**Efficient Regression-in-ratio Estimators in Simple Random Sampling with Empirical High and Low Extreme Maximum values**
Peter I. Ogunyinka1, Emmanuel F. Ologunleko2 and A. A. Sodipo2*1: Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Nigeria.
ogunyinka.peter@gmail.com : +234805-200-3783:* [*http://orcid.org/0000-0002-4811-7203*](http://orcid.org/0000-0002-4811-7203) *2: Department of Statistics, University of Ibadan, Ibadan, Nigeria.
sodipo.ademola@gmail.com: +234805-516-0345*

# AbstractSignificant improvement had been made to regression-in-ratio estimators in simple random sampling. However, such estimators would be over-estimated or under-estimated in the presence of extreme maximum or minimum value in the survey data, respectively. This study had proposed three regression-in-ratio estimators that minimized the effect of over-estimation or under-estimation in the estimates when there are extreme values in the survey data. The bias and the Mean Square Error (MSE) for the proposed estimator were developed. Theoretical comparison of the proposed estimators confirmed the conditional efficiency of the proposed estimators to the reviewed estimators. Empirical comparison, with twenty-six simulated populations comprising of high and low extreme maximum values, was used to ascertain the asymptotic characteristics of the proposed estimators to different magnitudes of extreme value. Two estimators, out of the three proposed estimators, proved to be more biased than the corresponding reviewed estimators while the third proposed estimator proved to be less bias than the corresponding reviewed estimator. The proposed estimators proved to be asymptotically efficient with smaller variances and Mean Square Errors (MSEs) over the reviewed estimators. Finally, the results of the percentage relative efficiency showed that the three proposed estimators were , and , respectively efficient over the corresponding reviewed estimators. Future study was proposed to test the proposed estimators when there exists high and low extreme minimum values in the data.

Keywords/Phrase: Regression-in-ratio estimators, maximum value, minimum value, simple random sampling, efficiency

**2010 AMS Classification**: 62D05

## 1.0 Introduction

Auxiliary information has proved significant in the estimation of population parameters in sample survey theory. Simple random sampling estimators maximize on the advantages of auxiliary information. Similarly, ratio, regression and product estimators are efficient over the conventional Simple Random Sampling Without Replacement (SRSWOR) estimator when there is high correlation between the study and auxiliary variables (Cochran, 1940; Robson, 1957 and Murthy, 1967). Mixed estimator combines two or more of ratio, regression or product estimation methods into one estimator (Mohanty, 1967 and Kiregrera, 1984). Ratio-in-regression, regression-in-ratio estimators and ratio-cum-regression estimators are few examples of mixed estimators. Mixed estimator has proved efficient over simple estimators.

There have been series of improvement proposed by many authors who have worked on the conventional ratio and regression estimators in estimating the population total or mean. Similar improvements on ratio estimator have been proposed by Kadilar and Cingi (2004, 2005 and 2006), Yan and Tian (2010), Subramani and Kumarapandiyan (2012a, 2012b, 2012c and 2012d), Jeelani *et al.* (2013) and Abbas *et al.* (2018). The authors have used the known values of the parameters like coefficient of variation, skewness, kurtosis, median, deciles of the auxiliary variables and the correlation coefficient between the study and auxiliary variables to improve on the estimation of the study variable. The recent work of Abbas *et al.* (2018) has proposed improvement on the ratio estimator by using the known maximum value of the auxiliary variable in the linear combination with the mean of the auxiliary variable to produce a uniformly smaller variance unbiased estimator.

Abbas *et al.* (2018) argues in the direction of Sarndal (1972) that extreme values (either maximum or minimum value) will cause over or under estimation, respectively. However, the work of Abbas *et al.* (2018) did not primarily, focused on the correction of this extreme value effect on the estimator since the methodology and conclusion of the study did not justify the aim of Sarndal (1972). Sarndal (1972) had argued in the direction of Godambe (1966 and 1969) to describe the uniqueness of Sample Survey (survey inference) Theory to General Statistical (Statistical Inference) Theory. Let denotes the ordered population units while denotes the ordered sample value obtained from the population.

The conventional sample mean using SRSWOR as shown in equation (1) is a Uniformly Minimum Variance Unbiased Estimator (UMVUE) of the population mean . However, if a priori information has confirmed that (and in the sample) is exceedingly large (maximum value), then using equation (1) will yield over-estimated population mean. Similarly, if the priori knowledge confirms that (and in the sample) is exceedingly small (minimum value), then using equation (1) will yield under-estimated population mean. These maximum and minimum values are called extreme values.

The effect of any of these extreme values could be destructive and disastrous in application. For instance, in the hypertension community of the medicine and surgery profession, it is established that hypertension is significantly correlated with cardiovascular risk (Currie and Delles, 2018). Similarly, in the ophthalmology community of the same profession, Bulpitt *et al.* (1975) and Chung *et al.* (2015) have reported significant relationship between high Intraocular Pressure (IOP) and glaucoma development. Both pre-hypertension and pre-high-IOP are measured with High Blood Pressure (HBP) at the threshold of 120/80 mmHg (mmHg means millimeters in mercury) (Currie and Delles, 2018). Back to the illustration of extreme value, if in a medical outreach, the average result of the Basic Medical Signs (BMS) test for ten sampled elderly patients (115/78, 195/125, 117/78, 115/76, 110/75, 110/75, 115/77, 115/77, 115/78 and 115/77) mmHg gives an average BP of 122/82 mmHg. The average systolic (numerator) is above 120 mmHg and the average diastolic (denominator) is above 80 mmHg. The doctor will conclude that an average examined patient has high tendency of cardiovascular and/or glaucoma risk. This conclusion is aided by the extreme observed BP of 195/125 mmHg in the sample data. The argument remains that while the presence of the extreme value is justified, the general conclusion is not justified. On the other hand, if the extreme maximum BP of 195/125 mmHg is removed from the sample BP analysis, the average BP will be 114/77 mmHg. This means that an average examined patient has no tendency of cardiovascular and/or glaucoma risk considering the normal threshold of less than 120/80 mmHg. Unfortunately, the conclusion will be misleading if such extreme BP value is removed from the analysis. In fact, such removal is not a good practice as it neglects the importance of randomness or probability. It is very important to note that when large sample size is considered, a reduced but non-negligible effect of extreme values(s) will be observed in the corresponding estimate.

Extreme value is referred to as outlier in the general statistical theory. Extreme value or outlier will introduce either over-estimation or under-estimation to any estimate. While there are statistical measures to detect and control the effect of outlier in general statistical theory, Godambe (1966 and 1969) and Sarndal (1972) have confirmed that such measures are inapplicable in Sample Survey theory because the Guass-Markov technique is basically inapplicable to sample survey theory. Godambe (1955) and Godambe and Joshi (1965) refer to this inapplicability as “Non-existence theorems”. Sarndal (1972) has proffered a unique solution to the correction of the extreme value effect in SRS in sample survey theory.



Figure 1: Diagrammatic explanation of the effect of extreme maximum and minimum values in sample survey.

Figure 1 gives the diagrammatic illustration on the challenge of extreme values and the solution proposed by Sarndal (1972). In figure 1, is the actual sample mean. However, due to extreme maximum or minimum values, was over-estimated as or under-estimated as , respectively. Sarndal (1972) had proposed the constant to be subtracted from or added to to obtain or , respectively. and are the new estimators, according to Sarndal (1972) which are closed to the supposed true estimator with significant reduction of the effect of the extreme values. Sarndal (1972) emphases that the new improved estimators and will remain unbiased just like the unbiased estimator but the variance will increase, which in general, will increase the MSE. This indicates that either under-estimation or over-estimation increases the variance rather than the biasness of the concerned estimators.

Khan and Shabbir (2013) seems to be the first study that applied the method of Sarndal (1972) to ratio, regression and product estimators. Few authors have used the proposed method of Sarndal (1972) to correct for the effect of extreme value in both the study variable and one auxiliary variable. Al-Hossain and Khan (2014) minimized the extreme value effects in ratio, regression and product estimators with two auxiliary variables. Finally, Khan *et al.* (2015) improved on the ratio-type estimators with extreme value effect. This study aims to improve on the recent work of
Abbas *et al.* (2018) ratio estimators by correcting the effect of the extreme values in the estimators using the method of Sarndal (1972). Similarly, this study will ascertain which of bias and variance is at the receiving end of the effect of extreme value in survey inference.

## 2.0 Methodology

### 2.1 Reviewed Ratio estimators

Abbas *et al.* (2018) has improved on the ratio estimators developed by Subramani and Kumarapandiyan (2012a) by replacing the known median value of the auxiliary variable with the known maximum value of the auxiliary variable. The ratio estimators are presented as

Where is the maximum value of the auxiliary variable. The general bias and the Mean Square Errors (MSEs) were presented as

and

where , where , , and .

This ratio type estimator was earlier mentioned by Muhanty (1967) as regression-in-ratio estimator. Though, Muhanty (1967) used two auxiliary variables. The presence of the regression estimator in the developed estimator is very obvious and significant. This study will refer to the developed estimators of Abbas *et al*. (2018) as regression-in-ratio estimators. Abbas *et al*. (2018) estimators did not consider the correction of extreme values.

The aim of this study is to improve on the three estimators developed by Abbas *et al*. (2018) by introducing the correction factor of Sarndal (1972) for the correction of extreme values in the survey data. Similarly, this study will be interested to investigate into the reaction of the proposed estimators to high and low extreme maximum values in the survey data.

**2.2: Proposed regression-in-ratio estimators**

In equation (2), this study assumed the presence of extreme maximum or minimum value in both the sample of the study and auxiliary variables. The new estimators based on the modification of equations (2), (3) and (4) are presented in equations (7), (9) and (11).

Equation (2) is modified as

This can be re-expressed as

Similarly, equation (3) is modified as

This is re-expressed as

Finally, equation (4) is modified as

This is re-expressed as

In equations (8), (10) and (12), .

To obtain the bias and the for in equation (7), substitute Similarly, , and would be required in the proof.

By substitution, equation (5) becomes where and Apply Taylor series of expansion and the expectation rule will give the bias equation as By further substitution and application of expectation principle, the bias equation is expressed as

The

The MSE of the estimator is expressed as Expand the bracket contents to obtain By further substitution and application of expectation principle, the MSE equation is expressed as

Such that here

Similarly, the corresponding bias and MSE of proposed estimator are presented as

and

Such that and

Finally, the corresponding bias and MSE of proposed estimator are presented as

and

Such that and

The general form of the bias (in equations (13), (15) and (17)) and the MSE (in equations (14), (16), and (18)) of the proposed estimators are, respectively, expressed as

and

The and are expressed as and

## 3.0 Theoretical and Empirical Comparison

### 3.1 Theoretical comparison of the proposed estimators with the corresponding reviewed estimators

This section compares the proposed estimators with the reviewed estimators based on the obtained MSE of each estimator.

1. Comparison of the proposed and the reviewed estimators

This comparison is made by This implies that Hence, Therefore, estimator will be efficient over estimator if .

1. Comparison of the proposed and the reviewed estimators

This comparison is made by This implies that Hence, Therefore, estimator will be efficient over estimator if .

1. Comparison of the proposed and the reviewed estimators

This comparison is made by This implies that Hence, Therefore, estimator will be efficient over estimator if .

### 3.2 Empirical comparison of estimators

*3.2.1 Explanation on HEMaV, HEMiV, LEMaV and LEMiV cases*

This study documented four empirical cases at which extreme values can be present in survey data. The conditions are High Extreme Maximum Values (HEMaV) case, High Extreme Minimum Values (HEMiV) case, Low Extreme Maximum Values (LEMaV) case and Low Extreme Minimum Values (LEMiV) case.

HEMaV distribution considered that both the study and the auxiliary variables had extreme values present in the data and it has positive (maximum) and very high (high) extreme values in the two variables, . These extreme values were sampled from the population data into the sample data in the HEMaV case analysis. The HEMiV distribution considered the extreme values case such that there exists negative (minimum) and high negative (high) extreme values in both the population and sample data of the study and the auxiliary variables. The LEMaV distribution considered the extreme values case where there is positive (maximum) but low positive (low) extreme values in both the population and sample data of the study and the auxiliary variables. Finally, the LEMiV case considered the extreme values case where there is negative (minimum) and low negative (Low) extreme values in both the population and sample data of the study and the auxiliary variables. Figure 2 explains the concept of HEMaV, HEMiV, LEMaV and LEMiV cases on a number line.



Figure 2: Calibrated number line for the explanation of HEMaV, HEMiV, LEMaV and LEMiV cases

The HEMaV simulation analysis injected very high and positive extreme values into the simulated data of such that the injected values were included into the sample of . This process was repeated for the twenty-six simulated populations.

*3.2.2: Conducting the simulation analysis*

This section compares the proposed estimators and the reviewed estimators using numerical case. An algorithm using R Language was developed for the simulation of data for the twenty six populations each with different population and sample sizes. The algorithm conducted the simulation and analysis in accordance to the following procedure.

1. Selection of different population and sample sizes for the twenty six populations in asymptotic procedure.
2. Simulation of data following normal distribution with pre-defined location and scalar values for each of the twenty six populations and for two population variables and with pre-defined high and positive correlation coefficient.
3. An extreme maximum value was injected into each of and such that these extreme values were also sampled into the sample variables and .
4. The extreme maximum value was classified into two types namely HIGH EXTREME MAXIMUM VALUE (HEMaV) and LOW EXTREME MAXIMUM VALUE (LEMaV). Hence, the analysis is divided into two parts in order to examine the reaction of the proposed estimators to both HEMaV and LEMaV.
5. For each of HEMaV and LEMaV, other necessary parameters were computed, the Bias, variances, MSE were, also, computed. Finally, the relative efficiencies were computed, for all the twenty six populations.

Tables 1 through 10 display the analysis results for the bias, MSEs and the percentage relative efficiencies for HEMaV case only. Explanations on the HEMaV case and the LEMaV case are presented in the subsequent section. The HEMiV and the LEMiV case analyses and discussion were not included in this study.

Table 1: Distribution of the population and sample sizes over the twenty-six simulated populations

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Population** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
|  | 5000 | 4800 | 4600 | 4400 | 4200 | 4000 | 3800 | 3600 | 3400 | 3200 | 3000 | 2800 | 2600 |
|  | 1667 | 1600 | 1533 | 1467 | 1400 | 1333 | 1267 | 1200 | 1133 | 1067 | 1000 | 933 | 867 |

Table 2: Distribution of the population and sample sizes over the twenty-six simulated populations (continuation)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Population** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** |
|  | 2400 | 2200 | 2000 | 1800 | 1600 | 1400 | 1200 | 1000 | 800 | 600 | 400 | 200 | 60 |
|  | 800 | 733 | 667 | 600 | 533 | 467 | 400 | 333 | 267 | 200 | 133 | 67 | 20 |

Table 3: HEMaV Analysis of the MSE of the proposed estimators for the 26 populations

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
|  | 3882.77 | 3769.28 | 7854.64 | 4690.25 | 3059.46 | 5708.78 | 4129.15 | 12065.60 | 10910.45 | 11247.47 | 7306.43 | 18202.45 | 7650.58 | 14875.81 |
|  | 3900.27 | 3786.39 | 7892.11 | 4712.12 | 3073.79 | 5737.87 | 4149.49 | 12131.10 | 10972.20 | 11311.95 | 7347.94 | 18312.91 | 7696.89 | 14972.39 |
|  | 3878.59 | 3765.00 | 7848.03 | 4685.04 | 3055.08 | 5702.35 | 4123.51 | 12054.88 | 10899.66 | 11235.86 | 7296.70 | 18185.11 | 7639.06 | 14857.76 |

Table 4: HEMaV Analysis of the MSE of the proposed estimators for the 26 populations (continuation)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** |
|  | 15018.53 | 12993.25 | 48473.33 | 31302.49 | 100816.49 | 51642.09 | 100226.15 | 176565.71 | 535243.52 | 1238196.57 | 1282460.48 | 23247862.93 |
|  | 15124.30 | 13084.59 | 48868.38 | 31556.81 | 101743.57 | 52135.43 | 101270.47 | 178791.69 | 542803.27 | 1260320.25 | 1316292.89 | 24447217.10 |
|  | 14998.52 | 12973.15 | 48424.63 | 31260.71 | 100717.59 | 51566.58 | 100089.01 | 176313.17 | 534508.69 | 1236081.30 | 1277996.50 | 23046194.53 |

Table 5: Rank of the average rank of the proposed estimators based on the MSE results in HEMaV case

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **Ranks of the average rank** |
|  | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | **2** |
|  | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | **3** |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |

Table 6: Rank of the average rank of both the proposed and reviewed estimators based on the bias results in HEMaV case

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **Ranks of the average rank** |
|  | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | **5** |
|  | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | **3** |
|  | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | **6** |
|  | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **4** |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |
|  | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | **2** |

Table 7: Rank of the average rank of both the proposed and reviewed estimators based on the variance results in HEMaV case

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **Average of rank** |
|  | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | **2** |
|  | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **5** |
|  | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | **3** |
|  | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | **6** |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |
|  | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | **4** |

Table 8: Rank of the average rank of both the proposed and reviewed estimators based on the MSE results in HEMaV case

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **Average of rank** |
|  | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | **2** |
|  | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **5** |
|  | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | **3** |
|  | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | **6** |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |
|  | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | **4** |

Table 9: Relative efficiency of the proposed estimators to the corresponding reviewed estimators for the 26 populations in HEMaV analysis

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Population** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
|  | 119.91 | 119.90 | 119.93 | 119.91 | 119.87 | 119.90 | 119.89 | 119.94 | 119.93 | 119.94 | 119.90 | 119.93 | 119.89 |
|  | 119.40 | 119.38 | 119.38 | 119.38 | 119.34 | 119.32 | 119.33 | 119.32 | 119.28 | 119.29 | 119.26 | 119.24 | 119.20 |
|  | 120.04 | 120.04 | 120.03 | 120.04 | 120.04 | 120.04 | 120.05 | 120.05 | 120.04 | 120.06 | 120.06 | 120.04 | 120.07 |

Table 10: Relative efficiency of the proposed estimators to the corresponding reviewed estimators for the 26 populations in HEMaV analysis (continuation)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Populations** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **Average** |
|  | 119.93 | 119.91 | 119.91 | 119.97 | 119.94 | 120.03 | 119.97 | 119.98 | 120.07 | 120.12 | 120.17 | 120.58 | 122.19 | **120.06** |
|  | 119.19 | 119.11 | 119.12 | 119.06 | 119.03 | 119.00 | 118.91 | 118.84 | 118.70 | 118.59 | 118.28 | 117.94 | 117.56 | **119.02** |
|  | 120.07 | 120.06 | 120.10 | 120.09 | 120.10 | 120.14 | 120.14 | 120.15 | 120.24 | 120.28 | 120.37 | 121.00 | 123.22 | **120.25** |

## 4.0: Discussion and Conclusion

### 4.1: Discussion

The distribution of the population and sample sizes in asymptotic order is shown in tables 1 and 2. Tables 3 and 4 show the MSE results for the twenty six populations while tables 5 and 6 show the rank of the average ranks for the obtained MSE results. The lower the MSE result, the higher the ranking of such estimator and vice versa. Table 5 shows the rank of the computed Mean Square Errors (MSEs) which was used to rank the proposed estimators. Tables 6 through 8 show the ranking of the proposed and the reviewed estimators based on the computed bias, variances and MSE, respectively. Finally, tables 9 and 10 show the results of the percentage relative efficiency analysis of the proposed and reviewed estimators for the twenty six simulated populations in HEMaV case. For tables 5 through 10, the average of the computed values for the twenty-six simulated populations were obtained in order to make decision on the concerned estimators. This numerical analysis focuses on the Low and High Extreme Maximum Values (LEMaV and HEMaV) analyses. It is very important to emphases that the algorithm used pre-defined positive and high correlation coefficients in the data simulation of the two variables used in the empirical analysis. However, after the injection of the extreme values in each of the two variables, the positive and high correlation coefficient was distorted and transformed to negative (in some cases) and low correlation coefficients. This article could only report the HEMaV analysis result due to space consumption but the general discussion and conclusion is based on the HEMaV and LEMaV analyses.

Table 5 shows the rank of the three proposed estimators based on the average of the ranks of the computed MSEs for the twenty six populations. The proposed estimators, , were rated as 2nd, 3rd and 1st, respectively. This implies that estimator is the most efficient estimator among the three proposed estimators while estimator is the least efficient estimator in the family of the proposed estimators.

Table 6 shows the rank of the proposed and reviewed estimators based on the computed bias for the twenty-six simulated populations in HEMaV case. The proposed and the reviewed estimators, , were rated as 5th, 3rd, 6th, 4th, 1st and 2nd, respectively. The results show that the reviewed estimator (with 3rd rank) is less bias to the corresponding proposed estimators (with 5th rank). Similarly, the reviewed estimator (with 4th rank) is less biased to the corresponding proposed estimator (with 6th rank). However, the reviewed estimator (with 2nd rank) is more bias than the corresponding proposed estimator (with 1st rank). This implies that the proposed estimators and are more bias than the corresponding reviewed estimators while is less bias than the corresponding reviewed estimator.

Table 7 shows the ranking of the proposed and reviewed estimators based on the computed variances for the twenty-six simulated populations. The proposed and reviewed estimators, , were rated as 2nd, 5th, 3rd, 6th, 1st and 4th, respectively. The proposed estimators, , were rated 2nd, 3rd and1st, respectively, as against the reviewed estimators with the rating of 5th, 6th and 4th, respectively. The same rating results were observed with the ranking of the six estimators based on the computed MSEs in table 8. In general, the proposed estimators proved efficient with lower variances and MSEs over the corresponding reviewed estimators.

Tables 9 and 10 show the percentage relative efficiency results of the proposed estimators to the corresponding reviewed estimators in the HEMaV case. It was observed that the proposed estimators and were averagely and relatively efficient over the corresponding reviewed estimators of and , respectively. Similarly, the proposed estimator was averagely relatively efficient over the corresponding reviewed estimator .

It is important to report that similar ranking results were observed in HEMaV case empirical analysis were also observed in the LEMaV empirical case analysis. Hence, the same conclusion was applied for both the HEMaV and LEMaV empirical analyses.

**4.2: Conclusion**

This study has reviewed the estimators developed by Abbas *et al.* (2018) with the view to correct the effect of the extreme high maximum and minimum values in the survey data. The method of Sarndal (1972) had been used for the correction of the extreme value effect in the estimators. Hence, this study had proposed three regression-in-ratio estimators and the corresponding biases, variances and mean square errors had been obtained. Theoretical comparison proved that the proposed estimators were efficient over the reviewed estimators. Similarly, the empirical comparison, following the simulation in R language for twenty six populations, had also confirmed that the proposed estimators proved efficient when there are low and high extreme maximum values only in the survey data. Sequel to these findings in this study, the proposed estimators are recommended over the reviewed estimators subject to the confirmation of the conditions of usage.

The study has written R language code for these analyses. The 623-line code had been uploaded to github repository as free and open access code. https://bit.ly/2GMG0Og. The code will only perform analyses based on the Low and High Extreme Maximum Values (LEMaV and HEMaV) cases. The code for Low and High Extreme Minimum Values (LEMiV and HEMiV) would be communicated in future research.

## 5.0: Suggestion for further study

This study had used the Low and High Extreme Maximum Values (LEMaV and HEMaV) in the empirical analyses. It is expected that future research would study the reaction of the proposed estimators to Low and High Extreme Minimum Values (LEMiV and HEMiV).

Sarndal (1972) has also developed solution to the challenge of extreme values in survey data. However, there is no method in survey statistics that is used to test (just like Komolgrov-Smirnov test in the general statistics) for significant extreme value in survey data. This study suggests for survey statistics method that can be used to test for significant extreme value effect in survey data before applying the extreme value correction method.

## **Acknowledgement**

The authors appreciate the professional contribution of Dr. Kayode Olurinola, the Chief Medical Director of Steksons and Sons Eye Clinic in Ijebu-Ode of Ogun State, Nigeria.

## References

1. A. Al-Hossain, and M. Khan. *Efficiency of ratio, product and regression estimators under maximu and minimum values, using two auxiliary variables*. The Scientific World Journal, Article ID:693782, 1-6, 2014.
2. B. Kiregrera. *Regression-type estimator using two auxiliary variables and model of double sampling from finite populations.* Metrika, **31**, 215-226, 1984.
3. C. Kadilar and H. Cingi. *A new estimator using two auxiliary variables.* Applied Mathematics and Computation. **162**, 901-908, 2005.
4. C. Kadilar and H. Cingi. *An improvement in estimating the population mean by using the correlation coefficient*. Hacettepe Journal of Mathematics and Statistics. **35**, 103-109, 2006.
5. C. Kadilar and H. Cingi. *Ratio estimators in simple random sampling.* Applied Mathematics and Computation. **151**, 893-902, 2004.
6. C. Sarndal. *Sample Survey theory Vs. general statistical theory: Estimation of the population mean.* Int. Stat. Rev., **40**(1), 1-12, 1972.
7. C. J. Bulpitt, C. Hodes and M.G. Everitt. *Intraocular pressure and systemic blood pressure in the elderly*. Brit. J. Ophthal. **59**, 717-720, 1975.
8. D. Robson. *Application of Multivariate Polykays to the Theory of unbiade Ratio Type Estimators.* Journal of American Statistical Association, **52**, 511-522, 1957.
9. G. Currie and C. Delles. *Blood pressure targets in the elderly.* Journal of Hypertension, **36**(2), 234-236, 2018.
10. H.J. Chung, H.B. Hwang and N.Y. Lee. *The association between primary open-angle glaucoma and blood pressure: Two aspects of hypertension and hypotension.* BioMed Research International, Hindawi Publishing Corporation, Article ID:827516, 1-7, <http://dx.doi.org/10.1155/2015/827516>, 2015.
11. J. Subramani and G. Kumarapandiyan. *Estimation of population mean using coefficient of variation and median of an auxiliary variable.* International Journal of Probability and Statistics, **1**, 111-1118, 2012b.
12. J. Subramani and G. Kumarapandiyan. *Estimation of population mean using known median and coefficient of kurtosis.* American Journal of Mathematics and Statistics, **2**, 101-107, 2012c.
13. J. Subramani, and G. Kumarapandiyan. *A class of almost unbiased modified ratio estimators for population mean with known population parameters.* Elixir Statistics, **44**, 7411-7415, 2012a.
14. J. Subramani, and G. Kumarapandiyan. *Modified ratio estimators using known median and coefficient of kurtosis.* American Journal of Mathematics and Statistics, **2**, 95-100, 2012d.
15. M. Khan, and J. Shabbir. *Some improved ratio, product and regression estimators of finite population mean when using minimum and maximum values.* The Scientific World Journal, Article ID:431868, 1-7, 2013.
16. M. Khan, S. Ullah, A. Al-Hossain and N. Bashir. *Improve ratio-type estimators using maximum and minimum values under simple random sampling scheme*. Hacettepe Journal of Mathematics and Statistics. **44**(4), 923-931, Doi: [10.15672/HJMS.2014297480](http://dx.doi.org/10.15672/HJMS.2014297480), 2015.
17. M.I. Jeelani, S. Moqbool, and S.A. Mir. *Modified ratio estimators of population mean using linear combination of coefficient of skewness and quartile deviation.* International Journalof Modern Mathematical Science, **6**, 172-183, 2013.
18. M.N. Murthy. *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta, India, 1967.
19. N. Abbas, M. Abid, M. Tahir, N. Abbas, and Z. Hussain. *Enhancing ratio estimators for estimating population mean using maximum value of auxiliary variable.* J. Natn. Sci. Foundation Sri Lanka **3**, 453-463, 2018.
20. S. Mohanty. *Combination of Regression and Ratio Estimate.* Jour. Ind. Statist. Assoc., **5**, 16-19, 1967.
21. V.P. Godambe and V.M. Joshi. *Admissibility and Bayes estimation in sampling from the population-I*. Ann. Math. Statist., **36**, 1707-1722, 1965.
22. V.P. Godambe. *A new approach to sampling from finite populations-I.* Sufficiency and Linear estimation. J. R. Statist. Soc. B, **28**, 310-319, 1966.
23. V.P. Godambe. *A unified theory of sampling from finite populations.* J. R. Statist. Soc. B, **17**, 268-278, 1955.
24. V.P. Godambe. *Some aspects of the theoretical developments in survey-sampling. New development in survey sampling*. New York, Wiley, 27-58, 1969.
25. W.G. Cochran. *Sampling Technique*. 3rd Edition, Wiley Eastern Limited, India, 1940.
26. Z. Yan and B. Tian. *Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. Information computing and applications.* ICICA 2010. Communications in Computer and Information science, **106**, Heidelberg, Berlin, Germany, 2010.