Minimization of Ratio Over-estimation Problem in Simple Random Sampling when Population Means are known

***Adewara, A.A.**

Department of Statistics, University of Ilorin, Ilorin, Nigeria

Received: August 8, 2016; *Revised: December 15, 2016;* *Accepted: December 23, 2016*

Abstract

In this study, the presence of over-estimation in the conventional ratio estimator $(\bar{y}_r)$ on mean estimator $(\bar{y})$ when population means are known is being contested. Three categories of data sets were used to justify this work. Category A is when the population mean of auxiliary variable is lower than the population mean of the variable of interest $(\bar{X} < \bar{Y})$, category B is when the population mean of auxiliary variable is greater than the population mean of the variable of interest $(\bar{X} > \bar{Y})$ while category C is when the population mean of auxiliary variable is the same as the population mean of the variable of interest $(\bar{X} = \bar{Y})$. It was observed that for all these categories evidences of over-estimation of $\bar{y}_r$ on $\bar{y}$ were recorded whenever $\rho < \frac{c_x}{2c_y}$. One of the earliest suggested alternatives, $\bar{y}_{str}$, also over estimated $\bar{y}$ when tested. Hence, an alternative, $\bar{y}_{aaar}$ which utilizes the regression estimate of the study and auxiliary variables (β_{xy}) under consideration was suggested and found to minimize over estimation of $\bar{y}_r$ on $\bar{y}$ whenever $\rho_{xy} > \frac{[c^2{}_y(\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 - 1) + (\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2}$, $0.1 \le \alpha < 0.7$. The conclusion from this study is that $\bar{y}_{aaar}$ may generally be used whenever problem of over estimation is encountered in ratio estimation.

Keywords: bias, estimator, mean square error, ratio, regression

1.0 Introduction

Let N and n be the population and sample sizes respectively, $\bar{X}$ and $\bar{Y}$ be the population means for the auxiliary variable (X) and the variable of interest (Y), $\bar{x}$ and $\bar{y}$ be the sample means based on the sample drawn. If the correlation between the study variable y and the auxiliary variable x is positive, the ratio method of estimation is quite effective. This paper is interested only in ratio method of estimation.

Then conventionally [1 – 2],

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{1}$$

$$bias(\bar{y}_r) = (\frac{N-n}{Nn})\bar{Y}[c^2{}_x - \rho_{xy}c_x c_y] \tag{2}$$

$$mse(\bar{y}_r) = (\frac{N-n}{Nn})\bar{Y}^2[c^2{}_x + c^2{}_y - 2\rho_{xy}c_x c_y] \tag{3}$$

$$mse(\bar{y}) = \bar{Y}^2(\frac{N-n}{Nn})c^2{}_y \tag{4}$$

***Corresponding Author: Tel: +234(0)8033813093; Email: aaadewara@gmail.com, aadewara@unilorin.edu.ng**

In sample surveys, supplementary information is often used to increase the precision of estimators [3–4]. Many authors have used auxiliary information for improved estimation of population mean of studied variable y [5–12]. The intent of this study is to detect the presence or otherwise of over estimation in $\bar{y}_r$ on $\bar{y}$ for using auxiliary information to improve estimation of population mean of study variable y under these three categories (Category A is when the population mean of auxiliary variable is lower than the population mean of the variable of interest $(\bar{X} < \bar{Y})$, category B is when the population mean of auxiliary variable is greater than the population mean of the variable of interest $(\bar{X} > \bar{Y})$ while category C is when the population mean of auxiliary variable is the same as the population mean of the variable of interest $(\bar{X} = \bar{Y})$) and how it can be minimized. This will be achieved by comparing $mse(\bar{y}_r)$ with $mse(\bar{y})$. Mean square error (mse) is used here because it is a good criterion normally in use in sample survey when efficiency between two estimators is to be determined.

One of the earliest alternative ratio estimators, $\bar{y}_{str}$, was suggested by Srivenkataramana and Tracy [12] as:

$$\bar{y}_{str} = \bar{y}\frac{N\bar{X} - n\bar{x}}{(N-n)\bar{X}} \tag{5}$$

where,

$$bias(\bar{y}_{str}) = (\frac{N-n}{Nn})\bar{Y}[(\frac{n}{N-n})^2 c^2{}_x - (\frac{n}{N-n})\rho_{xy} c_x c_y] \tag{6}$$

$$mse(\bar{y}_{str}) = (\frac{N-n}{Nn})\bar{Y}^2[(\frac{n}{N-n})^2 c^2{}_x + c^2{}_y - 2(\frac{n}{N-n})\rho_{xy} c_x c_y] \tag{7}$$

Adewara [8] suggested an alternative ratio estimator:

$$\bar{y}_{aaar} = \alpha(\frac{\bar{y}}{\bar{x}})(\bar{X} + \beta_{xy}) \tag{8}$$

where,

$$bias(\bar{y}_{aaar}) = \alpha(\frac{\bar{X} + \beta_{xy}}{\bar{X}})(\frac{N-n}{Nn})\bar{Y}(c^2{}_x - 2\rho c_x c_y) \tag{9}$$

$$mse(\bar{y}_{aaar}) = \alpha^2(\frac{\bar{X} + \beta_{xy}}{\bar{X}})^2(\frac{N-n}{Nn})\bar{Y}^2(c^2{}_x + c^2{}_y - 2\rho c_x c_y) \tag{10}$$

2.0 Materials and Methods

2.1 Data Sets Used

Three categories of data sets were used to justify this work (Table 1). Category A, when the population mean of auxiliary variable is lower than the population mean of the variable of interest $(\bar{X} < \bar{Y})$, category B, when the population mean of auxiliary variable is greater than the population mean of the variable of interest $(\bar{X} > \bar{Y})$ while category C is when the population mean of auxiliary variable is the same as the population mean of the variable of interest $(\bar{X} = \bar{Y})$.

2.2 Over-estimation of $\bar{y}_r$ on $\bar{y}$

Over estimation in $\bar{y}_r$ on $\bar{y}$ is calculated as:

$$mse(\bar{y}_r) - mse(\bar{y}) = (\frac{N-n}{Nn})\bar{Y}^2[c^2{}_x + c^2{}_y - 2\rho_{xy}c_x c_y] - (\frac{N-n}{Nn})\bar{Y}^2 c^2{}_y \quad (11)$$

It is recorded whenever,

$$\rho_{xy} < \frac{c_x}{2c_y} \quad (12)$$

That is, $mse(\bar{y}_r) > mse(\bar{y})$ (13)

Table 1: Data sets used for the Study

Population	**I**	**II**	**III**	**IV**
Source	Subramani and Kumarapandiyan [9]	Subramani and Kumarapandiyan [9]	Subramani and Kumarapandiyan [9]	Hyphothetical
Case	$\bar{X} < \bar{Y}$	$\bar{X} < \bar{Y}$	$\bar{X} > \bar{Y}$	$\bar{X} = \bar{Y}$
Population (N)	200	80	103	100
Sample (n)	20	20	40	40
$\bar{X}$	18.515	11.2646	556.5541	0.6
$\bar{Y}$	42.485	51.8264	62.6212	0.6
c_x	0.3763	0.7507	1.0963	0.0094
c_y	0.3321	0.3542	1.4588	0.0097
ρ_{xy}	0.8652	0.9413	0.7298	0.05
β_{xy}	1.7521	2.0434	0.1092	0.0516
$(\frac{n}{N-n})$	0.1111	0.3333	0.6349	0.6667
$(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2$	1.1982	1.3957	1.0004	1.1794

2.3 Over-estimation of $\bar{y}_{str}$ on $\bar{y}$ and $\bar{y}_r$

Over-estimation in $\bar{y}_{str}$ on $\bar{y}$ is calculated as:

$$mse(\bar{y}_{str}) - mse(\bar{y}) = (\frac{N-n}{Nn})\bar{Y}^2[(\frac{n}{N-n})^2 c^2{}_x + c^2{}_y - 2(\frac{n}{N-n})\rho_{xy}c_x c_y] - (\frac{N-n}{Nn})\bar{Y}^2 c^2{}_y \quad (14)$$

It is recorded whenever,

$$\rho_{xy} < \frac{[c_x(\frac{n}{N-n})]}{2c_y} \quad (15)$$

That is, $mse(\bar{y}_{str}) > mse(\bar{y})$ (16)

Over-estimation in $\bar{y}_{str}$ on $\bar{y}_r$ is calculated as:

$$mse(\bar{y}_{str}) - mse(\bar{y}_r) = (\frac{N-n}{Nn})\bar{Y}^2[(\frac{n}{N-n})^2 c^2{}_x + c^2{}_y - 2(\frac{n}{N-n})\rho_{xy} c_x c_y] - (\frac{N-n}{Nn})\bar{Y}^2[c^2{}_x + c^2{}_y - 2\rho_{xy} c_x c_y] \quad (17)$$

It is recorded whenever,

$$\rho_{xy} > \frac{[c_x((\frac{n}{N-n})^2 - 1)]}{2c_y[((\frac{n}{N-n}) - 1)]} \quad (18)$$

That is, $mse(\bar{y}_{str}) > mse(\bar{y}_r)$ (19)

2.4 Over-estimation Of $\bar{y}_{aaar}$ On $\bar{y}$

Over estimation in $\bar{y}_{aaar}$ on $\bar{y}$ is calculated as:

$$mse(\bar{y}_{aaar}) - mse(\bar{y}) = \alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2(\frac{N-n}{Nn})\bar{Y}^2(c^2{}_x + c^2{}_y - 2\rho c_x c_y) - (\frac{N-n}{Nn})\bar{Y}^2 c^2{}_y \quad (20)$$

It is recorded whenever,

$$\rho_{xy} < \frac{[c^2{}_y(\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 - 1) + (\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2}, 0.1 \le \alpha \le 1.0 \quad (21)$$

$\rho_{xy} < G$,

where G = $\frac{[c^2{}_y(\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 - 1) + (\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2}$ (22)

That is, $mse(\bar{y}_{aaar}) > mse(\bar{y})$ (23)

3.0 Results

The mean square errors obtained on $\bar{y}$ and $\bar{y}_r$ are presented in Table 2. The data revealed that for populations I and III, $mse(\bar{y}_r) < mse(\bar{y})$ while for populations II and IV $mse(\bar{y}_r) > mse(\bar{y})$. The effect of over estimation of $\bar{y}_r$ on $\bar{y}$ is felt because $\rho_{xy} < \frac{c_x}{2c_y}$

Table 3 presents the mean square errors obtained on $\bar{y}$ and $\bar{y}_{str}$. In this case, $mse(\bar{y}_{str}) < mse(\bar{y})$ for populations I, II and III while for population IV, $mse(\bar{y}_{str}) > mse(\bar{y})$. The effect of over estimation of $\bar{y}_{str}$ on $\bar{y}$ is felt because $\rho_{xy} < \frac{[c_x(\frac{n}{N-n})]}{2c_y}$.

Table 2:- Mean Square Errors obtained on $\bar{y}$ and $\bar{y}_r$

Estimator/Population	**I**	**II**	**III**	**IV**
Case	$\bar{X} < \bar{Y}$	$\bar{X} < \bar{Y}$	$\bar{X} > \bar{Y}$	$\bar{X} = \bar{Y}$
$mse(\bar{y})$	8.9582	12.6366	127.6079	0.00000051
$mse(\bar{y}_r)$	2.8953	18.9847	59.7030	0.0000009
ρ_{xy}	0.8652	0.9413	0.7298	0.0500
$\frac{c_x}{2c_y}$	0.3567	1.0597	0.3758	0.4845
Remark on $\bar{y}_r$ over $\bar{y}$	$\rho_{xy} > \frac{c_x}{2c_y}$	$\rho_{xy} < \frac{c_x}{2c_y}$ **(Over-estimation)**	$\rho_{xy} > \frac{c_x}{2c_y}$	$\rho_{xy} < \frac{c_x}{2c_y}$ **(Over-estimation)**

Table 3:- Mean Square Errors obtained on $\bar{y}$ and $\bar{y}_{str}$

Estimator/Population	**I**	**II**	**III**	**IV**
Case	$\bar{X} < \bar{Y}$	$\bar{X} < \bar{Y}$	$\bar{X} > \bar{Y}$	$\bar{X} = \bar{Y}$
$mse(\bar{y})$	8.9582	12.6366	127.6079	0.00000051
$mse(\bar{y}_{str})$	7.1488	2.1372	67.7895	*0.00000069*
ρ_{xy}	0.8652	0.9413	0.7298	0.0500
$\frac{[c_x(\frac{n}{N-n})]}{2c_y}$	0.0629	0.3532	0.2386	0.3230
Remark on $\bar{y}_{str}$ over $\bar{y}$	$\rho_{xy} > \frac{[c_x(\frac{n}{N-n})]}{2c_y}$	$\rho_{xy} > \frac{[c_x(\frac{n}{N-n})]}{2c_y}$	$\rho_{xy} > \frac{[c_x(\frac{n}{N-n})]}{2c_y}$	$\rho_{xy} < \frac{[c_x(\frac{n}{N-n})]}{2c_y}$ **(Over-estimation)**

The mean square errors obtained on $\bar{y}_r$ and $\bar{y}_{str}$ are presented in Table 4. Here, for populations II and IV, $mse(\bar{y}_{str}) < mse(\bar{y}_r)$. However, for populations I and III, $mse(\bar{y}_{str}) > mse(\bar{y}_r)$. The effect of over estimation of $\bar{y}_{str}$ on $\bar{y}_r$ is felt because

$$\rho_{xy} > \frac{[c_x((\frac{n}{N-n})^2 - 1)]}{2c_y[((\frac{n}{N-n}) - 1)]}$$

Table 5 shows the mean square errors obtained on $\bar{y}$, $\bar{y}_r$ and $\bar{y}_{aaar}$. From Table 2, the effect of over estimation of $\bar{y}_r$ on $\bar{y}$ is felt for populations II and IV whenever $\rho_{xy} < \frac{c_x}{2c_y}$ because $mse(\bar{y}_r) > mse(\bar{y})$. Hence, there is need to minimize this using $\bar{y}_{aaar}$, in other to determine "the level at which α will minimize over estimation of $\bar{y}_r$ on $\bar{y}$ ". The data showed that $mse(\bar{y}_{aaar}) < mse(\bar{y}) < mse(\bar{y}_r)$, whenever $0.1 \le \alpha < 0.7$.

The values of G when $0.1 \le \alpha \le 1.0$ obtained on $\bar{y}_{aaar}$ are presented in Table 6. Under this condition, the effect of over estimation of $\bar{y}_r$ on $\bar{y}$ felt in populations II and IV (Table 5; when $\rho_{xy} < G, 0.7 \le \alpha \le 1.0$) will be minimized whenever $\rho_{xy} > G$, $0.1 \le \alpha < 0.7$ (Table 6). Therefore, the information presented in Table 6 clearly answered the intent of this study which is "at what level of α will $\bar{y}_{aaar}$ minimizes over estimation of $\bar{y}_r$ on $\bar{y}$ ".

Table 4:- Mean Square Errors obtained on $\bar{y}_r$ and $\bar{y}_{str}$

Estimator/Population	**I**	**II**	**III**	**IV**
Case	$\bar{X} < \bar{Y}$	$\bar{X} < \bar{Y}$	$\bar{X} > \bar{Y}$	$\bar{X} = \bar{Y}$
$mse(\bar{y}_r)$	2.8953	18.9847	59.7030	0.0000009
$mse(\bar{y}_{str})$	***7.1488***	2.1372	***67.7895***	0.00000069
ρ_{xy}	0.8652	0.9413	0.7298	0.0500
$\frac{[c_x((\frac{n}{N-n})^2-1)]}{2c_y[((\frac{n}{N-n})-1)]}$	0.5596	1.4129	0.6143	0.8076
Remark on $\bar{y}_{str}$ over $\bar{y}_r$	$\rho_{xy} > \frac{[c_x((\frac{n}{N-n})^2-1)]}{2c_y[((\frac{n}{N-n})-1)]}$ **(Over-estimation)**	$\rho_{xy} < \frac{[c_x((\frac{n}{N-n})^2-1)]}{2c_y[((\frac{n}{N-n})-1)]}$	$\rho_{xy} > \frac{[c_x((\frac{n}{N-n})^2-1)]}{2c_y[((\frac{n}{N-n})-1)]}$ **(Over-estimation)**	$\rho_{xy} < \frac{[c_x((\frac{n}{N-n})^2-1)]}{2c_y[((\frac{n}{N-n})-1)]}$

Table 5:- Mean Square Errors obtained on $\bar{y}$, $\bar{y}_r$ and $\bar{y}_{aaar}$

Estimator/Population	**I**	**II**	**III**	**IV**
Case	$\bar{X} < \bar{Y}$	$\bar{X} < \bar{Y}$	$\bar{X} > \bar{Y}$	$\bar{X} = \bar{Y}$
$mse(\bar{y})$	8.9582	**12.6366**	127.6079	**0.00000051**
$mse(\bar{y}_r)$	2.8953	**18.9847**	59.7030	**0.0000009**
$mse(\bar{y}_{aaar})$ when				
α =1.0	3.4691	26.4969	59.7269	0.0000011
0.9	2.81	21.4625	48.3788	0.00000089
0.8	2.22	16.9581	38.2252	0.00000070
0.7	1.7	12.9835	29.2662	0.00000054
0.6	1.2489	**9.5389**	21.5017	**0.00000040**
0.5	0.8673	**6.6242**	14.9318	**0.00000028**
0.4	0.5551	**4.2395**	9.5563	**0.00000018**
0.3	0.3122	**2.3847**	6.6517	**0.000000099**
0.2	0.3188	**1.0599**	2.9563	**0.000000044**
0.1	0.0347	**0.2650**	0.7391	**0.000000011**

Table 6:- Values of G when $0.1 \le \alpha \le 1.0$ obtained on $\bar{y}_{aaar}$

α / pop	I	II	III	IV
1.0	0.6395	1.1266	0.3760	0.5630
0.9	0.5532	1.0869	0.22	0.4604
0.8	0.4324	1.0315	0.0019	0.3169
0.7	0.2562	0.9507	-0.3162	0.1077
0.6	-0.0152	**0.8261**	-0.8063	**-0.2147**
0.5	-0.4653	**0.6195**	-1.6192	**-0.7494**
0.4	-1.2939	**0.2392**	-3.1156	**-1.7337**
0.3	-3.0842	**-0.5825**	-6.3485	**-3.8603**
0.2	-8.1991	**-2.9301**	-15.5855	**-9.9364**
0.1	-35.8199	**-15.6072**	-65.4652	**-42.7470**
ρ_{xy}	0.8652	0.9413	0.7298	0.0500

4.0 Discussion and Conclusion

The effect of over estimation of $\bar{y}_r$ on $\bar{y}$ is felt for populations II and IV whenever $\rho_{xy} < \frac{c_x}{2c_y}$ because $mse(\bar{y}_r) > mse(\bar{y})$ (Table 2) while the effect of over estimation of $\bar{y}_{str}$ on $\bar{y}$ is felt only for population IV whenever $\rho_{xy} < \frac{[c_x(\frac{n}{N-n})]}{2c_y}$ because $mse(\bar{y}_{str}) > mse(\bar{y})$ (Table 3). Furthermore, the effect of over estimation of $\bar{y}_{str}$ on $\bar{y}_r$ is felt for populations I and III whenever $mse(\bar{y}_{str}) > mse(\bar{y}_r)$ because $\rho_{xy} > \frac{[c_x((\frac{n}{N-n})^2 - 1)]}{2c_y[((\frac{n}{N-n}) - 1)]}$ (Table 4). Finally, the effect of over estimation of $\bar{y}_r$ on $\bar{y}$ felt for populations II and IV has been minimized using $\bar{y}_{aaar}$ (Tables 5 and 6).

In conclusion therefore, for all the three categories ($\bar{X} < \bar{Y}$, $\bar{X} > \bar{Y}$ and $\bar{X} = \bar{Y}$) considered in this study, the flaw of over estimation of $\bar{y}_r$ on $\bar{y}$ recorded in populations II and IV when

$$\rho_{xy} < \frac{[c^2{}_y(\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 - 1) + (\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2}, 0.7 \le \alpha \le 1.0$$

will be minimized whenever

$$\rho_{xy} > \frac{[c^2{}_y(\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 - 1) + (\alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}+\beta_{xy}}{\bar{X}})^2}, \quad 0.1 \le \alpha < 0.7$$

using the ratio estimator, $\bar{y}_{aaar}$ as proposed by Adewara [8]. Hence, $mse(\bar{y}_{aaar}) < mse(\bar{y}) < mse(\bar{y}_r)$, whenever $0.1 \le \alpha < 0.7$

References

[1] Cochran, W.G. (1977). Sampling Techniques. 3rd Edition. John Wiley, New York.

[2] Okafor, F.C. (2002). Sample Survey Theory with Applications. First Edition, Afro-Orbit Publications Ltd. University of Nigeria, Nsukka, Nigeria.

[3] Adewara, A.A. (2006). Effect of improving both the auxiliary and variable of interest in ratio and product estimators. Proceedings of Pakistan Academy Science, Vol. 43, No. 4, pp. 275–278.

[4] Adewara, A.A., Rajesh, S. and Mukesh, K. (2012). Efficiency of Some Modified Ratio and Product Estimators Using Known Value of Some Population Parameters. International Journal of Applied Science and Technology, Vol. 2, No. 2, pp. 76-79.

[5] Adewara, A.A. (2015). Alternative to Tin Ratio Type Estimator. Al-Hikmah Journal of Pure and Applied Sciences, Vol. 2, pp. 7-13.

[6] Adewara, A.A. (2016a). Alternative Linear Combined Product Estimator of Population Mean in Simple Random Sampling. Journal of the Nigerian Association of Mathematical Physics, Vol. 33, No. 2, pp. 119–124.

[7] Adewara, A.A. (2016b). Alternative to Thakur and Gupta Linear Combined Product – Ratio Type Estimator of Population Mean in Simple Random Sampling. Journal of the Nigerian Association of Mathematical Physics, Vol. 33, No. 2, pp. 107–112.

[8] Adewara, A.A. (2016c). Some Improved Ratio Estimators Using Known Values of Some Population Parameters. Al-Hikmah Journal of Pure and Applied Sciences, Vol. 2, No. 2, pp. 89-97.

[9] Subramani, J. and Kumarapandiyan, G. (2012). Estimation of Population Mean Using Co-efficient of Variation and Median of an Auxiliary Variable. International Journal of Probability and Statistics, Vol. 1, No.4, pp. 111-118.

[10] Srivenkataramana, T. and Srinath, K.P. (1976). Ratio and Product Method of estimation in Sample Surveys when the two variables are moderately correlated. Vignana Bharathi, Vol. 2, pp. 54–58.

[11] Singh, R., Singh, V.K. and Adewara, A.A. (2013). Some Improved Estimators for Estimating Population Mean in Stratified Random Sampling. Journal of Scientific Research, Vol. 57, pp. 154–164.

[12] Strivenkataramana, T and Tracy, D.S. (1980). An alternative to ratio method in sample survey. Annals of the Institute of Statistical Mathematics, Vol. 32, pp. 111–120.