Statistics and Applications {ISSN 2452-7395 (online)}
Volume 15 Nos. 1&2, 2017 (New Series), pp 173-191

A Survey Technique for Estimating the Proportion and Sensitivity in a Stratified Dichotomous Finite Population

Tanveer A. Tarray[1] and Housila P. Singh[2]
[1]*Department of Computer Science and Engineering, Islamic University of Science and Technology-Awantipora-Kashmir*
[2]*School of Studies in Statistics, Vikram University, Ujjain*

Received: December 20, 2016; Revised: January 04, 2017; Accepted: January 29, 2017

Abstract

This paper proposes a two – stage stratified randomized response model based on Huang (2004) model and has a large gain in precision. It is also shown that the proposed model is more efficient than Kim and Warde (2004) and Huang (2004) stratified randomized response model. Numerical illustrations and graphs are also given in support of the present study.

Keywords: Stratified random sampling, Dichotomous population, Estimation of proportion, Maximum likelihood estimator and Mean squared error.

1. Introduction

In a survey of human population, questions requiring personal or controversial statements often run into trouble in terms of resistance. It is difficult to gather reliable data from interviewees and hard to raise the quality of responses when the survey topic is sensitive. To reduce the bias and to procure reliable data Warner (1965) developed a randomized response (RR) model to estimate a proportion for sensitive attributes including homosexuality, drug addiction or abortion. Greenberg et al. (1969) envisaged an unrelated question randomized response model using Warner's sensitive question and unrelated question. The randomized response model has been studied by various authors, Chaudhuri and Mukerjee (1988), Ryu et al. (1993), Tracy and Mangat (1996) and Fox and Tracy (1986). Some recent investigations like Mangat et al.(1997), Mahmood et al. (1998), Singh et al. (2000), Chang and Huang (2001), Javed et al. (2002), Huang (2004), Chang et al. (2004 a,b), Land et al. (2011), Singh and Tarray (2012, 2013), Tarray and Singh (2014, 2015, 2016) and Tarray (2016).

Consider a dichotomous population in which every person belongs to either to a sensitive group "A" or to the non – sensitive complement A^c. let a population with size N be divided into k disjoint groups (strata) with size N_i (i =1,2,…,k).

Let $\pi_{Si}(i=1,2,\ldots,k)$ be the proportion of people with the sensitive trait in a stratum i. Let $\pi_S=\sum_{i=1}^{k}w_i\pi_{Si}$ be the proportion of sensitive attributes in the whole

Corresponding author: Tanveer A. Tarray
E-mail: anveerstat@gmail.com

population, where w_i = (N_i/N) for (i = 1, 2, ...k) so that $w=\sum_{i=1}^{k} w_i = 1$. The aim of interest is to estimate π_S, the population proportion of individuals who are members of "A". To do so, a sample is selected by simple random sampling with replacement (SRSWR) from each stratum. Let n_i denote the number of units in the sample from stratum i and n denote the total number of units in the samples from all strata so that $n=\sum_{i=1}^{k} n_i$. In a direct response survey, the interviewee in the sample from stratum i is asked if they have membership "A" or A^c. let T_i be the probability that the respondents from the stratum i belonging to "A" report the truth. The respondents belonging to the non – sensitive group A^c have no reason to tell a lie. The usual sample proportion of "Yes" responses, $\hat{\pi}_{Di}$, is an estimator of π_{Si}. Thus the estimator of $\pi_S = \sum_{i=1}^{k} w_i \pi_{Si}$ is given by

$$\hat{\pi}_D = \sum_{i=1}^{k} w_i \hat{\pi}_{Di} . \tag{1.1}$$

Taking expectation of both sides of (1.1) we have

$$E\,(\hat{\pi}_D) = \sum_{i=1}^{k} w_i E\,(\hat{\pi}_{Di}) = \sum_{i=1}^{k} w_i \pi_{Si} T_i \ . \tag{1.2}$$

which gives the bias in $\hat{\pi}_D$ is

$$B\,(\hat{\pi}_D) = \sum_{i=1}^{k} w_i \pi_{Si} (T_i - 1) . \tag{1.3}$$

So the estimator $\hat{\pi}_D$ is a biased estimator of π_S .

The variance of $\hat{\pi}_D$ is given by

$$V(\hat{\pi}_D) = \sum_{i=1}^{k} w_i^2 \frac{\pi_{Si} T_i (1-\pi_{Si} T_i)}{n_i} . \tag{1.4}$$

So the mean square error (MSE) of the estimator $\hat{\pi}_D$ is given by

$$MSE(\hat{\pi}_D) = V\,(\hat{\pi}_D) + B((\hat{\pi}_D))^2$$

$$= \sum_{i=1}^{k} w_i^2 \frac{\pi_{Si} T_i (1-\pi_{Si} T_i)}{n_i} + \left(\sum_{i=1}^{k} w_i \pi_{Si} (T_i - 1) \right)^2 . \tag{1.5}$$

Information on π_{Si}'s for i = 1,2,…,k are usually unavailable. But if prior information on π_{Si} and T_i are available from past experience then it helps to derive the following optimal allocation formula.

Theorem 1.1: The optimal allocation of n to n_1, n_2... n_{k-1} and n_k to derive the minimum variance of the $\hat{\pi}_D$ subject to $n = \sum_{i=i}^{k} n_i$ is approximately given by

$$\frac{n_i}{n} = \frac{w_i \sqrt{\pi_{Si} T_i \left(1-\pi_{Si} T_i\right)}}{\sum_{i=1}^{k} w_i \sqrt{\pi_{Si} T_i \left(1-\pi_{Si} T_i\right)}} . \tag{1.6}$$

If we set (1.6) in (1.5) the minimum variance of the estimator $\hat{\pi}_D$ is given by

$$\mathrm{V}(\hat{\pi}_D) = \frac{1}{n}\left\{\sum_{i=1}^{k} w_i\sqrt{\pi_{Si}T_i(1-\pi_{Si}T_i)}\right\}^2 . \tag{1.7}$$

Thus the minimal mean square error of the estimator $\hat{\pi}_D$ is given by:

$$\mathrm{MSE}(\hat{\pi}_D) = \frac{1}{n}\left\{\sum_{i=1}^{k} w_i\sqrt{\pi_{Si}T_i(1-\pi_{Si}T_i)}\right\}^2 + \left(\sum_{i=1}^{k} w_i\pi_{Si}(T_i-1)\right)^2 . \tag{1.8}$$

Hong et al. (1994) suggested a stratified RR technique using a proportional allocation. A problem with the Hong et al. (1994) model is that it may cause a high cost because of the difficulty in obtaining a proportional sample from each stratum. To overcome this problem, Kim and Warde (2004) suggested a stratified RR technique using an optimal allocation which is more efficient than that using a proportional allocation. Kim and Elam (2005) have applied Kim and Warde's (2004) stratified RR model to Mangat and Singh's (1990) two stage RR model and studied the properties of their proposed model using optimum allocation. Kim and Warde (2005) have suggested a mixed randomized response model and extended this model to stratified sampling. Kim and Elam (2007) have envisaged stratified randomized response model for Greenberg et al.'s (1969) unrelated question RR model with its properties. Further Lee et al. (2013) have extended the estimation reported in Land et al. (2011) using a Poisson distribution and an unrelated question randomized response model reported in Greenberg et al. (1969).

In this paper we have developed a stratified randomized response model for Huang (2004) RR model. We have shown that the proposed stratified RR model is more efficient in terms of variance than Kim and warde's (2004) stratified RR model. Numerical illustrations and graphs are given in support of the present study.

2. Proposed model

In the proposed model, the finite population is partitioned into strata, and a sample is selected by simple random sampling with replacement (SRSWR) from each stratum. To get the full benefit from stratification, we suppose that the number of units in each stratum is known. An individual respondent in the sample from stratum i is required to reply to a direct query whether he / she bears "A" or not. When answering "No", the respondent is provided with a randomization device consisting of two statements (a) I am a member of "A" (sensitive group), and (b) I am not a member of "A", with probability P_i and (1-P_i) respectively. Since the respondents bearing "A" have no reason to tell a lie, it may reasonably be expected that they will be completely truthful in their answers, no matter whether a direct response or a randomized response procedure is adopted. Here it is assumed that the respondents belonging to "A" give totally honest response procedure, but with probability T_i following the usual direct response procedure in stratum i.

Let n_i denote the number of units in the sample from stratum i and n denote the total number of units in sample from all stratum so that $n = \sum_{i=1}^{k} n_i$. It is assumed that the probabilities T_i and P_i are set by the researcher, the probability of a "Yes" answer in a stratum i in the direct response procedure is given by:

$$\theta_{1i} = T_i\pi_{Si} \quad , \quad \text{for } (i = 1, 2 \ldots, k) \tag{2.1}$$

and in the randomized response procedure is given by

$$\theta_{2i} = P_i\pi_{Si}(1-T_i) + (1-P_i)(1-\pi_{Si})$$
$$= (2P_i - 1)\pi_{Si} - P_i\pi_{Si}T_i + (1-P_i) \quad (2.2)$$

where $(\theta_{1i}, \theta_{2i})$ are the proportion of "Yes" responses in the direct response and the proposed randomized response procedures respectively and π_{Si} is the proportion of respondents with the sensitive trait in the sample from stratum *i*. The proposed estimators for π_{Si} and T_i are respectively given by

$$\hat{\pi}_{Si} = \frac{P_i\hat{\theta}_{1i} + \hat{\theta}_{2i} - (1-P_i)}{(2P_i - 1)}, \quad for\ (\ i = 1, 2, ..., k) \quad (2.3)$$

and

$$\hat{T}_i = \frac{(2P_i - 1)\,\hat{\theta}_{1i}}{\{P_i\hat{\theta}_{1i} + \hat{\theta}_{2i} - (1-P_i)\}}, \quad for\ (\ i = 1, 2, ..., k), \quad (2.4)$$

where $\hat{\theta}_{ji}$, the observed proportion of "Yes" answers, in the binomial random variable with parameters n_i and θ_{ji} , j = 1,2.

The estimator $\hat{\pi}_{Si}$ is unbiased with the variance given by

$$V(\hat{\pi}_{Si}) = \frac{\pi_{Si}(1-\pi_{Si})}{n_i} + \frac{P_i(1-P_i)(1-\pi_{Si}T_i)}{n_i(2P_i - 1)} \quad (2.5)$$

Since the selection in different strata are made independently, the estimators for individual strata can be added together to obtain an estimators for the whole population. The estimator of $\pi_S = \sum_{i=1}^{k} w_i\pi_{Si}$, the proportion of respondents with the sensitive trait, is:

$$\hat{\pi}_S = \sum_{i=1}^{k} w_i\hat{\pi}_{Si} = \sum_{i=1}^{k} w_i\left[\frac{P_i\hat{\theta}_{1i} + \hat{\theta}_{2i} - (1-P_i)}{(2P_i - 1)}\right] \quad (2.6)$$

where we denote N to be the number of units in the whole population, N_i to be the total number of units in the stratum *i* and $w_i = (N_i/N)$ for (i = 1, 2, ...k) so that $w = \sum_{i=1}^{k} w_i = 1$.

Theorem 2.1: The proposed estimator $\hat{\pi}_S$ is an unbiased estimate for the population proportion π_S.

Proof: The unbiasedness of $\hat{\pi}_{Si}$ follows from $E(\hat{\theta}_{ji}) = \theta_{ji}$, j=1,2,…k. Thus the unbiasedness of $\hat{\pi}_S$ follows from taking the expected value of (2.4).

Theorem 2.2 The variance of the estimator $\hat{\pi}_S$ is:

$$V(\hat{\pi}_S) = \sum_{i=1}^{k} \frac{w_i^{\,2}}{n_i}\left[\pi_{Si}(1-\pi_{Si}) + \frac{P_i(1-P_i)(1-\pi_{Si}T_i)}{(2P_i - 1)^2}\right]. \quad (2.7)$$

It is known that the variance of the Kim and Warde (2004) estimator $\hat{\pi}_{kw}$ is given by

$$V(\hat{\pi}_{kw}) = \sum_{i=1}^{k} \frac{w_i^2}{n_i}\left[\pi_{Si}(1-\pi_{Si}) + \frac{P_i(1-P_i)}{(2P_i-1)^2}\right]. \tag{2.8}$$

From (2.5) and (2.6) we have

$$V(\hat{\pi}_{kw}) - V(\hat{\pi}_S) = \left[\sum_{i=1}^{k} \frac{w_i^2}{n_i} \frac{P_i(1-P_i)\pi_{Si}T_i}{(2P_i-1)^2}\right] > 0 \tag{2.9}$$

which follows that the proposed procedure cannot be less efficient than Kim and Warde's (2004) procedure. An unbiased estimator of the V($\hat{\pi}_S$) can easily be obtained which is given as follows.

Theorem 2.3: The unbiased estimator of the V($\hat{\pi}_S$) is given by:

$$\hat{V}(\hat{\pi}_S) = \hat{V}\left(\sum_{i=1}^{k} w_i\hat{\pi}_{Si}\right) = \sum_{i=1}^{k} w_i^2 \hat{V}(\hat{\pi}_{Si}) \tag{2.10}$$

where

$$\hat{V}(\hat{\pi}_{Si}) = \frac{1}{(n_i-1)}\left[\hat{\pi}_{Si}(1-\hat{\pi}_{Si}) + \frac{P_i(1-P_i)}{(2P_i-1)^2}\right] - \frac{P_i(1-P_i)\hat{\theta}_{1i}}{n_i(2P_i-1)}, \text{ for } i = 1,2,\ldots k \tag{2.11}$$

Information on π_{Si}'s for $i = 1,2,\ldots,k$ are usually unavailable. But if prior information on π_{Si} and T_i are available from past experience then we may derive the following optimal allocation formula.

Theorem 2.4: The optimal allocation of n to n_1, n_2... n_{k-1} and n_k to derive the minimum variance of the $\hat{\pi}_S$ subject to $n = \sum_{i=i}^{k} n_i$ is approximately given by

$$\frac{n_i}{n} = \frac{w_i\left[\pi_{Si}(1-\pi_{Si}) + \frac{P_i(1-P_i)(1-\pi_{Si}T_i)}{(2P_i-1)^2}\right]^{1/2}}{\sum_{i=1}^{k} w_i\left[\pi_{Si}(1-\pi_{Si}) + \frac{P_i(1-P_i)(1-\pi_{Si}T_i)}{(2P_i-1)^2}\right]^{1/2}}. \tag{2.12}$$

Proof: Follows from section 5.5 of Cochran (1977).

The minimal variance of the estimator $\hat{\pi}_S$ is given by:

$$V(\hat{\pi}_S) = \frac{1}{n}\left\{\sum_{i=1}^{k} w_i\left[\pi_{Si}(1-\pi_{Si}) + \frac{P_i(1-P_i)(1-\pi_{Si}T_i)}{(2P_i-1)^2}\right]^{1/2}\right\}^2. \tag{2.13}$$

Using (2.11), the unbiased minimal variance of the estimator $\hat{\pi}_S$ in (2.13) can be derived. To obtain the mean square error of the estimator $\hat{T}_i$, let us define $d_{1i} = (2P_i-1)\hat{\theta}_{1i}$ and $d_{2i} = P_i\hat{\theta}_{1i} + \hat{\theta}_{2i} - (1-P_i)$, it is easy to see that $E(d_{1i}) = (2P_i-1)\pi_{Si}T_i$ and $E(d_{2i}) = (2P_i-1)\pi_{Si}$. The estimator $\hat{T}_i$ can be expressed as $\hat{T}_i = d_{1i}/d_{2i}$, and we have $\hat{T}_i = E(d_{1i})/E(d_{2i})$. Further we define the following quantities

$$e_{1i} = \frac{d_{1i}-E(d_{1i})}{E(d_{1i})} \text{ and } e_{2i} = \frac{d_{2i}-E(d_{2i})}{E(d_{2i})}$$

assuming that $|e_{2i}| < 1$ so that the function $(1+e_{2i})^{-1}$ can be validly expanded as a power series. It can be easily proved that

$$E(e_{1i}^2)=\frac{\theta_{1i}(1-\theta_{1i})}{n_i\pi_{Si}^2T_i^2},\ E(e_{2i}^2)=\frac{\{P_i^2\theta_{1i}(1-\theta_{1i})+\theta_{2i}(1-\theta_{2i})-2P_i\theta_{1i}\theta_{2i}\}}{n_i(2P_i-1)^2\pi_{Si}^2}$$

and $$E(e_{1i}e_{2i})=\frac{\{P_i\theta_{1i}(1-\theta_{1i})-\theta_{1i}\theta_{2i}\}}{n_i(2P_i-1)\pi_{Si}^2T_i^2}.$$

The estimation error of the estimator $\hat{T}_i$ can be expressed as

$$\hat{T}_i-T_i=T_i(e_{1i}-e_{2i})+O_p(n_i^{-1/2})$$

Then we have the following theorem from Huang (2004, pp 78).

Theorem 2.5: The mean square error (MSE) of the estimator $\hat{T}_i$, up to terms of order (n^{-1}), is given by

$$MSE(\hat{T}_i)=\frac{T_i(1-T_i)}{n_i\pi_{Si}}+\frac{P_i(1-P_i)T_i^2(1-\pi_{Si}T_i)}{n_i(2P_i-1)^2\pi_{Si}^2}. \tag{2.14}$$

We denote T be the weighted probability $T=\sum_{i=1}^{k}w_iT_i$, where T_i is the probability that a respondent with the sensitive trait will report truthfully. Thus using (2.4) in $T=\sum_{i=1}^{k}w_iT_i$, we get an estimate of the weighted probability T as

$$\hat{T}=\sum_{i=1}^{k}w_i\hat{T}_i=\sum_{i=1}^{k}w_i\left[\frac{(2P_i-1)\hat{\theta}_{1i}}{\{P_i\hat{\theta}_{1i}+\hat{\theta}_{2i}-(1-P_i)\}}\right]. \tag{2.15}$$

From (2.12), the mean square error (MSE) of $\hat{T}$ to order (n^{-1}) is given by

$$MSE(\hat{T})=\frac{1}{n_i\pi_{Si}}\left\{\sum_{i=1}^{k}w_i^2\left[T_i(1-T_i)+\frac{P_i(1-P_i)T_i^2(1-\pi_{Si}T_i)}{(2P_i-1)^2\pi_{Si}}\right]\right\} \tag{2.16}$$

Theorem 2.6: The optimal allocation of n to n_1, n_2... n_{k-1} and n_k to derive the minimum variance of the $\hat{\pi}_S$ subject to $n=\sum_{i=i}^{k}n_i$ is approximately given by

$$\frac{n_i}{n}=\frac{w_i\left[\frac{T_i(1-T_i)}{\pi_{Si}}+\frac{P_i(1-P_i)T_i^2(1-\pi_{Si}T_i)}{(2P_i-1)^2\pi_{Si}^2}\right]^{1/2}}{\sum_{i=1}^{k}w_i\left[\frac{T_i(1-T_i)}{\pi_{Si}}+\frac{P_i(1-P_i)T_i^2(1-\pi_{Si}T_i)}{(2P_i-1)^2\pi_{Si}^2}\right]^{1/2}}. \tag{2.17}$$

The minimum mean square error (MSE) of $\hat{T}$ to order n^{-1} is given by

$$MSE(\hat{T})=\frac{1}{n}\left[\sum_{i=1}^{k}w_i\left\{\frac{T_i(1-T_i)}{\pi_{Si}}+\frac{P_i(1-P_i)(1-\pi_{Si}T_i)T_i^2}{(2P_i-1)^2\pi_{Si}^2}\right\}\right]^2. \tag{2.18}$$

3. Efficiency Comparisons

In this section we have made the comparison of proposed estimators $(\hat{\pi}_S, \hat{T})$ with direct randomized response estimator $\hat{\pi}_D$, Kim and Warde's (2004) randomized response estimator $\hat{\pi}_{kw}$ and Huang (2004) estimator $\hat{\pi}_H$ (say) in stratified randomized sampling numerically.

3.1 Comparison with direct response procedure

For two strata (i.e. k = 2) in the population and $T = T_1 = T_2$, we have computed the relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect to direct randomized response estimator $\hat{\pi}_D$ by using the formula:

$$RE(\hat{\pi}_S, \hat{\pi}_D) = \frac{\left[(w_1 A_1 + w_2 A_2)^2 + n(T-1)^2 \pi_S^2\right]}{\left(w_1\sqrt{A_1^*} + w_2\sqrt{A_2^*}\right)^2} \tag{3.1}$$

where $A_1 = \sqrt{\pi_{S1}T(1-\pi_{S1}T)}$, $A_2 = \sqrt{\pi_{S2}T(1-\pi_{S2}T)}$,

$$A_1^* = \left[\pi_{S1}(1-\pi_{S1}) + \frac{P_1(1-P_1)(1-\pi_{S1}T)}{(2P_1-1)^2}\right],$$

$$A_2^* = \left[\pi_{S2}(1-\pi_{S2}) + \frac{P_2(1-P_2)(1-\pi_{S2}T)}{(2P_2-1)^2}\right],$$

and $\pi_S = w_1\pi_{S1} + w_2\pi_{S2}$.

3.2 Comparison with Kim and Warde (2004) estimator

For two strata (i.e. k=2) in the population and $T = T_1 = T_2$, we have computed the relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect to Kim and Warde's (2004) estimator $\hat{\pi}_{kw}$ by using the formula:

$$RE(\hat{\pi}_S, \hat{\pi}_{kw}) = \frac{\left[\left(w_1\sqrt{B_1} + w_2\sqrt{B_2}\right)^2\right]}{\left(w_1\sqrt{A_1^*} + w_2\sqrt{A_2^*}\right)^2} \tag{3.2}$$

where

$$B_1 = \left[\pi_{S1}(1-\pi_{S1}) + \frac{P_1(1-P_1)}{(2P_1-1)^2}\right],$$

$$B_2 = \left[\pi_{S2}(1-\pi_{S2}) + \frac{P_2(1-P_2)}{(2P_2-1)^2}\right],$$

and $\pi_S = w_1\pi_{S1} + w_2\pi_{S2}$.

3.3 Comparison with Huang's (2004) estimator $\hat{\pi}_H$ in stratified random sampling

For two strata (i.e. k=2) in the population and $T = T_1 = T_2$, we have computed the relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect Huang's (2004) estimator $\hat{\pi}_H$ by using the formula:

$$RE(\hat{\pi}_S,\hat{\pi}_H)=\frac{\left[\pi_S(1-\pi_S)+\dfrac{P(1-P)(1-\pi_S T)}{(2P-1)^2}\right]}{\left(w_1\sqrt{A_1^*}+w_2\sqrt{A_2^*}\right)^2} \tag{3.3}$$

3.4 Comparison of the proposed estimator $\hat{T}$ with Huang's (2004) estimator $\hat{T}_H$

For two strata (i.e. k=2) in the population and $T = T_1 = T_2$, we have computed the relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect Huang's (2004) estimator $\hat{T}_H$ by using the formula:

$$RE(\hat{T},\hat{T}_H)=\frac{\left\{\dfrac{T(1-T)}{\pi_S}+\dfrac{P(1-P)T^2(1-\pi_S T)}{(2P-1)^2\pi_S^2}\right\}}{\left\{w_1\sqrt{D_1}+w_2\sqrt{D_2}\right\}^2}. \tag{3.4}$$

where $$D_1=\left\{\frac{T(1-T)}{\pi_S}+\frac{P_1(1-P_1)T^2(1-\pi_{S1}T)}{(2P_1-1)^2\pi_{S1}^2}\right\}$$

and $$D_2=\left\{\frac{T(1-T)}{\pi_S}+\frac{P_2(1-P_2)T^2(1-\pi_{S2}T)}{(2P_2-1)^2\pi_{S2}^2}\right\}.$$

4. Numerical illustrations

To judge the merits of the proposed estimators $(\hat{\pi}_S,\hat{T})$ over direct randomized response estimator $\hat{\pi}_D$, Kim and Warde's (2004) randomized response estimator $\hat{\pi}_{kw}$ and Huang (2004) estimators $(\hat{\pi}_H,\hat{T}_H)$ in stratified randomized sampling. We have computed the relative efficiencies (REs) of the estimators $(\hat{\pi}_S,\hat{T})$ with respect to the direct randomized response estimator $\hat{\pi}_D$, Kim and Warde's (2004) randomized response estimator $\hat{\pi}_{kw}$ and Huang (2004) estimators $(\hat{\pi}_H,\hat{T}_H)$ in stratified randomized sampling by using the formulae (3.1) - (3.4), for prior information on $\pi_{S1},\pi_{S2}, w_1, w_2, T$, T_1, T_2, k = 2, n = 1000 and different values of P, P_1, P_2 ;

Findings are displayed in Tables 1 to 4. Diagrammatic representations are also given in Figures 1 to 4.

It is observed from Tables 1 to 4 and Figures 1 to 4 that:

(i) The RE$(\hat{\pi}_S,\hat{\pi}_D)$ increases while the value of RE $(\hat{\pi}_S,\hat{\pi}_{kw})$, RE$(\hat{\pi}_S,\hat{\pi}_H)$, RE $(\hat{T},\hat{T}_H)$ decrease as both (P_1, P_2) increase and no trend is observed when π_S increases.

(ii) The values of RE$(\hat{\pi}_S,\hat{\pi}_D)$, RE $(\hat{\pi}_S,\hat{\pi}_{kw})$, RE$(\hat{\pi}_S,\hat{\pi}_H)$ and RE $(\hat{T},\hat{T}_H)$ are larger than 100. So the envisaged estimators $(\hat{\pi}_S,\hat{T})$ are more efficient than Kim and Warde's (2004) estimator $\hat{\pi}_{kw}$ and Huang (2004) stratified randomized response model with substantial gain in efficiency. Thus the suggested estimators $(\hat{\pi}_S,\hat{T})$ are recommended for their use in practice.

Table 1: Relative Efficiency of the Proposed Estimator $\hat{\pi}_S$ with respect to Direct Randomized Response Procedure $\hat{\pi}_D$.

π_S	π_{S1}	π_{S2}	n = 1000	P_1= 0.6	P_1=0.65	P_1=0.7	P_1=0.75	P_1=0.8	P_1=0.85	P_1=0.9	P_1=0.93
			T	P_2 =0.65	P_2 =0.7	P_2 = 0.75	P_2 = 0.8	P_2 = 0.85	P_2 = 0.9	P_2 = 0.95	P_2 = 0.95
0.40	0.38	0.43	0.10	25.99	55.39	95.39	146.02	207.30	279.27	362.00	401.97
0.42	0.38	0.43	0.13	37.43	72.61	118.56	175.15	242.22	319.60	407.06	426.11
0.40	0.38	0.43	0.15	23.64	50.33	86.55	132.22	187.25	251.56	325.02	360.39
0.42	0.38	0.43	0.17	34.63	67.11	109.40	161.32	222.61	292.98	372.09	389.29
0.40	0.38	0.43	0.19	21.82	46.40	79.69	121.54	171.79	230.26	296.73	328.63
0.52	0.48	0.53	0.21	50.08	96.61	156.56	229.15	313.47	408.49	513.09	535.39
0.50	0.48	0.53	0.23	32.16	68.05	116.07	175.50	245.52	325.17	413.46	454.89
0.52	0.48	0.53	0.25	46.12	88.82	143.64	209.71	286.02	371.48	464.91	484.78
0.50	0.48	0.53	0.27	29.53	62.40	106.23	160.25	223.57	295.19	374.06	410.93
0.52	0.48	0.53	0.29	42.24	81.23	131.08	190.87	259.54	335.94	418.87	436.49
0.60	0.58	0.63	0.31	40.44	85.32	144.99	218.21	303.60	399.60	504.60	553.26
0.62	0.58	0.63	0.33	57.20	109.92	177.19	257.71	349.96	452.28	562.97	586.23
0.60	0.58	0.63	0.35	36.90	77.71	131.72	197.62	273.96	359.14	451.52	494.11
0.62	0.58	0.63	0.37	52.04	99.78	160.39	232.47	314.41	404.54	501.14	521.39
0.60	0.58	0.63	0.39	33.44	70.29	118.82	177.71	245.43	320.40	401.01	437.97
0.72	0.68	0.73	0.41	67.75	130.53	211.07	308.09	420.07	545.33	682.10	710.71
0.70	0.68	0.73	0.43	43.72	92.16	156.35	234.80	325.82	427.60	538.23	589.11
0.72	0.68	0.73	0.45	61.15	117.47	189.23	274.94	372.90	481.27	598.14	622.50
0.70	0.68	0.73	0.47	39.26	82.54	139.54	208.67	288.14	376.06	470.53	513.67
0.72	0.68	0.73	0.49	54.68	104.73	168.03	242.96	327.71	420.37	519.06	539.55

Figure 1: Relative Efficiency of the Proposed Estimator $\hat{\pi}_S$ with respect to Direct Randomized Response Procedure $\hat{\pi}_D$.

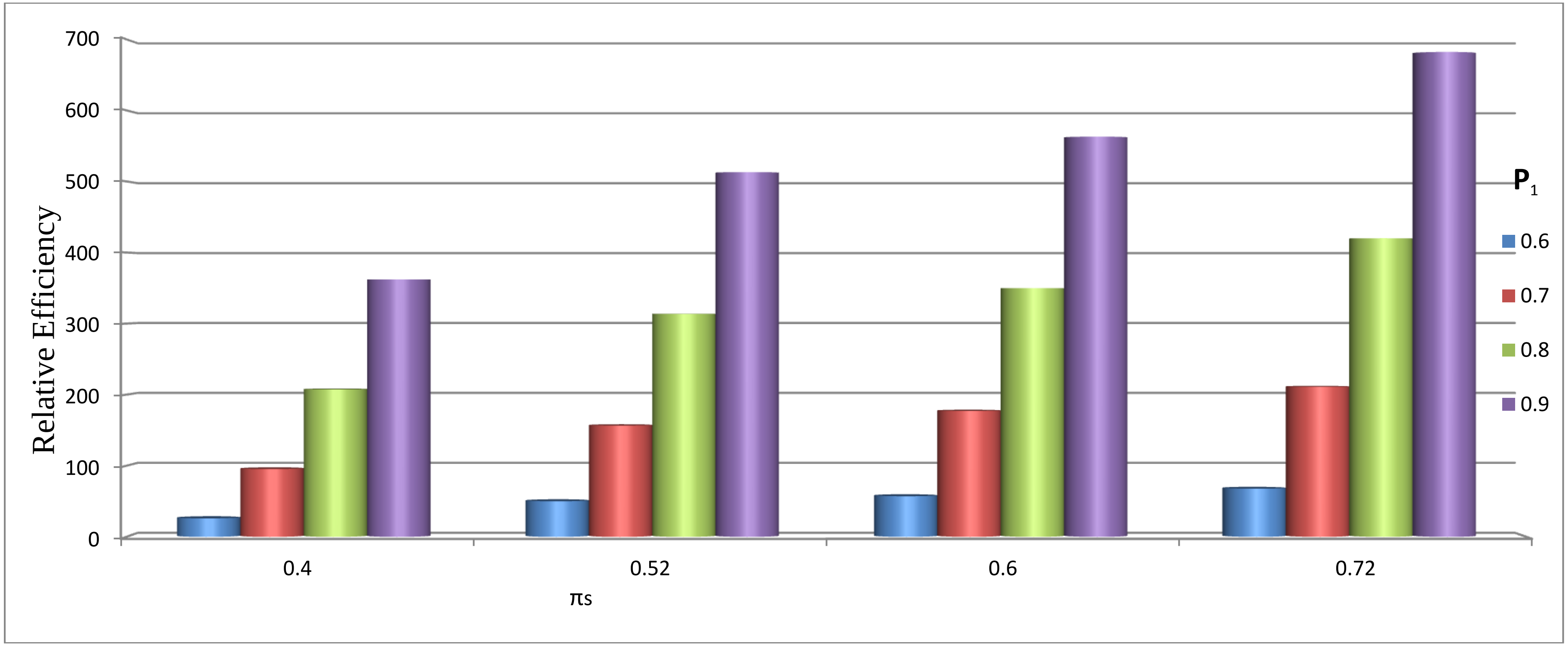


Table 2: Relative Efficiency of the Proposed estimator $\hat{\pi}_S$ with respect to Kim and Warde's (2004) Estimator $\hat{\pi}_{kw}$.

π_S	π_{S1}	π_{S2}	n = 1000	P_1= 0.6	P_1=0.65	P_1=0.7	P_1=0.75	P_1=0.8	P_1=0.85	P_1=0.9	P_1=0.93
			T	P_2 =0.65	P_2 =0.7	P_2 = 0.75	P_2 = 0.8	P_2 = 0.85	P_2 = 0.9	P_2 = 0.95	P_2 = 0.95
0.40	0.38	0.43	0.10	1.04	1.04	1.03	1.03	1.02	1.02	1.01	1.01
0.42	0.38	0.43	0.13	1.05	1.05	1.04	1.04	1.03	1.02	1.01	1.01
0.40	0.38	0.43	0.15	1.06	1.06	1.05	1.04	1.04	1.03	1.02	1.01
0.42	0.38	0.43	0.17	1.07	1.06	1.06	1.05	1.04	1.03	1.02	1.02
0.40	0.38	0.43	0.19	1.08	1.07	1.07	1.06	1.05	1.04	1.02	1.02
0.52	0.48	0.53	0.21	1.11	1.10	1.09	1.08	1.06	1.05	1.03	1.02
0.50	0.48	0.53	0.23	1.12	1.11	1.10	1.09	1.07	1.06	1.04	1.03
0.52	0.48	0.53	0.25	1.13	1.12	1.11	1.09	1.08	1.06	1.03	1.03
0.50	0.48	0.53	0.27	1.14	1.13	1.12	1.11	1.09	1.07	1.04	1.03
0.52	0.48	0.53	0.29	1.16	1.15	1.13	1.11	1.09	1.06	1.04	1.03
0.60	0.58	0.63	0.31	1.21	1.20	1.18	1.15	1.13	1.10	1.06	1.05
0.62	0.58	0.63	0.33	1.23	1.21	1.19	1.16	1.13	1.09	1.05	1.05
0.60	0.58	0.63	0.35	1.24	1.23	1.21	1.18	1.15	1.11	1.07	1.05
0.62	0.58	0.63	0.37	1.27	1.25	1.22	1.18	1.15	1.11	1.06	1.05
0.60	0.58	0.63	0.39	1.28	1.26	1.23	1.20	1.17	1.12	1.08	1.06
0.72	0.68	0.73	0.41	1.38	1.35	1.31	1.27	1.21	1.15	1.09	1.08
0.70	0.68	0.73	0.43	1.40	1.37	1.33	1.29	1.24	1.18	1.12	1.09
0.72	0.68	0.73	0.45	1.43	1.40	1.35	1.30	1.24	1.17	1.10	1.09
0.70	0.68	0.73	0.47	1.45	1.42	1.38	1.32	1.27	1.20	1.13	1.10
0.72	0.68	0.73	0.49	1.49	1.45	1.40	1.34	1.27	1.19	1.11	1.09

Figure 2: Relative Efficiency of the Proposed Estimator $\hat{\pi}_S$ with respect to Kim and Warde's (2004) Estimator $\hat{\pi}_{kw}$.

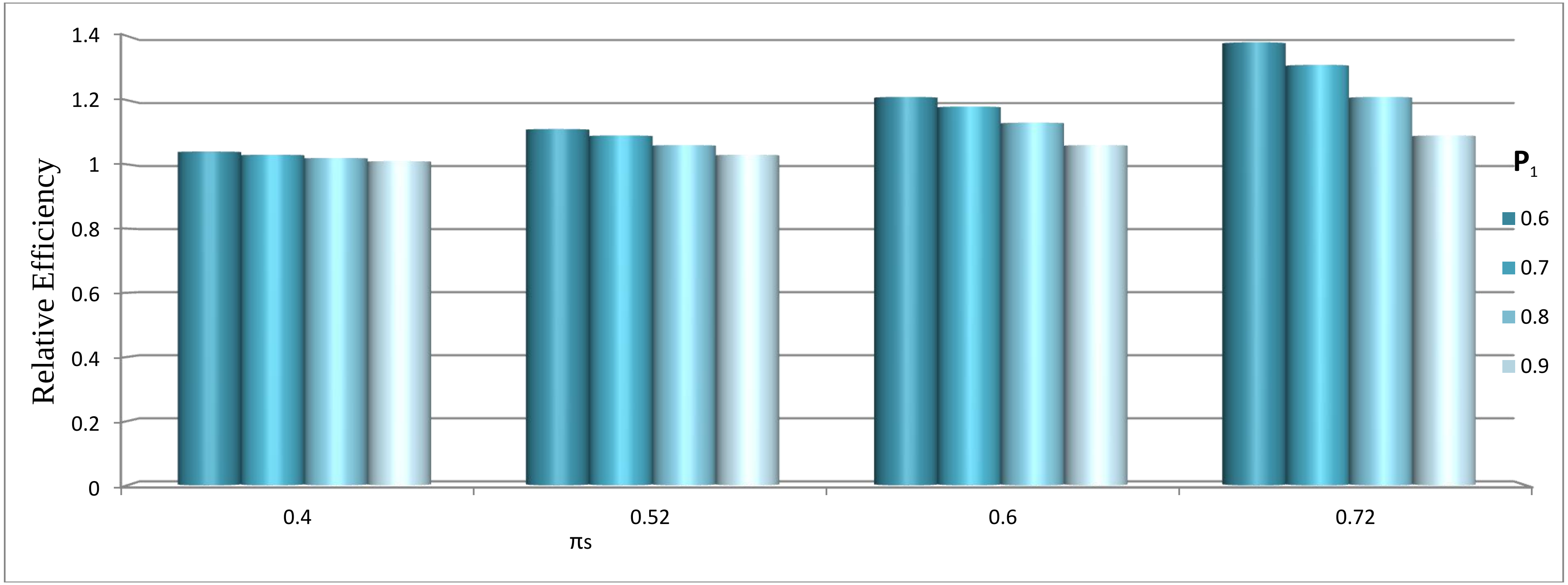


Table 3: Relative Efficiency of the Proposed Estimator $\hat{\pi}_S$ with respect to Huang (2004) Estimator $\hat{\pi}_H$.

π_S	π_{S1}	π_{S2}	n = 1000	P_1= 0.6	P_1=0.65	P_1=0.7	P_1=0.75	P_1=0.8	P_1=0.85	P_1=0.9	P_1=0.93
			T	P_2 =0.65	P_2 =0.7	P_2 = 0.75	P_2 = 0.8	P_2 = 0.85	P_2 = 0.9	P_2 = 0.95	P_2 = 0.95
0.40	0.38	0.43	0.10	1.23	1.17	1.13	1.11	1.09	1.08	1.07	1.03
0.42	0.38	0.43	0.13	1.70	1.47	1.35	1.28	1.23	1.20	1.17	1.07
0.40	0.38	0.43	0.15	1.23	1.17	1.13	1.11	1.09	1.08	1.07	1.03
0.42	0.38	0.43	0.17	1.70	1.47	1.35	1.28	1.23	1.20	1.17	1.06
0.40	0.38	0.43	0.19	1.23	1.17	1.13	1.11	1.09	1.08	1.07	1.03
0.52	0.48	0.53	0.21	1.69	1.46	1.34	1.27	1.22	1.19	1.16	1.06
0.50	0.48	0.53	0.23	1.23	1.17	1.13	1.10	1.09	1.07	1.07	1.03
0.52	0.48	0.53	0.25	1.69	1.46	1.34	1.27	1.22	1.19	1.16	1.06
0.50	0.48	0.53	0.27	1.23	1.16	1.13	1.10	1.09	1.07	1.06	1.03
0.52	0.48	0.53	0.29	1.69	1.46	1.34	1.27	1.22	1.18	1.16	1.06
0.60	0.58	0.63	0.31	1.23	1.16	1.13	1.10	1.09	1.07	1.06	1.03
0.62	0.58	0.63	0.33	1.69	1.46	1.34	1.27	1.22	1.18	1.16	1.06
0.60	0.58	0.63	0.35	1.23	1.16	1.13	1.10	1.08	1.07	1.06	1.03
0.62	0.58	0.63	0.37	1.69	1.45	1.33	1.26	1.21	1.18	1.15	1.06
0.60	0.58	0.63	0.39	1.23	1.16	1.12	1.10	1.08	1.07	1.06	1.02
0.72	0.68	0.73	0.41	1.69	1.46	1.34	1.27	1.22	1.19	1.16	1.06
0.70	0.68	0.73	0.43	1.23	1.16	1.13	1.10	1.09	1.07	1.06	1.03
0.72	0.68	0.73	0.45	1.69	1.45	1.34	1.26	1.22	1.18	1.16	1.06
0.70	0.68	0.73	0.47	1.23	1.16	1.13	1.10	1.08	1.07	1.06	1.03
0.72	0.68	0.73	0.49	1.68	1.45	1.33	1.26	1.21	1.18	1.15	1.06

Figure 3: Relative Efficiency of the Proposed Estimator $\hat{\pi}_S$ with respect to Huang (2004) Estimator $\hat{\pi}_H$.

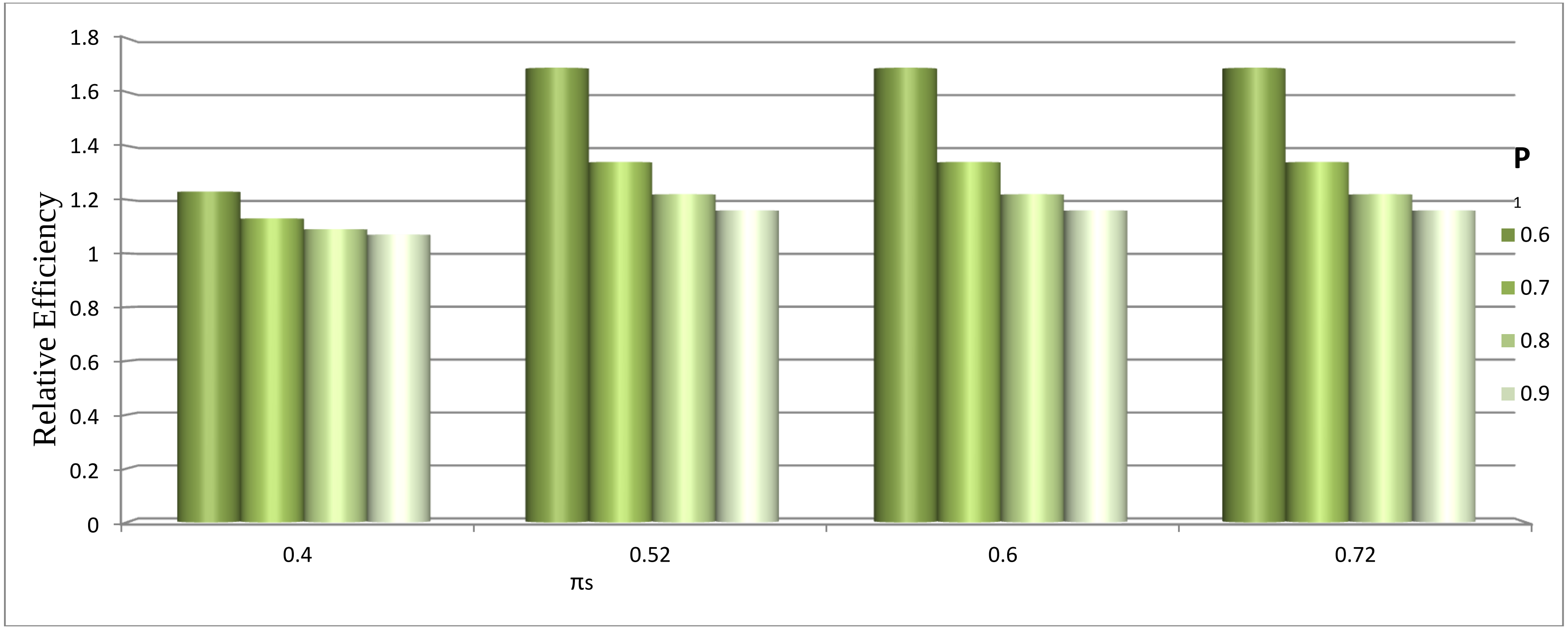
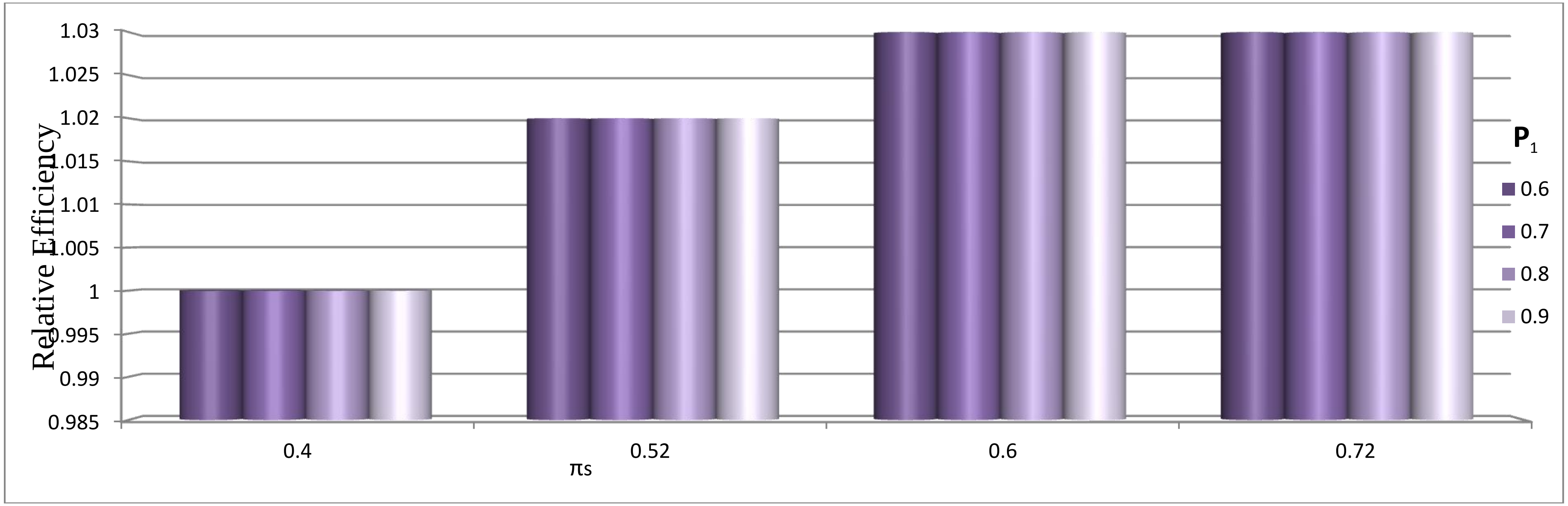


Table 4: Relative Efficiency of the Proposed Estimator $\hat{T}$ with respect to Huang (2004) Estimator $\hat{T}_H$.

π_S	π_{S1}	π_{S2}	n = 1000 T	P_1= 0.6 P_2 =0.65	P_1=0.65 P_2 =0.7	P_1=0.7 P_2 = 0.75	P_1=0.75 P_2 = 0.8	P_1=0.8 P_2 = 0.85	P_1=0.85 P_2 = 0.9	P_1=0.9 P_2 = 0.95	P_1=0.93 P_2 = 0.95
0.40	0.38	0.43	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.42	0.38	0.43	0.13	1.02	1.02	1.02	1.02	1.02	1.02	1.01	1.00
0.40	0.38	0.43	0.15	1.00	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.42	0.38	0.43	0.17	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.01
0.40	0.38	0.43	0.19	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.52	0.48	0.53	0.21	1.02	1.03	1.03	1.02	1.02	1.02	1.02	1.01
0.50	0.48	0.53	0.23	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.52	0.48	0.53	0.25	1.02	1.03	1.03	1.03	1.03	1.02	1.02	1.01
0.50	0.48	0.53	0.27	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.52	0.48	0.53	0.29	1.02	1.03	1.03	1.03	1.03	1.03	1.03	1.01
0.60	0.58	0.63	0.31	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.62	0.58	0.63	0.33	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.01
0.60	0.58	0.63	0.35	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.62	0.58	0.63	0.37	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.01
0.60	0.58	0.63	0.39	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.72	0.68	0.73	0.41	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.01
0.70	0.68	0.73	0.43	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.72	0.68	0.73	0.45	1.03	1.03	1.04	1.04	1.04	1.03	1.03	1.01
0.70	0.68	0.73	0.47	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
0.72	0.68	0.73	0.49	1.03	1.03	1.04	1.04	1.04	1.04	1.04	1.01

Figure 4: Relative Efficiency of the Proposed Estimator $\hat{T}$ with respect to Huang (2004) Estimator $\hat{T}_H$.



5. Discussion

This paper addresses the problem of estimating the proportion π_S of the population belonging to a sensitive group using randomized response technique in stratified sampling. An improved two – stage stratified randomized response model using Huang (2004) model has been proposed. It has been shown numerically that the proposed randomized response model is more efficient than the Kim and Warde (2004) and Huang (2004) stratified randomize response models. In addition to the gain in efficiency, the proposed method is more beneficial than the previous method in that a stratified randomize response method assists to solve the limitations of randomized response that is the loss of the individual characteristics of the respondents.

Acknowledgments

The authors are grateful to the editor in chief and to the learned referees for their valuable suggestions regarding improvement of this article.

References

Chang, H.J. and Huang, K.C. (2001). Estimation of proportion and sensitivity of a qualitative character. *Metrika*, **53**, 269-280.

Chang, H.J., Wang, C.L. and Huang, K.C. (2004a). Using randomized response to estimate the proportion and truthful reporting probability in a dichotomous finite population. *Journal of Applied Statistics*, **53**, 269-280.

Chang, H.J., Wang, C.L. and Huang, K.C. (2004b). On estimating the proportion of a qualitative sensitive character using randomized response sampling. *Quality and Quantity*, **38**, 675-680.

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response*: *Theory and Techniques*. Marcel- Dekker, New York, USA.

Cochran, W.G. (1977). *Sampling Technique*, 3rd *Edition*. New York: John Wiley and Sons, USA.

Fox, J.A. and Tracy, P.E. (1986): *Randomized Response*: *A method of Sensitive Surveys*. Newbury Park, CA: SAGE Publications.

Greenberg, B., Abul- Ela A., Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question randomized response: Theoretical framework. *Journal of the American Statistical Association*, **64**, 529-539.

Greenberg, B., Kubller, R.R., Abernathy, J.R., Horvitz, D.G. (1971). Applications of the RR technique in obtaining quantitative data. *Journal of the American Statistical Association*, **66**, 243-250.

Hedayat, A.S. and Sinha, B.K. (1991). *Design and Inference in finite Population Sampling*. New York: John Wiley and Sons, USA.

Hong, K., Yum, J. and Lee, H.(1994). A stratified randomized response technique. *Korean Journal of Applied Statistics*, **7**, 141-147.

Horvitz, D.G., Shah, B.V. and Simmons, W.R. (1967). The unrelated question randomized response model. *Proceedings of Social Statistical section. Journal of the American Statistical Association*, **62**, 65–72.

Huang, K.C. (2004). A survey technique for estimating the proportion and sensitivity in a dichotomous finite population. *Statistica Neerlandica*, **58**, 75-82.

Javed, M., Gupta, J.P. and Singh, S. (2002). An improved randomized response strategy using unrelated question. *Aligarh Journal of Statistics*, **22**, 17-25.

Kim, J.M. and Elam, M.E. (2005). A two- stage stratified Warner's randomized response model using optimal allocation. *Metrika*, **61**, 1-7.

Kim, J.M. and Elam, M.E. (2007). A stratified unrelated randomized response model. *Statistical Papers*, **48**, 215-233.

Kim, J.M. and Warde, W.D. (2004). A stratified Warner randomized response model. *Journal of Statistical Planning Inference*, **120**, 155-165.

Kim, J.M. and Warde, W.D. (2005). A mixed randomized response model. *Journal of Statistical Planning Inference,* **133**, 211-221.

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

Land, M., Singh, S. and Sedory, S.A. (2011). Estimation of a rare attribute using Poisson distribution. *Statistics*, **46(3)**, 351-360.

Lee, G.S., Uhm, D. and Kim, J.M. (2013). Estimation of a rare sensitive attribute in a stratified sample using Poisson distribution. *Statistics*, **47(3)**, 575-589.

Mahajan, P.K., Gunta, J.P., Singh, R. (1994). Determination of optimum strata boundaries for scrambled response. *Statistica*, **3**, 375-381.

Mahmood, M. Singh, S. and Horn, S. (1988). On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. *Biometrical Journal*, **40**, 237-242.

Mangat, N.S. (1994). An improved randomized response strategy. *Journal of Royal Statistical Society*, **B 56(1),** 93-95.

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

Mangat, N.S., Singh, R. and Singh, S. (1997). Violation of respondent's privacy in Moors model- its rectification through a random group strategy response model. *Communications in Statistics: Theory and Methods*, **26(3),** 243-255.

Ryu, J.B., Hong, K.H. and Lee, G.S. (1993). *Randomized Response Model*. Freedom Academy, Seoul, Korea.

Singh, H.P. and Tarray, T.A. (2012). A Stratified Unknown repeated trials in randomized response sampling. Common. *Korean Statistical Society*, **19(6)**, 751-759.

Singh, H.P. and Tarray, T.A. (2013). A modified survey technique for estimating the proportion and sensitivity in a dichotomous finite population. *International Journal of Advanced Scientific and Technical Research*,**3(6),** 459-472.

Singh, J. (1981). A note on estimation from randomized response surveys. *Journal of the Indian Society of Agricultural Statistics*, **33(2),** 89-93.

Singh, R. and Mangat, N.S. (1996). *Elements of Survey Sampling*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Singh, S. (1993). An alternative to Warner's randomized response technique. *Statistica*, annoLIII n.**1**, 67-69.

Singh, S. (2003). *Advanced Sampling Theory with Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Singh, S. and Singh, R. (1992). An alternative estimator for randomized response technique. *Journal of the Indian Society of Agricultural Statistics*, **44**, 149-154.

Singh, S., Singh, R. and Mangat. N.S. (2000). Some alternative strategies to Moor's model- in randomized response technique for eliminating evasive answer bias. *Journal of Statistical Planning Inference*, **83**, 243-255.

Tarray, T.A. (2016). *Statistical Sample Survey Methods and Theory*. Elite Publishers (onlinegatha.com), India.

Tarray, T.A., and Singh, H.P. (2014). A Proficient randomized response model. *Istatistika*: *Journal of Turkey Statistical Association*,**7(3),** 87-98.

Tarray, T.A., and Singh, H.P. (2015). A randomized response model for estimating a rare sensitive attribute in stratified sampling using Poisson distribution. *Model Assisted Statistics and Applications,* **10**, 345-360.

Tarray, T.A., and Singh, H.P. (2016). New procedures of estimating proportion and sensitivity using randomized response in a dichotomous finite population. *Journal of Modern Applied Statistical Methods,* **15(1),** 635-669.

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 Mukherjee. *Journal of Applied Statistical Science*, **4(2/3),**147–158.

Tracy, D.S. and Osahan, S.S. (1999). An improved randomized response technique. *Pakistan Journal of Statistics*, **15(1)**, 1–6.

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

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