Adapted Exponential Type Estimator in the Presence of Non-response

Ceren Ünal* and Cem Kadilar

Department of Statistics, Hacettepe University, 06800 Beytepe, Ankara, Turkey
E-mail: cerenunal@hacettepe.edu.tr
**Corresponding Author*

Received 15 August 2020; Accepted 19 March 2021;
Publication 22 June 2021

Abstract

In this article, we propose an estimator using the exponential function for the population mean in the case of non-response on both the study and the auxiliary variables. The equations for the Bias and Mean Square Error (MSE) are derived to the first order of approximation and theoretical comparisons are made with existing estimators in literature. Besides, we examine the efficiency of the proposed estimator according to the classical ratio and regression estimator, Hansen-Hurwitz unbiased estimator, and the estimator of Singh et al. (2009). Following theoretical comparisons, we infer that the proposed estimator is more efficient than compared estimators under the obtained conditions in theory. Moreover, these theoretical results are supported numerically by providing an empirical study on five different data sets.

Keywords: Exponential type estimator, auxiliary variable, non-response, population mean, efficiency.

1 Introduction

In sample surveys, it is well known that while estimating the population parameters, the information of the auxiliary variable is usually used in order

Journal of Reliability and Statistical Studies, Vol. 14, Issue 1 (2021), 311–322.
doi: 10.13052/jrss0974-8024.14115

to improve the efficiency of the estimators. In other words, the main aim of studies is to find out efficient estimators by using the auxiliary information. For this reason, many authors have proposed type of ratio, product, regression and exponential estimators using the auxiliary information in recent years. However, required information on different variables may not be obtained correctly and completely. This situation is named as non-response and this decreases the efficiency. Hansen and Hurwitz [1] considered a method in order to deal with this problem and introduced a new technique of sub-sampling the non-respondents. In this method, suppose that $S = (S_1, S_2, \ldots, S_N)$ consists of N units $(N_1 + N_2 = N)$ is composed of N_1 and N_2 belonging to responding units and non-responding units, respectively, and sample of size n is drawn without replacement (SRSWOR) which is divided into two groups as n_1 units are "responding group" and n_2 units are "not responding group". Here, (y_i, x_i) are the values of the study and auxiliary variables for the ith unit $(i = 1, 2, \ldots, 10)$ of the population, respectively. A sub-sample of size $r = {}^{n_2}/_{h}$ $(h > 1)$ units is randomly drawn from n_2 non-responding units where *h* is the inverse sampling rate at the second phase sample of size *n*. $W_1 = \frac{N_1}{N}$ and $W_2 = \frac{N_2}{N}$ are proportions for responding and non-responding for the population, respectively.

Hansen and Hurwitz [1] were the first to propose the unbiased estimator for the population mean in the presence of non-response. The unbiased estimator is given as

$$t_1 = w_1 \bar{y}_1 + w_2 \bar{y}_{2(r)}, \tag{1}$$

where $w_1 = \frac{n_1}{n}$ and $w_2 = \frac{n_2}{n}$ denote responding and non-responding proportions, respectively, for the sample. In addition, $\bar{y}_1$ and $\bar{y}_{2(r)}$ symbolize the sample means of the study variable y depending on n_1 and r units, respectively. The variance of the estimator in (1) is given by

$$V(t_1) = \bar{Y}^2 \left(\lambda C_y^2 + \frac{W_2(h-1)}{n} C_{y(2)}^2 \right), \tag{2}$$

Here, $f = \frac{n}{N}, \lambda = \frac{1-f}{n}, C_y^2 = \frac{S_y^2}{\bar{Y}^2}$. Also, $C_{y(2)}^2 = {S_{y(2)}^2} \big/ {\bar{Y}^2}$ is the coefficient of variation of the study variable for N_2 non-responding units of the population.

When non-response exists on both the study and the auxiliary variables and when the population mean of the auxiliary variable $(\bar{X})$ is known, some of important estimators in the presence of non-response in literature may be considered as follows:

Cochran [2] suggests a ratio estimator for the population mean as

$$t_2 = \frac{\bar{y}^*}{\bar{x}^*}\bar{X}, \tag{3}$$

and its MSE, up to the first order of approximation, is given by

$$MSE(t_2) = \bar{Y}^2\Bigg(\lambda(C_y^2 + C_x^2 - 2C_{yx}) + \frac{W_2(h-1)}{n}(C_{y(2)}^2 + C_{x(2)}^2 - 2\rho_{yx(2)}C_{y(2)}C_{x(2)})\Bigg), \tag{4}$$

where $C_x^2 = \frac{S_x^2}{\bar{X}^2}$, $C_{xy} = \rho_{xy}C_xC_y$, $C_{x(2)}^2 = \frac{S_{x(2)}^2}{\bar{X}^2}$ and $C_{yx(2)} = \rho_{yx(2)}C_{y(2)}C_{x(2)}$. Note that ρ_{xy} and $\rho_{yx(2)}$ are the population correlation coefficient of the response and non-response group between the study and auxiliary variables, respectively.

Using the technique of Hansen and Hurwitz, an exponential estimator, which is introduced by Singh et al. [3], is

$$t_3 = \bar{y}^* exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right), \tag{5}$$

when there are non-response units on both the study as well as the auxiliary variables. The MSE of the estimator in (5) is given by

$$MSE(t_3) = \bar{Y}^2\Bigg(\lambda C_y^2 + \lambda\frac{C_x^2}{4} - \lambda C_{yx} + \frac{W_2(h-1)}{n} \times \left(C_{y(2)}^2 + \frac{C_{x(2)}^2}{4} - \rho_{yx(2)}C_{y(2)}C_{x(2)}\right)\Bigg). \tag{6}$$

Cochran [2] also adapted classical regression estimator to the case of the incomplete information on the study and auxiliary variables as

$$t_4 = \bar{y}^* + b^*(\bar{X} - \bar{x}^*), \tag{7}$$

where $b^* = S_{xy}^* \big/ S_x^{*2}$. The equation of MSE, up to the first order of approximation, for the regression estimator in (7) is given by

$$MSE(t_4) = \bar{Y}^2\Bigg(\lambda C_y^2(1 - \rho_{xy}^2) + \frac{W_2(h-1)}{n} \times \left(C_{y(2)}^2 + \rho_{xy}^2\frac{C_y^2}{C_x^2}C_{x(2)}^2 - 2\rho_{xy}\frac{C_y}{C_x}C_{yx(2)}\right)\Bigg). \tag{8}$$

Main motivation of this study is to find out more efficient estimator than existing estimators in literature considering the non-response. For this reason, we adapt the estimator which is proposed by Vishwakarma et al. [4] to a new estimator considering the case of non-response units on both the study and the auxiliary variables for the estimation of the population mean. In Section 2, the expressions of the Bias and MSE for the adapted estimator are also obtained. Theoretical and numerical comparisons between the adapted estimator and existing estimators, such as Hansen-Hurwitz unbiased estimator, adapted classical ratio and regression estimators, Singh et al. [3] exponential estimator, are made in Sections 3 and 4, respectively.

2 The Adapted Estimator

We adapt the exponential type estimator which is proposed by Vishwakarma et al. [4] to a new estimator considering the case of non-response occurs on both the study and the auxiliary variables as follows. The adapted estimator is given as follows:

$$t_5 = \alpha \bar{y}^* + (1-\alpha)\bar{y}^* \exp\left(\frac{\bar{X}-\bar{x}^*}{\bar{X}+\bar{x}^*}\right). \tag{9}$$

To obtain the Bias and MSE of the proposed estimator in (9), we can write that

$$\bar{y}^* = \bar{Y}(1+e_0^*), \quad \bar{x}^* = \bar{X}(1+e_1^*),$$

then, we have $E(e_x^*) = E(e_y^*) = 0$, $E(e_x^{*^2}) = \lambda C_x^2 + \frac{W_2(h-1)}{n} C_{x(2)}^2$, $E(e_y^{*^2}) = \lambda C_y^2 + \frac{W_2(h-1)}{n} C_{y(2)}^2$ and $E(e_x^* e_y^*) = \lambda \rho_{xy} C_x C_y + \frac{W_2(h-1)}{n} \rho_{xy(2)}$ $C_{x(2)} C_{y(2)}$.

Now, expressing the estimator in (9), in terms of $e_i^*(i = x, y)$, we have

$$t_5 = \alpha \bar{Y}(1+e_0^*) + (1-\alpha)\bar{Y}(1+e_0^*) \exp\left(\frac{\bar{X}-\bar{X}-\bar{X}e_1^*}{\bar{X}+\bar{X}+\bar{X}e_1^*}\right), \tag{10}$$

$$= \bar{Y}(\alpha + \alpha e_0^*) + \bar{Y}(1+e_0^* - \alpha - \alpha e_0^*) \exp\left(-\frac{e_1^*}{2}\left(1+\frac{e_1^*}{2}\right)^{-1}\right), \tag{11}$$

$$= \bar{Y}\left(1 + e_0^* - \frac{e_1^*}{2} + \alpha\frac{e_1^*}{2} + \frac{3e_1^{*^2}}{8} - \frac{3\alpha e_1^{*^2}}{8} - \frac{e_0^* e_1^*}{2} + \frac{\alpha}{2} e_0^* e_1^*\right). \tag{12}$$

Expanding the right hand side of (12), to the first degree of approximation, we have

$$(t_5 - \bar{Y}) = \bar{Y}\left(e_0^* + e_1^*\left(\frac{\alpha}{2} - \frac{1}{2}\right) + e_1^{*^2}\left(\frac{3}{8} - \frac{3\alpha}{8}\right) + e_0^* e_1^*\left(\frac{\alpha}{2} - \frac{1}{2}\right)\right). \tag{13}$$

Taking the expectation of both sides of (13), we get the bias as

$$\begin{aligned} BIAS(t_5) = \bar{Y}\Bigg(&\frac{3}{8}(1-\alpha)\left(\lambda C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2\right) \\ &+ \left(\frac{\alpha - 1}{2}\right)\left(\lambda C_{yx} + \frac{W_2(h-1)}{n}\rho_{yx(2)}C_{y(2)}C_{x(2)}\right)\Bigg). \end{aligned} \tag{14}$$

Squaring both sides of (13), we have

$$(t_5 - \bar{Y})^2 = \bar{Y}^2\left(e_0^{*^2} + e_1^{*^2}\left(\frac{\alpha^2}{4} - \frac{\alpha}{2} + \frac{1}{4}\right) + e_0^* e_1^*(\alpha - 1)\right),$$

then taking expectation on both sides, we get the MSE as

$$\begin{aligned} MSE(t_5) = \bar{Y}^2\Bigg(&\lambda C_y^2 + \frac{W_2(h-1)}{n}C_{y(2)}^2 \\ &+ \left(\lambda C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2\right)\left(\frac{\alpha^2}{4} - \frac{\alpha}{2} + \frac{1}{4}\right) \\ &+ \left(\lambda \rho_{xy} C_x C_y + \frac{W_2(h-1)}{n}\rho_{yx(2)}C_{y(2)}C_{x(2)}\right)(\alpha - 1)\Bigg). \end{aligned} \tag{15}$$

The optimal value of α is obtained by minimizing the MSE in (15) as

$$\alpha^* = \frac{A_1 - 2A_2}{A_1}, \tag{16}$$

where $A_1 = E(e_1^{*^2}) = \lambda C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2$ and $A_2 = E(e_0^* e_1^*) = \lambda C_{yx} + \frac{W_2(h-1)}{n}\rho_{yx(2)}C_{y(2)}C_{x(2)}$.

Replacing α in (15) with the optimal value of α, given in (16), we have the minimum MSE of the estimator as

$$MSE_{\min}(t_5) = \bar{Y}^2 \left(\lambda \left(C_y^2 + \frac{A_2}{A_1} \left(\frac{A_2}{A_1} - 2\rho_{xy} \frac{C_y}{C_x} \right) C_x^2 \right) + \frac{W_2(h-1)}{n} \left(C_{y(2)}^2 + \frac{A_2}{A_1} \left(\frac{A_2}{A_1} - 2\rho_{xy(2)} \frac{C_{y(2)}}{C_{x(2)}} \right) C_{x(2)}^2 \right) \right). \tag{17}$$

Also, by substituting the obtained A_1 and A_2 equations, the minimum MSE of the estimator can be rewritten as follows:

$$MSE_{\min}(t_5) = \bar{Y}^2 \left[\left(\lambda C_y^2 + \frac{W_2(h-1)}{n} C_{y(2)}^2 \right) - \frac{\left(\lambda C_{xy} + \frac{W_2(h-1)}{n} \rho_{yx(2)} C_{y(2)} C_{x(2)} \right)^2}{\left(\lambda C_x^2 + \frac{W_2(h-1)}{n} C_{x(2)}^2 \right)} \right].$$

3 Efficiency Comparisons

In this section, we compare the MSE equation of the adapted estimator (t_5) with the MSE equations of the mentioned estimators, such as Hansen and Hurwitz [1] unbiased estimator, Cochran [2] classical ratio and regression estimators, Singh et al. [3] exponential estimator, mentioned in Section 1.

We find the efficiency conditions of the proposed estimator as follows:

(i) $MSE_{\min}(t_5) < MSE(t_1)$

$$\left(\lambda C_{xy} + \frac{W_2(h-1)}{n} C_{yx(2)} \right)^2 > 0 \tag{18}$$

(ii) $MSE_{\min}(t_5) < MSE(t_2)$

$$\left(\left(\lambda C_x^2 + \frac{W_2(h-1)}{n} C_{x(2)}^2 \right) - \left(\lambda C_{yx} + \frac{W_2(h-1)}{n} C_{yx(2)} \right) \right)^2 > 0 \tag{19}$$

(iii) $MSE_{\min}(t_5) < MSE(t_3)$

$$\left(\left(\lambda C_{xy} + \frac{W_2(h-1)}{n}\rho_{yx(2)}C_{y(2)}C_{x(2)}\right) - \frac{1}{2}\left(\lambda C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2\right)\right)^2 > 0 \tag{20}$$

(iv) $MSE_{\min}(t_5) < MSE(t_4)$

$$\left(\left(\frac{W_2(h-1)}{n}C_{x(2)}^2\rho_{xy}\frac{C_y}{C_x}\right) - \left(\frac{W_2(h-1)}{n}C_{yx(2)}\right)\right)^2 > 0 \tag{21}$$

The conditions between (18)–(21) are always satisfied, we infer that the proposed estimator, t_5, is more efficient than the compared estimators t_1, t_2, t_3 and t_4.

4 Numerical Illustration

To examine the appropriateness of the proposed estimator, we have used five popular different data sets considered by many researchers in literature. The descriptive statistics and results for each data set are given as follows:

The PREs of t_1, t_2, t_3, t_4 and t_5 for various values of *h* are presented in Tables 2–6 based on five populations, respectively.

We would like to remark that the PRE of the adapted estimator is more efficient than the other compared estimators, t_1, t_2, t_3 and t_4 in the presence

Table 1 Descriptive statistics for each data set

		Parameters										
	Population	N	n	W_2	$\bar{X}$	$\bar{Y}$	C_y	C_x	ρ_{yx}	$C_{y(2)}$	$C_{x(2)}$	$\rho_{yx(2)}$
1	Khare and Kumar [5]	96	25	0.25	1807.23	185.22	1.053	1.0633	0.904	0.528	0.853	0.895
2	Khare and Sinha [6]	96	40	0.25	144.87	137.92	1.32	0.81	0.77	2.08	0.94	0.72
3	Khare and Srivastava [7]	70	35	0.2	1755.53	981.29	0.6254	0.801	0.778	0.4087	0.574	0.445
4	Sinha and Kumar [8]	109	35	0.25	255.97	485.92	0.6559	0.6037	0.857	0.7335	0.6897	0.834
5	Sinha and Kumar [9]	109	35	0.25	41.24	485.92	0.6559	1.126	0.451	0.4785	1.166	0.714

Table 2 PREs of the Proposed Estimator (t_5) and Other Estimators for Population 1

	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
t_1	100.0000	100.0000	100.0000	100.0000	100.0000
t_2	425.4729	370.1973	332.8156	305.8494	285.4779
t_3	301.6963	310.1851	317.9189	324.9939	331.4910
t_4	419.9153	350.2990	306.3844	276.1563	254.0789
t_5	**491.1458**	**463.1675**	**447.5542**	**438.3615**	**432.8506**

Table 3 PREs of the Proposed Estimator (t_5) and Other Estimators for Population 2

	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
t_1	100.0000	100.0000	100.0000	100.0000	100.0000
t_2	202.2646	194.3660	190.6994	188.5823	187.2039
t_3	148.0884	144.4216	142.6821	141.6667	141.0011
t_4	219.9692	212.4751	208.9698	206.9382	205.6122
t_5	**220.6768**	**214.9752**	**212.6081**	**211.3493**	**210.5799**

Table 4 PREs of the Proposed Estimator (t_5) and Other Estimators for Population 3

	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
t_1	100.0000	100.0000	100.0000	100.0000	100.0000
t_2	124.3555	108.5754	98.8639	92.2849	87.5332
t_3	208.3451	188.8973	176.1676	167.1876	160.5133
t_4	209.0156	184.9004	169.7404	159.3284	151.7359
t_5	**210.8401**	**189.2205**	**176.1734**	**167.4633**	**161.2454**

Table 5 PREs of the Proposed Estimator (t_5) and Other Estimators for Population 4

	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
t_1	100.0000	100.0000	100.0000	100.0000	100.0000
t_2	351.9514	342.8085	337.4328	333.8937	331.3874
t_3	233.9950	232.7577	232.0054	231.4997	231.1363
t_4	359.2028	350.8473	345.8426	342.5690	340.2464
t_5	**359.4776**	**351.3462**	**346.5934**	**343.4756**	**341.2729**

of non-response. Furthermore, the PREs of the adapted estimator stands out for Population 1, 2 and 3, especially, according to the results in Tables 2–6. We also see that the PRE of the t_5 estimator decrease with the increasing values of *h* except the Population 5.

Table 6 PREs of the Proposed Estimator (t_5) and Other Estimators for Population 5

	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
t_1	100.0000	100.0000	100.0000	100.0000	100.0000
t_2	38.8760	37.0780	35.8300	34.9130	34.2109
t_3	107.8945	110.9662	113.3976	115.3701	117.0024
t_4	133.8007	140.4609	145.9319	150.5061	154.3873
t_5	**133.8633**	**140.6118**	**146.1823**	**150.8558**	**154.8319**

5 Conclusion

In this study, we propose a new exponential type estimator for the estimation of the population mean using the information of the auxiliary variable in the presence of non-response. Equations for the bias and minimum MSE of the proposed estimator are obtained. In theoretical comparisons, the proposed estimator is found more efficient than the estimators in literature, such as Hansen and Hurwitz [1] unbiased estimator, Cochran [2] adapted ratio and regression estimators, and Singh et al. [3] exponential estimator, under the obtained conditions. We use five data sets with the aim of supporting the results in theory and we show that the proposed estimator is quite efficient than other compared estimators as seen in Tables 2–6. Hence, the proposed estimator is recommended based on the theoretical and numerical results and can be used in applications.

References

[1] Hansen, M. H., and Hurwitz, W. N. (1946). The problem of non-response in sample surveys, *Journal of the American Statistical Association*, 41(236), 517–529.
[2] Cochran, W.G., *Sampling Techniques*, John Wiley and sons, New-York, 1977.
[3] Singh, R., Kumar, M., Chaudhary, M. K., and Smarandache, F. (2009). Estimation of mean in presence of non-response using exponential estimator. arXiv preprint arXiv:0906.2462.
[4] Vishwakarma, G. K., Singh, R., Gupta, P. C., and Pareek, S. (2016). Improved ratio and product type estimators of finite population mean in simple random sampling, *Investigación Operacional*, *37*(1), 70–76.
[5] Khare, B. B., and Kumar, S. (2009). Utilization of coefficient of variation in the estimation of population mean using auxiliary character in

the presence of non-response, *National Academy Science Letters-India*, 32(7–8), 235–241.
[6] Khare, B. B., and Sinha, R. R. (2009). On class of estimators for population mean using multi-auxiliary characters in the presence of non-response, *Statistics in Transition*, 10(1), 3–14.
[7] Khare, B. B., and Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response, *National Academy Science Letters*, 16, 111–111.
[8] Sinha, R. R., and Kumar, V. (2015a). Estimation of mean using double sampling the non-respondents with known and unknown variance, *International Journal of Computing Science and Mathematics*, 6(5), 442–458.
[9] Sinha, R. R., and Kumar, V. (2015b). Families of estimators for finite population variance using auxiliary character under double sampling the non-respondents, *National Academy Science Letters*, 38(6), 501–505.

Biographies

Ceren Ünal received her B.Sc. and M.Sc. degrees in Statistics from Hacettepe University, Turkey, in 2014 and 2017. She is a Ph.D. student at the same university. She is currently working as a research assistant in the Department of Statistics at Hacettepe University. Her research interest is Sampling.

Cem Kadilar was born in Ankara in Turkey in 1972. After graduation from higher school Ankara College, he received his B.Sc., M.Sc., and Ph.D. degrees in Statistics from Hacettepe University. He became Associate Professor in 2004 and Professor in 2010 at the same university. He has approximately 150 publications and more than 3500 citations. His interest areas are Sampling, Time Series Analysis, and Survival Analysis.