

Lawson et al., 2018

Volume 4 Issue 2, pp. 01-11

Date of Publication: 14th July, 2018

DOI-<https://dx.doi.org/10.20319/mijst.2018.42.0111>

This paper can be cited as: Lawson, N., Rachokarn, T., & Charurotkeerati, T. (2018). An Efficient General Family of Estimators for Population Means in Sampling with Non-Response. *Matter: International Journal of Science and Technology*, 4(2).01-11.

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

AN EFFICIENT GENERAL FAMILY OF ESTIMATORS FOR POPULATION MEANS IN SAMPLING WITH NON-RESPONSE

Nuanpan Lawson

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand
nuanpan.n@sci.kmutnb.ac.th

Thanapanang Rachokarn

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand
thanapanang@hotmail.com

Thitanont Charurotkeerati

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand
thitanon.c@sci.kmutnb.ac.th

Abstract

In estimating population means for a study variable from random sampling, it is often possible to reduce the bias and improve the efficiency of estimators by including known data on an auxiliary variable which is correlated with the study variable. Non-response is a common problem that occurs when estimating values of variables by random sampling as it can increase the bias and reduce the efficiency of estimators. In this paper, we propose a new family of estimators for population means of a study variable when non-response occurs in the sampling of the study variable and when information on an auxiliary variable is either known or can be obtained by non-response sampling. The asymptotic properties of the proposed estimators such as bias, mean

square error (MSE), and minimum mean square error have been derived up to a first order approximation. A numerical study of the new estimators shows that they are more efficient than other existing estimators.

Keywords

Family of estimators, Auxiliary variable, Non-response, Mean square error

1. Introduction

In statistics, if a study variable y is correlated with an auxiliary variable x , then estimators of properties of the auxiliary variable can be used to improve the precision of estimators for the study variable. For example, Cochran (1940) used the population mean of an auxiliary variable to develop a ratio estimator to improve the estimator for population mean of a study variable, whereas Robson (1957) and Murthy (1964) used the population mean of an auxiliary variable to develop a product estimator to improve the estimator of a study variable.

Searls (1964) used the known coefficient of variation C_y of a study variable y to improve the estimator of the population mean for that study variable. Searls's estimator \bar{y}'_s for the population mean \bar{Y} is given by

$$\bar{y}'_s = a_0 \bar{y}, \quad (1)$$

where \bar{y} is the sample mean of study variable y , $a_0 = [1 + \phi C_y^2]^{-1}$ and ϕ is a constant. Searls showed that for simple random sampling without replacement, the bias and MSE of (1) are given by

$$\text{Bias}(\bar{y}'_s) = -a_0 \phi C_y^2 \bar{Y}. \quad (2)$$

$$\text{MSE}(\bar{y}'_s) = a_0 \phi C_y^2 \bar{Y}^2 \quad (3)$$

and that (3) gives a smaller mean square error than the conventional unbiased estimator.

Sisodia and Dwivedi (1981), Pandey and Dubey (1988) improved the efficiency of ratio and product estimators, respectively, by using the known population coefficient of variation of an auxiliary variable. Singh et al. (1973) improved the efficiency of estimators of a study

variable by using the known coefficient of kurtosis of an auxiliary variable, while Upadhyaya and Singh (1999) used the coefficients of variation and kurtosis of an auxiliary variable (see e.g., Soponviwatkul and Lawson (2017)). Khoshnevisan et al. (2007) proposed a general family of estimators for population mean of a study variable based on known population properties for an auxiliary variable. They assumed that all sample estimators were obtained from full response sampling. The Khoshnevisan et al. (2007) estimator is as follows:

$$T_k = \bar{y} \left(\frac{a\bar{X} + b}{\alpha(a\bar{x} + b) + (1-\alpha)(a\bar{X} + b)} \right)^g, \quad (4)$$

where \bar{y} is the sample mean of study variable. The known auxiliary variable information included in (4) are the sample mean \bar{x} , the population mean \bar{X} , and $a \neq 0$ and b which are real numbers or functions of known auxiliary parameters. The constants α and g are real numbers which are to be determined.

A major problem with sampling is non-response, which can cause unbiased full-response estimators to be biased. Hansen and Hurwitz (1946) first suggested that sub-sampling of non-responding groups could be used to develop an unbiased estimator for the population mean by combining the available information from responding and non-responding groups. Hansen and Hurwitz (1946) used random sampling without replacement (SRSWOR) to select a sample of size n from a population of size N . Then, if n_1 units respond and n_2 units do not respond, they selected a sub-sample of size $r (= n_2 k^{-1}); k > 1$ from the non-responding sample and conducted personal interviews on this sub-sample to obtain information about study variable y . Hansen and Hurwitz (1946) then estimated the population mean \bar{Y} from the sample means \bar{y}_1 and \bar{y}'_2 of the responding sample and the sub-sample of the non-responding sample by the estimator

$$\bar{y}^* = w_1 \bar{y}_1 + w_2 \bar{y}'_2, \quad (5)$$

Hansen and Hurwitz showed that the estimator (5) is unbiased and that the variance is:

$$V(\bar{y}^*) = \phi S_y^2 + \phi^* S_{y(2)}^2, \quad (6)$$

where $\phi = (N-n)/Nn$, $\phi^* = w_2(k-1)/n$, $w_i = n_i/n; (i=1,2)$, $(S_y^2, S_{y(2)}^2)$ are the population variance of the study variable y and the population variance of the non-responding group under study variable y .

The method of Hansen and Hurwitz (1946) uses sample information from the study variable only. Khare and Srivastava (1993, 1995), Khare and Kumar (2009), Khare and Rehman (2015), Rachokarn and Lawson (2017a, b, c) have applied the method of Hansen and Hurwitz (1946) first to improve non-response estimators for auxiliary variables and then to improve non-response estimators for study variables based on the non-response auxiliary variable estimators.

In the present study, we first generalize the Searls type estimator (see equation (1)) to non-response sampling by using the Hansen and Hurwitz method to obtain a non-response estimate of the mean of the study variable. Then, by generalizing the Khoshnevisan et al. (2007) estimator (see equation (4)), we develop an improved family of population non-response estimators for a study variable by including non-response estimators of an auxiliary variable. Expressions for the bias and mean squared error of the proposed estimators have been obtained and the performance of the new estimators has been compared with the performance of previous estimators.

2. Proposed Estimators

In this section, we propose an improved estimator of population mean \bar{Y} of a study variable assuming that non-response occurs only in the study variable and that the population mean of an auxiliary variable is known. The proposed estimator is as follows:

$$\bar{y}_S^* = a_1 \bar{y}^*, (7)$$

where a_1 is a constant and \bar{y}^* is the sample mean obtained as in equation (5) by using the Hansen and Hurwitz (1946) non-response sampling method. The value of a_1 , for which the $MSE(\bar{y}_S^*)$ will be minimum is given by:

$$a_{1(opt.)} = \left[1 + \frac{N-n}{Nn} \frac{S_y^2}{\bar{Y}^2} + \phi^* \frac{S_{y(2)}^2}{\bar{Y}^2} \right]^{-1}, (8)$$

When S_y^2/\bar{Y}^2 and $S_{y(2)}^2/\bar{Y}^2$ do not differ significantly, one can approximate $S_{y(2)}^2/\bar{Y}^2 = C_y^2$ and neglect the terms of order $1/N$, as follows

$$a_{1(opt.)} = \left[1 + \frac{C_y^2}{n} (1 + \phi^*) \right]^{-1}. \quad (9)$$

Therefore, the new proposed estimator \bar{y}_S^* for estimating population mean \bar{Y} is .

$$\bar{y}_S^* = \left[1 + \frac{C_y^2}{n} (1 + \phi^*) \right]^{-1} \bar{y}, \quad (10)$$

which is of the ratio estimator type. The bias and MSE of (10) can be shown to be

$$Bias(\bar{y}_S^*) = (1 - A_2)\bar{Y}. \quad (11)$$

$$MSE(\bar{y}_S^*) = (1 - A_1) \frac{S_y^2}{n} + (1 - 2A_2)\phi^* S_{y(2)}^2, \quad (12)$$

Where $A_1 = \frac{C_y^2}{n} [1 - n^2 \phi^{*2}]$, $A_2 = \frac{C_y^2}{n} [1 + n\phi^*]$.

Based on the new estimator in (10) and the Khoshnevisan et al. (2007) estimator in equation (4)., we propose a family of improved non-response estimators for population mean \bar{Y} as follows:

$$T = \bar{y}_S^* \left[\frac{a\bar{X} + b}{\alpha(a\bar{x} + b) + (1 - \alpha)(a\bar{X} + b)} \right]^g, \quad (13)$$

where \bar{y}_S^* is defined in (10) and the remaining variables are defined in (4).

We now study the bias and error of the proposed ratio-type estimators in (13) for large samples. We assume that the errors $e_0^* = (\bar{y}^* - \bar{Y})/\bar{Y}$ and $e_1 = (\bar{x} - \bar{X})/\bar{X}$, are sufficiently small to justify using a Taylor series expansion. We use the following notation: $E(e_0^{*2}) = \phi C_y^2 + \phi^* C_{y(2)}^2$, $E(e_1^2) = \phi C_x^2$, $E(e_0^*, e_1) = \phi C_{yx}$.

To first order, the bias and MSE of the estimator, T , are given by

$$\text{Bias}(T) = (1 - A_2)\bar{Y} \left[\phi \left\{ \begin{array}{l} g \frac{(g+1)}{2} \alpha^2 \lambda^2 C_x^2 \\ -g\alpha\lambda C_{yx} \end{array} \right\} \right] - \bar{Y} A_2. \quad (14)$$

$$\text{MSE}(T) = (1 - 2A_2)\bar{Y}^2 \left[\phi \left\{ C_y^2 + g^2 \alpha^2 \lambda^2 C_x^2 - 2g\alpha\lambda C_{yx} \right\} \right] + 2\bar{Y}^2 A_2, \quad (15)$$

where $\lambda = a\bar{X} / (a\bar{X} + b)$, $C_x^2 = S_x^2 / \bar{X}^2$, $C_y^2 = S_y^2 / \bar{Y}^2$, $C_{yx} = S_{yx} / \bar{Y}\bar{X}$, $S_{yx} = \rho_{yx} S_y S_x$.

The minimum $\text{MSE}(T)$ is obtained for the value of $\alpha_{opt} = \theta_1 / \lambda g$, where $\theta_1 = \rho_{yx} C_y / C_x$.

Therefore, the minimum mean square error of the ratio estimator, T , is given by

$$\text{MSE}_{\min}(T) = (1 - 2A_2)\bar{Y}^2 \left[\frac{\phi C_y^2 (1 - \rho_{yx}^2)}{+ \phi^* C_{y(2)}^2} \right] + 2\bar{Y}^2 A_2. \quad (16)$$

3. Examples of the Proposed Estimators

Some examples of the proposed estimators are shown in Tables 1 and 2 for selected values of the constants and variables in (13). The estimators in Table 1 are of the ratio type and the estimators in Table 2 are of the product type.

Table 1: Examples of the proposed ratio estimators T

Ratio estimators	α	a	b
$T_1 = \bar{y}_S^* \left[\bar{X} / \bar{x} \right]$	1	1	0
$T_2 = \bar{y}_S^* \left[(\bar{X} + \sigma_x) / (\bar{x} + \sigma_x) \right]$	1	1	σ_x
$T_3 = \bar{y}_S^* \left[(\bar{X} + C_x) / (\bar{x} + C_x) \right]$	1	1	C_x
$T_4 = \bar{y}_S^* \left[(\bar{X} + \beta_1(x)) / (\bar{x} + \beta_1(x)) \right]$	1	1	$\beta_1(x)$
$T_5 = \bar{y}_S^* \left[(\bar{X} + \beta_2(x)) / (\bar{x} + \beta_2(x)) \right]$	1	1	$\beta_2(x)$

Up to first order, the bias and MSE of the ratio estimators in Table 1 are:

When $i=1$;

$$Bias(T_i) = (1 - A_2)\bar{Y} \left[\varphi(C_x^2 - C_{yx}) \right] - \bar{Y}A_2. \quad (17)$$

$$MSE(T_i) = (1 - 2A_2)\bar{Y}^2 \left[\varphi \{ C_y^2 + C_x^2 - 2C_{yx} \} + \varphi^* C_{y(2)}^2 \right] + 2\bar{Y}^2 A_2. \quad (18)$$

When $i=2, \dots, 5$;

$$Bias(T_i) = (1 - A_2)\bar{Y} \left[\varphi \left(\lambda_{(i-1)}^2 C_x^2 - \lambda_{(i-1)} C_{yx} \right) \right] - \bar{Y}A_2. \quad (19)$$

$$MSE(T_i) = (1 - 2A_2)\bar{Y}^2 \left[\varphi \left\{ C_y^2 + \lambda_{(i-1)}^2 C_x^2 - 2\lambda_{(i-1)} C_{yx} \right\} + \varphi^* C_{y(2)}^2 \right] + 2\bar{Y}^2 A_2. \quad (20)$$

Table 2: Examples of the proposed product estimators T

Product estimators	α	a	b
$T_6 = \bar{y}_S^* \left[\bar{x} / \bar{X} \right]$	1	1	0
$T_7 = \bar{y}_S^* \left[(\bar{x} + \sigma_x) / (\bar{X} + \sigma_x) \right]$	1	1	σ_x
$T_8 = \bar{y}_S^* \left[(\bar{x} + C_x) / (\bar{X} + C_x) \right]$	1	1	C_x
$T_9 = \bar{y}_S^* \left[(\bar{x} + \beta_1(x)) / (\bar{X} + \beta_1(x)) \right]$	1	1	$\beta_1(x)$
$T_{10} = \bar{y}_S^* \left[(\bar{x} + \beta_2(x)) / (\bar{X} + \beta_2(x)) \right]$	1	1	$\beta_2(x)$

Up to first order, the bias and MSE of the product estimators in Table 2 are:

When $j=6$;

$$Bias(T_j) = (1 - A_2)\bar{Y} \left[\varphi C_{yx} \right] - \bar{Y}A_2. \quad (21)$$

$$MSE(T_j) = (1 - 2A_2)\bar{Y}^2 \left[\varphi \{ C_y^2 + C_x^2 + 2C_{yx} \} + \varphi^* C_{y(2)}^2 \right] + 2\bar{Y}^2 A_2. \quad (22)$$

When $j=7, \dots, 10$;

$$Bias(T_j) = (1 - A_2)\bar{Y} \left[\varphi \lambda_{(j-6)} C_{yx} \right] - \bar{Y}A_2. \quad (23)$$

$$MSE(T_j) = (1 - 2A_2)\bar{Y}^2 \left[\begin{array}{l} \varphi \{ C_y^2 + \lambda_{(j-6)}^2 C_x^2 + 2\lambda_{(j-6)} C_{yx} \} \\ + \varphi^* C_{y(2)}^2 \end{array} \right] + 2\bar{Y}^2 A_2, \quad (24)$$

where $\lambda_1 = \bar{X} / (\bar{X} + \sigma_x)$, $\lambda_2 = \bar{X} / (\bar{X} + C_x)$, $\lambda_3 = \bar{X} / (\bar{X} + \beta_1(x))$, $\lambda_4 = \bar{X} / (\bar{X} + \beta_2(x))$.

4. Efficiency Comparisons

We compare the efficiency of the new estimators T with respect to other relevant estimators.

$$MSE(\bar{y}_S^*) - MSE_{\min}(T) \text{ if } (1 - A_1)C_y^2 - (1 - 2A_2)C_y^2(1 - \rho_{yx}^2) > 0 \quad (25)$$

$$MSE(T_1) - MSE_{\min}(T) \text{ if } (C_x - \rho_{yx}C_y)^2 > 0 \quad (26)$$

$$MSE(T_i) - MSE_{\min}(T); i = 2, \dots, 5 \text{ if } (\lambda_{(i-1)}C_x - \rho_{yx}C_y)^2 > 0 \quad (27)$$

$$MSE(T_6) - MSE_{\min}(T) \text{ if } (C_x + \rho_{yx}C_y)^2 > 0 \quad (28)$$

$$MSE(T_j) - MSE_{\min}(T); j = 7, \dots, 10 \text{ if } (\lambda_{(j-6)}C_x + \rho_{yx}C_y)^2 > 0 \quad (29)$$

When conditions (25) to (29) are satisfied, one can infer that the proposed estimator T is more efficient than other existing estimators.

5. Numerical illustrations

For numerical illustration, we use data in Khare and Sinha (2009) based on a population census of 96 rural villages in West Bengal, India in 1981. In their study, Khare and Sinha (2009) assumed that the study group was the number of agricultural laborers in a village and that the auxiliary data was the area of a village. They divided the 96 villages into a response group and a non-response group with the non-response group being 25% of villages with area greater than 160 hectares. The values of the parameters in the Khare and Sinha study were as follows:

$N = 96$, $n = 24$, $n_2 = 3, 6, 10$, $\bar{Y} = 137.93$, $\bar{X} = 144.87$, $w_2 = 0.25$, $S_y = 182.50$, $S_{y(2)} = 287.42$, $C_x = 0.81$, $C_{x(2)} = 0.94$, $\rho_{yx} = 0.77$, $\rho_{yx(2)} = 0.72$, $\beta_1(x) = 0.55$, $\beta_2(x) = 3.06$.

Table 3 shows a comparison of the percentage relative efficiency (PRE) of our proposed estimators with respect to \bar{y}^* for different values of the Hansen and Hurwitz non-response parameter k with the PRE of some other estimators.

Table 3: PREs of the new estimators T with respect to \bar{y}^*

Estimator	1/k			Estimator	1/k		
	1/2	1/3	1/4		1/2	1/3	1/4
\bar{y}^*	100.00 (1901.36)	100.00 (2761.88)	100.00 (3622.41)	T_6	132.48 (1435.17)	120.18 (2298.18)	114.60 (3160.85)
T_1	115.82 (1641.58)	110.27 (2504.75)	107.57 (3367.58)	T_7	165.63 (1147.98)	137.34 (2011.05)	126.05 (2873.77)
T_2	111.20 (1709.90)	107.34 (2573.05)	105.43 (3435.86)	T_8	132.85 (1431.21)	120.38 (2294.22)	114.75 (3156.89)
T_3	115.79 (1642.04)	110.25 (2505.22)	107.55 (3368.04)	T_9	132.73 (1432.48)	120.32 (2295.49)	114.70 (3158.16)
T_4	115.80 (1641.90)	110.25 (2505.07)	107.56 (3367.89)	T_{10}	133.85 (1420.52)	120.95 (2283.53)	115.14 (3146.20)
T_5	115.70 (1643.35)	110.19 (2506.52)	107.51 (3369.34)	T	215.83 (880.97)	158.36 (1744.09)	138.96 (2606.87)

* MSE (.).

From Table 3, the proposed estimator T which uses the coefficient of variation of the study variable was more efficient than all other estimators. Moreover, the table also shows that the PRE decreases as k increases

6. Conclusion

A new estimator for population means by adjusting Searls (1964)'s estimator using the idea of Hansen and Hurwitz (1946) was suggested. Moreover, we proposed an efficient general family of estimators to estimate population mean by adjusting Khoshnevisan et al. (2007)'s estimator when non-response occurs in the sampling of the study variable and when information on an auxiliary variable is either known or can be obtained by non-response sampling. The asymptotic properties of the proposed estimators such as bias, mean square error under study conditions were obtained and studied. In conclusion, the numerical illustration revealed that the efficiency of the new family of estimators was apparently better than other existing estimators. Therefore, one can recommend that the new estimators can be used in practice when nonresponse occurs on the variable of interest y only, while the information on an auxiliary variable is known. However the values of auxiliary variable x are not always available therefore

in future research we should extend our study to include cases where the parameter of auxiliary variable x is not known.

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: 262-275.
<https://doi.org/10.1017/S0021859600048012>
- Hansen, M.N. & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236):517- 529.
<https://doi.org/10.1080/01621459.1946.10501894>
- Khare, B.B. & Rehman, H.U. (2015). Improved ratio in regression type estimator for population mean using known coefficient of variation of the study character in the presence of non-response. *HCTL Open International Journal of Technology Innovations and Research*, 14: 1-7.
- Khare, B.B. & Sinha, R.R. (2009). On class of estimators for population mean using multi-auxiliary character in presence of non-response. *Statistics in Transition-New Series*, 10: 3-14.
- Khare, B.B. & Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response. *National Academy Science Letters-India*, 16(3): 111-114.
- Khare, B.B. & Srivastava, S. (1995). Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. *Proceedings-national Academy Science, India, Section A*, 65: 195-204.
- Khare, B.B. & Kumar, S. (2009). Transformed two phase sampling ratio and product type estimators for population mean in the presence of non-response. *Aligarh Journal of Statistics*, 29: 91-106.
- Khoshnevisan, M., Singh, R. Chauhan, P. Sawan N. & Smarandache, F. (2007). A general family of estimators for estimating population mean using known value of some population parameter(s). *Far East J. Theoretical Statistics*, 22: 181-191.
- Murthy, M.N. (1964). Product method of estimation. *Sankhya. Series A*, 26: 69-74.
- Pandey, B. N. & Dubey, V. (1988). Modified product estimator using coefficient of variation of auxiliary variable. *Assam Statistical Review*, 2: 64-66.
- Rachokarn, T. & Lawson, N. (2017a). Improvement in estimating the population mean using exponential type estimator in the presence of non-response. *International Journal of*

Scientific and Research Publications, Volume 7, Issue 6, 87-90.

<https://doi.org/10.5614/j.math.fund.sci.2017.49.3.6>

- Rachokarn, T. & Lawson, N. (2017b). An efficient general family of estimators for population mean in the presence of non-response. *Journal of Mathematical and Fundamental Sciences*, 49(3), 283-293.
- Rachokarn, T. & Lawson, N. (2017c). A class of ratio chain type exponential estimator for population mean in the presence of non-response. *International Journal of Agricultural and Statistical Sciences*, 13(2), 431-437.
- Robson, D.S. (1957). Application of multivariate polykeys to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 52: 511-522.
<https://doi.org/10.1080/01621459.1957.10501407>
- Searls, D.T. (1964). The Utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, 59: 1225-1226.
<https://doi.org/10.1080/01621459.1964.10480765>
- Singh, J. Pandey, B. N. & Hirano, K. (1973). On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Annals of the Institute of Statistical Mathematics*, 25: 51-55. <https://doi.org/10.1007/BF02479358>
- Sisodia, B.V.S. & Dwivedi, V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of the Indian Society of Agricultural Statistics*, 33: 13-18.
- Soponviwatkul, K & Lawson, N. (2017). New ratio estimators for estimating population mean in simple random sampling using a coefficient of variation, correlation coefficient and a regression coefficient. *Gazi University Journal of Science*, 30(4), 610-621.
- Upadhyaya, L.N. & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41: 627-636.
[https://doi.org/10.1002/\(SICI\)1521-4036\(199909\)41:5<627::AID-BIMJ627>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.0.CO;2-W)