DOI: http://doi.org/10.4038/sljastats.v22i1.8049

Sri Lankan Journal of Applied Statistics, Vol (22-1)

Estimator of Domain Mean Using Stratified Sampling in the Presence on Non-Response

Ashutosh[1*]

[1]Department of Statistics, Faculty of Science & Technology, MGKVP, Varanasi-221002

[*]Corresponding Author: kumarashubhustat@gmail.com

Received: 03rd June 2021 / Revised: 14th August 2021 / Published: 30th August 2021

ABSTRACT

The present work investigated a direct estimator of domain mean which helps for future work in the field of small area estimation. Our investigation will helpful when the availability of non-response in strata which may or may not be the same in all the domains. We discussed the proposed estimators for domain mean utilizing stratified sampling with non-response and also studied its properties. Proposed estimator has compared with a direct ratio estimator for domain mean utilizing stratified sampling with non-response where, non-response is available approximately (30%) in the domain. We considered two situations in the first, non-responses are approximately (30%). However, in second case, different non-responses like 20 % and 40 % in the strata 1 and 2 respectively. An empirical study has been carried out for the data Sarndal et al. (1992) in terms of the mean square error. We obtained that proposed generalized investigation is more efficient than ratio estimator in case i over case ii. It is analyzed that the direct generalized investigation is a better choice over direct ratio estimate with or without non-responses in both the cases.

Key words: Direct Generalized Stratified Estimator, Non-Response, Mean Square Error, Domain Mean, Auxiliary Character.

1. Introduction

Estimation of domain parameter plays significant role in recent years due to increasing the vital interest in the government policy and well-organized plan for, distribution of different facilities such as flood affected regions, coastal regions, socio-economic, depressed areas, topography, etc in the demanded regions. The main difficulties arise during estimation of the domains parameter when low number of sample units regarding the domains. When sample units in the domain is accessible.

The direct estimator for domain mean gives a better result over indirect estimates for study domains. However, units in the domain are not accessible. The indirect estimator is a better choice than the direct estimator (Gonzalez, 1973). In this relation, the imminent contributions have been discussed by Ghosh and Rao (1994) and Rehman (2008). Several investigations have been utilized the indirect estimator, especially based on synthetic estimator (Tikkiwal and Ghiya 2000, Rao 2003, Singh and Seth 2014 and Ashutosh 2020). The model-based approach has been studied (Purcell and Kish 1979 and Tikkiwal et al. 2013). However, estimation of domain mean utilizing stratified sampling (Clement et al. 2014 and Aditya et al. 2014).

If, we select a sample from interested domain rather than the population termed as a direct estimator. Most of the estimators based on direct method for domain estimation through the model based approach have been discussed (Cochran, 1977, Sarndal et al. 1992 and Salveti et al. 2012). Whenever, if the units in the study domain is reliable, but due to the high variation within the sample units in the study domain, the traditional estimators do not give good results, in such conditions, we use another method of estimation than the traditional estimator.

Due to complex types of demand in the industries, a typical situation arises the non-response units present in the strata within the domain. In such a scenario, estimates could not give precise result. Hence, we use the idea of Hansen and Hurwitz (1946). Alilah et al. (2020) have been discussed about the non-response with two stage sampling. In the present work, we proposed stratified based estimators for domain mean with non-response. We consider non-response on the study character. However, response units are available of auxiliary character corresponding to study character.

2. Methodology

Suppose L independent domains consist of the units U_a $(a=1,2,3,\ldots\ldots L)$ of sizes N_a. Each of the domains are stratified in to h^{th} strata of a^{th} domains $U_{h,a}$ have size $N_{h,a}$ $(a=1, 2, 3, \ldots\ldots L, h=1, 2, \ldots H)$. A random sample $s_{h,a}$ is selected in the h^{th} stratum and a^{th} domain of size $n_{h,a}$ of $U_{h,a}$ with size $N_{h,a}$ through simple random sampling without replacement. We represent the study character by y and auxiliary character by x. Where,

$$\sum_{a=1}^{L} N_a = N, \quad \sum_{a=1}^{L} n_a = n, \quad \sum_{h=1}^{H} N_{h,a} = N_a, \quad \sum_{h=1}^{H} n_{h,a} = n_a \quad \text{and}$$

$$\sum_{h=1}^{H} W_{h,a} = 1 \tag{1}$$

where, $\sum_{h=1}^{H} W_{h,a}$: shows sum of h^{th} stratum weights is equal to 1 for and a^{th} domains.

Notations are presenting here:

$\bar{Y}_a$: a^{th} domain mean of y based on N_a *observations*.

$\bar{X}_a$: a^{th} domain mean of x based on N_a observations.

$\bar{Y}_{h,a}$: a^{th} domain mean of h^{th} stratum of y based on $N_{h,a}$ observations.

$\bar{X}_{h,a}$: a^{th} domain mean of h^{th} stratum of x based on $N_{h,a}$ observations.

$\bar{y}_{h,a}$: Sample mean of a^{th} domain, h^{th} stratum of y based on $n_{h,a}$ observations.

$\bar{x}_{h,a}$: Sample mean of a^{th} domain, h^{th} stratum of x based on $n_{h,a}$ observations.

For estimation of non-respondent, we select a sample of h^{th} stratum and a^{th} domain of size $n_{h,a}$ from $N_{h,a}$. We notice that the selected sample have only $n_{1h,a}$ respondent units and $n_{2h,a}$ non-respondent units and obtain the value of the non-respondent units. We take a sub-sample of $r_{h,a}$ $(n_{2h,a} / g_{h,a}\,,\ g_{h,a} > 1)$ from non-respondent in h^{th} stratum $n_{2h,a}$ units by using Hansen and Hurwitz (1946) technique of non-respondent $n_{1h,a}$ and $r_{h,a}$ units on y and x are given by

$$\bar{y}^{*}_{h,a} = \frac{n_{1h,a}\,\bar{y}_{1h,a}}{n_{h,a}} + \frac{n_{2h,a}\,\bar{y}^{1}_{2h,a}}{n_{h,a}} \quad \text{and} \quad \bar{x}^{*}_{h,a} = \frac{n_{1h,a}\,\bar{x}_{ih,a}}{n_{h,a}} + \frac{n_{2h,a}\,\bar{x}^{1}_{2h,a}}{n_{h,a}} \tag{2}$$

Where, $(\bar{y}_{1h,a}, \bar{y}^{1}_{2h,a})$ and $(\bar{x}_{1h,a}, \bar{x}^{1}_{2h,a})$ denotes the mean of $n_{1h,a}$ and $r_{h,a}$ units of y and x.

We denote,

$n_{1h,a}$ = Number of respondent units in the h^{th} stratum and a^{th} domain of the study character.

$n_{2h,a}$ = Number of non-respondent units in h^{th} stratum and a^{th} domain auxiliary character.

$\bar{y}_{1h,a}$: Sample mean of h^{th} stratum and a^{th} domain of y on $n_{1h,a}$ observations.

$\bar{y}^{1}_{2h,a}$: Sample mean of h^{th} stratum and a^{th} domain of y on $n_{2h,a}$ observations.

$N_{1h,a}$ = Number of respondent units h^{th} stratum, a^{th} domain.

$N_{2h,a}$ = Number of non-respondent units of h^{th} stratum, a^{th} domain.

$\bar{x}_{1h,a}$: Sample mean of h^{th} stratum and a^{th} domain, of x based on $n_{1h,a}$ observations.

$\bar{x}^{1}_{2h,a}$: Sample mean of h^{th} stratum and a^{th} domain, of x based on $n_{2h,a}$ observations.

$W_{1h,a} = \dfrac{N_{1h,a}}{N_{h,a}}$ response rate of h^{th} stratum, a^{th} domain.

$W_{2h,a} = \dfrac{N_{2h,a}}{N_{h,a}}$ Non-response rate of h^{th} stratum, a^{th} domain.

Let study character y of i^{th} observation of a^{th} domain, h^{th} stratum $y_{hi,a}$ $(i = 1, 2, 3, \ldots, N_{h,a};\ a = 1, 2, 3, \ldots, L;\ h = 1, 2, 3, \ldots, H)$ of x of i^{th} observation of h^{th} stratum for a^{th} domain $x_{hi,a}$ $(i = 1, 2, 3, \ldots, N_{h,a};\ a = 1, 2, 3, \ldots, L;\ h = 1, 2, 3, \ldots, H)$ and their domain units $U_{hi,a}$ $(i = 1,2,3,\ldots., N_{h,a},\ a = 1,2,3,\ldots., L,\ h = 1,2,3,\ldots H)$.

Values of the h^{th} stratum variance, covariance and coefficient of variation of a^{th} domain of y and x are written as

$$S^2_{Yh,a} = \frac{1}{(N_{h,a} - 1)} \sum_{i=1}^{N_{h,a}} (y_{hi,a} - \bar{Y}_{h,a})^2, S^2_{Xh,a} = \frac{1}{(N_{h,a} - 1)} \sum_{i=1}^{N_{h,a}} (x_{hi,a} - \bar{X}_{h,a})^2,$$

$S_{YXh,a} = \frac{1}{(N_{h,a}-1)} \sum_{i=1}^{N_{h,a}} (y_{hi,a} - \overline{Y}_{h,a})(x_{hi,a} - \overline{X}_{h,a})$, $C_{Yh,a} = \frac{S_{Yh,a}}{\overline{Y}_{h,a}}$, $C_{Xh,a} = \frac{S_{Xh,a}}{\overline{X}_{h,a}}$ and $C_{XYh,a} = \frac{S_{YXh,a}}{\overline{X}_{h,a}\overline{Y}_{h,a}}$, $\sum_{h=1}^{H} W_{h,a}\overline{Y}_{h,a} = \overline{Y}_a$ and

$$\sum_{h=1}^{H} W_{h,a}\overline{X}_{h,a} = \overline{X}_a \ . \tag{3}$$

3. Estimation of Domain Mean Using Stratified Sampling

Stratified sampling is used when variation within the domain is high, so the domain is sub-divided according to homogeneous strata, which is low variation than the domain. Here, we are discussing the ratio and regression estimators for domain mean.

(i) Direct Ratio Estimator:

Ratio estimator is an updated form of the design-based estimator with incorporate auxiliary character. Khare et al. (2018) have been discussed a direct ratio estimator utilizing x. Tikkiwal and Ghiya (2000) have been discussed direct ratio estimator with stratified sampling is

$$T_{DR,st,a} = \frac{\overline{y}_{st,a}}{\overline{x}_{st,a}}\overline{X}_{h,a} \qquad \text{Tikkiwal and Ghiya (2000)} \tag{4}$$

where, a^{th} domain mean of y: $\overline{y}_{st,a} = \sum_{h=1}^{H} W_{h,a}\overline{y}_{h,a}$ and a^{th} domain mean of x: $\overline{x}_{st,a} = \sum_{h=1}^{H} W_{h,a}\overline{x}_{h,a}$

Bias and mean square error of $T_{DR,st,a}$

$$Bias(T_{DR,st,a}) = E\left(\frac{\overline{y}_{st,a}}{\overline{x}_{st,a}}\overline{X}_{h,a} - \overline{Y}_a\right)$$

by using large sample approximations, we have

$$Bias(T_{DR,st,a}) = \sum_{h=1}^{H} W_{h,a}\overline{Y}_{h,a}\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C_{YXh,a} - \overline{Y}_a \tag{5}$$

and

$$MSE(T_{DR,st,a}) = E\left(\frac{\bar{y}_{st,a}}{\bar{x}_{st,a}}\bar{X}_{h,a} - \bar{Y}_a\right)^2$$

$$= \left(\frac{\sum_{h=1}^{H} W_{h,a}\bar{y}_{h,a}}{\sum_{h=1}^{H} W_{h,a}\bar{x}_{h,a}}\bar{X}_{h,a} - \bar{Y}_a\right)^2$$

After solving, we get

$$MSE(T_{DR,st,a}) = \sum_{h=1}^{H} W_{h,a}^2 \bar{Y}_{h,a}^2 \frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}} \left\{C_{Yh,a}^2 + C_{Xh,a}^2 - 2C_{YXh,a}\right\} \tag{6}$$

(ii) Direct generalized estimator:

Generalized estimator is an application of the auxiliary character which is a modern type of the estimator. The performance of the generalized is more efficient than the traditional design estimator and ratio estimator using auxiliary character x. The direct generalized estimate for domain mean has been explained (Tikkiwal and Ghiya 2000)

$$T_{DG,st,a} = \bar{y}_{st,a}\left(\frac{\bar{x}_{st,a}}{\bar{X}_{h,a}}\right)^{\delta} \qquad \text{Tikkiwal and Ghiya (2000)} \tag{7}$$

Bias and MSE of $T_{DG,st,a}$ are given as:

$$Bias(T_{DG,st,a}) = \sum_{h=1}^{H} W_{h,a}\bar{Y}_{h,a}\frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}}\left\{\frac{\delta(\delta-1)}{2}C_{Xh,a}^2 + \delta C_{YXh,a}\right\} - \bar{Y}_a \tag{8}$$

and

$$MSE(T_{DG,st,a}) = \sum_{h=1}^{H} W_{h,a}^2 \bar{Y}_{h,a}^2 \frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}}\left\{C_{Yh,a}^2 + \delta^2 C_{Xh,a}^2 + 2\delta C_{YXh,a}\right\} \tag{9}$$

Now for optimum value of δ, the MSE of $T_{DG,st,a}$, partially differentiate w.r.to δ, equate to zero i.e., $\frac{\partial MSE(T_{DG,st,a})}{\partial \delta} = 0$. After solving, we have

$$\delta_{opt} = -\frac{C_{YXh,a}}{C_{Xh,a}^{2}} \tag{10}$$

Substitute δ_{opt} in the Equation (9), we have

$$MSE(T_{DG,st,a,opt}) = \sum_{h=1}^{H} W_{h,a}^{2} \overline{Y}_{h,a}^{2} \frac{N_{h,a} - n_{h,a}}{N_{h,a} n_{h,a}} C_{Yh,a}^{2} \left(1 - \rho_{YXh,a}^{2}\right) \tag{11}$$

4. Proposed Estimator

In this section, we investigate a problem of non-response, which is generally occurred due to diminished information or miss information about the units of the interested domains. In some of the situations, study domains contained high variation and non-response available in the auxiliary character. Hence, we propose a direct generalized estimator for domain mean through stratified sampling with non-response

$$T_{DG,st,\beta,a}^{*} = \overline{y}_{st,a}^{*} \left(\frac{\overline{x}_{st,a}^{*}}{\overline{\overline{X}}_{h,a}} \right)^{\beta}, \tag{12}$$

Where β is a chosen constant of a^{th} domain and the value of y of non-respondents can be written by

$\overline{y}_{st,a}^{*} = \sum_{h=1}^{H} W_{h,a} \overline{y}_{h,a}$, and a^{th} domain mean of x of non-respondents are given by $\overline{x}_{st,a}^{*} = \sum_{h=1}^{H} W_{h,a} \overline{x}_{h,a}$.

Members of the proposed estimator $T_{DG,st,\beta,a}^{*}$:

(i) $T_{DG,st,0,a}^{*} = \overline{y}_{st,a}^{*}$, if $\beta = 0$ (13)

(ii) $T_{DG,st,-1,a}^{*} = \frac{\overline{y}_{st,a}^{*}}{\overline{x}_{st,a}^{*}} \overline{X}_{h,a}$, if $\beta = -1$ (14)

(iii) $T_{DG,st,1,a}^{*} = \overline{y}_{st,a}^{*} \frac{\overline{x}_{st,a}^{*}}{\overline{\overline{X}}_{h,a}}$, if $\beta = 1$ (15)

(iv) $T^{*}_{DG,st,2,a} = \bar{y}^{*}_{st,a}\left(\frac{\bar{x}^{*}_{st,a}}{\bar{X}_{h,a}}\right)^{2}$, if $\beta = 2$ (16)

Bias and Mean Square Error of members of $T^{*}_{DG,st,\beta,a}$:

We assume the following assumption for large sample approximations:

$\bar{y}_{h,a} = \bar{Y}_{h,a}(1+\varepsilon_0)$, $\bar{x}_{h,a} = \bar{X}_{h,a}(1+\varepsilon_1)$, such that $E(\varepsilon_0) = 0$, $E(\varepsilon_1) = 0$,

$|\varepsilon_i| < 1;\ i = 0, 1$ $E(\varepsilon_0^2) = \frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}} C^2_{Yh,a} + \frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}} C^2_{2Yh,a}$,

$E(\varepsilon_1^2) = \frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}} C^2_{Xh,a} + \frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}} C^2_{2Yh,a}$

and $E(\varepsilon_0\varepsilon_1) = \frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}} C_{YXh,a} + \frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}} C_{2YXh,a}$.

Where, $g_{h,a} = \frac{n_{2h,a}}{r_{h,a}}$, $g_{h,a} \geq 1$ and $W_{2h,a} = \frac{N_{2h,a}}{N_{h,a}}$. (17)

(i) Bias and Mean Square Error of $T^{*}_{DG,st,-1,a}$:

After using the large sample approximations (17) in (14), we get

$$Bias(T^{*}_{DG,st,-1,a}) = \sum_{h=1}^{H} W_{h,a}\bar{Y}_{h,a}\left\{\frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}} C^2_{Xh,a} + \frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}} C^2_{2Yh,a}\right\} - \bar{Y}_a \tag{18}$$

and

$$MSE(T^{*}_{DG,st,-1,a}) = \sum_{h=1}^{H} W^2_{h,a}\bar{Y}^2_{h,a}\left[\frac{N_{h,a} - n_{h,a}}{N_{h,a}n_{h,a}}\left\{C^2_{Yh,a} + C^2_{Xh,a} - 2C_{YXh,a}\right\}\right.$$

$$\left. + \frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}\left\{C^2_{2Yh,a} + C^2_{2Xh,a} - 2C_{2YXh,a}\right\}\right] \tag{19}$$

Bias and MSE of $T^{*}_{DG,st,\beta,a}$:

Putting in the Equation (12) by using the (17) obtained the Bias and MSE, which are given as follows:

$$Bias\left(T_{DG,st,\beta,a}\right)=\sum_{h=1}^{H}W_{h,a}\bar{Y}_{h,a}\left[\frac{\beta(\beta-1)}{2}\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C^2_{Xh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C_{2Xh,a}\right\}\right.$$

$$\left.+\beta\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C_{XYh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C_{2YXh,a}\right\}\right]-\bar{Y}_a \qquad (20)$$

and

$$MSE\left(T_{DG,st,\beta,a}\right)=\sum_{h=1}^{H}W^2_{h,a}\bar{Y}^2_{h,a}\left[\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C^2_{Yh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C^2_{2Yh,a}\right\}+\right.$$

$$\left.\beta^2\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C^2_{Xh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C^2_{2Xh,a}\right\}+2\beta\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C_{YXh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C_{2YXh,a}\right\}\right] \qquad (21)$$

The optimum value of β, partially differentiate Equation (20) by β and equate to zero, we have

$$\beta_{opt}=-\frac{\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C^2_{Xh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C^2_{2Xh,a}\right\}}{\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C_{YXh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C_{2YXh,a}\right\}} \qquad (22)$$

Substituting β_{opt} in the Equation (22), we have

$$MSE\left(T_{DG,st,\beta,a,opt}\right)=\sum_{h=1}^{H}W^2_{h,a}\bar{Y}^2_{h,a}\left[\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}n_{h,a}}C^2_{Yh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C^2_{2Yh,a}\right\}\right.$$

$$-\frac{\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}\,n_{h,a}}C_{YXh,a}+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C_{2YXh,a}\right\}^2}{\left\{\frac{N_{h,a}-n_{h,a}}{N_{h,a}\,n_{h,a}}C_{Xh,a}^2+\frac{(g_{h,a}-1)W_{2h,a}}{n_{h,a}}C_{2Xh,a}^2\right\}}\Bigg] \quad (23)$$

5. Empirical Study

We take the Sweden municipalities MU284 (Sarndal et al. 1992, appendix B). The population is geographically sub-divided (domain) into eight different parts 1, 2, 3, 4, 5, 6, 7 and 8 having their sizes 25, 48, 32, 38, 56, 32, 41, 15 and 29 respectively. However, we considered only four domains 2, 4, 5 and 6 because these domains have large units compared to other domains. The proposed estimator is a kind of direct estimator. Then each of the domains is classified into homogeneous according to our convenient in to two strata: value of below 1500 (millions of kronor) and above 1500 (millions of kronor). We consider two cases i and ii of non-response (in both Population I and Population II).

Case i: If non-respondents are available in both strata (1 and 2) as well as in the domains (approximately 30%).

Case ii: If different non-respondents are available in both strata 1 and 2 approximately 20% and 40% respectively.

Population I

y: Real estate values according to 1984 assessment (in millions of kronor).

x: Total number of municipal employees in 1984.

Table 1: Value of parameters of the strata (1 and 2) and domains

Domain Values	**Domain**							
	48		38		56		41	
	1	2	1	2	1	2	1	2
$N_{h,a}$	30	18	15	23	25	31	19	22
$W_{h,a}$	0.625	0.375	0.395	0.605	0.446	0.554	0.463	0.537
$\overline{Y}_{h,a}$	1195.10	5930.722	1216.933	4023.261	961.84	4727.613	1060.947	3137.727
$\overline{X}_{h,a}$	569.833	3473.444	576.467	2825.478	439.920	3168.516	487.105	1628.864

$S^2_{Yh,a}$	161392	156245 49	32261. 78	128341 42	65851 .31	444824 56	10276 2.2	328283 5
$S^2_{Xh,a}$	63455. 18	703536 7	19051. 41	247862 85	14523 .99	676622 10	21797 .21	129212 0
$S_{YXh,a}$	81507. 67	989373 1	17316. 75	170581 98	21561 .94	534888 10	36389 .01	200112 0
$\rho_{YXh,a}$	0.805	0.943	0.698	0.956	0.697	0.975	0.769	0.972

Table 2: The parameter values of strata (1 and 2) for domains (1, 2, 3 and 4) in case i

Domain	**Strata**	$S^2_{2Yh,a}$	$S^2_{2Xh,a}$	$S_{2YXh,a}$	$g_{h,a}$	$n_{2h,a}$	$W_{h,a}$
1	1	48417.6	19036.6	24452.3	3	5	0.30
	2	4687365	2110610.1	2968119.3	2	3	0.30
2	1	9678.53	5715.42	5195.03	2	3	0.30
	2	3850242.6	7435885.5	5117459.4	2	4	0.30
3	1	19755.4	4357.2	6468.6	2	4	0.30
	2	13344736.8	20298663	16046643	3	5	0.30
4	1	30828.7	6539.2	10916.7	2	3	0.30
	2	984850.5	387636.0	600336	2	3	0.30

Table 3: The parameter values of strata (1 and 2) for domains (1, 2, 3 and 4) in case ii

Domain	**Strata**	$S^2_{2Yh,a}$	$S^2_{2Xh,a}$	$S_{2YXh,a}$	$g_{h,a}$	$n_{2h,a}$	$W_{h,a}$
1	1	32278.4	12691	16301.5	2	3	0.20
	2	6249819.6	2814146.8	3957492.4	3	4	0.40
2	1	6452.4	3810.3	3463.4	2	2	0.20
	2	5133656.8	9914514	6823279.2	3	5	0.40
3	1	13170.3	2904.8	4312.4	2	3	0.20
	2	17792982.4	27064884	21395524	4	6	0.40
4	1	20552.4	4359.4	7277.8	2	2	0.20
	2	1313134	516848	800448	3	5	0.40

Population II

Another population is considered (Sarndal et al. 1992, appendix B) which is classified in to four domains with stratum 1 and 2 according to the revenues less than 1500 (in millions of kronor) and revenues above 1500 (in millions of kronor).

y: Revenues of 1985 municipal taxation assessment (in millions of kronor).

x: 1985 population (in thousands).

Table 4: The parameter value of the strata for the domains (1, 2, 3 and 4)

Domain Values	**Domain**							
	48		32		38		56	
	1	2	1	2	1	2	1	2
$N_{h,a}$	22	26	14	18	14	24	29	27
$W_{h,a}$	0.458	0.542	0.436	0.564	0.368	0.632	0.518	0.482
$\overline{Y}_{h,a}$	65.318	376.153	67.5	260.61	75.857	376.5	63.448	498.704
$\overline{X}_{h,a}$	9.318	46.846	10.642	34.5	12.357	41.875	9.862	50.370
$S^2_{Yh,a}$	417.75	129916.4	275.9615	41200.84	143.362	466644.2	304.113	1565813
$S^2_{Xh,a}$	7.465	1801.415	4.555	544.971	4.863	2133.853	5.695	5924.088
$S_{YXh,a}$	53.846	10243.34	32.038	4559.147	23.747	30428.41	39.385	95437.92
$\rho_{YXh,a}$	0.964	0.670	0.904	0.962	0.899	0.964	0.946	0.991

Table 5: The parameter values of strata in case i for all domains (1, 2, 3 and 4)

Domain	**Strata**	$S^2_{2Yh,a}$	$S^2_{2Xh,a}$	$S_{2YXh,a}$	$g_{h,a}$	$n_{2h,a}$	$W_{h,a}$
1	1	125.33	2.234	16.154	2	3	0.30
	2	38974.92	540.424	3073.002	2	4	0.30
2	1	82.79	1.366	9.6114	2	2	0.30
	2	12360.25	163.49	1367.744	2	3	0.30
3	1	43.01	1.4589	7.1241	2	2	0.30
	2	139993.26	640.156	9128.523	2	4	0.30
4	1	91.234	1.7085	11.815	3	5	0.30
	2	469743.90	1777.226	28631.376	2	4	0.30

Table 6: Parameter values of strata (1 and 2) for each domain in the case ii

Domain	**Strata**	$S^2_{2Yh,a}$	$S^2_{2Xh,a}$	$S_{2YXh,a}$	$g_{h,a}$	$n_{2h,a}$	$W_{h,a}$
1	1	83.55	1.493	10.7692	2	2	0.20
	2	51966.56	720.566	4097.336	3	5	0.40
2	1	55.192	0.9110	6.4076	2	2	0.20
	2	16480.34	217.988	1823.658	2	4	0.40
3	1	28.67	0.9726	4.7494	2	2	0.20
	2	186657.68	853.541	12171.364	3	5	0.40
4	1	60.82	1.139	7.877	2	3	0.20
	2	626325.20	2369.635	38175.168	3	6	0.40

Table 7: MSE of estimators ($T_{DG,st,a}$), ($T_{DR,st,a}$) and MSE of ($T^*_{DG,\beta,st,a}$), ($T^*_{DG,-1,st,a}$) in both the case i and ii: (For Population I)

Estimator	**Domain**			
	1	2	3	4
$T_{DR,st,a}$	19637.51	212264.3	374598.8	5300.008
$T_{DG,st,a}$	14106.54	16130.59	23426.3	2831.272
	-0.612456*	-0.430573	-0.679005	-0.766475
	-0.823619*	-0.483321	-0.529823	-0.803970
$T^*_{DG,-1,st,a}$	25831.8	283289.8	591459.2	7187.966
$T^*_{DG,\beta,st,a}$	18559.97	21412.03	36860.98	3375.673
	-0.612456**	-0.430573	-0.679005	-0.804845
	-0.823619**	-0.483321	-0.529823	-0.80397
$T^*_{DG,-1,st,a}$	27174.06	398674.2	689628.4	7710.016
$T^*_{DG,\beta,st,a}$	19601.51	30111.45	42920.12	3655.99
	-0.612456***	-0.430573	-0.679005	-0.747439
	-0.823619***	-0.483321	-0.529823	-0.803970

* represent value of δ in the ($T_{DG,st,a}$), ** represent β in the proposed estimator in case i, *** represent β in the proposed estimator in case ii.

We obtained from Table 7, value of MSE of $T^*_{DG,\beta,st,a}$ is lower than the value of MSE of $T^*_{DG,-1,st,a}$ for all domains (1, 2, 3 and 4). It is seen that in both the cases i and ii β is different for stratum 1 in the domain 1. However, value of β is different for both stratum in the domains 2, 3 and 4 different, but it is not affected on the performance of the generalized estimator. But, we have seen that the generalized estimator with non-response under case i, is better as compared to cases ii in terms of MSE. It is also seen that the generalized estimator without non-response is lower value as compared to both the cases. Hence, empirically it proved the general theory of the non-response. This picture is also seen for ratio estimator in all the domains. We observed that the similar pattern have been seen for Tikkiwal and Ghiya estimator for all domains (1, 2, 3 and 4) when different value of δ in the strata (1 and 2).

Table 8: MSE of $T_{DR,st,a}$, $T_{DR,st,a}$ and $T^*_{DG,-1,st,a}$, $T^*_{DG,\beta,st,a}$ in both the cases i and ii of domains (1, 2, 3 and 4) (For Population II)

Estimator	**Domain**			
	1	2	3	4
$T_{DR,st,a}$	920.681	60.819	1529.028	2373.952
$T_{DG,st,a}$	809.055	54.475	544.417	261.900
	-1.028968*	-1.109029	-0.795542	-1.075035
	-0.708168*	-1.10748	-1.586005	-1.627167
$T^*_{DG,-1,st,a}$	1091.187	67.863	1798.264	3250.886
$T^*_{DG,\beta,st,a}$	958.421	59.165	639.723	358.618
	-1.028968**	-1.109029	-0.795542	-1.075035
	-0.708168**	-1.107482	-1.586005	-1.627167
$T^*_{DG,\beta,st,a}$	1204.381	71.886	2040.148	3981.104
$T^*_{DG,\beta,st,a}$	1057.929	62.945	725.854	439.048
	-1.028968***	-1.109029	-0.795542	-1.075035
	-0.708168***	-1.107482	-1.586005	-1.627167

* Represent the δ in the ($T_{DG,st,a}$), ** Represent the β in the proposed estimator in the case i, *** Represent value of β in the proposed estimator in the case ii.

The important points has been seen from Table 8 that the MSE of $T_{DG,st,a}$ is less than the MSE of $T_{DR,st,a}$ for the domains 1, 2, 3 and 4. However, value of β is different for both stratum in the domains 2, 3 and 4 different. It has seen that the generalized and ratio estimator with non-response under case i, is better as compared to cases ii in terms of MSE. It is also seen that the MSE of generalized and ratio estimators without non-response situation is lower than the MSE of generalized and ratio estimators with non-response. Results proved that the theory of the non-response and without non-response in the population I.

Table 9: Relative efficiency (RE) of $T_{DG,st,a}$ and $T_{DR,st,a}$ are given in terms of (%) for domains (1, 2, 3 and 4): (For both Populations I and II)

Population	**Estimator**	**Domain**			
		1	2	3	4
I	$T_{DR,st,a}$	100.000	100.000	100.000	100.000
	$T_{DG,st,a}$	139.209	1315.912	1599.052	187.200
II	$T_{DR,st,a}$	100.000	100.000	100.000	100.000
	$T_{DG,st,a}$	113.800	111.647	280.856	906.434

From Table 9, we can say that the RE (in terms %) of $T_{DG,st,a}$ is less than the RE (in terms of %) of $T_{DR,st,a}$ for all domains (1, 2, 3 and 4) in both the populations I and II.

Table 10: RE (in %) of the estimators $T^*_{DG,\beta,st,a}$ and $T^*_{DG,-1,st,a}$ in the presence of approximately (30%) non-response of $T^*_{DG,\beta,st,a}$ and $T^*_{DG,\beta,st,a}$ in both cases i and ii for all domains (1, 2, 3 and 4): (Population I and Population II)

Populations		**Estimator**	**Domain**			
			1	2	3	4
I		$T^*_{DG,-1,st,a}$	100.000	100.000	100.000	100.000
	i	$T^*_{DG,\beta,st,a}$	139.327	1323.040	1604.567	212.9343
	ii	$T^*_{DG,\beta,st,a}$	138.635	1323.995	1606.772	210.8873
II		$T^*_{DG,-1,st,a}$	100.000	100.000	100.000	100.000
	i	$T^*_{DG,\beta,st,a}$	113.853	114.703	281.100	906.504
	ii	$T^*_{DG,\beta,st,a}$	113.843	114.204	281.069	906.757

It is seen from the Table 10, the RE (in terms of %) of $T^*_{DG,\beta,st,a}$ is less than $T^*_{DG,-1,st,a}$ for all domains (1, 2, 3 and 4) in both the cases i and ii for population I and population II. The generalized estimator is better performed in the population I than population II which is due to the accessible value in the domain. The low effect of the non-respondent present within the stratum is than the between stratum.

6. Results

From Table 7 to Table 10, we analyzed that the MSE of $T^*_{DG,\beta,st,a}$ is lower than $T^*_{DG,-1,st,a}$ for all domains (1, 2, 3 and 4) in both the case i and ii for both population I and population II (see Table 9). MSE of $T^*_{DG,\beta,st,a}$ is greater than the $T_{DG,st,a}$ due to presence of non-response in the proposed estimator for domain mean $T^*_{DG,\beta,st,a}$ for domains (1, 2, 3 and 4) (for Populations I, II). RE of the proposed estimator for domain mean of $T^*_{DG,\beta,st,a}$ is less than in the cases i and ii for domains (1, 2, 3 and 4) for both populations I, II (see Table 10). RE of the proposed estimator for domain mean is higher for domain 2 and 3 than 1 and 4 due to high correlation between strata (For Population I). RE of the proposed estimator for domain mean of $T^*_{DG,\beta,st,a}$ is higher for domain 4 than the domains 1, 2 and 3 due to high correlation within domain and between strata (For Population II). RE of the proposed estimator for domain mean is lower than the ratio estimator in the discussed cases is approximately same because, due to due to loss of information in both the estimators are same.

7. Conclusion

The proposed estimator in the presence of non-response is better performed in the situation I than situation II. The proposed estimator for domain mean is better than the direct ratio estimator for domain mean with auxiliary character in the presence of non-response. Hence, the proposed estimator for domain mean is preferred over the direct ratio estimator for domain mean in the presence of non-response in domain and strata.

8. Applications

The present investigation may be used those areas like industry where missing units are present. It may also enhance for the scientific purpose of nuclear data, bio-statistical data and local areas problem related to socio-economic and lost due to COVID-19 (other diseases also), etc.

References

1. Aditya, K., Sud, U. C. and Ghrade, Y. (2014) Estimation of domain mean using two stage sampling with sub sampling of non-respondents. *Journal of Indian Social Agricultural Statistics. 68*, 1, 39-54.
2. Alilah, D. A., Onayango, C, O. and Ombaka, E. O. (2020) Efficiency of Domain Mean Estimators in the Presence of Non-response Using Two-Stage Sampling

with Non-linear and Linear Cost Function. *Annals of Data Science,* DOI: 10.1007/s40745-020-00312-x

3. Ashutosh. (2020): Estimation of domain mean using conventional synthetic estimator with two auxiliary characters. *Annals of Data Science,* DOI: 10.1007/s40745-020-00287-9
4. Climent, E. P, Udofia, G. A. and Enang, E. I. (2014) Estimation for domain in stratified sampling design in the presence of non-response. *American Journal of Mathematics & Statistics. 4, (2),* 65-71.
5. Cochran, W. G. (1977) *Sampling Techniques.* 3rd Edition, John Wiley and Sons.
6. Gonzalez, M. E. (1973) Use and evaluation of synthetic estimators, *Proceedings of the social Statistics, American Statistical Association,* 33-36.
7. Hansen, M. H. and Hurwitz, W. N. (1946) Problem of non-response in sample surveys. *Journal of American Statistician,* 41, 517-529.
8. Khare. B. B., Ashutosh. and S. Khare, (2018) Modified direct regression estimator for domain mean using auxiliary character. *International Journal of Applied Mathematics and Statistics. 57, (6),* 10-20.
9. Purcell, N. J., and Kish, L. (1979) Estimates for small domains. *Biometrics, 35,* 365-384.
10. Rahman, A. (2008) *A review of small area estimation problems and methodological developments.* NATSEM, University of Canada.
11. Ghosh, M. and Rao. J. N. K. (1994) Small area estimation: an appraisal. *Statistical Science, 9, (1),* 55-76.
12. Rao. J. N. K. (2003) *Small area estimation,* Wiley Inter-Science, John Wiley and Sons, New Jersey.
13. Salveti. et al. (2012) Model based direct estimator of small area distribution. *Australia & New Zealand Journal of Statistics. 54, (1),* 103-123.
14. Sarndal, C.E. Swensson, B. and Wretman, J.H. (1992) *Model assisted survey sampling,* Springer-Verlog, New York.
15. Singh, V. K. and Seth, S. K. (2014) On the use of one-parameter family of synthetic estimators for small areas, *Journal of Statistics Application & Probability. 3, (3),* 355-361.
16. Tikkiwal, G. C. and Ghiya. A (2000) A generalized class of synthetic estimators with application to crop acreage estimation for small domains. *Biometrical Journal. 42, (7),* 865-876.
17. Tikkiwal, G. C., Rai. P. K and Ghiya. A (2013) On the performance of generalized regression estimator for Small domains. *Communications in Statistics – Simulation and Computation. 42, (7),* 865-876.