STATISTICS IN TRANSITION new series, September 2017
Vol. 18, No. 3, pp. 375–392, DOI: 10.21307/stattrans-2016-076

POPULATION VARIANCE ESTIMATION USING FACTOR TYPE IMPUTATION METHOD

RANJITA PANDEY[1], KALPANA YADAV[2]

ABSTRACT

We propose a variance estimator based on factor type imputation in the presence of non-response. Properties of the proposed classes of estimators are studied and their optimality conditions are derived. The proposed classes of factor type ratio estimators are shown to be more efficient than some of the existing estimators, namely, the usual unbiased estimator of variance, ratio-type, dual to ratio type and ratio cum dual to ratio estimators. Their performances are assessed on the basis of relative efficiencies. Findings are illustrated based on a simulated and real data set.

Key words: auxiliary information, mean squared error, simple random sampling without replacement (SRSWOR).

1. Introduction

Estimation of population variance is of significant importance in the theory of estimation. Efficient variance estimation under auxiliary information has been widely discussed by various authors such as Das and Tripathi (1978), Srivenkatramana (1980), Isaki (1983), Singh et al. (1988), Singh and Katyar (1991), Rao and Shao (1992), Sarndal (1992), Agrawal and Sthapit (1995), Rao and Sitter (1995), Garcia and Cebrain (1996), Arcos et al. (2005), Kadilar and Cingi (2006, 2006a), Solanki and Singh (2013) and Yadav & Kadilar (2013).

A common aspect of data collection is the inability to record all items under a response variable. Amputing incomplete observations from the collected or available data and proceeding with statistical analysis of the restricted complete data set is the most common and convenient approach of handling missing data. However, the process of replacing missing items with plausible values called imputation is popular among data analysts as it enables construction of standard programs based on some probability sampling models, for substituting missing data with a point estimate. Such models have potential to reduce bias and improve

[1] Department of Statistics, University of Delhi, New Delhi. E-mail: ranjitapandey111@gmail.com.
[2] Department of Statistics, University of Delhi, New Delhi. E-mail: kalpana22yadav@gmail.com.

precision to a significant extent in comparison with the amputation approach. Rubin (1976), Fay (1991) and Rao (1996) have reviewed various imputation techniques.

Large sample surveys are mostly accompanied either by *unit* non-response, where a sampled subject refuses/is unable to provide information for some variables, or *item* non-response, where several units on the study variable are missing. Variance estimation after imputation has been studied by Kim et al. (2001), Raghunath and Singh (2006), Beaumont at al. (2011) and Singh and Solanki (2009-2010) using auxiliary information in the presence of random non-response. In the present paper, an improved factor type (FT) estimator of population variance based on an auxiliary variable is proposed, under non-response. Our work is motivated by the theoretical properties of FT estimator introduced by Singh and Shukla (1987).

2. Notations and estimators in literature

Let $\Omega = \{1, 2, ..., N\}$ be a finite population of N identifiable units. Let $(y_i, x_i),\ i = 1,2,3,..., N$ be the observed value of study variable and auxiliary variable for i^{th} individual from a finite population Ω. From a finite population of *N* identifiable units, a simple random units sample, *s*, of size *n* is drawn without replacement. *r* denotes the number of responding units in the sample *s*. The remaining *(n-r)* units are non-responding units.

The following notations for the population are defined for study and auxiliary variables respectively: $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$:

Population mean of the variables *Y* and *X*; $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ represent sample means of the study variable *y* and the auxiliary variable *x* respectively; $S_{y(N)}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2, S_{x(N)}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$: population variances of variables *Y* and *X*; $s_{y(n)}^2 = \frac{1}{n-1}\sum_{i=1}^{N}(y_i - \bar{y})^2, s_{x(n)}^2 = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$: sample variances of corresponding to variables *Y* and *X*; $s_{y(r)}^2 = \frac{1}{n-1}\sum_{i=1}^{N}(y_i - \bar{y})^2, s_{x(r)}^2 = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 =$ sample variances of responding units in the respective samples; $\mu_{ts} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^t (x_i - P)^s$, $\lambda_{ts} = \frac{\mu_{ts}}{\mu_{t0}^{t/2}\mu_{0s}^{s/2}}$, and the sampling fraction $f = \frac{n}{N}$. *B*(.) represents bias and *M* (.) represents mean squared error of the respective estimators.

To obtain the bias and M.S.E. of existing and suggested estimators we additionally consider
$s^2_{y(r)} = S^2_{y(N)}(1+e_0); s^2_{x(r)} = S^2_{x(N)}(1+e_1)$; $s^2_{x(n)} = S^2_{x(N)}(1+e_2)$ and $|e_i|<1; (i=0,1,2)$.
such that $E(e_0)=E(e_1)=E(e_2)=0$; $E(e_0^2)=M_1(\lambda_{40}-1)$, $E(e_1^2)=M_1(\lambda_{04}-1)$,
$E(e_2^2)=M_2(\lambda_{04}-1)$;
$E(e_0e_1)=M_1(\lambda_{22}-1), E(e_1e_2)=M_2(\lambda_{04}-1), E(e_0e_2)=M_2(\lambda_{22}-1)$ where

$$M_1 = \frac{1}{r}-\frac{1}{N},\ M_2 = \frac{1}{n}-\frac{1}{N},\ M_3 = M_1 - M_2 = \frac{1}{r}-\frac{1}{n}.$$

The variance of the usual unbiased variance estimator $S^2_{y(N)}$ is given by:

$$V\left(S^2_{y(N)}\right) = M_1 S^4_{y(N)}(\lambda_{40}-1) \tag{1}$$

Isaki (1982) (hereafter *IK*) discussed a *ratio type* variance estimator for estimating population variance and its properties. Under non-response we write

$$t_{IK} = s^2_{y(r)} \frac{S^2_{x(N)}}{s^2_{x(r)}} \tag{2}$$

The estimator t_{IK} is found to be biased and its M.S.E. is given by:

$$M(t_{IK}) = M_1 S^4_{y(N)}\left[(\lambda_{04}-1)+(\lambda_{40}-1)-2(\lambda_{22}-1)\right] \tag{3}$$

Srivenkataramana and Tracy (1980) (hereafter *SV*) have given a *dual to ratio* estimator for variance estimator in sample surveys. Under non-response it can be modified as:

$$t_{SV} = s^2_{y(r)}\left[\frac{S^2_{x(N)} - f s^2_{x(r)}}{(1-f)S^2_{x(N)}}\right] \tag{4}$$

The M.S.E. of t_{SV} is given by:

$$M(t_{SV}) = M_1 S^4_y\left[(\lambda_{40}-1)+g^2(\lambda_{04}-1)-2g(\lambda_{22}-1)\right] \text{where } g = \frac{f}{(1-f)} \tag{5}$$

Yadav and Kadilar (2013) (hereafter *YK*) proposed the *ratio-cum-dual to ratio* type estimator for the population variance of the study variable. The ratio-cum-dual type variance estimator under non-response is given by:

$$t_{YK} = s^2_{y(r)}\left[\alpha \frac{S^2_{x(N)}}{s^2_{x(r)}} + (1-\alpha)\frac{S^2_{x(N)} - f s^2_{x(r)}}{(1-f)S^2_{x(N)}}\right] \tag{6}$$

The M.S.E. of t_{YK} is given by:

$$M\left(t_{YK}\right)=M_1\, S_{y(N)}^4\left[\left(\lambda_{40}-1\right)+\alpha_1^2\left(\lambda_{04}-1\right)-2\alpha_1\left(\lambda_{22}-1\right)\right] \tag{7}$$

The M.S.E. of the proposed estimator is minimized for the optimum value α as

$$\alpha=\frac{K-g}{1-g} \text{ such that } K=\frac{\lambda_{22}-1}{\lambda_{04}-1}$$

$$M\left(t_{YK}\right)_{\min}=M_1 S_{y(N)}^4\left[\lambda_{40}-1-\frac{\left(\lambda_{22}-1\right)^2}{\lambda_{04}-1}\right] \tag{8}$$

3. Proposed estimators and their properties

Singh and Shukla (1987) proposed a family of FT ratio estimator of *population mean* for *complete sample* case. Unbiased, ratio, product and dual to ratio estimators are its special cases. An advantage of one-parameter class of estimators is that it requires only knowledge of the quantity $\rho\frac{C_y}{C_x}$ for making the best selection of the parameter. Population correlation coefficient between variables *Y* and *X* is represented by ρ and the respective coefficient of variation by C_y and C_x. The value of function $\rho\frac{C_y}{C_x}$ does not fluctuate considerably in repeated surveys and therefore could be guessed accurately from previous data or past experience or a pilot survey or otherwise [(Murthy (1967); Reddy, (1978)]. The *proposed variance estimator* is constructed as a function of some factors of the parameter termed as Factor-Type (F-T) estimator. This process of factorization makes it possible to yield more than one optimum value of the parameter so that at the same time bias of the estimator can also be controlled. The new class of FT ratio estimator for *population variance* of the study variable *under non-response* is proposed as:

$$t_{SSi}=s_{y(r)}^2\phi_i(k)\,;i=1,2,3 \tag{9}$$

where $\phi_1(k)=\frac{(A+C)S_{x(N)}^2+fBs_{x(n)}^2}{(A+fB)S_{x(N)}^2+Cs_{x(n)}^2}$; $\phi_2(k)=\frac{(A+C)s_{x(n)}^2+fBs_{x(r)}^2}{(A+fB)s_{x(n)}^2+Cs_{x(r)}^2}$ and

$$\phi_3(k)=\frac{(A+C)S_{x(N)}^2+fBs_{x(r)}^2}{(A+fB)S_{x(N)}^2+Cs_{x(r)}^2}.$$

where, $A=(k-1)(k-2)$, $B=(k-1)(k-4)$, $C=(k-2)(k-3)(k-4)$, $0\le k\le\infty$.

Assume, $\theta_1=\frac{fB}{A+fB+C},\theta_2=\frac{C}{A+fB+C}$ and $\theta=\theta_1-\theta_2$

The properties of the proposed family of estimators are presented through the following theorems:

Theorem 1:

(i) The estimator t_{SS1} for population variance could be written in terms of $e_i ; i = 0,1,2$ as

$$t_{SS1} = S^2_{y(N)}\left[1 + \theta e_2 - \theta\theta_2 e_2^2 + \theta e_0 e_2 + e_0\right] \tag{10}$$

with $B(t_{SS1}) = S^2_{y(N)} M_2 \theta\left[(\lambda_{22} - 1) - \theta_2(\lambda_{04} - 1)\right]$ (11)

and M.S.E. $M(t_{SS1}) = S^4_{y(N)}\left[M_1(\lambda_{40} - 1) + 2\theta M_2(\lambda_{22} - 1) + \theta^2 M_2(\lambda_{04} - 1)\right]$ (12)

The corresponding minimum M.S.E. at $\theta = \frac{-(\lambda_{22} - 1)}{(\lambda_{04} - 1)} = -P$ is given by

$$\left[M(t_{SS1})\right]_{\min} = S^4_{y(N)}\left[\frac{M_1(\lambda_{40} - 1)(\lambda_{04} - 1) - M_2(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{13}$$

(ii) The estimator t_{SS2} in terms of $e_i ; i = 0,1,2$ is

$$t_{SS2} = S^2_{y(N)}\left[1 + e_0 - \theta e_2 + \theta e_1 - \theta\theta_2 e_1^2 + \theta\theta_1 e_2^2 - \theta e_0 e_2 + \theta e_0 e_1 - \theta^2 e_1 e_2\right] \tag{14}$$

with $B(t_{SS2}) = S^2_{y(N)}(M_1 - M_2)\theta\left[(\lambda_{22} - 1) - \theta_2(\lambda_{04} - 1)\right]$ (15)

and $M(t_{SS2}) = S^4_{y(N)}\left[M_1(\lambda_{40} - 1) + \theta M_3\left[2(\lambda_{22} - 1) + \theta(\lambda_{04} - 1)\right]\right]$ (16)

The minimum mean squared error of t_{SS2} at $\theta = \frac{-(\lambda_{22} - 1)}{(\lambda_{04} - 1)} = -P$ is given by

$$\left[M(t_{SS2})\right]_{\min} = S^4_{y(N)}\left[\frac{M_1(\lambda_{40} - 1)(\lambda_{04} - 1) - M_3(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{17}$$

(iii) The estimator t_{SS3} for population variance could be written in terms of

$e_i ; i = 0,1,2$ as $t_{SS3} = S^2_{y(N)}\left[1 + e_0 + \theta e_1 - \theta\theta_2 e_1^2 + \theta e_0 e_1\right]$ (18)

with $B(t_{SS3}) = S^2_{y(N)} M_1 \theta\left[(\lambda_{22} - 1) - \theta_2(\lambda_{04} - 1)\right]$ (19)

and $M(t_{SS3}) = S^4_{y(N)} M_1\left[(\lambda_{40} - 1) + 2\theta(\lambda_{22} - 1) + \theta^2(\lambda_{04} - 1)\right]$ (20)

The minimum M.S.E. of t_{SS3} at $\theta = \frac{-(\lambda_{22} - 1)}{(\lambda_{04} - 1)} = -P$ is given by

$$[M(t_{SS3})]_{\min} = S_{y(N)}^4 M_1 \left[\frac{(\lambda_{40}-1)(\lambda_{04}-1)-(\lambda_{22}-1)^2}{(\lambda_{04}-1)} \right] \tag{21}$$

Proof: $t_{SS3} = s_{y(r)}^2 \phi_i(k); i = 1,2,3.$

Substituting the value of $\phi_i(k); i = 1,2,3$ and using the concept of large sample approximation, we get

$$t_{SS1} = S_{y(N)}^2 (1+e_0) \left[\frac{A+C+f\,B+f\,Be_2}{A+C+f\,B+C\,e_2} \right]$$

$$= S_{y(N)}^2 (1+e_0)(1+\theta_1 e_2)(1+\theta_2 e_2)^{-1}$$

$$t_{SS2} = S_{y(N)}^2 (1+e_0) \left[\frac{A+f\,B+C+C\,e_2+f\,Be_1}{A+f\,B+C+C\,e_1+f\,Be_2} \right] = S_{y(N)}^2 (1+e_0)(1+\theta_1 e_1+\theta_2 e_2)(1+\theta_1 e_2+\theta_2 e_1)^{-1}$$

$$t_{SS3} = S_{y(N)}^2 (1+e_0) \left[\frac{A+C+f\,B+f\,Be_1}{A+C+f\,B+C\,e_1} \right] = S_{y(N)}^2 (1+e_0)(1+\theta_1 e_1)(1+\theta_2 e_1)^{-1}$$

Using Taylor's expansion and ignoring terms of $o(n^{-1})$ and higher order leads to equations (10), (14) and (18).

Since we know that $B(t_{SSi}) = E(t_{SSi} - S_{y(N)}^2); i = 1, 2, 3.$

Therefore, $B(t_{SS1}) = S_{y(N)}^2 E[e_0 + \theta e_2 - \theta\theta_2 e_2^2 + \theta e_0 e_2]$

$$B(t_{SS2}) = S_{y(N)}^2 E[e_0 - \theta e_2 + \theta e_1 - \theta\theta_2 e_1^2 + \theta\theta_1 e_2^2 - \theta e_0 e_2 + \theta e_0 e_1 - \theta^2 e_1 e_2]$$

and $B(t_{SS3}) = S_{y(N)}^2 E[e_0 + \theta e_1 - \theta\theta_2 e_1^2 + \theta e_0 e_1]$

Substituting the values of $e_i; i = 0,1,2$ using section 2, and simplifying, equations (11), (15) and (19) are obtained.

Also, $M(t_{SSi}) = E(t_{SSi} - S_{y(N)}^2)^2; i = 1,2,3.$

Substituting the values of estimators and solving it, and ignoring higher order terms, we get

$$M(t_{SS1}) = S_{y(N)}^4 E[e_0^2 + 2\theta e_0 e_2 + \theta^2 e_2^2]$$

$$M(t_{SS2}) = S_{y(N)}^4 E[e_0^2 - 2\theta e_0 e_2 + 2\theta e_0 e_1 + \theta^2 e_2^2 - 2\theta^2 e_1 e_2 + \theta^2 e_1^2]$$

$$M(t_{SS3}) = S_{y(N)}^4 E[e_0^2 + 2\theta e_0 e_1 + \theta^2 e_1^2]$$

Substituting the expectations values of e_0 , e_1 and e_2 and solving it, leads to equations (12), (16) and (20).

Now, differentiating these expressions with respect to P and then equating to zero yields $\frac{d\,M(t_{SSi})}{dP} = 0 \Rightarrow P = \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$

Substituting the value of P in equation (12), (16) and (20), corresponding expressions for the minimum *M.S.E.s* are obtained.

Remark 1: Multiple choices of *k*:

The optimality condition $\theta = \theta_1 - \theta_2 = -P$ provides the equation

$$(P-1)k^3 + [P(f-8) + (f+9)]k^2 + [P(23-5f) - (26+5f)]k + [(4f-22)P + (4f+24)] = 0, \tag{22}$$

which is a cubic equation in k. Its roots are represented by k_1, k_2, k_3 (say), for which mean squared error is optimum. The best choice criterion for k, which controls the quantum of bias in the corresponding estimator, is outlined in the following algorithm:

Step I: Compute $\left|B(t_{SSi})_{k_j}\right| \qquad for \;\; i, j = 1,2,3$

Step II: For given *i*, choose k_j as $\left|B(t_{SSi})_{k_j}\right| = \min_{j=1,2,3}\left[\left|B(t_{SSi})\right|_{k_j}\right]$.

Remark 2: Factor-type ratio estimator (Singh and Shukla (1987)) for population variance of the study variable (without imputation) is defined as:

$$t_{SS} = \frac{(A+C)S_x^2 + fBs_x^2}{(A+fB)S_x^2 + Cs_x^2} \tag{23}$$

M.S.E. and minimum M.S.E. of estimator t_{SS} at $\theta = \frac{-(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$ are shown below:

$$M(t_{SS}) = S_y^4 M_1\left[(\lambda_{40} - 1) + \theta^2(\lambda_{04} - 1) + 2\theta(\lambda_{22} - 1)\right] \tag{24}$$

$$[M(t_{SS})]_{\min} = S_y^4 M_1\left[\frac{(\lambda_{40} - 1)(\lambda_{04} - 1) - (\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{25}$$

4. Comparisons

On pair-wise comparison of expressions for *M.S.E.s* (from section 2 and section 3) (i) among the proposed estimators (ii) between the proposed and some of the existing estimators, we obtain theoretical conditions of superiority, which are shown in Table 1 and Table 2.

Table 1. Comparison within Proposed estimators

Estimators (Minimum M.S.E.)	More efficient than (Minimum M.S.E.)	Condition
t_{SS2}	t_{SS1}	$\lambda_{22} > 1$
t_{SS3}		
t_{SS3}	t_{SS2}	

Table 2. Comparison within Proposed estimators and Traditional estimators

Estimators (Minimum M.S.E.)	More efficient than (Minimum M.S.E.)	Condition
t_{SS1}	$S^2_{y(N)}$	$\lambda_{22} > 1$
t_{SS2}		
t_{SS3}		
t_{SS1}	t_{IK}	$(\lambda_{22} - 1)^2 > \frac{M_1}{M_2} A$
t_{SS2}		$(\lambda_{22} - 1)^2 > \frac{M_1}{M_3} A$
t_{SS3}		$(\lambda_{22} - 1)^2 > A$
t_{SS1}	t_{SV}	$(\lambda_{22} - 1)^2 > \frac{M_1}{M_2} B$
t_{SS2}		$(\lambda_{22} - 1)^2 > \frac{M_1}{M_3} B$
t_{SS3}		$(\lambda_{22} - 1)^2 > B$
t_{SS1}	t_{YK}	$\lambda_{22} > 1$
t_{SS2}		$\lambda_{22} > 1$
t_{SS3}		is equal to

where $A = (\lambda_{04} - 1)[2(\lambda_{22} - 1) - (\lambda_{04} - 1)]$; $B = (\lambda_{04} - 1)g[2(\lambda_{22} - 1) - g(\lambda_{04} - 1)]$.

5. Simulation study

An artificial population [Source: Shukla and Thakur (2008)] of size N = 200 containing values of main variable Y and auxiliary variable X.

Parameters of the population are given as below:

$\bar{Y} = 42.485$; $\bar{X} = 18.515$; $S_y^2 = 199.0598$; $S_x^2 = 48.5375$; $\rho = 0.8652$;

$f = 0.3$; $\lambda_{22} = 2.47$; $\lambda_{04} = 3.74$; $\lambda_{40} = 2.56$, n = 60, r = 50

For the above data set, equation (22) provides three k-values: $k_1 =$ 1.54; $k_2 = 2.94$; $k_3 = 6.67$

The simulation process comprises the following steps:

Step 1: Draw a random sample of size n = 60 from the population of $N =$ 200 by SRSWOR.

Step 2: Discard 10 randomly chosen units from each sample corresponding to Y.

Step 3: Impute these discarded units of Y by the proposed methods and the available methods separately. Compute the value of different estimators and also for the proposed estimators.

Step 4: Repeat the above steps 30,000 times, which provides multiple sample-based estimates $\hat{t}_{s^2(1)}, \hat{t}_{s^2(2)} .... \hat{t}_{s^2(30,000)}$.

Step 5: Bias of $\hat{t}_1$ is obtained by

$$B\left(\hat{t}_{s^2}\right) = \frac{1}{30,000} \sum_{i=1}^{30,000} \left[s_{y(r)}^2 - S_{y(N)}^2\right] .$$

Step 6: Mean squared error of $\hat{y}$ is computed by

$$M\left(\hat{t}_{s^2}\right) = \frac{1}{30,000} \sum_{i=1}^{30,000} \left[s_{y(r)}^2 - S_{y(N)}^2\right]^2 .$$

Step 7: Percentage Relative efficiency (*PRE*) is computed from equation (26) and shown in Table 5:

$$PRE\left(*, t_{SSi}\right)_j = \frac{M[*]}{M\left[t_{SSi\,;\,i=1,2,3}\right]} \times 100 : j = 1, 2, 3, 4; \tag{26}$$

such that $*$ represents different existing methods.

Bias and M.S.E.*s* of the existing and proposed estimators computed from 30,000 repeated samples drawn by SRSWOR from population $N =$ 200 are shown in Table 3.

Table 3. Bias, Mean Squared Error of Different Suggested and Traditional Estimators

Traditional Estimators	**Bias**	**M.S.E.**	**Suggested Estimators**		**Bias**	**M.S.E.**
$S^2_{y(N)}$	-35.82	2417.27	t_{SS1}	$k_1 = 1.55$	7.03	1572.64
t_{IK}	-40.51	1914.42		$k_2 = 2.94$	3.32	1800.42
t_{SV}	-45.04	2130.76		$k_3 = 6.67$	6.48	1602.47
t_{YK}	-43.86	1934.08	t_{SS2}	$k_1 = 1.55$	20.61	1067.79
				$k_2 = 2.94$	20.37	1047.37
				$k_3 = 6.67$	20.57	1064.34
			t_{SS3}	$k_1 = 1.55$	1.27	1262.77
				$k_2 = 2.94$	-3.79	1511.06
				$k_3 = 6.67$	0.52	1293.07

Computational results for efficiency loss due to imputation is measured as $(LI)_i = \frac{M(t_{SSi})}{M(t_{SS})}$ such that, $M(t_{SSi})$ and $M(t_{SS})$ are the M.S.E.s of the proposed estimators with and without imputation (from Remark 2). The losses are reported in Table 4.

Table 4. Loss due to Imputation

Optimum k	$k_1 = 1.55$	$k_2 = 2.94$	$k_3 = 6.67$
$(LI)_1$	0.74	0.72	0.75
$(LI)_2$	0.68	0.75	0.70
$(LI)_3$	0.75	0.72	0.75

Table 5. P.R.E. of suggested estimators with respect to different Traditional estimators

Estimators	**Optimum k values**	$PRE\left(S_{y(N)}^2, t_{SSi}\right)_1$	$PRE\left(t_{IK}, t_{SSi}\right)_2$	$PRE\left(t_{SV}, t_{SSi}\right)_3$	$PRE\left(t_{YK}, t_{SSi}\right)_4$
t_{SS1}	$k_1 = 1.55$	153.71	121.73	135.49	122.98
	$k_2 = 2.94$	134.26	106.33	118.35	107.42
	$k_3 = 6.67$	150.85	119.47	132.97	120.69
t_{SS2}	$k_1 = 1.55$	226.38	179.29	199.55	181.13
	$k_2 = 2.94$	**230.79**	182.78	203.44	184.66
	$k_3 = 6.67$	227.11	179.87	200.19	181.72
t_{SS3}	$k_1 = 1.55$	191.43	151.61	168.74	153.16
	$k_2 = 2.94$	159.97	126.69	141.01	127.99
	$k_3 = 6.67$	186.94	148.05	164.78	149.57

5.1. Values of k for Unbiased Estimator $t_{SSi}; i = 1,2,3.$

For unbiased estimator,

$$B\left(t_{SSi\ ;\ i=1,2,3}\right) = 0 \Rightarrow S_{y(N)}^2 \theta\left[\left(\lambda_{22} - 1\right) - \theta_2\left(\lambda_{04} - 1\right)\right] = 0$$

$$\Rightarrow \left(\theta_1 - \theta_2\right)\left[\left(\lambda_{22} - 1\right) - \theta_2\left(\lambda_{04} - 1\right)\right] = 0 \qquad (27)$$

Case 1: $\theta_1 - \theta_2 = 0 \Rightarrow \dfrac{fB - C}{A + fB + C} = 0 \Rightarrow fB - C = 0$

$$\Rightarrow f(k-1)(k-4) - (k-2)(k-3)(k-4) = 0$$

$$\Rightarrow (k-4)\left[f(k-1) - (k-2)(k-3)\right] = 0 \qquad (28)$$

From (27) either $(k-4) = 0 \Rightarrow k = k_1' = 4$ (29)

$$\text{or } k^2 - (f+5)k + (f+6) = 0$$

the remaining two roots of k are

$$k_2' = \frac{(f+5) + \sqrt{(f+5)^2 - 4(f+6)}}{2} \qquad (30)$$

$$k_3' = \frac{(f+5) - \sqrt{(f+5)^2 - 4(f+6)}}{2} \qquad (31)$$

On putting the value of f for the above data set, we get

$$k_2' = 3.5 \tag{32}$$

$$k_3' = 1.8 \tag{33}$$

Case 2: $$\left[(\lambda_{22}-1)-\theta_2(\lambda_{04}-1)\right]=0 \Rightarrow \theta_2=\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} \tag{34}$$

Since we know that $\theta_2=\dfrac{C}{A+fB+C}=0$. Then, on equating it with $\theta_2=\dfrac{(\lambda_{22}-1)}{(\lambda_{04}-1)}$, we get a cubic equation in the form of k as follows:

$$(\lambda_{22}-\lambda_{04})k^3-\left[9\lambda_{04}+f(\lambda_{22}-1)-8\lambda_{22}-1\right]k^2$$
$$+\left[23\lambda_{22}-26\lambda_{04}-5f(\lambda_{22}-1)+3\right]k+\left[24\lambda_{04}-22\lambda_{22}+4f(\lambda_{22}-1)-2\right]=0 \tag{35}$$

On putting the values of λ_{22}, λ_{04} and f we get three different values of k .

$$k_4' = 1.72,\ k_5' = 2.60,\ k_6' = 6.19 \tag{36}$$

6. Real data analysis

A real data of size *N* = 66 is taken from Indian Institute of Sugarcane Research, which comprises annual production data (in '000 tonnes) represented as the auxiliary variable *X* and the corresponding cultivation area (in '000 ha.) represented as the study variable *Y*, over the time period of *1950-51* to *2015-16*. Parameters of the above population are given as below:

$\bar{Y} =$ 22.30; $\bar{X} = 193558.80$; $S_y^2 = 2278933.68$; $S_x^2 = 8658527591$; $\rho = 0.9904$;

$f = 0.3$; $\lambda_{22} = 1.23$; $\lambda_{04} = 1.77$; $\lambda_{40} = 1.35$, n = 20, r = 10

For the above data set, equation (22) provides three k-values: $k_1 =$ 1.68, $k_2 = 3.09$ and $k_3 =$ 5.23. Initially we selected 10,000 independent random samples of size n =20 from the above population of size $N =$ 66 by SRSWOR.

The empirical bias and M.S.E.*s* of the existing and proposed estimators computed from these repeated samples are shown in Table 6.

Table 6. Bias, Mean Squared Error of Different Suggested and Traditional Estimators

Traditional Estimators	Bias	M.S.E.	Suggested Estimators		Bias	M.S.E.
$S^2_{y(N)}$	-1.03E+06	1.24E+12	t_{SS1}	$k_1 = 1.68$	-1.01E+06	1.03E+12
t_{IK}	-1.34E+06	1.12E+12		$k_2 = 3.09$	-1.02E+06	1.05E+12
t_{SV}	-1.04E+06	1.08E+12		$k_3 = 5.23$	-1.02E+06	1.05E+12
t_{YK}	-1.12E+06	1.27E+12	t_{SS2}	$k_1 = 1.68$	-1.03E+06	1.04E+12
				$k_2 = 3.09$	-1.03E+06	1.04E+12
				$k_3 = 5.23$	-1.02E+06	1.02E+12
			t_{SS3}	$k_1 = 1.68$	-1.01E+06	1.03E+12
				$k_2 = 3.09$	-1.02E+06	1.05E+12
				$k_3 = 5.23$	-1.01E+06	1.04E+12

Table 7. Loss due to Imputation

Optimum k	$k_1 = 1.68$	$k_2 = 3.09$	$k_3 = 5.23$
$(LI)_1$	0.70	0.72	0.75
$(LI)_2$	0.77	0.76	0.70
$(LI)_3$	0.73	0.77	0.75

Table 8. P.R.E. of suggested estimators with respect to different Traditional estimators

Estimators	Optimum k values	$PRE\left(S^2_{y(N)}, t_{SSi}\right)_1$	$PRE\left(t_{IK}, t_{SSi}\right)_2$	$PRE\left(t_{SV}, t_{SSi}\right)_3$	$PRE\left(t_{YK}, t_{SSi}\right)_4$
t_{SS1}	$k_1 = 1.55$	120.39	108.74	104.85	123.30
	$k_2 = 2.94$	118.10	106.67	102.86	120.95
	$k_3 = 6.67$	118.10	106.67	102.86	120.95
t_{SS2}	$k_1 = 1.55$	119.23	107.69	103.85	122.12
	$k_2 = 2.94$	119.23	107.69	103.85	122.12
	$k_3 = 6.67$	121.57	109.80	105.88	**124.51**
t_{SS3}	$k_1 = 1.55$	120.39	108.74	104.85	123.30
	$k_2 = 2.94$	118.10	106.67	102.86	120.95
	$k_3 = 6.67$	119.23	107.69	103.85	122.12

6.1 Values of k for Unbiased Estimator $t_{SSi}; i=1,2,3.$

For unbiased estimator, $B(t_{SSi\ ;\ i=1,2,3})=0 \Rightarrow S^2_{y(N)}\theta[(\lambda_{22}-1)-\theta_2(\lambda_{04}-1)]=0$

$$\Rightarrow(\theta_1-\theta_2)[(\lambda_{22}-1)-\theta_2(\lambda_{04}-1)]=0 \qquad \ldots(37)$$

Case 1: $\theta_1-\theta_2=0 \Rightarrow \frac{fB-C}{A+fB+C}=0 \Rightarrow fB-C=0$

$$\Rightarrow f(k-1)(k-4)-(k-2)(k-3)(k-4)=0$$

$$\Rightarrow (k-4)[f(k-1)-(k-2)(k-3)]=0 \qquad (38)$$

From (28) either $(k-4)=0 \Rightarrow k=k_1^{'}=4$ (39)

$$\text{or} \quad k^2-(f+5)k+(f+6)=0$$

the remaining two roots of *k* are

$$k_2^{'}=\frac{(f+5)+\sqrt{(f+5)^2-4(f+6)}}{2} \qquad (40)$$

$$k_3^{'}=\frac{(f+5)-\sqrt{(f+5)^2-4(f+6)}}{2} \qquad (41)$$

On putting the value of f for the above data set, we get

$$k_2^{'}=3.5 \qquad (42)$$

$$k_3^{'}=1.8 \qquad (43)$$

Case 2: $[(\lambda_{22}-1)-\theta_2(\lambda_{04}-1)]=0 \Rightarrow \theta_2=\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)}$ (44)

Since we know that $\theta_2=\frac{C}{A+fB+C}=0$. Then, on equating it with $\theta_2=\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)}$, we get a cubic equation in the form of k as follows:

$$(\lambda_{22}-\lambda_{04})k^3-[9\lambda_{04}+f(\lambda_{22}-1)-8\lambda_{22}-1]k^2$$
$$+[23\lambda_{22}-26\lambda_{04}-5f(\lambda_{22}-1)+3]k+[24\lambda_{04}-22\lambda_{22}+4f(\lambda_{22}-1)-2]=0 \qquad (45)$$

On putting the values of λ_{22}, λ_{04} and *f* we get two different values of k .

$$k_4^{'}=0.72 \text{ and } k_5^{'}=7.53 \qquad (46)$$

7. Conclusion

The present paper suggests three new FT variance estimators under item non-response on the study variable, in a bivariate sample data. FT estimator, a generalized class of estimators for ratio, product, dual to ratio and the usual unbiased estimator are found to be more efficient than some existing estimators. The FT variance estimator maintains an optimum balance between reduction of bias and that of reducing M.S.E through *k*. We can choose k values for different pair of (f, P) values. Thus, the FT variance estimator could be made almost unbiased by an appropriate choice of multiple available k values.

Table 5 and Table 8 show P.R.E. of the suggested estimators with respect to different traditional estimators based on simulated and real data. It is observed from these tables that the proposed FT estimators prove to be better than the usual unbiased, ratio, dual to ratio and ratio cum dual to ratio estimators, under non-response. The proposed estimator t_{SS2} performs best among the three proposed estimators from the point of view of increasing efficiency. The three proposed FT type estimators are the best estimators in the sense of having the largest *PRE* among all the prevalent estimators discussed here.

Acknowledgments

Part of the work of the first author is supported by R&D Grant from University of Delhi.

REFERENCES

AGRAWAL, M. C., STHAPIT, A. B., (1995). Unbiased Ratio-Type Variance Estimation, Statistics and Probability Letters, 25, pp. 361–364.

ARCOS, A., RUEDA, M., MARTINEZ, M. D., GONZALEZ, S., ROMAN, Y., (2005). Incorporating the auxiliary information available in variance estimation, Applied Mathematics and Computation, 160, pp. 387–399.

BEAUMONT, J. F., HAZZIA, D., BOCCI, C., (2011). On variance estimation under auxiliary value imputation in sample surveys. Statistica Sinica, 21, pp. 515–537.

BEAUMONT, J. F., HAZIZA, D., BOCCI, C., (2011). On variance estimation under Auxiliary value imputation in sample surveys, Statistica Sinca, 21, pp. 515–537.

DAS, A. K., TRIPATHI, T. P., (1978). Use of auxiliary information in estimating the finite population variance, Sankhya, C, 40, pp. 139–148.

FAY, B. E., (1991). A design based pros, predictive on missing data variance. Proceedings of the annual Research Conference, U.S., Bureau of the census, pp. 429–440.

GARCIA, M. R., CEBRAIN, A. A., (1996). Repeated substitution method: The ratio estimator for the population variance, Metrika, 43, pp. 101–105.

ISAKI, C. T., (1983). Variance estimation using auxiliary information, Journal of the American Statistical Association, 78, pp. 117–123.

KADILAR, C., CINGI, H., (2006). Ratio estimators for the population variance in simple and stratified random sampling, Applied Mathematics and Computation, 173, pp. 1047–1059.

KADILAR, C., CINGI, H., (2006a). Improvement in Variance Estimation using Auxiliary Information, Hacettepe Journal of Mathematics and Statistics, 35 (1), pp. 111–115.

KIM, J. K., BRICK, J. M., FULLER, W. A., KALTON, G., (2001). On the bias of the multiple imputation variance estimator in survey sampling, Journal of the Royal Statistical Society series B, 68, pp. 509–521.

MURTHY, M. N., (1967). Sampling Theory and Methods. Statistical Publishing Society, Calcutta, India.

RAGHUNATH, A. AND SINGH, S. (2006). Estimation of variance from missing data, Metron-International Journal of Statistics, LXIV, 2, pp. 161–177.

RAO, SITTER, (1995). Variance estimation under two phase sampling with application to imputation for missing data. Biometrika, 82, pp. 453–460.

RAO, J.N.K. (1996). On variance estimation with imputed survey data. Journal of the American Statistical Association, 91 (434), pp. 499-506.

RAO, J. N. K., SHAO J., (1992). Jackknife variance estimation with survey data under hot deck imputation. Biometrika, 79 (4), 811–822.

REDDY, V. N., (1978). A Study on the use of Prior Knowledge on Certain Population Parameters in Estimation. Sankhya, C, 40, pp. 29–37.

RUBIN, D. B., (1976). Inference and missing data, Biometrika, 63(3), pp. 581–592.

SARANDAL, C. E., (1992). Methods for estimating the precision of survey estimates when imputation has been used, Survey Methodology, 18, pp. 241–252.

SHUKLA, D., THAKUR, N. S., (2008). Estimation of mean with imputation of missing data using factor-type estimator, Statistics in Transition, 9, 1, pp. 33–48.

SINGH, H. P., UPADHYAYA, L. N., NAMJOSHI, U. D., (1988). Estimation of finite population variance, Current Science, 57, pp. 1331–1334.

SINGH, H. P., KATYAR, N. P., (1991). Variance estimation through the mean square successive differences and sample variance using a priori information, Journal of the Indian Society of Agricultural Statistics, 43, 1, pp. 16–29.

SINGH, H. P., SOLANKI, R. S., (2009). Estimation of finite population variance using auxiliary information in presence of random non-response, Gujarat Statistical Review, 1, pp. 37–637.

SINGH, H. P., SOLANKI, R. S., (2010). Estimation of finite population variance using auxiliary information in presence of random non-response, Gujarat Statistical Review, 2, pp. 46–58.

SINGH, V. K., SHUKLA, D., (1987). One parameter family of factor-type ratio estimator, Metron, 45, 1-2, pp. 273–283.

SOLANKI, R. S., SINGH, H. P., (2013). An improved class of estimators for the population variance, Model Assisted Statistics and Applications 8, 3, pp. 229–238.

SRIVENKATRAMANA, T., TRACY, D. S., (1980). An alternative to ratio method in sample surveys, Annals of the Institute of Statistical Mathematics, 32, pp. 111–120.

YADAV, S. K., KADILAR, C., (2013). Improved exponential type ratio estimator of population variance, Revista colombiana de Estadistica, 36, 1, pp. 145–152.