EFFECT OF NEGLIGIBILITY OF KNOWN AUXILIARY VARIABLE COEFFICIENT OF VARIATION IN PRODUCT ESTIMATION

A. A. Adewara

Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria

Corresponding Author: A. A. Adewara, aaadewara@gmail.com, aadewara@unilorin.edu.ng.

ABSTRACT: If the auxiliary variable coefficient of variation (c_x) is known but negligible, then what comes of the mean square error (mse) of our conventional product and mean per unit estimators. In this study, if c_x is known but negligible, that is, if it tend towards zero ($c_x \to 0$), the mse of both the conventional product and mean per unit estimator becomes equal. The question that will quickly come to mind here is, how often does known c_x becomes negligible?. It at times do occur in sample survey, so there must be statistical awareness on this in case we experience such in our researches, just like when the population size (N) is large, then, the finite population correction (fpc) becomes negligible and hence, it tends towards zero ($f = \frac{n}{N} \to 0$). If $c_x \to 0$, then virtually the mse of some of the existing proposed alternatives may tend towards the mse of the mean per unit estimator. When this occur, over estimation problem in product estimation does not arise but if known c_x is not negligible, Adewara ([Ade16]) proposed an alternative product estimator, $\bar{y}_{aaap}$, which utilizes c_x which was found to minimize over estimation of $\bar{y}_p$ on $\bar{y}$ whenever

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

$0.1 \le \alpha < 0.7$

KEYWORDS: product, coefficient of variation, estimator, mean square error, bias.

1. INTRODUCTION

It is well known that the use of auxiliary information at the estimation stage provides efficient estimators of the parameter (s) of the study character y. When the population mean $\bar{X}$ of the auxiliary character x is known, a large number of estimators such as ratio, product and regression estimators and their modifications, have been suggested by various authors. Das and Tripathi ([DT80]) have advocated that the coefficient of variation c_x of the auxiliary character x is also available in many practical situations. Keeping this in view, Sisodia and Dwivedi ([SD81]), Singh and Upadhyaya ([SU86]) and Pandey and Dubey ([PD88]) have made the use of coefficient of variation c_x along with the population mean $\bar{X}$ in estimating the population mean $\bar{Y}$ of y ([ST05]). These are some of the earlier researchers on alternative estimators.

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, c_x and c_y be the coefficient of variation for auxiliary and variable of interest respectively.

Conventionally ([Mur61, Coc77]),

$$\bar{y}_p = \frac{\bar{y}\bar{x}}{\bar{X}} \quad (1)$$

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

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

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

$$bias(\bar{y}) = 0 \quad (5)$$

The intent of this study is to determine the effect of negligibility of known c_x on product and mean per unit estimator.

Adewara ([Ade16]) suggested an alternative product estimator:

$$\bar{y}_{aaap} = \alpha(\frac{\bar{y}\bar{x}}{\bar{X}+c_x}) \quad (6)$$

where,

$$bias(\bar{y}_{aaap}) = \alpha\left(\frac{\bar{X}}{\bar{X}+c_x}\right)\left(\frac{N-n}{Nn}\right)\bar{Y}(c^2{}_x + 2\rho c_x c_y) \quad (7)$$

$$mse(\bar{y}_{aaap}) = \alpha^2\left(\frac{\bar{X}}{\bar{X}+c_x}\right)^2\left(\frac{N-n}{Nn}\right)\bar{Y}^2(c^2{}_x + c^2{}_y + 2\rho c_x c_y) \quad (8)$$

2.0. WHEN c_x IS KNOWN BUT NEGLIGIBLE ($c_x \to 0$):

If $c_x \to 0$, eqs (2,3,7 and 8) becomes

$$bias(\bar{y}_p) = 0 \quad (9)$$

$$mse(\bar{y}_p) = \bar{Y}^2\left(\frac{N-n}{Nn}\right)c^2{}_y \quad (10)$$

$$bias(\bar{y}_{aaap}) = 0 \quad (11)$$

$$mse(\bar{y}_{aaap}) = \alpha^2\bar{Y}^2\left(\frac{N-n}{Nn}\right)c^2{}_y \quad (12)$$

One can deduce that, $bias(\bar{y}) = bias(\bar{y}_p) = bias(\bar{y}_{aaap}) = 0$ and $mse(\bar{y}) = mse(\bar{y}_p) = mse(\bar{y}_{aaap}) = \bar{Y}^2\left(\frac{N-n}{Nn}\right)c^2{}_y$, whenever $\alpha = 1$ for $\bar{y}_{aaap}$

2.1. WHEN c_x IS NOT NEGLIGIBLE ($c_x \neq 0$):

If $c_x \neq 0$, then eqs. 2, 3, 7 and 8 remains.

The data sets used to justify this section are as shown in the table below:

Table 1: Data sets used

Population	I	II	III
Source	([Mad77])	Hypho-thetical	([PD88])
Population (N)	30	100	20
Sample (n)	6	30	8
$\bar{X}$	75.4313	0.2	18.8
$\bar{Y}$	7.6375	0.3	19.55
c_x	0.0986	0.0036	0.1281
c_y	0.2278	0.0036	0.1445
ρ_{xy}	-0.6823	-0.05	-0.9199
$\left(\frac{N-n}{Nn}\right)$	0.13333	0.0233	0.075
$\left(\frac{n}{N-n}\right)$	0.25	0.4286	0.6667
$\left(\frac{\bar{X}}{\bar{X}+c_x}\right)$	0.9987	0.9823	0.9932

2.2 OVER ESTIMATION OF $\bar{y}_p$ ON $\bar{y}$.

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

$$mse(\bar{y}_p) - mse(\bar{y}) = \left(\frac{N-n}{Nn}\right)\bar{Y}^2[c^2{}_x + c^2{}_y + 2\rho_{xy}c_x c_y] - \left(\frac{N-n}{Nn}\right)\bar{Y}^2 c^2{}_y \quad (13)$$

It is recorded whenever,

$$\rho_{xy} < -\frac{c_x}{2c_y} \quad (14)$$

That is,

$$mse(\bar{y}_p) > mse(\bar{y}) \quad (15)$$

2.3 OVER ESTIMATION OF $\bar{y}_{aaap}$ ON $\bar{y}$

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

$$mse(\bar{y}_{aaap}) - mse(\bar{y}) = \alpha^2\left(\frac{\bar{X}}{\bar{X}+c_x}\right)^2\left(\frac{N-n}{Nn}\right)\bar{Y}^2(c^2{}_x + c^2{}_y + 2\rho c_x c_y) - \left(\frac{N-n}{Nn}\right)\bar{Y}^2 c^2{}_y \quad (16)$$

It is recorded whenever,

$$\rho_{xy} > \frac{\left[c^2{}_y\left(1-\alpha^2\left(\frac{\bar{X}}{\bar{X}+c_x}\right)^2\right) - \alpha^2\left(\frac{\bar{X}}{\bar{X}+c_x}\right)^2 c^2{}_x\right]}{2c_y c_x \alpha^2\left(\frac{\bar{X}}{\bar{X}+c_x}\right)^2}, \quad 0.1 \le \alpha \le 1.0 \quad (17)$$

That is,

$$mse(\bar{y}_{aaap}) > mse(\bar{y}) \quad (18)$$

3.0 RESULTS

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

Estimator / population	I	II	III
$mse(\bar{y})$	0.4035	0.000000027	0.5985
$mse(\bar{y}_p)$	0.2408	**0.000000052**	0.0927
ρ_{xy}	-0.6823	-0.05	-0.9199
$-\dfrac{c_x}{2c_y}$	-0.2164	-0.50	-0.4433
Remark on $\bar{y}_p$ over $\bar{y}$	$\rho_{xy} < -\dfrac{c_x}{2c_y}$	$\rho_{xy} > -\dfrac{c_x}{2c_y}$ **(Over estimation)**	$\rho_{xy} < -\dfrac{c_x}{2c_y}$

Here, the effect of over estimation of $\bar{y}_p$ on $\bar{y}$ is felt for population II whenever $\rho_{xy} > -\dfrac{c_x}{2c_y}$ because $mse(\bar{y}_p) > mse(\bar{y})$. For populations I and III, over estimation is not felt because $\rho_{xy} < -\dfrac{c_x}{2c_y}$ and $mse(\bar{y}_p) < mse(\bar{y})$

Table 3: Mean Square Errors obtained on $\bar{y}$, $\bar{y}_p$ and $\bar{y}_{aaap}$

Estimator / population	I	II	III
$mse(\bar{y})$	0.4035	0.000000027	0.5985
$mse(\bar{y}_p)$	0.2408	0.000000052	0.0927
ρ_{xy}	-0.6823	-0.05	-0.9199
$mse(\bar{y}_{aaap})$ when α =1.0	**0.2405**	0.000000050	**0.0921**
0.9	**0.1948**	0.000000041	**0.0746**
0.8	**0.1539**	0.000000032	**0.0589**
0.7	**0.1178**	**0.000000025**	**0.0451**
0.6	**0.0866**	**0.000000018**	**0.0332**
0.5	**0.0601**	**0.000000013**	**0.0230**
0.4	**0.0385**	**0.000000008**	**0.0147**
0.3	**0.0216**	**0.000000005**	**0.0083**
0.2	**0.0096**	**0.000000002**	**0.0037**
0.1	**0.0024**	**0.0000000005**	**0.0092**

We can see in Table 2 that the effect of over estimation of $\bar{y}_p$ on $\bar{y}$ is felt for population II whenever $\rho_{xy} > -\dfrac{c_x}{2c_y}$ because $mse(\bar{y}_p) > mse(\bar{y})$. Hence, there is need to minimize this using, $\bar{y}_{aaap}$, therefore, the intent here is " when the known auxiliary variable coefficient of variation (c_x) is not negligible, at what level of α will $\bar{y}_{aaap}$ minimizes over estimation of $\bar{y}_p$ on $\bar{y}$ ".

Table 4:- Values of

$$\frac{[c^2{}_y(1-\alpha^2(\frac{\bar{X}}{\bar{X}+c_x})^2)-\alpha^2(\frac{\bar{X}}{\bar{X}+c_x})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}}{\bar{X}+c_x})^2}$$

when $0.1 \le \alpha \le 1.0$ obtained on $\bar{y}_{aaap}$

α / pop	I	II	III
1.0	**-0.2149**	-0.4819	**-0.4394**
0.9	**0.0564**	-0.3603	**-0.3062**
0.8	**0.4369**	-0.1904	**-0.1200**
0.7	**0.9890**	**0.0574**	**0.1517**
0.6	**1.8414**	**0.4393**	**0.5702**
0.5	**3.2551**	**1.0725**	**1.2642**
0.4	**5.8576**	**2.2383**	**2.5419**
0.3	**11.4804**	**4.7571**	**5.3024**
0.2	**27.5453**	**11.9534**	**13.1896**
0.1	**114.2960**	**50.8135**	**55.7801**
ρ_{xy}	-0.6823	-0.05	-0.9199

Here, the effect of over estimation of $\bar{y}_p$ on $\bar{y}$ felt in population II shown in Table 2 above when

$$\rho_{xy} > \frac{[c^2{}_y(1-\alpha^2(\frac{\bar{X}}{\bar{X}+c_x})^2)-\alpha^2(\frac{\bar{X}}{\bar{X}+c_x})^2 c^2{}_x]}{2c_y c_x \alpha^2(\frac{\bar{X}}{\bar{X}+c_x})^2}$$

$, 0.7 \le \alpha \le 1.0$ (as shown in Table 4) will be minimized whenever

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

$0.1 \le \alpha < 0.7$

4.0 DISCUSSION

Table 2 shows that the effect of over estimation of $\bar{y}_p$ on $\bar{y}$ is felt for population II while in Tables 3 and 4, one can see that this effect of over estimation will be minimized by $\bar{y}_{aaap}$ whenever $0.1 \le \alpha < 0.7$

5.0 CONCLUSION

In conclusion therefore, when the known auxiliary variable coefficient of variation (c_x) is not negligible, Adewara ([Ade16]) product estimator, $\bar{y}_{aaap}$, may be used to minimize the effect of over estimation of $\bar{y}_p$ on $\bar{y}$ whenever $0.1 \le \alpha < 0.7$.

REFERENCES

[Ade16] **A. A. Adewara** - *Improved Product Estimators Using Known Values of Some Population Parameters*. Journal of the Nigerian Association of Mathematical Physics, 35, pp. 445 – 450, 2016.

[Coc77] **W. G. Cochran** - *Sampling Techniques*. 3rd Edition. John Wiley, New York, 1977.

[DT80] **A. K. Das, T. P. Tripathi** - *Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known*. Sankhya, C, 42, pp. 76 -86, 1980.

[Mad77] **G. S. Maddala** - *Econometrics*. McGraw Hills Pub. Co. New York, 1977.

[Mur67] **M. N. Murthy** - *Sampling theory and Methods*. Statistical Publishing Society, Calcuta, India, 1967.

[PD88] **B. N. Pandey, V. Dubey** - *Modified product estimator using coefficient of variation of auxiliary variate*. Assam Statistical Review, 2, part 2, pp. 64 – 66, 1988.

[SD81] **B. V. Sisoda, V. K. Dwivedi** - *A modified ratio estimator using coefficient of variation of auxiliary variable*. Journal India Society of Agricultural Statistics, New Delhi, 33, pp. 13 – 18, 1981.

[ST05] **H. P. Singh, R. Tailor** - *Estimation of finite population mean with known coefficient of variation of an auxiliary character*. STATISTICA, anno LXV, n, 3, pp. 301 – 313, 2005.

[SU86] **H. P. Singh, L. N. Upadhyaya** - *A dual to modified ratio estimator using coefficient of variation of auxiliary variable*. Proceedings National Academy of Sciences, India, 56, A, part 4, pp. 336 – 340, 1986.