A Modified Generalized Two-phase Exponential Estimator To Estimate Finite Population Mean

Parikshita Khatua and Gopabandhu Mishra
Department of Statistics, Utkal University,Bhubaneswar-751004, India

Abstract: *Using two-phase sampling mechanism, a modified generalized exponential estimator of finite population mean in the presence of the available knowledge on the auxiliary variable x is suggested i.e. the population mean of auxiliary variable x is not known. The suggested estimator is found to be more efficient than usual two-phase ratio, product, regression estimators and also modified exponential ratio and product estimator suggested by Singh et.al (2007). Further a numerical study is made to support the theoretical comparison.*
Keywords: *Auxiliary variable, Exponential Estimator,Mean Square Error, Two-Phase Sampling,*

I. Introduction

In survey sampling, it is common to estimate finite population mean $\bar{Y}$ of a study variable of y. When information on auxiliary variable x, highly correlated with y, is readily available on all units of population, it is well known that ratio estimator (for high positive correlation), product estimator (for high negative correlation) and regression estimator (for high correlation) can be used to increase the efficiency, incorporating the knowledge of population mean $\bar{X}$ .However, in certain practical situation when the population mean of auxiliary variable $\bar{X}$ is not known a priori, the technique of two-phase sampling fruitfully applied .The usual procedure in such situation, a larger sample of size n' to estimate the population mean $\bar{X}$ while a sub-sample size n is drawn from the first phase sample n' to observe the characteristic under study.

Consider a finite population U={1,2,3, ... , N}. Let y and x be two real variable assuming the value y_i and x_i on the ith unit (i = 1,2,3,…….,N). Now consider y be the study variable and x be the auxiliary variable. We consider simple random sampling without replacement (SRSWOR) design to draw samples in each phase of two-phase sampling set-up. The first phase sample s' ($s' \subset u$) of fixed size n' is drawn to observe x only. The second phase sample s ($s \subset s'$) of fixed size n is drawn to observe y and x for given s, (n<n').

$$Let, \ \bar{x} = \frac{1}{n} \sum_{i \in s} x_i \ , \qquad \bar{y} = \frac{1}{n} \sum_{i \in s} y_i \quad and \qquad \bar{x}' = \frac{1}{n'} \sum_{i \in s'} x_i$$

Now the usual two phase ratio, product and regression estimators are given by ,

$$t_{dr} = \frac{\bar{y}}{\bar{x}} \bar{x}' \qquad\qquad (1.1)$$

$$t_{dp} = \frac{\bar{y} \ \bar{x}}{\bar{x}'} \qquad\qquad (1.2)$$

$$t_{dlr} = \bar{y} + b_{yx}(\bar{x}' - \bar{x}) \qquad\qquad (1.3)$$

where b_{yx} is the sample regression coefficient of 'y' on 'x', calculated from the data based on second phase sample of size 'n'.

The mean square error (MSE) of the estimators given in (1.1), (1.2) and (1.3) to first order of approximation are respectively given by

$$MSE \ (t_{dr}) = \bar{Y}^2[\theta \ C_y^2 + (\theta - \theta')(C_x^2 - 2C_{yx})] \qquad\qquad (1.4)$$

$$MSE \ (t_{dp}) = \bar{Y}^2[\theta \ C_y^2 + (\theta - \theta')(C_x^2 + 2C_{yx})] \qquad\qquad (1.5)$$

$$MSE \ (t_{dlr}) = \bar{Y}^2 C_y^2[\theta \ (1 - \rho^2) + \theta' \rho^2] \qquad\qquad (1.6)$$

where, $\theta = \left(\frac{1}{n} - \frac{1}{N}\right), \quad \theta' = \left(\frac{1}{n'} - \frac{1}{N}\right)$

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2}, \qquad C_x^2 = \frac{S_x^2}{\bar{X}^2}, \qquad C_{yx} = \frac{S_{yx}}{\bar{Y}\bar{X}} = \rho C_y C_x \ ,$$

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2, \qquad S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2,$$

$$S_{yx} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X}) = \rho \ S_y S_x \ .$$

ρ being the population correlation coefficient between 'y' and 'x'.

II.   Two-Phase Exponential Ratio And Product Estimator

Singh et.al(2007) have suggested modified exponential ratio and product estimators to estimate finite population mean $\bar{Y}$ of study variable y in presence of auxiliary variable x. As Singh et.al(2007) assumed that the population mean $\bar{X}$ of auxiliary variable x is not known, they used the two-phase mechanism to estimate the population mean $\bar{Y}$.

The modified exponential ratio and product estimator suggested by Singh et.al (2007) to estimate population mean $\bar{Y}$ under two-phase sampling are given by,

$$t_{der} = \bar{y}.\,exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) , \tag{2.1}$$

$$t_{dep} = \bar{y}.\,exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) , \tag{2.2}$$

The MSE of the estimators t_{der} and t_{dep} to first order of approximation respectively are given by ,

$$MSE\,(t_{der}) = \bar{Y}^2\left[\theta\,C_y^2 + (\theta - \theta')\left(\frac{C_x^2}{4} - C_{yx}\right)\right] \tag{2.3}$$

$$MSE\left(t_{dep}\right) = \bar{Y}^2\left[\theta\,C_y^2 + (\theta - \theta')\left(\frac{C_x^2}{4} + C_{yx}\right)\right] \tag{2.4}$$

III.   Modified Generalized Exponential Estimator

We propose a generalized exponential estimator to estimate finite population mean $\bar{Y}$ under two-phase sampling scheme with a assumption that the population mean of auxiliary variable x , $\bar{X}$ is not known.

$$t_{dge} = \bar{y}[d_1 + d_2(\bar{x}' - \bar{x})]exp\left[\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right] , \tag{3.1}$$

Where $d_1\,\,and\,\,d_2$ are suitable chosen constants or statistics and $d_1 +\,\,d_2$ not necessarily equal to unity.

Define, $\delta_{\bar{y}} = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$, $\delta_{\bar{x}} = \frac{\bar{x} - \bar{X}}{\bar{X}}$ and $\delta_{\bar{x}'} = \frac{\bar{x}' - \bar{X}}{\bar{X}}$

$\Rightarrow E(\delta_{\bar{y}}) =\,\,E(\delta_{\bar{x}}) =\,\,E(\delta_{\bar{x}'}) = 0,$

$E\left(\delta_{\bar{y}}^2\right) = \theta C_y^2$, $E\left(\delta_{\bar{x}}^2\right) = \theta\,C_x^2$, $E(\delta_{\bar{y}},\delta_{\bar{x}}) = \theta\rho C_y C_x$

$E(\delta_{\bar{x}'}^2) = \theta'C_x^2$, $E(\delta_{\bar{x}},\delta_{\bar{x}'}) = \theta'C_x^2$, $E(\delta_{\bar{y}},\delta_{\bar{x}'}) = \theta'\rho C_y C_x$

Now the MSE of the estimator t_{dge} to first order of approximation is given by

$$MSE\left(t_{dge}\right) = E\left(t_{dge} - \bar{Y}\right)^2$$

$$= E\left[\bar{y}\{d_1 + d_2(\bar{x}' - \bar{x})\}\,exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) - \bar{Y}\right]^2$$

$$= \bar{Y}^2 E\left[(d_1 - 1) + d_1\,\delta_{\bar{y}} + \left(\delta_{\bar{x}'} - \delta_{\bar{x}}\right)\left(\frac{1}{2}d_1 + d_2\bar{X}\right)\right]^2$$

$$= \bar{Y}^2\left[(d_1 - 1)^2\,\,+ (\theta - \theta')C_x^2\left(\frac{1}{2}d_1 + d_2\bar{X}\right)^2 + \theta d_1^2 C_y^2\right.$$

$$\left.- (\theta - \theta')C_{yx}(d_1^2 + 2d_1 d_2\bar{X})\right] \tag{3.2}$$

Now the optimum value of $MSE\,(\,t_{dge})$ is obtained substituting the optimum values of $d_1\,\,and\,\,d_2$. Considering the equation (3.2) , using the principles of maxima and minima, we get the optimum values of $d_1\,\,and\,\,d_2$. Let the optimum value of $d_1\,\,and\,\,d_2$ be d_1^* and d_2^* respectively.
Now,

$$d_1^* = \frac{1}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]} \tag{3.3}$$

$$d_2^* = \frac{\left(2\rho\,C_y - C_x\right)}{\left[1 + C_y^2\{\theta - (\theta - \theta')\rho^2\}\right]2\bar{X}\,C_x} \tag{3.4}$$

The optimum $MSE\,\left(t_{dge}\right)$ is obtain by substituting d_1^* and d_2^* in (3.2)

$$MSE\left(t_{dge}\right)_{opt} = \frac{S_y^2[\theta - (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]} \tag{3.5}$$

IV.   Comparison Of Efficiency

(i) Comparison of mean per unit estimator and the two-phase generalized exponential estimator optimum $\left(t_{dge}\right)_{opt}$.

$$V(\bar{y}) - MSE\left(t_{dge}\right)_{opt} = \theta S_y^2 - \frac{S_y^2[\theta - (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}$$

Since, the denominator

$$1 + C_y^2[\theta - (\theta - \theta')\rho^2] \geq 1 \tag{4.1}$$

We consider,

$$\theta S_y^2 - S_y^2[\theta - (\theta - \theta')\rho^2] = (\theta - \theta')\rho^2 \geq 0 \tag{4.2}$$

From, (4.1) and (4.2), we have $MSE\left(t_{dge}\right) \leq V(\bar{y})$.

Hence, the two-phase generalized exponential estimator always more efficient than mean per unit estimator.

(ii) Comparison of two phase ratio estimator (t_{dr}) and two-phase generalized exponential estimator $\left(t_{dge}\right)_{opt}$.

From, (1.4) and (3.5), we consider

$$MSE\ (t_{dr}) - MSE\ \left(t_{dge}\right)$$

$$= \bar{Y}^2\left[\theta\ C_y^2 + (\theta - \theta')\left(C_x^2 - 2C_{yx}\right)\right] - \frac{S_y^2[\theta - (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}$$

Since, $1 + C_y^2[\theta - (\theta - \theta')\rho^2] \geq 1$ \tag{4.3}

We consider,

$$\bar{Y}^2\left[\theta\ C_y^2 + (\theta - \theta')\left(C_x^2 - 2C_{yx}\right)\right] - S_y^2[\theta - (\theta - \theta')\rho^2]$$

$$= (\theta - \theta')\ \bar{Y}^2\left(\rho\ C_y - C_x\right)^2 \geq 0 \tag{4.4}$$

Hence, considering (4.3) and (4.4), we have $MSE\ \left(t_{dge}\right)_{opt} \leq MSE\ (t_{dr})$

i.e. the two-phase generalized exponential estimator is more efficient than conventional two-phase ratio estimator.

(iii) Comparison of two-phase product estimator (t_{dp}) and two-phase generalized exponential estimator $\left(t_{dge}\right)_{opt}$

From, (1.5) and (3.5), we consider

$$MSE\ (t_{dp}) - MSE\left(t_{dge}\right)_{opt}$$

$$= \bar{Y}^2\left[\theta\ C_y^2 + (\theta - \theta')\left(C_x^2 + 2C_{yx}\right)\right] - \frac{S_y^2[\theta - (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}$$

Since, $1 + C_y^2[\theta - (\theta - \theta')\rho^2] \geq 1$ \tag{4.5}

We consider

$$\bar{Y}^2\left[\theta\ C_y^2 + (\theta - \theta')\left(C_x^2 + 2C_{yx}\right)\right] - S_y^2[\theta - (\theta - \theta')\rho^2]$$

$$= (\theta - \theta')\ \bar{Y}^2\left(\rho\ C_y + C_x\right)^2 \geq 0 \tag{4.6}$$

Hence, considering (4.5) and (4.6), we have

$$MSE\ \left(t_{dge}\right) \leq MSE\ (t_{dp})$$

i.e. two-phase generalized exponential estimator is more efficient than conventional two-phase product estimator $\left(t_{dge}\right)_{opt}$.

(iv) Comparison two-phase regression estimator (t_{dlr}) and two-phase generalized exponential estimator.

From, (1.6) and (3.5), we consider

$$MSE\ (t_{dlr}) - MSE\left(t_{dge}\right)_{opt}$$

$$= \bar{Y}^2 C_y^2[\theta\ (1 - \rho^2) + \theta'\rho^2] - \frac{S_y^2[\theta - (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}$$

$$= \bar{Y}^2 C_y^2[\theta\ (1 - \rho^2) + \theta'\rho^2]\left[1 - \frac{1}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}\right] \tag{4.7}$$

Since, $1 + C_y^2[\theta - (\theta - \theta')\rho^2] \geq 1$

From, (4.7) we have $MSE\left(t_{dge}\right)_{opt} \leq MSE\ (t_{dlr})$

i.e. two-phase generalized exponential estimator is more efficient than two-phase regression estimator.

(v) Comparison two-phase exponential ratio estimator (t_{der}) and two-phase generalized exponential estimator $(t_{dge})_{opt}$

From, (2.3) and (3.5), we consider

$MSE\,(t_{der}) - MSE(t_{dge})_{opt}$

$$= \bar{Y}^2\left[\theta\,C_y^2 + (\theta - \theta')\left(\frac{C_x^2}{4} - C_{yx}\right)\right] - \frac{S_y^2[\theta - (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}$$

Since ,

$$1 + C_y^2[\theta - (\theta - \theta')\rho^2] \geq 1 \tag{4.8}$$

We consider

$$= \bar{Y}^2\left[\theta\,C_y^2 + (\theta - \theta')\left(\frac{C_x^2}{4} - C_{yx}\right)\right] - S_y^2[\theta - (\theta - \theta')\rho^2]$$

$$= (\theta - \theta')\bar{Y}^2\left(\rho C_y - \frac{1}{2}C_x\right)^2 \geq 0 \tag{4.9}$$

From, (4.8) and (4.9), we have

$MSE(t_{dge})_{opt} \leq MSE\,(t_{der})$

i.e. two-phase generalized exponential estimator always more efficient than two-phase exponential ratio estimator suggested by Singh et.al(2007)

(vi) Comparison two-phase exponential product estimator (t_{dep}) and e two-phase generalized exponential estimator $(t_{dge})_{opt}$

From, (2.4) and (3.5) we consider

$MSE\,(t_{dep}) - MSE(t_{dge})$

$$= \bar{Y}^2\left[\theta\,C_y^2 + (\theta - \theta')\left(\frac{C_x^2}{4} + C_{yx}\right)\right] - \frac{S_y^2[\theta + (\theta - \theta')\rho^2]}{1 + C_y^2[\theta - (\theta - \theta')\rho^2]}$$

Since , $1 + C_y^2[\theta - (\theta - \theta')\rho^2] \geq 1$ \qquad (4.10)

We consider

$$\bar{Y}^2\left[\theta\,C_y^2 + (\theta - \theta')\left(\frac{C_x^2}{4} + C_{yx}\right)\right] - S_y^2[\theta - (\theta - \theta')\rho^2]$$

$$= (\theta - \theta')\bar{Y}^2\left(\rho C_y + \frac{1}{2}C_x\right)^2 \geq 0 \tag{4.11}$$

From, (4.10) and (4.11), we have

$MSE(t_{dge})_{opt} \leq MSE\,(t_{dep})$

i.e. two-phase exponential estimator always more efficient than two-phase exponential product estimator suggested by Singh et.al(2007)

V.   Numerical Comparison

To study efficiency of the estimator we have considered six population data sets. For better eye view, we consider three population data sets with positive correlation between y and x and three population data sets with negative correlation between y and x. The percent relative efficiency of different estimators are calculated with respect to mean per unit estimator $\bar{y}$.

Population I: Cochran (1977), P.325

y: No. of Persons ,   x: No of Rooms,   N=10,   n'=6,   n=4

$\bar{Y} = 101.1$, $C_y^2 = 0.0233$, $C_x^2 = 0.0182$, $C_{yx} = 0.0134$, $\rho = 0.651$

Population II: Murthy (1967), P.228

y: Output ,   x: Fixed Capital ,   N=80 , n'=70 , n= 30

$\bar{Y} = 5182.64$, $C_y^2 = 0.1255$, $C_x^2 = 0.5635$, $C_{yx} = 0.2503$, $\rho = 0.9413$

Population III: Das (1988)

y: No of agricultural laborers for 1971, x: No of agricultural laborers for 1961,N=278, n'=70, n=30,

$\bar{Y} = 39.0680$, $C_y^2 = 2.0883$, $C_x^2 = 2.6237$, $C_{yx} = 1.6883$, $\rho = 0.7213$

Population IV: Steel and Torrie (1960) p.282

y: Log of leaf burn in sacs,   x: chlorine percentage , N=30, n'=12,   n=4,

$\bar{Y} = 0.6860$, $C_y^2 = 0.2306$, $C_x^2 = 0.5614$, $C_{yx} = -0.1798$, $\rho = -0.4996$

Population V: Drapper and Smith: p.352

y: Average atmospheric temperature, degrees Fahrenheit, x: Average wind velocity in miles per hour

N=25, n'=12,   n=8,

$\bar{Y} = 52.60, \qquad C_y^2 = 0.1077, \quad C_x^2 = 0.07615, \quad C_{yx} = -0.0558, \qquad \rho = -0.6163$

Population VI: Gujurati p.259
y: Year to year percentage change in the index of hourly earnings,   x: the unemployment rate(%) N=12, n'=8, n=5,

$\bar{Y} = 4.066, \qquad C_y^2 = 0.0977, \quad C_x^2 = 0.0535, \quad C_{yx} = -0.0519, \qquad \rho = -0.718$

Table 1: Percent relative efficiency with respect to mean per unit estimator ($\bar{y}$) when $\rho > 0$

Population	Estimator			
	t_{dr}	t_{der}	t_{dlr}	t_{dge}
1	125.91	126.81	130.86	131.21
2	72.37	297.90	307.77	308.87
3	130.03	146.34	149.98	156.19

Table 2: Percent relative efficiency with respect to mean per unit estimator ($\bar{y}$) when $\rho < 0$

Population	Estimator			
	t_{pr}	t_{dep}	t_{dlr}	t_{dge}
4	59.77	115.15	123.76	128.76
5	119.62	120.103	122.88	123.80
6	149.47	133.93	149.53	150.68

VI.   Conclusion
Table 1 and Table 2 clearly show that the modified two-phase estimator is more efficient than mean per unit estimator, two-phase ratio estimator, two-phase ratio type exponential estimator, two-phase product estimator, two-phase product type exponential estimator and two-phase regression estimator.

References
[1]   Bhal , S. and Tuteja, R.K. Ratio and product type exponential estimator, *Information and optimization Sciences, 12,(1991) 159-163*
[2]   Cochran, W.G. Sampling Techniques 3rd Edition, *Wiley* Estern Limited, India.(1977)
[3]   Das,A.K. Contribution to theory of sampling strategies based on auxiliary information. *Unpublished   Ph .D thesis.(1977)*
[4]   Drapper,N .R and Smith ,H. Applied Regression Analysis. *John Wiley & sons, New-York.(1966)*
[5]   Gujurati, D . Essential of Economics, *2nd Edition Mc Graw-Hill International* .(1999)
[6]   Murthy , M.N. Sampling Theory and Methods, *Statistical Publishing Society, Calcutta.*(1977)
[7]   Singh H.P. and Vishwakarma, G.K. Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian journal of statistics, 36, No.3,(2007) 217-225.*
[8]   Steel R.G.D. and Torrie , J.H. . Principle and Procedures of Statistics, *MC-Graw Hill, Newyork.(1960)*
[9]   Sukhatme, P.V.Sukhatme, B.V.Sukhatme, S. and Asok C. Sampling Theory of Surveys with Applications, *Indian Society of Agricultural Statistics, New-Delhi.(1984)*