Hacettepe Journal of Mathematics and Statistics
Volume 46 (2) (2017), 325 – 338

A new improved estimator of population mean in partial additive randomized response models

Nilgun Ozgul*† and Hulya Cingi‡

Abstract

In this study, we have developed a new improved estimator for the population mean estimation of the sensitive study variable in Partial Additive Randomized Response Models (RRMs) using two non-sensitive auxiliary variables. The mean squared error of the proposed estimator is derived and compared with other existing estimators based on the auxiliary variable. The proposed estimator is compared with [19],[5] and [13] estimators in performing a simulation study and is found to be more efficient than other existing estimators using non-sensitive auxiliary variable. The results of the simulation study are discussed in the final section.

Keywords: Randomized response models; Sensitive question, Auxiliary variable, Efficiency, Mean square error.

Received : 23.06.2016 *Accepted* : 02.08.2016 *Doi* : 10.15672/HJMS.2017.414

1. Introduction

Respondents sometimes come across sensitive questions, such as gambling, alcoholism, sexual and physical abuse, drug addiction, abortion, tax evasion, illegal income, mobbing, political view, doping usage, homosexual activities and many others. Respondents often do not respond truthfully, or even refuse to answer when asked directly such sensitive questions. Obtaining valid and reliable information, the researchers commonly use the randomized response models (RRMs). Starting from the pioneering work of [21], many versions of RRMs have been developed that can deal with both proportion and mean estimations. Standard RRMs have been primarily used with surveys that usually require a “yes ”or “no”response to a sensitive question, or a choice of responses from a set of nominal categories. Nevertheless, the literature on RRMs is comprised of various studies dealing with situations where the response to a sensitive question results in a quantitative

*Hacettepe University, Department of Statistics, Beytepe, 06800, Ankara, Turkey, Email: nozgul@hacettepe.edu.tr
†Corresponding Author.
‡Hacettepe University, Department of Statistics, Beytepe, 06800, Ankara, Turkey, Email: hcingi@hacettepe.edu.tr

variable [5]. Quantitative RRMs are used to estimate the mean value of some behavior in a population. These RRMs are sub-classified in either additive models or multiplicative models. In additive models, respondents are asked to scramble their responses using a randomization device, such as a deck of cards. Each of the cards in the deck has a number. The numbers in the deck follow a known probability distribution, such as Normal, Chi-square, Uniform, Poisson, Binomial, or Weibull etc. The respondent is asked to add his "real response "to the "number listed on card "he/she picked, and then report only the sum to the interviewer. Multiplicative RRMs are similar to the additive RRMs. Again, a deck of cards with known probability distribution is used, but now when the respondents scramble their responses, they are asked to report the product of the "real response "and the "number listed on the selected card ". The interviewer cannot see the card and can simply record the reported number [18]. RRMs can also be categorized by how the respondents are instructed to randomize. If all respondents are asked to randomize their response, the model is characterized as a "full randomization model". If some of the respondents are instructed to randomize their response, the model is characterized as a "partial randomization model ".

[21] for the first time, introduced the randomization method for the proportion of a population characterized by a sensitive variable. Later, [22], [8] and [20] extended [21] "s approach to RRMs by estimating the mean of sensitive quantitative variables. Since then, a large number of RRMs have been developed to estimate the mean of quantitative variables. [7], [14] proposed additive RRMs while [6], [1], [16], [9], [10] and [11] proposed multiplicative RRMs. Later, [17] and [4] proposed mixed RRMs with combining additive and multiplicative techniques.

In sampling theory, it is known that there is a considerable reduction in Mean Square Error (MSE) equation when auxiliary information is used, in particular when the correlation between the study variable and the auxiliary variable is high [2]. In recent years, auxiliary information for mean estimation of sensitive variable has been used in RRMs. [5], [19] and [13] suggested regression, ratio, regression cum ratio, respectively, using the auxiliary variable for estimating of the quantitative sensitive variable. Choice of scrambling mechanism plays an important role in quantitative response models. In the RRMs literature, additive models are more effective and user-friendly than multiplicative RRMs and also partial randomization models are more efficient than full randomization models. Therefore, in this paper, we focus on additive partial RRMs for quantitative data. We propose a new improved estimator for the population mean of the sensitive study variable in partial additive RRMs using two non-sensitive auxiliary variables. The remaining part of the paper is organized as follows. In section 2, we present the notations about additive RRMs for quantitative data and introduce various estimators using the auxiliary variable for the unknown mean of a sensitive variable in RRMs. In section 3, we introduce the proposed estimator. In section 4, the proposed estimator is compared with other existing estimators with a simulation study in RRMs and we obtain specific results suggesting that proposed estimator is more efficient than other estimators. Section 5 concludes the paper.

2. Notations and Various Existing Estimators in Partial Additive RRMs for Quantitative Data

Let Y be the study variable, a sensitive variable which cannot be observed directly. Let X be a non-sensitive auxiliary variable which has positive correlation with Y. For example, the sensitive study variable Y may be the annual household income and X may be the annual rental value. Let a random sample of size n be drawn from a finite population. For the i_{th} unit $(U_1, U_2, \ldots, U_N)$, let y_i and x_i be the values of the study variable Y and auxiliary variable X, respectively.

[5] proposed a class of regression estimator for the mean of sensitive variable using a non-sensitive auxiliary variable. In their approach, to estimate μ_y, a sample of n individuals is selected from the population and each respondent is asked to perform a Bernoulli trial with a probability of success P. If this is successful, the respondent then gives the true values of both Y and X. In the case of failure, the respondent gives their answers by using the values given in S and R which are the various randomized designs for Y and X variables, respectively.

Table 1. General Scheme

Reported Responses	(Z, U)	
Variables	(Y, X)	with probability P
Randomized Designs	(S, R)	with probability (1-P)

Under the general scheme, given in Table 1, regression estimator in [5] is,

$$\hat{\mu}_{DP} = \frac{\bar{z} + b\left(\mu_u - \bar{u}\right) - c}{h}, \qquad (h \neq 0) \tag{2.1}$$

where $\bar{z} = \frac{1}{n} \sum\limits_{i=1}^{n} z_i$ is the sample mean of the reported responses for the sensitive variable. Here, the reported response for the sensitive variable is given by $Z = PY + (1-P)(Y + W)$. Here W is the scrambling variable which has pre-assigned distribution. W is the scrambling variable with known true mean μ_w and known variance S_w^2. $\bar{u} = \dfrac{1}{n} \sum\limits_{i=1}^{n} u_i$ is the sample mean of the reported responses for the non-sensitive auxiliary variable. b is suitably selected real constant. Here, c and h depend exclusively on the randomized design.

The variance of $\hat{\mu}_{DP}$ is

$$\hat{\mu}_{DP} = \frac{\bar{z} + b\left(\mu_u - \bar{u}\right) - c}{h}, \qquad (h \neq 0) \tag{2.2}$$

where $S_z^2 = \dfrac{\sum\limits_{i=1}^{N} \left(z_i - \mu_z\right)^2}{N-1}$ and $S_u^2 = \dfrac{\sum\limits_{i=1}^{N} \left(u_i - \mu_u\right)^2}{N-1}$ are the population variances of z and u, respectively, $S_{zu} = \frac{\sum\limits_{i=1}^{N} (z_i - \mu_z)(u_i - \mu_u)}{N-1}$ is the population covariance between z and u, $B = \frac{S_{zu}}{S_u^2}$ is the population regression coefficient between z and $u . E\left(\bar{z}\right) = \mu_z = \frac{1}{N} \sum\limits_{i=1}^{N} z_i$ and $E\left(\bar{u}\right) = \mu_u = \frac{1}{N} \sum\limits_{i=1}^{N} u_i$ are the population means of Z and U, respectively. Substituting the population regression coefficient $B = \dfrac{S_{zu}}{S_u^2}$, minimum variance of unbiased estimator $\hat{\mu}_{DP}$ is

$$Var(\hat{\mu}_{DP})_{\min} = \frac{S_z^2}{nh^2}\left(1 - \rho_{zu}^2\right). \tag{2.3}$$

where $\rho_{zu} = \dfrac{S_{zu}}{S_z S_u}$ is the population correlation coefficient between Z and U.

[19] proposed a ratio estimator for the mean of sensitive variable using a non-sensitive auxiliary variable for full randomization additive model. In their model, the respondent is asked to provide true responses for X. The estimator in [19] is

$$\hat{\mu}_{SR} = \bar{z}\left(\frac{\mu_x}{\bar{x}}\right). \tag{2.4}$$

where $\bar{z}$ is the sample mean of the reported responses for the sensitive variable ($Z = Y + W$), $E(\bar{x}) = \mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the known population mean of non-sensitive auxiliary variable, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of non-sensitive auxiliary variable.

The bias and MSE of $\hat{\mu}_{SR}$, under first order of the approximation, are

$$Bias(\hat{\mu}_{SR}) \cong \lambda\mu_z\left(C_x^2 - C_{zx}\right). \tag{2.5}$$

and

$$MSE\ (\hat{\mu}_{SR}) \cong \lambda\mu_z^2\left[C_z^2 + C_x^2 - 2\rho_{xz}C_xC_z\right]. \tag{2.6}$$

where $\lambda = \frac{1}{n} - \frac{1}{N}$, $C_z = \frac{S_z}{\mu_z}$ and $C_x = \frac{S_x}{\mu_x}$ are the coefficients of variation of Z and X, respectively, ρ_{xz} is the correlation coefficient between X and Z.

[13] proposed regression-cum-ratio estimator for full randomization additive model. Regression-cum-ratio estimator in [13] is

$$\hat{\mu}_{GRR} = \left[b_1\bar{z} + b_2\left(\mu_x - \bar{x}\right)\right]\left(\frac{\mu_x}{\bar{x}}\right) \tag{2.7}$$

where $\bar{z}$ is the sample mean of the reported responses for the sensitive variable ($Z = Y + W$), $E(\bar{x}) = \mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the known population mean of non-sensitive auxiliary variable, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of non-sensitive auxiliary variable, b_1 and b_2 are constants.

The bias and MSE of $\hat{\mu}_{GRR}$, under first order of the approximation, are

$$Bias(\hat{\mu}_{GRR}) \cong (b_1 - 1)\,\mu_z + b_1\lambda\mu_z\left\{C_x^2 - C_{zx}\right\} + b_2\lambda\mu_xC_x^2. \tag{2.8}$$

and

$$\begin{aligned} MSE\left(\hat{\mu}_{GRR}\right) \cong{}& (b_1 - 1)^2\mu_z^2 + \lambda\left[b_1^2\mu_z^2\left\{C_z^2 + 3C_x^2 - 4\rho_{zx}C_zC_x\right\}\right.\\ &+ b_2^2\mu_x^2C_x^2 - 2b_2\mu_z\mu_xC_x^2 - 2b_1\mu_z^2\left\{C_x^2 - \rho_{zx}C_zC_x\right\}\\ &\left.- 2b_1b_2\mu_z\mu_x\left\{\rho_{zx}C_zC_x - 2C_x^2\right\}\right]. \end{aligned} \tag{2.9}$$

Differentiating (2.9) with respect to b_1 and b_2, the following optimum values which minimize the MSE are

$$b_{1(opt)} = \frac{1 - \lambda C_x^2}{1 - \lambda\{C_x^2 - C_z^2\left(1 - \rho_{zx}^2\right)\}} \tag{2.10}$$

and

$$b_{2(opt)} = \frac{\mu_z}{\mu_x}\left\{1 + b_{1(opt)}\left(\frac{\rho_{zx}C_z}{C_x} - 2\right)\right\}. \tag{2.11}$$

Substituting the optimum values of b_1 and b_2 in (2.10) and (2.11) , the minimum MSE of $\hat{\mu}_{GRR}$ is

$$MSE(\hat{\mu}_{GRR})_{\min} \cong \frac{\lambda\mu_z^2 C_z^2\left(1-\rho_{zx}^2\right)\left(1-\lambda C_x^2\right)}{\lambda C_z^2\left(1-\rho_{zx}^2\right)+\left(1-\lambda C_x^2\right)}. \tag{2.12}$$

3. Suggested Improved Estimator in Partial Additive RRMs

Applying the general formulation of [5], we propose an improved estimator for the mean of sensitive variable using two non-sensitive auxiliary variables (X and M). For example, the sensitive study variable Y may be the annual household income and X may be the annual rental value and M may be the number of vehicles in household. In our approach, to estimate μ_y, the procedure works as follows: the respondent gives the true values of the sensitive variable and non-sensitive variables (Y, X, M) with known probability P, whereas provides the scrambled responses with known probability (1-P).

The distribution of the responses is illustrated as: Here, Y is the sensitive variable of

Table 2. General Scheme for Proposed Estimator

Reported Responses	(Z, U, V)	
Variables	(Y, X, M)	with probability P
Randomized Designs	(S, R, L)	with probability (1-P)

interest with unknown mean μ_y and unknown variance S_y^2 , X and M are non-sensitive variables with known means μ_x and μ_m . Z is the reported response for the sensitive variable Y which is given by $Z = PY + (1-P)(Y+W)$, U and V are the reported responses for the first non-sensitive variable X and the second non-sensitive variable M, respectively. Here W is the scrambling variable which has pre-assigned distribution. W is the scrambling variable with known true mean μ_w and known variance S_w^2 . S, R and L are the various randomized designs for Y, X and M variables, respectively. The reported responses for auxiliary variables changes to which randomized design adopted. Under the general scheme presented in Table 2, we propose the following improved estimator based on a SRSWOR sample $(z_1,u_1,v_1),(z_2,u_2,v_2),\ldots,(z_n,u_n,v_n)$ of n responses

$$\hat{\mu}_{NHR} = \frac{\bar{z}_R - c}{h}, \qquad (h \neq 0) \tag{3.1}$$

here

$$\bar{z}_R = k_1\bar{z}\left(\frac{\mu_u}{k_2\bar{u}+(1-k_2)\mu_u}\right)\left(\frac{\mu_v}{k_3\bar{v}+(1-k_3)\mu_v}\right) \qquad (h \neq 0) \tag{3.2}$$

where $\bar{z} = \dfrac{1}{n}\sum\limits_{i=1}^{n} z_i$ is the sample mean of the reported responses for sensitive variable and $\bar{u} = \dfrac{1}{n}\sum\limits_{i=1}^{n} u_i$ and $\bar{v} = \frac{1}{n}\sum\limits_{i=1}^{n} v_i$ are the sample means of reported responses for first and second auxiliary variables, respectively. $E(\bar{u}) = \mu_u = \dfrac{1}{N}\sum\limits_{i=1}^{N} u_i$ and $E(\bar{v}) = \mu_v = \dfrac{1}{N}\sum\limits_{i=1}^{N} v_i$ are the population means of first and second reported auxiliary variables, respectively. k_1, k_2 and k_3 are the constants. Here, c and h values for partial additive RRMs obtained as:

$$\begin{aligned} \mu_z &= P\mu_y + (1-P)\left(\mu_y+\mu_w\right) \\ \mu_z &= \mu_y + (1-P)\mu_w. \end{aligned} \tag{3.3}$$

from (3.3) c and h values are given:

$$\mu_y = \mu_z - (1-P)\mu_w. \tag{3.4}$$

Here, $c = (1-P)\mu_w$, $h = 1$

To obtain the *MSE* equation for the proposed estimator, we define the following relative error terms [3]

$$e_0 = \frac{(\bar{z}-\mu_z)}{\mu_z}, e_1 = \frac{(\bar{u}-\mu_u)}{\mu_u}, e_2 = \frac{(\bar{v}-\mu_v)}{\mu_v}. \tag{3.5}$$

such that

$$\begin{aligned} E\left(e_0\right) = E\left(e_1\right) = E\left(e_2\right) = 0;\ E\left(e_0^2\right) = \lambda C_z^2, E\left(e_1^2\right) = \lambda C_u^2, E\left(e_2^2\right) = \lambda C_v^2, \\ E\left(e_0 e_1\right) = \lambda \rho_{zu} C_z C_u, E\left(e_0 e_2\right) = \lambda \rho_{zv} C_z C_v, E\left(e_1 e_2\right) = \lambda \rho_{uv} C_u C_v, \end{aligned} \tag{3.6}$$

where $C_v^2 = \dfrac{S_v^2}{\mu_v^2}$, $S_v^2 = \dfrac{\sum\limits_{i=1}^{N}\left(v_i-\mu_v\right)^2}{N-1}$, $\rho_{zu} = \dfrac{S_{zu}}{S_z S_u}$, $\rho_{zv} = \dfrac{S_{zv}}{S_z S_v}$, $\rho_{uv} = \dfrac{S_{uv}}{S_u S_v}$.

Expressing (3.2) in terms of e's in (3.5) and retaining terms in e's up to first order, we have,

$$\begin{aligned} \bar{z}_R - \mu_z &= k_1\mu_z\left(1+e_0\right)\left(\frac{\mu_u}{k_2\mu_u\left(1+e_1\right)+(1-k_2)\mu_u}\right)\left(\frac{\mu_v}{k_3\mu_v\left(1+e_2\right)+(1-k_3)\mu_v}\right) - \mu_z \\ &= k_1\mu_z\left(1+e_0\right)\left(1+k_2e_1\right)^{-1}\left(1+k_3e_2\right)^{-1} - \mu_z \\ &= k_1\mu_z\left(1+e_1\right)\left(1-k_2e_1+...\right)\left(1-k_3e_2+...\right) - \mu_z \\ &= (k_1-1)\mu_z + k_1\mu_z\left(e_0 - k_2\, e_1 - k_3e_2 - k_2e_0\ e_1 - k_3e_0e_2 + k_2k_3\, e_1e_2\right). \end{aligned} \tag{3.7}$$

Taking expectation of both sides of (3.7) and using notations in (3.6), the bias equation of the estimator $\bar{z}_R$

$$Bias\left(\bar{z}_R\right) = (k_1-1)\mu_z - k_1\lambda\mu_z\left(k_2\rho_{zu}C_zC_u + k_3\rho_{zv}C_zC_v\ -\ k_2k_3\rho_{uv}C_uC_v\right). \tag{3.8}$$

using (3.1) and (3.8) , we obtain the bias equation of the estimator $\hat{\mu}_{NHR}$

$$Bias\left(\hat{\mu}_{NHR}\right) = (k_1-1)\mu_z -\ k_1\lambda\mu_z\left(k_2\rho_{zu}C_zC_u + k_3\rho_{zv}C_zC_v\ -\ k_2k_3\rho_{uv}C_uC_v\right) - (1-P)\mu_w. \tag{3.9}$$

Retaining terms in e's up to first order, taking the square of both sides of (3.7) and expectation and using notations in (3.6) , the *MSE* equation of the estimator $\bar{z}_R$

$$\begin{aligned} E(\bar{z}_R-\mu_z)^2 =& E\left((k_1-1)^2\mu_z^2 + k_1^2\mu_z^2e_0^2 + k_1^2k_2^2\mu_z^2e_1^2 + k_1^2k_3^2\mu_z^2e_2^2 \right. \\ & \left. - 2k_1^2k_2\mu_z^2e_0e_1, -2k_1^2k_3\mu_z^2e_0e_2\ +2k_1^2k_2k_3\mu_z^2e_1e_2\right)\} \\ MSE\left(\bar{z}_R\right) =& (k_1-1)^2\mu_z^2 + \lambda k_1^2S_z^2 + \lambda k_1^2S_z^2\left(k_2^2C_u^2 + k_3^2C_v^2 - 2k_2\rho_{zu}C_zC_u \right. \\ & \left. - 2k_3\rho_{zv}C_zC_v\ +\ 2k_2k_3\rho_{uv}C_uC_v\right). \end{aligned} \tag{3.10}$$

using (3.1) and (3.10), we obtain the *MSE* equation of the estimator $\hat{\mu}_{NHR}$

$$\begin{aligned} MSE\left(\hat{\mu}_{NHR}\right) =& (k_1-1)^2\mu_z^2 + \lambda k_1^2S_z^2 + \lambda k_1^2S_z^2\left(k_2^2C_u^2 + k_3^2C_v^2 - 2k_2\rho_{zu}C_zC_u \right. \\ & \left. - 2k_3\rho_{zv}C_zC_v\ +\ 2k_2k_3\rho_{uv}C_uC_v\right) \end{aligned} \tag{3.11}$$

where $S_{uv} = \dfrac{\sum\limits_{i=1}^{N}\left(u_i-\mu_u\right)\left(v_i-\mu_v\right)}{N-1}$ is the covariance between u and v.

Differentiating (3.11) with respect to k_1, k_2 and k_3, and then by setting the resulting equations to zero, we obtain the following equations:

$$\frac{\partial MSE\left(\hat{\mu}_{NHR}\right)}{\partial k_1} = 2\left(k_1-1\right)\mu_z^2 + 2\lambda k_1 S_z^2 + 2\lambda k_1 \mu_z^2\left(k_2^2C_u^2 + k_3^2C_v^2 - 2k_2\rho_{zu}C_zC_u - 2k_3\rho_{zv}C_zC_v + 2k_2k_3\rho_{uv}C_uC_v\right) = 0. \tag{3.12}$$

$$\frac{\partial MSE\left(\hat{\mu}_{NHR}\right)}{\partial k_2} = k_2C_u^2 - \rho_{zu}C_zC_u + k_3\rho_{uv}C_uC_v = 0. \tag{3.13}$$

$$\frac{\partial MSE\left(\hat{\mu}_{NHR}\right)}{\partial k_3} = k_3C_v^2 - \rho_{zv}C_zC_v + k_2\rho_{uv}C_uC_v = 0. \tag{3.14}$$

Solving the (3.12), (3.13) and (3.14) simultaneously, we get the optimum values which minimize theMSE equation.

$$k_{1(opt)} = \frac{2-\lambda C_z^2}{2\left(1-\lambda C_z^2 R_{zu.v}^2\right)} \tag{3.15}$$

where $R_{zu.v}^2 = \dfrac{\rho_{zu}^2 + \rho_{zv}^2 - 2\rho_{zu}\rho_{zv}\rho_{uv}}{1-\rho_{uv}^2}$

$$k_{2(opt)} = \frac{\rho_{zu}\left(C_z/C_v\right) - \rho_{zv}\rho_{uv}\left(C_z/C_u\right)}{1-\rho_{uv}^2}. \tag{3.16}$$

$$k_{3(opt)} = \frac{\rho_{zv}\left(C_z/C_v\right) - \rho_{zu}\rho_{uv}\left(C_u^2/C_zC_v\right)}{1-\rho_{uv}^2}. \tag{3.17}$$

Substituting the optimum values of k_1, k_2 and k_3 in (3.11), the minimum MSE of $\hat{\mu}_{NHR}$ is

$$MSE_{\min}\left(\hat{\mu}_{NHR}\right) = \lambda S_z^2\left[\frac{R_{zu.v}^2 - \frac{1}{2}\left(\lambda C_z^2 - 1\right)}{1-\lambda C_z^2 R_{zu.v}^2} - \frac{R_{zu.v}^2}{C_z^2}\left(\frac{1-\lambda C_z^2/2}{1-\lambda C_z^2 R_{zu.v}^2}\right)^2\right]. \tag{3.18}$$

MSE and mean equations change depending on the specified models. We specified three partial additive models. In the first model M1, the additive model is applied for the sensitive variable while the direct method is utilized for the non-sensitive auxiliary variables {Z=PY+(1-P)(Y+W), U=X, V=M}. In the second model M2, additive model is applied for the sensitive variable and both of two non-sensitive auxiliary variables {Z=PY+(1-P)(Y+W), U=PX+(1-P)(X+T), V=PM+(1-P)(M+K)}. In the third model M3, the additive model is applied for the sensitive variable and the first non-sensitive auxiliary variable while the direction method is utilized for second non-sensitive auxiliary variable {Z=PY+(1-P)(Y+W), U=PX+(1-P)(X+T), V=M}. Here W, T and K are the scrambling variables which have pre-assigned distributions. W is the scrambling variable with known true mean μ_w and known variance S_w^2 in S design, T is the scrambling variable with known true mean μ_t and known variance S_t^2 in R design, K is the scrambling variable with known true mean μ_k and known variance S_k^2 in L design [15]

Mean equations which will be used in MSE equation in (3.18) for M1 is obtained as below: Mean equation of z is

$$\begin{aligned}\mu_z &= P\mu_y + (1-P)\left(\mu_y+\mu_w\right)\\ \mu_z &= \mu_y + (1-P)\mu_w.\end{aligned} \tag{3.19}$$

Mean equation of u is

$$\begin{aligned}\mu_u &= P\mu_x + (1-P)\mu_x \\ \mu_u &= \mu_x.\end{aligned} \tag{3.20}$$

Mean equation of v is

$$\begin{aligned}\mu_v &= P\mu_m + (1-P)\mu_m \\ \mu_v &= \mu_m.\end{aligned} \tag{3.21}$$

Variance equation of z which will be used in MSE equation in (3.18) for M1 is obtained as,

$$\begin{aligned}S_z^2 &= PE\left(Y^2\right) + (1-P)E\left\{(Y+W)^2\right\} - \mu_z^2 \\ &= S_y^2 + (1-P)S_w^2 + P(1-P)\mu_w^2 \\ S_z^2 &= S_y^2 + (1-P)\,\mu_w^2\left(C_w^2 + P\right).\end{aligned} \tag{3.22}$$

The correlation equation between Z and U which will be used in MSE equation in (3.18) for M1 is obtained as

$$\rho_{zu} = \frac{S_{yx}}{S_x\sqrt{S_y^2 + (1-P)\,\mu_w^2\,(C_w^2+P)}} \tag{3.23}$$

for $S_{zu} = S_{yx}$, $S_z = \sqrt{S_y^2 + (1-P)\,\mu_w^2\,(C_w^2+P)}$, $S_u = S_x$.

The correlation equation between Z and V which will be used in MSE equation in (3.18) for M1 is obtained as

$$\rho_{zv} = \frac{S_{ym}}{S_x\sqrt{S_y^2 + (1-P)\,\mu_w^2\,(C_w^2+P)}} \tag{3.24}$$

for $S_{zv} = S_{ym}$, $S_z = \sqrt{S_y^2 + (1-P)\,\mu_w^2\,(C_w^2+P)}$, $S_v = S_m$.

The correlation equation between U and V which will be used in MSE equation in (3.18) for M1 is obtained as

$$\rho_{uv} = \frac{S_{xm}}{S_x S_m} = \rho_{xm} \tag{3.25}$$

for $S_{uv} = S_{xm}$, $S_u = S_x$, $S_v = S_m$.

Mean, variance and correlation equations which will be used in MSE equation in (3.18) for M2 and M3 is obtained similarly given as before. Mean, variance and correlation equations for three models are given in Table 3.

4. Simulation Study

In this section, a simulation study is presented to show the performance of the proposed estimator in comparison to other estimators using the auxiliary variable for partial additive RRMs. The proposed estimator,$\hat{\mu}_{NHR}$, is compared with $\hat{\mu}_{DP}$ in [5], $\hat{\mu}_{SR}$ in [19], and $\hat{\mu}_{GRR}$ in [13]. It is known that RRMs based on the auxiliary variable are practically indistinguishable and always perform better than RRMs in which auxiliary variables are not used. For this reason, the estimators without using auxiliary variables are not included in the simulation study [5]. We generate three finite populations of size 10000 from multivariate normal distribution. Three populations have theoretical means of $[Y, X, M]$ as $\mu = [5, 5, 5]$ and have different variance-covariance matrices. The populations are generated as the levels of correlation between the variables. The correlation levels are considered as low, medium and high. The covariance matrices and the correlations are presented in (4.2), (4.3) and (4.4). The scrambling variable W is considered to be a normal random variable with mean equal to zero and standard deviation is equal to 0.30. The scrambling variables, T and K, are normal random variables with mean equal

Table 3. Special Partial Additive Randomized Models

	S	R	L	Mean	Variance and Correlation Equations
M1	Y+W	X	M	$\mu_z = \mu_y + (1-P)\mu_w$ $\mu_u = \mu_x$ $\mu_v = \mu_m$ $c = (1-P)\mu_w, h = 1$	$S_z^2 = S_y^2 + (1-P)\,\mu_w^2\,(C_w^2 + P)$ $\rho_{zu} = \dfrac{S_{yx}}{S_x\sqrt{S_y^2 + (1-P)\,\mu_w^2\,(C_w^2+P)}}$ $\rho_{zv} = \dfrac{S_{ym}}{S_m\sqrt{S_y^2 + (1-P)\,\mu_w^2\,(C_w^2+P)}}$ $\rho_{uv} = \dfrac{S_{xm}}{S_x S_m} = \rho_{xm}$
M2	Y+W	X+T	M+K	$\mu_z = \mu_y + (1-P)\mu_w$ $\mu_u = \mu_x + (1-P)\mu_t$ $\mu_v = \mu_m + (1-P)\mu_k$ $c = (1-P)\mu_w, h = 1$	$S_z^2 = S_y^2 + (1-P)\,\mu_w^2\,(C_w^2 + P)$ $\rho_{zu} = \dfrac{S_{yx} + P(1-P)\mu_w\mu_t}{\sqrt{\{S_y^2 + (1-P)\mu_w^2\,(C_w^2+P)\}\,\{S_x^2 + (1-P)\mu_t^2\,(C_t^2+P)\}}}$ $\rho_{zv} = \dfrac{S_{my} + P(1-P)\mu_w\mu_k}{\sqrt{\{S_y^2 + (1-P)\mu_w^2\,(C_w^2+P)\}\,\{S_m^2 + (1-P)\mu_k^2\,(C_k^2+P)\}}}$ $\rho_{uv} = \dfrac{S_{xm} + P(1-P)\mu_t\mu_k}{\sqrt{\{S_x^2 + (1-P)\mu_t^2\,(C_t^2+P)\}\,\{S_m^2 + (1-P)\mu_k^2\,(C_k^2+P)\}}}$
M3	Y+W	X+T	M	$\mu_z = \mu_y + (1-P)\mu_w$ $\mu_u = \mu_x + (1-P)\mu_t$ $\mu_v = \mu_m$ $c = (1-P)\mu_w, h = 1$	$S_z^2 = S_y^2 + (1-P)\,\mu_w^2\,(C_w^2 + P)$ $\rho_{zu} = \dfrac{S_{yx} + P(1-P)\mu_w\mu_t}{\sqrt{\{S_y^2 + (1-P)\mu_w^2\,(C_w^2+P)\}\,\{S_x^2 + (1-P)\mu_t^2\,(C_t^2+P)\}}}$ $\rho_{zv} = \dfrac{S_{ym}}{S_m\sqrt{\{S_y^2 + (1-P)\mu_w^2\,(C_w^2+P)\}}}$ $\rho_{uv} = \dfrac{S_{xm}}{S_m\sqrt{\{S_x^2 + (1-P)\mu_t^2\,(C_t^2+P)\}}}$

to zero and standard deviations are equal to 0.20. We use the simulation studies of [5] and [13] to determine the parameters that are to be easier to compare.

The variance-covariance matrix define as:

$$(4.1) \qquad \sum_i = \begin{bmatrix} \mathrm{S}_{\mathrm{yi}}^2 & S_{yxi} & S_{ymi} \\ S_{xyi} & \mathrm{S}_{\mathrm{xi}}^2 & S_{xmi} \\ S_{myi} & S_{mxi} & \mathrm{S}_{\mathrm{mi}}^2 \end{bmatrix}, \quad i = 1,\, 2,\, 3.$$

The variance-covariance matrices and the correlation coefficients for each population are given below.

Population I

There are low correlations between the sensitive variable and non-sensitive auxiliary variables in population I. The variance-covariance matrix for population I is

$$(4.2) \qquad \sum_1 = \begin{bmatrix} 9.0 & 1.8 & 1.5 \\ 1.8 & 4.0 & 0.8 \\ 1.5 & 0.8 & 4.0 \end{bmatrix}, \ \rho_{yx} = 0.30, \rho_{ym} = 0.25.$$

Population II

There are medium correlations between the sensitive variable and non-sensitive auxiliary variables in population II. The variance-covariance matrix for population II is

$$(4.3) \qquad \sum_2 = \begin{bmatrix} 9.0 & 3.6 & 3.1 \\ 3.6 & 4.0 & 1.2 \\ 3.1 & 1.2 & 4.0 \end{bmatrix}, \ \rho_{yx} = 0.60, \rho_{ym} = 0.52$$

Population III

There are high correlations between the sensitive variable and non-sensitive auxiliary variables in population III. The variance-covariance matrix for population III is

$$\sum_{3} = \begin{bmatrix} 9.0 & 5.4 & 4.2 \\ 5.4 & 4.0 & 2.0 \\ 4.2 & 2.0 & 4.0 \end{bmatrix}, \rho_{yx} = 0.90, \rho_{ym} = 0.70 \tag{4.4}$$

The process is repeated 5000 times and for different sample sizes: n = 50, 100, 200, 300, 500. The value of the design parameter P changes from 0.10 to 0.90 with an increment of 0.1. We observe small differences in efficiency with almost each value of the design parameter when auxiliary variable is utilized in partial additive RRMs in our simulation study. Thus, we only present the simulation results for P=0.20. That means when partial additive model is utilized, 0.20 percent of the respondents give direct answers, the rest of the respondents use the scrambling devices. The performance of the estimators is measured by the simulated mean square error as:

$$MSE\left(\hat{\mu}\right) = \frac{1}{5000} \sum_{i=1}^{5000} \left(\hat{\mu}_i - \mu_y\right)^2 \tag{4.5}$$

where $\hat{\mu}$ is the estimate of μ_y on the $i_t h$ sample.

Simulation results are summarized in Tables 4, 5 and 6.

In Tables 4-6, theoretical and empirical *MSE* values of the estimators according to degree of the correlation between the sensitive and non-sensitive variables are given for specified models, respectively. For three models, in all circumstances, regardless of both degree of correlation and sample size, the proposed estimator is always more efficient than $\hat{\mu}_{DP}$,$\hat{\mu}_{SR}$ and $\hat{\mu}_{GRR}$ estimators. For population I and II, where there are low and medium correlations between the variables, respectively, the differences between the *MSE* values of the proposed estimator and the other estimators are small. For population III, where there are high correlations between the variables, the *MSE* values of the proposed estimator are relatively smaller than the *MSE* values of other estimators. We can say that when the degree of the correlation between the variables increases, the efficiency of the proposed estimator increases. We also note that the *MSE* values of the estimators are smaller when the sample size increases and that is an expected result.

Table 4. Theoretical and empirical MSEs of the estimators according to degree of the correlation between the sensitive and non-sensitive variables under Model 1 for P=0.20.

		Population I		Population II		Population III	
	Estimators	MSE		MSE		MSE	
n		Theoretical	Empirical	Theoretical	Empirical	Theoretical	Empirical
50	$\hat{\mu}_{DP}$	0.1684	0.1728	0.1197	0.1187	0.0358	0.0372
	$\hat{\mu}_{SR}$	0.1894	0.1916	0.1203	0.1191	0.0472	0.0480
	$\hat{\mu}_{GRR}$	0.1672	0.1754	0.1191	0.1187	0.0358	0.0374
	$\hat{\mu}_{NHR}$	**0.1617**	**0.1705**	**0.0979**	**0.1047**	**0.0053**	**0.0055**
100	$\hat{\mu}_{DP}$	0.0838	0.0840	0.0595	0.0608	0.0178	0.0186
	$\hat{\mu}_{SR}$	0.0942	0.0945	0.0599	0.0613	0.0235	0.0239
	$\hat{\mu}_{GRR}$	0.0828	0.0840	0.0594	0.0610	0.0178	0.0186
	$\hat{\mu}_{NHR}$	**0.0807**	**0.0834**	**0.0488**	**0.0499**	**0.0026**	**0.0027**
200	$\hat{\mu}_{DP}$	0.0415	0.0408	0.0295	0.0300	0.0088	0.0091
	$\hat{\mu}_{SR}$	0.0466	0.0465	0.0296	0.0302	0.0116	0.0120
	$\hat{\mu}_{GRR}$	0.0414	0.0411	0.0291	0.0297	0.0088	0.0191
	$\hat{\mu}_{NHR}$	**0.0400**	**0.0404**	**0.0241**	**0.0237**	**0.0013**	**0.0013**
300	$\hat{\mu}_{DP}$	0.0274	0.0272	0.0194	0.0193	0.0058	0.0060
	$\hat{\mu}_{SR}$	0.0308	0.0308	0.0196	0.0193	0.0077	0.0079
	$\hat{\mu}_{GRR}$	0.0273	0.0272	0.0194	0.0193	0.0058	0.0060
	$\hat{\mu}_{NHR}$	**0.0264**	**0.0262**	**0.0159**	**0.0158**	**0.0008**	**0.0008**
500	$\hat{\mu}_{DP}$	0.0161	0.0163	0.0114	0.0115	0.0034	0.0035
	$\hat{\mu}_{SR}$	0.0181	0.0183	0.0115	0.0116	0.0045	0.0046
	$\hat{\mu}_{GRR}$	0.0161	0.0163	0.0114	0.0116	0.0034	0.0035
	$\hat{\mu}_{NHR}$	**0.0155**	**0.0159**	**0.0093**	**0.0095**	**0.0005**	**0.0005**

Table 5. Theoretical and empirical MSEs of the estimators according to degree of the correlation between the sensitive and non-sensitive variables under Model 2 for P=0.20.

		Population I		Population II		Population III	
	Estimators	MSE		MSE		MSE	
n		Theoretical	Empirical	Theoretical	Empirical	Theoretical	Empirical
50	$\hat{\mu}_{DP}$	0.1670	0.1722	0.1183	0.1188	0.0357	0.0375
	$\hat{\mu}_{SR}$	0.1883	0.1905	0.1191	0.1180	0.0459	0.0466
	$\hat{\mu}_{GRR}$	0.1659	0.1740	0.1178	0.1173	0.0357	0.0370
	$\hat{\mu}_{NHR}$	**0.1601**	**0.172**	**0.0976**	**0.1045**	**0.0068**	**0.0070**
100	$\hat{\mu}_{DP}$	0.0831	0.0835	0.0589	0.0606	0.0178	0.0188
	$\hat{\mu}_{SR}$	0.0937	0.0938	0.0593	0.0605	0.0229	0.0233
	$\hat{\mu}_{GRR}$	0.0835	0.0847	0.0588	0.0600	0.0178	0.0186
	$\hat{\mu}_{NHR}$	**0.0799**	**0.0831**	**0.0486**	**0.0499**	**0.0034**	**0.0034**
200	$\hat{\mu}_{DP}$	0.0411	0.0405	0.0291	0.0297	0.0088	0.0090
	$\hat{\mu}_{SR}$	0.0464	0.0462	0.0293	0.0298	0.0113	0.0116
	$\hat{\mu}_{GRR}$	0.0411	0.0408	0.0294	0.0300	0.0088	0.0190
	$\hat{\mu}_{NHR}$	**0.0396**	**0.0402**	**0.0241**	**0.0237**	**0.0016**	**0.0016**
300	$\hat{\mu}_{DP}$	0.0271	0.0270	0.0192	0.0191	0.0058	0.0059
	$\hat{\mu}_{SR}$	0.0306	0.0307	0.0194	0.0191	0.0075	0.0077
	$\hat{\mu}_{GRR}$	0.0271	0.0271	0.0197	0.0193	0.0058	0.0060
	$\hat{\mu}_{NHR}$	**0.0261**	**0.0269**	**0.0159**	**0.0158**	**0.0011**	**0.0011**
500	$\hat{\mu}_{DP}$	0.0160	0.0162	0.0113	0.0114	0.0034	0.0035
	$\hat{\mu}_{SR}$	0.0180	0.0182	0.0114	0.0115	0.0044	0.0045
	$\hat{\mu}_{GRR}$	0.0159	0.0162	0.0114	0.0116	0.0034	0.0035
	$\hat{\mu}_{NHR}$	**0.0153**	**0.0158**	**0.0093**	**0.0095**	**0.0006**	**0.0006**

Table 6. Theoretical and empirical MSEs of the estimators according to degree of the correlation between the sensitive and non-sensitive variables under Model 3 for P=0.20.

		Population I		**Population II**		**Population III**	
	Estimators	**MSE**		**MSE**		**MSE**	
n		**Theoretical**	**Empirical**	**Theoretical**	**Empirical**	**Theoretical**	**Empirical**
50	$\hat{\mu}_{DP}$	0.1684	0.1728	0.1197	0.1187	0.0358	0.0372
	$\hat{\mu}_{SR}$	0.1895	0.1916	0.1203	0.1191	0.0472	0.0480
	$\hat{\mu}_{GRR}$	0.1672	0.1718	0.1191	0.1187	0.0358	0.0373
	$\hat{\mu}_{NHR}$	**0.1608**	**0.1727**	**0.0978**	**0.1047**	**0.0062**	**0.0064**
100	$\hat{\mu}_{DP}$	0.0838	0.0840	0.0595	0.0608	0.0178	0.0186
	$\hat{\mu}_{SR}$	0.0942	0.0945	0.0899	0.0613	0.0235	0.0239
	$\hat{\mu}_{GRR}$	0.0835	0.0847	0.0594	0.0609	0.0178	0.0186
	$\hat{\mu}_{NHR}$	**0.0802**	**0.0835**	**0.0487**	**0.0500**	**0.0031**	**0.0031**
200	$\hat{\mu}_{DP}$	0.0414	0.0408	0.0294	0.0300	0.0088	0.0091
	$\hat{\mu}_{SR}$	0.0466	0.0465	0.0296	0.0302	0.0116	0.0120
	$\hat{\mu}_{GRR}$	0.0414	0.0411	0.0291	0.0297	0.0088	0.0191
	$\hat{\mu}_{NHR}$	**0.0398**	**0.0405**	**0.0241**	**0.0237**	**0.0015**	**0.0015**
300	$\hat{\mu}_{DP}$	0.0273	0.0272	0.0194	0.0193	0.0058	0.0060
	$\hat{\mu}_{SR}$	0.0308	0.0308	0.0196	0.0193	0.0077	0.0079
	$\hat{\mu}_{GRR}$	0.0273	0.0272	0.0194	0.0194	0.0058	0.0060
	$\hat{\mu}_{NHR}$	**0.0262**	**0.0264**	**0.0159**	**0.0158**	**0.0010**	**0.0010**
500	$\hat{\mu}_{DP}$	0.0161	0.0163	0.0114	0.0116	0.0034	0.0035
	$\hat{\mu}_{SR}$	0.0181	0.0183	0.0115	0.0116	0.0045	0.0046
	$\hat{\mu}_{GRR}$	0.0161	0.0163	0.0114	0.0116	0.0034	0.0035
	$\hat{\mu}_{NHR}$	**0.0154**	**0.0159**	**0.0093**	**0.0095**	**0.0006**	**0.0006**

5. Conclusion

In the RRMs literature, there are several estimators based on a non-sensitive auxiliary variable. In this paper, we propose a new improved estimator based on two non-sensitive auxiliary variables for the population mean of a sensitive variable in partial additive RRMs. The proposed estimator is more efficient than other existing estimators in all circumstances, regardless of which the model is applied. The proposed estimator can be considered reliable and may lead the researcher to find a suitable estimator for RRMs. The estimation of the mean of a sensitive variable can be improved by using more non-sensitive auxiliary variables. It is proved that RRMs based on two or more auxiliary variables are certainly more efficient than those with one auxiliary variable. We show that the efficiency of the proposed estimator can be quite distinctive if the correlation between the study and the auxiliary variables is high. Additionally, the additive RRMs are more efficient and user friendly than the multiplicative RRMs. Thus, we substitute our proposed estimator to three specific partial additive RRMs and we compare these newly-generated models. In the future work, the study can be extended by combining the additive and multiplicative RRMs for the proposed estimator using more than one auxiliary variable based on different sampling methods.

References

[1] Bar-Lev S.K., Bobovitch E., Boukai B. *A note on randomized response models for quantitative data.* Metrika **60**, 255-260, 2004.

[2] Cingi, H. Ornekleme Kuramı. Ankara: 3.Baski, Bizim Buro Basimevi, 2009.

[3] Cingi, H. and Kadilar, C.*Advances in Sampling Theory-Ratio Method of Estimation* .Bentham Science Publishers,**121**, 2009.

[4] Diana G., Perri P.F.*New Scrambled Response Models For Estimating The Mean of A Sensitive Quantitative Character.*Journal Of Applied Statistics ,**37** (11), 1875-1890, 2010.

[5] Diana G., Perri P.F. *A class of Estimators for Quantitative Sensitive Data.*Statistical Papers,**52**, 633-650, 2011.

[6] Eichhorn, B. H. and Hayre, L. S.*Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data.* Journal of Statistical Planning and Inference , **7** (1) ,307-316, 1983.

[7] Gjesvang, C.R., Singh,S. *An improved Randomized Response Model:Estimation of Mean.* Journal of Applied Statistics, **36** (12),1361-1367, 2009.

[8] Greenberg B.G., Kuebler R.R., Abernathy J.R., Horvitz D.G.*Application of the randomized response technique in obtaining quantitative data.* Journal of American Statistical Association , **66**, 243-250, 1971.

[9] Gupta S., Gupta B., Singh S. *Estimation of sensitivity level of personal interview survey questions.* J. Stat Plan Inference ,**100**, 239-247, 2002.

[10] Gupta, S.N. and Shabbir, J. *Sensitivity estimation for personal interview survey Questions.* Statistica,**64**, 643-653, 2004.

[11] Gupta, S. N. and Shabbir,J. *On the Estimation of Population Mean and Sensitivity in Two-stage Optional Randomized Response Model.* Journal of Indian Society of Agricultural Statistics, **61**, 164-168, 2007.

[12] Gupta S., Shabbir J., Sousa R. and Sehra S. *Mean and sensitivity estimation in optional randomized response model.*Communications in Statistics , **140**, 2870-2874, 2010.

[13] Gupta, S., Shabbir, J., Sousa, R., Real, P. C. *Estimation of the mean of a sensitive variable in the presence of auxiliary information.* Communications in Statistics, Theory and Methods, **41**, 1-12, 2012.

[14] Hussain Z., Shabbir, J.*Improved Estimation Procedures for the Mean of Sensitive Variable using Randomized Response Model* . Pak.J.Statist.**25**(2), 205-220, 2009.

[15] Ozgul, N. *Proportion and Mean Estimators in Randomized Response Models*, Ph.D. Thesis. Hacettepe University, Ankara, 2013.

[16] Ryu J.-B, Kim J.-M, Heo T-Y, Park C. G. *On stratified Randomized Response Sampling.* Model Assisted Statistics and Application **1**(1): 31-36, 2005.

[17] Saha, A.*A simple randomized response technique in complex surveys.* Metron **LXV**,59-66, 2007.

[18] Sehra, Supriti, M.A. *Two-Stage Optional Randomized Response Models, Master of Art Dissertation.* The University of North Carolina, Greensboro, U.S.A, 2008.

[19] Sousa, R., Shabbir, J. Real, P. C. and Gupta, S.*Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information* . Journal of Statistical Theory and Practice, **4**(3), 495-507, 2010.

[20] Thornton, B. and Gupta, S. N. *Comparative Validation of a Partial (versus Full) Randomized Response Technique: Attempting to Control for Social Desirability Response Bias to Sensitive Questions.* Individual Differences Research, **2**, 214-224, 2004.

[21] Warner, S.L. Randomized response: *A survey technique for eliminating evasive answer bias.* Journal of the American Statistical Association, **60**, 63-69, 1965.

[22] Warner, S.L. *The Linear Randomized Response Model*, Journal of the American Statistical Association **66**(336), 884-888, 1971.