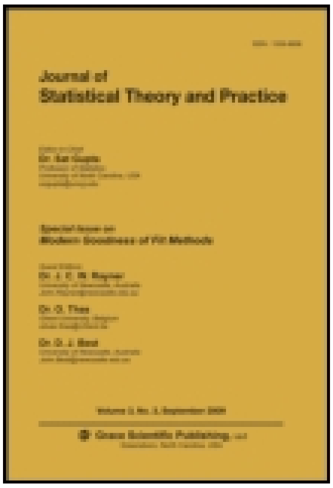

Journal of Statistical Theory and Practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ujsp20

Ratio Estimation of the Mean of a Sensitive Variable in the Presence of Auxiliary Information

Rita Sousa [a], Javid Shabbir [b], Pedro Corte Real [c] & Sat Gupta [d]
[a] New University of Lisbon & Statistics Portugal, Lisbon, Portugal
[b] Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan
[c] New University of Lisbon, Lisbon, Portugal
[d] University of North Carolina, Greensboro, NC, USA
Published online: 30 Nov 2011.

© Grace Scientific Publishing

Journal of
Statistical Theory and Practice
Volume 4, No. 3, September 2010

Ratio Estimation of the Mean of a Sensitive Variable in the Presence of Auxiliary Information

Rita Sousa, *New University of Lisbon & Statistics Portugal, Lisbon, Portugal. Email: rita.sousa@ine.pt.*

Javid Shabbir, *Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan. Email: jsqau@yahoo.com*

Pedro Corte Real, *New University of Lisbon, Lisbon, Portugal. Email: parcr@fct.unl.pt*

Sat Gupta, *University of North Carolina at Greensboro, NC, USA. Email: sngupta@uncg.edu*

Received: February 22, 2010 *Revised:* July 31, 2010

Abstract

We propose a ratio estimator for the mean of sensitive variable utilizing information from a non-sensitive auxiliary variable. Expressions for the *Bias* and *MSE* of the proposed estimator (correct up to first and second order approximations) are derived. We show that the proposed estimator does better than the ordinary RRT mean estimator that does not utilize the auxiliary information. We also show that there is hardly any difference in the first order and second order approximations for *MSE* even for small sample sizes. We also generalize the proposed estimator to the case of transformed ratio estimators but these transformations do not result in any significant reduction in *MSE*. An extensive simulation study is presented to evaluate the performance of the proposed estimator. The procedure is also applied to some financial data (purchase orders (sensitive variable) and gross turn-over (non-sensitive variable)) in 2009 for 5090 companies in Portugal from a survey on Information and Communication Technologies (ICT) usage.

AMS Subject Classification: 62D05.

Key-words: Ratio estimator; Randomized response technique; Mean square error; Absolute relative bias.

1. Introduction

In survey research, there are many situations when the primary variable of interest (Y) is sensitive and direct observation on this variable may not be possible. However, we may

* 1559-8608/10-3/$5 + $1pp - see inside front cover

be able to directly observe a highly correlated auxiliary variable (X). For example, Y may be the number of abortions a woman might have had in her life and X may be her age. Similarly Y may be the total purchase orders in a year for a company and X may be the total turn-over for that company in that year. In such cases, one will be able to observe X directly but will have to rely on some randomized response technique to collect information on Y. In such situations, mean of Y can be estimated by using one of many randomized response techniques (RRT) but this estimator can be improved considerably by utilizing information from the auxiliary variable X. Many authors have presented ratio estimators when both Y and X are directly observable. These include Kadilar and Cingi (2006), Turgut and Cingi (2008), Singh and Vishwakarma (2008), Koyuncu and Kadilar (2009) and Shabbir and Gupta (2010).

Also, many authors have estimated the mean of a sensitive variable when the primary variable is sensitive and there is no auxiliary variable available. These include Eichhorn and Hayre (1983), Gupta and Shabbir (2004), Gupta, Gupta and Singh (2002), Saha (2008) and Gupta, Shabbir and Sehra (2010).

In this paper, we propose a ratio estimator where the RRT estimator of the mean of Y is further improved by using information on an auxiliary variable X. Expressions for the *Bias* and *MSE* for the proposed estimator are derived, correct up to both the first order and second order approximations. It is shown that the two approximations are very similar even for moderate sample size. We also observe that there is considerable reduction in *MSE* when auxiliary information is used, particularly when the correlation between the study variable and the auxiliary variable is high.

Let Y be the study variable, a sensitive variable which cannot be observed directly. Let X be a non-sensitive auxiliary variable which is strongly correlated with Y. Let S be a scrambling variable independent of Y and X. The respondent is asked to report a scrambled response for Y given by $Z = Y + S$ but is asked to provide a true response for X. Let a random sample of size n be drawn without replacement from a finite population $U = U_1, U_2, \ldots, U_N$. For the ith unit ($i = 1, 2, \ldots, N$), let y_i and x_i respectively be the values of the study variable Y and auxiliary variable X. Moreover, let $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{z} = \frac{\sum_{i=1}^{n} z_i}{n}$ be the sample means and $\bar{Y} = E(Y)$, $\bar{X} = E(X)$ and $\bar{Z} = E(Z)$ be the population means for Y, X and Z, respectively. We assume that $\bar{X}$ is known and $\bar{S} = E(S) = 0$. Thus, $E(Z) = E(Y)$. Let us also define $\delta_z = \frac{\bar{z}-\bar{Z}}{\bar{Z}}$ and $\delta_x = \frac{\bar{x}-\bar{X}}{\bar{X}}$, such that $E(\delta_i) = 0$, $i = z, x$.

If information on X is ignored, then an unbiased estimator of μ_Y is the ordinary sample mean ($\bar{z}$) given by (1.1) below.

$$\hat{\mu}_Y = \bar{z}. \tag{1.1}$$

The mean square error (*MSE*) of $\hat{\mu}_Y$ is given by

$$MSE\left(\hat{\mu}_Y\right) = \frac{1-f}{n}\left(S_y^2 + S_s^2\right), \tag{1.2}$$

where $f = n/N$, $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ and $S_s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(s_i - \bar{S})^2$.

2. The Proposed Estimator

We propose the following ratio estimator for estimating the population mean of the study variable Y using the auxiliary variable X:

$$\hat{\mu}_R = \bar{z}\left(\frac{\bar{X}}{\bar{x}}\right) = \bar{Z}(1+\delta_z)(1+\delta_x)^{-1}. \tag{2.1}$$

Using Taylor's approximation and retaining terms of order up to 4, (2.1) can be rewritten as

$$\hat{\mu}_R - \bar{Z} \cong \bar{Z}\left\{\delta_z - \delta_x - \delta_z\delta_x + \delta_x^2 - \delta_x^3 + \delta_x^4 + \delta_z\delta_x^2 - \delta_z\delta_x^3\right\}. \tag{2.2}$$

Under the assumption of bivariate normality (see Sukhatme and Sukhatme, 1970), we have $E\left(\delta_z^2\right) = \frac{1-f}{n}C_z^2$, $E\left(\delta_x^2\right) = \frac{1-f}{n}C_x^2$, $E\left(\delta_x\delta_z\right) = \frac{1-f}{n}C_{zx}$, where $C_{zx} = \rho_{zx}C_zC_x$ and C_z and C_x are the coefficients of variation of Z and X, respectively. Also we have:

$$E(\delta_z\delta_x^3) = \left(\frac{1-f}{n}\right)^2 3\rho_{zx}C_zC_x^3, \quad E(\delta_z^2\delta_x^2) = \left(\frac{1-f}{n}\right)^2 (1+2\rho_{zx}^2)C_z^2C_x^2,$$

$$E(\delta_x^4) = \left(\frac{1-f}{n}\right)^2 3C_x^4, \quad E(\delta_z\delta_x^2) = E(\delta_z^2\delta_x) = E(\delta_x^3) = 0$$

and

$$C_z^2 = C_y^2 + \frac{S_s^2}{\bar{Y}^2}, \quad \rho_{zx} = \frac{\rho_{yx}}{\sqrt{1+\frac{S_s^2}{S_y^2}}}.$$

Recognizing that $\bar{Z} = \bar{Y}$ in Equation (2.2), we can get expressions for the *Bias* of $\hat{\mu}_R$, correct up to second order of approximation, as given by

$$Bias^{(2)}(\hat{\mu}_R) \cong Bias^{(1)}(\hat{\mu}_R) + 3\left(\frac{1-f}{n}\right)^2 \bar{Y}\left[C_x^4 - \rho_{zx}C_zC_x^3\right], \tag{2.3}$$

where

$$Bias^{(1)}(\hat{\mu}_R) \cong \frac{1-f}{n}\bar{Y}\left(C_x^2 - \rho_{zx}C_zC_x\right). \tag{2.4}$$

is the *Bias* corresponding to first order of approximation.

Similarly from (2.2), *MSE* of $\hat{\mu}_R$, correct to second order of approximation, is given by

$$MSE^{(2)}(\hat{\mu}_R) = E(\hat{\mu}_R - \bar{Z})^2 \cong \bar{Z}^2 E\left\{\delta_z - \delta_x - \delta_z\delta_x + \delta_x^2 - \delta_x^3 + \delta_x^4 + \delta_z\delta_x^2 - \delta_z\delta_x^3\right\}^2$$

or

$$MSE^{(2)}(\hat{\mu}_R) \cong \bar{Z}^2 E\left\{\delta_z^2 + \delta_x^2 - 2\delta_z\delta_x + 3\delta_z^2\delta_x^2 + 3\delta_x^4 - 6\delta_z\delta_x^3 - 2\delta_z^2\delta_x + 4\delta_z\delta_x^2 - 2\delta_x^3\right\}$$

Since $\bar{Z} = \bar{Y}$, we have

$$MSE^{(2)}(\hat{\mu}_R) \cong MSE^{(1)}(\hat{\mu}_R) + 3\bar{Y}^2 \left(\frac{1-f}{n}\right)^2 C_x^2 \left[(1+2\rho_{zx}^2)C_z^2 + 3C_x^2 - 6\rho_{zx}C_zC_x\right], \quad (2.5)$$

where

$$MSE^{(1)}(\hat{\mu}_R) \cong \frac{1-f}{n}\bar{Y}^2 \left(C_z^2 + C_x^2 - 2\rho_{zx}C_zC_x\right) \quad (2.6)$$

is the *MSE* corresponding to the first order approximation. The difference between the two approximations for *MSE* is given by

$$3\bar{Y}^2 \left(\frac{1-f}{n}\right)^2 C_x^2 \left[(1+2\rho_{zx}^2)C_z^2 + 3C_x^2 - 6\rho_{zx}C_zC_x\right]$$

and it converges to zero as $n \to N$. Our simulation results in Section 3 will also confirm this pattern.

According to the first order of approximation, $MSE^{(1)}(\hat{\mu}_R) < MSE(\hat{\mu}_Y)$ if

$$\left(\rho_{zx} - \frac{1}{2}\frac{C_x}{C_z}\right) > 0. \quad (2.7)$$

If 2nd order approximation is used, we can easily see that $MSE^{(2)}(\hat{\mu}_R) < MSE(\hat{\mu}_Y)$ if

$$2\rho_{zx}\frac{C_z}{C_x} + 3\frac{1-f}{n}\left[6\rho_{zx}C_zC_x - 3C_x^2 - (1+2\rho_{zx}^2)C_z^2\right] > 1. \quad (2.8)$$

3. A Simulation Study

In this section, we conduct a simulation study with particular focus on the following two issues:

a. How does the ratio estimator $\hat{\mu}_R$ compare with the RRT mean estimator $\hat{\mu}_Y$;
b. How do the *Bias* and *MSE* for the ratio estimator, correct up to second order of approximation, compare with the *Bias* and *MSE* expressions correct up to first order of approximation.

We considered 3 bivariate normal populations with different covariance matrices to represent the distribution of (*Y, X*). The scrambling variable *S* is taken to be a normal distribution with mean equal to zero and standard deviation equal to 10% of the standard deviation of *X*. The reported response is given by *Z* = *Y*+*S*.

All of the simulated populations have theoretical mean of [*Y, X*] as $\mu = [2\ 2]$ and covariance matrices as given below.

Population 1

$$N = 1000$$
$$\sigma^2 = \begin{bmatrix} 9 & 1.9 \\ 1.9 & 4 \end{bmatrix}, \quad \rho_{XY} = 0.3167.$$

Population 2

$$N = 1000$$
$$\sigma^2 = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}, \quad \rho_{XY} = 0.6708.$$

Population 3

$$N = 1000$$
$$\sigma^2 = \begin{bmatrix} 6 & 3 \\ 3 & 2 \end{bmatrix}, \rho_{XY} = 0.8660.$$

For each population we considered five sample sizes: n = 20, 50, 100, 200 and 300.

The absolute relative bias (ARB) for the two estimators is given by $|Bias(\hat{\mu}_Y)/\bar{Y}|$ and $|Bias(\hat{\mu}_R)/\bar{Y}|$. We estimate ARB using 5000 samples of size n selected from each population. The empirical ARB values for both estimators are given in Table 3.1. As expected, ARB generally decreases as the sample size increases, with some exceptions due to random fluctuations.

Table 3.1. Empirical absolute relative bias for RRT mean estimator and ratio estimator (bold).

Population		**Empirical *ARB***				
N	ρ_{XY}	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 300$
1000	0.3549	0.0021	0.0011	0.0010	0.0018	0.0016
		0.0223	**0.0071**	**0.0057**	**0.0006**	**0.0009**
	0.6965	0.0010	0.0021	0.0014	0.0014	0.0011
		0.0193	**0.0061**	**0.0015**	**0.0029**	**0.0021**
	0.8783	0.0012	0.0013	0.0008	0.0013	0.0011
		0.0181	**0.0063**	**0.0023**	**0.0026**	**0.0020**

The RRT mean estimator should generally perform better than the ratio estimator because this is an unbiased estimator. Nevertheless, the ratio estimator produces fairly good results.

The theoretical ARB results for the ratio estimator, correct to first and second degree of approximation, are presented in Table 3.2.

One can see that second order approximation as compared to first order approximation does not result in major difference in ARB even for modest sample size of n = 20 and 50.

Table 3.3 below gives empirical and theoretical *MSE's* for the ratio estimator based on both the 1st order and 2nd order approximations. As we see from the table, there is hardly a difference between the two approximations even for small samples. We use the following expression to find the percent relative efficiency (PRE) of ratio estimator as compared to the RRT mean estimator:

$$PRE = \frac{MSE(\hat{\mu}_Y)}{MSE(\hat{\mu}_R)} \times 100\,.$$

All the percent relative efficiencies are greater than 100 indicating that the ratio estimator is better than the RRT mean estimator.

Table 3.2. Theoretical absolute relative bias for ratio estimator based on 1st and 2nd order (bold) approximation.

Population		Theoretical *ARB*				
N	ρ_{XY}	$n=20$	$n=50$	$n=100$	$n=200$	$n=300$
1000	0.3549	0.0224	0.0087	0.0041	0.0018	0.0011
		0.0258	**0.0092**	**0.0042**	**0.0019**	**0.0011**
	0.6965	0.0155	0.0060	0.0029	0.0013	0.0007
		0.0167	**0.0062**	**0.0029**	**0.0013**	**0.0007**
	0.8783	0.0142	0.0055	0.0026	0.0012	0.0007
		0.0153	**0.0057**	**0.0026**	**0.0012**	**0.0007**

Table 3.3. *MSE* correct to 1st and 2nd order approximations and *PRE* for the ratio estimator relative to the RRT mean estimator.

Population			*MSE* Estimation			*MSE* Condition		*PRE*	
N	ρ_{XY}	n	Empirical	1st-Order	2nd-Order	1st-Order	2nd-Order	1st-Order	2nd-Order
1000	0.3549	20	0.5782	0.4462	0.5249	0.0340	0.6947	104.86	89.15
		50	0.1837	0.1730	0.1848		0.9464		98.15
		100	0.0819	0.0820	0.0846		1.0304		101.57
		200	0.0358	0.0364	0.0370		1.0723		103.37
		300	0.0219	0.0212	0.0214		1.0863		103.98
	0.6965	20	0.3434	0.3036	0.3327	0.4785	2.9075	171.63	156.65
		50	0.1202	0.1177	0.1221		3.0887		165.49
		100	0.0548	0.0558	0.0568		3.1492		168.67
		200	0.0248	0.0248	0.0250		3.1794		170.30
		300	0.0152	0.0145	0.0145		3.1894		170.85
	0.8783	20	0.1178	0.1012	0.1139	0.5919	2.9424	309.31	274.89
		50	0.0406	0.0392	0.0412		3.0185		294.99
		100	0.0183	0.0186	0.0190		3.0439		302.36
		200	0.0083	0.0083	0.0083		3.0565		306.18
		300	0.0050	0.0048	0.0048		3.0608		307.48

MSE comparison condition based on 1st order approximation given in expression (2.7) .
MSE comparison condition based on 2nd order approximation given in expression (2.8) .

There are small differences between *MSE* values based on 1st and 2nd order approximations for smaller sample sizes (n=20 and 50) but the *MSE* values are very similar when the sample size is larger. We can also note that the ratio estimator gets more and more efficient as the coefficient of correlation between X and Y increases. We can further note that for small correlation values, the ratio estimator may not be better than the RRT mean estimator, particularly so if sample size is small.

4. Numerical Example

We now compare the RRT mean estimator and the ratio estimator using a real data set. The data come from a sample from the survey on Information and Communication Technologies (ICT) usage in enterprises in 2009 with seat in Portugal (Smilhily and Storm, 2010). This survey intends to promote the development of the national statistical system in the information society and to contribute to a deeper knowledge about the usage of ICT by enterprises. The target population covers all industries with one and more persons employed in the sections of economic activity C to N and S, from NACE (Statistical Classification of Economic Activities in the European Community) Rev. 2 (Eurostat, 2008). The data are essentially collected using Electronic Data Interchange, applying direct connection between information systems at the respondent and the National Statistics Institute. For some enterprises the paper questionnaire is still used. The questions in the structural business surveys mainly deal with characteristics that can be found in the organisations' annual reports and financial statements, such as employment, turnover and investment.

In our application the study variable Y is the purchase orders in 2009, collected by the ICT survey in that year. This is typically a confidential variable for enterprises, only known from business surveys. The auxiliary variable X is the turnover of each enterprise. This information can be easily obtained from enterprise records available in the public domain, as administrative information. In 2009 the population survey contained approximately 278000 enterprises and we know the value of X for all these enterprises. The purchase orders information was collected in the ICT survey and we have the values of Y for 5090 enterprises (which answered this question in the ICT survey in 2009). For this study, these 5090 enterprises are considered as our population. The scrambling variable S is taken to be a normal random variable with mean equal to zero and standard deviation equal to 10% of the standard deviation of X, that is $\sigma_S = 0.1\sigma_X$ The reported response is given by $Z = Y{+}S$ (the purchase order value plus a random quantity). The variables Y and X are strongly correlated so we can take advantage of this correlation by using the ratio estimator. In the next tables we present the results for the RRT mean estimator and for the ratio estimator for different sample sizes.

<u>Population Characteristics</u>:

$N = 5090$, $\rho_{XY} = 0.9832$

$\mu_X = 32.53$, $\mu_Y = 26.06$, $\sigma_X = 183.42$, $\sigma_Y = 67.07$ (in millions of Euros)

$\gamma_1^X = 31.54$, $\gamma_1^Y = 36.12$, $\gamma_2^X = 1481.08$, $\gamma_2^Y = 1839.13$, where γ_1 and $\gamma_2$are the coefficients of *skewness* and *kurtosis*, respectively. We use the following sample sizes in our simulation study:

n = 100, 200, 300, 400, 500, 1000, 1500 and 2000.

The empirical *ARB* values for both estimators, based on 5000 iterations, are given in Table 4.1. As expected, the bias decreases as the sample size increases, except for some random fluctuation. We expect the RRT mean estimator to perform better than the ratio estimator because this is an unbiased estimator, however, we don't see major differences between the two for larger samples.

The theoretical *ARB* results for the ratio estimator, correct to 1st degree of approximation, are presented in Table 4.2. We use only the 1st order approximations from here on since the 1st and 2nd order approximations are very similar, as we have seen earlier.

Table 4.3 presents the results for the empirical *MSE* estimates, the theoretical estimates, correct to 1st degree of approximation, and the *PRE* of ratio estimator relative to the RRT mean estimator.

Clearly the ratio estimator performs better than the RRT mean estimator for the real data also. The effect of sample size on the *PRE* calculation is neutralized when 1st order approximation is used, as can be seen from Equations (1.2) and (2.6).

Table 4.1. Empirical absolute relative bias for the RRT mean estimator and theratio estimator(bold).

Population		**Empirical *ARB***							
N	ρ_{XY}	$n=100$	$n=200$	$n=300$	$n=400$	$n=500$	$n=1000$	$n=1500$	$n=2000$
5090	0.9832	0.0219	0.0002	0.0096	0.0107	0.0163	0.0145	0.0106	0.0096
		0.0284	**0.0198**	**0.0171**	**0.0183**	**0.0166**	**0.0149**	**0.0127**	**0.0121**

Table 4.2. Theoretical absolute relative bias for ratio estimator correct to 1st order approximation.

Population		**Theoretical *ARB***							
N	ρ_{XY}	$n=100$	$n=200$	$n=300$	$n=400$	$n=500$	$n=1000$	$n=1500$	$n=2000$
5090	0.9832	0.0368	0.0180	0.0118	0.0086	0.0068	0.0030	0.0018	0.0011

Table 4.3. *MSE*, correct to 1st order approximation, and *PRE* for the ratio estimator relative to the RRT mean estimator.

Population			***MSE* Estimation**		***PRE***
N	ρ_{XY}	n	**Empirical**	***Theoretical***	
5090	0.9832	100	12.8924	15.2630	2286.36
		200	6.4608	7.4786	
		300	4.4498	4.8838	
		400	3.5279	3.5864	
		500	2.7380	2.8079	
		1000	1.4117	1.2510	
		1500	0.8805	0.7321	
		2000	0.6033	0.4726	

5. Transformed Ratio Estimators

Now consider the transformed ratio estimator

$$\hat{\mu}_{TR} = \bar{z}\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right), \tag{5.1}$$

where c and d are the unit-free parameters, which may be quantities such as the coefficient of *skewness* and coefficient of *kurtosis* for X. Many researchers have used transformed ratio estimators. These include Sisodia and Dwivedi (1981), Singh, Panday and Hirano (1973), Kulkarni (1977), Upadhyaya and Singh (1999), Upadhyaya, Singh and Singh (2000) and Chandra and Singh (2005).

We can rewrite (5.1) using relative error terms in the form

$$\hat{\mu}_{TR} = \bar{z}\left(1+\delta_z\right)\left(1+\eta\,\delta_x\right)^{-1}, \tag{5.2}$$

where $\eta = \frac{c\bar{X}}{c\bar{X}+d}$.

Expanding (5.2), the *Bias,* correct to first order of approximation, is given by

$$Bias^{(1)}(\hat{\mu}_{TR}) \cong \frac{1-f}{n}\bar{Y}\left\{\eta^2 C_x^2 - \eta\rho_{zx}C_z C_x\right\}. \tag{5.3}$$

By (2.4) and (5.3) $Bias^{(1)}(\hat{\mu}_{TR}) < Bias^{(1)}(\hat{\mu}_R)$ if

$$(\eta - 1)\left\{\rho_{zx} - \frac{(\eta+1)C_x}{C_z}\right\} > 0. \tag{5.4}$$

Similarly MSE of $\hat{\mu}_{TR}$, to first order of approximation, is given by

$$MSE^{(1)}(\hat{\mu}_{TR}) \cong \frac{1-f}{n}\bar{Y}^2\left\{C_z^2 + \eta^2 C_x^2 - 2\eta\rho_{zx}C_z C_x\right\}. \tag{5.5}$$

By (2.6) and (5.5), $MSE^{(1)}(\hat{\mu}_{TR}) < MSE^{(1)}(\hat{\mu}_R)$ if

$$(\eta - 1)\left\{\rho_{zx} - \frac{(\eta+1)C_x}{2C_z}\right\} > 0. \tag{5.6}$$

Now we conduct a simulation study with particular focus on the comparison between the ratio estimator $\hat{\mu}_R$ and the transformed ratio estimator $\hat{\mu}_{TR}$. We consider the same three bivariate normal populations as in the previous simulation study (Section 3). The scrambling variable *S* is taken to be a normal random variable with mean equal to zero and the standard deviation equal to 10% of the standard deviation of *X*. The reported response is given by *Z* = *Y*+*S*. To compare these estimators, we present the results for the RRT mean estimator ($\hat{\mu}_Y$), the ratio estimator ($\hat{\mu}_R$) and for transformed ratio estimator $\hat{\mu}_{TRi}(i = 1,2,3,4)$ with 4 different combinations of parameters *c* and *d*:

1. $\hat{\mu}_{TR1} = \bar{z}\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right)$, where *c* = 1 and *d* = coefficient of *skewness*;
2. $\hat{\mu}_{TR2} = \bar{z}\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right)$, where *c* = 1 and *d* = coefficient of *kurtosis*;
3. $\hat{\mu}_{TR3} = \bar{z}\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right)$, where *c* = coefficient of *skewness* and *d* = coefficient of *kurtosis*;
4. $\hat{\mu}_{TR4} = \bar{z}\left(\frac{c\bar{X}+d}{c\bar{x}+d}\right)$, where *c* = coefficient of *kurtosis* and *d* = coefficient of *skewness*.

The empirical *ARB* values for these six estimators are given in Table 5.1.

The empirical *ARB* results in the table 5.1 and the theoretical *ARB* results, to 1st degree of approximation, in the table 5.2 indicate that the transformed ratio estimators do not produce major reductions in *ARB* as compared to the ratio estimator when sample size is large. Some reduction is observed for small sample size when using transformations where the additive parameter (*d*) is the *kurtosis*.

Tables 5.3 and 5.4 present the results for the empirical *MSE* estimates and for the theoretical estimates, correct to 1st order of approximation, respectively. Both tables indicate that modest gains can be achieved by using transformations where the additive parameter (*d*) is the coefficient of *skewness*.

Table 5.1. Empirical absolute relative bias.

Population			Empirical *ARB*					
N	ρ_{XY}	n	$\hat{\mu}_Y$	$\hat{\mu}_R$	$\hat{\mu}_{TR1}$	$\hat{\mu}_{TR2}$	$\hat{\mu}_{TR3}$	$\hat{\mu}_{TR4}$
1000	0.3209	20	0.0002	0.0337	0.0435	0.0006	0.0026	0.0366
		50	0.0007	0.0118	0.0146	0.0002	0.0019	0.0126
		100	0.0003	0.0052	0.0065	0.0000	0.0009	0.0056
		150	0.0000	0.0032	0.0040	0.0001	0.0003	0.0035
		200	0.0012	0.0025	0.0030	0.0008	0.0016	0.0027
		300	0.0020	0.0041	0.0045	0.0023	0.0021	0.0043
	0.6746	20	0.0011	0.0122	0.0113	0.0111	0.0018	0.0119
		50	0.0004	0.0042	0.0038	0.0037	0.0015	0.0041
		100	0.0001	0.0022	0.0021	0.0021	0.0005	0.0022
		150	0.0005	0.0016	0.0015	0.0016	0.0002	0.0016
		200	0.0010	0.0005	0.0005	0.0001	0.0013	0.0005
		300	0.0015	0.0013	0.0014	0.0011	0.0016	0.0014
	0.8684	20	0.0006	0.0120	0.0115	0.0108	0.0013	0.0119
		50	0.0005	0.0041	0.0039	0.0036	0.0012	0.0040
		100	0.0001	0.0018	0.0017	0.0017	0.0005	0.0018
		150	0.0002	0.0010	0.0010	0.0012	0.0001	0.0010
		200	0.0009	0.0004	0.0004	0.0001	0.0011	0.0004
		300	0.0014	0.0010	0.0011	0.0010	0.0015	0.0010

Table 5.2. Theoretical absolute relative bias to 1st order approximation.

Population			*ARB* (1st Order)				
N	ρ_{XY}	n	$\hat{\mu}_R$	$\hat{\mu}_{TR1}$	$\hat{\mu}_{TR2}$	$\hat{\mu}_{TR3}$	$\hat{\mu}_{TR4}$
1000	0.3209	20	0.0248	0.0310	0.0017	0.0031	0.0267
		50	0.0096	0.0120	0.0006	0.0012	0.0103
		100	0.0046	0.0057	0.0003	0.0006	0.0049
		150	0.0029	0.0036	0.0002	0.0004	0.0031
		200	0.0020	0.0025	0.0001	0.0003	0.0022
		300	0.0012	0.0015	0.0001	0.0001	0.0013
	0.6746	20	0.0124	0.0116	0.0108	0.0031	0.0121
		50	0.0048	0.0045	0.0042	0.0012	0.0047
		100	0.0023	0.0021	0.0020	0.0006	0.0022
		150	0.0014	0.0013	0.0012	0.0004	0.0014
		200	0.0010	0.0009	0.0009	0.0003	0.0010
		300	0.0006	0.0006	0.0005	0.0001	0.0006
	0.8684	20	0.0123	0.0118	0.0108	0.0020	0.0121
		50	0.0048	0.0046	0.0042	0.0008	0.0047
		100	0.0023	0.0022	0.0020	0.0004	0.0022
		150	0.0014	0.0014	0.0013	0.0002	0.0014
		200	0.0010	0.0010	0.0009	0.0002	0.0010
		300	0.0006	0.0006	0.0005	0.0001	0.0006

Table 5.3. Empirical *MSE* for the RRT mean estimator, the ratio estimator and and for the transformed ratio estimators.

Population			Empirical *MSE*					
N	ρ_{XY}	n	$\hat{\mu}_Y$	$\hat{\mu}_R$	$\hat{\mu}_{TR1}$	$\hat{\mu}_{TR2}$	$\hat{\mu}_{TR3}$	$\hat{\mu}_{TR4}$
0.1000	0.3209	20	0.4445	0.5799	0.6584	0.4097	0.4686	0.6010
		50	0.1702	0.1881	0.2000	0.1546	0.1793	0.1915
		100	0.0831	0.0872	0.0914	0.0750	0.0875	0.0884
		150	0.0524	0.0546	0.0571	0.0475	0.0551	0.0554
		200	0.0376	0.0395	0.0412	0.0343	0.0394	0.0400
		300	0.0216	0.0223	0.0232	0.0196	0.0227	0.0225
	0.6746	20	0.4915	0.3226	0.3197	0.3956	0.5167	0.3216
		50	0.1883	0.1165	0.1148	0.1497	0.1979	0.1159
		100	0.0920	0.0558	0.0548	0.0728	0.0967	0.0554
		150	0.0580	0.0352	0.0346	0.0460	0.0608	0.0350
		200	0.0414	0.0254	0.0250	0.0330	0.0434	0.0253
		300	0.0238	0.0145	0.0142	0.0189	0.0250	0.0144
	0.8684	20	0.2957	0.1117	0.1083	0.1973	0.3119	0.1106
		50	0.1133	0.0396	0.0382	0.0743	0.1195	0.0391
		100	0.0553	0.0188	0.0181	0.0361	0.0584	0.0186
		150	0.0349	0.0118	0.0114	0.0228	0.0367	0.0117
		200	0.0249	0.0086	0.0083	0.0164	0.0263	0.0085
		300	0.0143	0.0049	0.0047	0.0094	0.0151	0.0048

Table 5.4. Theoretical *MSE* to 1st order of approximation for the RRT mean estimator, the ratio estimator and for the transformed ratio estimators.

Population			Theoretical *MSE* (1st Order)				
N	ρ_{XY}	n	$\hat{\mu}_R$	$\hat{\mu}_{TR1}$	$\hat{\mu}_{TR2}$	$\hat{\mu}_{TR3}$	$\hat{\mu}_{TR4}$
1000	0.3209	20	0.4496	0.4672	0.3994	0.4663	0.4548
		50	0.1743	0.1811	0.1549	0.1808	0.1763
		100	0.0826	0.0858	0.0734	0.0857	0.0835
		150	0.0520	0.0540	0.0462	0.0539	0.0526
		200	0.0367	0.0381	0.0326	0.0381	0.0371
		300	0.0214	0.0222	0.0190	0.0222	0.0217
	0.6746	20	0.2939	0.2884	0.3885	0.5162	0.2961
		50	0.1140	0.1118	0.1507	0.2002	0.1132
		100	0.0540	0.0530	0.0714	0.0948	0.0536
		150	0.0340	0.0333	0.0449	0.0597	0.0338
		200	0.0240	0.0235	0.0317	0.0421	0.0238
		300	0.0140	0.0137	0.0185	0.0246	0.0139
	0.8684	20	0.0984	0.0947	0.1920	0.3113	0.0971
		50	0.0381	0.0367	0.0744	0.1207	0.0377
		100	0.0181	0.0174	0.0353	0.0572	0.0178
		150	0.0114	0.0110	0.0222	0.0360	0.0112
		200	0.0080	0.0077	0.0157	0.0254	0.0079
		300	0.0047	0.0045	0.0091	0.0148	0.0046

Table 5.5 gives the percent relative efficiency (*PRE*) of various transformed ratio estimators relative to the ratio estimator based on 1st order approximation.

Table 5.5. *PRE* for the transformed ratio estimator relative to the ratio estimator based on 1st order of approximation.

Population		***PRE* (1st Order)**			
N	ρ_{XY}	$\hat{\mu}_{TR1}$	$\hat{\mu}_{TR2}$	$\hat{\mu}_{TR3}$	$\hat{\mu}_{TR4}$
	0.3209	96.24	112.56	96.41	98.86
1000	0.6746	101.93	75.65	56.94	100.64
	0.8684	103.87	51.24	31.60	101.28

We can observe that the transformed ratio estimators that utilize the parameter d as coefficient of *skewness* result in higher *PRE* as compared to the ratio estimator when the correlation is larger. This was expected based on the condition in (5.6) and Table 5.6 below. Note that the transformed ratio estimator performs better than the ratio estimator when the condition in (5.6) is satisfied.

Table 5.6. Calculations for the expression in (5.6).

Population		**Condition (*MSE* - 1st Order)**			
N	ρ_{XY}	$\hat{\mu}_{TR1}$	$\hat{\mu}_{TR2}$	$\hat{\mu}_{TR3}$	$\hat{\mu}_{TR4}$
	0.3209	−0.0299	0.0855	−0.0285	−0.0088
1000	0.6746	0.0127	−0.2160	−0.5074	0.0043
	0.8684	0.0108	−0.2751	−0.6259	0.0037

6. Conclusions

We can observe from this study that the estimation of the mean of a sensitive variable can be improved by using a non-sensitive auxiliary variable. The ratio estimators, in spite of being biased, can have much better *PRE* as compared to the RRT mean estimator. Our simulation study and the numerical example show that this improvement can be quite substantial if the correlation between the study variable and the auxiliary variable is high. We also note that there is hardly any difference in the *Bias* or *MSE* of the proposed estimator when using 1st or 2nd order approximation. It is further noticed that the transformed ratio estimators produce very minimal gain over the ordinary ratio estimator.

References

Chandra, P., Singh, H.P., 2005. A family of estimators for population variance using knowledge of kurtosis of an auxiliary variable in sample surveys. *Statistics in Transition*, 7(1), 27–34

Eichhron, B.H., Hayre, L.S., 1983. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, 7, 307–316.

Eurostat, 2008. NACE Rev. 2 - Statistical classification of economic activities in the European Community. *Official Publications of the European Communities,* 112–285 and 306–311

Gupta, S.N., Gupta, B.C., Singh, S., 2002. Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference,* 100, 239–247

Gupta, S., Shabbir, J., 2004. Sensitivity estimation for personal interview survey questions. *Statistica,* 64, 643–653

Gupta, S., Shabbir, J., Sehra, S., 2010, Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference,* 140(10), 2870–2874

Kadilar, C., Cingi, H., 2006. Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters,* 19(1), 75–79

Koyuncy, N., Kadilar, C., 2009. Efficient estimators for the population mean. *Hacettepe Journal of Mathematics and Statistics,* 38(2), pp. 217–225

Kulkarni, S.P., 1977. A note on modified ratio estimator using transformation. *Journal of the Indian Society of Agricultural Statistics,* 30(2), 125–128

Saha, A., 2008. A randomized response technique for quantitative data under unequal probability sampling. *Journal of Statistical Theory and Practice,* 2(4), 589–596

Shabbir, J., Gupta, S., 2010. Estimation of the finite population mean in two-phase sampling when auxiliary variables are attribute. *Hacettepe Journal of Mathematics and Statistics,* 39(1), 121–129

Singh, H.P., Vishwakarma, G., 2008. Some families of estimators of variance of stratified random sample mean using auxiliary information. *Journal of Statistical Theory and Practice,* 2(1), 21–43

Singh, J., Panday, B.N., Hirano, K., 1973. On the utilization of known coefficient of kurtosis in the estimation procedure of variable. *American Institute of Statistical Mathematics,* 25, 51–55.

Sisodia, B.V.S., Dwivedi, V.K., 1981. A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of the Indian Society of Agricultural Statistics*, 33, 13–18

Smilhily, M., Storm, H., 2010. ICT usage in enterprises 1 - 2009. *Eurostat Publications,* Issue 1.

Sukhatme, P.V., Sukhatme, B.V., 1970. *Sampling Theory of Surveys with Applications*. Iowa State University Press.

Turgut, Y., Cingi, H., 2008. New generalized estimators for the population variance using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 37(2), 177–184

Upadhyaya, L.N., Singh, H.P., 1999. Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41(5), 627–636

Upadhyaya L.N., Singh, G.N., Singh, H.P., 2000. Use of transformed auxiliary variable in the estimation of population ratio in sample surveys. *Statistics in Transition*, 4(6), 1019–1027.