African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 25-34)
www.abjournals.org

EFFECT OF VARYING CORRELATION VALUES ON THE EFFICIENCY OF SOME SELECTED POPULATION VARIANCE ESTIMATOR USING KNOWN VALUES OF AN AUXILIARY VARIABLE

Etaga Harrison O[1], Etaga Cecilia N[2] and Osuoha Chizoba S[1]

[1]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

[2]Department of Educational Foundation, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

ABSTRACT: *The use of Auxiliary variable in estimating the parameters of a study variable has been on the increase. Several authors have worked on some estimator combining various population values of the auxiliary variable and the study variable. There is need to compared some of these estimators under varying correlation values. Twelve estimators were considered. Two distributions were used to do the comparison. Three correlation levels were considered. The performance criteria used is the Mean Square Error (MSE). A sample of size 30 was simulated. The results showed that estimator T_4 is the best under Geometric distribution but the worst under the uniform distribution. Estimator T_{12} is the best under the Uniform while it took 7th in Geometric distribution. Estimators T_1, T_3, T_4, T_8, T_9, T_{10} and T_{12} under the Geometric distribution are not affected by correlation while T_2, T_5, T_6 and T_7 were affected by correlation. Under the Uniform distribution, only estimator T_3 and T_{11} had a little effect at high correlation. All other estimators are not affected by correlation. Almost all the estimators are affected by distribution except T_1.*

KEYWORDS: Correlation, Estimator, Auxiliary, Variable, MSE, Bias

INTRODUCTION

The estimation of population variance using auxiliary information have been widely researched on in the past. The use of auxiliary variable is either at the design stage or estimation stage. The paper takes a look at the use of auxiliary variable at the estimation stage. A lot of work has been done on this in the past by various author using various information of the auxiliary variable. The use of the auxiliary variable information in estimating the population parameter is for the purpose of obtaining a more efficient estimate. To use an auxiliary variable in estimating the population parameter, the two, auxiliary variable and the study variable must be correlated. Most of the earlier works have not consider the severity of the degree of correlation on the parameter estimates. There is need therefore the check this by varying the lever of correlation between the auxiliary variable and the study variable. Thus, in this paper, an attempt has been made to review the performances of some selected estimators at varying levels of correlation. According to Rajesh (2008), the use of auxiliary variable can increase the precision of an estimator when study variable, Y is highly correlated with auxiliary variable, X. He explained that there are situations when information is available in form of attribute which is highly correlated with Y; for instance, Sex and height of the persons, amount of milk produced and breed of the cow, amount of yield of wheat crop and variety of wheat. He proposed some

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 25-34)
www.abjournals.org

ratio estimators for estimating the population mean of the variable under study, which make use of the information regarding the population proportion possessing certain attributes.

Various authors like Isaki (1983), Singh et al (1988), Upadhyaya and Singh (1999), Kadilar and Cingi (2006), Solanki *et al.*, (2015) , Etaga *et al* (2019), Etaga *et al* (2015) and Singh and Kumar (2015) have paid their attention towards the enhanced estimation of population variance of the study variable Y. In the words of Cochran "The correlation between the auxiliary variable and the study variable will serve as an advantage to increase or speed up the precision of the estimation" (Cochran, 1977). Good auxiliary variables are those which are similar to the study variable and can be taken as good proxies of the study variable (Nasir *et al*, 2018). The appropriate use of auxiliary information in probability sampling designs yields considerable reduction in the variance of the estimators of population parameter. Much of the data that are statistically analyzed are collected in surveys. Use of auxiliary information can increase the precision of an estimator when the study variable Y is highly correlated with the auxiliary variables X. The sample variance is the most appropriate estimator of population variance. Tonui et al (2017) proposed some eight estimators, the performances of these estimators differ. There is need to further investigate these performances on varying level of correlation between the Auxiliary variable and the studied variable.

MATERIAL AND METHOD

This study is limited to the efficiency of twelve estimators of population variance proposed by (Tonui, *et al*., 2017) under two (2) distributions namely, Uniform, and, Geometric distributions.

The Ratio Estimators

1) $\hat{t}_1 = s_y^2$

2) $\hat{t}_2 = s_y^2 \frac{S_y^2}{s_x^2}$

3) $\hat{t}_3 = s_y^2 \left[\frac{S_x^2 + kx}{s_x^2 + kx}\right]$

4) $\hat{t}_4 = s_y^2 \left[\frac{S_x^2 - Cx}{s_x^2 - Cx}\right]$

5) $\hat{t}_5 = s_y^2 \left[\frac{S_x^2 - kx}{s_x^2 - kx}\right]$

6) $\hat{t}_6 = s_y^2 \left[\frac{S_x^2 kx - Cx}{s_x^2 kx - Cx}\right]$

7) $\hat{t}_7 = s_y^2 \left[\frac{S_x^2 Cx - Kx}{s_x^2 Cx - Kx}\right]$

8) $\hat{t}_8 = s_y^2 \left[\dfrac{S_x^2 + Mx}{s_x^2 + Mx}\right]$

9) $\hat{t}_9 = s_y^2 \left[\dfrac{S_x^2 + Q_1}{s_x^2 + Q_1}\right]$

10) $\hat{t}_{10} = s_y^2 \left[\dfrac{S_x^2 + Q_3}{s_x^2 + Q_3}\right]$

11) $\hat{t}_{11} \; = s_y^2 \left[\dfrac{CxS_x^2 + M_x}{Cxs_x^2 + M_x}\right]$

12) $\hat{t}_{12} = \; \hat{S}_{P_m}^2 \; = s_y^2 \left[\dfrac{S_x^2 kx + M_x^2}{s_x^2 kx + M_x^2}\right]$

Bias of the Estimators

1) $Bias\ (t_1) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \psi_1\ (\psi_1 - \frac{\lambda_{22}-1}{k_1-1})\}, where\ \psi_1 = 0$

2) $Bias\ (t_2) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \psi_2\ (\psi_2 - \frac{\lambda\ _{22}-1}{k_x-1})\}$

$$= \frac{1-f}{n} S_y^2 [k_X - 1) - (\lambda_{22} - 1)], \;\; where\ \psi_2 = 1$$

3) $Bias\ (t_3) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \psi_3\ (\psi_3 - \frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_3 = \frac{S_x^2}{S_x^2+K_x}$

4) $Bias\ (t_4) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_4(\psi_4 - \frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_4 = \frac{S_x^2}{S_x^2-C_x}$

5) $Bias\ (t_5) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_5(\psi_5 - \frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_5 = \frac{S_x^2}{S_x^2-k_x}$

6) $Bias\ (t_6) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_6(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_6 = \frac{S_x^2 K_x}{S_x^2 k_x - C_x}$

7) $Bias\ (t_7) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_7(\psi_7 - \left(\frac{\lambda_{22}-1}{k_x-1}\right))\}, where\ \psi_7 = \frac{S_x^2 C_x}{S_x^2 C_x - k_x}$

8) $Bias\ (t_8) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_8(\psi_8 - \left(\frac{\lambda_{22}-1}{k_x-1}\right))\}, where\ \psi_8 = \frac{S_x^2}{S_x^2+M_x}$

9) $Bias\ (t_9) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_9(\psi_9 - \left(\frac{\lambda_{22}-1}{k_x-1}\right))\}, where\ \psi_9 = \frac{S_x^2}{S_x^2+Q_1}$

10) $Bias\ (t_{10}) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_1(\psi_{10} - \left(\frac{\lambda_{22}-1}{k_x-1}\right))\}, where\ \psi_{10} = \frac{S_x^2}{S_x^2+Q_3}$

11) $Bias\ (t_{11}) \; = \; \frac{1-f}{n} S_y^2 \ \{(K_x - 1)\ \{\psi_{11}(\psi_{11} - [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_{11} = \; \frac{C_x S_x^2}{C_x S_x^2+M_x}$

12) $Bias\ (t_{12})\ =\ \hat{S}^2_{Pm} = \frac{1-f}{n} S^2_y\ \{(K_x - 1)\ , \ell^*\{\ell^* - \frac{\lambda_{22}-1}{k_x-1})\}$

MSE of the Estimators

1) $MSE\ (t_1)\ = Var\ (t_1) = \frac{1-f}{n} S^4_y\ \{(K_y - 1) +\ (k_x - 1)\ \psi_1[\psi_1 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})$

$= \frac{1-f}{n} S^4_y\ \{(K_y - 1), where\ \psi_1 = 0$

2) $MSE\ (t_2)\ = \frac{1-f}{n} S^4_y\ \{(K_y - 1) +\ (k_x - 1)\ \psi_2[\psi_2 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\} = \frac{1-f}{n} S^4_y\ \{(K_y - 1) + (k_x - 1)\text{-}2(\lambda_{22} - 1)],\quad where\ \psi_2 = 1$

3) MSE $(t_3)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \{k_x - 1\}\psi_3(\psi_3 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_3 = \frac{S^2_x}{S^2_x+K_x}$

4) MSE $(t_4)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_4\{k_x - 1\}[\psi_4 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_4 = \frac{S^2_x}{S^2_x-C_x}$

5) MSE $(t_5)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_5\{k_x - 1\}[\psi_5 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_5 = \frac{S^2_x}{S^2_x-k_x}$

6) MSE $(t_6)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_6\{k_x - 1\}[\psi_6 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_6 = \frac{S^2_x K_x}{S^2_x K_x-C_x}$

7) MSE $(t_7)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_7\{k_x - 1\}[\psi_7 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_7 = \frac{S^2_x C_x}{S^2_x C_x-k_x}$

8) MSE $(t_8)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_8\{k_x - 1\}[\psi_8 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_8 = \frac{S^2_x}{S^2_x+M_x}$

9) MSE $(t_9)\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_9\{k_x - 1\}[\psi_9 - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_9 = \frac{S^2_x}{S^2_x+Q_1}$

10) MSE $(t_{10})\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_{10}\{k_x - 1\}[\psi_{10} - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_{10} = \frac{S^2_x}{S^2_x+Q_3}$

11) MSE $(t_{11})\ = \frac{1-f}{n} S^4_y[\{k_y - 1\} + \psi_{11}\{k_x - 1\}[\psi_{11} - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\}, where\ \psi_{11} = \frac{C_x S^2_x}{C_x S^2_x+M_x}$

12) MSE$(\hat{S}^2_{Pm}) = \frac{1-f}{n} S^2_y\ \{(K_y - 1)\ , \ell^*(K_x - 1)\ [\ell^* - 2\ [(\frac{\lambda_{22}-1}{k_x-1})\},$

where

$S^2_y(s^2_y)\ =\ Population(sample)\ variance\ of\ Y,$

$S^2_x(s^2_x) =\ Population(sample)\ variance\ of\ X$

$\bar{Y}\ (\bar{y}) = Population(sample)\ Mean\ of\ Y.$

$\bar{X}\ (\bar{x}) = Population(sample)\ Mean\ of\ X.\ S^2_{xy}$ = Covariance between Y and X

Defining parameters as follows;

$\mu_{rs} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^r (y_i - \bar{y})^s}{N-1}$ $\lambda_{22} = \frac{\mu_{rs}}{\mu_{20}^{\frac{r}{s}} \mu_{02}^{\frac{r}{s}}}$, $\mu_{20} = S_y^2, \mu_{02} = S_x^2$ *and* $\mu_{11} = Sxy$;

$\lambda_{22} = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$, $\lambda_{21} = \frac{\mu_{21}}{\mu_{20}\mu_{02}^{\frac{1}{2}}}$ such that;

$C_y = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{20}}{\bar{Y}^2}$ is the coefficient of variation of the study variable y

$C_x = \frac{S_x^2}{\bar{X}^2} = \frac{\mu_{02}}{\bar{X}^2}$ is the coefficient of variation for the auxiliary variable x

$\ell xy = \frac{S_{xy}}{S_x S_y} = \frac{\mu_1}{\sqrt{\mu_{20}}\sqrt{\mu_{02}}}$, coefficient of correlation between x and y

$k(y) = \lambda_{40} = \frac{\mu_{40}}{\mu_{20}^2}$ coefficient of kurtosis for the study variable, $k(x) = \lambda_{04} = \frac{\mu_{40}}{\mu_{02}^2}$ Coefficient of kurtosis for the auxiliary variable

Data

The data for this work was simulated from two distributions. The Uniform and Geometric distributions. Three correlation values where chosen, 0.55, 0.75, and 0.95. The sample size use was 30.

RESULTS

Table 1: Estimates for the Twelve Estimators for Uniform Distribution for Various Correlation Values.

	0.55	0.75	0.95
Estimators	**Estimates**	**Estimates**	**Estimates**
T1	16.074	17.919	18.918
T2	23.63724	19.88065	15.92193
T3	22.49406	19.62767	16.19364
T4	-63.3626	28.8521	8.896491
T5	25.27572	20.22155	15.59001
T6	31.20339	20.92629	14.38766
T7	23.73856	19.90311	15.90125
T8	18.20188	18.56804	17.65529
T9	18.35564	18.6159	17.57183
T10	18.04473	18.45877	17.74209
T11	22.43162	19.61796	16.18023
T12	16.27492	17.98753	18.7674

Table 2: Bias for the Twelve Estimators for Uniform Distribution at various Correlation Values

Correlations	0.55	0.75	0.95
Estimators	**Bias**	**Bias**	**Bias**
T1	0	0	0
T2	1.159504	1.244124	1.081771
T3	0.972759	1.021859	0.912786
T4	11.84239	12.73677	23.61051
T5	1.419322	1.565377	1.314741
T6	2.25558	2.297025	2.456951
T7	1.175865	1.264546	1.095406
T8	0.275604	0.273828	0.275903
T9	0.298778	0.300228	0.301856
T10	-1.46164	-1.81422	-2.82993
T11	1.021986	1.076553	0.971878
T12	0.021056	0.020953	0.022696

Table 3: MSE for the Twelve Estimators for Uniform Distribution at Various Correlation Values.

MSE Uniform	Rank0.55	MSE	Rank0.75	MSE	Rank0.95	MSE
T1	2	8.358176	2	8.358176	2	8.358176
T2	8	38.11136	8	39.49294	8	36.49875
T3	7	33.80881	7	34.48553	6	32.55469
T4	12	254.1618	12	268.8312	12	480.1688
T5	10	43.98216	10	46.58313	10	41.83165
T6	11	62.27042	11	62.29111	11	66.8975
T7	9	38.48471	9	39.94838	9	36.81396
T8	4	16.60803	4	16.43284	4	16.65925
T9	5	17.23158	5	17.1293	5	17.36181
T10	3	15.97256	3	14.89152	3	15.9511
T11	6	32.73085	6	31.70396	7	33.37709
T12	1	7.911479	1	8.076474	1	7.753718

Table 4: Estimates for the Twelve Estimators for Geometric Distribution at Various Correlation Values

Correlations	0.55	0.75	0.95
Estimators	**Estimates**	**Estimates**	**Estimates**
T1	7.913	8.116	8.189
T2	9.145124	8.718074	7.451799
53T3	8.287879	8.365999	7.822825
T4	7.782003	8.044034	8.282259
T5	6.955432	6.641432	63.82603
T6	12.47158	43.11112	13.3932
T7	9.19013	8.731756	7.440738
T8	8.693766	8.513447	7.665816
T9	8.821647	8.572343	7.615596
T10	8.581656	8.446593	7.744253
T11	9.134244	8.713222	7.456172
T12	8.988015	8.622702	7.556289

Table 5: Bias for the Twelve Estimators for Geometric Distribution at Various Correlation Values

Correlations	0.55	0.75	0.95
Estimators	**Bias**	**Bias**	**Bias**
T1	0	0	0
T2	2.832623	1.882368	1.531091
T3	0.359908	0.395988	0.385623
T4	0.018605	0.005632	-0.00196
T5	2.581188	17.05202	104.0108
T6	20.00166	237.6411	20.29312
T7	3.007692	1.958902	1.579185
T8	1.299096	0.899046	0.765471
T9	1.688391	1.149823	0.918139
T10	0.471897	-0.01219	-0.2225
T11	2.79223	1.856871	1.513438
T12	2.250504	1.385408	1.119193

Table 6: MSE for the Twelve Estimators for Geometric Distribution at Various Correlation Values

MSE Uniform	Rank0.55	MSE	Rank0.75	MSE	Rank0.95	MSE
T1	2	11.90516	2	11.90516	2	11.90516
T2	10	31.79113	9	25.4914	9	23.13818
T3	3	14.68265	3	15.02342	3	14.98526
T4	1	11.8777	1	11.79071	1	11.76317
T5	8	27.70243	11	120.6183	12	687.9685
T6	12	147.0971	12	1593.14	11	141.3047
T7	11	32.98369	10	26.02093	10	23.47413
T8	5	21.27521	5	18.62627	5	17.73649
T9	6	23.96071	6	20.39138	6	18.82423
T10	4	19.14231	4	16.84362	4	16.24286
T11	9	29.30201	8	23.25729	8	21.08736
T12	7	27.70133	7	21.82439	7	19.99478

Table 7: Summary of Ranking

GEOMETRIC				UNIFORM			
	R = 0.55	R = 0.75	R = 0.95	MSE Uniform	Rank 0.55	Rank 0.75	Rank 0.95
T1	2	2	2	T1	2	2	2
T2	10	9	9	T2	8	8	8
T3	3	3	3	T3	7	7	6
T4	1	1	1	T4	12	12	12
T5	8	11	12	T5	10	10	10
T6	12	12	11	T6	11	11	11
T7	11	10	10	T7	9	9	9
T8	5	5	5	T8	4	4	4
T9	6	6	6	T9	5	5	5
T10	4	4	4	T10	3	3	3
T11	9	8	8	T11	6	6	7
T12	7	7	7	T12	1	1	1

CONCLUSION

Based on the ranks of the estimators for various correlation values, it can be concluded that

- Estimator T_4 is the best under Geometric distribution but the worst under the uniform distribution.
- Estimator T_{12} is the best under the Uniform while it took 7^{th} in Geometric.
- Estimators T_1, T_3, T_4, T_8, T_9, T_{10} and T_{12} under the Geometric distribution are not affected by correlation while T_2, T_5, T_6 and T_7 were affected by correlation.
- Under the Uniform distribution, only estimator T_3 and T_{11} had a little effect at high correlation. All other estimators are not affected by correlation.
- Almost all the estimators are affected by distribution except T_1.

Future Research

More work can be done on other classes of estimators not considered here. Other level of correlations can be tried while more distribution could be introduced to check for their effects. The sample sizes can also be varied.

REFERENCES

Cochran, W.G. (1977). Sampling Techniques, *3rd edition. Wiley, New York.*

Etaga H. O. , Nwosu, C. R and Udechukwu C. (2015**)** A comparative analysis of some almost unbiased modified ratio estimators for population mean with known population parameters of the auxiliary variable. Journal of Mathematics and Applied Statistics (IJSAR-JMAS) Vol 2, No1. Pp 34-47 ISSN 2408-7637

Etaga Harrison. O, Etaga Cecilia. C, Eriobu Nkiru. O and Ezeiruaku Perpetua C. (2019) Comparative Analysis of Some Selected Classes of Ratio Estimators, European Journal of Statistics and Probability, Vol.7, No.1, pp, 1-41, ECRTD-UK

Isaki, C.T. (1983). Variance Estimation using Auxiliary Information, *Journal of the American Statistical Association, 78:381, 117-123,* Doi: 10.1080/01621459.10477939

Kadilar, C. and Cingi, H. (2006). New ratio Estimators using Correlation Coefficient. Interstat. http://interstat.stat journals.net/year/2006/articles/060300472

Nasir, A., Abid, M., Tahir, M., and Zawar, M. (2018). Enhancing ratio estimators for estimating population mean using maxi mum value of auxiliary variable. *http://dox.doi.org/10.4038/jnsfsr.v46i3.8498*

Singh, H.P., and Kumar, M. (2015). Improved ratio type estimators of finite population variance using quartiles. *Mathematics and Statistics, 44(3):747-754.*

Singh, H.P.,Upadhyaya, L.N., and Namjashi, U.D. (1988).Variance Estimation in Simple Random Sampling using Auxiliary Information. Dept of Mathematics and Stat. College of Agric. Eng. Juwahrial Nehrukrishi Vishwa Vidyalaya Jabal per 482004 India. Vol.578 No.24

Solanki, S.R., Singh, H.P., and Surya, K.P. (2015). Improved ratio-type estimators of finite population variance. *Hacettepe J. Math. Stat 44 (3), 747-754.*

Tonui, M.K. Odhiambo, R.O., and Orwa, G.O. (2017). Estimation of Population Variance Using the Coefficient of Kurtosis and Median of Auxiliary Variable under Simple Random Sampling http://doi.org/10.4236/ojs. 2017.76066.

Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed Auxiliary Variable in estimating the finite population mean. *Biometrical Journal 41 (5), 627-636*