that a higher precision is obtained when the sample fraction is high.
At level one which is best, the proposed estimator has a higher precision than estimators $\bar{y}_0$, $\bar{y}_C$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$.
At level two, the proposed estimator has a higher precision than estimators $\bar{y}_0$, $\bar{y}_C$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$.
At level three, the proposed estimator has a higher precision than estimators $\bar{y}_0$, $\bar{y}_{SC}$, $\bar{y}_{SM}$ and $\bar{y}_{SA}$.

Table 3.19: Relative Efficiency of the proposed estimator with respect to existing estimators at all levels

	Relative efficiency at all levels		
Estimator	**1 (k = 0.7)**	**2 (k = 0.57)**	**3 (k=0.4)**
$\bar{y}_0$	216.9405	248.1267	310.7893
$\bar{y}_C$	102.3134	104.1056	98.1901
$\bar{y}_{SC}$	80.9196	89.15688	109.4087
$\bar{y}_{SM}$	109.7707	121.9329	146.6754
$\bar{y}_{SA}$	315.4441	371.4229	504.3325
$\hat{\bar{y}}_p$	100	100	100

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 339-344 www.ijaem.net ISSN: 2395-5252

Table 3.19 above shows that at level one,
The proposed estimator $\hat{\bar{y}}_p$is approximately 217% more efficient over $\bar{y}_0$
The proposed estimator is approximately 102% more efficient over $\bar{y}_C$
The proposed estimator is approximately 19% less efficient over $\bar{y}_{SC}$
The proposed estimator $\hat{\bar{y}}_p$is approximately 110% more efficient over $\bar{y}_{SM}$
The proposed estimator $\hat{\bar{y}}_p$is approximately 315% more efficient over $\bar{y}_{SA}$

Table 3.19 above reveals that at level two:
The proposed estimator $\hat{\bar{y}}_p$is approximately 248% more efficient over $\bar{y}_0$
The proposed estimator is approximately 104% more efficient over $\bar{y}_C$
The proposed estimator is approximately 11% less efficient over $\bar{y}_{SC}$
The proposed estimator is approximately 122% more efficient over $\bar{y}_{SM}$
The proposed estimator $\hat{\bar{y}}_p$is approximately 371% more efficient over $\bar{y}_{SA}$

Table 3.19 above shows that at level three:
The proposed estimator $\hat{\bar{y}}_p$is approximately 311% efficient over $\bar{y}_0$
The proposed estimator is approximately 2% less efficient over $\bar{y}_C$
The proposed estimator is approximately 109% efficient over $\bar{y}_{SC}$
The proposed estimator is approximately 147% efficient over $\bar{y}_{SM}$
The proposed estimator $\hat{\bar{y}}_p$is approximately 504% efficient over $\bar{y}_{SA}$

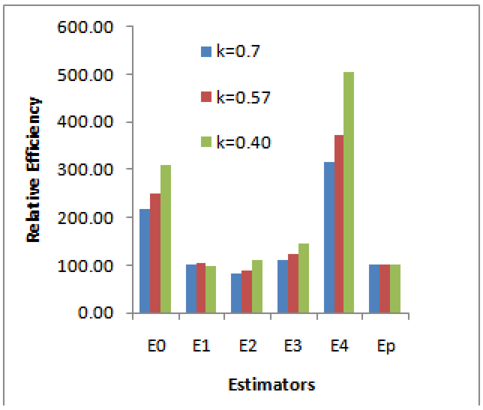

Figure 3.4: Relative efficiency (%) of the proposed estimator to other estimators

Figure 4.4 above reveals that as k decreases, the relative efficiency of the proposed estimator to the existing estimators gets extremely high. The relative efficiency of the newly developed estimator to the existing estimators is minimal at level one.

I. CONCLUSION

In this paper, we proposed a ratio type estimator for double sampling with two auxiliary variables. When comparing our proposed estimator to well-known estimators of double sampling from equations 3.1 to 3.5, it is clear that our proposed estimator is more precise than all other estimators mentioned in Section 2, implying that our estimator provides a more accurate estimate of population mean.

REFERENCES

[1]. Armstrong, J. and St-Jean, H. (1983). Generalized regression estimator for two-phase sample of tax records, SurveyMethodology, 20, 91-105.
[2]. Bahl, S., &Tuteja, R. K. (1991). Ratio and product type exponential estimators. Journal of information and optimization sciences, 12(1), 159-164
[3]. Khare BB, Srivastava U, Kumar K (2013) A generalized chain ratio in regression estimator for population mean using twoauxiliary characters in sample survey. J Sci Res Banaras Hindu Univ Varanasi 57:147–153
[4]. B. B. Khare and Habib R. (2015) .Modified Generalized Chain Ratio in Regression TypeEstimator for Population Mean using Two Phase Sampling in the Presence of Non-Response, Journal of Statistics Applications & Probability, 4, No. 2, 289-295
[5]. Cochran, W. G., (1940). Some properties of estimators based on sampling scheme with varying probabilities, Austral. J. Statist., 17, pp. 22–28.
[6]. Cohran, W. G. (1977). Sampling Techniques, 3rd Ed. New York: Wiley.
[7]. Hansen M.H and W. N. Hurwit (1943). On the Theory Of Sampling from Finite Populations," The Annals of Mathematical Statistics, Vol. 14, No. 4,, pp. 333-362.
[8]. Khan, M. and Shabbir, J. (2013). A Ratio Type Estimator for the Estimation of Population Variance Using Quartiles and its Functions of an Auxiliary Variable, Journal of Statistics Applications & Probability, 2(3), 157-162.
[9]. Khan M. (2016) A ratio chain-type exponential estimator for finite population

mean using double sampling, Khan SpringerPlus (2016) 5:86
[10]. Robson D. S., “Application of Multivariate Polykays, The Theory of Unbiased Ratio Type Estimators,” Journal of the American Statistical Association, Vol. 52, No. 280, 1957, pp. 511-522.
[11]. Samiuddin, M., &Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. Pakistan journal of statistics-all series-, 23(2), 99.
[12]. Singh B.K &SanjibChoudhury (2012) “Exponential Chain Ratio and Product Type Estimators for Finite Population Mean Under Double Sampling Scheme” Global Journal of Science Frontier Research. Vol 12, Issue 6, Version 1.0
[13]. Singh, G.N., &Majhi, D. (2014) “Some chain-type exponential estimators of population mean in two-phase sampling. Statistics in Transition, 15(2), 221-230
[14]. Singh, R.K & Ahmed, A (2015) “Improved exponential ratio and product type estimators for finite population mean under double sampling scheme. International Journal of Scientific and Engineering Research, 6(4), 509-514
[15]. Singh, H. P. and Espejo, M. R. (2003). On linear regression and ratio-product estimation of a finite population mean, The Statistician, 1, 59-67.
[16]. Srivastava, S., Rani, Khare, B. B. and Srivastava, S. R. (1990). “A generalized chain ratio estimator for mean of finite population”. Journal of the Indian Society of Agricultural Statistics, 42(1), 108-117.