Stochastic Modeling and Applications
Vol.26 No. 1 (January-June, 2022)
ISSN: 0972-3641

Received: 12th February 2022 *Revised: 04th April 2022* *Selected: 29th April 2022*

ESTIMATION OF POPULATION VARIANCE IN THE PRESENCE OF NON-RESPONSE BY IMPUTING REGRESSION -TYPE ESTIMATORS FOR THE VARIANCE IN THE NON-RESPONSE STRATUM

RAJ KUMARI BAHL AND M.C. AGRAWAL

ABSTRACT: Non-response to surveys is emerging as a very mammoth problem. We are in an era where extracting information from respondents requires a lot of effort. We, in this paper, propose two estimators of the population variance in the presence of non-response by imputing regression-type estimators for the variance in the non-response stratum. We also compare the twin estimators with an existing unbiased estimator of the population variance as well as two estimators of population variance that make use of ratio-type estimators for the population variance in the non-response stratum proposed by Bahl and Agrawal (2021). For this purpose, we derive the variance of each of the three estimators. We also obtain the conditions under which the proposed estimators perform better than the existing estimators. Illustrative examples have been furnished to probe and assess the performance of the proposed estimators.

Keywords: Non-response, Response and non-response strata, Estimators of population variance in presence of non-response, Regression-type estimators, Variance of estimators of population variance in the presence of non-response.

1. Introduction

Suppose that we have a finite population of size N from which a random sample s of size n is drawn without replacement. Let y_i (i = 1, 2, ..., N) denote the value of the characteristic of interest y on the unit i. Drawing the sample dichotomizes the population into sampled and non-sampled parts to be denoted by s and $\overline{s}$ respectively with respective sizes n and N−n.

Non-response has long been identified to be a major problem in survey sampling and when the sample is drawn, n_1 units respond while remaining n_2 (= n −n_1) units do not furnish any response. When non-response is present in the initial attempt, Hansen and Hurwitz (1946) propose the following double sampling scheme for estimating the population mean:

(a) select a simple random sample of size n and then mail the questionnaire to these sample units.

(b) out of the n_2 non-responding units, select a sub-sample of size $m = \frac{n_2}{k}$ (k>1) using simple random sampling and contact them through personal interviews.

Hansen and Hurwitz (1946) build their theory based on the assumption that the population of size N consists of two strata of 'respondents' and 'non-respondents' having respective sizes N_1 and N_2 (= $N - N_1$).

Let

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

and

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N} (y_i - \bar{Y})^2$$

be the population mean and the population variance of the survey variable y.

Further, let

$$\mu_{0r} = \frac{1}{N-1}\sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^r, \qquad r = 1, 2, 3, \ldots$$

denote the rth order population moment of the survey variable y.

Next, let $(\bar{Y}_1, S_{y_1}^2)$ and $(\bar{Y}_2, S_{y_2}^2)$ denote the coupling of the respective population means and variances with regard to the strata of respondents and non-respondents. The population mean $\bar{Y}$ may be written as

$$\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$$

where W_1 and W_2 are the population proportions of the response and non-response strata respectively, i.e.

$$W_1 = \frac{N_1}{N} \quad \text{and} \quad W_2 = \frac{N_2}{N}.$$

Further, let

$$\mu_{0r}^{(1)} = \frac{1}{N_1 - 1}\sum_{i=1}^{N_1} \left(y_i - \bar{Y}_1\right)^r, \quad r = 1, 2, 3, \ldots$$

and

$$\mu_{0r}^{(2)} = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2} \left(y_i - \bar{Y}_2\right)^r, \quad r = 1, 2, 3, \ldots.$$

be the r^{th} order population moments in respect of the strata of respondents and non-respondents.

Besides, let $(\bar{y}, s_y^2), (\bar{y}_1, s_{y_1}^2), (\bar{y}_2, s_{y_2}^2)$ and $(\bar{y}_{m_2}, s_{y_{m_2}}^2)$ denote the means coupled with variances based on n, n_1, n_2 and m units respectively, where

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$s_{y_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2$$

$$s_{y_2}^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2} (y_i - \bar{y}_2)^2$$

and

$$s_{y_{m_2}}^2 = \frac{1}{m-1}\sum_{i=1}^{m} (y_i - \bar{y}_{m_2})^2.$$

Next, we define

$$\beta_2(y)=\frac{\mu_{04}}{\mu_{02}^2},\quad j=1,2$$

as the population coefficient of kurtosis of the characteristic of interest, i.e., y.

Finally, let

$$\beta_2(y_j)=\frac{\mu_{04}^{(j)}}{\mu_{02}^{2(j)}},\ j=1,2$$

where j = 1 and 2 stand respectively for the response and non-response strata.

Hansen and Hurwitz (1946) proposed the following unbiased estimator of the population mean $\bar{Y}$

$$\bar{y}^*=w_1\bar{y}_1+w_2\bar{y}_{m_2} \tag{1.1}$$

where w_1 and w_2 are the sample proportions of the responding and non-responding units and are given by

$$w_1=\frac{n_1}{n}\quad\text{and}\quad w_2=\frac{n_2}{n}.$$

In the presence of non-response, Cochran (1977, p.374), Rao (1986), and Särndal et al. (1992, p.583) suggest the use of auxiliary variable x with a view to enhance the possibilities of achieving improved results. Let x_i (i = 1, 2, ..., N) be the measurement of the i^{th} unit on x. Corresponding to the above population-based quantities $\bar{Y},\bar{Y}_1,\bar{Y}_2,S_y^2,S_{y_1}^2,S_{y_2}^2,\ \mu_{0r},\mu_{0r}^{(1)},\mu_{0r}^{(2)},\beta_2(y),\beta_2(y_1)$ and $\beta_2(y_2)$ for the characteristic of interest, let $\bar{X},\bar{X}_1,\bar{X}_2,S_x^2,S_{x_1}^2,S_{x_2}^2,\ \mu_{r0},\mu_{r0}^{(1)},\mu_{r0}^{(2)}$, $\beta_2(x_1)$ and $\beta_2(x_2)$ be their counterparts in respect of the x-variable and, similarly, corresponding to the sample-based quantities $\bar{y},\bar{y}_1,\bar{y}_2,s_y^2,s_{y_1}^2,s_{y_2}^2$ and $s_{y_{m_2}}^2$, let $\bar{x},\bar{x}_1,\bar{x}_2,s_x^2,s_{x_1}^2,s_{x_2}^2$ and $s_{x_{m_2}}^2$ represent their counterparts with respect to the x-variable.

Apart from this, let

$$\mu_{rs}=\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\bar{X}\right)^r\left(y_i-\bar{Y}\right)^s\qquad r,s=1,2,3,\ldots$$

denote the joint $(r,s)^{th}$ order population moments for the characteristic of interest and auxiliary characteristic.

Further, let

$$\mu_{rs}^{(j)}=\frac{1}{N_j-1}\sum_{i=1}^{N_j}(x_i-\bar{X}_j)^r(y_i-\bar{Y}_j)^s,\quad j=1,2;\ r,s=1,2,3,\ldots$$

denote the joint $(r,s)^{th}$ order population moments for the characteristic of interest and auxiliary characteristic with regards to the strata of respondents and non-respondents.

Finally, let

$$\theta_j = \frac{\mu_{22}^{(j)}}{\mu_{20}^{(j)}\mu_{02}^{(j)}}, \quad j = 1, 2;$$

Although the problem of estimation of population mean in the presence of non response has engaged considerable attention, the estimation of population variance in the presence of non-response remains little explored and, hence, we take up the same to focus on finding suitable estimator of population variance in the presence of non-response in the ensuing sections.

2. Existing Estimators of Population Variance in the Presence of Non-Response

2.1 An unbiased estimator

Rao (1990) and Agrawal and Sthapit (2002) have made use of an unbiased estimator of population variance in regards to the estimation of variance of estimator of population mean in the presence of non-response. Bahl and Agrawal (2021) have derived the same estimator by an alternative approach consisting of the dichotomy of the population into response and non-response stratum. The said authors address the estimator by $s_y'^2$ and present it as

$$s_y'^2 = \frac{N}{N-1}\left[\left\{w_1 - \frac{1}{N} - \frac{(N-n)}{N(n-1)}w_2\right\}s_{y_1}^2 + \left\{\frac{N-1}{N}w_2 - \frac{k(N-1)}{N(n-1)}w_2\right\}s_{y_{m_2}}^2 + \frac{n(N-1)}{N(n-1)}w_1 w_2(\bar{y}_1 - \bar{y}_{m_2})^2\right]$$

(2.1.1)

with an alternative form given as

$$s_y'^2 = p_1 s_{y_1}^2 + p_2 s_{y_2}^2 + p_3(\bar{y}_1 - \bar{y}_{m_2})^2$$
(2.1.2)

where

$$p_1 = \frac{nw_1 - 1}{n-1}, \qquad p_2 = w_2 - \frac{kw_1}{n-1}, \qquad p_3 = \frac{n}{n-1}w_1 w_2$$
(2.1.3)

The same authors have also obtained the variance of this estimator to terms of $O(1/n)$, given as

$$V(s_y'^2) \doteq \frac{\lambda}{n}W_1W_2\left\{\left(S_{y_1}^2 - S_{y_2}^2\right) - (W_1 - W_2)\left(\bar{Y}_1 - \bar{Y}_2\right)^2\right\}^2 + \frac{\lambda}{n}W_1 S_{y_1}^4\{\beta_2(y_1) - 1\}$$
$$+\left(k - \frac{n}{N}\right)W_2\frac{S_{y_2}^4}{n}\{\beta_2(y_2) - 1\} + \frac{4}{n}(\bar{Y}_1 - \bar{Y}_2)^2 W_1 W_2\left[\lambda W_2 S_{y_1}^2 + \left(k - \frac{n}{N}\right)W_1 S_{y_2}^2\right]$$
$$+\frac{4W_1W_2}{n}(\bar{Y}_1 - \bar{Y}_2)\left[\lambda\mu_{03}^{(1)} - \left(k - \frac{n}{N}\right)\mu_{03}^{(2)}\right] \qquad (2.1.4)$$

where

$$\lambda = \frac{N-n}{N-1}$$

2.2. Two estimators of population variance imputing ratio estimators in non-response stratum

The estimator ${s'_y}^2$ discussed above does not make use of the available auxiliary information. An efficient use of auxiliary information appears in Bahl and Agrawal (2021) who have propounded the following two estimators of population variance in presence of non-response by employing ratio-type estimators for the variance in the non-response stratum.

$$S_{yr1}^2 = \frac{N}{N-1}\left[\left\{w_1 - \frac{1}{N} - \frac{(N-n)}{N(n-1)}w_2\right\}s_{y_1}^2 + \left\{\frac{N-1}{N}w_2 - \frac{k(N-1)}{N(n-1)}w_2\right\}\frac{s_{y_{m_2}}^2}{s_{x_{m_2}}^2}s_{x_2}^2 + \frac{n(N-1)}{N(n-1)}w_1 w_2(\bar{y}_1 - \bar{y}_{m_2})^2\right]$$

(2.2.1)

and

$$S_{yr2}^2 = \frac{N}{N-1}\left[\left\{w_1 - \frac{1}{N} - \frac{(N-n)}{N(n-1)}w_2\right\}s_{y_1}^2 + \left\{\frac{N-1}{N}w_2 - \frac{k(N-1)}{N(n-1)}w_2\right\}\frac{s_{y_{m_2}}^2}{s_{x_{m_2}}^2}S_{x_2}^2 + \frac{n(N-1)}{N(n-1)}w_1 w_2(\bar{y}_1 - \bar{y}_{m_2})^2\right]$$

(2.2.2)

which have been alternatively expressed as

$$S_{yr1}^2 = p_1 s_{y_1}^2 + p_2\frac{s_{y_{m_2}}^2}{s_{x_{m_2}}^2}s_{x_2}^2 + p_3(\bar{y}_1 - \bar{y}_{m_2})^2$$

(2.2.3)

and

$$S_{yr2}^2 = p_1 s_{y_1}^2 + p_2\frac{s_{y_{m_2}}^2}{s_{x_{m_2}}^2}S_{x_2}^2 + p_3(\bar{y}_1 - \bar{y}_{m_2})^2$$

(2.3.4)

where p_i (i= 1,2,3) are defined in (2.1.3).

The authors have also worked out the variances of the above estimators to terms of $O(1/n)$ and these are

$$V(S_{yr1}^2) \doteq \frac{\lambda}{n}W_1W_2\left\{\left(S_{y_1}^2 - S_{y_2}^2\right) - (W_1 - W_2)\left(\bar{Y}_1 - \bar{Y}_2\right)^2\right\}^2 + \frac{\lambda}{n}W_1S_{y_1}^4\{\beta_2(y_1)-1\} + \left(k - \frac{n}{N}\right)W_2\frac{S_{y_2}^4}{n}\{\beta_2(y_2)-1\}$$
$$+\frac{4}{n}(\bar{Y}_1 - \bar{Y}_2)^2W_1W_2\left[\lambda W_2S_{y_1}^2 + \left(k - \frac{n}{N}\right)W_1S_{y_2}^2\right] + \frac{4W_1W_2}{n}(\bar{Y}_1 - \bar{Y}_2)\left[\lambda\mu_{03}^{(1)} - \left(k - \frac{n}{N}\right)\mu_{03}^{(2)}\right]$$
$$+W_2S_{y_2}^4\left(\frac{k-1}{n}\right)\left(\beta_2(x_2) - 2\theta_2 + 1\right) + 4W_1W_2(\bar{Y}_1 - \bar{Y}_2)\left(\frac{k-1}{n}\right)\frac{S_{y_2}^2}{S_{x_2}^2}\mu_{21}^{(2)} \qquad (2.3.5)$$

and

$$V(S_{yr2}^2) \doteq \frac{\lambda}{n} W_1 W_2 \left\{ \left(S_{y_1}^2 - S_{y_2}^2\right) - (W_1 - W_2)\left(\bar{Y}_1 - \bar{Y}_2\right)^2 \right\}^2 + \frac{\lambda}{n} W_1 S_{y_1}^4 \{\beta_2(y_1) - 1\}$$

$$+ \left(k - \frac{n}{N}\right) W_2 \frac{S_{y_2}^4}{n} \{\beta_2(y_2) + \beta_2(x_2) - 2\theta_2\} + \frac{4}{n}(\bar{Y}_1 - \bar{Y}_2)^2 W_1 W_2 \left[\lambda W_2 S_{y_1}^2 + \left(k - \frac{n}{N}\right) W_1 S_{y_2}^2\right]$$

$$+ \frac{4 W_1 W_2}{n} (\bar{Y}_1 - \bar{Y}_2) \left\{ \lambda \mu_{03}^{(1)} - \left(k - \frac{n}{N}\right) \mu_{03}^{(2)} + \left(k - \frac{n}{N}\right) \frac{S_{y_2}^2}{S_{x_2}^2} \mu_{21}^{(2)} \right\} \quad (2.3.6)$$

We now propose two estimators similar to S_{yr1}^2 and S_{yr2}^2 considered above by imputing regression-type estimators in the non-response stratum.

3. Two estimators of population variance imputing regression estimators in non-response stratum

Continuing with the framework of dichotomization of the population into response and non-response strata attributed to Bahl and Agrawal (2021), we impute two different regression estimators

$$s_{y_{m_2}}^2 + b_{m_2}^2 \left(s_{x_2}^2 - s_{x_{m_2}}^2\right)$$

(3.1)

and

$$s_{y_{m_2}}^2 + b_{m_2}^2 \left(S_{x_2}^2 - s_{x_{m_2}}^2\right)$$

(3.2)

where

$$b_{m_2} = \frac{s_{xy_{m_2}}}{s_{x_{m_2}}^2}$$

for estimating the variance $S_{y_2}^2$ in the non-response stratum. Here, for the use of the estimator in (3.2), it has been assumed that the auxiliary information is available on all the N_2 units constituting the non-response stratum.

Using (3.1) and (3.2) to replace $s_{y_{m_2}}^2$ in (2.5), we propose two new estimators of the population variance in the presence of non-response, viz.

$$S_{yr1}^2 = \frac{N}{N-1}\left[\left\{w_1 - \frac{1}{N} - \frac{(N-n)}{N(n-1)} w_2\right\} s_{y_1}^2 + \left\{\frac{N-1}{N} w_2 - \frac{k(N-1)}{N(n-1)} w_2\right\} \frac{s_{y_{m_2}}^2}{s_{x_{m_2}}^2} s_{x_2}^2 + \frac{n(N-1)}{N(n-1)} w_1 w_2 (\bar{y}_1 - \bar{y}_{m_2})^2\right]$$

(3.3)

and

$$S^2_{yr2} = \frac{N}{N-1}\left[\left\{w_1 - \frac{1}{N} - \frac{(N-n)}{N(n-1)}w_2\right\}s^2_{y_1} + \left\{\frac{N-1}{N}w_2 - \frac{k(N-1)}{N(n-1)}w_2\right\}\frac{s^2_{y_{m_2}}}{s^2_{x_{m_2}}}S^2_{x_2} + \frac{n(N-1)}{N(n-1)}w_1w_2(\bar{y}_1 - \bar{y}_{m_2})^2\right]$$

(3.4)

which can be alternatively written as

$$S^2_{yr1} = p_1s^2_{y_1} + p_2\frac{s^2_{y_{m_2}}}{s^2_{x_{m_2}}}s^2_{x_2} + p_3(\bar{y}_1 - \bar{y}_{m_2})^2$$

(3.5)

and

$$S^2_{yr2} = p_1s^2_{y_1} + p_2\frac{s^2_{y_{m_2}}}{s^2_{x_{m_2}}}S^2_{x_2} + p_3(\bar{y}_1 - \bar{y}_{m_2})^2$$

(3.6)

where p_i (i= 1,2,3) are defined in (2.7). Note that, for k = 1, S^2_{yr1} will reduce to s'^2_y.

Now, we work out the expected value and bias of each of the above two estimators to the first degree of approximation.

5. An Empirical Investigation

To provide numerical evidence in support of the findings of Section 4, we consider a finite population for which we examine different combinations of proportions of respondents and non-respondents in the population i.e., W_1 and W_2 in Tables 1, 2 and 3. The following data set pertaining to this population assumes that both the proposed estimators are unbiased to the first degree of approximation. The population quantities for the response stratum are

$S^2_{y_1}$ = 144, $\mu^{(1)}_{03}$ = 200, $\beta_2(y_1)$ = 2

and the corresponding quantities for the non-response stratum are

$S^2_{y_2}$ = 1000, $\mu^{(2)}_{03}$ = 250, $\beta_2(y_2)$ = 5.5

while $\beta_2(x_2)$ (and consequently θ_2 under the condition (3.11)) is assumed to take one of the three possible values, i.e., $\beta_2(x_2)$ = 1.5 or 3 or 4.5. It is clear from (3.14) and the above data that the maximum value that can be attained by $\beta_2(x_2)$ is 5.5. Further, since the conditions for better performance of any of the proposed estimators depends on the difference say, D of the population means $\bar{Y}_1$ and $\bar{Y}_2$, as borne out by (4.2), we consider for each of the three tables presented below values of D (= $\bar{Y}_1 - \bar{Y}_2$) as -25, -50, 0, 25 and 50.

The three tables have been prepared to highlight the performance of the three estimators s'^2_y, S^2_{yr1} and S^2_{yr2} discussed in Sections 2 and 3 by computing their variances for four chosen values of *f* (= n/N) and k = 2, 3 and 4 corresponding to each of the aforesaid three specified values of $\beta_2(x_2)$. These tables also present the percent gains in precision of the two competing estimators S^2_{yr1} and S^2_{yr2} relative to the existing estimator $s_y'^2$ denoted by G_1 and G_2 and the percent gain of S^2_{yr2} relative to S^2_{yr1} being denoted by G_3. More explicitly, let

$$G_1 = \frac{V\left(s_y'^2\right) - V(S^2_{yr1})}{V(S^2_{yr1})} \times 100$$

$$G_2 = \frac{V\left(s_y'^2\right) - V(S^2_{yr2})}{V(S^2_{yr2})} \times 100$$

and

$$G_3 = \frac{V(S^2_{yr1}) - V(S^2_{yr2})}{V(S^2_{yr2})} \times 100.$$

Table 1: Percent gains of S^2_{yr1} and S^2_{yr2} relative to $s_y'^2$ and of S^2_{yr2} relative to S^2_{yr1} for different combinations of f, $\beta_2(x_2)$ and k and varying values of D when $W_1 = 0.25$ and $W_2 = 0.75$

D = -50										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.05	G_1	5.6	7.7	8.7	24.8	36.1	42.6	52.5	85.1	107.2
	G_2	11.6	11.7	11.8	63.3	64.3	64.9	204.4	210.6	213.7
	G_3	5.7	3.8	2.8	30.9	20.7	15.6	99.6	67.8	51.4
0.10	G_1	5.8	7.8	8.8	25.7	37.0	43.4	54.7	87.9	110.2
	G_2	11.6	11.7	11.8	63.4	64.4	64.9	204.9	211.0	214.1
	G_3	5.5	3.6	2.7	30.0	20.0	15.0	97.1	65.5	49.4
0.25	G_1	6.3	8.3	9.2	28.6	40.0	46.1	62.6	97.8	120.4
	G_2	11.7	11.8	11.8	63.7	64.7	65.1	206.6	212.5	215.3
	G_3	5.0	3.2	2.4	27.3	17.6	13.0	88.5	57.9	43.1
0.50	G_1	7.5	9.2	10.0	35.3	46.1	51.3	82.4	120.4	142.3
	G_2	11.7	11.8	11.9	64.3	65.1	65.5	210.1	215.3	217.6

	G_3	3.9	2.4	1.7	21.4	13.0	9.4	70.0	43.1	31.1
D = -25										
0.05	G_1	6.0	8.2	9.3	28.1	41.2	48.7	61.8	103.1	132.6
	G_2	12.5	12.5	12.6	74.8	75.5	75.8	292.0	297.8	300.8
	G_3	6.1	4.0	3.0	36.4	24.3	18.2	142.2	95.9	72.3
D = -25										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.10	G_1	6.2	8.3	9.4	29.1	42.2	49.6	64.5	106.8	136.7
	G_2	12.5	12.6	12.6	74.9	75.5	75.8	292.4	298.2	301.1
	G_3	5.9	3.9	2.9	35.5	23.4	17.5	138.5	92.6	69.5
0.25	G_1	6.8	8.8	9.8	32.4	45.6	52.7	74.3	119.9	150.7
	G_2	12.5	12.6	12.6	75.0	75.7	76.0	294.0	299.6	302.3
	G_3	5.4	3.4	2.5	32.2	20.6	15.2	126.0	81.7	60.5
0.50	G_1	8.0	9.8	10.6	40.2	52.7	58.9	99.6	150.7	181.8
	G_2	12.5	12.6	12.6	75.4	76.0	76.2	297.4	302.3	304.4
	G_3	4.2	2.5	1.8	25.1	15.2	10.9	99.1	60.5	43.5
D = 0										
0.05	G_1	5.9	8.0	9.1	28.8	42.3	50.2	64.2	108.4	140.8
	G_2	12.2	12.3	12.4	77.1	78.1	78.6	320.3	329.8	334.7
	G_3	6.0	4.0	3.0	37.6	25.1	18.9	156.1	106.2	80.5
0.10	G_1	6.1	8.2	9.2	29.8	43.4	51.2	67.0	112.5	145.4
	G_2	12.2	12.3	12.4	77.2	78.2	78.6	321.1	330.5	335.3
	G_3	5.8	3.8	2.9	36.6	24.3	18.1	152.1	102.6	77.4
0.25	G_1	6.7	8.7	9.7	33.2	47.0	54.5	77.5	126.9	161.1
	G_2	12.3	12.3	12.4	77.5	78.4	78.8	323.6	332.7	337.2
	G_3	5.3	3.4	2.5	33.2	21.4	15.8	138.7	90.7	67.4

0.50	G_1	7.9	9.7	10.5	41.3	54.5	60.9	104.6	161.1	196.4
	G_2	12.3	12.4	12.4	78.0	78.8	79.1	329.1	337.2	340.7
	G_3	4.1	2.5	1.8	26.0	15.8	11.3	109.7	67.4	48.7
D = 25										
0.05	G_1	5.4	7.4	8.3	27.3	39.8	47.1	60.6	100.5	128.9
	G_2	11.2	11.2	11.3	71.8	72.5	72.8	278.1	283.8	286.7
	G_3	5.4	3.6	2.7	35.0	23.3	17.5	135.5	91.4	69.0
D = 25										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.10	G_1	5.6	7.5	8.5	28.2	40.8	48.0	63.2	104.2	132.8
	G_2	11.2	11.2	11.3	71.9	72.5	72.9	278.5	284.2	287.1
	G_3	5.3	3.5	2.6	34.0	22.5	16.8	131.9	88.2	66.3
0.25	G_1	6.1	7.9	8.8	31.5	44.1	51.0	72.7	116.8	146.3
	G_2	11.2	11.3	11.3	72.1	72.7	73.0	280.1	285.6	288.2
	G_3	4.8	3.1	2.3	30.9	19.8	14.6	120.0	77.9	57.6
0.50	G_1	7.2	8.8	9.5	38.9	51.0	56.8	97.2	146.3	175.9
	G_2	11.2	11.3	11.3	72.4	73.0	73.2	283.4	288.2	290.3
	G_3	3.7	2.3	1.6	24.1	14.6	10.5	94.5	57.6	41.5
D = 50										
0.05	G_1	4.6	6.2	7.0	23.4	33.9	39.9	50.5	81.3	101.9
	G_2	9.3	9.4	9.5	58.7	59.7	60.2	189.4	195.2	198.2
	G_3	4.5	3.0	2.3	28.6	19.2	14.5	92.3	62.9	47.7
0.10	G_1	4.7	6.3	7.1	24.2	34.8	40.7	52.6	83.9	104.7
	G_2	9.3	9.4	9.5	58.8	59.8	60.2	189.9	195.6	198.5
	G_3	4.4	2.9	2.2	27.8	18.5	13.9	90.0	60.7	45.8
0.25	G_1	5.1	6.7	7.5	26.9	37.5	43.1	60.1	93.2	114.2
	G_2	9.4	9.5	9.5	59.1	60.0	60.	191.	197.	199.

							4	4	0	7
	G_3	4.0	2.6	1.9	25.3	16.4	12.1	82.0	53.7	39.9
0.50	G_1	6.1	7.5	8.1	33.2	43.1	47.9	78.7	114.2	134.3
	G_2	9.4	9.5	9.6	59.6	60.4	60.8	194.8	199.7	201.9
	G_3	3.1	1.9	1.4	19.9	12.1	8.7	64.9	39.9	28.8

Table 2: Percent gains of S_{yr1}^2 and S_{yr2}^2 relative to $s_y'^2$ and of S_{yr2}^2 relative to S_{yr1}^2 for different combinations of f, $\beta_2(x_2)$ and k and varying values of D when $W_1 = W_2 = 0.50$

D = -50										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.05	G_1	4.4	5.9	6.8	17.1	24.4	28.5	33.4	50.7	61.2
	G_2	8.9	9.0	9.1	39.8	40.7	41.2	95.3	98.4	100.0
	G_3	4.3	2.9	2.2	19.4	13.1	9.9	46.4	31.7	24.1
0.10	G_1	4.5	6.1	6.9	17.6	25.0	29.0	34.6	52.1	62.6
	G_2	8.9	9.0	9.1	39.9	40.8	41.2	95.5	98.6	100.2
	G_3	4.2	2.8	2.1	18.9	12.7	9.5	45.2	30.6	23.1
0.25	G_1	4.9	6.4	7.2	19.6	26.8	30.6	39.0	56.9	67.1
	G_2	8.9	9.1	9.1	40.1	41.0	41.4	96.3	99.4	100.8
	G_3	3.8	2.5	1.8	17.2	11.2	8.3	41.3	27.1	20.2
0.50	G_1	5.8	7.2	7.8	23.9	30.6	33.8	49.3	67.1	76.3
	G_2	9.0	9.1	9.2	40.6	41.4	41.8	98.2	100.8	102.0
	G_3	3.0	1.8	1.3	13.5	8.3	6.0	32.7	20.2	14.6
D = -25										
0.05	G_1	5.5	7.6	8.6	24.4	35.7	42.2	51.4	83.5	105.6
	G_2	11.4	11.6	11.7	61.8	63.3	64.1	195.6	204.3	208.8
	G_3	5.6	3.7	2.8	30.1	20.4	15.4	95.3	65.8	50.2
0.10	G_1	5.7	7.7	8.8	25.2	36.5	43.0	53.5	86.4	108.6
	G_2	11.4	11.6	11.	62.0	63.4	64.	196.	204.	209.

				7			2	3	9	4
	G_3	5.4	3.6	2.7	29.3	19.7	14.8	93.0	63.6	48.3
0.25	G_1	6.3	8.2	9.2	28.1	39.5	45.7	61.3	96.2	118.8
	G_2	11.5	11.7	11.7	62.4	63.8	64.4	198.6	207.0	211.2
	G_3	4.9	3.2	2.3	26.7	17.4	12.9	85.1	56.5	42.2
D = -25										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.50	G_1	7.4	9.2	10.0	34.8	45.7	51.0	80.9	118.8	140.8
	G_2	11.6	11.7	11.8	63.2	64.4	65.0	203.6	211.2	214.6
	G_3	3.9	2.3	1.7	21.1	12.9	9.3	67.9	42.2	30.7
D = 0										
0.05	G_1	5.8	7.9	9.0	28.0	41.5	49.4	62.0	105.4	137.4
	G_2	11.9	12.1	12.2	74.4	76.2	77.1	294.5	311.2	320.2
	G_3	5.8	3.9	2.9	36.2	24.5	18.5	143.5	100.2	77.0
0.10	G_1	6.0	8.1	9.1	29.0	42.6	50.4	64.8	109.4	141.9
	G_2	12.0	12.1	12.2	74.5	76.3	77.2	295.8	312.5	321.3
	G_3	5.7	3.8	2.8	35.3	23.7	17.8	140.1	97.0	74.1
0.25	G_1	6.5	8.6	9.6	32.4	46.2	53.7	75.0	123.6	157.5
	G_2	12.0	12.2	12.3	75.0	76.8	77.6	300.2	316.6	324.8
	G_3	5.1	3.3	2.5	32.2	20.9	15.5	128.7	86.3	65.0
0.50	G_1	7.8	9.6	10.4	40.5	53.7	60.3	101.6	157.5	192.9
	G_2	12.1	12.3	12.3	76.1	77.6	78.3	310.0	324.8	331.7
	G_3	4.0	2.5	1.8	25.4	15.5	11.2	103.3	65.0	47.4
D = 25										
0.05	G_1	4.5	6.1	7.0	23.0	33.5	39.5	49.4	79.8	100.4
	G_2	9.2	9.3	9.4	57.4	58.7	59.5	181.6	189.6	193.9
	G_3	4.5	3.0	2.3	27.9	18.9	14.3	88.5	61.1	46.6
0.10	G_1	4.6	6.3	7.1	23.8	34.3	40.	51.5	82.5	103.

							3			2
	G2	9.2	9.3	9.4	57.5	58.8	59.5	182.2	190.2	194.4
	G3	4.4	2.9	2.2	27.2	18.3	13.7	86.3	59.0	44.9
0.25	G1	5.1	6.7	7.4	26.5	37.1	42.7	58.9	91.7	112.7
	G2	9.2	9.4	9.5	57.8	59.2	59.8	184.4	192.2	196.0
	G3	4.0	2.6	1.9	24.8	16.1	12.0	79.0	52.4	39.2
D = 25										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.50	G1	6.0	7.4	8.0	32.7	42.7	47.6	77.3	112.7	132.9
	G2	9.3	9.5	9.5	58.6	59.8	60.3	189.1	196.0	199.2
	G3	3.1	1.9	1.4	19.5	12.0	8.6	63.0	39.2	28.5
D = 50										
0.05	G1	2.9	3.9	4.4	15.3	21.7	25.2	31.1	46.9	56.4
	G2	5.8	5.9	5.9	34.9	35.7	36.1	86.1	89.0	90.4
	G3	2.8	1.9	1.4	17.0	11.5	8.7	41.9	28.6	21.8
0.10	G1	3.0	4.0	4.5	15.8	22.2	25.7	32.2	48.2	57.7
	G2	5.8	5.9	5.9	34.9	35.8	36.2	86.3	89.2	90.6
	G3	2.7	1.8	1.4	16.5	11.1	8.3	40.9	27.7	20.9
0.25	G1	3.2	4.2	4.7	17.5	23.8	27.1	36.2	52.5	61.7
	G2	5.8	5.9	5.9	35.2	36.0	36.3	87.1	89.8	91.2
	G3	2.5	1.6	1.2	15.1	9.8	7.3	37.3	24.5	18.2
0.50	G1	3.8	4.7	5.1	21.2	27.1	29.8	45.7	61.7	69.9
	G2	5.9	5.9	6.0	35.6	36.3	36.6	88.8	91.2	92.3
	G3	2.0	1.2	0.9	11.9	7.3	5.2	29.6	18.2	13.2

Table 3: Percent gains of S^2_{yr1} and S^2_{yr2} relative to S'^2_y and of S^2_{yr2} relative to S^2_{yr1} for different combinations of f, $\beta_2(x_2)$ and k and varying values of D when $W_1 = 0.75$ and $W_2 = 0.25$

D = -50			
f	Percent	$\beta_2(x_2)$ = 1.5	$\beta_2(x_2)$ = 3.0	$\beta_2(x_2)$ = 4.5

	gain	k			k			k		
		2	3	4	2	3	4	2	3	4
0.05	G_1	2.9	4.0	4.7	10.2	14.8	17.4	18.6	27.9	33.5
	G_2	5.8	6.1	6.3	22.0	23.4	24.2	44.1	47.5	49.4
	G_3	2.8	2.0	1.5	10.7	7.5	5.8	21.5	15.3	11.9
D = -50										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.10	G_1	3.0	4.1	4.8	10.5	15.1	17.7	19.3	28.7	34.3
	G_2	5.8	6.1	6.3	22.1	23.5	24.3	44.3	47.8	49.7
	G_3	2.7	1.9	1.4	10.5	7.3	5.6	21.0	14.8	11.5
0.25	G_1	3.3	4.4	5.0	11.7	16.3	18.8	21.7	31.2	36.6
	G_2	5.9	6.2	6.3	22.5	23.9	24.6	45.2	48.7	50.4
	G_3	2.5	1.7	1.3	9.6	6.5	4.9	19.4	13.3	10.1
0.50	G_1	4.0	5.0	5.5	14.4	18.8	20.9	27.2	36.6	41.4
	G_2	6.1	6.3	6.5	23.3	24.6	25.2	47.3	50.4	51.9
	G_3	2.0	1.3	0.9	7.8	4.9	3.6	15.8	10.1	7.4
D = -25										
0.05	G_1	4.8	6.6	7.6	19.6	28.8	34.1	39.5	62.7	77.9
	G_2	9.7	10.1	10.2	47.1	49.2	50.3	123.1	131.6	136.2
	G_3	4.7	3.2	2.5	22.9	15.8	12.1	60.0	42.4	32.8
0.10	G_1	4.9	6.7	7.7	20.3	29.5	34.8	41.1	64.7	80.0
	G_2	9.8	10.1	10.2	47.3	49.4	50.5	123.7	132.2	136.8
	G_3	4.6	3.1	2.4	22.4	15.3	11.6	58.6	41.0	31.6
0.25	G_1	5.4	7.2	8.1	22.7	31.9	36.9	46.7	71.5	86.9
	G_2	9.8	10.2	10.3	47.8	49.9	50.9	126.0	134.4	138.7
	G_3	4.2	2.8	2.1	20.5	13.6	10.2	54.0	36.6	27.7
0.50	G_1	6.5	8.1	8.8	28.1	36.9	41.3	60.8	86.9	101.4
	G_2	10.0	10.3	10.4	49.1	50.9	51.7	131.0	138.7	142.3
	G_3	3.3	2.1	1.5	16.4	10.2	7.4	43.7	27.7	20.3

D = 0										
0.05	G_1	5.6	7.8	8.9	27.2	40.6	48.6	59.8	102.1	133.7
	G_2	11.6	11.9	12.1	71.5	74.2	75.6	269.6	292.4	305.1
	G_3	5.7	3.8	2.9	34.8	23.9	18.2	131.4	94.2	73.4
D = 0										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.10	G_1	5.8	7.9	9.0	28.2	41.7	49.6	62.5	106.1	138.1
	G_2	11.7	11.9	12.1	71.7	74.4	75.7	271.4	294.2	306.7
	G_3	5.5	3.7	2.8	34.0	23.1	17.5	128.5	91.3	70.8
0.25	G_1	6.4	8.5	9.5	31.6	45.3	52.9	72.4	120.0	153.6
	G_2	11.7	12.0	12.1	72.4	75.0	76.3	277.3	300.0	311.8
	G_3	5.0	3.3	2.4	31.0	20.5	15.3	118.8	81.8	62.4
0.50	G_1	7.6	9.5	10.3	39.6	52.9	59.7	98.4	153.6	189.0
	G_2	11.9	12.1	12.2	74.0	76.3	77.3	290.7	311.8	321.9
	G_3	4.0	2.4	1.7	24.7	15.3	11.0	96.9	62.4	46.0
D = 25										
0.05	G_1	3.5	4.8	5.5	18.0	26.3	31.1	37.4	58.9	72.9
	G_2	7.0	7.3	7.4	42.5	44.4	45.4	113.0	120.6	124.8
	G_3	3.4	2.3	1.8	20.7	14.3	10.9	55.0	38.9	30.0
0.10	G_1	3.6	4.9	5.6	18.7	27.0	31.7	38.9	60.8	74.8
	G_2	7.1	7.3	7.4	42.6	44.5	45.5	113.6	121.2	125.4
	G_3	3.3	2.3	1.7	20.2	13.8	10.5	53.8	37.6	28.9
0.25	G_1	4.0	5.2	5.9	20.8	29.1	33.6	44.2	67.0	81.0
	G_2	7.1	7.4	7.5	43.1	45.0	45.9	115.6	123.1	127.0
	G_3	3.1	2.0	1.5	18.5	12.3	9.2	49.5	33.6	25.4
0.50	G_1	4.7	5.9	6.4	25.7	33.6	37.5	57.2	81.0	94.1

	G_2	7.3	7.5	7.5	44.2	45.9	46.6	120.1	127.0	130.3
	G_3	2.4	1.5	1.1	14.7	9.2	6.7	40.0	25.4	18.6
D = 50										
0.05	G_1	1.5	2.1	2.5	8.7	12.5	14.7	16.9	25.2	30.2
	G_2	3.0	3.2	3.3	18.4	19.6	20.3	39.3	42.3	43.9
	G_3	1.5	1.0	0.8	9.0	6.3	4.9	19.1	13.6	10.6
D = 50										
f	Per-cent gain	$\beta_2(x_2)$ = 1.5			$\beta_2(x_2)$ = 3.0			$\beta_2(x_2)$ = 4.5		
		k			k			k		
		2	3	4	2	3	4	2	3	4
0.10	G_1	1.6	2.2	2.5	9.0	12.8	15.0	17.5	25.9	30.8
	G_2	3.0	3.2	3.3	18.5	19.7	20.4	39.5	42.5	44.1
	G_3	1.4	1.0	0.8	8.8	6.1	4.7	18.7	13.2	10.2
0.25	G_1	1.7	2.3	2.7	10.0	13.8	15.8	19.6	28.1	32.9
	G_2	3.1	3.3	3.3	18.8	20.0	20.6	40.3	43.3	44.8
	G_3	1.3	0.9	0.7	8.1	5.5	4.1	17.3	11.8	9.0
0.50	G_1	2.1	2.7	2.9	12.2	15.8	17.6	24.6	32.9	37.1
	G_2	3.2	3.3	3.4	19.5	20.6	21.1	42.0	44.8	46.1
	G_3	1.1	0.7	0.5	6.5	4.1	3.0	14.0	9.0	6.6

It emerges from the above tables that the gain in precision of s_{yr1}^2 over $s_y'^2$ increases with k regardless of the values of other parameters while for the estimator s_{yr2}^2, the gain in precision over $s_y'^2$ is at best marginal with increasing k. However, the extent of gain scored by s_{yr2}^2 relative to s_{yr1}^2 over $s_y'^2$ is higher and, at times, for some configurations of parameters, it is markedly higher, thus rendering the estimator s_{yr2}^2 a frontrunner, despite the fact that the gain of s_{yr2}^2 over s_{yr1}^2 diminishes with increasing value of k. The tables also reflect that, for a fixed k, the gains in precision of s_{yr1}^2 and s_{yr2}^2 over $s_y'^2$ increases with increasing value of the sampling fraction f while the gain of s_{yr2}^2 over s_{yr1}^2 shows a dip for this case. It is clearly brought out by the above tables that the increase in measure of peakedness (i.e., $\beta_2(x_2)$) of the distribution of x-variable in the non-response stratum impacts in a positive manner the gains in precision of s_{yr1}^2 and s_{yr2}^2 over $s_y'^2$. Our computation based on the difference (distance) between stratum means, i.e., $D = \bar{Y}_1 - \bar{Y}_2$ = -50, -25, 0, 25, 50 leads us to conclude that the estimators S_{yr1}^2 and S_{yr2}^2 score optimally over $s_y'^2$ when D = 0 and so does S_{yr2}^2 over S_{yr1}^2 when D = 0. However, a transition from negative to positive values of $D\,(\neq 0)$, results in lower gains in precision be it G_1 or G_2 or G_3. This is clearly reflected by condition (4.2) coupled with (4.7). Finally Tables 1, 2 and 3 highlight an intuitive fact that an increase in the value of W_1 (which is proportion of respondents in the population) leads to a dip in gain in precision, be it G_1 or G_2 or G_3.

References

Agrawal, M.C. and Sthapit, A.B. (2002): 'Efficient Non-response-adapted and Imputation-based Ratio-type Estimators', *Jour. of Combinatorics, Information and System Sciences*, **27,** 253-270.

Bahl, R.K. and Agrawal, M.C. (2021): 'Estimation of Population Variance in the Presence of Non-Response by Imputing Regression -Type Estimators for the Variance in the Non-Response Stratum', Stochastic Modeling and Applications, July-December 2021, 25(2)

Cochran, W.G. (1977): *Sampling Techniques*, 3rd ed., John Wiley and Sons, New York.

Hansen, M.H. and Hurwitz, W.N. (1946): 'The Problem of Non-response in Sample Surveys', *Jour. Amer. Statist.Assoc,* **41,** 517-529.

Isaki, C.T. (1983), Variance estimation using auxiliary information, J. Amer. Statist. Assoc., **76**, 117-123

Rao, P.S.R.S. (1990): 'Regression Estimators with Sub-Sampling of Non-respondents', *Data Quality Control: Theory and Pragmatics,* Ed. By Guanna E. Leipens and V.R.R. Uppulusis, Marcel Dekker, Inc., New York, 191-208.

Särndal, C.E., Swensson, B. and Wretman, J. (1992): *Model Assisted Survey Sampling*, Springer Verlag, New York.

About the authors

Raj Kumari Bahl is an Associate Professor at the Department of Statistics, Ramjas College, University of Delhi, India. She did her Ph.D. from University of Edinburgh, U.K. Her areas of interest are Actuarial Statistics and Survey Sampling. Her E-mail address is: rkbstats@yahoo.co.in

M.C.Agrawal has retired as a Professor of Statistics from the Department of Statistics, University of Delhi, India. His area of interest is Survey Sampling. His E-mail address is: mc_agrawal@yahoo.com