ISSN 2250-3153

Improvement in Estimating the Population Mean Using Exponential Type Estimator in the Presence of Non-Response

ThanapanangRachokarn andNuanpan Lawson*

Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

*Corresponding author: nuanpan.n@sci.kmutnb.ac.th

Abstract- This paper proposes a class of exponential type estimator to estimate the population mean $\bar{Y}$ in the presence of non-response adapting the estimators suggested by Singh et.al [1], Kumar and Bhougal [2], and Rachokarn and Lawson [3]. The general expressions for bias and MSE of the proposed class of estimator have been obtained. In addition, theoretical and numerical studies were used in order to access the performance of proposed class of estimator. The results of this study showed that the proposed class of estimator is always efficient under the percent relative efficiencies (PREs) criterion as compared to other relevant estimators.

Index Terms- Exponential estimator, Study variable, Auxiliary variable, Non-response.

I. Introduction

In the survey sampling practice, many statisticians attempt to improve the performance of interest estimators by focusing on the uses the benefits of auxiliary information. Unfortunately, in many practical of survey sampling, the required data cannot be collected from all units in the sample, which is an important problem caused by non-response. An estimate derived from such incomplete data may be lead to misleading interpretation and conclusions. In order to deal with such situations, the technique of Hansen and Hurwitz [4] is utilized.

Let (Y_i, X_i) be the non-negative values for the i$^{\text{th}}$unit of the population $U=(U_1,U_2,...,U_N)$ on the study variableyand the auxiliary variable x with their population means $(\bar{Y},\bar{X})$. Supposed the population U of size N is divided in N_1 responding units and $N_2=N-N_1$ non-responding units. Using simple random sampling without replacement (SRSWOR), a sample of size n is drawn from the population of size N which observed that n_1 responding units and n_2 non-responding units. From n_2 non-responding units, a sub-sample of size $r=n_2/k; k>1$ is randomly drawn again by making extra efforts. Therefore, the estimator for $\bar{Y}$ and $\bar{X}$ based on n_1+r proposed by Hansen and Hurwitz [4] are given respectively by:

$$\bar{y}^* = (n_1/n)\bar{y}_1 + (n_2/n)\bar{y}_{2r}, \tag{1}$$

$$\bar{x}^* = (n_1/n)\bar{x}_1 + (n_2/n)\bar{x}_{2r}, \tag{2}$$

The estimator (1) and (2) are unbiased with variance

$$V(\bar{y}^*) = \lambda S_y^2 + \lambda^* S_{y(2)}^2, \tag{3}$$

$$V(\bar{x}^*) = \lambda S_x^2 + \lambda^* S_{x(2)}^2, \tag{4}$$

where $(\bar{y}_1,\bar{x}_1)$ are the sample meansthat based on n_1 of study and auxiliary variables respectively.In the other hand $(\bar{y}_{2r},\bar{x}_{2r})$ are respectively sub sample means based on r of study and auxiliary variables.

$$\lambda = (N-n)/Nn,\ \lambda^* = W_2(k-1)/n,\ W_2 = N_2/N,$$
$$S_y^2 = \sum_{i=1}^{N}(y_i-\bar{Y})^2/(N-1),\ S_{y(2)}^2 = \sum_{i=1}^{N_2}(y_i-\bar{Y}_2)^2/(N_2-1),$$
$$S_x^2 = \sum_{i=1}^{N}(x_i-\bar{X})^2/(N-1),\ S_{x(2)}^2 = \sum_{i=1}^{N_2}(x_i-\bar{X}_2)^2/(N_2-1).$$

Using technique of Hansen and Hurwitz [4], Singh et.al [1] introduced an exponential ratio and product type estimators for the population mean $\bar{Y}$ when non-response occurs, arerespectively given by

$$t_1 = \bar{y}^* \exp\left(\frac{\bar{X}-\bar{x}^*}{\bar{X}+\bar{x}^*}\right) \tag{5}$$

$$t_2 = \bar{y}^* \exp\left(\frac{\bar{x}^*-\bar{X}}{\bar{x}^*+\bar{X}}\right) \tag{6}$$

Later, Kumar and Bhougal[2] extended the Singh et.al [1] by proposing a class of ratio-product estimators for estimating population mean $\bar{Y}$ of the study variable y in presence of non-response, as

$$t_3 = \bar{y}^* \left[\gamma \exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*} \right) + (1-\gamma) \exp\left(\frac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}} \right) \right] \quad (7)$$

where γ is the arbitrary constant number.

Recently, Rachokarn and Lawson [3] proposed a family of estimator to estimate the population mean $\bar{Y}$ for incomplete data due to non-response problems, as

$$t_4 = \bar{y}^* \left(\frac{a\bar{X} + b}{\alpha(a\bar{x}^* + b) + (1-\alpha)(a\bar{X} + b)} \right)^g \quad (8)$$

where $a(a \neq 0)$ and b are either real numbers or functions of population parameters, and α, g are suitably chosen constants.

To the first degree of approximation, the mean squared errors (MSEs) and minimum MSE of the estimator t_1, t_2, t_3, and t_4 are respectively given by

$$MSE(t_1) = \bar{Y}^2 \left[A_1 + \frac{A_2}{4} - A_3 \right] \quad (9)$$

$$MSE(t_2) = \bar{Y}^2 \left[A_1 + \frac{A_2}{4} + A_3 \right] \quad (10)$$

$$Min.MSE(t_3) = Min.MSE(t_4) = \bar{Y}^2 \left[A_1 - \frac{A_3^2}{A_2} \right] \quad (11)$$

where

$$A_1 = \lambda C_y^2 + \lambda^* C_{y(2)}^2,\ A_2 = \lambda C_x^2 + \lambda^* C_{x(2)}^2,\ A_3 = \lambda C_{yx} + \lambda^* C_{yx(2)},$$
$$C_y = S_y / \bar{Y},\ C_x = S_x / \bar{X},\ C_{yx} = S_{yx} / \bar{Y}\,\bar{X},\ C_{yx(2)} = S_{yx(2)} / \bar{Y}\,\bar{X}.$$

In this study, the authors propose a class of exponential type estimator for population mean $\bar{Y}$ of a study variable $y$under the assumption that the population mean $\bar{X}$ of auxiliary variable x is known and information on the study and auxiliary variables is incomplete. Motivation for the proposed class of estimator is based on Singh et.al [1], Kumar and Bhougal[2], and Rachokarn and Lawson [3]. The general expressions for bias and MSE of the proposed class of estimator have been obtained up to the first order of approximation.In addition, comparative studies of the proposed class of estimator with other relevant estimators have been considered through the theoretical and numerical studies, which show the performance of the proposed class of estimator was clearly more efficient than the other relevant estimators.

II. THE PROPOSED CLASS OF ESTIMATOR

On getting motivation from Singh et.al [1], Kumar and Bhougal[2], and Rachokarn and Lawson [3] when the population mean $\bar{X}$ of auxiliary variable x is known in advance.A class of exponential type estimator to estimate the population mean $\bar{Y}$ in the presence of non-responsehas been proposed, as

$$t_5 = \bar{y}^* \left(\frac{a\bar{X} + b}{\alpha(a\bar{x}^* + b) + (1-\alpha)(a\bar{X} + b)} \right)^g \left[\gamma \exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*} \right) + (1-\gamma) \exp\left(\frac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}} \right) \right] \quad (12)$$

where $\bar{y}^*$ are the Hansen and Hurwitz [4] estimator for population mean of the study variable.

<u>Note that:</u> the proposed class of estimator t_5 can reduce to the estimator t_1, t_2, and t_3 when one replace the parameters $g = 0$ and $\gamma = 1$, $g = 0$ and $\gamma = 0$, and $g = 0$ into equation (12) respectively.

III. BIAS AND MSE OF THE PROPOSED CLASS OF ESTIMATOR

Whenever the sample size is large compared to the population size,one can retain only the terms of order n^{-1} and ignore all terms of order greater than or equal to n^{-2}. To study the large sample properties of the proposed estimator, the following notations are used: $\bar{y}^* = \bar{Y}(1 + e_0^*)$, $\bar{x}^* = \bar{X}(1 + e_1^*)$.
Then, one have $E(e_0^*) = E(e_1^*) = 0$, $E(e_0^{*2}) = \lambda C_y^2 + \lambda^* C_{y(2)}^2 = A_1$,
$E(e_1^{*2}) = \lambda C_x^2 + \lambda^* C_{x(2)}^2 = A_2$, $E(e_0^* e_1^*) = \lambda C_{yx} + \lambda^* C_{yx(2)} = A_3$.

The equation (12) can be rewritten in terms of e's, as

$$t_5 = \bar{Y}(1 + e_0^*)(1 + \alpha\tau e_1^*)^{-g} \left[\gamma \exp\left\{ \frac{-e_1^*}{2} \left(1 + \frac{e_1^*}{2} \right)^{-1} \right\} + (1-\gamma) \exp\left\{ \frac{e_1^*}{2} \left(1 + \frac{e_1^*}{2} \right)^{-1} \right\} \right] \quad (13)$$

where $\tau = a\bar{X} / (a\bar{X} + b)$.

Expanding the right hand side of equation (13), multiplying out, and then subtracting $\bar{Y}$ on both sides, one have

$$t_5 - \bar{Y} = \bar{Y} \begin{pmatrix} 1 + e_0^* - (\gamma - \frac{1}{2} + g\alpha\tau)e_1^* \\ +(\frac{\gamma}{2} - \frac{1}{8} + \gamma g\alpha\tau - \frac{g\alpha\tau}{2} + \frac{g(g+1)}{2})\alpha^2\tau^2)e_1^{*2} \\ -(\gamma - \frac{1}{2} + g\alpha\tau)e_0^* e_1^* \end{pmatrix} - \bar{Y} \quad (14)$$

The Bias of the proposed class of estimator to the first order of approximation is obtained by taking expectations on both sides of equation (14), one have

$$Bias(t_5) = \frac{\bar{Y}}{8}\left[(4\gamma - 1 + 8\gamma g\alpha\tau - 4g\alpha\tau + 4g(g+1)\alpha^2\tau^2)A_2 - 4(2\gamma - 1 + 2g\alpha\tau)A_3\right] \quad (15)$$

Similarly, MSE of the proposed class of estimator is obtained by squaring and taking expectations on both sides of equation (14), one get

$$MSE(t_5) = \bar{Y}^2\left[A_2 + (\gamma - \frac{1}{2} + g\alpha\tau)^2 A_2 - 2(\gamma - \frac{1}{2} + g\alpha\tau)A_3\right] \quad (16)$$

The MSE of the proposed class of estimator in equation (16) is minimized when

$$\gamma = \frac{2A_3 - (2g\alpha\tau - 1)A_2}{2A_2} = \gamma_{(opt.)} \quad (17)$$

Substituting the equation (17) in equation (16), one gets the minimum MSEof the proposed class of estimator given by

$$Min.MSE(t_5) = MSE(t_{5(opt.)}) = \bar{Y}^2\left[A_1 - \frac{A_3^2}{A_2}\right] \quad (18)$$

IV. COMPARISON OF THE PROPOSED CLASS OF ESTIMATOR WITH THE RELEVANT ONES

In this section, the authors compare the efficiency of the proposed class of estimator $t_{5(opt.)}$ with the various relevant estimators such as the unbiased estimator $\bar{y}^*$, the estimator t_1, t_2, t_3, and t_4 by considering expressions of MSE from these estimators up to the first order of approximation. The details are as follows:

$$MSE(\bar{y}^*) - Min.MSE(t_5) = A_3^2 > 0 \quad (19)$$

$$MSE(t_1) - Min.MSE(t_5) = (A_2 - 2A_3)^2 > 0 \quad (20)$$

$$MSE(t_2) - Min.MSE(t_5) = (A_2 + 2A_3)^2 > 0 \quad (21)$$

$$Min.MSE(t_3) = Min.MSE(t_4) = Min.MSE(t_5) = \bar{Y}^2\left[A_1 - \frac{A_3^2}{A_2}\right] \quad (22)$$

Therefore, when the conditions (19) to (21) are satisfied, one can infer that the proposed class of estimator $t_{5(opt.)}$ will be more efficient than the other relevant estimators.

V. NUMERICAL STUDY

To study the performance of the estimator discussed in the present paper, the authors considered the data set given in Hazra[5]. The data consists of a list of 96 villages in West Bengal India along with their populations in 1981, assumed that the number of agricultural labors and the area of the village were taken as the study and auxiliary variables respectively. In this study, the 24 villages, whose area was greater than 160 hectares, have been considered to contain non-response. The population parameters and the constants computed from this population are given as follows:

$N = 96,\ n = 40,\ \bar{X} = 144.87,\ \bar{Y} = 137.93,\ C_x = 0.81,\ C_y = 1.32,\ \rho_{yx} = 0.77,$
$C_{x(2)} = 0.94,\ C_{y(2)} = 2.08.$

The criteria for comparing the efficiency of estimators in this study, thepercent relative efficiencies (PREs) of different estimators with respect to the unbiased estimator $\bar{y}^*$ for various values of k have been used and presented in Table 1 as follows:

Table 1. PREs of different estimators with respect to $\bar{y}^*$ for various values of k

Estimator	$1/k$			
	1/5	1/4	1/3	1/2
$\bar{y}^*$	100.00 (2550.98)	100.00 (2034.67)	100.00 (1518.35)	100.00 (1002.04)
t_1	141.99 (1796.56)	143.01 (1422.70)	144.76 (1048.84)	148.45 (674.99)
t_2	70.72 (3607.23)	70.26 (2895.84)	69.51 (2184.45)	68.02 (1473.06)
t_3, t_4, $t_{5(opt.)}$	213.79 (1193.20)	215.05 (946.15)	217.42 (698.36)	223.16 (449.03)

*Figures in parenthesis give the MSE (.).

From Table 1, one observed that the PREs of estimator t_1, t_3, t_4, and $t_{5(opt.)}$ decreases as the value of k increases, while the PRE of estimator t_2 increase as the value of k increases. It is also noted that the estimator t_3, t_4, and $t_{5(opt.)}$ under minimum conditions perform better than the estimator $\bar{y}^*$, t_1, and t_2. Therefore, one can infer that, besides the estimator $t_{5(opt.)}$, the estimator t_3 and t_4 can use alsoin practice.

VI. CONCLUSION

The present article considers the problem of estimating the population mean $\bar{Y}$ of a study variable $y$in the presence of non-responsewhen non-response occurs on both the study and auxiliary variables while the population mean $\bar{X}$ of auxiliary variable is known. Following Singh et.al [1], Kumar and Bhougal [2], and Rachokarn and Lawson [3], a class of exponential type estimator has been proposed and their properties are studied. The minimum MSEof the proposed class of estimator is also obtained. Furthermore, one also investigated the relation between the proposed class of estimator and other relevant estimators in terms of the percent relative efficiencies (PREs). The results from the theoretical and numerical studies can be confirmed that the proposed class of estimator have the best efficiency as same as the estimator proposed by Kumar and Bhougal (2011), and Rachokarn and Lawson (2016) and it has efficiency better than the estimator considered by Singh et.al (2010).Therefore, one can recommend that the proposed class of estimator is efficient and should be used in practical surveys.

.

REFERENCES

[1] Singh, R., Chauhan, P., Sawan, N. and F. Smarandache (2010). Estimation of Mean in Presence of Non Response Using Exponential Estimator, Multispace & Multistructure Neutrosophic Transdisciplinarity. 4, 768-758.

[2] Kumar, S. and Bhougal, S. (2011). Estimation of the population mean in presence of non-response, CSAM, 18, 537-548.

[3] Rachokarn, T., N. Lawson (2016). An efficient family of estimators for the population mean using auxiliary information in the presence of missing observations, Paper presented at the 2nd international conference on mathematics, engineering, and industrial application 2016, August, 10-12, Songkhla, Thailand.

[4] Hansen, M. H. and W. N. Hurwitz (1946). The problem of non-response in sample surveys, J. Am. Stat. Assoc. 41, 517-529.

[5] Hazra, R. (2015). Combination of two ratio type estimator for estimating population mean using auxiliary variable with double sampling in presence of non-response, Bull. Math. & Stat. Res 3 (2), 33-41.

AUTHORS

First Author -ThanapanangRachokarn, PhD student, Department of Applied Statistics, Faculty of Applied Science, King Mongkut’s University ofTechnology North Bangkok, Bangkok 10800, Thailand
Second Author -Nuanpan Lawson, Assistant Professor, Department of Applied Statistics, Faculty of Applied Science, King Mongkut’s University ofTechnology North Bangkok, Bangkok 10800, Thailand

Correspondence Author -nuanpan.n@sci.kmutnb.ac.th