A GENERAL PROCEDURE FOR ESTIMATING THE MEAN OF A SENSITIVE VARIABLE USING AUXILIARY INFORMATION

Tanveer A. Tarray*[1] and Housila P. Singh**
*Department of Mathematics, Islamic University of Science and Technology, Awantipora, Pulwama, Kashmir - India,192122
**School of Studies in Statistics, Vikram University, Ujjain – India, 456010.

ABSTRACT
This paper proposes classes of estimators for the mean of sensitive variable utilizing information on a non – sensitive auxiliary variable. Expressions for the biases and mean square errors of the suggested estimators correct up to first order of approximation are derived. It has been shown that the suggested new class of estimators based on the known population mean $\overline{X}$ and variance S_x^2 of the auxiliary variable X are better than conventional unbiased estimators which do not utilize the auxiliary information, Sousa et al.'s (2010) ratio estimator and Gupta et al.'s (2012) regression estimator under a very realistic condition.

KEYWORDS: Class of estimators, Randomized response technique; Mean Square error, Bias, Auxiliary variable.

MSC: 62D05.

RESUMEN
Este trabajo propone clases de estimadores para la media de una variable sensitiva usando información sobre un variable auxiliar no – sensitiva. Expresiones para los sesgos y errores cuadráticos medios de los estimadores sugeridos, corregidos hasta el primer orden de aproximación, son derivados. Se demuestra que la nueva clase de estimadores sugerida, basada en al conocimiento de la media $\overline{X}$ y varianza S_x^2 de la variable auxiliar X, son mejores que estimadores insesgados tradicionales que no usan la información auxiliar, como el estimador de razón de Sousa et al. (2010) y el estimador de regresión de Gupta et al. (2012) bajo condiciones muy realísticas.

1.INTRODUCTION

In survey sampling, it is well recognized that the use of auxiliary information results in substantial gain in efficiency over the estimators which do not utilize such information. Ratio, product, regression estimators and their many rectifications have been discussed in the literature. In survey research, direct reliable observation on the variable of interest Y is sometimes not possible because the variable may be sensitive in nature such as habitual tax evasion, reckless driving, indiscriminate gambling, abortion, etc. However we may be able to directly observe a highly correlated auxiliary variable X, for instance see Sousa et al. (2010, 2012) and Koyuncu et al. (2014). Eichhorn and Hayre (1983), Bar – Lev et al. (2004), Perri (2008) and many others have estimated the mean of a sensitive variable when the study variable is sensitive and there is no auxiliary variable. Sousa et al. (2010), Gupta et al. (2012), Koyuncu et al. (2014) and Tarray and Singh (2014) have proposed mean estimators based on randomized response technique (RRT) models in the presence of an auxiliary variable that can be observed directly.

In this paper we have made an effort for developing the classes of estimators of the population mean $\overline{Y}$ of the sensitive variable Y using two different situations: (i) when the population mean $\overline{X}$ of the auxiliary variable X is known; and (ii) when both population mean $\overline{X}$ and variance S_x^2 of the auxiliary variable X are known.

Let Y be the variable under study, a sensitive variable which can't be observed directly. Let X be a non – sensitive auxiliary variable that have a positive correlation with the study variable Y. Let S be a scrambling variable independent of the study variable Y and the auxiliary variable X. The respondent is asked to report a scrambled response for Y given by Z_a =Y+S, but is asked to give a true response for the auxiliary variable X. To obtain the second response, Hussain (2012) advocated the use of

[1] [1]Corresponding Authors e-mail ID: tanveerstat@gmail.com

subtractive model Z_s = Y-S. Here we suggest the generalized additive model Z_g = Y+gS for giving the scrambled response, but is asked to provide a true response for X, where g is a suitably choosen constant such that |g| <1. It is to be noted that for g =1 , Z_g reduces to additive model Z_a = Y+S while for g = -1 it reduces to subtractive model Z_s =Y-S. Suppose a simple random sample of size n is drawn without replacement from a finite population U=(U_1,U_2,…U_N). For the i[th] unit (i=1,2,…,N), let y_i and x_i respectively, be the values of the study variable Y and the auxiliary variable X. Further, let $\bar{y}=(1/n)\sum_{i=1}^{n} y_i$, $\bar{x}=(1/n)\sum_{i=1}^{n} x_i$, $\bar{z}_g=(1/n)\sum_{i=1}^{n} z_{gi}$ be the sample means and $\bar{Y}=E(Y)$, $\bar{X}=E(X)$, $\bar{Z}_g=E(Z_g)$ be the population mean for Y, X and Z_g respectively. We assume that the population mean $\bar{X}$ of the auxiliary variable X is known and $\bar{S}=E(S)=0$. Thus, $E(Z_g)=E(Y)$. If information on auxiliary variable X is ignored, then an unbiased estimator of the population mean $\bar{Y}$ is given by

$$\hat{\bar{Y}}=\bar{z}_g=\frac{1}{n}\sum_{i=1}^{g} z_{gi} \qquad (1.1)$$

The variance / mean square error (MSE) (ignoring finite population correction terms) of $\hat{\bar{Y}}$ is given by

$$MSE(\hat{\bar{Y}})=\frac{1}{n}(S_y^2+g^2S_s^2) \ , \qquad (1.2)$$

where $S_y^2=\frac{1}{(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2$, and $S_s^2=\frac{1}{(N-1)}\sum_{i=1}^{N}(s_i-\bar{S})^2$.

We note that for $g=\pm 1$, the MSE in (1.2) is given by

$$MSE(\bar{Z}_a)=MSE(\bar{Z}_s)=\frac{1}{n}(S_y^2+S_s^2) \ , \qquad (1.3)$$

where $\bar{z}_a=\frac{1}{n}\sum_{i=1}^{n} z_{ai}$ and $\bar{z}_s=\frac{1}{n}\sum_{i=1}^{n} z_{si}$.

From (1.2) and (1.3) we have

$$MSE(\bar{Z}_a)\{=MSE(\bar{Z}_s)\}=\frac{S_s^2}{n}(1-g^2) \qquad (1.4)$$

which is positive if

$1-g^2 > 0$

i.e. if $g^2 < 1$

i.e. if $|g| < 1$ (1.5)

Thus putting the condition |g| < 1 in the proposed additive model Z_g = Y+gS to be better than usual additive model Z_a = Y+S and the subtractive model Z_s = Y-S is justified.

When the population mean $\bar{X}$ of the auxiliary X is known, Sousa et al. (2010) proposed a ratio estimator based on usual additive model Z_a = Y+S for the population mean $\bar{Y}$ of the sensitive variable Y as

$$\hat{\mu}_R=\bar{z}_a\left(\frac{\bar{X}}{\bar{x}}\right) \ . \qquad (1.6)$$

Based on the proposed model Z_g = Y+gS, |g| <1, using information on the population mean $\bar{X}$, Singh and Tarray (2014) suggested a ratio estimator for the population mean $\bar{Y}$ as

$$\hat{\mu}_{R(1)}=\bar{z}_g\left(\frac{\bar{X}}{\bar{x}}\right) \ . \qquad (1.7)$$

For g =1 in (1.7) , $\hat{\mu}_{R(1)}$ reduces to the estimator $\hat{\mu}_R$ in (1.6) due to Sousa et al. (2010).

Sousa et al. (2010) suggested the transformed ratio estimator

$$\hat{\mu}_{R(2)} = \bar{z}_a\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right), \tag{1.8}$$

where c and d are the unit – free parameters, which may be quantities such as the coefficient of skewness and coefficient kurtosis of the auxiliary variable X.

Motivated by Sousa et al. (2010) we define a transformed ratio estimator for the population mean $\bar{Y}$ of the sensitive variable Y as

$$\hat{\mu}_{R(3)} = \bar{z}_g\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right), \tag{1.9}$$

where (c,d) are same as defined for (1.8).

Several other estimators like:

$$\hat{\mu}_{R(4)} = \bar{z}_g\left(\frac{\bar{X}}{\{\bar{X}+\alpha(\bar{x}-\bar{X})\}}\right),\ \hat{\mu}_{R(5)} = \bar{z}_g\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha},\ \hat{\mu}_{R(6)} = \bar{z}_g \exp\left(\frac{\alpha(\bar{X}-\bar{x})}{(\bar{X}+\bar{x})}\right),$$

$$\hat{\mu}_{R(7)} = \bar{z}_g\left(2-\left\{\frac{\bar{X}}{\bar{x}}\right\}^{\alpha}\right),\ \hat{\mu}_{R(8)} = \bar{z}_g\left(\frac{\bar{X}^*}{\{\alpha\bar{x}^*+(1-\alpha)(\bar{x}^*-\bar{x}^*)\}}\right)^{\eta},\ \hat{\mu}_{R(9)} = \bar{z}_g \exp\left(\frac{(\bar{X}^*-\bar{x}^*)}{(\bar{X}^*+\bar{x}^*)}\right),$$

$$\hat{\mu}_{R(10)} = \bar{z}_g \exp\left(1-\frac{\alpha\bar{X}}{(\bar{X}+(\alpha-1)\bar{x})}\right) = \bar{z}_g \exp\left(\frac{\bar{X}-\bar{x}}{(\bar{X}+(\alpha-1)\bar{x})}\right)$$

etc. can be proposed for estimating the population mean $\bar{Y}$ in the presence of known value of the population mean $\bar{X}$ of the auxiliary variable x, where $\bar{x}^* = c\bar{x}+d, \bar{X}^* = c\bar{X}+d$, and (α,c,d,η) being suitably chosen constants.

Keeping the form of the estimators reported in section 1 in mind, we think that defining different estimators and study their properties separately, it would be more appropriate to consider a unified approach of defining a class of estimators and discuss its properties. Such a class of estimators for estimating the population mean $\bar{Y}$ using known value of $\bar{X}$ in the absence of scrambling variable has been studied by Srivastava (1971). Motivated by this arguments and the procedure adopted by Srivastava (1971) we have proposed a class of estimators for population mean $\bar{Y}$ of the sensitive variable Y in the presence of scrambling variable S utilizing the knowledge on population mean $\bar{X}$ of the auxiliary variable X and its properties are studied in sections 2 and 2.1. Further motivated by Srivastava and Jhajj (1981) we extend the class of estimators (2.1) to one which depend also upon the ratio of sample variance to the population variance of the auxiliary variable X and show that the asymptomatic mean square error can be lower than that attained by an estimator of the class $\hat{\mu}_h$ in (2.1).

We assume that a simple random sample of size n is drawn from a finite population of size N. For simplicity we assume that the population size N is large as compared to the sample size n so that finite population correction terms are ignored. We write

$S_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{X})^2, \mu_{pqt} = \frac{1}{(N)}\sum_{j=1}^{N}(z_{gj}-\bar{Z}_g)^p(x_j-\bar{X})^q(y_j-\bar{Y})^t$, ((p,q) being non – negative integers), $C_y^2 = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{002}}{\bar{Y}^2}, C_x^2 = \frac{S_x^2}{\bar{X}^2} = \frac{\mu_{020}}{\bar{X}^2}, \lambda = \frac{\mu_{120}}{\bar{Z}_g S_x^2} = \frac{\mu_{120}}{\bar{Y}S_x^2}$,

$\gamma_1 = \frac{\mu_{030}}{S_x^3}, \beta_1(x) = \gamma_1^2, \beta_2(x) = \frac{\mu_{040}}{S_x^4}$ and $\rho_{yx} = \frac{S_{yx}}{S_x S_y}$ is the correlation coefficient between y and x.

Let $w = \bar{z}_g/\bar{Z}_g$, $u = \bar{x}/\bar{X}$ and $\nu = s_x^2/S_x^2$. Then we have

$$E(w)=E(u)=E(\nu)=1, E\{(w-1)^2\}=\left(\frac{1}{n}\right)C_{z_g}^2 = \frac{(S_y^2+g^2S_s^2)}{n\bar{Y}^2},$$

$$E\{(u-1)^2\}=\left(\frac{1}{n}\right)C_x^2, E\{(w-1)(u-1)\}=\left(\frac{1}{n}\right)\rho_{yx}C_yC_x$$

$$=\left(\frac{1}{n}\right)\rho_{xz_g}C_{z_g}C_x$$

and up to terms of order n^{-1},

$$E\{(\nu-1)^2\}=\left(\frac{1}{n}\right)(\beta_2(x)-1), E\{(w-1)(\nu-1)\}=\left(\frac{1}{n}\right)\lambda,$$

$$E\{(u-1)(\nu-1)\}=\left(\frac{1}{n}\right)C_x\gamma_1.$$

2. A CLASS OF ESTIMATORS FOR $\bar{Y}$ BASED ON THE KNOWN POPULATION MEAN $\bar{X}$ OF THE AUXILIARY VARIABLE X

Following the same procedure as adopted by Srivastava (1971) we define a class of estimators for the population mean $\bar{Y}$ as

$$\hat{\mu}_h = \bar{z}_g h(u) \tag{2.1}$$

where $u=\bar{x}/\bar{X}$, h (.) is a parametric function such that h (1) = 1 and which satisfies certain regularity conditions.

Up to terms of order n^{-1}, the bias and mean square error (ignoring finite population correction terms) of the proposed class of estimators $\hat{\mu}_h$ are respectively given by

$$B(\hat{\mu}_h)=\left(\frac{\bar{Y}}{2n}\right)\left[2\rho_{xz_g}C_{z_g}C_x h_1(1)+C_x^2h_{11}(1)\right]$$

$$=\left(\frac{\bar{Y}}{2n}\right)\left[2\rho_{yx}C_yC_x h_1(1)+C_x^2h_{11}(1)\right] \tag{2.2}$$

and

$$MSE(\hat{\mu}_h)=\left(\frac{\bar{Y}^2}{n}\right)\left[C_{z_g}^2+C_x^2h_1^2(1)+2\rho_{xz_g}C_{z_g}C_x h_1(1)\right]$$

$$=\left(\frac{\bar{Y}^2}{n}\right)\left[C_y^2(1+g^2r^2)+C_x^2h_1^2(1)+2\rho_{yx}C_yC_x h_1(1)\right] \tag{2.3}$$

where $h_1(u)$ and $h_{11}(u)$ are respectively the first and second order partial derivatives of the function h (u),

$$C_{z_g}^2=C_y^2(1+g^2r^2),\ r=(S_s/S_y),\ \rho_{xz_g}=\frac{S_{xz_g}}{S_{z_g}S_x}=\frac{S_{yx}}{S_xS_y\sqrt{(1+g^2r^2)}}=\frac{\rho_{yx}}{\sqrt{(1+g^2r^2)}},$$

$$\rho_{yx}=\frac{S_{yx}}{S_xS_y}, S_{xz_g}=\frac{1}{(N-1)}\sum_{i=1}^{N}(z_{gi}-\bar{Z}_{gi})(x_i-\bar{X}), S_{z_g}^2=\frac{1}{(N-1)}\sum_{i=1}^{N}(z_{gi}-\bar{Z}_{gi})^2=S_y^2(1+g^2r^2),$$

$$S_{yx}=\frac{1}{(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X}), S_x^2=\frac{1}{(N-1)}\sum_{i=1}^{N}(x_i-\bar{X})^2, \text{ and } C_y^2=S_y^2/\bar{Y}^2.$$

The biases and MSEs of the estimators $\hat{\mu}_R$ and $\hat{\mu}_{R(i)}$ $(i=1 \text{ to } 10)$ can be easily obtained from the expressions (2.2) and (2.3) just by putting the suitable values of the partial derivatives $h_1(1)$ and $h_{11}(1)$.

The $MSE(\hat{\mu}_h)$ at (2.3) is minimized for

$$h_1(1) = -\rho_{xz_g}\left(\frac{C_{z_g}}{C_x}\right) = -\rho_{yx}\left(\frac{C_y}{C_x}\right) \quad (2.4)$$

Thus the resulting minimum $MSE(\hat{\mu}_h)$ is given by

$$\text{min.} MSE(\hat{\mu}_h) = \frac{S_y^2}{n}(1+g^2r^2-\rho_{yx}^2) \quad (2.5)$$

which equals to the mean square error of the linear regression estimator

$$\bar{y}_{lrg} = \bar{z}_g + \beta_g(\bar{X}-\bar{x}) \quad (2.6)$$

where $\beta_g = \frac{S_{xz_g}}{S_x^2} = \frac{S_{yx}}{S_x^2} = \beta = \rho_{yx}\frac{S_y}{S_x}$ is the population regression coefficient of Z_g (or Y) on X.

Assuming linear relationship between Y and X; Tarray and Singh (2014) proposed the following regression estimator for the population mean $\bar{Y}$ of the sensitive variable Y as

$$\hat{y}_{lr} = \bar{z}_g + \hat{\beta}_g(\bar{X}-\bar{x})$$

whose approximate MSE is equal to the regression estimator defined in (2.6), where

$$\hat{\beta}_g = \frac{s_{xz_g}}{s_x^2}, s_{xz_g} = \frac{1}{(n-1)}\sum_{i=1}^{n}(z_{gi}-\bar{z}_{gi})(x_i-\bar{x}) \text{ and } s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

It is to be mentioned here that for g=1, the estimator due to Tarray and Singh (2014) reduces to the regression estimator

$$\hat{y}_{lra} = \bar{z}_a + \hat{\beta}_{xz_a}(\bar{X}-\bar{x})$$

where $\hat{\beta}_{xz_a}$ is the sample regression coefficient between Z_a (Y+X) and X.

We note that the linear regression estimator $\bar{y}_{lrg}$ in (2.6) is not a member of the class defined by (2.1). It can be proved that even for the wider class of estimators

$$\hat{\mu}_H = H(\bar{z}_g, u) \quad (2.7)$$

of the population mean $\bar{Y}$ of the sensitive variable Y, where H(.,.) is a function satisfying suitable conditions, the lower bound of the asymptotic mean square error is equal to the asymptotic mean square error of the linear regression estimator $\bar{y}_{lrg}$ and is not reduced. It should be mentioned here that $\bar{y}_{lrg}$ is a member of the class of estimators $\hat{\mu}_h$ given by (2.7). The estimator $\hat{\mu}_R$ and $\hat{\mu}_{R(i)}$ (i=1 to10) are members of both the classes of estimators $\hat{\mu}_h$ in (2.1) and $\hat{\mu}_H$ in (2.7).

2.1 Efficiency comparison

To the first degree of approximation, the MSE of the ratio estimator $\hat{\mu}_{R(1)}$ (ignoring fpc term) proposed by Singh and Tarray (2014) is given by

$$MSE(\hat{\mu}_{R(1)}) = \left(\frac{\bar{Y}^2}{n}\right)\left[C_y^2(1+g^2r^2)+C_x^2-2\rho_{yx}C_yC_x\right] \quad (2.8)$$

For g =1, $MSE(\hat{\mu}_{R(1)})$ reduces to the MSE of the ratio estimator $\hat{\mu}_R$ due to Sousa et al. (2010) as

$$MSE(\hat{\mu}_R) = \left(\frac{\bar{Y}^2}{n}\right)\left[C_y^2(1+r^2)+C_x^2-2\rho_{yx}C_yC_x\right]. \quad (2.9)$$

We below give the efficiency comparison of the proposed class of estimators $\hat{\mu}_h$ with that of the conventional estimator $\hat{\overline{Y}}$, Tarray and Singh (2014) ratio estimator $\hat{\mu}_{R(1)}$ and Sousa et al. (2010) ratio estimator $\hat{\mu}_R$ when the optimum value $-\rho_{yx}(C_y/C_x)$ of $h_1(1)$ does not coincide with its true value.
From (1.2) and (2.5) we have

$$MSE(\hat{\mu}_h)-MSE(\hat{\overline{Y}})=\left(\frac{\overline{Y}^2}{n}\right)C_x^2h_1(1)\left[h_1(1)+2\rho_{yx}(C_y/C_x)\right]$$

which is less than zero if

$$h_1(1)\left[h_1(1)+2\rho_{yx}(C_y/C_x)\right]<0$$

$$\left.\begin{array}{l}\text{i.e. if either } 0<h_1(1)<-2\rho_{yx}(C_y/C_x)\\ \qquad\text{or } -2\rho_{yx}(C_y/C_x)<h_1(1)<0\end{array}\right\} \tag{2.10}$$

From (2.5) and (2.8) we have

$$MSE(\hat{\mu}_h)-MSE(\hat{\mu}_{R(1)})=\left(\frac{\overline{Y}^2}{n}\right)C_x^2(h_1(1)-1)\left[h_1(1)+2\rho_{yx}(C_y/C_x)+1\right]$$

which fewer than zero if

$$(h_1(1)-1)\left[h_1(1)+\{1+2\rho_{yx}(C_y/C_x)\}\right]<0$$

$$\left.\begin{array}{l}\text{i.e. if either } 1<h_1(1)<-\{1+2\rho_{yx}(C_y/C_x)\}\\ \qquad\text{or } -\{1+2\rho_{yx}(C_y/C_x)<h_1(1)<1\end{array}\right\} \tag{2.11}$$

Further from (2.5) and (2.9) we have

$$MSE(\hat{\mu}_h)-MSE(\hat{\mu}_R)=\left(\frac{\overline{Y}^2}{n}\right)\left[C_y^2(g^2-1)+C_x^2\left[(h_1^2(1)-1)+2\rho_{yx}(C_y/C_x)(h_1(1)-1)\}\right]\right]$$

(2.12)
which is negative if

$$\left[C_y^2(g^2-1)+C_x^2(h_1^2(1)-1)\{h_1(1)+2\rho_{yx}(C_y/C_x)+1\}\right]<0$$

$$\text{i.e. if} \qquad (g^2-1) \text{ and } C_x^2(h_1^2(1)-1)\{h_1(1)+2\rho_{yx}(C_y/C_x)+1\}<0$$

$$\left.\begin{array}{l}\text{i.e. if either } 1<h_1(1)<-\{1+2\rho_{yx}(C_y/C_x)\},\ |g|<1\\ \qquad\text{or } -\{1+2\rho_{yx}(C_y/C_x)\}<h_1(1)<1,\quad |g|<1\end{array}\right\} \tag{2.13}$$

Thus the proposed class of estimators $\hat{\mu}_h$ is better than the conventional estimator $\hat{\overline{Y}}$, Tarray and Singh (2014) ratio estimator $\hat{\mu}_{R(1)}$ and Sousa et al. (2010) ratio estimator $\hat{\mu}_R$ respectively if the conditions (2.10), (2.11) and (2.13) hold good.

Now we compare the proposed class of estimators $\hat{\mu}_h$ with that of the conventional estimator $\hat{\overline{Y}}$, Tarray and Singh (2014) ratio estimator $\hat{\mu}_{R(1)}$ and Sousa et al. (2010) ratio estimator $\hat{\mu}_R$ in the situation where the optimum value $-\rho_{yx}(C_y/C_x)$ of $h_1(1)$ exactly coincide with its true value, which is possible from the past data or the experience gathered in due course of time for instance, see Srivastava (1967), Reddy (1973) , Srivenkataramana and Tracy (1980) and Singh and Ruiz Esepjo (2003) .
From (1.2), (2.5), (2.8) and (2.9) we have

$$MSE(\hat{\overline{Y}})-\min.MSE(\hat{\mu}_h)=\left(\frac{S_y^2}{n}\right)\rho_{yx}^2>0 \tag{2.14}$$

$$MSE(\hat{\mu}_{R(1)}) - \min .MSE(\hat{\mu}_h) = \left(\frac{\overline{Y}^2}{n}\right)(C_x - \rho_{yx}C_y)^2 \tag{2.15}$$

$$> 0 \text{ provided } C_x \neq \rho_{yx}C_x \text{ (i.e.} R \neq \beta)$$

$$MSE(\hat{\mu}_R) - \min .MSE(\hat{\mu}_h) = \left(\frac{\overline{Y}^2}{n}\right)[r^2C_y^2(1-g^2) + (C_x - \rho_{yx}C_y)^2] \tag{2.16}$$

$$> 0 \text{ provided } |g| < 1 \text{ and } R \neq \beta,$$

where $R = \overline{Y}/\overline{X}$ and β is the population regression coefficient of Y on X.

It follows from (2.14) to (2.16) that the proposed class of estimators $\hat{\mu}_h$ is more efficient than:

(i) the conventional estimator $\hat{\overline{Y}}$,

(ii) Tarray and Singh's (2014) estimator $\hat{\mu}_{R(1)}$ unless $R = \beta$. In the case of $R = \beta$, both the estimators $\hat{\mu}_h$ and $\hat{\mu}_{R(1)}$ are equal efficient.

(iii) Sousa et al.'s (2010) estimator $\hat{\mu}_R$ when |g| <1 and $R \neq \beta$.

3. THE CLASS OF ESTIMATORS BASED ON KNOWN POPULATION MEAN $\overline{X}$ AND VARIANCE S_x^2 OF THE AUXILIARY VARIABLE X

Suppose that the population mean $\overline{X}$ and variance S_x^2 of the auxiliary variable X are known. In such situations, we propose a class of estimators of $\overline{Y}$ as

$$\hat{\mu}_t = \bar{z}_g t(u, \nu) , \tag{3.1}$$

where $t(u, \nu)$ is a parametric function of u and ν such that t(1,1) =1. Whatever sample is chosen, let (u, ν) assume values in a bounded closed convex subset, R, of the two – dimensional real space containing the point (1,1). The function $t(u, \nu)$ is continuous and bounded having continuous and bounded first and second order partial derivatives in R.

Since there are only a finite number of samples therefore under the above conditions, the expectation and the mean square error of the estimators of the class $\hat{\mu}_t$ exist. Expanding $t(u, \nu)$ about the point (1,1) in a second – order Taylor's series, we have that , $E(\hat{\mu}_t) = \overline{Y} + O(n^{-1})$, and so the bias is of the order of n^{-1}. Thus mean square error of $\hat{\mu}_t$ up to terms of order n^{-1} is

$$\begin{aligned} MSE(\hat{\mu}_t) &= \left(\frac{\overline{Y}^2}{n}\right)\Big[C_{z_g}^2 + C_x^2 t_1^2(1,1) + (\beta_2(x) - 1)t_2^2(1,1) + 2\rho_{xz_g} C_{z_g} C_x t_1(1,1) \\ &\quad + 2\lambda t_2(1,1) + 2C_x\gamma_1 t_1(1,1)t_2(1,1)\Big], \\ &= \left(\frac{\overline{Y}^2}{n}\right)\Big[(1/\overline{Y}^2)(S_y^2 + g^2S_s^2) + C_x^2 t_1^2(1,1) + (\beta_2(x) - 1)t_2^2(1,1) \\ &\quad + 2\rho_{yx}C_xC_y t_1(1,1) + 2\lambda t_2(1,1) + 2C_x\gamma_1 t_1(1,1)t_2(1,1)\Big], \end{aligned} \tag{3.2}$$

where $t_1(u, \nu)$ and $t_2(u, \nu)$ denote the first partial derivatives of the function $t(u, \nu)$. The mean square error of $\hat{\mu}_t$ at (3.2) is minimized for

$$\left.\begin{aligned} t_1(1,1) &= \frac{(\lambda\gamma_1 - \rho_{yx}C_y(\beta_2(x)-1))}{(\beta_2(x)-\beta_1(x)-1)C_x} \\ t_1(1,1) &= \frac{(\rho_{yx}C_y\gamma_1 - \lambda)}{(\beta_2(x)-\beta_1(x)-1)} \end{aligned}\right\} \quad (3.3)$$

Thus the resulting minimum mean square error of the estimator $\hat{\mu}_t$ up to terms of order n^{-1} is given by

$$\min.MSE(\hat{\mu}_t) = \left(\frac{1}{n}\right)\bar{Y}^2\left[C_{z_g}^2(1-\rho_{xz_g}^2) - \frac{(\rho_{xz_g}C_{z_g}\gamma_1 - \lambda)^2}{(\beta_2(x)-\beta_1(x)-1)}\right]$$

$$= \left(\frac{1}{n}\right)\bar{Y}^2\left[C_{z_g}^2(1-\rho_{xz_g}^2) - \frac{(\rho_{xz_g}C_y\gamma_1 - \lambda)^2}{(\beta_2(x)-\beta_1(x)-1)}\right]$$

$$= \left(\frac{1}{n}\right)\left[S_y^2(1+g^2r^2-\rho_{yx}^2) - \frac{\bar{Y}^2(\rho_{yx}C_y\gamma_1 - \lambda)^2}{(\beta_2(x)-\beta_1(x)-1)}\right]$$

$$= MSE(\bar{y}_{lrg}) - \frac{\bar{Y}^2}{n}\frac{(\rho_{yx}C_y\gamma_1 - \lambda)^2}{(\beta_2(x)-\beta_1(x)-1)} \quad (3.4)$$

which is, of course, fewer than the minimum mean squared error of any estimator of the class (2.1), with equality iff $\rho_{yx}C_y\gamma_1 = \lambda$. This shows that the proposed class of estimators $\hat{\mu}_t$ would be worth using when the relationship between Z_g (or Y) and X is markedly non – linear and $(\beta_2(x)-\beta_1(x)-1)$ is small.

It is to be noted that for g =0, the class of estimators $\hat{\mu}_t$ reduces to the class of estimators $\hat{\mu}_t^* = \bar{y}t(u,v)$ envisaged by Srivastava and Jhajj (1981) in the absence of scrambling variable S. If we set g =0 in (3.4), we get the minimum MSE of the class of estimator $\hat{\mu}_t^*$ up to terms of order n^{-1} as

$$\min.MSE(\hat{\mu}_t^*) = \left(\frac{1}{n}\right)\left[S_y^2(1-\rho_{yx}^2) - \frac{\bar{Y}^2(\rho_{yx}C_y\gamma_1 - \lambda)^2}{(\beta_2(x)-\beta_1(x)-1)}\right]. \quad (3.5)$$

Comparing (3.4) and (3.5) we observe that there is increase in the MSE($\hat{\mu}_t$) (i.e. in the MSE of the proposed class of estimators $\hat{\mu}_t$) due to the presence of scrambling variable S which is expected too. Thus whatever be the value of S, a respondent chooses, the effect of scrambling will be small if he/she selects the value of g in the neighborhood of 'zero'.

Remark 3.1 – The proposed class of estimators $\hat{\mu}_t$ of the population mean $\bar{Y}$ of the sensitive variable Y are very large. Any parametric function $t(u,v)$ satisfying certain regularity conditions and t(1,1)=1, can generate an estimator of the class $\hat{\mu}_t$. For example, the following estimators:

$$\hat{\mu}_{t(1)} = \bar{z}_g u^{\alpha}v^{\beta},\ \hat{\mu}_{t(2)} = \bar{z}_g\frac{\{1+\alpha(u-1)\}}{\{1-\beta(v-1)\}},$$

$$\hat{\mu}_{t(3)} = \bar{z}_g[1+\alpha(u-1)+\beta(v-1)], \hat{\mu}_{t(4)} = \bar{z}_g\{1-\alpha(u-1)-\beta(v-1)\}^{-1},$$

$$\hat{\mu}_{t(S)} = \bar{z}_g e^{\alpha(u-1)+\beta(v-1)}$$

etc. are the members of the suggested class of estimators $\hat{\mu}_t$.

It can be easily seen that the optimum values of the parameters α and β in all the five estimators are same and are given by the right – hand sides of (3.3).

It is easily shown that if we consider a wider class of estimators

$$\hat{\mu}_T = T(\bar{z}_g, u, \nu) \tag{3.6}$$

of the population mean $\bar{Y}$ of the sensitive variable Y, where the function $T(\bar{z}_g, u, \nu)$ satisfies $T(\bar{Y},1,1) = \bar{Y}$, $T_1(\bar{Y},1,1) = 1$ and $T_1(\bar{Y},1,1) = \left.\frac{\partial T(\bar{z}_g, u, \nu)}{\partial \bar{z}_g}\right|_{(\bar{Y},1,1)}$, (denoting the first partial derivative of the function $T(\bar{z}_g, u, \nu)$ with respect to $\bar{z}_g$ about the point ($\bar{Y}$, 1,1)), the minimum mean square error of $\hat{\mu}_t$ is equal to (3.4) and is not reduced. We note that the regression type estimator $t_1 = \{(\bar{z}_g + \alpha(u-1) + \beta(\nu-1)\}$ is a member of the class (3.6) but not of the class (3.1).

Remark 3.2 – It should be mentioned that the efficient use of the two suggested class $\hat{\mu}_h$ in (2.1) and $\hat{\mu}_t$ in (3.1) presumes that the optimum values of $h_1(1)$ and $t_i(1,1)$ (i=1,2) are known. But these values are functions of unknown population parameters. Singh (1982) and Srivastava and Jhajj (1983) have shown that the estimators of the class with estimated values of optimum parameters obtained by their consistent estimators, attain the same minimum mean square error of the estimators of the class based on optimum values, up to the first order of approximation.

Alternatively, to use such estimators in practice, one has to use some guessed values of the parameters such as $C_y, C_x, \rho_{yx}, \lambda, \gamma_1$ and $\beta_2(x)$, either through past experience or through a pilot sample survey. It may be noted that even if the values of the parameters used in the estimator are not exactly equal to their optimum values as given by (2.4) and (3.3) but are close enough, the resulting estimator will be better than the usual unbiased estimator $\bar{z}_g$, as has been illustrated by Das and Tripathi (1978, sec. 3). The relevant references in this context are Srivastava and Jhajj (1980), Sampath (1988), Upadhyaya et al. (2004) and Jhajj et al. (2005).

3.1 The bias of the proposed class of estimators $\hat{\mu}_t$

To obtain the bias of the estimator $\hat{\mu}_t$, we will have to strengthen the conditions on $t(u,\nu)$ of section 3 assuming that its third order partial derivative also exists and are continuous and bounded. Then expanding $t(u,\nu)$ about the point (1,1) in a third order Taylor's series, taking expectation of $\hat{\mu}_t$ and retaining terms up to terms of order n^{-1} , we obtain

$$\begin{aligned} B(\hat{\mu}_t) &= E(\hat{\mu}_t) - \bar{Y} \\ &= \left(\frac{\bar{Y}}{2n}\right)\left[2\rho_{yx}C_xC_y t_1(1,1) + 2\lambda t_2(1,1) + C_x^2 t_{11}(1,1) \right. \\ &\quad \left. + (\beta_2(x)-1)t_{22}(1,1) + 2C_x\gamma_1 t_{12}(1,1)\right] \end{aligned} \tag{3.7}$$

where $t_{11}(u,\nu)$, $t_{12}(u,\nu)$ and $t_{22}(u,\nu)$ denote the second partial derivatives of $t(u,\nu)$.

It is observed from (3.7) that the bias of $\hat{\mu}_t$ depends also up on the second order partial derivatives of the function $t(u,\nu)$ at the point (1,1) and hence will be different for different estimators of the class. For the sake of the completeness we below give the biases of the five estimators $\hat{\mu}_{t(i)}$, $i = 1$ to 5 up to terms of order n^{-1}, are given by

$$B(\hat{\mu}_{t(1)}) = \left(\frac{\bar{Y}}{2n}\right)\left\{2\rho_{yx}C_yC_x\alpha + 2\lambda\beta + C_x^2\alpha(\alpha-1) + (\beta_2(x)-1)\beta(\beta-1) + 2C_x\gamma_1\alpha\beta\right\} \tag{3.8}$$

$$B(\hat{\mu}_{t(2)}) = \left(\frac{\bar{Y}}{2n}\right)\left\{2\rho_{yx}C_yC_x\alpha + 2\lambda\beta + 2(\beta_2(x)-1)\beta^2 + 2C_x\gamma_1\alpha\beta\right\} \tag{3.9}$$

$$B(\hat{\mu}_{t(3)}) = \left(\frac{\overline{Y}}{2n}\right)\left\{2\rho_{yx}C_yC_x\alpha + 2\lambda\beta\right\} \quad (3.10)$$

$$B(\hat{\mu}_{t(4)}) = \left(\frac{\overline{Y}}{2n}\right)\left\{\rho_{yx}C_yC_x\alpha + \lambda\beta + C_x^2\alpha^2 + (\beta_2(x) - 1)\beta^2 + 2C_x\gamma_1\alpha\beta\right\} \quad (3.11)$$

$$B(\hat{\mu}_{t(5)}) = \left(\frac{\overline{Y}}{2n}\right)\left\{2\rho_{yx}C_yC_x\alpha + 2\lambda\beta + C_x^2\alpha^2 + (\beta_2(x) - 1)\beta^2 + 2C_x\gamma_1\alpha\beta\right\} \quad (3.12)$$

4. CONCLUSION

This article proposed two classes of estimators in the cases (i) when population mean $\overline{X}$ of the auxiliary variable X is known; and (ii) when both the population mean $\overline{X}$ and the population variance S_x^2 are known. Sousa et al. (2010) in their study have shown that second order approximation as compared to first order approximation does not result in major difference in absolute relative bias (ARB) and even MSE for moderate sample size. Keeping this in mind we have derived the bias and MSE expressions of the proposed classes of estimators up to first order of approximation only. Asymptotic optimum estimator (AOE) is identified in each suggested class along with mean square error formula. It has been shown that the proposed class of estimators based on both the population mean $\overline{X}$ and population variance S_x^2 of the auxiliary variable X is more efficient than Sousa et al's (2010) ratio estimator, Gupta et al.'s (2012) regression estimator and the conventional estimators. The proposed study is very sound in theoretical point of view.

ACKNOWLEDGEMENTS: The authors are thankful to the Editor – in- Chief - Professor C.N. Bouza, and the two anonymous learned referees for their valuable suggestions regarding improvement of the paper.

RECEIVED NOVEMBER , 2014
REVISED MAY, 2015

REFERENCES

[1] BAR –LEV S.K.E., BOBOVITCH and BOUKAI B. (2004): A note on Randomized response models for quantitative data. **Metrika**, 60, 225-250.

[2] CHRISTOFIDEs T.C. (2003): A generalized randomized response technique. **Metrika**, 57, 195-200.

[3] COCHRAN W.G. (1977): **Sampling Technique,** *3rd Edition.***,** JohnWiley and Sons, New York

[4] DAS A.K. and TRIPATHI T.P. (1978): Use of auxiliary information in estimating the finite population variance. **Sankhya -C**, 40, 139-148.

[5] EICHHORN B.H. and HAYRE .LS. (1983): Scrambled randomized response methods for obtaining sensitive quantitative data. **Journal of Statistics Planning and Inference,** 7, 307-316.

[6] GUPTA S., SHABBIR J., SOUSA R. and REAL P.C. (2012): Estimation of mean of a sensitive variable in the presence of auxiliary information. **Communication of Statistics Theory and Methods,** 41, 2394-2404.

[7] HUSSAIN Z. (2012): Improvement of Gupta and Thornton scrambling model through double use of randomization device. **International Journal Academic Research Business and Social Science**, 2,91-97.

[8] JHAJJ H.S., SHARMA M.K. and GROVER L.K. (2005): An efficient class of chain estimators of population variance under sub sampling scheme. **Journal of Japan Statistics Society,** 35, 273 – 286.

[9] KOYUNCU N., GUPTA S. and SOUSA R. (2014): Exponential – type estimator of the mean of a sensitive variable in the presence of non sensitive auxiliary information. **Communication of Statistics Theory and Methods,** 43, 1583-2014.

[10] PERRI P.F. (2008) Modified randomized devices for Simmons' model. **Model Assisted Statistics and Applications,** 3, 233–239.

[11] REDDY V.N. (1973): On ratio and product method of estimation. **Sankhya - B,** 35, 307–316

[12] SAMPATH S. (1988): Generalized optimum class of ratio type estimators in double sampling. **Gujarat Statistics Review,** 15 (2), 25-32.

[13] SINGH H.P. and RUIZ-ESPEJO M. (2003): On linear regression and ratio-product estimation of a finite population mean. **The Statistician,** 52, (1), 59-67.

[14] SINGH R.K. (1982): On estimating ratio and product of population parameters. **Calcutta Statistics Association Bulletin ,** 31, 69-76.

[15] SRIVASTAVA S.K. (1967): An estimator using auxiliary information. **Calcutta Statistics Association Bulletin ,** 16,121–132.

[16] SRIVASTAVA S.K. (1971): A general estimator for the mean of a finite population using multi – auxiliary information. **Journal of American Statistical Association,** 66 , 404 – 407.

[17] SRIVASTAVA S.K. and JHAJJ H.S. (1980): A class of estimators using auxiliary information for estimating finite population variance. **Sankhya- C,** 42, 87-96.

[18] SRIVASTAVA S.K. and JHAJJ H.S.. (1981): A class of estimators of the population mean in survey sampling using auxiliary information. **Biometrika,** 68, 341-343.

[19] SRIVASTAVA S.K. and JHAJJ H.S. (1983): A class of estimators of the population mean using multi – auxiliary information. **Calcutta Statistics Association Bulletin ,**32, 47-56.

[20] SRIVENKATARAMANA T. and TRACY D.S. (1980) An alternative to ratio method in sample survey. **Annals of Institutional Statistics and Mathematics**, 32,111–120.

[21] SOUSA R., SHABBIR J., REAL P.C. and GUPTA S. (2010): Ratio estimation of the mean of a sensitive variable in the presence of Auxiliary information. **Journal of Statistics Theory and Practice,** 4, 495-07.

[22] TARRAY T. A. and SINGH H.P. (2014): An improved estimation procedure of the mean of a sensitive variable using auxiliary information. **Communicated**

[23] UPADHYAYA L.N., SINGH, H.P. AND SINGH S. (2004): A class of estimators for estimating the variance of the ratio estimator. **Journal of Japan Statistics Society,** 34, 47-63.