Generalized Forced Quantitative Randomized Response Model

Oluseun Odumade and Sarjinder Singh

Department of Statistics and Computer Networking, St. Cloud State University, St. Cloud, MN 56301-4498, USA.
E-mail: sarjinder@yahoo.com

Abstract

A new generalized forced quantitative randomized response (GFQRR) model for estimating the population total of a sensitive variable is proposed and studied under a unified setup. The bias and variance expressions are derived under unequal probability sampling design. It is shown that the models due to Eichhorn and Hayre (1983), Bar-Lev, Bobovitch, and Boukai (2004), Liu and Chow (1976a, 1976b), Stem and Steinhorst (1984), and Gjestvang and Singh (2005) are special cases of the proposed GFQRR model. Numerical illustrations are carried out to show the performance of the proposed GFQRR model.

Keywords: Randomized response sampling, estimation of population total, sensitive quantitative variable.

1. Introduction

The problem of estimation of the population total of a sensitive quantitative variable is well known in survey sampling. Warner (1965) was the first to suggest an ingenuous method to estimate the proportion of sensitive characters like induced abortions, drug used etc., through a randomization device like a deck of cards, spinners etc. such that the respondents' privacy should be protected. A rich growth of literature can be found in Fox and Tracy (1986), and Tracy and Mangat (1996). Mangat and Singh (1990) proposed a two-stage randomized response model. Leysieffer and Warner (1976), and Lanke (1975, 1976) studied different randomized response procedures at equal level of protection of the respondents, and later Nayak (1994), Bhargava (1996), Zou (1997), Bhargava and Singh (2001, 2002) and Moors (1997) found that Mangat and Singh (1990) and Warner (1965) models remain equally efficient at equal protection. Note that this result is not true for all the randomized response models. Bhargava (1996), the detail is available in Singh (2003) on page no. 939-941, shows that Mangat (1994) model remains more efficient than Warner (1965) model at equal protection. Note that Mangat (1994) model is a special case of Kuk (1990) model. Mangat (1994) model is further improved and studied by Gjestvang and Singh (2006). Eichorn and Hayre (1983) 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 estimator of the population total $Y = \sum_{i\in\Omega} Y_i$ under the simple random and with replacement (SRSWR) sampling is given by:

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

with variance:

$$V(\hat{Y}_{\text{EH}}) = \frac{N^2}{n}\sigma_y^2 + \frac{N^2}{n}C_\gamma^2\bar{Y}^2(1+C_y^2) \tag{1.2}$$

where $C_\gamma^2 = \gamma^2/\theta^2$, $\bar{Y} = Y/N$ and $C_y = \sigma_y/\bar{Y}$. We shall now discuss a randomized response model recently studied by Bar-Lev, Bobovitch, and Boukai (2004), which we say BBB model hereafter. 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 $E(S) = \theta$ and $V(S) = \gamma^2$ 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 1.1**.

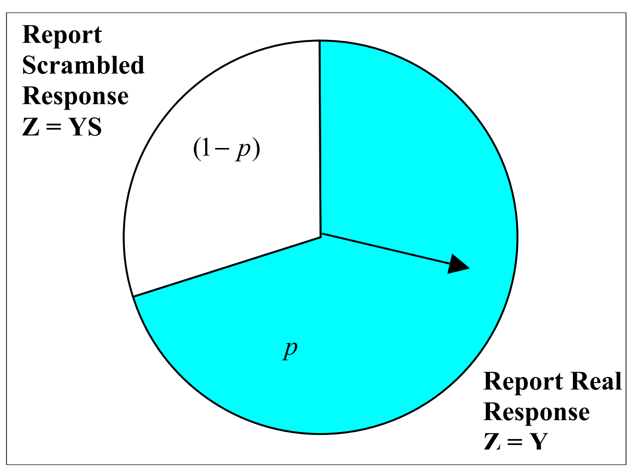


Fig. 1.1. BBB randomized response device.

An unbiased estimator of population total Y is given by:

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

with variance under SRSWR sampling given by:

$$V[\hat{Y}_{(BBB)}] = \frac{N^2}{n} \bar{Y}^2 [C_y^2 + (1 + C_y^2) C_s^2(p)] \quad (1.5)$$

where

$$C_s^2(p) = \frac{(1-p)\theta^2(1+C_\gamma^2) + p}{[(1-p)\theta + p]^2} - 1$$

In the next section, we suggest a new generalized forced quantitative randomized response (GFQRR) model.

2. Proposed GFQRR model

Consider a population Ω consisting of N units. Let Y_i, $i = 1,2,...,N$, be the value of the i^th^ population unit of the sensitive quantitative variable. Our aim is to estimate the population total $Y = \sum_{i\in\Omega} Y_i$. Let $\pi_i, i \in \Omega$ be the probability of including the i^th^ unit from the population Ω in the sample s with probability design $p(s)$. The i^th^ respondent selected in the sample is requested to rotate a spinner having three statements:

(i) report the real value of the sensitive variable, Y_i, with probability p_1

(ii) report the scrambled response SY_i, with probability p_2

(iii) report the fixed response F, with probability p_3.

where S is a scrambling variable and its distribution is assumed to be known. In other words, if E_R is the expected value and V_R is the variance over the randomization device used in a survey, then $E_R(S) = \theta$ and $V_R(S) = \gamma^2$ are assumed to be positive and known. Conclusively, the distribution of the i^th^ response is given by:

$$Z_i = \begin{cases} Y_i, & \text{with probability } p_1 \\ SY_i, & \text{with probability } p_2 \\ F, & \text{with probability } p_3 \end{cases} \quad (2.1)$$

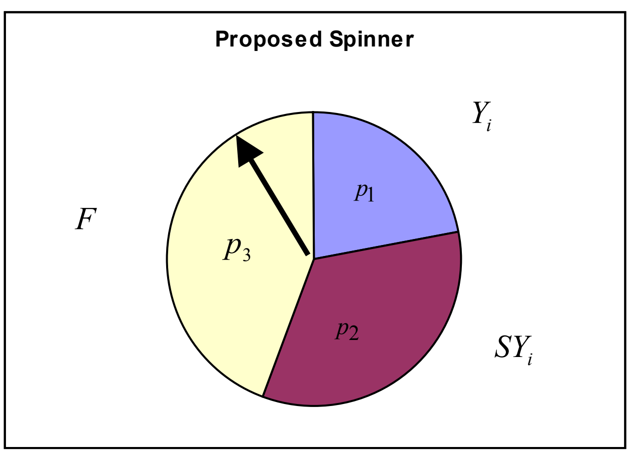


Fig. 2.1. GFQRR spinner.

Consequently, we have the following theorem.

Theorem 2.1. An unbiased estimator of the population total Y is given by:

$$\hat{Y}_p = \sum_{i\in s} d_i \left(\frac{Z_i - p_3 F}{p_1 + p_2\theta} \right) \quad (2.2)$$

where $d_i = \pi_i^{-1}$ are called design weights.

Proof. Let E_p and E_R be the expected values over the design p and the randomization device, say spinner, thus we have:

$$E(\hat{Y}_p) = E_p E_R \left[\sum_{i\in s} d_i \left(\frac{Z_i - p_3 F}{p_1 + p_2\theta} \right) \right] = E_p(\sum_{i\in s} d_i Y_i) = Y$$

which proves the theorem.

Theorem 2.2. The minimum variance of the proposed estimator $\hat{Y}_p$ is given by:

$$\min V(\hat{Y}_p) = \frac{1}{2} \sum_{i\neq j\in\Omega} \sum \Theta_{ij} (d_i Y_i - d_j Y_j)^2 + \frac{1}{(p_1 + p_2\theta)^2} \left[\{p_1 + p_2(\gamma^2 + \theta^2) - (p_1 + p_2\theta)^2\} \sum_{i\in\Omega} d_i Y_i^2 - \frac{(p_1 + p_2\theta)^2 (\sum_{i\in\Omega} d_i Y_i)^2}{(1-p_3)(\sum_{i\in\Omega} d_i)} \right] \quad (2.3)$$

where $\Theta_{ij} = (\pi_i \pi_j - \pi_{ij})$.

Proof. Let V_R and V_p denote the variance over the randomization device, say spinner, and over the design, we have:

$$V(\hat{Y}_p) = V_p E_R \left[\sum_{i\in s} d_i \left(\frac{Z_i - p_3 F}{p_1 + p_2\theta} \right) \right] + E_p V_R \left[\sum_{i\in s} d_i \left(\frac{Z_i - p_3 F}{p_1 + p_2\theta} \right) \right]$$

$$= V_p \left[\sum_{i\in s} d_i Y_i \right] + E_p \sum_{i\in s} d_i^2 \left(\frac{V_R(Z_i)}{(p_1 + p_2\theta)^2} \right)$$

$$= \frac{1}{2} \sum_{i\neq j\in\Omega} \sum \Theta_{ij} (d_i Y_i - d_j Y_j)^2 + \frac{1}{(p_1 + p_2\theta)^2} \left[\{p_1 + p_2\theta^2(1 + C_\gamma^2) - (p_1 + p_2\theta)^2\} \sum_{i\in\Omega} d_i Y_i^2 + p_3(1-p_3)F^2 \sum_{i\in\Omega} d_i - 2p_3 F(p_1 + p_2\theta) \sum_{i\in\Omega} d_i Y_i \right] \quad (2.4)$$

On differentiating (2.4) with respect to F and setting equal to zero, we have:

$$F = \frac{(p_1 + p_2\theta) \sum_{i\in\Omega} d_i Y_i}{(1-p_3) \sum_{i\in\Omega} d_i} \quad (2.5)$$

On substituting (2.5) into (2.4) we get (2.3), it proves the theorem.

In the next section, we show that the BBB, Eichhorn and Hayre (1983) and Liu ad Chow (1976a, 1976b) models are special cases of the proposed GFQRR model.

2.1 Special Cases

Case I. If $p_1 = 0$, $p_2 = 1$, and $p_3 = 0$, the proposed GFQRR model reduces to the Eichhorn and Hayre (1983) model.
Case II. If $p_1 = p$, $p_2 = (1-p)$, and $p_3 = 0$, the proposed GFQRR model reduces to the BBB model.
Case III. Note that a quantitative forced alternative randomization device, due to Liu and Chow (1976a, 1976b), is valid only for estimating the proportion of a sensitive attribute in a population unlike the proposed model, which estimates the average of a quantitative sensitive variable. Interestingly, note that if X_i is a qualitative variable, take 1 and 0 value for a sensitive and non-sensitive attribute in the population, set $Z = 0$ as forced "no" answer, and set $F = 1$ as forced "yes" answer, then the present model is reduced to an optimized forced alternative randomizing device proposed by Stem and Steinhorst (1984).

2.2. Relative efficiency

Under simple random and without replacement (SRSWOR) sampling, we have $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/N(N-1)$. The percent relative efficiency (RE) of the proposed GFQRR model under SRSWOR sampling with respect to the BBB model under SRSWR sampling design is given by:

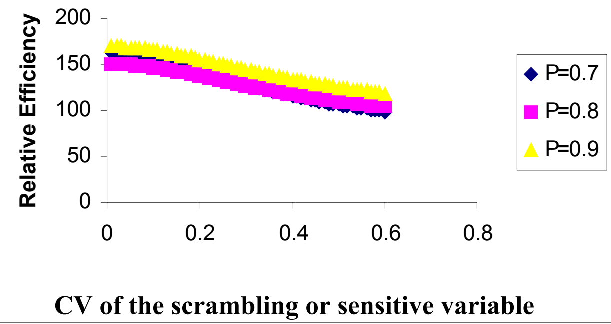
$$\mathrm{RE}(\mathrm{BBB}, \hat{Y}_p)_{\mathrm{srswor}} = \frac{\left\{C_y^2 + (1+C_\gamma^2)C_s^2(p)\right\} \times 100\%}{\frac{N(1-f)}{(N-1)} C_y^2 + C_s^2(p_1, p_2)(1+C_y^2) - \frac{1}{(1-p_3)}} \qquad (2.2.1)$$

where

$$C_s^2(p_1, p_2) = \frac{p_1 + p_2\theta^2(1+C_\gamma^2)}{(p_1 + p_2\theta)^2} - 1 .$$

We observe through simulation that the relative efficiency is highly sensitive towards the mean value of the scrambling variable θ. If we consider a very large value of θ, then the relative efficiency $\mathrm{RE}(\mathrm{BBB}, \hat{Y}_p)_{\mathrm{srswor}}$ of the proposed estimator with respect to the BBB model converges to 100% as the value of the scrambling variable's coefficient of variation also becomes large. Following Cochran (1977), the value of the coefficient of variation should be around 10% for any consistent and practicable data sets. Thus, we decided to choose $N = 10{,}000$, $n = 100$, three values of $p_1 = p = 0.7, 0.8, 0.9$, $p_2 = 2(1-p_1)/3$ and $p_3 = (1 - p_1 - p_2)$. If $\theta = 10$ and the values of the coefficient of variations of the scrambling variable and sensitive variable were kept same, that is, $C_y = C_\gamma$ were chosen between 0.01 and 0.60 with a step of 0.01. Then, the percent relative efficiency of the GFQRR model with respect to the BBB model is shown in the Figure 2.2.

Fig. 2.2. RE of the GFQRR model with respect to the BBB model



If we change $\theta = 1$, and keep the other parameters at the same level, then the results are presented in Figure 2.3.

Fig. 2.3. RE of the proposed GFQRR model with respect to the BBB model

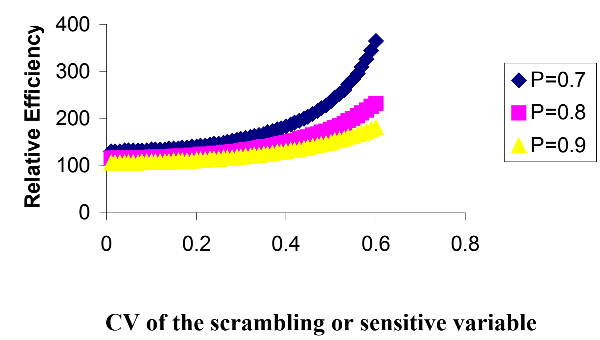


Figure 2.3 shows that if the mean value θ of the scrambling variable is less than one, then more gain is expected from the proposed model at higher values of the coefficient of variations of the scrambling variable or sensitive variable. Note that for higher value of θ, the proposed GFQRR model may perform pitiable than the BBB model, thus the proposed model could be more beneficial if it is used with a scrambling variable having the mean value θ close to one as used by Gupta, Gupta and Singh (2000). The proposed model may perform better for higher value of coefficient of variation of the scrambling variable in a situation as shown in Figure 2.3. Singh and Mathur (2005) have considered situations where the values of the coefficient of variations of the scrambling variable and the sensitive variable can be between 0 and 6 with a step of 0.1.

Now the estimator (2.2) depends upon F, which in turn depends upon Y_i values, and hence it is not practicable estimator. To overcome this difficulty, we consider a new strategy discussed in the next section.

3. Practical GFQRR Model

In this case, we suggest to take two independent random samples s_1 and s_2 from the population Ω using the sampling designs $p(s_1)$ and $p(s_2)$, respectively. In the first sample s_1, each respondent selected is requested to experience the spinner as shown in Figure 3.1:

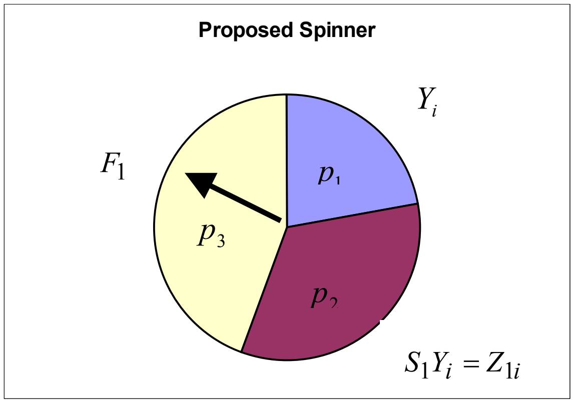


Fig. 3.1. GFQRR spinner for the first sample**.**

Note that the value of F_1 has to be decided before doing the survey based on the parameters to be used in the second spinner used in the second independent survey. Here, this proposed GFQRR model differs from the existing randomization devises. In other words, although both samples are independent, the devices are dependent.

Consequently, the distribution of the i[th] response in the first sample s_1 is given by:

$$Z_{1i} = \begin{cases} Y_i, & \text{with probability } p_1 \\ S_1 Y_i, & \text{with probability } p_2 \\ F_1, & \text{with probability } p_3 \end{cases} \tag{3.1}$$

where S_1 is a scrambling variable such that $E_R(S_1) = \theta_1$, $V_R(S_1) = \gamma_1^2$ and $C_{\gamma_1}^2 = \gamma_1^2 / \theta_1^2$ are assumed to be known.

In the second independent random sample s_2, each respondent selected is requested to experience the spinner as shown in Figure 3.2:

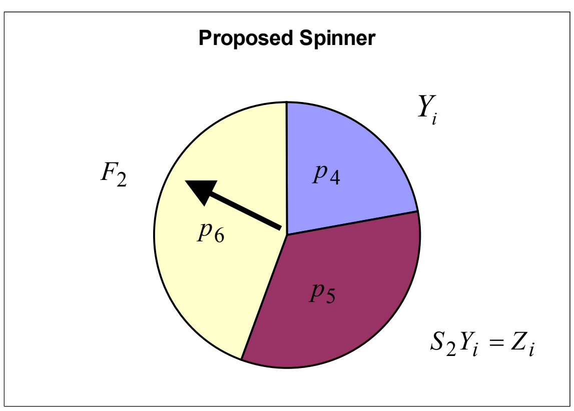


Fig. 3.2. GFQRR spinner for the second sample**.**

In this case, the distribution of the i[th] response in the second sample s_2 is given by:

$$Z_{2i} = \begin{cases} Y_i, & \text{with probability } p_4 \\ S_2 Y_i, & \text{with probability } p_5 \\ F_2, & \text{with probability } p_6 \end{cases} \tag{3.2}$$

where

$$p_6 F_2 = p_3 F_1 \tag{3.3}$$

and S_2 is a scrambling variable such that $E_R(S_2) = \theta_2$, $V_R(S_2) = \gamma_2^2$ and $C_{\gamma_2}^2 = \gamma_2^2 / \theta_2^2$ are assumed to be known.

Then we have the following theorem:

Theorem 3.1. An unbiased estimator of the population total Y is given by:

$$\hat{Y}_{ff} = \frac{1}{\Delta}\left[\sum_{i \in s_1} d_{1i} Z_{1i} - \sum_{i \in s_2} d_{2i} Z_{2i}\right] \tag{3.4}$$

where $\Delta = (p_1 - p_4) + (p_2\theta_1 - p_5\theta_2)$ and $d_{1i} = \pi_{1i}^{-1}, d_{2i} = \pi_{2i}^{-1}$ are the design weights used in the first and second sample respectively; and $\theta_1 = E_R(S_1)$ and $\theta_2 = E_R(S_2)$ are the known means of the scrambling variables S_1 and S_2 used in the first and second sample, respectively.

Proof. Taking expected value on both sides of (3.4) we have:

$$E(\hat{Y}_{ff}) = E\left[\frac{\sum_{i \in s_1} d_{1i} Z_{1i} - \sum_{i \in s_2} d_{2i} Z_{2i}}{\Delta}\right]$$

$$= E_p E_R\left[\frac{\sum_{i \in s_1} d_{1i} Z_{1i} - \sum_{i \in s_2} d_{2i} Z_{2i}}{(p_1 - p_4) + (p_2\theta_1 - p_5\theta_2)}\right]$$

$$= \left[\frac{\sum_{i \in \Omega}(p_1 + p_2\theta_1)Y_i + p_3 F_1 N - \sum_{i \in \Omega}(p_4 + p_5\theta_2)Y_i - p_5 F_2 N}{(p_1 - p_4) + (p_2\theta_1 - p_5\theta_2)}\right]$$

$$= \sum_{i \in \Omega} Y_i = Y$$

which proves the theorem.

Theorem 3.2. The minimum variance of the proposed estimator $\hat{Y}_{ff}$ is given by:

$$\text{Min.V}(\hat{Y}_{ff}) =$$

$$= \frac{1}{\Delta^2}\left[\frac{(p_1 + p_2\theta_1)^2}{2} \sum_{i \neq j \in \Omega} \sum \Theta_{1ij}(d_{1i}Y_i - d_{1j}Y_j)^2\right.$$

$$+ \frac{(p_4 + p_5\theta_2)^2}{2} \sum_{i \neq j \in \Omega} \sum \Theta_{2ij}(d_{2i}Y_i - d_{2j}Y_j)^2$$

$$+ \{\Psi_1 - (p_1 + p_2\theta_1)\} \sum_{i \in \Omega} d_{1i} Y_i^2$$

$$+ \{\Psi_2 - (p_4 + p_5\theta_2)\} \sum_{i \in \Omega} d_{2i} Y_i^2$$

$$
-\frac{p_3\left\{\begin{array}{l}(p_1+p_2\theta_1)\{\sum_{i\in\Omega} d_{1i}Y_i-\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})(d_{1i}Y_i-d_{1j}Y_j)\}\\ +(p_4+p_5\theta_2)\{\sum_{i\in\Omega} d_{2i}Y_i-\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})(d_{1i}Y_i-d_{1j}Y_j)\}\end{array}\right\}^2}{\left\{\begin{array}{l}(1-p_3)\sum_{i\in\Omega} d_{1i}+\frac{p_3^2}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})^2\\ +\{p_3(1-p_6)/p_6\}\sum_{i\in\Omega} d_{2i}+\frac{p_6}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{2ij}(d_{2i}-d_{2j})^2\end{array}\right\}}\Bigg]
\tag{3.5}
$$

where $\Theta_{1ij}=(\pi_{1i}\pi_{1j}-\pi_{1ij})$, $\Theta_{2ij}=(\pi_{2i}\pi_{2j}-\pi_{2ij})$, $\Psi_1=\{p_1+p_2\theta_1^2(1+C_{\gamma_1}^2)\}$ and $\Psi_2=\{p_4+p_5\theta_2^2(1+C_{\gamma_2}^2)\}$.

Proof. Let V_R and V_p denote the variance over the randomization device and over the designs used in the independent samples, then:

$$
V(\hat{Y}_{ff})=\frac{1}{\Delta^2}\left[V\left(\sum_{i\in s_1} d_{1i}Z_{1i}\right)+V\left(\sum_{i\in s_2} d_{2i}Z_{2i}\right)\right]
\tag{3.6}
$$

Now we have:

$$
\begin{aligned}
V\left(\sum_{i\in s_1} d_{1i}Z_{1i}\right)&=V_pE_R\left(\sum_{i\in s_1} d_{1i}Z_{1i}\right)+E_pV_R\left(\sum_{i\in s_1} d_{1i}Z_{1i}\right)\\
&=V_p\left((p_1+p_2\theta_1)\sum_{i\in s_1} d_{1i}Y_i+p_3F_1\sum_{i\in s_1} d_{1i}\right)+E_p\left(\sum_{i\in s_1} d_{1i}^2V_R(Z_{1i})\right)\\
&=(p_1+p_2\theta_1)^2\left\{\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}Y_i-d_{1j}Y_j)^2\right\}\\
&\quad+\{p_1+p_2(\gamma_1^2+\theta_1^2)-(p_1+p_2\theta_1)^2\}\sum_{i\in\Omega} d_{1i}Y_i^2\\
&\quad+F_1^2\left\{p_3(1-p_3)\sum_{i\in\Omega} d_{1i}+p_3^2\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})^2\right\}\\
&\quad-2p_3F_1(p_1+p_2\theta_1)\left\{\sum_{i\in\Omega} d_{1i}Y_i-\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})(d_{1i}Y_i-d_{1j}Y_j)\right\}
\end{aligned}
\tag{3.7}
$$

Similarly,

$$
\begin{aligned}
V\left(\sum_{i\in s_2} d_{2i}Z_{2i}\right)&=(p_4+p_5\theta_2)^2\left\{\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{2ij}(d_{2i}Y_i-d_{2j}Y_j)^2\right\}\\
&\quad+\{p_4+p_5(\gamma_2^2+\theta_2^2)-(p_4+p_5\theta_2)^2\}\sum_{i\in\Omega} d_{2i}Y_i^2\\
&\quad+F_2^2\left\{p_6(1-p_6)\sum_{i\in\Omega} d_{2i}+p_6^2\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{2ij}(d_{2i}-d_{2j})^2\right\}\\
&\quad-2p_6F_2(p_4+p_5\theta_2)\left\{\sum_{i\in\Omega} d_{2i}Y_i-\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{2ij}(d_{2i}-d_{2j})(d_{2i}Y_i-d_{2j}Y_j)\right\}
\end{aligned}
\tag{3.8}
$$

On substituting (3.7) and (3.8) into equation (3.6) and using the relation (3.3) and then setting:

$$
\frac{dV(\hat{Y}_{ff})}{dF_1}=0
$$

we have:

$$
F_1=\frac{\left[\begin{array}{l}(p_1+p_2\theta_1)\left\{\sum_{i\in\Omega} d_{1i}Y_i-\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})(d_{1i}Y_i-d_{1j}Y_j)\right\}\\ +(p_4+p_5\theta_2)\left\{\sum_{i\in\Omega} d_{2i}Y_i-\frac{1}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{2ij}(d_{2i}-d_{2j})(d_{2i}Y_i-d_{1j}Y_j)\right\}\end{array}\right]}{\left[\begin{array}{l}(1-p_3)\sum_{i\in\Omega} d_{1i}+\frac{p_3^2}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{1ij}(d_{1i}-d_{1j})^2\\ +\{p_3(1-p_6)/p_6\}\sum_{i\in\Omega} d_{2i}+\frac{p_6}{2}\sum_{i\neq j\in\Omega}\sum\Theta_{2ij}(d_{2i}-d_{2j})^2\end{array}\right]}
\tag{3.9}
$$

The use of (3.9) in (3.6) leads to (3.5), and which proves the Theorem 3.2. Under simple random and with replacement (SRSWR) sampling the results reduce to Gjestvang and Singh (2005). Note that it is not easy to suggest an unbiased estimator of variance if the value of F_1 is unknown.

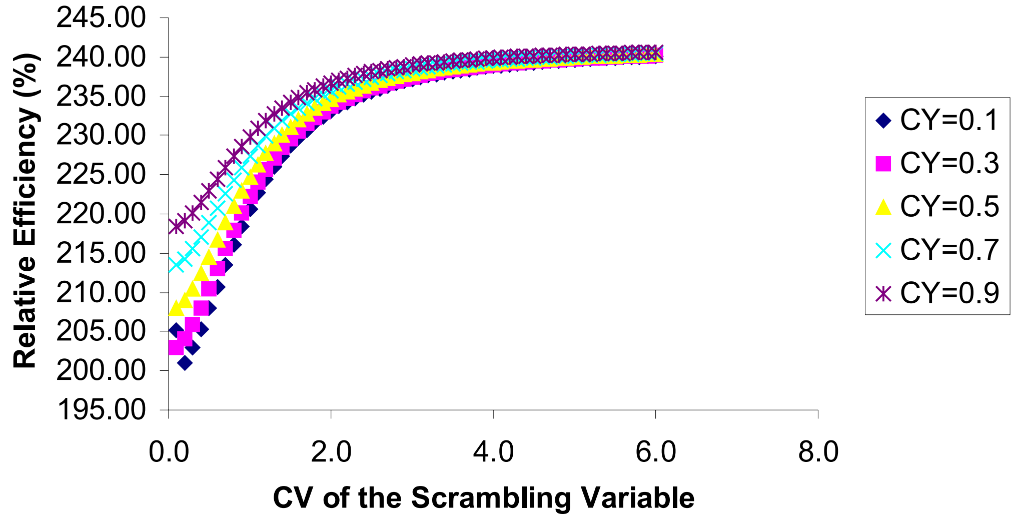
3.1. Relative efficiency

Assuming $n_1=n_2=n/2$, then the percent relative efficiency (RE) of the proposed GFQRR model under SRSWOR sampling with respect to BBB model under SRSWR sampling is given by:

$$
\begin{aligned}
RE(\hat{Y}_{BBB},\hat{Y}_{ff})&=\frac{V(\hat{Y}_{BBB})}{V(\hat{Y}_{ff})}\times 100\%\\
&=\frac{\Delta^2\{C_y^2+(1+C_y^2)C_p^2\}\times 100\%}{\left\{\begin{array}{l}(1-f/2)(p_1+p_2\theta_1+p_4+p_5\theta_2)\left(\frac{N}{N-1}\right)C_y^2\\ +(1+C_y^2)\{\Psi_1-(p_1+p_2\theta_1)+\Psi_2\\ -(p_4+p_5\theta_2)\}-\frac{p_3\{(p_1+p_2\theta_1)+(p_4+p_5\theta_2)\}^2}{\{(1-p_3)+p_3(1-p_6)/p_6\}}\end{array}\right\}}
\end{aligned}
\tag{3.1.1}
$$

The relative efficiency expression in (3.1.1) depends upon several choices. Thus, to look at the behavior of the performance of the proposed GFQRR model with respect to BBB model, we considered a situation where $N=10{,}000$, $n=100$, $\theta=500$, $\theta_1=100$, $\theta_2=900$, $P_1=P=0.8$ (equal protection in the both GFQRR and BBB models), $P_2=2(1-P_1)/3$, $P_4=0.2$, and $P_5=2(1-P_4)/3$. The value of the coefficient of variation C_y of the sensitive variable was changed between 0.1 and 0.9 with a step of 0.2 as shown in Figure 3.3. The values of the coefficient of variation of the three scrambling variables were kept same between 0.1 and 6 with a step of 0.1 by following Singh and Mathur (2005), that is $C_\gamma=C_{\gamma_1}=C_{\gamma_2}$. If the value of coefficient of variation of the scrambling variable becomes more than 2, then the RE becomes almost constant.

Fig. 3.3. Relative efficiency of the proposed GFQRR model with respect to BBB model



More gains are expected if the value of coefficient of variation of the study variable is high, say 0.9, and the value of the coefficient of variation of the scrambling variable is near 0.1. In a real survey, the practicable values of coefficient of variations of the scrambling and sensitive variables are around 0.1 by following Cochran (1977). For such situations, the relative efficiency is shown in Figures 2.2, 2.3 and 3.3. Thus, for these types of practical situations, it is always possible to adjust the randomization devices such that the proposed GFQRR model performs better than the BBB model.

Summary

The proposed generalized forced quantitative randomized response (GFQRR) model has been found to be more efficient than the recently developed BBB model. In addition to that the proposed GFQRR model could be used under more advanced sampling schemes such as: Simple random without replacement (SRSWOR) sampling, Probability proportional to size and without replacement (PPSWOR) sampling, and hence has more practical utility than the BBB model.

Acknowledgement: The authors are thankful to Miss Oludare Odumade for editing the entire manuscript.

References

Bar-Lev, S.K., Bobovitch, E. and Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika*, 255-260.

Bhargava, M. (1996). *An investigation into the efficiencies of certain randomized response strategies*. Unpublished Ph.D. thesis submitted to Punjab Agricultural University, Ludhiana, India.

Bhargava, M. and Singh, R. (2001). Efficiency comparison of certain randomized response schemes with U-model. *J. Indian Soc. Agric. Statist.*, 54(1), 19-28.

Bhargava, M. and Singh, R. (2002). On the efficiency comparison of certain randomized response strategies. *Metrika*, 55(3), 191-197.

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

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

Fox, J.A. and Tracy, P.E. (1986). *Randomized response: A method for sensitive surveys*. SAGE Publications.

Gjestvang, C. R. and Singh, S. (2005). Forced quantitative randomized response model: A new device. *Metrika* (*Revised submitted*).

Gjestvang, C.R and Singh, S. (2006). A new randomized response model. *J. Roy. Statist. Soc.,* B, 68(3), 523-530.

Gupta, S., Gupta, B.C. and Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *J. Statist. Plann. & Infer.*, 100, 209-220.

Kuk, A.Y.C. (1990). Asking sensitive questions indirectly. *Biometrika*, 77, 436-438.

Liu, P.T. and Chow, L.P.(1976a). A new discrete quantitative randomized response model. *J. Amer. Statist. Assoc.,* 71, 72-73.

Liu, P.T. and Chow, L.P. (1976b). The efficiency of the multiple trial randomized response technique. *Biometrics,* 32, 607-618.

Lanke, J. (1975). On the choice of the unrelated question in Simons version of randomized response. *J. Amer. Statist. Assoc.,* 70, 80--83.

Lanke, J. (1976). On the degree of protection in randomized interviews. *Int. Statist. Rev.*, 44, 197-203.

Leysieffer, F.W. and Warner, S. L. (1976). Respondent jeopardy and optimal designs in randomized response models. *J. Amer. Statist. Assoc*., 71, 649-656.

Mangat, N.S. (1994). An improved randomized response strategy. *J. R. Statist. Soc*., B, 56, 93-95.

Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, 77 (2): 439-442.

Moors, J.J.A. (1997). *A critical evaluation of Mangat's two-step procedure in randomized response*. Discussion paper at Center for Economic Research, Tilburg University, The Netherlands.

Nayak, T.K. (1994). On randomized response surveys for estimating a proportion. *Commun. Statist.-Theory Meth.,* 23(1), 3303-3321.

Singh, H.P. and Mathur, N. (2005). Estimation of population mean when coefficient of variation is known using scrambled response technique. *J. Statist. Plann. Infer.*, 131, 135-144.

Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael "Selected" Amy.* Kluwer Academic Publishers, The Netherlands.

Stem, D.E. and Steinhorst, R.K. (1984). Telephone interview and mail questionnaire applications of the randomized response model. *J. Amer. Statist. Assoc.*,555-564.

Tracy, D.S. and Mangat, N.S. (1996). Some developments in randomized response sampling during the last decade - A follow up of review by Chaudhuri and Mukerjee. *J. Applied Statist. Sci.*, 4(2/3),147-158.

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

Zou, G. (1997). Two-stage randomized response procedures as single stage procedures. *Australian J. Statistics*, 39, 235-236.