STATISTICS IN TRANSITION new series, December 2019
Vol. 20, No. 4, pp. 181–189, DOI 10.21307/stattrans-2019-041
Submitted – 18.01.2018; Paper ready for publication – 16.08.2018

A NEW RATIO ESTIMATOR: AN ALTERNATIVE TO REGRESSION ESTIMATOR IN SURVEY SAMPLING USING AUXILIARY INFORMATION

Mir Subzar[1], S. Maqbool[2], T. A. Raja[2], Prayas Sharma[3]

ABSTRACT

The most dominant problem in the survey sampling is to obtain the better ratio estimators for the estimation of population mean or population variance. Estimation theory is enhanced by using the auxiliary information in order to improve on designs, precision and efficiency of estimators. A modified class of ratio estimator is suggested in this paper to estimate the population mean. Expressions for the bias and the mean square error of the proposed estimators are obtained. Both analytical and numerical comparison has shown the suggested estimator to be more efficient than some existing ones. The bias of the suggested estimator is also found to be negligible for the population under consideration, indicating that the estimator is as good the regression estimator and better than the other estimators under consideration.

Key words: ratio type estimators, auxiliary information, bias, mean square error, simple random sampling, efficiency.

AMS Subject Classification: 62D05

1. Introduction

In sample surveys, auxiliary information on the finite population under study is quite often available from previous experience, census or administrative databases. The sampling literature describes different procedures for using auxiliary information to improve the sampling design and/or obtain more efficient estimators. The use of auxiliary information at the estimation stage has been dealt at great dealt at great length for improving estimation in sample surveys. In sample surveys, auxiliary information is used at selection as well as estimation stages to improve the design as well as obtaining more efficient estimators. Increased precision can be obtained when the study variable Y is highly correlated with auxiliary variable X .

[1] Division of Agricultural Statistics, SKUAST-Kashmir (190025), India. E-mail: subzarstat@gmail.com.
[2] Division of Agricultural Statistics, SKUAST-Kashmir (190025), India.
[3] Department of Decision Sciences, University of Petroleum and Engineering Studies, Dehradun. E-mail: prayassharma02@gmail.com.

Usually, in a class of efficient estimators, the estimators with minimum variance or mean square error is regarded as the most efficient estimator. A good estimator can also be described by the value of its bias. An estimator with minimum absolute bias is regarded as a better estimator among others in the class (Rajesh et al., 2011).

When the population mean of an auxiliary variable is known, so many estimators for the population parameters of study variable have been discussed in literature. The literature on survey sampling describes a great variety of techniques for using auxiliary information by means of ratio, product and regression methods.

If the regression line of the character of interest Y on the auxiliary variable, X is through the origin and when correlation between study and auxiliary variables is positive (high), then the ratio estimate of mean or total may be used (Cochran 1940).

On the other hand, if the regression line used for the estimate does not pass through the origin but makes an intercept along the y-axis, the regression estimation is used (Okafor, 2002). Furthermore, when correlation between study variable on the auxiliary variable passes through a suitable neighbourhood of the origin, in which case, the efficiencies of these estimators are almost equal. When the population parameters of the auxiliary variable X such as population mean, coefficient of variation, coefficient of kurtosis, coefficient of skewness, median are known, a number of modified estimators such as modified ratio estimators, modified product estimators and modified linear regression estimators have been proposed and is widely acceptable in the literature.

This paper is another attempt in solving this problem. An alternative ratio estimator for population mean of the study variable Y (see Sharma and Singh (2014,15), which is more efficient than some of the existing estimators is suggested using the information on one auxiliary variable, X, that is highly correlated with the study variable.

2. Review of the existing estimators

To enhance effective comparison, we summarize below some existing estimators, their biases and mean square errors.

Consider a finite population of N distinct and identifiable units $G = \{G_1, G_2, G_3, ..., G_N\}$. Let a sample of size n be drawn from the population by simple random sampling without replacement. Suppose that interest is to obtain a ratio estimate of the mean of a random variable Y from the sample using a related variable X as supplementary information and assuming that the total of X is known from sources outside the survey.

Table 1. Some existing estimators, their biases and mean square errors

S/N	Estimator	Bias	Mean square error
1	$\bar{y}$	0	$\frac{1-f}{n}\bar{Y}^2 C_y^2$
2	$\bar{y}_{cl} = \frac{\bar{y}}{\bar{x}}\bar{X}$ Classical ratio	$\frac{1-f}{n}\bar{Y}[C_x^2 - \rho C_x C_y]$	$\frac{1-f}{n}\bar{Y}^2[C_y^2 + C_x^2 - 2\rho C_x C_y]$
3	$\bar{y}_{SK} = \frac{\bar{y}}{\bar{x}+M_d}\bar{X} + M_d$ Subramani and Kumarapandiyan	$\frac{1-f}{n}\bar{Y}[aC_x^2 - \rho a C_x C_y]$	$\frac{1-f}{n}\bar{Y}^2[C_y^2 + C_x^2 a(a-2\theta)]$
4	$\bar{y}_{KC} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}}\bar{X}$ Kadilar and Cingi	$\frac{1-f}{n}\bar{Y}C_x^2$	$\frac{1-f}{n}\bar{Y}^2[C_x^2 + C_y^2(1-\rho^2)]$
5	$\bar{y}_{reg} = \bar{y} + b(\bar{X}-\bar{x})$ Regression estimator	0	$\frac{1-f}{n}\bar{Y}^2 C_y^2(1-\rho^2)$

Where

$C_x = \frac{s_x}{\bar{X}}$; $C_y = \frac{s_y}{\bar{Y}}$; the coefficients of variation of the auxiliary variable, X and the response variable, Y ;

$\rho = \frac{Sxy}{S_x S_y}$; the correlation coefficient between X and Y ;

$a = \frac{\bar{X}}{\bar{X}+M_d}$, $m = \frac{\bar{X}}{\bar{Y}}$; $B = \frac{S_{xy}}{S_x^2}$; the regression coefficient;

$\theta = \frac{\rho C_y}{C_x}$ and $f = \frac{n}{N}$, where $S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$,

$S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$; the population variances of the auxiliary and study variables respectively;

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

$\bar{X} = N^{-1}\sum_{i=1}^{N} x_i$, $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$; population means of the auxiliary and study variables;

$\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$, $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$; sample means of the auxiliary and study variables are respectively defined wherever they appear.

3. Suggested estimator

The proposed ratio estimator is obtained by forming linear combination of Subramani and Kumarapandiyan (2012) and Kadilar and Cingi (2004) estimators as shown below:

$$\bar{y}_{pr} = \frac{\alpha\bar{y}(\bar{X} + M_d)}{\bar{x} + M_d} + \frac{\beta(\bar{y} - b(\bar{X} + \bar{x}))}{\bar{x}}\bar{X} \qquad (1)$$

Such that $\alpha + \beta = 1$

3.1. Bias and mean square error of the proposed estimator

To obtain the approximate expression for the bias and the mean squared error for the proposed ratio estimator, let

$$\bar{x} = \bar{X}(1 + e_x); \quad \bar{y} = \bar{Y}(1 + e_y). \qquad (2)$$

Where

$$e_x = \frac{\bar{x} - \bar{X}}{\bar{X}}, \quad e_y = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$$

So that,

$$E(e_x) = E(e_y) = 0, \quad E(e_x^2) = \frac{(1-f)}{n}C_x^2, \quad E(e_y^2) = \frac{(1-f)}{n}C_y^2 \qquad (3)$$

$$E(e_x e_y) = \frac{(1-f)}{n}\rho C_x C_y = \frac{(1-f)}{n}\theta C_x^2$$

Therefore, expressing (1) in terms of (2), we obtain

$$\bar{y}_{pr} = \frac{\alpha\bar{Y}(1+e_y)(\bar{X} + \beta_1)}{\bar{X}(1+e_x) + \beta_1} + (1-\alpha)\left[\frac{(\bar{Y}(1+e_y) + b[\bar{X} - \bar{X}(1+e_x))]}{\bar{X}(1+e_x)}\bar{X}\right]$$

$$= \frac{\alpha\bar{Y}(1+e_y)(\bar{X} + \beta_1)}{(\bar{X} + \beta_1) + \bar{X}e_x} + (1-\alpha)\left[\bar{Y}(1+e_y)(1+e_x)^{-1} - b\bar{X}e_x(1+e_x)^{-1}\right]$$

$$= \alpha\bar{Y}(1+e_y)(1+ae_x)^{-1} + \bar{Y}(1-e_x+e_x^2+e_y-e_ye_x) - \bar{Y}(\alpha-\alpha e_x+\alpha e_x^2+\alpha e_y-\alpha e_ye_x)$$
$$-b\bar{X}e_x + b\bar{X}e_x^2 + ab\bar{X}e_x - ab\bar{X}e_x^2$$

By Taylor Series approximation up to order 2, the expression becomes

$$\bar{y}_{pr} = \alpha\bar{Y}(1-ae_x+z^2e_x^2+e_y-we_ye_x) + \bar{Y}(1-e_x+e_x^2+e_y-e_ye_x)$$
$$+\bar{Y}(\alpha e_x-\alpha-\alpha e_x^2-\alpha e_y+\alpha e_ye_x) - B\frac{\bar{X}}{\bar{Y}}e_x + B\frac{\bar{X}}{\bar{Y}}e_x^2 + \alpha B\frac{\bar{X}}{\bar{Y}}e_x - \alpha B\frac{\bar{X}}{\bar{Y}}e_x^2$$

$$= \bar{Y}[\alpha-\alpha ae_x+\alpha a^2e_x^2+\alpha e_y-\alpha we_ye_x+1-e_x+e_x^2+e_y-e_ye_x-\alpha+\alpha e_x$$
$$-\alpha e_x^2-\alpha e_y+\alpha e_ye_x - B\frac{\bar{X}}{\bar{Y}}e_x + B\frac{\bar{X}}{\bar{Y}}e_x^2 + \alpha B\frac{\bar{X}}{\bar{Y}}e_x - \alpha B\frac{\bar{X}}{\bar{Y}}e_x^2]$$

The expression for the bias of this estimator to first order approximation is obtained as follows:

$$B(\bar{y}_{pr}) = E(\bar{y}_{pr} - \bar{Y})$$

$$= E[\bar{Y}(1+e_y+(\alpha BM-BM-\alpha-\alpha a)e_x+(\alpha-\alpha a-1)e_ye_x+(\alpha a^2+1-\alpha+BM-\alpha BM)e_x^2-\bar{Y})]$$
$$= \frac{1-f}{n}\bar{Y}[(\alpha-\alpha a-1)\rho C_yC_x+(\alpha a^2+1-\alpha+BM-\alpha BK)C_x^2 \quad (4)$$

$$MSE(\bar{y}_{pr}) = E(\bar{y}_{pr} - \bar{Y})^2$$

$$= E[\bar{Y}(1+e_y+(\alpha BM-BM-\alpha-\alpha a)e_x+(\alpha-\alpha a-1)e_ye_x+(\alpha a^2+1-\alpha+BM-\alpha BM)e_x^2-\bar{Y})]^2$$
$$= \frac{1-f}{n}\bar{Y}^2[C_y^2+2(\alpha BM-BM-\alpha-\alpha a)\rho C_yC_x+(\alpha BM-BM-\alpha-\alpha a)^2C_x^2] \quad (5)$$

3.2. Optimal conditions for the proposed estimator

To obtain the value of α that minimizes the MSE, we take partial derivative of equation (5) with respect to α and equate to zero as follows:

$$\frac{\partial MSE(\bar{y}_{pr})}{\partial\alpha} = \frac{1-f}{n}\bar{Y}^2[C_y^2+2(\alpha BM-BM-\alpha-\alpha a)\rho C_yC_x+(\alpha BM-BM-\alpha-\alpha a)^2C_x^2]=0$$
$$\Rightarrow \rho C_xC_y + \alpha(BM+1-a)C_x^2 - (BM+1)C_x^2 = 0$$

$$\Rightarrow \alpha = \frac{(BM+1)C_x^2 - \rho C_xC_y}{(BM+1-a)C_x^2} \quad (6)$$

Substituting for (6) in (5) gives the optimal MSE for $\bar{y}_{pr}$ as:

$$MSE(\bar{y}_{pr}) = \frac{1-f}{n}\bar{Y}^2[C_y^2(1-\rho^2)] \tag{7}$$

4. Efficiency comparison

In order to compare the efficiency of the various existing estimators with that of proposed estimators, we require the expressions of mean square error of these estimators, up to first order approximation. An analytical comparison of the proposed estimator with three of the existing estimators namely: the classical, Subramani and Kumarapandiyan (2012) and Kadilar and Cingi (2004) estimators are carried out.

4.1. Efficiency comparison of proposed and classical

In this section, the analytical condition under which the proposed estimator will be more efficient than classical ratio estimator is established.

$$\begin{aligned} MSE(\bar{y}_{pr}) - MSE(\bar{y}_{cl}) &= \frac{1-f}{n}\bar{Y}^2[C_y^2(1-\rho^2)] - \frac{1-f}{n}\bar{Y}^2[C_y^2 + C_x^2 - 2\rho C_x C_y] \\ &= \frac{1-f}{n}\bar{Y}^2(C_y^2 - C_y^2\rho^2 - C_y^2 - C_x^2 + 2\rho C_x C_y] \\ &= \frac{1-f}{n}\bar{Y}^2(2\rho C_x C_y - C_x^2 - C_y^2\rho^2) \\ &= -\left[\frac{1-f}{n}\bar{Y}^2(C_y\rho - C_x)^2\right] \end{aligned} \tag{8}$$

Since the expression in the square bracket is always positive, we conclude that the proposed estimator will always be more efficient than the classical ratio estimator.

4.2. Efficiency comparison of proposed and Subramani and Kumarapandiyan

$$\begin{aligned} MSE(\bar{y}_{pr}) - MSE(\bar{y}_{SK}) &= \frac{1-f}{n}\bar{Y}^2[C_y^2(1-\rho^2)] - \frac{1-f}{n}\bar{Y}^2[C_y^2 + C_x^2 a(a-2\theta)] \\ &= \frac{1-f}{n}\bar{Y}^2[C_y^2 - C_y^2 - C_x^2 a\left(a - \frac{2\rho C_y}{C_x}\right)] \\ &= \frac{1-f}{n}\bar{Y}^2[-C_y^2\rho^2 - C_x^2 a\left(a - \frac{2\rho C_y}{C_x}\right)] \end{aligned}$$

$$= -\left\{\frac{1-f}{n}\bar{Y}^2[C_x^2\rho^2 + C_x^2 a(a-2\theta)]\right\} \tag{9}$$

Therefore, for the proposed estimator to be more efficient than Yan and Tian (2010), the terms in the second bracket must be positive. This implies that:

$$C_x^2\rho^2 + C_x^2 a(a-2\theta) > 0 \tag{10}$$

4.3. Efficiency comparison of proposed and Kadilar and Cingi

$$MSE(\bar{y}_{pr}) - MSE(\bar{y}_{KC}) = \frac{1-f}{n}\bar{Y}^2[C_y^2(1-\rho^2)] - \frac{1-f}{n}\bar{Y}^2[C_x^2 + C_y^2(1-\rho^2)]$$
$$= -\left[\frac{1-f}{n}\bar{Y}^2 C_x^2\right] \tag{11}$$

Since the expression in the square bracket of equation (11) is always positive, it therefore means that the proposed estimator will always be more efficient than Kadilar and Cingi (2004) estimator of population mean.

5. Numerical comparison

In this section, to study the performance of the estimator presented in this work, we consider empirical population. The source of the population is Singh and Chaudhary (1986) and the values of requisite population parameters are given. We compare the efficiency of the proposed estimator with the existing estimators using the known population data.

Table 2. Data Statistics for population

Parameters	Population	Parameters	Population
N	34	C_x	0.7531
n	20	Md	150
$\bar{Y}$	856.4117	β_1	1.1823
$\bar{X}$	199.4412	$\theta = \rho(C_y/C_x) = BU$	0.50620
ρ	0.4453	$a = \bar{X}/\bar{X} + Md$	0.58204
S_y	733.1407	$M = \bar{X}/\bar{Y}$	0.23288
C_y	0.8561	$B = \rho S_y/S_x$	2.17333
S_x	150.2150	$R = \bar{Y}/\bar{X}$	4.29406

Table 3. Estimators, biases, MSE and % relative efficiency for population.

Population			
Estimator	**MSE**	**Bias of the estimator**	**% Relative efficiency**
$\bar{y}_{cl}$	12557.99	5.658	100.00
$\bar{y}_{SK}$	10236.38	3.293	122.67
$\bar{y}_{KC}$	19977.62	11.457	62.86
$\bar{y}_{pr}$	10165.43	-0.069	123.53

6. Discussion

Optimal mean square error (MSE) of the proposed estimators given in Equation (7) has the same expression as the MSE of the regression estimator which is known to be more efficient than the ratio and the product estimators. The comparison of the suggested estimator with the three existing estimators are derived analytically and these comparisons show that the suggested estimators are more efficient than the classical ratio (1940), Kadilar and Cingi (2004) estimators and preferred over the Subramani and Kumarapandiyan (2012) estimator when the condition stated in the equation (10) is satisfied.

From empirical study, results in the Table 3 reveals that our suggested estimators has lower mean square error than the classical ratio (1940), Kadilar and Cingi (2004) and Subramani and Kumarapandiyan (2012) in the population under consideration, showing that the suggested estimator is more efficient than all the other estimators under consideration. This due to the fact that the suggested estimator is equally as efficient as the regression estimator and confirms Cochran (1940), Robson (1957), Murthy (1967) and Perri (2005) assertion that the regression estimator is generally more efficient than the ratio and product estimators.

Analyses of biases have also shown that the suggested estimator have smallest bias than the all other estimators under consideration. From the Table 3, also from bias point of view, bias is negligible and agrees with the assertion of the Okafor (2002) that any estimator with relative bias less than 10% is considered to have a negligible bias.

7. Conclusion

Since the from the equation (7) the suggested estimator gives the same precision as the regression estimator and is consistently better in terms of bias and efficiency then the three estimators under consideration, the suggested estimator can always be used as an alternative to the regression estimator and gives a better replacement to some existing ratio estimators.

REFERENCES

COCHRAN, W. G., (1940). The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. Journal of Agricultural Science, 30, pp. 262–275.

KADILAR, C., CINGI, H., (2004). Ratio estimators in simple random sampling. Applied Mathematics and Computation, 151, pp. 893–902.

MURTHY, M. N., (1967). Sampling theory and methods, Statistical Publishing Society, Calcutta, India.

SINGH, D., CHAUDHARY, F. S., (1986). Theory and Analysis of Sample Survey Designs, New Age International Publisher.

SUBRAMANI, J., KUMARAPANDIYAN, G., (2012). Estimation of population mean using known median and co-efficient of skewness. American Journal of Mathematics and Statistics, 2(5), pp. 101–107.

ROBSON, D. S., (1957). Application of multivariate Polykays to the theory of unbiased ratio type estimation. Journal of American Statistical Association, 52, pp. 411–422.

OKAFOR, F. C., (2002). Sample Survey Theory with Applications (1st ed.), N sukka, Nigeria, Afro-Orbis.

PERRI, P. F., (2005). Combining two Auxiliary Variables in Ratio-cum-product type Estimators. Proceedings of Italian Statistical Society. Intermediate meeting on Statistics and Environment, Messina, 21-23 September, pp. 193–196.

RAJESH, T., RAJESH, P., JONG-MIND, K., (2011). Ratio-cum-product Estimators of Population mean using known Parameters of Auxiliary Variables. Communications of the Korean Statistical Society, 18(2), pp. 155–164.

SHARMA, P., SINGH, R., (2014). Improved Ratio Type Estimators Using Two Auxiliary Variables under Second Order Approximation, Mathematical Journal of Interdisciplinary Sciences, Vol. 2, No. 2, pp. 193–204.