ESTIMATION OF FINITE POPULATION VARIANCE USING SCRAMBLED RESPONSES IN THE PRESENCE OF AUXILIARY INFORMATION

Sarjinder Singh[1], Stephen A. Sedory[1] and Raghunath Arnab[2*]
[1]Texas A & M University-Kingsville, Kingsville, TX 78363, USA
E-mail: sarjinder@yahoo.com
[2]Department of Statistics, University of Botswana, P.BagUB 00705, Gaborone, Botswana. (* Honorary Research Fellow, University of Kwa-Zulu Natal) **E-mail:** arnabr@mopipi.ub.bw

ABSTRACT

In this paper, a new estimator for estimating the finite population variance of a sensitive variable based on scrambled responses collected using a randomization device is introduced. The estimator is then improved by using known auxiliary information. The estimators due to Das and Tripathi (1978: Sankhya) and Isaki (1983: JASA) are shown to be special cases of the proposed estimator. Numerical simulations are performed to study the magnitude of the gain in efficiency when using the estimator with auxiliary information with respect to the estimator based only on the scrambled responses. An idea to extend the present work from SRSWOR design to more complex design is also given.

Key words: Randomized response sampling; Sensitive variables; Variance estimation; Efficiency; Finite population; Auxiliary information.

1. INTRODUCTION

The collection of data through personal interview surveys on sensitive issues such as induced abortions, drug abuse and family income is a serious issue. For example, some questions are sensitive: (a) By how much did you underreport your income on your 2009 tax return? (b) How many abortions have you had? (c) How many children have you molested? (e) Do you use illegal drugs? Randomized response techniques are one way to get people to answer truthfully. Horvitz *et al.* (1967) and Greenberg *et al.* (1971) have extended Warner's (1965) model to the case where the responses to the sensitive question are quantitative rather than a simple 'yes' or 'no'. The respondent selects, by means of a randomization device, one of the two questions: one being the sensitive question, the other being unrelated. However, there are several difficulties which arise when using this unrelated question method. The main one is choosing the unrelated question. As Greenberg *et al.* (1971) note, it is essential that the mean and variance of the responses to the unrelated question be close to those for the sensitive question: otherwise, it will often be possible to recognize from the response which question was selected. However, the mean and variance of the responses to the sensitive question are unknown, making it difficult to choose good unrelated question. A second difficulty is that in some cases the answers to the unrelated question may be more rounded or regular, making it possible to recognize which question was answered. For example, Greenberg *et al.* (1971) considered the sensitive question: about how much money did the head of this

household earn last year. This was paired with the question: about how much money do you think the average head of a household of your size earns in a year. An answer such as \$26,350 is more likely to be in response to the unrelated question, while an answer such as \$18,618 is almost certainly in response to the sensitive question. A third difficulty is that some people are hesitant to disclose their answer to the sensitive question even though they know that the interviewer cannot be sure that the sensitive question was selected. For example, some respondents may not want to reveal their income even though they known that the interviewer can only be 0.75 certain, say, that the figure given is the respondent's income. These difficulties are no longer present in the scrambled randomized response method introduced by Eichhorn and Hayre (1983). This method we summarize as follows:

Each respondent scrambles their response Y by multiplying it by a random variable S and then reveals only the scrambled result $Z = YS$ to the interviewer. Thus the scrambled randomized response model maintains the privacy of the respondents. The variable S is called a scrambling variable and its distribution is known. In particular, the quantities $E(S) = \theta$ and $\gamma_a = E(S-\theta)^a$ for $a = 2,3,4$ are known.

Diana and Perri (2009a, 2009b, 2010, 2011) rightly pointed out that in direct question survey techniques when dealing with non-sensitive questions, it is very common to use auxiliary information to improve estimation strategies. A very limited effort has been made to make use of auxiliary information to improve the estimators of sensitive variables, as one might see in referring to Singh *et al.* (1996), Strachan *et al.* (1998), Tracy and Singh (1999), Son *et al.* (2008) and Singh and Kim (2011). An extensive review of the literature on randomized response sampling can be found in a recent monograph by Chaudhuri (2011). To our knowledge, no one has made any attempt to study an estimator of the finite population variance using scrambled responses on the study variable and making use of an auxiliary variable to improve the estimator.

1.1 NOTATION

Assume that a simple random sample and without replacement (SRSWOR) of size n is drawn from the given population of N units. Let the value of the sensitive study variable, Y and the auxiliary variable, X, for the i^{th} unit $(i = 1,2,...,N)$ of the population be denoted by Y_i and X_i, and the value for the i^{th} unit in the sample $(i = 1,2,...,n)$ by y_i and x_i, respectively. In the population, we define a few parameters of the sensitive study variable Y_i and X_i the related auxiliary variable as follows. Let $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ denote the population mean of the sensitive study variable and let $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ be the population mean of the auxiliary variable X. In this paper, we consider the problem of estimating the finite population variance (or population mean square error) of the sensitive study variable Y defined by: $S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ in the presence of the

known population mean $\overline{X} = N^{-1}\sum_{i=1}^{N} X_i$ and the known population variance $S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$ of the auxiliary variable, X. Let the higher ordered central moments of the study variable Y and the auxiliary variable X be given by

$$\mu_{ab} = \frac{1}{(N-1)}\sum_{i=1}^{N}(Y_i - \overline{Y})^a (X_i - \overline{X})^b$$

for $a, b = 0,1,2,3,4$ etc.

Let $Z_i = Y_i S$ denote the scrambled response from the i[th] sampled unit for $i = 1,2,...,n.$ Here we differ from the Das and Tripathi (1978) and Isaki (1983) estimators in that, instead of observing direct response Y_i on the sensitive study variable, we observe the scrambled response, $Z_i = Y_i S$. The inadmissibility of the usual estimator of finite population variance in case of true responses on the study variable has been studied by Strauss (1982).

2. NAIVE ESTIMATOR OF THE FINITE POPULATION VARIANCE

Following Eichhorn and Hayre (1983) the mean of the response, $\overline{Y}$, can be estimated from a sample of scrambled Z values by using the known knowledge of the distribution of the scrambling variable *S.* Let Z_i, $i = 1,2,3,...,n$ be the observed scrambled responses. Then the sample variance of the scrambled responses is given by:

$$s_z^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Z_i^2 - \frac{\left(\sum_{i=1}^{n} Z_i\right)^2}{n}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} Y_i^2 S_i^2 - \frac{\sum_{i=1}^{n} Y_i^2 S_i^2 + \sum_{i\neq j=1}^{n}\sum^{n} Y_i S_i Y_j S_j}{n}\right]$$

Let E_R denote the expected value over the randomization device. Taking expected value E_R on both sides, we get:

$$E_R\left(s_z^2\right) = \frac{\gamma_2}{n}\sum_{i=1}^{n} Y_i^2 + \theta^2 s_y^2 \qquad (1)$$

The proof of Equation (1) is available on request from the authors.

We thus have the following theorem.

Theorem 2.1. An unbiased estimator of the finite population variance is given by

$$s_y^{*2} = \frac{1}{\theta^2}\left[s_z^2 - \frac{\gamma_2}{n(\gamma_2 + \theta^2)}\sum_{i=1}^{n} Z_i^2\right] \qquad (2)$$

Proof. Available on request from the authors**.**

We call the above the naive estimator of the variance. Next we have the following corollaries:

Corollary 2.1.The variance of s_y^{*2} over the randomization device is given by

$$V_R(s_y^{*2}) = \frac{\left(\gamma_4 + 4\gamma_3\theta + 4\theta^2\gamma_2 - \gamma_2^2\right)}{n^2(\gamma_2+\theta^2)^2}\sum_{i=1}^{n} Y_i^4 + \frac{(\gamma_2^2 + 2\gamma_2\theta^2)}{n^2(n-1)^2\theta^4}\sum_{i\neq j=1}^{n}\sum^{n} Y_i^2 Y_j^2 + \frac{2\gamma_2}{n^2(n-1)^2\theta^2}\sum_{i\neq j\neq k=1}^{n}\sum^{n}\sum^{n} Y_i^2 Y_j Y_k + \frac{4\left(\gamma_3\theta + 2\gamma_2\theta^2\right)}{n^2(n-1)(\gamma_2+\theta^2)\theta^2}\sum_{i\neq j=1}^{n}\sum^{n} Y_i^3 Y_j \quad (3)$$

Proof. Available on request from the authors.

Corollary 2.2. The expected value of $V_R(s_y^{*2})$ over the sampling design $P(s)$ is given by

$$E_P\left[V_R\left(s_y^{*2}\right)\right] = \frac{\left(\gamma_4 + 4\gamma_3\theta + 4\theta^2\gamma_2 - \gamma_2^2\right)}{n(\gamma_2+\theta^2)^2}\frac{1}{N}\sum_{i=1}^{N} Y_i^4 + \frac{(\gamma_2^2 + 2\gamma_2\theta^2)}{n(n-1)\theta^4}\frac{1}{N(N-1)}\sum_{i\neq j=1}^{N}\sum^{N} Y_i^2 Y_j^2 + \frac{2\gamma_2\,(n-2)}{n(n-1)\theta^2 N(N-1)(N-2)}\sum_{i\neq j\neq k=1}^{N}\sum^{N}\sum^{N} Y_i^2 Y_j Y_k + \frac{4\left(\gamma_3\theta + 2\gamma_2\theta^2\right)}{n(\gamma_2+\theta^2)\theta^2 N(N-1)}\sum_{i\neq j=1}^{N}\sum^{N} Y_i^3 Y_j \quad (4)$$

Proof. Available on request from the authors.

Theorem 2.2. The variance of the estimator s_y^{*2} is given by

$$V(s_y^{*2}) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\mu_{40} - \mu_{20}^2\right) + \frac{\left(\gamma_4 + 4\gamma_3\theta + 4\theta^2\gamma_2 - \gamma_2^2\right)}{n^2(\gamma_2+\theta^2)^2}\frac{n}{N}\sum_{i=1}^{N} Y_i^4 + \frac{(\gamma_2^2 + 2\gamma_2\theta^2)}{n^2(n-1)^2\theta^4}\frac{n(n-1)}{N(N-1)}\sum_{i\neq j=1}^{N}\sum^{N} Y_i^2 Y_j^2 + \frac{2\gamma_2\, n(n-1)(n-2)}{n^2(n-1)^2\theta^2 N(N-1)(N-2)}\sum_{i\neq j\neq k=1}^{N}\sum^{N}\sum^{N} Y_i^2 Y_j Y_k + \frac{4\left(\gamma_3\theta + 2\gamma_2\theta^2\right)n(n-1)}{n^2(n-1)(\gamma_2+\theta^2)\theta^2 N(N-1)}\sum_{i\neq j=1}^{n}\sum^{n} Y_i^3 Y_j \quad (5)$$

Proof. Available on request from the authors.

3. DIFFERENCE TYPE ESTIMATOR OF THE FINITE POPULATION VARIANCE

Following Das and Tripathi (1978) and Isaki (1983), we define a new difference type estimator of the finite population variance as:

$$\hat{\sigma}_v^{*2} = s_y^{*2} + B\left(S_x^2 - s_x^2\right) \quad (6)$$

where B is a constant to be determined such that the variance of the estimator $\hat{\sigma}_v^{*2}$ is minimum.

If $B = s_y^{*2} \big/ s_x^2$ then the estimator $\hat{\sigma}_v^{*2}$ becomes ratio type estimator given by

$$\hat{\sigma}_{rat}^{*2} = s_y^{*2}\left(\frac{S_x^2}{s_x^2}\right) \tag{7}$$

If $B = \dfrac{\alpha s_y^{*2}}{\alpha s_x^2 + (1-\alpha)S_x^2}$, where α is a constant, then the estimator $\hat{\sigma}_v^{*2}$ becomes Das and Tripathi (1978) type estimator given by:

$$\hat{\sigma}_{dt}^{*2} = \frac{s_y^{*2} S_x^2}{\alpha s_x^2 + (1-\alpha)S_x^2} \tag{8}$$

Now we have the following theorems:

Theorem 3.1. The proposed estimator $\hat{\sigma}_v^{*2}$ is an unbiased estimator of the finite population variance S_y^2 if B is a known constant.

Proof. Available on request from the authors.

Theorem 3.2. The minimum variance of the proposed estimator $\hat{\sigma}_v^{*2}$ is given by

$$\text{Min. V}\left(\hat{\sigma}_v^{*2}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\mu_{40} - \mu_{20}^2\right)\left[1 - \frac{(\mu_{22} - \mu_{20}\mu_{02})^2}{\left(\mu_{40} - \mu_{20}^2\right)\left(\mu_{04} - \mu_{02}^2\right)}\right]$$

$$+ \frac{\left(\gamma_4 + 4\gamma_3\theta + 4\theta^2\gamma_2 - \gamma_2^2\right)}{n(\gamma_2 + \theta^2)^2}\frac{1}{N}\sum_{i=1}^{N} Y_i^4 + \frac{(\gamma_2^2 + 2\gamma_2\theta^2)}{n(n-1)\theta^4}\frac{1}{N(N-1)}\sum_{i\neq}^{N}\sum_{j=1}^{N} Y_i^2 Y_j^2$$

$$+ \frac{2\gamma_2\,(n-2)}{n(n-1)\theta^2 N(N-1)(N-2)}\sum_{i\neq}^{N}\sum_{j\neq}^{N}\sum_{k=1}^{N} Y_i^2 Y_j Y_k + \frac{4\left(\gamma_3\theta + 2\gamma_2\theta^2\right)}{n(\gamma_2 + \theta^2)\theta^2 N(N-1)}\sum_{i\neq}^{N}\sum_{j=1}^{N} Y_i^3 Y_j \tag{9}$$

Proof. Available on request from the authors.

4. REGRESSION TYPE ESTIMATOR

In practice the value of B is unknown so the proposed estimator $\hat{\sigma}_v^{*2}$ becomes difficult to implement in practice. Thus we suggest a linear regression type estimator $\hat{\sigma}_{lr}^{*2}$ given by

$$\hat{\sigma}_{lr}^{*2} = s_y^{*2} + \hat{B}\left(S_x^2 - s_x^2\right) \tag{10}$$

where

$$\hat{B} = \frac{\hat{\mu}_{22}^{*} - s_y^{*2} s_x^2}{\hat{\mu}_{04} - s_x^4} \tag{11}$$

with $(n-1)s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $(n-1)\hat{\mu}_{04} = \sum_{i=1}^{n}(x_i - \bar{x})^4$, and

$$s_y^{*2} = \frac{1}{\theta^2}\left[s_z^2 - \frac{\gamma_2}{n(\gamma_2 + \theta^2)}\sum_{i=1}^{n} Z_i^2\right].$$

An unbiased estimator of μ_{22} is given in the following Lemma.

Lemma 4.1. An unbiased estimator of μ_{22} is given by:

$$\hat{\mu}_{22}^{*} = \frac{1}{(n-1)(\gamma_2+\theta^2)}\left[\sum_{i=1}^{n}(Z_i-\bar{Z})^2(x_i-\bar{x})^2 + \frac{\gamma_2}{\theta^2}\left\{\sum_{i\neq j=1}^{n}\sum^{n}\alpha_{ij}Z_iZ_j\right\}\right] \quad (12)$$

where

$$\alpha_{ij} = \frac{1}{n^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{(x_i-\bar{x})+(x_j-\bar{x})}{n}.$$

Proof. Available on request from the authors.

The value of $\hat{B} = (\hat{\mu}_{22}^{*} - s_y^{*2}s_x^2)\big/(\hat{\mu}_{04} - s_x^4)$ depends upon scrambled responses, thus it will increase the variance of the linear regression estimator $\hat{\sigma}_{lr}^{*2}$ compared to $\hat{\sigma}_v^{*2}$. It will also make it difficult to find the value of $V_R\left(\hat{\sigma}_{lr}^{*2}\right)$, which would be equal to:

$$V_R\left(\hat{\sigma}_{lr}^{*2}\right) = V_R(s_y^{*2}) + \frac{(S_x^2-s_x^2)^2}{(\hat{\mu}_{04}-s_x^4)^2}V_R\left(\hat{\mu}_{22}^{*}-s_y^{*2}s_x^2\right) + 2\frac{(S_x^2-s_x^2)}{(\hat{\mu}_{04}-s_x^4)}Cov_R\left(s_y^{*2},\hat{\mu}_{22}^{*}-s_y^{*2}s_x^2\right)$$

and in fact it makes it difficult to find the exact variance $V(\hat{\sigma}_{lr}^{*2})$ of the linear regression type estimator. Thus, in Section 6 we consider comparing the linear regression type estimator $\hat{\sigma}_{lr}^{*2}$ with the naive unbiased estimator s_y^{*2} through simulation study.

5. RELATIVE EFFICIENCY OF THE DIFFERENCE ESTIMATOR

The proposed difference type estimator $\hat{\sigma}_v^{*2}$ will be more efficient than the naive estimator s_y^{*2} if

$$Min.V\left(\hat{\sigma}_v^{*2}\right) < V(s_y^{*2})$$

that is, if

$$\left(\frac{1}{n}-\frac{1}{N}\right)\left(\mu_{40}-\mu_{20}^2\right)\left[1-\frac{(\mu_{22}-\mu_{20}\mu_{02})^2}{\left(\mu_{40}-\mu_{20}^2\right)\left(\mu_{04}-\mu_{02}^2\right)}\right] < \left(\frac{1}{n}-\frac{1}{N}\right)\left(\mu_{40}-\mu_{20}^2\right)$$

or equivalently if

$$\frac{(\mu_{22}-\mu_{20}\mu_{02})^2}{\left(\mu_{40}-\mu_{20}^2\right)\left(\mu_{04}-\mu_{02}^2\right)} > 0 \quad (13)$$

which is always true. Thus the proposed estimator $\hat{\sigma}_v^{*2}$ is always more efficient than the naive estimator s_y^{*2}. In order to look at the magnitude of the relative efficiency of the estimator $\hat{\sigma}_v^{*2}$ with respect to the naive estimator s_y^{*2}, we compute the percent relative efficiency defined as:

$$\mathrm{RE}(1) = \frac{V(s_y^{*2})}{V(\hat{\sigma}_v^{*2})} \times 100\% \tag{14}$$

Assume that the sensitive variable y_i and the auxiliary variable x_i are related to each other with the linear model defined by:

$$y_i = Rx_i + e_i x_i^g \tag{15}$$

where $e_i \sim N(0,1)$. The auxiliary variable $x_i \sim G(a,b)$ is generated from the gamma distribution by using the ISML subroutine RNGAM with parameters $a = 2.2$ and $b = 3.5$. We generate a population of size $N = 5000$ units from the model for a given value of g and R. We assumed the scrambling variable S has a beta distribution with parameters α and β. Thus, the first four moments of the scrambling variable S are given by:

$$E(S) = \theta = \frac{\alpha}{\alpha + \beta} \tag{16}$$

$$V(S) = \gamma_2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{17}$$

$$\gamma_3 = E(S-\theta)^3 = \gamma_2^{3/2}\left[\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{\sqrt{\alpha\beta}(\alpha+\beta+2)}\right] \tag{18}$$

and

$$\gamma_4 = E(S-\theta)^4 = \gamma_2^2\left[\frac{6\{\alpha^3 - \alpha^2(2\beta-1) + \beta^2(\beta+1) - 2\alpha\beta(\beta+2)\}}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} + 3\right] \tag{19}$$

We computed the value of the percent RE(1) for different values of sample sizes n, R, g, α and β as shown in Table 5.1. The FORTRAN code used to produce these results is available on request from the authors. We note that Δ and ρ_{xy} which appear in the Table 5.1 are given by:

$$\Delta = \frac{(\mu_{22} - \mu_{20}\mu_{02})}{\sqrt{\mu_{40} - \mu_{20}^2}\ \sqrt{\mu_{04} - \mu_{02}^2}} \quad \text{and} \quad \rho_{xy} = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}.$$

Table 5.1. RE of the difference estimator $\hat{\sigma}_v^{*2}$ over the naive estimator s_y^{*2}.

g	R	n	α	β	RE(1)	Δ	ρ_{xy}		g	R	n	α	β	RE(1)	Δ	ρ_{xy}
0.0	0.5	100	1.5	0.5	290.7	0.9408	0.9369		1.5	2.0	100	1.5	0.5	155.2	0.5629	0.3083
0.0	0.5	200	1.5	0.5	288.2	0.9408	0.9369		1.5	2.0	100	2.0	0.5	159.8	0.5629	0.3083
0.0	0.5	300	1.5	0.5	285.3	0.9408	0.9369		1.5	2.0	200	1.5	0.5	155.4	0.5629	0.3083
0.0	0.5	400	1.5	0.5	282.4	0.9408	0.9369		1.5	2.0	200	2.0	0.5	160.1	0.5629	0.3083
0.0	1.0	100	1.5	0.5	326.8	0.9844	0.9831		1.5	2.0	300	1.5	0.5	155.6	0.5629	0.3083
0.0	1.0	200	1.5	0.5	323.0	0.9844	0.9831		1.5	2.0	300	2.0	0.5	160.5	0.5629	0.3083
0.0	1.0	300	1.5	0.5	318.8	0.9844	0.9831		1.5	2.0	400	1.5	0.5	155.9	0.5629	0.3083
0.0	1.0	400	1.5	0.5	314.5	0.9844	0.9831		1.5	2.0	400	2.0	0.5	160.9	0.5629	0.3083
0.0	1.5	100	1.5	0.5	333.2	0.9930	0.9924		1.5	2.5	100	1.5	0.5	155.6	0.5645	0.3684

0.0	1.5	200	1.5	0.5	329.1	0.9930	0.9924		1.5	2.5	100	2.0	0.5	161.1	0.5645	0.3684
0.0	1.5	300	1.5	0.5	324.6	0.9930	0.9924		1.5	2.5	200	1.5	0.5	155.9	0.5645	0.3684
0.0	1.5	400	1.5	0.5	320.1	0.9930	0.9924		1.5	2.5	200	2.0	0.5	161.4	0.5645	0.3684
0.0	2.0	100	1.5	0.5	335.0	0.9960	0.9957		1.5	2.5	300	1.5	0.5	156.1	0.5645	0.3684
0.0	2.0	200	1.5	0.5	330.8	0.9960	0.9957		1.5	2.5	300	2.0	0.5	161.9	0.5645	0.3684
0.0	2.0	300	1.5	0.5	326.2	0.9960	0.9957		1.5	2.5	400	1.5	0.5	156.3	0.5645	0.3684
0.0	2.0	400	1.5	0.5	321.6	0.9960	0.9957		1.5	2.5	400	2.0	0.5	162.3	0.5645	0.3684
0.0	2.5	100	1.5	0.5	335.6	0.9975	0.9972		2.0	0.5	100	1.5	0.5	162.0	0.5854	0.0662
0.0	2.5	200	1.5	0.5	331.3	0.9975	0.9972		2.0	0.5	100	2.0	0.5	165.3	0.5854	0.0662
0.0	2.5	300	1.5	0.5	326.7	0.9975	0.9972		2.0	0.5	200	1.5	0.5	162.3	0.5854	0.0662
0.0	2.5	400	1.5	0.5	322.0	0.9975	0.9972		2.0	0.5	200	2.0	0.5	165.7	0.5854	0.0662
0.5	0.5	100	1.5	0.5	166.6	0.6340	0.6952		2.0	0.5	300	1.5	0.5	162.5	0.5854	0.0662
0.5	0.5	100	2.0	0.5	227.1	0.6340	0.6952		2.0	0.5	300	2.0	0.5	166.0	0.5854	0.0662
0.5	0.5	200	1.5	0.5	166.7	0.6340	0.6952		2.0	0.5	400	1.5	0.5	162.8	0.5854	0.0662
0.5	0.5	200	2.0	0.5	229.6	0.6340	0.6952		2.0	0.5	400	2.0	0.5	166.4	0.5854	0.0662
0.5	0.5	300	1.5	0.5	166.7	0.6340	0.6952		2.0	1.0	100	1.5	0.5	161.8	0.5849	0.0827
0.5	0.5	300	2.0	0.5	232.3	0.6340	0.6952		2.0	1.0	100	2.0	0.5	165.2	0.5849	0.0827
0.5	0.5	400	1.5	0.5	166.7	0.6340	0.6952		2.0	1.0	200	1.5	0.5	162.1	0.5849	0.0827
0.5	0.5	400	2.0	0.5	235.2	0.6340	0.6952		2.0	1.0	200	2.0	0.5	165.5	0.5849	0.0827
0.5	1.0	100	1.5	0.5	244.8	0.8349	0.8879		2.0	1.0	300	1.5	0.5	162.4	0.5849	0.0827
0.5	1.0	200	1.5	0.5	243.8	0.8349	0.8879		2.0	1.0	300	2.0	0.5	165.9	0.5849	0.0827
0.5	1.0	300	1.5	0.5	242.8	0.8349	0.8879		2.0	1.0	400	1.5	0.5	162.6	0.5849	0.0827
0.5	1.0	400	1.5	0.5	241.6	0.8349	0.8879		2.0	1.0	400	2.0	0.5	166.3	0.5849	0.0827
0.5	1.5	100	1.5	0.5	289.4	0.9121	0.9451		2.0	1.5	100	1.5	0.5	161.7	0.5844	0.0991
0.5	1.5	200	1.5	0.5	287.3	0.9121	0.9451		2.0	1.5	100	2.0	0.5	165.0	0.5844	0.0991
0.5	1.5	300	1.5	0.5	285.0	0.9121	0.9451		2.0	1.5	200	1.5	0.5	161.9	0.5844	0.0991
0.5	1.5	400	1.5	0.5	282.6	0.9121	0.9451		2.0	1.5	200	2.0	0.5	165.4	0.5844	0.0991
0.5	2.0	100	1.5	0.5	311.5	0.9465	0.9680		2.0	1.5	300	1.5	0.5	162.2	0.5844	0.0991
0.5	2.0	200	1.5	0.5	308.6	0.9465	0.9680		2.0	1.5	300	2.0	0.5	165.8	0.5844	0.0991
0.5	2.0	300	1.5	0.5	305.5	0.9465	0.9680		2.0	1.5	400	1.5	0.5	162.5	0.5844	0.0991
0.5	2.0	400	1.5	0.5	302.2	0.9465	0.9680		2.0	1.5	400	2.0	0.5	166.1	0.5844	0.0991
0.5	2.5	100	1.5	0.5	322.8	0.9643	0.9791		2.0	2.0	100	1.5	0.5	161.5	0.5839	0.1155
0.5	2.5	200	1.5	0.5	319.5	0.9643	0.9791		2.0	2.0	100	2.0	0.5	164.9	0.5839	0.1155
0.5	2.5	300	1.5	0.5	315.8	0.9643	0.9791		2.0	2.0	200	1.5	0.5	161.8	0.5839	0.1155
0.5	2.5	400	1.5	0.5	312.0	0.9643	0.9791		2.0	2.0	200	2.0	0.5	165.3	0.5839	0.1155
1.0	1.0	100	2.0	0.5	158.2	0.5416	0.5051		2.0	2.0	300	1.5	0.5	162.0	0.5839	0.1155
1.0	1.0	200	2.0	0.5	158.7	0.5416	0.5051		2.0	2.0	300	2.0	0.5	165.6	0.5839	0.1155
1.0	1.0	300	2.0	0.5	159.3	0.5416	0.5051		2.0	2.0	400	1.5	0.5	162.3	0.5839	0.1155
1.0	1.0	400	2.0	0.5	159.8	0.5416	0.5051		2.0	2.0	400	2.0	0.5	166.0	0.5839	0.1155
1.0	1.5	100	1.5	0.5	161.8	0.5948	0.6555		2.0	2.5	100	1.5	0.5	161.4	0.5833	0.1317
1.0	1.5	100	2.0	0.5	185.2	0.5948	0.6555		2.0	2.5	100	2.0	0.5	164.8	0.5833	0.1317
1.0	1.5	200	1.5	0.5	162.0	0.5948	0.6555		2.0	2.5	200	1.5	0.5	161.6	0.5833	0.1317
1.0	1.5	200	2.0	0.5	186.3	0.5948	0.6555		2.0	2.5	200	2.0	0.5	165.2	0.5833	0.1317
1.0	1.5	300	1.5	0.5	162.2	0.5948	0.6555		2.0	2.5	300	1.5	0.5	161.9	0.5833	0.1317
1.0	1.5	300	2.0	0.5	187.3	0.5948	0.6555		2.0	2.5	300	2.0	0.5	165.5	0.5833	0.1317
1.0	1.5	400	1.5	0.5	162.4	0.5948	0.6555		2.0	2.5	400	1.5	0.5	162.1	0.5833	0.1317
1.0	1.5	400	2.0	0.5	188.5	0.5948	0.6555		2.0	2.5	400	2.0	0.5	165.9	0.5833	0.1317
1.0	2.0	100	1.5	0.5	177.9	0.6490	0.7548		2.5	0.5	100	1.5	0.5	159.4	0.5779	0.0649
1.0	2.0	100	2.0	0.5	228.5	0.6490	0.7548		2.5	0.5	100	2.0	0.5	162.4	0.5779	0.0649
1.0	2.0	200	1.5	0.5	178.1	0.6490	0.7548		2.5	0.5	200	1.5	0.5	159.6	0.5779	0.0649
1.0	2.0	200	2.0	0.5	230.6	0.6490	0.7548		2.5	0.5	200	2.0	0.5	162.7	0.5779	0.0649
1.0	2.0	300	1.5	0.5	178.2	0.6490	0.7548		2.5	0.5	300	1.5	0.5	159.8	0.5779	0.0649
1.0	2.0	300	2.0	0.5	232.9	0.6490	0.7548		2.5	0.5	300	2.0	0.5	163.1	0.5779	0.0649
1.0	2.0	400	1.5	0.5	178.3	0.6490	0.7548		2.5	0.5	400	1.5	0.5	160.1	0.5779	0.0649
1.0	2.0	400	2.0	0.5	235.4	0.6490	0.7548		2.5	0.5	400	2.0	0.5	163.4	0.5779	0.0649
1.0	2.5	100	1.5	0.5	195.1	0.6991	0.8202		2.5	1.0	100	1.5	0.5	159.3	0.5778	0.0684

1.0	2.5	100	2.0	0.5	298.1	0.6991	0.8202		2.5	1.0	100	2.0	0.5	162.4	0.5778	0.0684
1.0	2.5	200	1.5	0.5	195.2	0.6991	0.8202		2.5	1.0	200	1.5	0.5	159.6	0.5778	0.0684
1.0	2.5	200	2.0	0.5	302.8	0.6991	0.8202		2.5	1.0	200	2.0	0.5	162.7	0.5778	0.0684
1.0	2.5	300	1.5	0.5	195.2	0.6991	0.8202		2.5	1.0	300	1.5	0.5	159.8	0.5778	0.0684
1.0	2.5	300	2.0	0.5	307.9	0.6991	0.8202		2.5	1.0	300	2.0	0.5	163.0	0.5778	0.0684
1.0	2.5	400	1.5	0.5	195.2	0.6991	0.8202		2.5	1.0	400	1.5	0.5	160.0	0.5778	0.0684
1.0	2.5	400	2.0	0.5	313.4	0.6991	0.8202		2.5	1.0	400	2.0	0.5	163.4	0.5778	0.0684
1.5	0.5	100	1.5	0.5	155.0	0.5626	0.1069		2.5	1.5	100	1.5	0.5	159.3	0.5776	0.0720
1.5	0.5	100	2.0	0.5	158.0	0.5626	0.1069		2.5	1.5	100	2.0	0.5	162.3	0.5776	0.0720
1.5	0.5	200	1.5	0.5	155.2	0.5626	0.1069		2.5	1.5	200	1.5	0.5	159.5	0.5776	0.0720
1.5	0.5	200	2.0	0.5	158.3	0.5626	0.1069		2.5	1.5	200	2.0	0.5	162.7	0.5776	0.0720
1.5	0.5	300	1.5	0.5	155.4	0.5626	0.1069		2.5	1.5	300	1.5	0.5	159.8	0.5776	0.0720
1.5	0.5	300	2.0	0.5	158.6	0.5626	0.1069		2.5	1.5	300	2.0	0.5	163.0	0.5776	0.0720
1.5	0.5	400	1.5	0.5	155.6	0.5626	0.1069		2.5	1.5	400	1.5	0.5	160.0	0.5776	0.0720
1.5	0.5	400	2.0	0.5	159.0	0.5626	0.1069		2.5	1.5	400	2.0	0.5	163.4	0.5776	0.0720
1.5	1.0	100	1.5	0.5	154.9	0.5622	0.1769		2.5	2.0	100	1.5	0.5	159.2	0.5775	0.0755
1.5	1.0	100	2.0	0.5	158.3	0.5622	0.1769		2.5	2.0	100	2.0	0.5	162.3	0.5775	0.0755
1.5	1.0	200	1.5	0.5	155.1	0.5622	0.1769		2.5	2.0	200	1.5	0.5	159.5	0.5775	0.0755
1.5	1.0	200	2.0	0.5	158.6	0.5622	0.1769		2.5	2.0	200	2.0	0.5	162.6	0.5775	0.0755
1.5	1.0	300	1.5	0.5	155.3	0.5622	0.1769		2.5	2.0	300	1.5	0.5	159.7	0.5775	0.0755
1.5	1.0	300	2.0	0.5	158.9	0.5622	0.1769		2.5	2.0	300	2.0	0.5	163.0	0.5775	0.0755
1.5	1.0	400	1.5	0.5	155.6	0.5622	0.1769		2.5	2.0	400	1.5	0.5	160.0	0.5775	0.0755
1.5	1.0	400	2.0	0.5	159.3	0.5622	0.1769		2.5	2.0	400	2.0	0.5	163.3	0.5775	0.0755
1.5	1.5	100	1.5	0.5	154.9	0.5622	0.2443		2.5	2.5	100	1.5	0.5	159.2	0.5774	0.0790
1.5	1.5	100	2.0	0.5	158.9	0.5622	0.2443		2.5	2.5	100	2.0	0.5	162.3	0.5774	0.0790
1.5	1.5	200	1.5	0.5	155.1	0.5622	0.2443		2.5	2.5	200	1.5	0.5	159.4	0.5774	0.0790
1.5	1.5	200	2.0	0.5	159.2	0.5622	0.2443		2.5	2.5	200	2.0	0.5	162.6	0.5774	0.0790
1.5	1.5	300	1.5	0.5	155.4	0.5622	0.2443		2.5	2.5	300	1.5	0.5	159.7	0.5774	0.0790
1.5	1.5	300	2.0	0.5	159.5	0.5622	0.2443		2.5	2.5	300	2.0	0.5	162.9	0.5774	0.0790
1.5	1.5	400	1.5	0.5	155.6	0.5622	0.2443		2.5	2.5	400	1.5	0.5	159.9	0.5774	0.0790
1.5	1.5	400	2.0	0.5	159.9	0.5622	0.2443		2.5	2.5	400	2.0	0.5	163.3	0.5774	0.0790

Clearly, the proposed difference type estimator remains more efficient than the naive estimator in many practical situations as shown in Table 5.1. The values of the parameters g and R have been chosen in such a way that the value of the correlation coefficient ρ_{xy} is made to change from a very high to a very low value so that the effect of a relationship between the auxiliary and the study variable may be examined. In Table 5.1 the value of ρ_{xy} ranges from 0.9972 to 0.0649. Note that when the value of ρ_{xy} is 0.0649 then the average relative efficiency RE(1) is 161.32% with a standard deviation of 1.75%; and when the value of ρ_{xy} increases to 0.9972 the average RE(1) is 328.90% with a standard deviation of 5.83%.

A graphical representation of the RE(1) versus the value of the correlation coefficient ρ_{xy} is given in Fig 5.1. It is clear that if the value of ρ_{xy} is more than 0.5 then the RE(1) goes on increasing. Thus the use of auxiliary information with higher correlation with the study variable helps to improve the estimator of finite population variance in the case of scrambled responses.

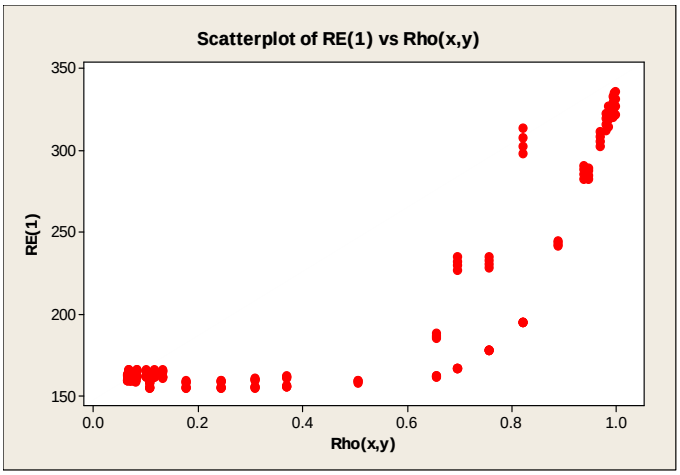

Fig. 5.1. RE(1) value versus the value of the correlation coefficient**.**

Figure 5.2 shows that as the value of ρ_{xy} increases beyond 0.5 the value of Δ also increases, which in fact helps the difference estimator of the finite population variance to produce efficient results.

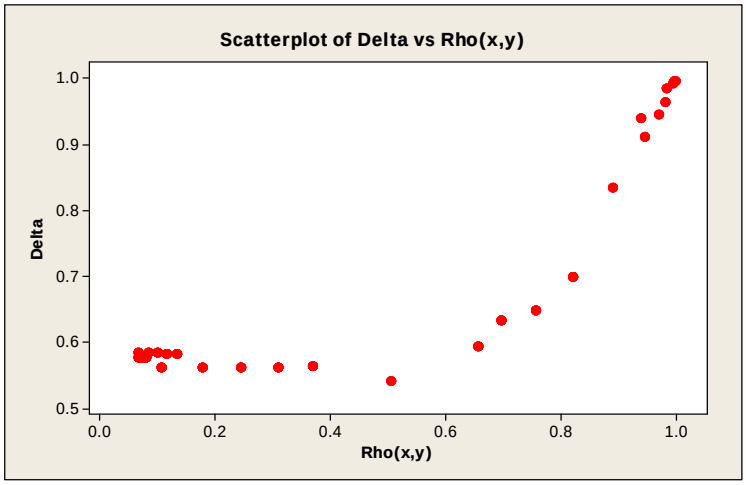

Fig. 5.2. The value of Δ and ρ_{xy}.

Table 5.2 has been designed to illustrate the simultaneous effect of g and R on the RE(1) value. For $g = 0.0$, as the value of R increases from 0.5 to 2.5, the RE(1) increases from 286.65% to 328.90%; for $g = 0.5$ as the value of R increases from 0.5 to 2.5 then the RE(1) increases from 198.90% to 317.51; for $g = 1.0$, as the value of R increases from 1.0 to 2.5, the RE(1) increases from 159.01% to 250.4%; Note that for $g = 1.0$ and $R = 0.5$, the RE(1) is missing indicating that the difference estimator performs less efficient in such a situation. For $g = 1.5$, as the value of R increases from 0.5 to 2.5, the RE(1) increases from 156.88% to 158.82%; for $g = 2.0$, as the value of R increases from 0.5 to 2.5, the RE(1) decreases from 164.13% to 163.54%; and for $g = 2.5$, as the value of R increases from 1.0 to 2.5, the RE(1) decreases from 161.32% to 161.16%.

Table 5.2. Simultaneous effect of g and R on RE(1).

			R		
g	0.5	1.0	1.5	2.0	2.5
0.0	286.65	320.76	326.76	328.38	328.90
0.5	198.90	243.25	286.08	306.95	317.51
1.0	-	159.01	174.46	205.00	250.40
1.5	156.88	156.99	157.32	157.92	158.82
2.0	164.13	163.97	163.82	163.67	163.54
2.5	161.32	161.28	161.24	161.20	161.16

Table 5.3. Simultaneous effect of g and R on ρ_{xy}.

			R		
g	0.5	1.0	1.5	2.0	2.5
0.0	0.93691	0.98309	0.99239	0.99570	0.99724
0.5	0.69515	0.88788	0.94512	0.96797	0.97914
1.0	-	0.50505	0.65550	0.75483	0.82018
1.5	0.10693	0.17688	0.24425	0.30827	0.36836
2.0	0.06620	0.08270	0.09913	0.11547	0.13173
2.5	0.06489	0.06843	0.07196	0.07549	0.07920

From Table 5.2 and Table 5.3, one could reach the same conclusion that, when the value of g is close to zero and the value of R is close to 2.0, the proposed difference estimator is likely to perform better than the naïve estimator. Note that for $n = 100$ the average RE(1) value is 195.87%; for $n = 200$ the average RE(1) value is 195.70%; for $n = 300$ the average RE(1) value is 195.49%; and for $n = 400$ the average RE(1) value is 195.29%. This indicates that the RE(1) value does not depend upon the value of the sample size. The value of β is fixed at 0.5, and, if the value of α changes from 1.5 to 2.0, the average RE(1) value decreases from 208.32% to 177.75%.

6. SIMULATION STUDY FOR THE REGRESSION ESTIMATOR

In this section, we again consider the case where the sensitive variable y_i and the auxiliary variable x_i are related to each other by the linear model defined as:

$$y_i = Rx_i + e_i x_i^g \qquad (20)$$

where $e_i \sim N(0,1)$. The auxiliary variable $x_i \sim G(a,b)$ is generated from the gamma distribution by using the ISML subroutine RNGAM with parameters $a = 2.2$ and $b = 3.5$. We generate a population of size $N = 5000$ units from the model for a given value of g and R. Then from the given population of size $N = 5000$, we used the ISML subroutine CALL RNSRI(NS, NP, IR) to select a SRSWOR sample of size n and both the study variable y_i and the auxiliary variable x_i are observed for $i = 1,2,3,...,n$. Next we used a subroutine CALL

RNBET(NS, A, B, S) to generate n values of a scrambling variable s_i, $i = 1,2,...,n$ from the beta distribution with a given choice of α and β. In other words, we assumed the scrambling variable $S \sim B(\alpha, \beta)$. Then we obtained the scrambled responses on the study variable as $z_i = y_i s_i$, $i = 1,2,...,n$ from a given sample. We repeated this process $T = 2000$ times. By using information from the ordered pairs $(x_i, z_i), i = 1,2,....,n$, of the t-th sample, $t = 1,2,...,T$, we computed the three estimators $\hat{\theta}_{0|t} = s^{*2}_{y|t}$; $\hat{\theta}_{1|t} = \hat{\sigma}^{*2}_{lr|t}$ and $\hat{\theta}_{2|t} = \hat{\sigma}^{*2}_{v|t}$. No doubt theoretically the estimators $\hat{\theta}_{0|t}$ and $\hat{\theta}_{2|t}$ are unbiased estimators of the parameter of interest S_y^2, while the estimator $\hat{\theta}_{1|t}$ is a biased estimator. In order to see the performance of the proposed estimators, we computed the simulated relative bias in each of the three estimators as follows:

$$RB\left(\hat{\theta}_j\right) = \frac{\frac{1}{T}\sum_{t=1}^{T}\hat{\theta}_{j|t} - S_y^2}{S_y^2} \times 100\% \qquad (21)$$

We also computed the relative efficiencies of the linear regression and the difference type estimator with respect to the usual estimator as:

$$RE(0, j) = \frac{\sum_{t=1}^{T}\left[\hat{\theta}_{0|t} - S_y^2\right]^2}{\sum_{t=1}^{T}\left[\hat{\theta}_{j|t} - S_y^2\right]^2}, \text{ for } j = 1,2. \qquad (22)$$

We simulated the values for different parameters g, R, α, and β and the sample size n such that the absolute value of the simulated relative bias remained less than 10% for all three of the estimators, and such that the percent relative efficiency $RE(0,1)$ remained more than 100% and the value of the percent relative efficiency $RE(0,2)$ remained more than 125%. We also counted whether any of the three estimates $\hat{\theta}_{0|t}$, $\hat{\theta}_{1|t}$ and $\hat{\theta}_{2|t}$ took on negative values, but fortunately none of the cases led to negative estimates. The results so obtained are presented in Table 6.1. The FORTRAN Code used in this simulation is available on request from the authors. The values of $RB(\hat{\theta}_0)$ and $RB(\hat{\theta}_2)$ are supposed to be equal to zero, but bias due to simulation remains because we have not selected all possible samples. Also the randomized response arising from the Beta distribution is also used only once to scramble the dataset in a given sample. No doubt the $RB(\hat{\theta}_1)$ reflects the relative bias as that is expected from a linear regression type estimator. For all choices of g, R, α, and β, the percent relative efficiency of the difference type estimator remains higher than that of the linear regression type estimator. The difference in the percent relative efficiencies of the two estimators is clearly observed through our simulation in Table 6.1.

Table 6.1. Comparison of three estimators of the finite population variance.

g	R	α	β	n	$RE(0,1)$	$RE(0,2)$	$RB(\hat{\theta}_1)$	$RB(\hat{\theta}_2)$	$RB(\hat{\theta}_0)$
0.0	0.5	1.0	0.5	200	198.8	851.1	-2.844	1.374	-4.010
0.0	1.0	1.0	0.5	200	198.2	679.2	-3.078	1.663	-4.333
0.0	1.0	2.0	0.5	200	119.0	914.2	-8.211	-2.963	-8.958
0.0	1.5	1.0	0.5	200	193.9	732.4	-3.211	1.652	-4.471
0.0	1.5	2.0	0.5	200	117.7	803.7	-8.719	-3.337	-9.460
0.0	2.0	1.0	0.5	200	191.0	783.0	-3.288	1.624	-4.545
0.0	2.0	2.0	0.5	200	117.2	759.1	-8.946	-3.515	-9.684
0.0	2.5	1.0	0.5	200	189.2	823.2	-3.338	1.600	-4.590
0.0	2.5	2.0	0.5	200	116.9	735.0	-9.073	-3.618	-9.808
0.5	0.5	1.0	0.5	200	118.7	255.0	-8.557	-5.838	-9.322
0.5	0.5	1.5	0.5	200	116.9	891.6	-4.845	-1.754	-5.238
0.5	0.5	2.0	0.5	200	149.5	552.2	-1.999	1.040	-2.444
0.5	0.5	0.5	1.0	300	222.9	360.3	-2.332	-1.834	3.481
0.5	0.5	1.0	1.0	300	221.3	720.9	5.693	3.155	8.470
0.5	0.5	1.5	1.0	300	146.4	578.0	7.522	3.785	9.100
0.5	0.5	1.5	1.5	300	182.3	774.3	2.896	-1.405	3.910
0.5	0.5	2.0	1.0	300	128.1	720.9	7.482	3.154	8.469
0.5	0.5	1.0	1.0	400	148.6	376.7	6.745	4.237	8.223
0.5	0.5	1.5	1.0	400	126.5	365.7	7.429	4.369	8.355
0.5	0.5	2.0	1.0	400	120.0	440.8	6.950	3.626	7.612
0.5	1.0	0.5	1.0	200	192.1	597.0	2.610	1.481	-3.618
0.5	2.0	1.5	1.5	400	207.2	216.2	2.795	-2.736	4.024
1.0	0.5	0.5	1.5	300	148.9	595.9	6.917	3.458	8.440
1.0	0.5	1.5	2.0	300	109.4	843.3	7.265	2.617	7.600
1.0	0.5	1.5	1.0	400	110.3	272.7	9.023	5.738	9.474
1.0	0.5	2.0	1.0	400	109.7	377.5	7.351	3.963	7.700
1.0	1.0	0.5	1.0	400	164.1	336.5	7.632	5.330	9.777
1.0	1.0	1.5	1.0	400	116.9	380.8	8.437	4.675	9.122
1.0	1.0	2.0	1.0	400	114.8	575.8	7.117	3.178	7.625
1.0	1.5	2.0	0.5	200	121.0	179.8	-2.337	1.917	-2.571
1.0	1.5	2.0	2.0	200	120.2	714.2	-6.540	-2.683	-7.171
1.0	1.5	0.5	2.0	400	114.6	451.4	9.060	4.566	9.700
1.0	1.5	1.0	1.0	400	148.2	921.6	6.290	2.522	7.656
1.0	2.0	1.5	0.5	200	108.3	472.8	-8.821	-4.222	-9.181
1.0	2.0	2.0	0.5	200	135.8	193.1	-2.474	2.075	-2.883
1.0	2.5	1.0	0.5	200	127.6	476.5	-8.655	-4.479	-9.776
1.0	2.5	2.0	0.5	200	139.5	343.0	-2.912	1.858	-3.440
1.0	2.5	1.5	1.0	300	264.8	472.5	3.402	-2.547	5.535
1.0	2.5	2.0	1.0	300	173.1	494.5	4.237	-2.507	5.575
1.0	2.5	1.5	1.0	400	211.3	149.8	2.295	-2.725	3.336
1.5	0.5	0.5	1.0	300	108.5	623.9	5.643	-2.353	5.877
1.5	1.0	1.0	1.0	300	103.4	571.3	5.784	-2.461	5.882
1.5	1.5	1.0	1.0	300	108.8	866.4	6.056	-2.146	6.316
1.5	2.0	0.5	2.0	300	104.7	248.4	5.131	-3.332	5.251
1.5	2.5	0.5	2.0	300	118.5	353.1	5.217	-3.023	5.680
1.5	2.5	1.5	1.0	300	111.8	139.3	4.457	-3.991	4.711
1.5	2.5	1.5	1.5	400	103.7	914.1	9.577	3.225	9.752

The value of RE(0,2) remains consistently higher than the value of RE(0,1), as expected. In all the situations listed in the table the absolute value of the percent relative bias remains less than 10% which is acceptable by following Cochran (1977). A closer study of results indicates that for $g = 0$, the average RE(0,1) value is 160.2% with a standard deviation of 40.4% and the average value of RE(0,2) is 786.8% with a standard deviation of 70.7%; for $g = 0.5$, the average

RE(0,1) value is 160.0% with a standard deviation of 40.0% and the average value of RE(0,2) is 526.9% with a standard deviation of 210.5%; for $g = 1.0$, the average RE(0,1) value is 141.03% with a standard deviation of 41.26% and the average value of RE(0,2) is 458.4% with a standard deviation of 214.4%; and for $g = 1.5$, the average RE(0,1) value is 108.49% with a standard deviation of 5.39% and the average value of RE(0,2) is 531.00% with a standard deviation of 298%. As the value of g increases the value of the correlation coefficient ρ_{xy} decreases as shown in table 5.1. Hence the average value of RE(0,1) decreases from 160.2% to 108.49% as the value of g increases from 0.0 to 1.5. If $R = 0.5$ then the average value of RE(0,1) is 145.11% with a standard deviation of 38.74% and the average value of RE(0,2) is 564.8% with a standard deviation of 211.0%; if $R = 1.0$ then the average value of RE(0,1) is 144.10% with a standard deviation of 39.80% and the average value of RE(0,2) is 579.3% with a standard deviation of 191.9%; if $R = 1.5$ then the average value of RE(0,1) is 132.10% with a standard deviation of 30.00% and the average value of RE(0,2) is 667.10% with a standard deviation of 262.60%; if $R = 2$ then the average value of RE(0,1) is 144.00% with a standard deviation of 44.30% and the average value of RE(0,2) is 445.00% with a standard deviation of 271.00%; and if $R = 2.5$ then the average value of RE(0,1) is 155.60% with a standard deviation of 52.70% and the average value of RE(0,2) is 490.10% with a standard deviation of 264.7%. Finally if $n = 200$ then the average value of RE(0,1) is 145.87% with a standard deviation of 34.83% and that of RE(0,2) is 618.70% with a standard deviation of 236.4%; if $n = 300$ then the average value of RE(0,1) is 150.20% with a standard deviation of 51.70% and that of RE(0,2) is 557.5% with a standard deviation of 214.9%; and if $n = 400$ then the average value of RE(0,1) is 138.10% with a standard deviation of 34.83% and that of RE(0,2) value is 444.60% with a standard deviation of 235.4%.

7. COMPLEX SURVEY DESIGN

Let a sample s of size n be selected from a population by using an arbitrary sampling design $P(s)$ with inclusion probabilities π_i for the ith unit and π_{ij} for the ith and jth unit $i \neq j = 1,2,...,N$. Assuming π_{ij}'s are positive for all $i \neq j$, an unbiased estimator of the population variance is given

$$S_z^2 = \frac{1}{N(\gamma_2 + \theta^2)} \sum_{i \in s} \frac{z_i^2}{\pi_i} - \frac{1}{N(N-1)\theta^2} \sum_{i \neq} \sum_{j \in s} \frac{z_i z_j}{\pi_{ij}}$$

Das and Tripathi (1978) and Isaki (1983) type variance estimator for the complex surveys design can be obtained as:

$$S_z^{*2} = S_z^2 + B^* \left(S_x^2 - \hat{S}_x^2 \right)$$

where $\hat{S}_x^2 = \frac{1}{N} \sum_{i \in s} \frac{x_i^2}{\pi_i} - \frac{1}{N(N-1)} \sum_{i \neq} \sum_{j \in s} \frac{x_i x_j}{\pi_{ij}}$ and $B*$ is a suitably chosen constant.

REFERENCES

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* A Chapman & Hall, CRC Press, Boca Raton, FL.

Cochran, W.G. (1977). *Sampling Techniques.* 3rd Ed., John Wiley and Sons, NY.

Das, A.K. and Tripathi, T.P. (1980). Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known. *Sankhy* $\bar{a}$ C,42,76--86.

Diana, G. and Perri, P.F. (2009a). Estimating a sensitive proportion through randomized response procedures based on auxiliary information. *Statistical Papers*, 50, 661-672.

Diana, G. and Perri, P.F. (2009b). A class of estimators for quantitative sensitive data. *Statistical Papers* (DOI: 10.1007/ s00362 - 009 - 0273 - 1) (Published online: August 20, 2009)

Diana, G. and Perri, P.F. (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, 37, 1875-1890.

Diana, G. and Perri, P.F. (2011). A calibration-based approach to sensitive data: a simulation study. *Journal of Applied Statistics*, (In press, obtained through personal contacts)

Eichhorn, B.H. and Hayre, L.S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *J. Statist. Planning Infer.*,**7**, 307--316.

Greenberg, B.G., Kuebler, R.R., Abernathy, J.R. and Horvitz, D.G. (1971). Application of the randomized response technique in obtaining quantitative data. *J. Amer. Statist. Assoc.*, 66, 243--250.

Horvitz, D.G., Shah, B.V., and Simmons, W.R. (1967). The unrelated question randomized response model. *Proc. Social. Statist. Sect*., ASA, 65-72.

Isaki, C.T. (1983).Variance estimation using auxiliary information. *J. Amer. Statist. Assoc.* ,78, 117--123.

Singh, S., Joarder, A.H. and King, M.L. (1996). Regression analysis using scrambled responses. *Australian Journal of Statistics*, 38, 201-211.

Singh, S. and Kim, Jong-Min (2011). Pseudo empirical log-likelihood estimator using scrambled responses. *Statistics and Probability Letters*, 81, 345-351.

Son, C.-K., Hong, K.-H., Lee, G.-S., and Kim, J.-M. (2008). The calibration of stratified randomized response estimators. *Communications of the Korean Statistical Society*, 14, 597-603.

Strachan, R., King, M.L. and Singh, S. (1998).Likelihood-based estimation of the general model with scrambled responses. *Australian and New Zealand Journal of Statistics*, 40, 279-290.

Strauss, I. (1982). On the admissibility of estimators for the finite population variance. *Metrika*, 29, 195-202.

Tracy, D.S. and Singh, S. (1999). Calibration estimators in randomized response survey. *Metron*, LVII, 47-68.

Warner, S.L. (1965): Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Stat. Assoc.*, 60, 63-69.