Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.12785/jsapl/020202

On the Use of Compromised Imputation for Missing data using Factor-Type Estimators

*Priyanka Singh**, *Ajeet Kumar Singh and V. K. Singh*

Department of Statistics, Banaras Hindu University, Varanasi, 221005, India

Received: 18 Nov. 2014, Revised: 24 Dec. 2014, Accepted: 28 Dec. 2014
Published online: 1 May 2015

Abstract: There are several methods of handling missing data in sample surveys, which is a typical problem of non-response. *Imputation* (fill-in) method is one of the methods to deal with non-response. The term Imputation refers to the process of assigning one or more values to an item when there is no reported value for that item. Many forms of imputation are available, including mean imputation, ratio method of imputation, hot deck imputation, cold deck imputation, regression imputation, etc. In recent past, a number of efficient compromised imputation strategies has been proposed by several survey statisticians.This paper suggests a one-parameter family of estimators, popularly known as Factor-Type Estimator (FTE), with compromised imputation strategy and discusses its properties. The proposed strategy has been observed to be more precise than other compromised estimators under optimality conditions. To support the discussed results, the relative efficiencies of the estimator have been obtained using four sets of empirical data.

Keywords: Compromised imputation, one-parameter family of estimators, optimum estimator, relative efficiency.

1 Introduction

Basic sampling theory assumes that the variable of interest is measured on every unit in the sample without error; but errors may arise in many situations. Besides sampling error, which is an essential part of a sample survey, sometimes non-sampling errors and particularly, non coverage is a quite serious problem because of the simple reason that the sample tends to be unrepresentative of the population and the estimates are biased (Thompson, [20]). Non-response (or non-coverage) is an inherent characteristic of any type of population and, therefore, cannot be eliminated by any means, rather, efficient methods are to be developed for estimating population parameters with the help of missing data so obtained.

A common technique for handling non-response is imputation, where the missing values are filled in to create a complete data set that can be analysed with traditional analysis methods. It is important to note that usually sample surveys are considered with the goal of making inferences about population quantities such as means, variances, correlations and regression coefficients, and the values of individual cases in the data set are not the main interest. Thus, the objective of *Imputation* is not to get the best possible predictions of the missing values, but to replace them by plausible values in order to exploit the information in the recorded variables in the incomplete cases for inference about population parameters (Little and Rubin,[4]).

Mean imputation, hot deck imputation, regression imputation, ratio imputation are all single imputation in the sense that a single value is imputed for every missing value to produce a complete data set. To deal with missing values effectively Kalton *et al*[1] and Sande[9] suggested imputation methods that make an incomplete data set structurally complete and its analysis. Lee *et al* ([2], [3]) used the information on an auxiliary variable for the purpose of imputation. Based on auxiliary variable,recently Singh and Horn [14] and Singh *et al* [15] suggested some compromised methods of imputation.

The purpose of this paper is to (i) suggest a one parameter family of estimators for population mean using compromised imputation strategy under the assumption of presence of non-response in the population and with the aid of information on an ancillary characteristic, (ii) discuss some of its salient properties,(iii) show its supremacy over some existing

* Corresponding author e-mail: priya.bhu.vns@gmail.com

compromised estimators and(iv) apply the proposed strategy on some empirical populations for illustration purpose. The overall study reveals that the proposed imputation strategy works efficiently in a variety of situations.

2 The problem and notations:

Let a simple random sample S of size n without replacement be drawn from a finite population $U=(y_1,y_2,...,y_N)$ of size N and with study characteristic Y. Let $(\overline{Y},\ \overline{X})$ be the population mean for the study variable Y and auxiliary variable X respectively. It is presumed that the sample consists of r responding units $(r<n)$ belonging to a set R and (n-r) non-responding units belonging to the set R^c. Further, let for every unit $i\in R$, the value y_i is observed and for the unit $i\in R^c$, the value y_i is missing for which suitable imputed value is to be derived. For this purpose, the ith value of the auxiliary variable is used as a source of imputation for missing data when $i\in R^c$.
In what follows, we shall use the following notations:

Z: Stands for either variable Y or variable X.
$\overline{z}_n$: Sample mean based on n observations for variable Z.
$\overline{z}_r$: Sample mean of the responding units based on r observations for the variable Z.
S_Z^2: Population mean square for the variable Z.
C_Z: Coefficient of variation (CV) for the variable Z ; C_Z=$\dfrac{S_Z}{\overline{Z}}$.
ρ: Coefficient of correlation between variables Y and X in the population.
$y_{.i}$: Imputed value for the ith value of y_i $(i=1,2,3....n)$.
$\theta_{n,N}$, $\theta_{r,N}$, $\theta_{r,n}$: Finite population corrections (fpc); $(\frac{1}{n}-\frac{1}{N})$, $(\frac{1}{r}-\frac{1}{N})$, $(\frac{1}{r}-\frac{1}{n})$ respectively.

3 Some imputation strategies :

Before suggesting the proposed imputation strategy, we shall mention here some existing imputation strategies for readiness of the material which has a direct relevance with the present work. We shall denote by (D,T) a sampling strategy where D stands for simple random sampling without replacement sampling scheme and T for an estimator for population mean $\overline{Y}$. Followings are some imputation methods and corresponding sampling strategies :

(a).$(D,\overline{y}_r$): ***Mean method***
Here

$$y_{.i}=\begin{cases} y_i \text{ if } & i\in R \\ \overline{y}_r \text{ if } & i\in R^c \end{cases} \tag{1}$$

The corresponding point estimator and its bias, B(.) and mean square error (MSE), M(.) are derived as

$$\overline{y}_n=\frac{1}{n}\sum_{i\in s}y_{.i}=\overline{y}_r \tag{2}$$

$$B[\overline{y}_r]=0 \tag{3}$$

$$M[\overline{y}_r]=\theta_{r,N}\overline{Y}^2C_Y^2. \tag{4}$$

(b).$(D,\overline{y}_{RAT})$: ***Ratio method***

$$y_{.i}=\begin{cases} y_i & \text{if} \quad i\in R \\ \hat{b}x_i & \text{if} \quad i\in R^c \end{cases} \tag{5}$$

where $\hat{b}=\dfrac{\sum_{i\in R}y_i}{\sum_{i\in R}x_i}$. Then the point estimator, its bias and MSE are given by:

$$\overline{y}_{RAT}=\overline{y}_r\frac{\overline{x}_n}{\overline{x}_r} \tag{6}$$

$$B(\overline{y}_{RAT})=\theta_{r,n}\overline{Y}\left[C_X^2-\rho C_YC_X\right] \tag{7}$$

$$M(\bar{y}_{RAT}) = \theta_{r,N}\bar{Y}^2 C_Y{}^2 + \theta_{r,n}[\bar{Y}]^2 \left[C_X^2 - 2\rho C_X C_Y\right] \tag{8}$$

(c).$(D, \bar{y}_{COMP})$: ***Compromised method(Singh and Horn, [14])***

$$y_{.i} = \begin{cases} \alpha \frac{n}{r} y_i + (1-\alpha)\hat{b}x_i & \text{if} \quad i \in R \\ (1-\alpha)\hat{b}x_i & \text{if} \quad i \in R^c \end{cases} \tag{9}$$

The point estimator is

$$\bar{y}_{COMP} = \alpha \bar{y}_r + (1-\alpha)\hat{b}\bar{x}_n \tag{10}$$

α being a suitable constant with

$$B(\bar{y}_{COMP}) = (1-\alpha)\,\theta_{r,n}\bar{Y}\left[C_X^2 - \rho C_Y C_X\right] \tag{11}$$

$$M(\bar{y}_{COMP}) = \theta_{r,N}\bar{Y}^2 C_Y^2 + \theta_{r,n}\bar{Y}^2\left[(1-\alpha)^2 C_X^2 - 2(1-\alpha)\rho C_X C_Y\right] \tag{12}$$

Some other compromised estimators are proposed by Singh *et al* [16] and Singh *et al* [15].

4 (D, T_k):*Proposed imputation strategy and compromised estimator :*

Based on an unknown constant $k > 0$, we now propose the following imputation strategy and corresponding one parameter family of estimators, T_k as

$$y_{.i} = \begin{cases} k\frac{n}{r} y_i + (1-k)\,\phi_k & \text{if} \quad i \in R \\ (1-k)\phi_k & \text{if} \quad i \in R^c \end{cases} \tag{13}$$

where $\phi_k = \bar{y}_r\ \psi\{k, \bar{x}_r, \bar{x}_s\}$. Here $\psi\{k, \bar{x}_r, \bar{x}_s\}$ is a function of k, $\bar{x}_r$ and $\bar{x}_s$ such that

$$\psi\{k, \bar{x}_r, \bar{x}_s\} = \frac{\eta\{t_1(k)\}}{\eta\{t_2(k)\}}; \tag{14}$$

$\eta\,(t_i(k)) = t_i(k) + (1 - t_i(k))\dfrac{\bar{x}_r}{\bar{x}_s}$; $i = 1, 2.$
$t_1(k)$=$\dfrac{fB}{(A+fB+C)}$, $t_2(k)$=$\dfrac{C}{(A+fB+C)}$;
A=$(k-1)(k-2)$, B=$(k-1)(k-4)$, C=$(k-2)(k-3)(k-4)$
and $f = \dfrac{n}{N}$, $\bar{x}_s = s\bar{x}_r + (1-s)\bar{X}$, $s = \dfrac{r}{r+n}$.
The corresponding point estimator for population mean $\bar{Y}$, T_k is then obtained as:

$$T_k = \bar{y}_r\left[k + (1-k)\psi\{k, \bar{x}_r, \bar{x}_s\}\right] \tag{15}$$

Remark 1: It is evident that T_k defines a family of estimators under compromised imputation for missing values in the sample; k being the parameter. In fact, a comparison of T_k with $\bar{x}_{0F}(t)$, defined in Singh *et al* [19], reveals that T_k could be considered a FTE, initially defined by Singh and Shukla [17] and Shukla [10], with compromised imputation in the presence of non-response.The other contribution on FTE are due to Shukla *et al* [12] and Singh and Shukla [18].

Remark 2: Letting the value of $k = 1$ and 4, it is seen that T_1= T_4 = $\bar{y}_r$, the estimator under mean method of imputation and for $k = 2$, T_2= $\bar{y}_r\left[2 - \dfrac{\bar{x}_s}{\bar{x}_r}\right]$, an estimator equivalent to Sahai and Ray [8] estimator: $\bar{y}_{SR} = \bar{y}_n\left[2 - \left(\dfrac{\bar{x}_n}{\bar{X}}\right)^{\alpha}\right]$

5 Properties of the proposed family :

5.1 Theorem 1:

The bias and MSE of the proposed strategy (D,T_K) to the terms of order $O(n^{-1})$ are given by

$$B[T_K] = \bar{Y}\theta_{r,N}(1-k)(d_1-d_2)\left\{\rho C_Y C_X - d_2 C_X^2\right\} \tag{16}$$

$$M[T_k] = \bar{Y}^2\theta_{r,N}\left[C_Y^2 + (1-k)^2(d_1-d_2)^2 C_X^2 + 2(1-k)(d_1-d_2)\rho C_Y C_X\right] \tag{17}$$

where $d_1 = \dfrac{A+fsB+C}{A+fB+C}$, $d_2 = \dfrac{A+fB+Cs}{A+fB+C}$.

The proof of the theorem is given in the Appendix.

Corollary 1: The bias and MSE of strategies $(D,\bar{y}_r)$ and (D,T_2) are given by

$$B[\bar{y}_r] = 0 \tag{18}$$

$$V[\bar{y}_r] = \theta_{r,N}\bar{Y}^2 C_Y^2 \tag{19}$$

$$B[T_2] = 2\bar{Y}\theta_{r,N}s\left\{\rho C_Y C_X - C_X^2\right\} \tag{20}$$

$$M[T_2] = \bar{Y}^2\theta_{r,N}\left\{C_Y^2 + (1+2s)^2 C_X^2 + 2(1+2s)\rho C_Y C_X\right\} \tag{21}$$

Letting k=1 and thereby d_1 =1 , d_2 =s and k=2 and hence d_1 =$-2s$, d_2 =1 in (16) and (17), the above expressions are straight forward.

5.2 Optimum estimator in the family:

Theorem 2: *The optimum choices of the parameter k which minimizes $M[T_k]$ are the real and positive roots of the equation*

$$(1-k)(d_1-d_2) = -\rho\frac{C_Y}{C_X} = -V(say) \tag{22}$$

and minimum MSE is given by

$$M_{min}[T_k] = \theta_{r,N}\bar{Y}^2 C_Y^2\left(1-\rho^2\right) \tag{23}$$

Proof: Re-writing $M[T_k]$ as

$$M[T_k] = \bar{Y}^2\theta_{r,N}\left\{C_Y^2 + H^2 C_X^2 + 2H\rho C_Y C_X\right\} \tag{24}$$

where H= $(1-k)(d_1-d_2)$ and realising that H is a function of k, in order to obtain optimum choices of k, we differentiate (24) with respect to H and equate to zero. Hence we have

$$\frac{\partial M[T_k]}{\partial H} = \theta_{r,N}\bar{Y}^2\left[2HH' C_X^2 + 2H'\rho C_Y C_X\right] = 0 \tag{25}$$

where $H' = \dfrac{\partial H}{\partial k}$, Since $H' \neq 0$,from (25), we have;

$$H = (1-k)(d_1-d_2) = -\rho\frac{C_Y}{C_X} = -V \tag{26}$$

Thus (22) follows. Further, substituting H from (26) to (17), we obtain (23).

Remark 3: A close look of the equation (22) reveals that it is a fourth degree equation in k. Therefore, for a given population (known value of V), one will get four optimum choices of k for which $M[T_k]$ would be minimum, having same values. However, the equation might yield some negative and imaginary values of k. Shukla [10] has pointed out that the equation yields at least one optimum value of $k>0$. On simplification of the equation (22) we have:

$$-(1-s)k^4+\{V+(1-s)(10+f)\}k^3-\{V(8-f)+(1-s)(35+6f)\}k^2 \\ +\{V(23-5f)+(1-s)(50+9f)\}k-\{V(22-4f)+4(1-s)(6+f)\}=0 \quad (27)$$

Remark 4: Since, for a given population, equation (27) might yield more than one real and positive root, the question arises as to how the appropriate choice of optimum k could be done amongst these values. A criterion for selecting suitable value may be set as follows :
"Out of all real and positive roots of the equation , select that optimum k which makes $|B[T_k]|$ smallest. "
Thus, using the proposed strategy, one can put control on the bias of the estimator along with minimising MSE.

6 Comparison of different strategies:

6.1 Basis of expressions of MSEs

On the basis of expressions of MSEs of different strategies, as discussed under sections 3 and 4, a comparison of the strategies can be made under optimality conditions.

6.2 Comparison of (D,T_k) with $(D,\bar{y}_r)$

From expression (4) and (23), we have $M_{min}[T_k]<V(\bar{y}_r)$ when

$$\bar{Y}^2\theta_{r,N}\rho^2C_Y^2>0, \quad (28)$$

which is always true. Therefore, proposed imputation strategy is superior to mean imputation method. It is also a trivial result, as $\bar{y}_r$ is a member of the proposed family.

6.3 Comparison of (D,T_k) with $(D,\bar{y}_{RAT})$

A comparison of expressions (8) with (23) reveals that

$M_{min}[T_k]<M(\bar{y}_{RAT})$ if

$$\bar{Y}^2\left\{\theta_{r,n}(C_X-\rho C_Y)^2+\theta_{n,N}\rho^2C_Y^2\right\}>0 \quad (29)$$

which always holds, implying that the proposed strategy under optimality condition is always preferable over $(D,\bar{y}_{RAT})$.

6.4 Comparison of (D,T_k) with $(D,\bar{y}_{COMP})$

Before comparing $(D,\bar{y}_{COMP})$ with (D,T_k), let us find the $M_{min}[\bar{y}_{COMP}]$. It is obtained as

$$M_{min}[\bar{y}_{COMP}]=\theta_{r,N}\bar{Y}^2C_Y^2-\theta_{r,n}\bar{Y}^2C_Y^2\rho^2 \quad (30)$$

for $\alpha=1-V$. Now comparing $(D,\bar{y}_{COMP})$ with (D,T_k) under optimality conditions, we see that $M_{min}[T_k]<M_{min}[\bar{y}_{COMP}]$ if

$$\theta_{n,N}\bar{Y}^2C_Y^2\rho^2>0 \quad (31)$$

which is always true. Thus (D,T_k) gives more efficient strategies as compared to compromised imputed strategy $(D,\bar{y}_{COMP})$ proposed by Singh and Horn [14] under the corresponding optimality conditions.

7 Empirical Study :

We now present the comparisons of different strategies on the basis of four data sets:

Population I : (Murthy, [6])

The data gives the number of absentees (Y) and number of workers (X) for the 43 factories. For the data, we get $\bar{Y}$= 9.651, $\bar{X}$=79.465, S_Y^2 = 43.137,S_X^2 = 1330.255, ρ=0.661. We take n=20.

Population II : (Mukhopadhyay, [5])

Here the data represents the quantity of raw materials required (in lakhs of bales) (Y) and number of labourers (in thousands) (X) in 20 jute mills. The following set of values were obtained $\bar{Y}$= 41.50, $\bar{X}$=441.95, S_Y^2 =95.737, S_X^2 =10215.21, ρ=0.6521. We take n=7.

Population III : (Singh and Chaudhary,[13])

The population is related to the area under wheat in the region during 1974, (Y) and during (1973), (X) in 34 villages. The following population values were obtained:

$\bar{Y}$= 199.441, $\bar{X}$=208.882, S_Y^2 = 22564.557, S_X^2 = 22652.046, ρ=0.9801. We take n=18.

Population IV: (Shukla *et al*,[11])

An artificial population of size 200 containing values of main variable Y and auxiliary variable X is given with the values: $\bar{Y}$= 42.485, $\bar{X}$=18.515, S_Y^2 = 199.060, S_X^2 = 48.538, ρ=0.8652. We take $n = 30$.

Since in a sample of any size, the number of respondents may vary from 0 to n (and accordingly the value of s, $0 \leq s \leq 0.5$), it is not out of place to study the behaviour of MSE of T_K under optimality conditions for each of the populations. Table 1 presents the optimum values of the parameter along with minimum MSE and bias for some selected values of r in each population.

The table reveals the following facts: (i) For each of the populations, there exist two positive real roots of the equation (27), (ii) with increasing number of respondents in the sample, the MSE decreases drastically and (iii) an appropriate choice of optimum k can be made under the criterion given in Remark 4.

Table:2 presents a comparison of different imputation strategies:$(D,\bar{y}_r)$, $(D,\bar{y}_{RAT})$, $(D,\bar{y}_{COMP})$ with the proposed strategy (D,T_k) under the optimality conditions, in terms of MSE,taking r=6, 5, 10 and 22 for populations I, II, III and IV respectively.

8 Conclusion :

The work presented a compromised imputation strategy under the scheme D and corresponding point estimator, utilizing the information on an auxiliary variable on the basis of FTE. The salient features of the strategy have some extra advantage over other existing estimators. On the basis of populations of different structures, a comparative study for the efficiency of the proposed strategy with some existing strategies showed that it is always preferable over other estimators.

Table 1: Optimum k and corresponding MSE for different values of r

Population	Value of $r(s)$	k_{opt}	$\lvert Bias\rvert$	$MinimumMSE$
I, N=43,n=20	2(0.09)	3.1364	0.09957	11.585
		5.2175	0.28508	
	4(0.17)	3.1125	0.03968	5.510
		5.3031	0.13042	
	6(0.23)	3.0902	0.02082	3.485
		5.3879	0.07951	
II, N=20,n=7	1(0.13)	3.0978	0.5298	52.287
		4.8455	0.1277	
	3(0.3)	3.0345	0.14251	15.595
		5.0106	0.04505	
	5(0.42)	2.9782	0.0694	8.256
		5.1726	0.02684	
III, N=34,n=18	2(0.1)	3.1364	4.2433	418.327
		5.2175	17.946	
	4(0.18)	3.1125	1.579	196.091
		5.3031	8.0802	
	6(0.25)	3.0902	0.7543	122.012
		5.3879	4.8456	
	10(0.36)	3.0949	0.1955	62.749
		5.7358	2.3292	
IV, N=200,n=30	4(0.12)	2.7224	0.0487	12.263
		4.7733	0.00927	
	12(0.29)	2.6238	0.00212	3.921
		4.9296	0.00706	
	20(0.4)	2.5338	0.0046	2.252
		5.084	0.00615	
	22(0.42)	2.5132	0.0052	2.025
		5.1223	0.00597	

Table 2: Minimum MSE, optimum values of the parameters for the estimators $\bar{y}_r$, $\bar{y}_{RAT}$, $\bar{y}_{COMP}$ and T_k

Estimator	Population	Opt. values of parameter and min MSE	R.E.
$\bar{y}_r$	I	6.186	100
	II	14.3605	100
	III	1592.792	100
	IV	8.053	100
$\bar{y}_{RAT}$	I	3.989	155.07
	II	12.586	114.09
	III	629.992	252.82
	IV	6.419	125.45
$\bar{y}_{COMP}$	I	3.988	155.11
	II	12.034	119.33
	III	629.440	253.04
	IV	6.247	128.90
T_k	I	3.485	177.50
	II	8.252	174.02
	III	62.762	2537.82
	IV	2.025	397.67

Acknowledgement

The authors are grateful to the referees for referring the paper and for providing useful suggestions.

Appendix

We have

$$T_k = \bar{y}_r\left[k+(1-k)\psi\{k,\bar{x}_r,\bar{x}_s\}\right] \tag{32}$$

where

$$\psi\{k,\bar{x}_r,\bar{x}_s\} = \frac{\eta\{t_1(k)\}}{\eta\{t_2(k)\}} \tag{33}$$

and

$$\eta\{t_i(k)\} = t_i(k)+\{1-t_i(k)\}\frac{\bar{x}_r}{\bar{x}_s}; i=1,2. \tag{34}$$

Re-writing $\eta\{t_i(k)\}$ in terms of A, B and C and substituting in (32), T_k becomes:

$$T_k = \bar{y}_r\left[k+(1-k)\left\{\frac{(A+C)\bar{x}_r+fB\bar{x}_s}{(A+fB)\bar{x}_r+C\bar{x}_s}\right\}\right] \tag{35}$$

Now using the large sample approximations
$\bar{y}_r = \bar{Y}(1+e_0)$, $\bar{x}_r = \bar{Y}(1+e_1)$
and the concept of two-phase sampling following Rao and Sitter [7] under the mechanism of missing completely at random (MCAR) for given r and n, we have:

$$E(e_0)=E(e_1)=0, E(e_0^2)=\theta_{r,N}C_Y^2, E(e_1^2)=\theta_{r,N}C_X^2, E(e_0e_1)=\theta_{r,N}\rho C_Y C_X \tag{36}$$

Converting the expressions (35) in terms of e_0 and e_1 and letting $d_1=\dfrac{A+fsB+C}{A+fB+C}$ and $d_2=\dfrac{A+fB+Cs}{A+fB+C}$, we can write T_k, retaining only up to the second power of e_0 and e_1 as

$$T_k = \bar{Y}\left[1+e_0+(1-k)(d_1-d_2)\left(e_1+e_0e_1-d_2e_1^2\right)\right]. \tag{37}$$

The expressions (37) obtained assuming that $|d_2e_1|<1$. Since for any choice of k, $|d_2|$ is always less than 1 and $|e_1|<1$, hence $|d_2e_1|<1$ is a valid assumption. Taking expectation of both the sides of (36) and realising that $B[T_k]=E[T_k]-\bar{Y}$, we have the expression (16), using the results (36). Similarly, under the large sample approximations,

$$M[T_k] = E\left[T_k-\bar{Y}\right]^2 = \bar{Y}^2E\left[e_0+(1-k)(d_1-d_2)e_1\right]^2 \tag{38}$$

which when solved, yields the expression (17).

References

[1] Kalton, G., Kasprzyk, D. and Santos, R.(1981): Issues of Non-Response and Imputation in the Survey of Income and Programme Participation, in *Current Topics in Survey Sampling*(eds.D.Krewski,R.Platek and J.N.K.Rao), Academic Press, New York, $455-480$

[2] Lee, H., Rancourt, E. and Sarndal, C. E. (1994): Experiments with variance estimation from survey data with imputed values, *Journal of Official Statistics*, $10(3)$, $231-243$.

[3] Lee, H., Rancourt, E. and Sarndal, C.E.(1995): Variance estimation in the presence of imputed data for the generalised estimation system, *Proceedings of the American Statistical Association,Social Survey Research Methods Section*, $384-389$.

[4] Little, R. J. A.and Rubin, D.B.(2002): Statistical Analysis with Missing Data. Wiley Series in Probability and Statistics, Hoboken, New Jersey.

[5] Mukhopadhyay, P. (2000): Theory and Methods of Survey Sampling. Prentice Hall of India Pvt. Ltd., New Delhi.

[6] Murthy, M. N. (1967): Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

[7] Rao, J.N.K.and Sitter, R.R.(1995): Variance estimation under two phase sampling with application to imputation for missing data. *Biometrika* ,82,$453-460$

[8] Sahai, A. and Ray, S.K.(1980): An efficient estimator using auxiliary information,*Metrika*, 27, $271-275$.

[9] Sande, I. G. (1979): A personal view of hot deck imputation procedures, Survey Methodology, 5, $238-246$.

[10] Shukla, D. (1988): On some Factor - Type Estimators for Population parameters in Sample Surveys, Unpublished Ph.D Thesis submitted to Banaras Hindu University, Varanasi, India .

[11] Shukla, D., Singhai, R and Thakur, N.S.(2011): A new imputation method for missing attribute values in data mining, *Journal of Applied Computer Science and Mathematics* 10(5), 14 – 19.

[12] Shukla,D., Singh, V. K .and Singh, G. N. (1991): On the use of transformation in factor-type estimator, *Metron*, 49(1-4),359-361.

[13] Singh, D and Chaudhari, F. S.(1986): Theory and Analysis of Sample Survey Designs, Wiley Eastern Limited, New Delhi.

[14] Singh, S and Horn, S. (2000): Compromised imputation in survey sampling, *Metrika*, 51,267 – 276.

[15] Singh, A. K, Singh, P. and Singh, V. K.(2014): Exponential-Type Compromised Imputation in Survey Sampling,*Journal of the Statistics Applications and Probability*,3(2),211-217.

[16] Singh, G.N., Kumari P., Kim, J. M. and Singh, S. (2010): Estimation of population mean using imputation techniques in sample surveys, *Journal of the Korean Statistical Society*, 39,67 – 74.

[17] Singh, V. K. and Shukla, D. (1987): One Parameter family of factor - type ratio estimator, *Metron*, 45(1-2), 273-283

[18] Singh, V. K. and Shukla, D. (1993): An efficient one parameter family of factor- type estimator in sample survey, *Metron*, 51(1-2), 139 – 159

[19] Singh, V. K. Singh, G.N. and Shukla, D.(1994): A class of chain ratio estimator with two auxiliary variables under double sampling scheme,*Sankhya Series B*, 46(2)209 – 221.

[20] Thompson, S. K. (2002): Sampling, Wiley, New York, 2nd Edition.