ISSN 1684-8403
Journal of Statistics
Volume 22, 2015. pp. 1-8

A Modified Class of Ratio and Product Estimators of the Population Mean in Simple Random Sampling using Information on Auxiliary Variable

Emmanuel John Ekpenyong [1] *and Ekaete Inyang Enang* [2]

Abstract

In sample surveys, information on auxiliary variables is usually used to improve on designs, precision and efficiency of Estimators. A modified class of Ratio Estimators is proposed in this paper, while a new class of Product Estimators is suggested to estimate the population mean. The efficiencies of these Estimators are compared and the most efficient Estimator, which is also unbiased, is identified among these Estimators. Numerical illustration is employed to validate this claim.

Keywords

Ratio estimator, Efficiency, Simple random sampling, Bias, Mean square error, Auxiliary variable

1. Introduction

It is a well-known fact that suitable use of auxiliary information results in a considerable improvement of the precision and accuracy of the Estimator. In this regard, Ratio, Product and Regression Estimators are widely employed if there is a high correlation between the variable of interest y and the auxiliary variable x. It has been established that, in general, the regression Estimator is more efficient than the Ratio and Product Estimators, except when the Regression equation of the variable under study on the auxiliary variable passes through the origin, in which case, the efficiencies of the Estimators are almost equal.

[1] Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria.
Email: ekpesstat@yahoo.com
[2] Department of Mathematics/Statistics/Computer Science, University of Calabar, Calabar, Nigeria.
Email: ekkaass@yahoo.com

In view of this, extensive work has been done in the direction of improving the performance of the classical Ratio and Product Estimators.

This paper modifies a class of Ratio Estimators proposed by Kadilar and Cingi (2004, 2006) and further proposes a class of Product Estimators from the modified classes of Ratio Estimators. Kadilar and Cingi (2004, 2006) used information on auxiliary and study variables such as correlation coefficient, coefficient of variation, coefficient of Kurtosis, etc, to develop a more efficient class of Ratio Estimators. Koyuncu and Kadilar (2009) also suggested a family of Estimators with minimum mean square errors. Many contributions on the improvement of Ratio and Product Estimators can be found in the works of Singh (2003), Singh and Tailor (2003), Upadhyaya and Singh (1999) and many others.

2. Modified class of Estimtors

Adapted from Kadilar and Cingi (2004, 2006), the proposed class of Estimators is given by:

$$\bar{y}_m = \left[\bar{y} + \lambda(\bar{X} - \bar{x})\right]\left[\frac{a\bar{X} + k}{a\bar{x} + k}\right]^{\alpha} \tag{2.1}$$

where

$\bar{y}$ is the sample mean of the variable of interest Y, λ may be some information of the auxiliary variable or functions of information of the auxiliary variable, a and k are real constants or information of auxiliary variable X and α can be -1 or 1. It should be noted that if $\alpha = 1$, then the Estimator is a Ratio Estimator; and if $\alpha = -1$, then the Estimator becomes a Product Estimator.

To obtain the Bias and Mean Square Error of the class of Estimators, we let:

$$e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}},\ e_1 = \frac{\bar{x} - \bar{X}}{\bar{X}},\ E(e_0^2) = \frac{1-f}{n}C_y^2, E(e_1^2) = \frac{1-f}{n}C_x^2,\ E(e_0e_1) = \frac{1-f}{n}\rho C_y C_x,\ f = \frac{n}{N} \tag{2.2}$$

$C_x = \frac{S_x}{\bar{X}}$, the coefficient of variation of the auxiliary variable

$C_y = \frac{S_y}{\bar{Y}}$, the coefficient of variation of the study variable

$\rho = \frac{S_{xy}}{S_x S_y}$, the correlation coefficient between auxiliary and study variables

${S_x}^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$, population variance of auxiliary variable

${S_y}^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$, population variance of study variable

$S_{xy} = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})$, population covariance between the auxiliary and study variables

where
$\bar{y}$ and $\bar{Y}$ are sample and population means of the character of interest, $\bar{x}$ and $\bar{X}$ are sample and population means of auxiliary information, C_y *and* C_x are coefficients of variation of Y and X respectively and ρ is the correlation coefficient between X and Y.

Using equation (2.2), (2.1) can be transformed as:

$$\bar{y}_m = \left[\bar{Y}\left(1+e_0\right) - \lambda\bar{X}e_1\right]\left[1+\gamma e_1\right]^{-\alpha} \tag{2.3}$$

where

$$\gamma = \frac{a\bar{X}}{a\bar{X}+k}$$

By Taylor's series approximation up to order 2 of $\left[1+\gamma e_1\right]^{-\alpha}$ in equation (2.3) and simplifying, we have

$$\bar{y}_m \simeq \bar{Y}\left[1-\alpha\gamma e_1 + \frac{\alpha(\alpha+1)}{2}\gamma^2 e_1^2 + e_0 - \alpha\gamma e_0 e_1 - \frac{\lambda\bar{X}}{\bar{Y}}e_1 + \frac{\lambda\bar{X}}{\bar{Y}}\alpha\gamma e_1^2\right] \tag{2.4}$$

Therefore, the Bias of the Estimator is,

$$B(\bar{y}_m) = E(\bar{y}_m - \bar{Y}) = \frac{1-f}{n}\bar{Y}\left\{\left[\frac{\alpha(\alpha+1)}{2}\gamma^2 + \frac{\lambda\bar{X}}{\bar{Y}}\alpha\gamma\right]C_x^2 - \alpha\gamma\rho C_x C_y\right\} \tag{2.5}$$

It's Mean Square Error is obtained by Taylor's series approximation as:

$$MSE(\bar{y}_m) = E(\bar{y}_m - \bar{Y})^2 = E\left[\bar{Y}\left(e_0 - \left(\alpha\gamma + \frac{\lambda\bar{X}}{\bar{Y}}\right)e_1\right)^2\right]$$

$$\simeq \bar{Y}^2\frac{1-f}{n}\left[C_y^2 - 2\alpha\gamma\rho C_x C_y - 2\lambda\frac{\bar{X}}{\bar{Y}}\rho C_x C_y + \alpha^2\gamma^2 C_x^2 + 2\alpha\gamma\lambda\frac{\bar{X}}{\bar{Y}}C_x^2 + \frac{\lambda^2\bar{X}^2}{\bar{Y}^2}C_x^2\right]$$

$$\cong \frac{1-f}{n}\bar{Y}^2[C_y^2 - 2(\alpha\gamma+\lambda M)\rho C_y C_x + (\alpha\gamma+\lambda M)^2 C_x^2],\ M = \frac{\bar{X}}{\bar{Y}} = \frac{1}{R} \tag{2.6}$$

To obtain the value of λ that would make equation (2.6) a minimum, we differentiate equation (2.6) with respect to λ and obtain:

$$\lambda_{opt} = \lambda^* = \frac{\bar{Y}\rho C_y}{\bar{X}C_x} - \frac{\bar{Y}\alpha\gamma}{\bar{X}} = R\left(\frac{\rho C_y}{C_x} - \alpha\gamma\right) = B - R\alpha\gamma \tag{2.7}$$

where $R = \frac{\bar{Y}}{\bar{X}}$

From equation (2.7), the optimum value of λ is a function of γ and α.

Substitution of equation (2.7) in (2.6) gives the optimum Mean Square Error of the class of Estimators as:

$$MSE(\bar{y}_m^*) = \frac{1-f}{n}\bar{Y}^2 C_y^2 (1-\rho^2) \tag{2.8}$$

It is clearly seen that the Mean Square Error of the class of Estimators at optimum λ is the same as that of the Regression Estimator of the population mean. Moreover, the members of this class at optimum λ are almost unbiased. This can be obtained by substituting equation (2.7) into (2.5). Table 1 shows some of the members of the class of Ratio and Product Estimators.

3. Efficiency comparison

The efficiencies of the members of this class of Ratio Estimators are compared and evaluated.

One member of the class $\bar{y}_{mi}$ is more efficient than the other $\bar{y}_{mj}$ if:

$$MSE(\bar{y}_{mi}) < MSE(\bar{y}_{mj})$$

$$\Rightarrow \frac{1-f}{n}\bar{Y}^2[C_y^2 - 2(\alpha\gamma_i + \lambda_i M)\rho C_y C_x + (\alpha\gamma_i + \lambda_i M)^2 C_x^2] < \frac{1-f}{n}\bar{Y}^2[C_y^2 -$$
$$2(\alpha\gamma_j + \lambda_j M)\rho C_y C_x + (\alpha\gamma_j + \lambda_j M)^2 C_x^2]$$

$$\Rightarrow -2[(\alpha\gamma_i + \lambda_i M) - (\alpha\gamma_j + \lambda_j M)]C + [(\alpha\gamma_i + \lambda_i M)^2 - (\alpha\gamma_j + \lambda_j M)^2] < 0,$$

where $C = \frac{\rho C_y}{C_x}$

$$\Rightarrow [(\alpha\gamma_i + \lambda_i M) - (\alpha\gamma_j + \lambda_j M)][(\alpha\gamma_i + \lambda_i M) + (\alpha\gamma_j + \lambda_j M) - 2C] < 0$$

$$\Rightarrow \textit{either}\ \ \alpha(\gamma_i - \gamma_j) + M(\lambda_i - \lambda_j) < 0 \text{ and } \alpha(\gamma_i + \gamma_j) + M(\lambda_i + \lambda_j) - 2C > 0$$

$$\text{Or}\quad \alpha(\gamma_i - \gamma_j) + R(\lambda_i - \lambda_j) > 0 \text{ and } \alpha(\gamma_i + \gamma_j) + R(\lambda_i + \lambda_j) - 2C < 0$$

$$\Rightarrow \textit{either}\ \ \alpha(\gamma_i - \gamma_j) < -M(\lambda_i - \lambda_j) \text{ and } \alpha(\gamma_i + \gamma_j) > 2C - M(\lambda_i + \lambda_j)$$

$$\text{Or}\quad \alpha(\gamma_i - \gamma_j) > -M(\lambda_i - \lambda_j) \text{ and } \alpha(\gamma_i + \gamma_j) < 2C - M(\lambda_i + \lambda_j) \tag{3.1}$$

4. Numerical illustration

The data sets in Table 2 used by some authors illustrate the efficiencies and performances of these Estimators. According to results given in Table 3, the performance of Estimators depend on population, except the case where λ is minimized. In population I, $\bar{y}_{mopt}$ is the most efficient Estimator, followed by Upadhyaya and Singh (1999) Estimator and the third is the simple sample mean Estimator. Population II has $\bar{y}_{mopt}$ as the best Estimator, which is followed by Kadilar and Cingi (2006) Estimator, $\bar{y}_{m10}$. Optimum Estimator is also the best one for Populations III and IV followed by the Estimator proposed by Singh and Tailor (2003) and sample mean.

5. Conclusion

From Table 1, it is observed that the modified class of Ratio Estimators proposed can produce different types of Ratio and Product Estimators, when different information on the auxiliary variables is being substituted. Some members of this class of Ratio Estimators include the Kadilar and Cingi (2004, 2006) types of Ratio Estimators, the simple sample mean, the classical Ratio Estimator and many others. In addition, Table 3 presents the numerical illustration of the efficiency of some members of this class of Estimators. The results show that the most efficient and unbiased member of this class is obtained when the optimum value of λ is substituted into the modified Ratio Estimator given in $\bar{y}_{m10}$. It is also noted that the performances of the members of the class of Estimators depend on population, except at optimum parameter λ. A class of Product Estimators, which was not suggested by Kadilar and Cingi (2006), is also proposed. Therefore, $\bar{y}_{m10opt}$ becomes the most efficient Estimator of the proposed family of Ratio Estimators. For further research, conditions for variation in performances of members of this class of Estimators based on populations could be theoretically and numerically established.

Table 1: Some members of the class of Ratio and Product Estimators

Ratio Estimators				**Product Estimators**			
a	k	λ	$\alpha = 1$	$\alpha = -1$	a	k	λ
0	1	0	$\bar{y}_{m1} = \bar{y}$ Sample mean	$\bar{y}_{m1} = \bar{y}$ Sample mean	0	1	0
1	0	0	$\bar{y}_{m2} = \frac{\bar{y}}{\bar{x}}\bar{X}$ Classical Ratio Estimator	$\bar{y}_{m2} = \frac{\bar{y}\bar{x}}{\bar{X}}$ Classical Product Estimator	1	0	0
1	ρ	0	$\bar{y}_{m3} = \frac{\bar{y}}{\bar{x}+\rho}(\bar{X}+\rho)$ Singh and Tailor (2003)	$\bar{y}_{m3} = \frac{\bar{y}}{\bar{X}+\rho}(\bar{x}+\rho)$ Singh and Tailor (2003)	1	ρ	0
$\beta_2(x)$	C_x	0	$\bar{y}_{m4} = \frac{\bar{y}}{\bar{x}\beta_2(x)+C_x}(\bar{X}\beta_2(x)+C_x)$ Upadhyaya and Singh (1999)	$\bar{y}_{m4} = \frac{\bar{y}}{\bar{x}\beta_2(x)+C_x}(\bar{X}\beta_2(x)+C_x)$ Upadhyaya and Singh (1999)	$\beta_2(x)$	C_x	0
1	0	b	$\bar{y}_{m5} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}}\bar{X}$ Kadilar and Cingi (2004)	$\bar{y}_{m5} = \frac{\bar{y}+b(\bar{x}-\bar{X})}{\bar{X}}\bar{x}$	1	0	-b
1	ρ	b	$\bar{y}_{m6} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}+\rho}(\bar{X}+\rho)$ Kadilar and Cingi (2006)	$\bar{y}_{m6} = \frac{\bar{y}+b(\bar{x}-\bar{X})}{\bar{X}+\rho}(\bar{x}+\rho)$	1	ρ	-b
1	C_x	b	$\bar{y}_{m7} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}+C_x}(\bar{X}+C_x)$ Kadilar and Cingi (2004)	$\bar{y}_{m7} = \frac{\bar{y}+b(\bar{x}-\bar{X})}{\bar{X}+C_x}(\bar{x}+C_x)$	1	C_x	-b
$\beta_2(x)$	C_x	b	$\bar{y}_{m8} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}\beta_2(x)+C_x}(\bar{X}\beta_2(x)+C_x)$ Kadilar and Cingi (2004)	$\bar{y}_{m8} = \frac{\bar{y}+b(\bar{x}-\bar{X})}{\bar{X}\beta_2(x)+C_x}(\bar{x}\beta_2(x)+C_x)$	$\beta_2(x)$	C_x	-b
C_x	ρ	b	$\bar{y}_{m9} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}C_x+\rho}(\bar{X}C_x+\rho)$ Kadilar and Cingi (2006)	$\bar{y}_{m9} = \frac{\bar{y}+b(\bar{x}-\bar{X})}{\bar{X}C_x+\rho}(\bar{x}C_x+\rho)$	C_x	ρ	-b

Ratio Estimator				**Product Estimator**			
ρ	$\beta_2(x)$	b	$\bar{y}_{m10} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}\rho+\beta_2(x)}(\bar{X}\rho+\beta_2(x))$ Kadilar and Cingi (2006)	$\bar{y}_{m10} = \frac{\bar{y}+b(\bar{x}-\bar{X})}{\bar{X}\rho+\beta_2(x)}(\bar{x}\rho+\beta_2(x))$	ρ	$\beta_2(x)$	-b
a	k	λ_{opt}	$\bar{y}_{mopt} = \frac{\bar{y}+\lambda_{opt}(\bar{X}-\bar{x})}{a\bar{x}+k}(a\bar{X}+k)$	$\bar{y}_{mopt} = \frac{\bar{y}+\lambda_{opt}(\bar{x}-\bar{X})}{a\bar{X}+k}(a\bar{x}+k)$	a	k	- λ_{opt}

Table 2: Data sets

Source	**Population**	**N**	**n**	$\bar{Y}$	$\bar{X}$	C_y	C_x	ρ	$\beta_2(x)$
Murthy (1967)	I	80	20	51.8264	11.2646	0.3542	0.7507	0.9413	0.06339
Kadilar and Cingi (2006)	II	200	50	500	25	15	2	0.9	50
Koyuncuand Kadilar (2009)	III	923	180	436.4345	11440.5	1.7183	1.8645	0.9543	18.7208
Murthy (1967)	IV	80	20	51.8264	2.8513	0.3542	0.9484	0.915	1.3005

Table3: MSE's of some Ratio-type members of the classes

Estimator	**Population I**	**Population II**	**Population III**	**Population IV**
$\bar{y}_{m1}$	12.63661	843750	2515.074	12.63661
$\bar{y}_{m2}$	18.979306	656250	267.64407	41.315067
$\bar{y}_{m3}$	14.450269	662262.321	267.58455	17.6849057
$\bar{y}_{m4}$	1.54673417	656525.597	267.637855	20.7801102
$\bar{y}_{m5}$	58.20310614	175312.5	3185.891797	92.65447161
$\bar{y}_{m6}$	49.7857254	174288.141	3185.39784	53.981439
$\bar{y}_{m7}$	51.3317244	173172.582	3184.88927	53.0725982
$\bar{y}_{m8}$	14.92976825	175264.615	3185.802687	59.50835213
$\bar{y}_{m9}$	47.40141771	174786.7378	3185.589301	52.63550853
$\bar{y}_{m10}$	57.53046	161757.2	3175.725	42.40431
$\bar{y}_{mopt}$	1.439996	160312.5	224.625	2.056924

References

1. Kadilar, C. and Cingi, H. (2004). Ratio Estimators in Simple Random Sampling. *Applied Mathematics and Computation*, **151**, 893-902.
2. Kadilar, C. and Cingi, H. (2006). A new Ratio Estimator using correlation coefficient. *Staistics on Internet*, 1-11.
   Retrieved from:
   http://interstat.statjournals.net/YEAR/2006/articles/0603004.pdf.
3. Koyuncu, N. and Kadilar, C. (2009). Efficient Estimators for the population mean. *Hacettepe Journal of Mathematics and Statistics*, **38(2)**, 217-225.
4. Murthy, M. N. (1967). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta, India.
5. Singh, G. N. (2003). On the Improvement of Product Method of Estimation in Sample Surveys. *Journal of Indian Society and Agricultural Statistics*, **56(3)**, 265-267.
6. Singh, H. P. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, **6**, 555-560.
7. Upadhyaya, L. N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, **41**, 627-636.