Thailand Statistician
July 2020; 18(3): 281-289
http://statassoc.or.th
Contributed paper

Simulation Study of Ratio Type Estimators in Stratified Random Sampling Using Multi-Auxiliary Information

Vishwantra Sharma* and Sunil Kumar
Department of statistics, University of Jammu, Jammu, India.
*Corresponding author; e-mail: vishwantrasharma07@gmail.com

Received: 12 July 2018
Revised: 11 February 2019
Accepted: 14 May 2019

Abstract

The paper addresses the problem of estimating the population mean of the study variable in stratified random sampling by using multi-auxiliary variable. In this paper, we proposed a ratio type estimator for estimating the population mean of the study variable by using multi-auxiliary variables. Stratified random sampling is taken into consideration. The expressions for the bias and mean square error (MSE) of the proposed estimator have been derived. The proposed estimator is compared with other existing estimators in terms of efficiency. An empirical study with the aid of simulation has also been carried out to validate the theoretical results obtained. The theoretical and empirical studies reveal that the proposed estimator performs better than existing estimators in the literature.

Keywords: Ratio estimator, bias, mean squared error, multi-auxiliary variables.

1. Introduction

In sample survey, we must consider the factor that has impact on the efficacy, cost and precision of the survey. The foremost thing that comes to one's mind when we are in this field is that which sampling procedure we need to choose. We know sample should be selected in such a manner that it represents the entire population of interest. Having such sample reduces the sampling error. The first plan available in literature is simple random sampling but it is not suitable when population is heterogeneous in nature. In such case, stratified random sampling plays an important role by overcoming the drawbacks of simple random sampling. In this plan the heterogeneous population is stratified into homogeneous sub-populations, defined as strata, such that within strata population should be homogeneous and between strata population should be heterogeneous, then a random sample of predetermined size is drawn from each stratum.

The use of auxiliary information improves the efficiency of the estimator. The auxiliary variable is taken into consideration in this study because it carries information about the study variable. Correlation between the two variables is very high. For example, in an agricultural survey, total cultivated area is known in advance and it acts as an auxiliary variable. D'Orazio and Catanese (2016) studied stratification in business and agriculture surveys in the presence of different auxiliary variables and concluded that more the number of auxiliary variables correlated with the target variable, the higher will be the benefits in using them for stratification. Haq and Shabbir (2013), Kadilar and Cingi

(2003), Shabbir and Gupta (2006) have proposed estimators in stratified sampling using information on single auxiliary variable. Bahl and Kumar (2000) proposed ratio and product type estimators for estimating the mean of the finite population using multi-auxiliary variables. Tailor et al. (2012) and Koyuncu and Kadilar (2009) proposed estimators of population mean using two auxiliary variables in stratified random sampling. Singh and Kumar (2012) have proposed improved estimator of population mean using two auxiliary variables in stratified random sampling. We have made comparison of our proposed estimator with Kadilar and Cingi (2003), Sisodia and Dwivedi (1981) and Singh and Kakran (1993) by extending their estimators for multi-auxiliary variables.

Main aim of this paper is to propose an exponential type estimator for estimating the population mean of the study variable in stratified random sampling using multiple auxiliary variables. Section 2, is devoted for the notation and preliminaries used in the paper, followed by Section 3 where we discussed the existing estimators from different authors in similar situations. In Section 4, the proposed estimator is derived along with its properties and optimized value. In Section 5, comparison of the proposed estimator with existing estimators is being made and conditions under which the proposed estimators performs better than the other estimators in terms of efficiency is obtained. In order to justify the use of these variables computational results are obtained by simulation through R software. Section 6 is devoted to the discussion based on the simulation study. Paper is concluded in Section 7.

2. Notations and Preliminaries

Consider a finite population $U=(U_1,U_2,....U_N)$ of size 'N' which are partitioned into 'l' homogenous subgroups, called strata, such that h^{th} strata consist of N_h units, where $h=1,2,...,l$ and $\sum_{h=1}^{l} N_h = N.$ From each stratum, we select a sample of size $n_h,$ by simple random sampling without replacement such that $\sum_{h=1}^{l} n_h = n,$ here n_h is the h^{th} stratum sample size. Let 'Y' be the study variable and $x_i,(i=1,2,...,k)$ are the auxiliary variables taking values y_{hj} and $x_{hij},(h=1,2,...,l;$ $j=1,2,...,N_h).$ Then, we define

$\bar{x}_{hi} = \frac{1}{n_h}\sum_{j=1}^{n_h} x_{hij};$ sample mean of the observation taken from h^{th} stratum for variable x_i,

$\bar{y}_h = \frac{1}{n_h}\sum_{j=1}^{n_h} y_{hj};$ sample mean of the observation taken from h^{th} stratum for variable y,

$\bar{X}_{hi} = \frac{1}{N_h}\sum_{j=1}^{N_h} x_{hij};$ h^{th} stratum mean for auxiliary variable x_i,

$\bar{Y}_h = \frac{1}{N_h}\sum_{j=1}^{N_h} y_{hj};$ h^{th} stratum mean for study variable y,

$\bar{X}_i = \sum_{h=1}^{l} w_h \bar{X}_{hi};$ population mean of the auxiliary variable x_i,

$\bar{Y} = \sum_{h=1}^{l} w_h \bar{Y}_h;$ population mean of the study variable y,

$\bar{x}_{sti} = \sum_{h=1}^{l} w_h \bar{x}_{hi};$ unbiased estimator of the population mean $\bar{X}_i$,

$\bar{y}_{st} = \sum_{h=1}^{l} w_h \bar{y}_h$; unbiased estimator of the population mean $\bar{Y}$,

$S_{yh}^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (y_{hj} - \bar{Y}_h)^2$; population mean square of h^{th} stratum for the study variable y,

$S_{xhi}^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (x_{hij} - \bar{X}_{hi})^2$; population mean square of h^{th} stratum for the auxiliary variable x_i,

$w_h = \frac{N_h}{N}$; stratum weight of the h^{th} stratum.

To obtain the bias and MSE, we define

$$\bar{y}_{st} = \sum_{h=1}^{l} w_h \bar{y}_h = \bar{Y}(1+e_0), \quad \bar{x}_{sti} = \sum_{h=1}^{l} w_h \bar{x}_{hi} = \bar{X}(1+e_i),$$

such that, $E(e_0) = E(e_i) = 0$, $E(e_0^2) = \sum_{h=1}^{l} w_h^2 \delta_h \frac{S_{yh}^2}{\bar{Y}^2}$, $E(e_i^2) = \sum_{h=1}^{l} w_h^2 \delta_h \frac{S_{xhi}^2}{\bar{X}_i^2}$, $E(e_0 e_i) = \sum_{h=1}^{l} w_h^2 \delta_h \frac{S_{yxhi}^2}{\bar{Y}\bar{X}_i}$,

$\delta_h = \frac{1}{n_h} - \frac{1}{N_h}$.

3. Estimators in Literature

The estimators available in the literature are discussed in this section along with proposed one by extending them for multi-auxiliary variables.

1) Kadilar-Cingi estimator

Kadilar and Cingi (2003) computed sample mean of the variables in stratified random sampling method as

$$\bar{y}_{st} = \sum_{h=1}^{l} w_h \bar{y}_h \quad ; \bar{x}_{st} = \sum_{h=1}^{l} w_h \bar{x}_h$$

and proposed a combined ratio estimator as

$$\bar{y}_{KC} = \frac{\bar{y}_{st}}{\bar{x}_{sti}} \bar{X}. \tag{1}$$

Proposed estimator under multiple auxiliary variables set up is

$$t_{KC} = \bar{y}_{st} \sum_{i=1}^{k} \lambda_i \frac{\bar{X}_i}{\bar{x}_{sti}}, \tag{2}$$

where λ_i 's $(i = 1, 2, ..., k)$ are weights assigned to auxiliary variables such that $\sum_{i=1}^{k} \lambda_i = 1$.

The bias and MSE of the above estimator is written as

$$B(t_{KC}) = \sum_{h=1}^{l} w_h^2 \delta_h \sum_{i=1}^{k} \lambda_i \frac{1}{\bar{X}_i} \left(R_i S_{xhi}^2 - S_{yxhi} \right), \tag{3}$$

$$MSE\left(t_{KC}\right) = \sum_{h=1}^{l} w_h^2 \delta_h \left(S_{yh}^2 + \sum_{i=1}^{k} \lambda_i^2 R_i^2 S_{xhi}^2 - 2 \sum_{i=1}^{k} \lambda_i R_i S_{yxhi} \right), \tag{4}$$

where $R_i = \frac{\bar{Y}}{\bar{X}_i}$.

2) Sisodia-Dwivedi estimator

Sisosdia and Dwivedi (1981) suggested a modified ratio estimator, considering that the coefficient of variation C_x is known

$$\bar{y}_{SD} = \bar{y}\left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right). \tag{5}$$

In stratified random sampling, Kadilar and Cingi (2003) proposed this estimator as

$$\bar{y}_{stSD} = \bar{y}_{st}\left\{\frac{\sum_{h=1}^{l} w_h\left(\bar{X}+C_x\right)}{\sum_{h=1}^{l} w_h\left(\bar{x}+C_x\right)}\right\}. \tag{6}$$

In stratified sampling when we consider the case of multi-auxiliary variable, we suggest the estimator as

$$t_{SD} = \bar{y}_{st}\sum_{i=1}^{k}\lambda_i\left\{\frac{\sum_{h=1}^{l} w_h(\bar{X}+C_x)}{\sum_{h=1}^{l} w_h(\bar{x}+C_x)}\right\}. \tag{7}$$

where λ_i's $(i=1,2,...,k)$ are weights assigned to auxiliary variables such that $\sum_{i=1}^{k}\lambda_i = 1.$

The bias and MSE of the above estimator can be written as

$$B(t_{SD}) = \sum_{h=1}^{l} w_h^2\delta_h\sum_{i=1}^{k}\lambda_i\frac{1}{\bar{X}_{SDi}}\left(R_{SDi}S_{xhi}^2 - S_{yxhi}\right), \tag{8}$$

$$MSE\left(t_{SD}\right) = \sum_{h=1}^{l} w_h^2\delta_h\left(S_{yh}^2 + \sum_{i=1}^{k}\lambda_i^2 R_{SDi}^2 S_{xhi}^2 - 2\sum_{i=1}^{k}\lambda_i R_{SDi}S_{yxhi}\right), \tag{9}$$

where $R_{SDi} = \dfrac{\bar{Y}}{\bar{X}_{SDi}}$ and $\bar{X}_{SDi} = (\bar{X}_i + C_{xi}).$

3) Singh-Kakran estimator

Motivated by Sisodia and Dwivedi (1981), Singh and Kakran (1993) proposed ratio-type estimator, considering that the coefficient of kurtosis of auxiliary variable $\beta_2(x)$ is known as

$$\bar{y}_{SK} = \bar{y}\left(\frac{\bar{X}+\beta_2(x)}{\bar{x}+\beta_2(x)}\right). \tag{10}$$

In stratified random sampling, Kadilar and Cingi (2003) proposed this estimator as

$$\bar{y}_{stSK} = \bar{y}_{st}\left\{\frac{\sum_{h=1}^{l} w_h\left(\bar{X}_h+\beta_{2h}(x)\right)}{\sum_{h=1}^{l} w_h\left(\bar{x}_h+\beta_{2h}(x)\right)}\right\}. \tag{11}$$

In stratified sampling when we consider the case of multi-auxiliary variable, we suggest this estimator as

$$t_{SK} = \bar{y}_{st} \sum_{i=1}^{k} \lambda_i \left\{ \frac{\sum_{h=1}^{l} w_h \left(\bar{X}_{hi} + \beta_{2hi}(x) \right)}{\sum_{h=1}^{l} w_h \left(\bar{x}_{hi} + \beta_{2hi}(x) \right)} \right\}, \tag{12}$$

where λ_i 's $(i = 1, 2, ..., k)$ are weights assigned to auxiliary variables such that $\sum_{i=1}^{k} \lambda_i = 1.$

The MSE and bias of the above estimator can be written as

$$B(t_{SK}) = \sum_{h=1}^{l} w_h^2 \delta_h \sum_{i=1}^{k} \lambda_i \frac{1}{\bar{X}_{SKi}} (R_{SKi} S_{xhi}^2 - S_{yxhi}), \tag{13}$$

$$MSE(t_{SK}) = \sum_{h=1}^{l} w_h^2 \delta_h \left(S_{yh}^2 + \sum_{i=1}^{k} \lambda_i^2 R_{SKi}^2 S_{xhi}^2 - 2 \sum_{i=1}^{k} \lambda_i R_{SKi} S_{yxhi} \right), \tag{14}$$

where $R_{SKi} = \frac{\bar{Y}}{\bar{X}_{SKi}}$ and $\bar{X}_{SKi} = \left(\bar{X}_i + \beta_{2i}(x) \right).$

4. The Proposed Estimator

In survey sampling, auxiliary variable increases the precision of the estimator when variable of interest is highly correlated with it. Bahl and Tuteja (1991) estimated mean of ratio and product type estimators using information on single auxiliary variable. Bahl and Kumar (2000) extended the work and proposed ratio and product type estimators under Simple random sampling for estimating the mean of the finite population using multi-auxiliary variables as

$$\hat{\bar{Y}}_{\text{Re}M} = \bar{y} \sum_{t=1}^{k} w_t \exp\left(\frac{\bar{X}_t - \bar{x}_t}{\bar{X}_t + \bar{x}_t} \right), \tag{15}$$

$$\hat{\bar{Y}}_{P\text{e}M} = \bar{y} \sum_{t=1}^{k} w_t \exp\left(\frac{\bar{x}_t - \bar{X}_t}{\bar{x}_t + \bar{X}_t} \right), \tag{16}$$

and $w_1, w_2, ..., w_t;\ t = 1, 2, ..., k$ are weights assigned to auxiliary variables such that $\sum_{t=1}^{k} w_t = 1.$

By using multiple auxiliary variables we estimate the population mean $\bar{Y}$ as

$$t_{Rst} = \bar{y}_{st} \sum_{t=1}^{k} \lambda_i \exp\left(\frac{\bar{X}_i - \bar{x}_{sti}}{\bar{X}_i + \bar{x}_{sti}} \right), \tag{17}$$

where $\lambda_1, \lambda_2, ..., \lambda_k$ are weights assigned to auxiliary variables such that $\sum_{i=1}^{k} \lambda_i = 1.$

Expanding the right hand side of (17) in terms of e 's and neglecting the terms having power greater than two, we get

$$(t_{Rst} - \bar{Y}) = \bar{Y} \sum_{i=1}^{k} \lambda_i \left(e_0 - \frac{e_i}{2} - \frac{e_0 e_i}{2} + \frac{3e_1^2}{8} \right). \tag{18}$$

Taking expectation on both sides of (18), one can obtain the bias of t_{Rst} to the first degree of approximation as

$$B(t_{Rst}) = \bar{Y} \sum_{h=1}^{l} w_h^2 \delta_h \sum_{i=1}^{k} \lambda_i \frac{1}{\bar{X}_i} \left(\frac{3}{8} R_i S_{xhi}^2 - S_{yxhi} \right). \tag{19}$$

Squaring both sides of (18) and neglecting terms of e 's having power greater than two, we have

$$(t_{Rst}-\bar{Y})^2=\bar{Y}^2\left\{\sum_{i=1}^{k}\lambda_i\left(e_0-\frac{e_i}{2}-\frac{e_0e_i}{2}+\frac{3e_1^2}{8}\right)\right\}^2. \tag{20}$$

Taking expectation to both sides of (20), we get the MSE of t_{Rst} as

$$MSE(t_{Rst})=\sum_{h=1}^{l}w_h^2\delta_h\left(S_{yh}^2+\frac{1}{4}\sum_{i=1}^{k}\lambda_i^2R_i^2S_{xhi}^2-\sum_{i=1}^{k}\lambda_iR_iS_{yxhi}\right), \tag{21}$$

where $R_i=\dfrac{\bar{Y}}{\bar{X}_i}$.

Differentiating (21) with respect to λ_i and equating to zero, hence we have the optimum value of λ_i as

$$\hat{\lambda}_i=2\frac{\beta_{yxhi}}{R_i}. \tag{22}$$

Substituting the optimum value of $\hat{\lambda}_i$, in (22), we get the optimum value of mean squared error of t_{Rst} , which is given by

$$MSE(t_{Rst})=\sum_{h=1}^{l}w_h^2\delta_h\left(S_{yh}^2+\frac{1}{4}\sum_{i=1}^{k}\hat{\lambda}_i^2R_i^2S_{xhi}^2-\sum_{i=1}^{k}\hat{\lambda}_iR_iS_{yxhi}\right). \tag{23}$$

5. Efficiency Comparisons

We compare our estimator with Kadilar and Cingi (2003), Sisodia and Dwivedi (1981) and Singh and Kakran (1993), and conclude that

(i) $MSE(t_{KC})-MSE(t_{Rst})>0$, if $3\sum_{i=1}^{k}\lambda_iR_iA>4B,$ (24)

where $A=\sum_{h=1}^{l}w_h^2S_{xhi}^2$ and $B=\sum_{h=1}^{l}w_h^2S_{yxhi}$.

(ii) $MSE(t_{SD})-MSE(t_{Rst})>0$, if $\theta_1A>\theta_2B,$ (25)

where $\theta_1=\sum_{i=1}^{k}\lambda_i^2R_{SDi}^2-\frac{1}{4}\sum_{i=1}^{k}\lambda_i^2R_i^2$ and $\theta_2=2\sum_{i=1}^{k}\lambda_iR_{SDi}-\sum_{i=1}^{k}\lambda_iR_i$.

(iii) $MSE(t_{SK})-MSE(t_{Rst})>0$, if $\theta_3A>\theta_4B,$ (26)

where $\theta_3=\sum_{i=1}^{k}\lambda_i^2R_{SKi}^2-\frac{1}{4}\sum_{i=1}^{k}\lambda_i^2R_i^2$ and $\theta_2=2\sum_{i=1}^{k}\lambda_iR_{SKi}-\sum_{i=1}^{k}\lambda_iR_i$.

When the conditions given in (24), (25) and (26) are satisfied, then our proposed estimator will be more efficient than the other considered estimators. Further, in the next section, the theoretical results are supported with simulation study.

6. Simulation Study

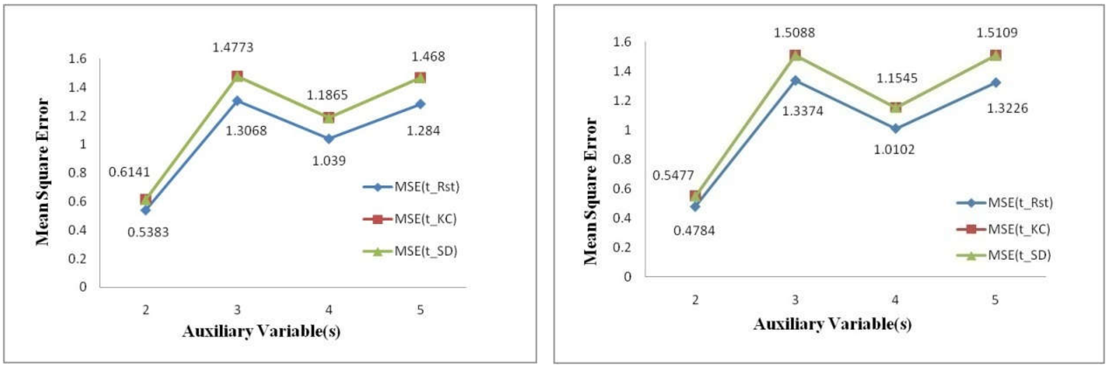
For validating the theoretical results obtained in the above section, we use the R software, version 3.5.0 (R Core Team 2018). The pseudo-populations of different sizes are generated by using the R software. For our study, we have generated sixteen populations for different number of auxiliary variables. For two auxiliary variables we will have four populations. Two populations with four strata;

one is with equal stratum size; other is with unequal stratum size and two populations with six strata; one is with equal stratum size and other one is with unequal stratum size. Similarly, we studied these populations for three, four and five auxiliary variables. All the populations are simulated from normal distribution. The obtained results are shown in the Table 1. MSE's of the proposed estimator and existing estimators for 4 equal and unequal size stratum and 6 equal and unequal sizes stratum are plotted against auxiliary variable to have clear vision about the applicability of the proposed estimator.

Table 1 MSE's of the estimators for different auxiliary variables and strata

No. of auxiliary variables	No. of strata	Strata type	$MSE(t_{Rst})$	$MSE(t_{KC})$	$MSE(t_{SD})$
2	4	Equal	0.5383	0.6141	0.6141
		Unequal	0.4784	0.5477	0.5477
	6	Equal	0.1381	0.1578	0.1578
		Unequal	0.1397	0.1594	0.1594
3	4	Equal	1.3068	1.4773	1.4773
		Unequal	1.3374	1.5088	1.5088
	6	Equal	0.2293	0.2619	0.2619
		Unequal	0.2440	0.2786	0.2786
4	4	Equal	1.0390	1.1865	1.1865
		Unequal	1.0102	1.1545	1.1545
	6	Equal	0.3236	0.3705	0.3705
		Unequal	0.3146	0.3603	0.3603
5	4	Equal	1.2840	1.4680	1.4680
		Unequal	1.3226	1.5109	1.5109
	6	Equal	0.4286	0.4901	0.4901
		Unequal	0.4165	0.4770	0.4770

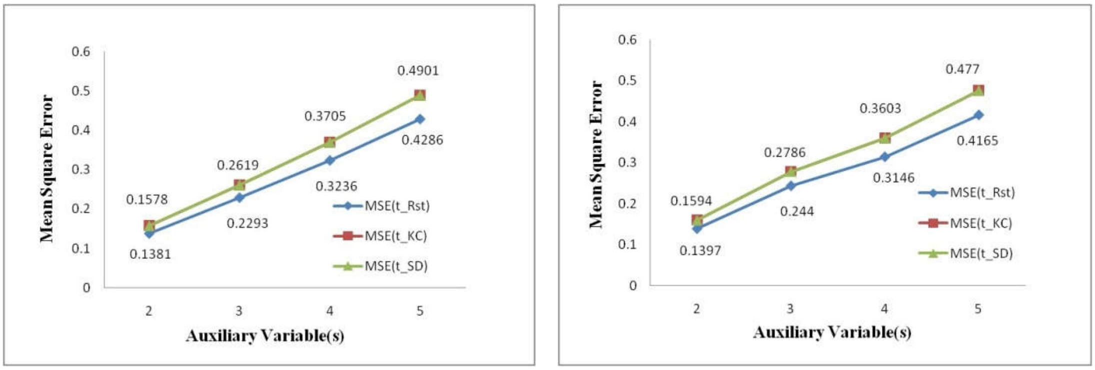
From Table 1, it can easily be seen that the generated proposed estimator outperformed all the other existing estimators.



(a) (b)

Figure 1 MSE of estimators for different auxiliary variables (a) for 4 equal size strata and (b) for 4 unequal size strata

Figure 1 depicts the comparison of MSE of proposed estimator with the Kadilar and Cingi (2003) and Sisodia and Dwivedi (1981) when population is divided into four strata. Figure 2 gives the same comparison when population is divided into six strata, respectively. The case of equal and unequal stratum size is also being taken into consideration.



(a) (b)

Figure 2 MSE of estimators for different auxiliary variables (a) for 6 equal size strata and (b) for 6 unequal size strata

From Table 1 as well as different figures we envisaged that the proposed estimator has the minimum MSE among the other considered estimators. There is not much difference in the value of MSE for equal and unequal sizes of strata. Also we can see that when we increase the number of strata, estimators are showing much better results in terms of MSE where as in case of equal and unequal strata MSE's are showing the same trend for different auxiliary variables. MSE of the proposed estimator is minimum in case of two auxiliary variables. Thus, increasing the auxiliary variables within the stratum is not a good choice, as MSE of proposed estimator is showing increasing trend in this case.

In Table 2, we have the bias of the proposed estimator for different auxiliary variable, for different strata as well as for equal and unequal strata type. We can see that for equal and unequal strata type the value of bias doesn't have much difference but if we increased stratum number then bias of the proposed estimator reduces.

7. Conclusions

In this paper we have studied the properties of ratio type estimator in stratified random sampling using multi-auxiliary variables. We derived the MSE of our proposed estimator and also modified some previous estimators and compared them with our proposed estimator theoretically and empirically. The bias for proposed estimator has been derived and from empirical study we can conclude that by increasing stratum number the bias of the proposed estimator reduces. Through simulation study, it has been made clear that the proposed estimator performs better than the other considered estimators. Thus, the proposed estimator is recommended for use in practice if the conditions derived are satisfied.

Acknowledgments

The authors would like to thanks the referees for their through review, which improved the quality of the paper.

Table 2 Bias of the proposed estimator for different auxiliary variables and strata.

No. of auxiliary variables	No. of strata	Strata type	Bias(t_{Rst})
2	4	Equal	0.12349
		Unequal	0.12196
	6	Equal	0.03225
		Unequal	0.03120
3	4	Equal	0.34257
		Unequal	0.37382
	6	Equal	0.07026
		Unequal	0.07173
4	4	Equal	0.30516
		Unequal	0.30474
	6	Equal	0.09771
		Unequal	0.09445
5	4	Equal	0.12139
		Unequal	0.12137
	6	Equal	0.03644
		Unequal	0.03588

References

Bahl S, Tuteja RK. Ratio and product type exponential estimator. J Inform Optim Sci. 1991; 12(1): 159-164.

Bahl S, Kumar M. Estimation of finite population mean using multi-auxiliary variables. J Stat Manag Sys. 2000; 3(1): 67-74.

D'Orazio M, Catanese E. A simple approach of stratification of units in multipurpose in business and agriculture surveys. Istat Working Papers. 2016; Report No. 10.

Haq A, Shabbir J. Improved family of ratio estimators in simple and stratified random sampling, Commu Stat Theory Methods. 2013; 42(5): 782-799.

Kadilar C, Cingi H. Ratio estimator in stratified random sampling. Biomet Jour. 2003; 45(2): 218-225.

Koyuncu N, Kadilar C. Family of estimators of population mean using two auxiliary variable in stratified random sampling. Commu Stat Theory Methods. 2009; 139(8): 2552-2558.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: Austria; 2018 [cited 2018 Feb 15]. Available from: https://www.R-project.org.

Shabbir J, Gupta S. A new estimator of population mean in stratified sampling. Commu Stat Theory Methods. 2006; 35: 1201-1209.

Singh, HP and Kakran, MS. A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. Revised version submitted to J Ind Soc Agri Stat, New Delhi, India (1993).

Singh R, Kumar M. Improved estimators of population mean using two auxiliary variables in stratified random sampling, Pak J Stat Oper Res. 2012; 8(1): 65-72.

Sisodia BVS, Dwivedi VK. A modified ratio estimator using coefficient of variation of auxiliary variable. J Ind Soc Agri Stat. 1981; 33:13-19.

Tailor R, Chouhan S, Tailor R, Garg N. A ratio-cum product estimator of population mean in stratified random sampling using two auxiliary variables. Statistica. 2012; 72(3): 287-297.