REVISTA INVESTIGACION OPERACIONAL VOL. 40, NO. 3, 374-385, 2019

A MODIFIED EFFICIENT RANDOMIZED RESPONSE MODEL

Housila P. Singh and Swarangi M. Gorey
School of Studies in Statistics
Vikram University, Ujjain-456010
M.P., India

ABSTRACT
In this paper, the problem of estimating the population total of a sensitive quantitative variable has been considered. Odumade and Singh (2009) have given a RR model to estimate the population total of sensitive quantitative variable. We have given a modified RR model. It is noticed that the RR model envisaged by Odumade and Singh (2009) suffers with a difficulty that the optimum value of the constant involved in their RR model depends on the value of unknown parameter of the study variable, which lacks the utility of their RR model. To overcome such a difficulty Odumade and Singh (2009) have also suggested another estimator of the population total which depends on two RR models based on two independent random samples s_1 and s_2 from the population Ω using the sampling design $p(s_1)$and $p(s_2)$ respectively; and also on a relation between the two constants involved in the two RR models. We should add here that drawing the two independent samples from a population constructing two RR models resultantly obtaining an estimator for a population total Y increase the cost of the survey. Keeping this in view we have suggested a RR model and developed an estimator for population total based on a single sample. The proposed RR model is free from the difficulty involved in Odumade and Singh (2009) model. We have shown theoretically that the proposed randomized response model is more efficient than Bar-Lev et al (2004) and Odumade and Singh (2009) RR models. This fact has also been supported through numerical illustration.

KEYWORDS: Randomized response model; Unequal probability sampling scheme; Simple random sampling without replacement scheme;Simple random sampling with replacement.

MSC: 62D05.

RESUMEN
En este paper, el problema de estimar el total de la población de una variable sensitiva y cuantitativaes considerado. Odumade y Singh (2009) desarrollaron un modelo de RR para estimar el total de la población. Nosotros presentamosuna modificación delmodelo RR. Como se nota, el modelo RR esbozado por Odumade y Singh (2009) tiene la dificultad de obtener que el óptimo valor de la constante, envuelta en su modelo de RR depende del valor desconocido del parámetros de la variable de estudio, lo que lastra la utilidad de su modelo. Para superar esto Odumade and Singh (2009) también sugirieron otro estimador del total de la población que depende de dos muestras independientes s_1 and s_2, de la población Ω, usando respectivamente diseños muestrales $p(s_1)$ y $p(s_2)$; y también de la relación entre dos constantes envueltas en los dos modelos de RR. Nosotros debemos apuntar que la selección de dos muestras independientes de la población resulta en la construcción de dos modelos de RR para obtener un estimador del total poblacional Y incrementa el costo del encuestaje . Teniendo esto a la vista sugerimos un modelo de RR y desarrollamos un estimador del total de la población basado en una sola muestra. Tal propuesta no tiene la dificultad que aparece en el modelo de Odumade y Singh (2009) . Probamos teóricamente que el modelo de respuesta aleatorizadas propuesto es mas eficinte que los dr Bar-Lev et al (2004) y Odumade y Singh (2009). Este hecho es soportado con ilustraciones numéricas.

PALABRAS CLAVE: modelo de respuestas aleatorizadas, esquema de probabilidades desiguales , esquema de muestreo simple aleatorio sin reemplazo, muestreo simple aleatorio con reemplazo**.**

1. INTRODUCTION

In survey methodology whenever the study variable is sensitive in nature either because it pertains to something that is too personal or stigmatizing or illegal, randomized response (RR) techniques are used to collect the data. A typical Randomized Response method was proposed by Warner (1965)to protect survey responder's privacy and to thus reduce a major source of bias (evasive answers or refusing to respond) in estimating the prevalence of sensitive characteristics in surveys of human populations.

Since its introduction there have been several extensions to the theory and use of the RR procedure (see Horvitz et al., (1967); Greenberg et al., (1969); Moors, (1997); Mangat and Singh, (1990); Kuk (1990); Mangat, (1994), Nayak (1994); Bhargava (1996);,Zou (1997); Bhargava and Singh (2001, 2002); Gjestvang and Singh (2006); Kim and Elam (2005); and Kim and Warde (2005)). Eichhorn and Hayre(1983), which is further studied by Arnab (1995, 1996), suggested a multiplicative model to collect information on sensitive quantitative variables like income, tax evasion, amount of drug used etc. According to them, each respondent in the sample is requested to report the scrambled response $Z_i = SY_i$, where Y_i is the real value of the sensitive quantitative variable, and S is the scrambling variable whose distribution is assumed to be known. In other words, $E_R(S) = \theta$ and $V_R(S) = \gamma^2$ are assumed to be known and positive. Then an unbiased estimator of the population total under the simple random and with replacement (SRSWR) sampling is given by:

$$\hat{y}_{EH} = \frac{N}{n}\sum_{i=1}^{n}\frac{Z_i}{\theta} \tag{1.1}$$

with variance:

$$V(\hat{y}_{EH}) = \frac{N^2\bar{Y}^2}{n}\left[C_y^2 + C_\gamma^2\left(1 + C_y^2\right)\right], \tag{1.2}$$

where $C_\gamma^2 = \gamma^2/\theta^2$, $\bar{Y} = Y/N$ and $C_y = \sigma_y/\bar{Y}$.

Review of Bar-Lev et al (2004) and Odumade and Singh (2009) RR Models

- **Bar-Lev, Bobovitch, and Boukai(2004) RR Model**

Bar-Lev et al (2004) proposed a quantitative randomized response (RR) procedure which generalizes that of Eichhorn and Hayre (1983). In BBB model, the distribution of the responses is given by

$$Z_i = \begin{cases} Y_i S & \text{with probability } (1-p) \\ Y_i & \text{with probability } p. \end{cases} \tag{1.3}$$

In other words, each respondent is requested to rotate a spinner unobserved by the interviewer, and if the spinner stops in the shaded area, then the respondent is requested to report the real response on the sensitive variable, say Y_i; and if the spinner stops in the non-shaded area, then the respondent is requested to report the scrambled response, say $Y_i S$, where S is any scrambling variable and its distribution is assumed to be known. Assume that $E_R(S) = \theta$ and $V_R(S) = \gamma^2$ are known. Let p be the proportion of the shaded area of the spinner and $(1-p)$ be the non-shaded area of the spinner as shown in Figure 1.

An unbiased estimator of population total Y is given by:

$$\hat{Y}_{BBB} = \frac{N}{n\{(1-p)\theta + p\}}\sum_{i=1}^{n} Z_i \tag{1.4}$$

with variance under SRSWR sampling given by

$$V\left(\hat{Y}_{BBB}\right) = \frac{N^2\bar{Y}^2}{n}\left[C_y^2 + (1 + C_y^2)C_p^2\right], \tag{1.5}$$

where

$$C_p^2 = \frac{(1-p)\theta^2(1 + C_\gamma^2) + p}{[(1-p)\theta + p]^2} - 1. \tag{1.6}$$

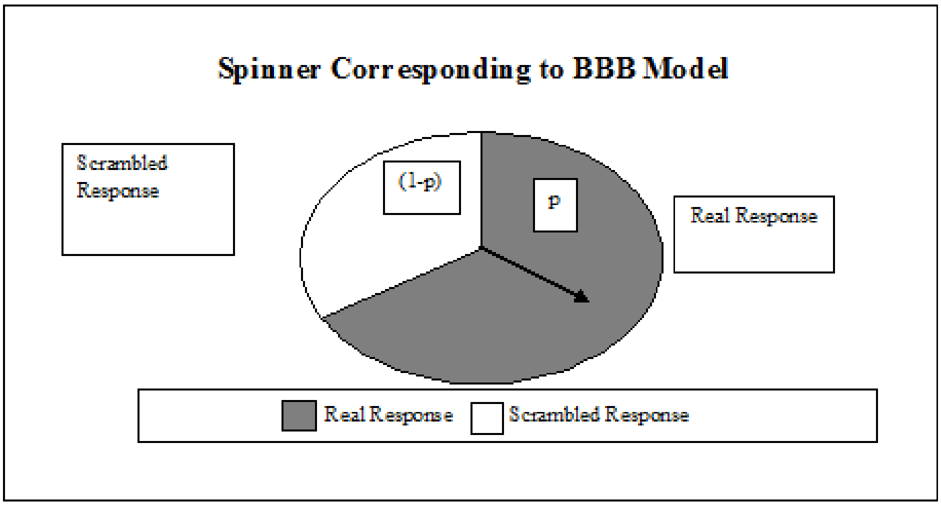

Figure. 1: Bar-Lev, Bobovich and Boukai (2004, BBB) randomized response device.

- **Odumade and Singh (2009) RR Model**

In this randomized response model, the distribution of the responses is given by

$$Z_i = \begin{cases} Y_i S + k & \text{with probability } (1-p) \\ Y_i & \text{with probability } p \end{cases} \tag{1.7}$$

In other words, each respondent is requested to rotate a spinner unobserved by the interviewer, and if the spinner stops in the shaded area, then the respondent is requested to report the real response on the sensitive variable, say Y_i; and if the spinner stops in the non-shaded area, then the respondent is requested to report the scrambled response, say $Y_i S + k$, where *S* is any scrambling variable and its distribution is assumed to be known, and k is assumed to be known constant . Assume that $E_R(S) = \theta$ and $V_R(S) = \gamma^2$ are known. Let p be the proportion of the shaded area of the spinner and (1-p) be the non shaded area of the spinner as shown in Figure 2.

An unbiased estimator of the population total *Y* is given by

$$\hat{Y}_{OS} = \frac{\sum_{i \in s} d_i \left[Z_i - k(1-p) \right]}{\{p + (1-p)\theta\}}, \tag{1.8}$$

with the variance

$$V(\hat{Y}_{OS}) = \frac{1}{2} \sum_{i \neq j} \sum_{\in \Omega} \Theta_{ij} \left(d_i Y_i - d_j Y_j \right)^2 + C_p^2 \sum_{i \in \Omega} d_i Y_i^2 + \frac{p(1-p)}{(p + \theta(1-p))^2} \left[k^2 \sum_{i \in \Omega} d_i + 2k(\theta - 1) \sum_{i \in \Omega} d_i Y_i \right], \tag{1.9}$$

where $d_i = \pi_i^{-1}$, $\pi_i = Pr(i \in s)$, $i \in \Omega$ be the probability of including the *ith* unit from the population Ω with sample *s* with probability design *p(s)* $\Theta_{ij} = (\pi_i \pi_j - \pi_{ij})$ and $\pi_{ij} = \Pr(i, j \in s)$ denote the probability of including both i^{th} and j^{th} units in the sample.

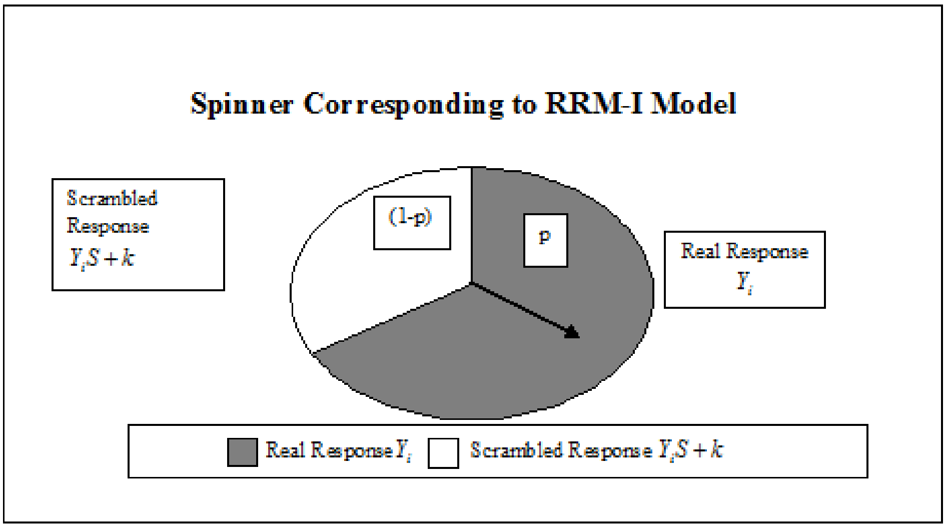

Figure 2: Odumade and Singh (2009) randomized response model

The variance of $\hat{Y}_{OS}$ at (1.9) is minimum when

$$k = \frac{(\theta - 1)\sum_{i\in\Omega} d_i Y_i}{\sum_{i\in\Omega} d_i}. \quad (1.10)$$

Thus the resulting minimum MSE of the estimator $\hat{Y}_{OS}$ is given by

$$min.V(\hat{Y}_{OS}) = \frac{1}{2}\sum_{i\neq j}\sum_{\in\Omega}\Theta_{ij}(d_iY_i - d_jY_j)^2 + C_p^2\sum_{i\in\Omega} d_iY_i^2 - \frac{\psi\left(\sum_{i\in\Omega} d_iY_i\right)^2}{\sum_{i\in\Omega} d_i}, \quad (1.11)$$

where $\psi = \frac{p(1-p)(\theta-1)^2}{\{p+(1-p)\theta\}^2}$.

It is observed from (1.10) that the optimum value of k depends on the unknown parameter of the study variable y which lacks the utility of the Odumade and Singh (2009) estimator $\hat{Y}_{OS}$ defined by (1.8). To overcome this difficulty we have suggested an alternative randomized response model. The proposed model is free from such a difficulty and more efficient than the Bar-Lev et al (2004) and Odumade and Singh (2009) randomized response models.

2. PROPOSED RANDOMIZED RESPONSE MODEL

In the proposed model we request each respondent to rotate a spinner as demonstrated in Figure 3.
In the proposed randomized response model, the distribution of responses is given by

$$Z_i = \begin{cases} Y_i[S - p(\theta - 1)] & with\ probability\ (1-p) \\ Y_i[(1-p)\theta + p] & with\ probability\ p. \end{cases} \quad (2.1)$$

In other words, each respondent is requested to rotate a spinner unobserved by the interviewer, and if the spinner stops in the shaded area, then the respondent is requested reported the scrambled response $Y_i[(1-p)\theta + p]$, and if the spinner stops in the non-shaded area, then the respondent is requested to report the scrambled response $Y_i[S - p(\theta - 1)]$, where S is any scrambling variable and its distribution is assumed

to be known. Assume that $E_R(S)=\theta$ and $V_R(S)=\gamma^2$ are known. Let *p* be the proportion of the shaded area of the spinner and *(1-p)* be the non-shaded area of the spinner, as shown in Figure 3.
Here we note that the proposed randomized response model at (2.1) can be used in practice without any difficulty as the mean θ of the scrambling variable *S* and probability *p* are known.

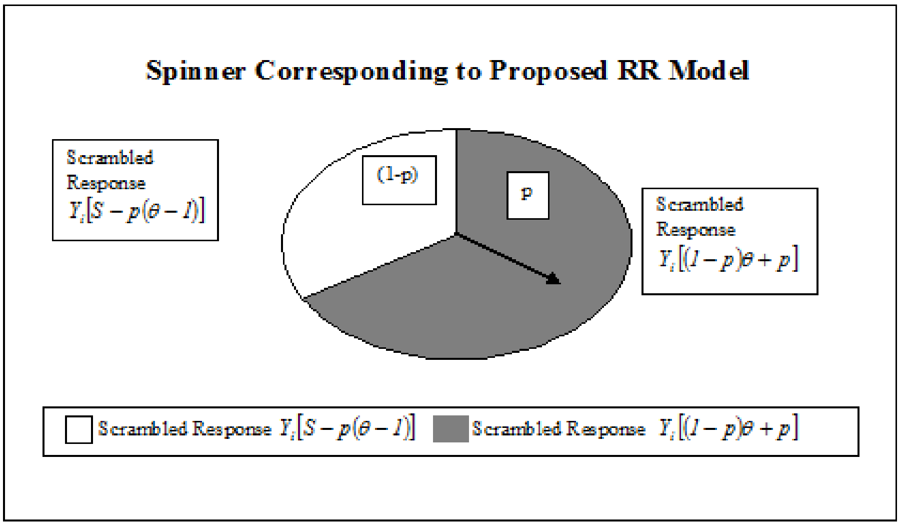

Figure 3: Proposed randomized response device

Consider a population Ω consisting of *N* units. Let $Y_i,\ i=1,2,\ldots,N,$ be the value of the *ith* population unit of the sensitive quantitative variable. It is desired to estimator the population total $Y=\sum_{i\in\Omega}Y_i$. Let $\pi_i=P_r(i\in s), i\in\Omega$ be the probability of including the *ith* unit from the population Ω in the sample *s* with probability design *p(s)*. Then we have the following theorems.

Theorem 2.1- An unbiased estimator of the population total *Y* based on proposed randomized response model at (2.1) is given by

$$\hat{Y}_{SG}=\frac{\sum_{i\in s}d_iZ_i}{[p+(1-p)\theta]} \quad (2.2)$$

where $d_i=\pi_i^{-1}$.

Proof- Let E_p and E_R be the expected values over the design *p(s)* and the randomization device, say spinner, thus we have

$$E(\hat{Y}_{SG})=E_pE_R(\hat{Y}_{SG})$$

$$=E_pE_R\sum_{i\in s}d_i\left\{\frac{Z_i}{p+(1-p)\theta}\right\}$$

$$=E_p\sum_{i\in s}d_i\left[\frac{Y_i\{(\theta-p(\theta-1))(1-p)+((1-p)\theta+p)p\}}{(p+(1-p)\theta)}\right]$$

$$=E_p\left[\sum_{i\in s}d_iY_i\right]=\sum_{i\in\Omega}Y_i=Y$$

which shows that the proposed estimator $\hat{Y}_{SG}$ is unbiased for the population total *Y*. Thus the theorem is proved.

Theorem 2.2- The variance of the proposed estimator $\hat{Y}_{SG}$ is given by

$$V\left(\hat{Y}_{SG}\right)=\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\left(\pi_i\pi_j-\pi_i\pi_j\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2+\left(C_p^2-\psi\right)\sum_{i\in\Omega}d_iY_i^2 \quad ,(2.3)$$

where $\pi_{ij}=P_r\left(i,j\in s\right)$ denotes the probability of including both *ith* and *jth* units in the sample, and

$$\psi=\frac{p(1-p)(\theta-1)^2}{\{p+(1-p)\theta\}^2}.$$

Proof- Let V_p and V_R denote, respectively, the variance over the design *p(s)* and over the randomization device, say, spinner, we have

$$V\left(\hat{Y}_{SG}\right)=E_pV_R\left(\hat{Y}_{SG}\right)+V_pE_R\left(\hat{Y}_{SG}\right),$$

$$=E_pV_R\left[\sum_{i\in s}d_i\left\{\frac{Z_i}{p+(1-p)\theta}\right\}\right]+V_pE_R\left[\sum_{i\in s}d_i\left\{\frac{Z_i}{p+(1-p)\theta}\right\}\right],$$

$$=E_p\left[\sum_{i\in s}d_i^2\frac{V_R(Z_i)}{(p+(1-p)\theta)^2}\right]+V_p\left[\sum_{i\in s}d_iY_i\right]. \quad (2.4)$$

Note that

$$V_R(Z_i)=E_R\left(Z_i^2\right)-\left(E_R(Z_i)\right)^2$$

Now

$$E_R\left(Z_i^2\right)=(1-p)Y_i^2\left[E_R\left(S^2\right)+p^2(\theta-1)^2-2p(\theta-1)E_R(S)\right]$$
$$+pY_i^2\left[(1-p)^2\theta^2+p^2+2p(1-p)\theta\right]$$
$$=Y_i^2\left[(1-p)\left\{\left(\theta^2+\gamma^2\right)+p^2(\theta-1)^2-2p\theta(\theta-1)\right\}\right.$$
$$\left.+p\left\{(1-p)^2\theta^2+p^2+2p(1-p)\theta\right\}\right]$$
$$=Y_i^2\left[(1-p)\left\{\theta^2\left(1+C_\gamma^2\right)+p^2(\theta-1)^2-2p\theta(\theta-1)\right\}\right.$$
$$\left.+p\left\{(1-p)^2\theta^2+p^2+2p(1-p)\theta\right\}\right]$$

So we have

$$V_R(Z_i)=Y_i^2\left[(1-p)\left\{\theta^2\left(1+C_\gamma^2\right)+p^2(\theta-1)^2-2p\theta(\theta-1)\right\}\right.$$
$$\left.+p\left\{(1-p)^2\theta^2+p^2+2p(1-p)\theta\right\}-\{p+(1-p)\theta\}^2\right]$$
$$=Y_i^2\left[\left\{(1-p)\theta^2\left(1+C_\gamma^2\right)+p\right\}-p(1-p)(\theta-1)^2-\{p+(1-p)\theta\}^2\right]$$
$$=Y_i^2\{p+(1-p)\theta\}^2\left[\frac{\left\{(1-p)\theta^2\left(1+C_\gamma^2\right)+p\right\}}{\{p+(1-p)\theta\}^2}-1-\frac{p(1-p)(\theta-1)^2}{\{p+(1-p)\theta\}^2}\right]$$
$$=Y_i^2\{p+(1-p)\theta\}^2\left(C_p^2-\psi\right). (2.5)$$

Thus we have

$$V\left(\hat{Y}_{SG}\right)=\sum_{i\in\Omega} d_i \frac{V_R(Z_i)}{\left(p+(1-p)\theta\right)^2}+\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2$$

$$=\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2+\left(C_p^2-\psi\right)\sum_{i\in\Omega} d_i Y_i^2$$

which proves the theorem.

Under SRSWOR sampling, $\pi_i=\frac{n}{N}$ and $\pi_{ij}=\frac{n(n-1)}{N(N-1)}$.The variance of the proposed estimator $\hat{Y}_{SG}$ under SRSWOR sampling is given by

$$V\left(\hat{Y}_{SG}\right)=\frac{N^2\bar{Y}^2}{n}\left[(1-f)C_y^2+\left\{1+\left(1-\frac{1}{N}\right)C_y^2\right\}\left(C_p^2-\psi\right)\right] \quad (2.6)$$

where $f=\frac{n}{N}$ denotes the sampling fraction.

3. EFFICIENCY COMPARISON

Putting $k=0$, then the variance in (1.9) reduces to the variance of the estimator $\hat{Y}_{BBB}=\frac{\sum_{i\in s} d_i Z_i}{\{p+(1-p)\theta\}}$ based on BBB model under unequal probability sampling design is given by

$$V\left(\hat{Y}_{BBB}\right)=\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2+C_p^2\sum_{i\in\Omega} d_i Y_i^2 \quad .(3.1)$$

From (2.3) and (3.1) we have

$$V\left(\hat{Y}_{BBB}\right)-V\left(\hat{Y}_{SG}\right)=\left[C_p^2\sum_{i\in\Omega} d_i Y_i^2-\left(C_p^2-\psi\right)\sum_{i\in\Omega} d_i Y_i^2\right]$$

$$=\psi\sum_{i=\Omega} d_i Y_i^2 \quad (3.2)$$

which is always positive. Thus the proposed model is always superior to the BBB model.
From (1.11) and (2.3) we have

$$min.V\left(\hat{Y}_{OS}\right)-V\left(\hat{Y}_{SG}\right)=\left[C_p^2\sum_{i\in\Omega} d_i Y_i^2-\psi\frac{\left(\sum_{i\in\Omega} d_i Y_i\right)^2}{\left(\sum_{i\in\Omega} d_i\right)}-\left(C_p^2-\psi\right)\sum_{i\in\Omega} d_i Y_i^2\right]$$

$$=\psi\left\{\sum_{i\in\Omega} d_i Y_i^2-\frac{\left(\sum_{i\in\Omega} d_i Y_i\right)^2}{\left(\sum_{i\in\Omega} d_i\right)}\right\}=\psi\sum_{i\in\Omega} d_i\left\{Y_i-\frac{\left(\sum_{i\in\Omega} d_i Y_i\right)}{\left(\sum_{i\in\Omega} d_i\right)}\right\}^2 \quad (3.3)$$

which is always positive. Thus the proposed randomized response model is superior to the Odumade and Singh (2009) randomized response model.

To look at the relative efficiency of the suggested estimator $\hat{Y}_{SG}$ under simple random sampling without replacement (SRSWOR) sampling scheme with respect to the BBB model under simple random sample with replacement (SRSWOR) scheme and the Odumade and Singh (2009) model under simple random sample without replacement sampling, we have resorted to some empirical experiments for different choices of parameters.

4. EMPIRICAL COMPARISONS

Under SRSWR sampling the variance of the unbiased estimator $\hat{Y}_{BBB}$ of the population total Y due to Bar-Lev et al (2004) is given by

$$V\left(\hat{Y}_{BBB}\right)=\frac{N^2\overline{Y}^2}{n}\left[C_y^2+\left(1+C_y^2\right)C_p^2\right] \quad (4.1)$$

Under SRSWOR, the minimum variance of the Odumade and Singh (2009) estimator $\hat{Y}_{OS}$ is given by

$$min.V\left(\hat{Y}_{OS}\right)=\frac{N^2\overline{Y}^2}{n}\left[(1-f)C_y^2+\left\{1+\left(1-\frac{1}{N}\right)C_y^2\right\}C_p^2-\psi\right]. \quad (4.2)$$

Using (2.6) and (4.1), the percent relative efficiency (PRE) of the proposed estimator $\hat{Y}_{SG}$ with respect to Bar-Lev et al (2004) estimator $\hat{Y}_{BBB}$ can be expressed as

$$PRE\left(\hat{Y}_{SG},\hat{Y}_{BBB}\right)=\frac{V\left(\hat{Y}_{BBB}\right)_{SRSWR}}{V\left(\hat{Y}_{SG}\right)_{SRSWOR}}\times 100$$

$$=\frac{\left[C_y^2+\left(1+C_y^2\right)C_p^2\right]}{\left[(1-f)C_y^2+\left\{1+\left(1-\frac{1}{N}\right)C_y^2\right\}\left(C_p^2-\psi\right)\right]}\times 100. \quad (4.3)$$

If $N\to\infty$, then $f\to 0$ and $\frac{1}{N}\to 0,$ then the percent relative efficiency in (4.3) reduces to :

$$PRE\left(\hat{Y}_{SG},\hat{Y}_{BBB}\right)=\frac{\left[C_y^2+\left(1+C_y^2\right)C_p^2\right]}{\left[C_y^2+\left\{1+C_y^2\right\}\left(C_p^2-\psi\right)\right]}\times 100. \quad (4.4)$$

Using (2.6) and (4.2), the PRE of the proposed estimator $\hat{Y}_{SG}$ with respect to Odumade and Singh (2009) estimator $\hat{Y}_{OS}$ is given by

$$PRE\left(\hat{Y}_{SG},\hat{Y}_{OS}\right)=\frac{V\left(\hat{Y}_{OS}\right)_{SRSWOR}}{V\left(\hat{Y}_{SG}\right)_{SRSWOR}}\times 100$$

$$=\frac{\left[(1-f)C_y^2+\left\{1+\left(1-\frac{1}{N}\right)C_y^2\right\}C_p^2-\psi\right]}{\left[(1-f)C_y^2+\left\{1+\left(1-\frac{1}{N}\right)C_y^2\right\}\left(C_p^2-\psi\right)\right]}\times 100. \quad (4.5)$$

If $N\to\infty$, then $f\to 0$ and $\frac{1}{N}\to 0,$ then the percent relative efficiency in (4.5) reduces to :

$$PRE\left(\hat{Y}_{SG},\hat{Y}_{OS}\right)=\frac{\left[(1-f)C_y^2+\left(1+C_y^2\right)C_p^2-\psi\right]}{\left[(1-f)C_y^2+\left(1+C_y^2\right)\left(C_p^2-\psi\right)\right]}\times 100. \quad (4.6)$$

It is observed from PRE formulae (4.3)-(4.6) that these formulae depend on four parameters C_γ, p, C_y and θ. Odumade and Singh (2009) have mentioned that in any real survey the value of coefficient of variations of the scrambling variable and the study variable is expected to lie between 0.1 and 0.9, and the value of p in the randomization device may vary from 0.7 to 0.9. The value of θ depends upon the choice of the investigator based on the nature of the sensitive variable under study. Thus we have computed the $PRE\left(\hat{Y}_{SG}, \hat{Y}_{BBB}\right)$ and $PRE\left(\hat{Y}_{SG}, \hat{Y}_{BBB}\right)$ for various choices of parameters as displayed in Table 1. We have considered the values of parameters in numerical illustration similar to those as taken by Odumade and Singh (2009) as $C_\gamma = 0.1(0.2)0.9$, $C_y = 0.1(0.2)0.9$, p=0.7,0.9 and $\theta = 2, 20, 200, 20{,}000.$

Table 1: PREs of the proposed estimator $\hat{Y}_{SG}$ with respect to Bar-Lev et al 's (2004) estimator $\hat{Y}_{BBB}$.

p	C_γ	C_y	θ	PRE	θ	PRE	θ	PRE	θ	PRE
0.7	0.1	0.1	2	830.88	20	4710.04	200	5414.55	20000	5496.12
		0.3		238.58		1645.09		2060.33		2112.66
		0.5		160.00		844.84		1070.50		1099.70
		0.7		136.99		574.93		724.49		744.02
		0.9		127.33		456.10		570.15		585.10
	0.3	0.1		268.36		774.20		844.86		852.85
		0.3		184.83		622.59		700.79		709.82
		0.5		147.09		483.30		557.71		566.58
		0.7		131.64		396.57		462.89		470.96
		0.9		124.30		345.43		404.75		412.07
	0.5	0.1		166.30		349.01		373.88		376.68
		0.3		147.78		324.91		351.68		354.73
		0.5		132.92		294.49		322.54		325.77
		0.7		124.54		269.36		297.46		300.74
		0.9		119.88		251.35		278.89		282.14
	0.7	0.1		134.73		227.96		240.56		241.98
		0.3		128.86		221.28		234.47		235.97
		0.5		122.68		211.85		225.68		227.26
		0.7		118.36		203.06		217.27		218.90
		0.9		115.62		196.10		210.46		212.12
	0.9	0.1		121.24		177.64		185.24		186.10
		0.3		118.89		175.13		182.96		183.85
		0.5		116.03		171.40		179.53		180.45
		0.7		113.75		167.71		176.07		177.03
		0.9		112.15		164.64		173.15		174.12
0.9	0.1	0.1	2	663.20	20	6822.99	200	8140.84	20000	8287.71
		0.3		186.61		3068.74		4785.47		5027.15
		0.5		136.59		1660.52		2898.39		3097.90
		0.7		122.39		1126.31		2040.71		2196.93
		0.9		116.50		880.34		1613.55		1742.42
	0.3	0.1		287.58		982.10		1078.27		1089.01
		0.3		166.22		856.57		999.87		1015.81
		0.5		132.37		715.11		896.69		917.97
		0.7		120.74		610.39		807.65		832.14
		0.9		115.58		541.22		741.63		767.63
	0.5	0.1		180.37		422.22		454.87		458.54
		0.3		145.02		403.80		444.00		448.45
		0.5		126.32		378.12		427.77		433.28
		0.7		118.08		354.51		411.64		418.08
		0.9		114.03		336.06		398.13		405.27
	0.7	0.1		143.27		265.06		281.44		283.28
		0.3		130.41		260.09		278.56		280.61
		0.5		120.55		252.66		274.08		276.45
		0.7		115.15		245.27		269.42		272.09
		0.9		112.20		239.06		265.35		268.27
	0.9	0.1		126.79		200.02		209.86		210.96
		0.3		121.23		198.17		208.79		209.98
		0.5		115.90		195.33		207.11		208.42
		0.7		112.47		192.39		205.33		206.76
		0.9		110.40		189.84		203.74		205.28

Table 2: PREs of the proposed estimator $\hat{Y}_{SG}$ with respect to Odumade and Singh's (2009) estimator $\hat{Y}_{OS}$

s	C_{γ}	C_y	θ	PRE	θ	PRE	θ	PRE	θ	PRE
0.7	0.1	0.1	2	107.24	20	145.64	200	152.62	20000	153.43
		0.3		111.44		227.58		261.86		266.18
		0.5		112.00		248.97		294.10		299.94
		0.7		112.16		256.18		305.37		311.79
		0.9		112.23		259.36		310.40		317.09
	0.3	0.1		101.67		106.68		107.37		107.45
		0.3		107.00		143.15		149.61		150.35
		0.5		109.42		176.66		191.54		193.32
		0.7		110.40		197.53		219.34		221.99
		0.9		110.87		209.83		236.38		239.66
	0.5	0.1		100.66		102.47		102.71		102.74
		0.3		103.94		118.57		120.78		121.03
		0.5		106.58		138.90		144.51		145.15
		0.7		108.07		155.70		164.94		166.01
		0.9		108.90		167.73		180.05		181.51
	0.7	0.1		100.34		101.27		101.39		101.41
		0.3		102.38		110.01		111.10		111.23
		0.5		104.54		122.37		125.14		125.45
		0.7		106.04		133.89		138.56		139.10
		0.9		106.99		143.00		149.43		150.17
	0.9	0.1		100.21		100.77		100.84		100.85
		0.3		101.56		106.20		106.85		106.92
		0.5		103.21		114.28		115.91		116.09
		0.7		104.52		122.27		125.02		125.33
		0.9		105.44		128.93		132.74		133.17
0.9	0.1	0.1	2	105.58	20	166.56	200	179.61	20000	181.07
		0.3		107.15		345.13		486.87		506.83
		0.5		107.32		412.10		659.68		699.58
		0.7		107.36		437.51		738.22		789.60
		0.9		107.38		449.21		777.34		835.01
	0.3	0.1		101.86		108.73		109.69		109.79
		0.3		105.47		162.47		174.30		175.62
		0.5		106.47		223.02		259.34		263.59
		0.7		106.82		267.85		332.72		340.77
		0.9		106.97		297.45		387.14		398.77
	0.5	0.1		100.80		103.19		103.51		103.55
		0.3		103.72		125.08		128.40		128.77
		0.5		105.26		155.62		165.55		166.66
		0.7		105.94		183.70		202.49		204.60
		0.9		106.28		205.64		233.42		236.61
	0.7	0.1		100.43		101.63		101.80		101.81
		0.3		102.51		113.22		114.74		114.91
		0.5		104.11		130.53		134.82		135.29
		0.7		104.98		147.77		155.72		156.59
		0.9		105.46		162.23		174.00		175.30
	0.9	0.1		100.27		100.99		101.09		101.10
		0.3		101.75		108.11		108.98		109.08
		0.5		103.18		119.07		121.42		121.68
		0.7		104.10		130.38		134.64		135.11
		0.9		104.65		140.21		146.43		147.11

It is observed from Table 1 that :

(i) for fixed values of (p, C_{γ}, θ), the $PRE(\hat{Y}_{SG}, \hat{Y}_{BBB})$ decreases as the coefficient of variation C_y of the study variable Y increases. Larger Gain in efficiency is observed when coefficient of variation (C_y, C_{γ}) is small which is expected too. Similar trend is observed for fixed values of (p, C_{γ}, θ), and increasing value of C_{γ}.

(ii) for fixed values of (p, C_γ, C_y), the $PRE(\hat{Y}_{SG}, \hat{Y}_{BBB})$ increases considerably as the value of θ increases .Higher gain in efficiency is observed when both the coefficients of variation (C_y, C_γ) of the study variable Y and scrambling variable s are small.
(iii) for fixed values of $(C_y, C_\gamma, \theta = 2)$, the $PRE(\hat{Y}_{SG}, \hat{Y}_{BBB})$ decreases as p increases while for $\theta = 20,200.20,000$ the trend is reverse.
Table 2 shows that
(i) for fixed values of (p, C_γ, θ), the $PRE(\hat{Y}_{SG}, \hat{Y}_{OS})$ increases as the coefficient of variation C_y of the study variable Y increases. The gain in efficiency by using the proposed estimator $\hat{Y}_{SG}$ over the estimator $\hat{Y}_{OS}$ due to Odumade and Singh (2009) is substantial when the value of coefficient of variation C_y of the study variable Y is large.
(ii) for fixed values of (p, C_y, θ), the $PRE(\hat{Y}_{SG}, \hat{Y}_{OS})$ decreases as the value of coefficient of variation C_γ of the scrambling variable s increases. Larger in gain in efficiency is observed when C_γ is small.
(iii) for fixed values of (p, C_γ, C_y), the $PRE(\hat{Y}_{SG}, \hat{Y}_{OS})$ increases substantially for increasing value of θ.
(iv) for fixed value of $(C_y, C_\gamma, \theta = 2)$, the $PRE(\hat{Y}_{SG}, \hat{Y}_{OS})$ decreases as the value of p increases while for $\theta = 20,200.20,000$ the opposite trend is observed.
It is observed from Tables 1 and 2 that the larger gain in efficiency is observed by using the proposed estimator $\hat{Y}_{SG}$ over Bar-Lev et al (2004) estimator $\hat{Y}_{BBB}$ as compared to Odumade and Singh's (2009) estimator $\hat{Y}_{OS}$. We further note from Tables 1 and 2 that the values of $PRE(\hat{Y}_{SG}, \hat{Y}_{BBB})$ and $PRE(\hat{Y}_{SG}, \hat{Y}_{OS})$ are larger than 100%. Thus the proposed estimator $\hat{Y}_{SG}$ (i.e. proposed randomized response model) is more efficient than the Bar-Lev et al (2004) estimator $\hat{Y}_{BBB}$ (i.e. the randomized response model due to Bar-Lev et al (2004)) and Odumade and Singh (2009) estimator $\hat{Y}_{OS}$ (i.e. randomized response model due to Odumadenad Singh (2009)). Thus proposed randomized response model (or estimator $\hat{Y}_{SG}$) is recommended for its use in practice without any reservation.

5. CONCLUSION

The problem of estimation of population total Y of the sensitive quantitative variable y is an important issue. Odumade and Singh (2009) have tackled this issue via defining randomized response technique and then proposed an unbiased estimator for the population total Y. The optimal estimator due to Odumade and Singh (2009) depends on the unknown population parameter under investigation which prevents the use of their estimator in practice. To overcome this difficulty we have proposed an unbiased estimator of the population total through defining a randomized response model. The proposed estimator is free from such a difficulty. We have compared the proposed estimator with that of Bar-Lev et al (2004) and Odumade and Singh (2009) estimators. Theoretically and empirically. It is found that the proposed randomized model is more efficient than the Bar-Lev et al (2004) and Odumade and Singh (2009) randomized response models. So our recommendation is in the favor of the proposed randomized response model in practice.

RECEIVED: AUGUST, 2017
REVISED: FEBRUARY, 2018

REFERENCES:

[1] ARNAB, R. (1995): On admissibility and optimality of sampling strategies in randomized response surveys. **Sankhya:** Series B, 57:385–390.

[2] ARNAB, R. (1996): Randomized response trials: a unified approach for qualitative data. **Communications in Statistics Theory and Methods,**25(6), 1173–1183.
[3] BAR-LEV, S. K., BOBOVITCH, E. and BOUKAI, B. (2004): A Note On Randomized Response Models For Quantitative Data. **Metrika**, 60,255-260.
[4] BHARGAVA, M. (1996):**An Investigation into the Efficiencies of Certain Randomized ResponseStrategies.**Unpublished Ph.D. thesis submitted to Punjab Agricultural University,Ludhiana, India.
[5] BHARGAVA, M., and SINGH, R. (2001): Efficiency comparison of certain randomized responseschemes with U-model.**Journal of the Indian Society of Agricultural Statistics**,54,19–28.
[6] BHARGAVA, M. and SINGH, R. (2002): On the efficiency comparison of certain randomized response strategies. **Metrika**, 55, 191-197.
[7] EICHHORN, B. H., and HAYRE, L. S. (1983): Scrambled Randomized Response Methods For Obtaining Sensitive Quantitative Data. **Journal of Statistical Planning and Inference**, 7,307-316.
[8] GJESTVANG, C. R., and SINGH, S. (2006): A new randomized response model. **Journal of the RoyalStatistical Society B,**68,523–530.
[9] GREENBERG, B.G., ABUL-ELA, A.L.A., SIMONS, W.R., and HORVITZ, D.G. (1969): The unrelated question randomizedresponse model: Theoretical frame work.**Journal of American Statistical Association**, 64,520–539.
[10] HORVITZ, D.G., SHAH, B.V. and SIMMONS, W.R. (1967): The unrelated question randomized response model.**Proceeding of Social Statistics Section. American Statistical Association**, 65–72.
[11] KIM, J.-M., and ELAM, M. E. (2005): A two-stage stratified Warner's randomized response modelusing optimal allocation. **Metrika**,61,1–7.
[12] KIM, J.-M. and WARDE, W. D. (2005): A mixed randomized response model. **Journal of StatisticalPlanning and Inference**,133,211–221.
[13] KUK, A. Y. C. (1990): Asking sensitive questions indirectly. Biometrika,77,436–438.
[14] MANGAT, N. S. and SINGH, R. (1990): An alternative randomized response procedure. **Biometrika**,77, 439–442.
[15] MANGAT, N. S. (1994): An improved randomized response strategy. **Journal of the RoyalStatistical Society B**,56,93–95.
[16] MOORS, J.J.A. (1997):**A critical evaluation of Mangat's two-step procedure in randomized response.**A Discussion paper at Center for Economic Research, Tilburg University, The Netherlands.
[17] NAYAK, T. K. (1994): On randomized response surveys for estimating a proportion.**Communication Statistics—Theory and Methods**, 23,3303–3321.
[18] ODUMADE, O. AND SINGH, S. (2009): Improved Bar-Lev, Bobovitch, And Boukai Randomized Response Models, **Communication InStatistics- Simulation and Computation**,38, 473-502.
[19] WARNER, S.L. (1965): Randomized Response: A Survey Technique For Eliminating Evasive Answer Bias. **Journal of AmericanStatistical Association**, 60, 63-69.
[20] ZOU, G. (1997): Two-stage randomized response procedures as single stage procedures.**Australian Journal of Statistics,**39,235–236.