Special Issue of Engineering and Scientific International Journal ISSN 2394-7187(Online) ISSN 2394-7179 (Print)
Proceedings of the UGC Sponsored - National Conference on "Recent Statistical Computing Techniques and their Applications"
Department of Statistics and Department of Computer Science, Ramanujan College,University of Delhi, Delhi.
March 11-12, 2016. **Published on January 2017**

Cogitation of Incompleteness in the midst of Imputation in Longitudinal Surveys for Population Mean

Kumari Priyanka[#1], Richa Mittal[#2]
[1]*Department of Mathematics, Shivaji College, University of Delhi, New Delhi, India-110027*
[1]*priyanka.ism@gmail.com*, [2]*sovereignricha@gmail.com*

Abstract— Encountering non-response is quite prone in sample surveys however smart be the design, which sours the analysis and hence the results. An effort has been made to exploit the non-response by using a completely fresh approach of imputation technique to estimate the population mean in two occasion successive sampling, utilizing completely known auxiliary information which is dynamic in nature and pronto over the occasions. The properties of proposed estimator have been analysed. Various cases in which non-response may creep are discussed. Extensive simulation algorithms have been designed and are applied to justify the theoretical results.

Keywords— Imputation; Non-Response; Dynamic Auxiliary Information; Successive Sampling; Exponential Type Estimators; Mean Squared Error; Bias; Optimum Replacement Strategy.

1. Introduction

Sophisticated sample surveys are being designed to prevent the non-response of sample units but it is hard to prevent completely due to the pure stochastic nature of incompleteness. Missing data makes the analysis more miserable when the data has to be collected and analyzed on more than one occasion. The problem of sampling on two successive occasions was initiated by Jessen (1942), and latter this idea was explored by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Feng and Zou (1997), Singh and Singh (2001), Singh and Priyanka (2008), Singh et al. (2013a), Bandyopadhyay and Singh(2014), Priyanka and Mittal (2014, 2015a, 2015b), Priyanka et al. (2015) and many others.

Longitudinal surveys are mainly about observing characteristics on more than one chance (occasion) so that the dynamics of the characteristic could be understood over a period so as to infer about the behaviors and patterns. In this process a variety of literature has been put on using many explanative twists, definitely enriching the field of study and a vast literature is available for dealing with non-response while sampling over successive occasion. One may cite Rubin (1976), Sande (1979), Kalton et al. (1981), Kalton and Kasprzyk (1982), Singh and Singh (1991) by considering complete data set and discarding all those units for which information was not available for at least one time. Also Lee et al. (1994, 1995), Singh and Horn (2002), Ahmed et al. (2006), Singh and Priyanka (2007), Singh (2009) and Singh et al. (2013b) can be seen for various new estimators for estimation of parameters by method of imputation using additional auxiliary information in successive sampling.

Various ideas have been dug into conceiving that auxiliary information utilized remains stable in nature while sampling over successive occasions but when difference (gap) between two occasions is sufficiently large, the nature of auxiliary variable may not sustain to be stable. In such a case the nature of auxiliary characteristic turns to be dynamic over the period of observation. So a completely fresh approach has been made using imputation technique while sampling over two successive occasions to negotiate with the ill effects of non-response. MCAR has been assumed implicitly and a more worthy estimator for population mean while sampling over successive occasion using additional auxiliary information which is changing (dynamic) over the period of observation, by imputing missing data in the presence of non-response. The properties of the proposed estimator have been elaborated theoretically considering that (i) non-response may arise on both occasions, (ii) it may occur only at first occasion or (iii) it may occur only at second occasion while comparing the proposed estimator with estimator having complete response, proposed by Priyanka and Mittal (2016). A Simulation study has also been put through to substantiate the practicability of the proposed estimator.

2. Survey Design and Analysis

2.1 Notations

Let $U = (U_1, U_2, ..., U_N)$ be the N element finite population, which has been sampled over two occasions. The characters under study is denoted by x(y) on the first (second) occasion, respectively. It is assumed that information on a dynamic (varying) auxiliary variable $z_1 (z_2)$, with the known population mean, is available on first (second) occasion. We assume that there is non-response at both the occasions. A simple random sample without replacement s_n of n units has been drawn on the first occasion. Let the number of responding unit out of n sampled units, which are drawn at the first occasion, be denoted by r_1, the set of responding units in s_n by R_1 and that of non-responding by R_1^c. A random sub-sample s_m of m = nλ unit is retained (matched) for its use on the current (second) occasion from the units which responded (r_1) at the first occasion and it is intuitive that these matched units will be completely responding at the current (second) occasion as well. A fresh simple random sample (without replacement), s_u of u = n - m = nμ units, is drawn on the second occasion from the non-sampled units of the population so that the sample size on the second occasion remains the same i.e. n. Let the number of responding units out of u sampled units which are drawn afresh at current occasion, be denoted by r_2, the set of responding unit in s_u by R_2, and that of non-responding units by R_2^c. λ and μ (λ + μ =1) are the fractions of matched and fresh sample, respectively, at the current(second) occasion. For every unit $i \in R_j$ (j =1, 2), the values x_i (y_i) are observed, but for the units $i \in R_j^c$ (j =1, 2) the values x_i (y_i) are missing and instead imputed values are derived. The following notations have been used hereafter:

$\bar{X}, \bar{Y}, \bar{Z}_1, \bar{Z}_2$: Population means of the variables x, y, z_1 and z_2 respectively.

$\bar{y}_u, \bar{z}_u, \bar{y}_{r_2}, \bar{z}_2(r_2), \bar{x}_m, \bar{y}_m, \bar{z}_1(m), \bar{z}_2(m), \bar{x}_{r_1}, \bar{z}_1(r_1)$: Sample mean of respective variate based on the sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz_1}, \rho_{xz_2}, \rho_{yz_1}, \rho_{yz_2}, \rho_{z_1z_2}$: Correlation coefficient between the variables shown in suffices.

$S_x^2, S_y^2, S_{z_1}^2, S_{z_2}^2$: Population mean squared of variables x, y, z_1 and z_2 respectively.

$f_1 = \left(\frac{r_1}{n}\right), f_2 = \left(\frac{r_2}{u}\right)$: The fraction of respondents at first and second occasions respectively.

$t_1 = (1 - f_1), t_2 = (1 - f_2)$: The fraction of non- respondents at first and second occasions respectively.
Sample mean of respective variate based on the sample sizes shown in suffice.

2.2 Formulation of the Proposed Estimator T

To estimate the population mean $\bar{Y}$ on the current (second) occasion, an estimator T_u has been proposed considering that non-response occurs at current occasion and the missing values occurring in the sample of size u are replaced by imputed values. Hence, the following imputation method has been proposed to cope up with the problem of non-response in sample To estimate the population mean $\bar{Y}$ on the current (second) occasion, an estimator T_u has been proposed considering that non-response occurs at current occasion and the missing values occurring in the sample of size u are replaced by imputed values. Hence, the following imputation method has been proposed to cope up with the problem of non-response in sample s_u :

$$y_{.i} = \begin{cases} y_i & \text{if} \quad i \in R_2 \\ \frac{1}{u - r_2}\left\{u\,\bar{y}_{r_2} \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(r_2)}{\bar{Z}_2 + \bar{z}_2(r_2)}\right) - r_2\,\bar{y}_{r_2}\right\} & \text{if} \quad i \in R_2^c \end{cases} \tag{1}$$

where $\bar{y}_{r_2} = \frac{1}{r_2}\sum_{i \in R_2} y_i$ and $\bar{z}_2(r_2) = \frac{1}{r_2}\sum_{i \in R_2} z_{2_i}$.

and hence the estimator for $\bar{Y}$ is given by

$$T_u = \bar{y}_{r_2} \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(r_2)}{\bar{Z}_2 + \bar{z}_2(r_2)}\right) \tag{2}$$

The second estimator T_m is based on sample size m = nλ common to the both occasions utilizing information retained from first occasion. Since non-response is assumed to be occurring on first occasion as well so the missing values occurring in the sample of size n are replaced by imputed values. The following imputation technique has been suggested

$$x_{.i} = \begin{cases} x_i & \text{if} \quad i \in R_1 \\ \frac{1}{n - r_1}\left\{n\,\bar{x}_{r_1} \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(r_1)}{\bar{Z}_1 + \bar{z}_1(r_1)}\right) - r_1\,\bar{x}_{r_1}\right\} & \text{if} \quad i \in R_1^c \end{cases} \tag{3}$$

where $\bar{x}_{r_1} = \frac{1}{r_1}\sum_{i \in R_1} x_i$ and $\bar{z}_1(r_1) = \frac{1}{r_1}\sum_{i \in R_1} z_{1_i}$.

Considering above proposed imputation method the estimator based on sample s_n is altered to

$$\bar{x}_n^* = \frac{1}{n}\sum_{i \in s_n} x_{.i} = \bar{x}_{r_1} \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(r_1)}{\bar{Z}_1 + \bar{z}_1(r_1)}\right) \tag{4}$$

Therefore, the estimator based on sample size m common to both occasions which utilizes the missing values by above method of imputation is given by

$$T_m = \bar{y}_m^*\left(\frac{\bar{x}_n^*}{\bar{x}_m^*}\right) \tag{5}$$

Special Issue of Engineering and Scientific International Journal ISSN 2394-7187(Online) ISSN 2394-7179 (Print)
Proceedings of the UGC Sponsored - National Conference on "Recent Statistical Computing Techniques and their Applications"
Department of Statistics and Department of Computer Science, Ramanujan College,University of Delhi, Delhi.
March 11-12, 2016. **Published on January 2017**

where $\bar{y}_m^{''} = \bar{y}_m \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(m)}{\bar{Z}_2 + \bar{z}_2(m)}\right)$, $\bar{x}_m^{''} = \bar{x}_m \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(m)}{\bar{Z}_1 + \bar{z}_1(m)}\right)$

and $\bar{x}_n^{*} = \bar{x}_{r_1} \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(r_1)}{\bar{Z}_1 + \bar{z}_1(r_1)}\right)$.

Considering the convex combination of the two estimators T_u and T_m, we have the final estimator of population mean $\bar{Y}$ on the current occasion as

$$T = \alpha T_u + (1-\alpha) T_m \tag{6}$$

where α $(0 \le \alpha \le 1)$ is a constant to be determined so as to minimize the mean squared error of the proposed estimators T.

2.3 *Properties of the Proposed Estimators T*

The properties of the proposed estimators T are derived under the following large sample approximations

$\bar{y}_{r_2} = \bar{Y}(1+e_0)$, $\bar{y}_m = \bar{Y}(1+e_1)$, $\bar{x}_m = \bar{X}(1+e_2)$, $\bar{x}_{r_1} = \bar{X}(1+e_3)$,

$\bar{z}_2(r_2) = \bar{Z}_2(1+e_4)$, $\bar{z}_2(m) = \bar{Z}_2(1+e_5)$, $\bar{z}_1(m) = \bar{Z}_1(1+e_6)$

and $\bar{z}_1(r_1) = \bar{Z}_1(1+e_7)$ such that $|e_i| < 1 \ \forall \ i = 0, ..., 7$.

2.4. *Bias and Mean Squared Error of the Estimators T*

The estimators T_u and T_m are exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population mean $\bar{Y}$. Therefore, the final estimator T defined in equation (6) is also biased estimator of $\bar{Y}$. The bias $B(.)$ and mean squared error $M(.)$ of the proposed estimator T are obtained (ignoring finite population corrections) and thus we have following theorems:

*Theorem 2.4.1.*Bias of the estimator T to the first order of approximations is obtained as

$$B(T) = \alpha B(T_u) + (1-\alpha) B(T_m) \tag{7}$$

where $$B(T_u) = \frac{1}{r_2}\bar{Y}\left(\frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\bar{Z}_2}\right) \tag{8}$$

and

$$B(T_m) = \bar{Y}\left(\frac{1}{m}\left(\frac{C_{2000}}{\bar{X}^2} - \frac{1}{8}\frac{C_{0020}}{\bar{Z}_1^2} + \frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{C_{1100}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{1010}}{\bar{X}\bar{Z}_1} + \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2} + \frac{1}{2}\frac{C_{0110}}{\bar{Y}\bar{Z}_1} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\bar{Z}_2} - \frac{1}{4}\frac{C_{0011}}{\bar{Z}_1\bar{Z}_2}\right)\right.$$
$$\left. + \frac{1}{r_1}\left(\frac{1}{8}\frac{C_{0020}}{\bar{Z}_1^2} - \frac{C_{2000}}{\bar{X}^2} + \frac{C_{1100}}{\bar{X}\bar{Y}} + \frac{1}{2}\frac{C_{1010}}{\bar{X}\bar{Z}_1} - \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2} - \frac{1}{2}\frac{C_{0110}}{\bar{Y}\bar{Z}_1} + \frac{1}{4}\frac{C_{0011}}{\bar{Z}_1\bar{Z}_2}\right)\right) \tag{9}$$

where $C_{rstq} = E\left[(x_i - \bar{X})^r (y_i - \bar{Y})^s (z_{1_i} - \bar{Z}_1)^t (z_{2_i} - \bar{Z}_2)^q\right]$;

$(r, s, t, q) \ge 0$.

*Theorem 2.4.2.*Mean squared error of the estimator T to the first order of approximations is obtained as

$$M(T) = \alpha^2 M(T_u) + (1-\alpha)^2 M(T_m) + 2\alpha(1-\alpha)Cov(T_u, T_m) \tag{10}$$

$$M(T_u) = \frac{1}{r_2} A_1 S_y^2 \tag{11}$$

$$M(T_m) = \left(\frac{1}{m}A_2 + \frac{1}{r_1}A_3\right) S_y^2 \tag{12}$$

where

$A_1 = (5/4) - \rho_{yz_2}$, $A_2 = (5/2) - 2\rho_{yx} + \rho_{yz_1} - \rho_{yz_2} - \rho_{xz_1} + \rho_{xz_2} - (1/2)\rho_{z_1z_2}$

$A_3 = 2\rho_{yx} - \rho_{yz_1} + \rho_{xz_1} - \rho_{xz_2} + (1/2)\rho_{z_1z_2} - (5/4)$ and $Cov(T_u, T_m) = 0$.

2.5 *Minimum Mean Squared Error of the Proposed Estimator T*

Since the mean squared error of the estimator T given in equation (10) is a function of unknown constant α, therefore, it has been minimized with respect to α and subsequently the optimum value of α and hence optimum mean squared error of the estimator T are given respectively as

$$\alpha_{opt.} = M(T_m) / (M(T_u) + M(T_m)) \tag{13}$$

$$M(T)_{opt.} = (M(T_u) \cdot M(T_m)) / (M(T_u) + M(T_m)) \tag{14}$$

Further, substituting the value of the mean squared error of the estimators defined in equations (2) and (5) in equation (13) and (14) respectively, the simplified values of $\alpha_{opt.}$ and $M(T)_{opt.}$ are obtained as

$$\alpha_{opt.} = \frac{\mu f_2 [\mu A_3 - (f_1 A_2 + A_3)]}{[\mu^2 f_2 A_3 - \mu(f_1 f_2 A_2 + f_2 A_3 - f_1 A_1) - f_1 A_1]} \tag{15}$$

$$M(T)_{opt.} = [\mu C_1 - C_2] S_y^2 / n [\mu^2 C_3 - \mu C_4 - C_5] \tag{16}$$

where $C_1 = A_1 A_3$, $C_2 = f_1 A_1 A_2 + A_1 A_3$, $C_3 = f_2 A_3$, $C_4 = f_2 A_3 + f_1 f_2 A_2 - f_1 A_1$,

$C_5 = f_1 A_1$ and μ is the fraction of the sample drawn afresh at the current(second) occasion.

Remark 2.5.1: $M(T)_{opt.}$ derived in equation (16) is a function of μ. To estimate the population mean on each occasion the better choice of μ are 1(case of no matching); however, to estimate the change in mean from one occasion to other, μ should be 0(case of complete matching). But intuition suggests that the optimum choices of μ are desired to devise the amicable strategy for both the problems simultaneously.

2.6 Optimum Replacement Strategies for the Estimator T

In order to find the optimum value of μ, the mean squared error $M(T)_{opt.}$ in equation (16) have been minimized with respect to μ so obtained is one of the two roots given by

$$\hat{\mu} = \left(D_2 \pm \sqrt{D_2^2 - D_1 D_3}\right) \Big/ D_1 \tag{17}$$

where $D_1 = C_1 C_3$, $D_2 = C_2 C_3$, $D_3 = C_1 C_5 + C_2 C_4$.

The real value of $\hat{\mu}$ exist, iff $D_2^2 - D_1 D_3 \geq 0$. For any situation, which satisfies these conditions, two real values of $\hat{\mu}$ may be possible , hence to choose a value of $\hat{\mu}$, it should be taken care of that $0 \leq \hat{\mu} \leq 1$, all other values of $\hat{\mu}$ are inadmissible. For a situation having both the values of $\hat{\mu}$ as real, the minimum of the two can be considered as best choice. Hence, substituting the minimum real value of $\hat{\mu}$ say μ_0 from equation (17) in equation (16), the minimum value of the mean squared error of the estimator T with respect to both the parameters α as well as μ is obtained as

$$M(T)^*_{opt.} = [\mu_0 C_1 - C_2] S_y^2 \Big/ n \left[\mu_0^2 C_3 - \mu_0 C_4 - C_5\right] \tag{18}$$

3. Special Cases

3.1 Case I: When there is Non-Response only at the First Occasion (Previous Occasion)

When there is a presence of non-response, the proposed estimator T for population mean $\bar{Y}$ changes to

$$T_1 = \varphi T_u^\theta + (1-\varphi) T_m \tag{19}$$

where $T_u^\theta = \bar{y}_u \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(u)}{\bar{Z}_2 + \bar{z}_2(u)}\right)$ and T_m is defined in equation (5) and $\varphi\,(0 \leq \varphi \leq 1)$ is a real constant to be determined so as to minimize the mean squared error of the estimator T_1.

In this case, the optimum value of fraction of sample drawn afresh and the minimum mean squared error of the estimator T_1 at the admissible value of $\hat{\mu}$ are derived respectively as

$$\hat{\mu} = \left(D_5 \pm \sqrt{D_5^2 - D_4 D_6}\right) \Big/ D_4 = \mu_1 (say)$$

$$M(T_1)^*_{opt.} = [\mu_1 C_6 - C_7] S_y^2 \Big/ n \left[\mu_1^2 A_3 - \mu_1 C_8 - C_9\right] \tag{20}$$

Where,

$D_4 = A_3 C_6$, $D_5 = A_3 C_7$, $D_6 = C_6 C_9 + C_7 C_8$, $C_6 = A_1 A_3$, $C_7 = f_1 A_1 A_2 + A_1 A_3$

$C_8 = A_3 + f_1 A_2 - f_1 A_1$, $C_9 = f_1 A_1$ and $f_1 = r_1 / n$.

3.2 Case II: When there is Non-Response only at the Second (Current) Occasion

The estimator for population mean $\bar{Y}$ at the current occasion in the presence of non-response at current occasion is given by

$$T_2 = \psi T_u + (1-\psi) T_m^\theta \tag{21}$$

where $T_m^\theta = \bar{x}_n^* \left(\frac{\bar{y}_m^*}{\bar{x}_m^*}\right)$,

$\bar{y}_m^* = \bar{y}_m \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(m)}{\bar{Z}_2 + \bar{z}_2(m)}\right)$, $\bar{x}_m^* = \bar{x}_m \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(m)}{\bar{Z}_1 + \bar{z}_1(m)}\right)$,

$\bar{x}_n^* = \bar{x}_n \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(n)}{\bar{Z}_1 + \bar{z}_1(n)}\right)$ and T_u is defined in equation (2)

and $\psi\,(0 \leq \psi \leq 1)$ is a real constant to be determined so as to minimize the mean squared error of the estimator T_2.

In this case, the optimum value of fraction of sample drawn afresh and the minimum mean squared error of the estimator T_2 at the admissible value of $\hat{\mu}$ are derived respectively as

$$\hat{\mu} = \left(D_8 \pm \sqrt{D_8^2 - D_7 D_9}\right) \Big/ D_7 = \mu_2 (say)$$

$$M(T_2)^*_{opt.} = [\mu_2 C_{10} - C_{11}] S_y^2 \Big/ n \left[\mu_2^2 C_{12} - \mu_2 C_{13} - A_1\right] \tag{22}$$

where

$D_7 = C_{10} C_{12}$, $D_8 = C_{11} C_{12}$, $D_9 = A_1 C_{10} + C_{11} C_{13}$, $C_{10} = A_1 A_3$, $C_{11} = A_1 A_2 + A_1 A_3$

$C_{12} = f_2 A_3$, $C_{13} = f_2 A_3 + f_2 A_2 - A_1$ and $f_2 = r_2 / u$.

4. Efficiency Comparison

The percent relative loss in the efficiency of the proposed estimators T has been recorded to infer about the effect of incompleteness in the data over the occasions with respect to the estimator T_{CR} (Priyanka and Mittal (2016)) under the same circumstances but for complete response over the occasions which is described as

$$T_{CR} = \xi T_u^\theta + (1-\xi) T_m^\theta \tag{23}$$

Where, T_u^θ and T_m^θ have been defined in equation (19) and

Special Issue of Engineering and Scientific International Journal ISSN 2394-7187(Online) ISSN 2394-7179 (Print)
Proceedings of the UGC Sponsored - National Conference on ***"Recent Statistical Computing Techniques and their Applications"***
Department of Statistics and Department of Computer Science, Ramanujan College,University of Delhi, Delhi.
March 11-12, 2016. ***Published on January 2017***

(21) and $\xi\,(0 \le \xi \le 1)$ is a real constant to be determined so as to minimize the mean squared error of the estimator T_{CR}. The optimum mean squared error for the estimator T_{CR} with respect to ξ as well as μ is obtained as

$$M(T_{CR})^{*}_{opt.} = \left[\mu^{*} G_1 - G_2\right] S_y^2 \Big/ n \left[\mu^{*2} B_3 - \mu^{*} G_3 - B_1\right] \quad (24)$$

where $\mu^{*} = \left(H_2 \pm \sqrt{H_2^2 - H_1 H_3}\right) \Big/ H_1$,

$$H_1 = B_3 G_1,\; H_2 = B_3 G_2,\; H_3 = B_1 G_1 + G_2 G_3,\; G_1 = B_1 B_3,$$

$$G_2 = B_1 B_2 + B_1 B_3,\quad G_3 = B_3 + B_2 - B_1,\quad B_1 = (5/4) - \rho_{yz_2},$$

$$B_2 = (5/2) - 2\rho_{yx} + \rho_{yz_1} - \rho_{yz_2} - \rho_{xz_1} + \rho_{xz_2} - (1/2)\rho_{z_1 z_2}$$

$$\text{and } B_3 = 2\rho_{yx} - \rho_{yz_1} + \rho_{xz_1} - \rho_{xz_2} + (1/2)\rho_{z_1 z_2} - (5/4).$$

The percent relative loss in precision of the estimators T, T_1 and T_2 with respect to the estimator T_{CR} under their respective optimality conditions are given by

$$\left.\begin{aligned} L_0 &= \frac{M(T)^{*}_{opt.} - M(T_{CR})^{*}_{opt.}}{M(T)^{*}_{opt.}} \times 100 \\ L_1 &= \frac{M(T_1)^{*}_{opt.} - M(T_{CR})^{*}_{opt.}}{M(T_1)^{*}_{opt.}} \times 100 \\ L_2 &= \frac{M(T_2)^{*}_{opt.} - M(T_{CR})^{*}_{opt.}}{M(T_2)^{*}_{opt.}} \times 100 \end{aligned}\right\} \quad (25)$$

5. Numerical Illustrations and Simulation

5.1 Empirical Study

Population Source: [Free access to the data by Statistical Abstracts of the United States] Empirical validation of theoretical results has been elaborated by means of a natural population. The population I consist of N=51 states of United States. Let y_i be net summer capacity during 2008 in the i^{th} state of U. S., x_i denote the net summer capacity during 2000 in the i^{th} state of U. S., z_{1_i} denote the residential consumption of electric power during 2000 and z_{2_i} denote the residential consumption of electric power during 2008. The empirical analysis of the proposed estimators has been shown in Table 1 for various choices for fraction of non-response over the successive occasions.

5.2 Generalization of Empirical Study

A more generalized study has also been done to show the impact of the proposed estimators under different fractions of non-response and choices of correlation coefficients of study and auxiliary variables. The results obtained are shown in Table 2. Here for the sake of convenience, we have considered $\rho_{yz_1} = \rho_{yz_2} = \rho_1$ and $\rho_{xz_1} = \rho_{xz_2} = \rho_2$.

Table 1: Empirical results when the proposed estimators T, T_1 and T_2 have been compared to the estimator T_{CR}

$t_1=0.30,\quad t_2=0.30$						
μ^{*}	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.6773	0.6018	0.7347	0.4312	19.80	3.03	10.31
$t_1=0.20,\quad t_2=0.20$						
μ^{*}	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.6773	0.6050	0.6968	0.5062	11.86	1.03	6.41
$t_1=0.25,\quad t_2=0.15$						
μ^{*}	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.6773	0.6540	0.7158	0.5387	10.56	2.03	4.25

Table 2: Generalized empirical results while the proposed estimators T, T_1 and T_2 have been compared to the estimator T_{CR}

$t_1=0.10$ and $t_2=0.10$								
ρ_{yx}, $\rho_{z_1 z_2}$		**0.5**						
ρ_2	ρ_1	μ^{*}	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.4	**0.4**	0.81	0.63	0.50	0.59	11.49	5.62	11.99
	0.5	0.78	0.58	0.49	0.54	12.54	6.88	13.34
	0.6	0.75	0.55	0.47	0.50	13.71	8.20	14.87
0.5	**0.4**	0.84	0.63	0.50	0.59	11.90	6.07	12.40
	0.5	0.81	0.58	0.49	0.54	13.14	7.51	13.93
	0.6	0.78	0.55	0.47	0.50	14.51	9.05	15.66
0.6	**0.4**	0.83	0.63	0.50	0.59	12.21	6.40	12.71
	0.5	0.83	0.58	0.49	0.54	13.59	7.99	14.37
	0.6	0.80	0.55	0.47	0.50	15.12	9.70	16.26
0.7	**0.4**	0.87	0.63	0.50	0.59	12.45	6.65	12.95
	0.5	0.85	0.58	0.49	0.54	13.94	8.36	14.72
	0.6	0.82	0.55	0.47	0.50	15.60	10.21	16.74
$t_1=0.25$ and $t_2=0.20$								
ρ_{yx}, $\rho_{z_1 z_2}$		**0.5**						
ρ_2	ρ_1	μ^{*}	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.4	**0.4**	0.81	0.81	0.58	0.75	18.83	4.13	18.92
	0.5	0.78	0.74	0.57	0.66	18.72	4.87	19.66
	0.6	0.75	0.69	0.56	0.59	19.07	5.59	20.75

0.5	**0.4**	0.84	0.81	0.58	0.75	18.91	4.58	19.30
	0.5	0.81	0.74	0.57	0.66	19.27	5.52	20.21
	0.6	0.78	0.69	0.56	0.59	19.82	6.46	21.48
0.6	**0.4**	0.83	0.81	0.58	0.75	19.20	4.92	19.59
	0.5	0.83	0.74	0.57	0.66	19.69	6.01	20.62
	0.6	0.80	0.69	0.56	0.59	20.39	7.13	22.04
0.7	**0.4**	0.87	0.81	0.58	0.75	19.42	5.18	19.81
	0.5	0.85	0.74	0.57	0.66	20.01	6.39	20.94
	0.6	0.82	0.69	0.56	0.59	20.84	7.66	22.49

Note: The values for μ^*, μ_0, μ_1 and μ_2 have been rounded off up to two places of decimal for presentation.

Table 3: Generalized empirical results while the proposed estimators T, T_1 and T_2 have been compared to the estimator T_{CR}

$t_1=0.10$ and $t_2=0.10$								
ρ_{yx}, $\rho_{z_1z_2}$	**0.6**							
ρ_2	ρ_1	μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.4	**0.4**	0.91	0.83	0.81	0.81	9.85	3.03	9.89
	0.5	0.87	0.68	0.65	0.65	1.93	4.83	11.21
	0.6	0.83	0.61	0.57	0.57	12.35	6.60	12.95
0.5	**0.4**	0.93	0.83	0.81	0.81	9.95	3.14	9.99
	0.5	0.89	0.68	0.65	0.65	11.18	5.09	11.46
	0.6	0.85	0.61	0.57	0.57	12.79	7.07	13.38
0.6	**0.4**	0.94	0.83	0.81	0.81	10.02	3.21	10.07
	0.5	0.90	0.68	0.65	0.65	11.36	5.28	11.63
	0.6	0.87	0.61	0.57	0.57	13.11	7.41	13.70
0.7	**0.4**	0.94	0.83	0.81	0.81	10.07	3.27	10.12
	0.5	0.91	0.68	0.65	0.65	11.49	5.43	11.77
	0.6	0.88	0.61	0.57	0.57	13.35	7.67	13.95
$t_1=0.25$ and $t_2=0.20$								
ρ_{yx}, $\rho_{z_1z_2}$	**0.6**							
ρ_2	ρ_1	μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.4	**0.4**	0.91	**	0.60	**	-	2.41	-
	0.5	0.87	0.90	0.59	0.87	18.85	3.73	18.93
	0.6	0.83	0.78	0.58	0.71	1902	4.94	19.59
0.5	**0.4**	0.93	**	0.60	**	-	2.52	-
	0.5	0.89	0.90	0.59	0.87	19.08	4.00	19.16
	0.6	0.85	0.78	0.58	0.71	19.42	5.41	19.99
0.6	**0.4**	0.94	**	0.60	**	-	2.60	-
	0.5	0.90	0.90	0.59	0.87	19.24	4.19	19.32
	0.6	0.87	0.78	0.58	0.71	19.72	5.76	20.29
0.7	**0.4**	0.94	**	0.60	**	-	2.66	-
	0.5	0.91	0.90	0.59	0.87	19.36	4.34	19.44
	0.6	0.88	0.78	0.58	0.71	19.94	6.02	20.51

5.3 Monte Carlo Simulation

The population II comprise of N = 51 states of United States. Let y_i be the net electric power generation during 2008 in the i^{th} state of U. S., x_i be the net electric power generation during 2000 in the i^{th} state of U. S., z_{1_i} denote the net summer capacity during 2000 in the i^{th} state of U. S. and z_{2_i} denote the net summer capacity during 2008 in the i^{th} state of U. S.

Monte Carlo simulation has been performed on population II, for better analysis considering different choices of t_1 and t_2.

5.3.1 Simulation Algorithm

(i) Choose 5000 samples of size n=25 using simple random sampling without replacement on first occasion for both the study and auxiliary variable.

(ii) For $f_1 = 0.88$, choose $r_1 = 22$ responding units out of n=25 samples units.

(iii) Calculate sample mean $\bar{x}_{r_1|k}$ and $\bar{z}_{1|k}(r_1)$ for k =1, 2, - - -, 5000.

(iv) Retain m=15 units out of each $r_1 = 22$ sample units of the study and auxiliary variables at the first occasion.

(v) Calculate sample mean $\bar{x}_{m|k}$ and $\bar{z}_{1|k}(m)$ for k= 1, 2, - - -, 5000.

(vi) Select u=10 units using simple random sampling without replacement from N-n=26 units of the population for study and auxiliary variables at second (current) occasion.

(vii) For $f_2 = 0.90$, choose $r_2 = 9$ responding units out of u=10 samples units.

(viii) Calculate sample mean $\bar{y}_{r_2|k}$, $\bar{y}_{m|k}$, $\bar{z}_{2|k}(m)$ and $\bar{z}_{2|k}(r_2)$ for k = 1, 2, - - -, 5000.

(ix) Iterate the parameter α from 0.1 to 0.9 with a step of 0.2.

(x) Iterate ξ from 0.1 to 0.9 with a step of 0.1 within (ix).

(xi) Calculate the percent relative loss in efficiencies of the proposed estimator T, T_1 and T_2 with respect to estimator T_{CR} as

$$L(T) = \frac{\sum_{k=1}^{5000} [T_{|k} - T_{CR|k}]^2}{\sum_{k=1}^{5000} [T_{|k}]^2} \times 100,$$

$$L(T_1) = \frac{\sum_{k=1}^{5000} [T_{1|k} - T_{CR|k}]^2}{\sum_{k=1}^{5000} [T_{1|k}]^2} \times 100$$

and $$L(T_2) = \frac{\sum_{k=1}^{5000} [T_{2|k} - T_{CR|k}]^2}{\sum_{k=1}^{5000} [T_{2|k}]^2} \times 100, \quad k=1, 2, ..., 5000.$$

Table 4: Simulation result when the proposed estimator T is compared with the estimator T_{CR} when non-response occurs on both the occasion

α \ ξ	SET	0.1	0.3	0.5	0.7
0.1	I	46.05	66.92	39.65	-20.54
	II	31.76	55.63	21.70	-59.24
	III	25.14	49.85	22.09	-75.52
0.2	I	41.00	62.83	34.92	-35.84
	II	22.79	50.56	14.35	-79.91
	III	14.44	42.69	7.15	-82.36
0.3	I	38.96	60.92	31.50	-40.43
	II	17.24	47.62	9.23	-91.28
	III	7.94	38.42	0.211	-100.57
0.4	I	38.93	60.81	32.26	-41.78
	II	18.38	47.74	10.57	-89.35
	III	7.79	38.07	-1.01	-114.80
0.5	I	43.25	62.87	36.25	-33.20
	II	24.81	51.11	16.44	-76.86
	III	11.57	42.22	4.43	-106.78
0.6	I	49.73	66.09	42.63	-19.59
	II	33.26	56.74	25.91	-57.18
	III	21.53	48.56	14.57	-83.80
0.7	I	56.23	70.48	49.70	-4.89
	II	42.93	63.18	36.78	-34.08
	III	32.87	55.59	26.18	-56.75
0.8	I	62.35	75.10	57.24	10.94
	II	52.18	69.19	46.56	-12.58
	III	43.92	62.51	38.34	-31.46
0.9	I	68.17	78.73	64.14	24.08
	II	60.26	74.18	55.75	6.61
	III	53.71	68.74	48.86	9.51

I: n=25, $\mu = 0.40$, $t_1=0.28$, $t_2=0.30$
II: n=25, $\mu = 0.40$, $t_1=0.16$, $t_2=0.20$
III: n=25, $\mu = 0.40$, $t_1=0.12$, $t_2=0.10$

Table 5: Simulation result when the proposed estimator T_1 is compared with the estimator T_{CR} when non- response occurs only on first occasion

φ \ ξ	SET	0.1	0.3	0.5	0.7
0.1	I	50.48	66.33	42.76	-12.67
	II	39.01	58.29	32.93	-42.17
	III	15.55	46.11	13.02	-78.92
0.2	I	42.35	62.20	38.16	-25.80
	II	27.35	51.74	23.25	-63.11
	III	3.85	38.69	1.42	-105.64
0.3	I	36.84	58.90	32.90	-39.54
	II	18.80	47.64	16.23	-78.27
	III	-4.81	33.18	-8.18	-125.97
0.4	I	33.01	56.58	29.03	-47.21
	II	15.93	44.81	12.45	-86.36
	III	-7.52	31.00	-11.53	-132.10
0.5	I	31.89	55.96	28.12	-48.72
	II	16.13	45.46	13.17	-84.67
	III	-3.76	33.62	-6.90	-123.98
0.6	I	33.77	57.31	30.42	-43.88
	II	21.25	49.17	18.03	-72.92
	III	5.63	39.17	2.71	-102.72
0.7	I	38.97	60.54	35.32	-33.70
	II	29.43	54.49	26.19	-55.78
	III	18.00	47.01	15.71	-75.88
0.8	I	45.31	64.55	42.39	-19.92
	II	38.56	60.40	35.62	-35.71
	III	30.99	55.73	28.19	-49.95
0.9	I	52.39	68.83	49.50	-5.54
	II	47.67	66.10	45.13	-15.94
	III	42.67	63.03	39.97	-24.82

I: n=25, $\mu = 0.40$, $t_1=0.28$
II: n=25, $\mu = 0.40$, $t_1=0.16$
III: n=25, $\mu = 0.40$, $t_1=0.12$

6. Rendition of Results

The performance of an estimator in successive sampling in the presence of non-response is generally judged on the basis of percent relative loss in efficiency (lesser is the loss better is the estimator) and in terms of optimum value of fraction of fresh sample to be drawn afresh on current(second) occasion which directly related to the cost of survey. Following interpretation can be drawn from Tables 1- 6,

Table 6: Simulation result when the proposed estimator T_2 is compared with the estimator T_{CR} when non-response occurs only on second occasion

ψ \ ξ	SET	0.1	0.3	0.5	0.7
0.1	**I**	38.23	59.08	31.01	-45.26
	II	37.14	57.99	28.97	-48.73
	III	35.67	59.41	27.97	-49.80
0.2	**I**	39.89	59.63	30.71	-45.18
	II	36.79	57.84	27.28	-52.52
	III	33.84	58.05	24.82	-54.67
0.3	**I**	42.86	61.30	34.87	-37.21
	II	38.89	59.65	30.58	-46.86
	III	35.46	58.75	27.50	-50.78
0.4	**I**	47.48	65.36	41.25	-24.02
	II	43.37	62.44	35.71	-33.67
	III	38.81	60.30	31.41	-42.55
0.5	**I**	54.24	69.72	48.85	-7.38
	II	49.43	66.35	42.67	-19.20
	III	44.53	63.44	36.95	-30.51
0.6	**I**	60.24	74.16	55.69	6.91
	II	56.03	70.84	50.14	-3.87
	III	50.54	67.35	43.65	-16.55
0.7	**I**	66.19	77.92	62.46	21.03
	II	61.97	74.79	56.88	9.71
	III	56.78	71.62	51.09	-2.07
0.8	**I**	71.42	81.29	68.36	32.87
	II	67.38	78.11	63.25	22.38
	III	62.51	75.22	57.45	11.42
0.9	**I**	75.87	83.97	73.02	42.92
	II	71.87	81.26	68.50	33.36
	III	67.34	78.70	63.38	23.12

I: n=25, $\mu = 0.40$, $t_2=0.40$
II: n=25, $\mu = 0.40$, $t_2=0.30$
III: n=25, $\mu = 0.40$, $t_2=0.20$

6.1 Results based on Empirical Study

- From Table 1 we can see that the values of μ_0, μ_1, and μ_2 exist for various choices of fraction of non-response over two successive occasions which completely signifies the utility of a dynamic natured auxiliary character.
- Also from Table 1, we identify that percent relative loss L_0, L_1 and L_2 exist each combination of t_1 and t_2 and when non response occurs at both occasion the percent relative loss is more as compared to non-response on first or second occasion only.
- We can also conclude from the Table 1 that, the percent relative loss in efficiency is not very much significant when the proposed estimators T, T_1 and T_2 are compared to estimator T_{CR}.

6.2 Results Extracted from the Generalized Study for Various Combinations of Correlation Coefficients

- In Table 2 and Table 3, we see that the values of $\mu_0, \mu_1, \mu_2, L_0, L_1$ and L_2 exist for almost every combination of coefficient of correlation of study and auxiliary characteristics considering various possibilities of non-response that can creep in a sample survey over successive occasions.
- We also see that the proposed estimators work efficiently when auxiliary character which is dynamic in nature conceives a moderate or low correlation with the study character over the successive occasions.
- We also identify that as the amount of correlation between study and auxiliary character is increased, the proposed estimators intend to provide a lesser fraction of sample to be drawn afresh at current occasion.

6.3 Results Based on Simulation Study in Table 4, Table 5 and Table 6

- We can see that for fixed choice of ξ, percent relative loss first decreases and then increases with increasing value of α and φ respectively while for all fixed choices of ξ, percent relative loss $L(T_2)$ increases as ψ increases.
- The percent relative loss $L(T), L(T_1)$ and $L(T_2)$, for fixed choice of α, φ and ψ respectively, first increase as ξ increases and then decreases with increasing value of ξ.
- As we decrease the fraction of non-response in the sample on first and second occasion, the percent relative loss $L(T), L(T_1)$ and $L(T_2)$ decrease for all combinations of α, φ and ψ with ξ respectively.

10. Conclusion

The proposed estimators have been analysed considering a detailed study in the presence of non-response utilizing additional auxiliary information which is dynamic in nature over the successive occasions. Loss in efficiency is very much plebeian when non-response is encountered in the sample survey. Although percent relative loss is encountered for various fractions of non-response on two occasions using the proposed method of imputation while utilizing dynamic auxiliary character but a negative loss is also available for various choices of parameters. This signifies that the proposed estimators emerge to be better than the estimator due to Priyanka and Mittal (2016) for such combinations of parameters and hence the proposed method of imputation is fruitful to cope with the non-response. The proposed imputation techniques prove to be worthy from the point of cost as well when correlation between study and auxiliary character is considered moderate or even low. Hence, it is observed that the proposed imputation methods deal the sour effect of non-response excellently, therefore, the proposed estimators may be recommended for encouraging their practical use by survey practitioners.

References

[1] A. Bandyopadhyay and G. N. Singh, "On the use of two auxiliary variables to improve the precision of estimate in two-occasion successive sampling", Int. J. Math. Stat., Vol. 15, pp. 73-88, 2014.
[2] A. R. Eckler, "Rotation Sampling", Ann. Mathe. Stat., Vol. 26, pp. 664-685, 1955.
[3] G. Kalton, D. Kasprzyk and R. Santos, "Issues of non-response and imputation of income and program participation", In :D. Krevoski, R. Platek, J.N.K. Rao eds. Current Topic in Survey Sampling, New York Academic Press 1981, pp.455-480.
[4] G. Kalton and D. Kasprzyk, "Imputating for missing survey responses", Proc. Sec. Surv. Res. Meth. Amer. Statist. Assoc. 1982, pp. 22-31.
[5] G. N. Singh and K. Priyanka, "Effect of non-response on current occasion in search of good rotation patterns on successive occasions", Statist. Transit. New Ser., Vol. 8, pp. 273-292, 2007.
[6] G. N. Singh and K. Priyanka, "Search of good rotation patterns to improve the precision of estimates at current occasion", Comm. Stat. – Theo. Meth., Vol. 37, pp.337 – 348, 2008.
[7] G. N. Singh, K. Priyanka, S. Prasad, S. Singh, and J. M. Kim, "A class of estimators for estimating the population variance in two occasion rotation patterns", Comm. Statis. App. Meth., Vol. 20, pp. 247-257, 2013a.
[8] G. N. Singh, D. Majhi, S. Prasad and F. Homa, "Assessment of Non-response under ratio method of imputation in two occasion successive sampling", J. statist. theo. Appl. Vol. 12, pp. 403-418, 2013b.
[9] H. D. Patterson, "Sampling on successive occasions with partial replacement of units", J. Royal Statis. Soci. Vol. 12, pp. 241-255, 1950.
[10] H. Lee, E. Rancourt and C. E. Särndall, "Experiments with variance estimation from survey data with imputed values", J. official Statist., Vol. 10, pp. 231-243, 1994.
[11] H. Lee, E. Rancourt and C. E. Särndall, "Variance estimation in the presence of imputed data for generalized estimation system", Proc. Amer. Statist. Assoc. 1995, pp. 384-389.
[12] I. G. Sande, "A personal view of hot-deck imputation procedure", Surv. Methodo. Statist. Can., Vol. 5, pp. 238-258, 1979.
[13] L. Gordon, "Successive sampling in finite populations", The Ann. Stat., Vol. 11, pp. 702 – 706, 1983.
[14] M. S. Ahmed, O. Al-Titi, Z. Al-Rawi, and W. Abu-Dayyeh, "Estimation of a population mean using different imputation methods", Statist. Transit., Vol. 7, pp. 1247-1264, 2006.
[15]. K. Priyanka, and R. Mittal, "Effective rotation patterns for median estimation in successive sampling". Statis. Trans., Vol. 15, pp. 197-220, 2014.
[16]. K. Priyanka, R. Mittal, and J. Min-Kim, "Multivariate Rotation Design for Population Mean in sampling on Successive Occasions". Comm. Statis. Appli. Meth., Vol. 22, pp.445-462, 2015.
[17]. K. Priyanka, and R. Mittal, "Estimation of Population Median in Two-Occasion Rotation Sampling". J. Statis. Appli. Proba. Lett., Vol. 2, pp. 205-219, 2015a.
[18]. K. Priyanka, and R. Mittal, "A Class of Estimators for Population Median in two Occasion Rotation Sampling". HJMS, Vol. 44, pp. 189 – 202, 2015b.
[19]. K. Priyanka, and R. Mittal, "New Approaches using Exponential Type Estimator with Cost Modelling for Population Mean on Successive Waves". Statis. Trans., (Accepted) (2016).
[20] R. Arnab and F. C. Okafor, "A note on double sampling over two occasions", Paki. J. Stat. Vol. 8, pp. 9-18, 1992.
[21] R. B. Rubin, "Inference and missing data", Biometrika., Vol. 63, pp. 581-592, 1976.
[22] R. D. Narain, "On the recurrence formula in sampling on successive occasions", J. Indi. Soci. Agri. Stat. Vol. 5, pp. 96-99, 1953.
[23] R. J. Jessen, "Statistical investigation of a sample survey for obtaining farm facts", Iowa Agri. Experi. Stat. Road Bull. 304, pp. 1-104, 1942.
[24] R. S. Biradar and H. P Singh, "Successive sampling using auxiliary information on both occasions", Cal. Statist. Assoc. Bull. Vol. 51, pp. 234-251, 2001.
[25] R. Singh and N. Singh, Imputation methods in two-dimensional survey. Recent Advances in Agricultural Statistics Research, Wiley Eastern Ltd, New Delhi, 1991.
[26] S. Feng and G. Zou, "Sample rotation method with auxiliary variable", Comm. Stat.- Theo. Meth. Vol. 26, pp. 1497-1509, 1997.
[27] S. Singh, "A new method of imputation in survey Sampling", Statist. Vol. 43, pp. 499-511, 2009.