Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 73(3) 2019 211–216

New Ratio Estimators for Population Mean using Robust Regression

Mir Sabzar, S. Maqbool, T.A. Raja and M.A. Bhat

Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, Srinagar, J&K, India

Received 14 July 2017; Revised 07 June 2019; Accepted 08 August 2019

SUMMARY

The present paper deals with proposing the linear regression ratio type estimators for estimating the population mean in simple random sampling without replacement when there are outliers in the data and these estimators are not affected by these outliers as we are using here robust regression, which is obtained by Huber M-estimation. The expressions for mean square error are obtained, compared with the estimators in literature and analysed that our estimators are efficient class of estimators. We also provide numerical example and simulation study using gamma distribution to support the theoretical results.

Keywords: Population median, Correlation coefficient, Coefficient of variation, Robust regression, Mean square error, Efficiency.

1. INTRODUCTION

Linear regression analysis is one of the important statistical tool that is used for many fields and in linear regression analysis the parameters of interest in the model are almost estimated by the method of least square method of estimation, but there is a problem whenever there are outliers in the population and in that case least square method of estimation is not as accurate as it is very sensitive to outliers, as outliers have more serious effect on statistical inference. Keeping this serious problem of outliers in mind another technique was developed known as Robust regression instead of Least square method of estimation which gives information about the observation whether it is valid to be retained or thrown out, as the basic aim of the robust regression analysis is that the model should be fitted in such a way that it gives us the majority of information about the data. Many properties of robust regression technique are efficiency, breakdown point and bounded influence which are used to define the reliability of the technique in a theoretical sense. As we know that estimation part is one of the important objective of the sampling theory and it is enhanced by using the supplementary information and one of the property (efficiency) of the robust regression tells us how well this technique performs better than the least square method after cleaning the data i.e., without outliers and the another property of robust regression technique (breakdown point) gives us information regarding the stability of the estimators when there are large number of outliers present in the data (Hampel, [6]) and this measure is known as global robustness in this sense. Ratio estimators is more suitable type of estimator for estimating the finite population mean when the information of auxiliary variable is known and that is positively correlated with the study variable. But there is a problem for estimating the population mean by the classical ratio type estimators as when there are outliers present in the data as these estimators are very much sensitive to the outliers in the data (Chatterjee [4]). To overcome this problem of classical ratio type estimator which is too much sensitive to outliers, we in this article use Huber M- estimates, in lieu of least square method of estimation. For more detailed discussion one may go through the following Hampel [7], Kadilar [8], [9], [10], [11], [12] and [13], Singh [15], Sisodia and Dwivedi [16] and Upadhyaya and Singh [17].

Corresponding author: Mir Sabzar
E-mail *address*: subzarstat@gmail.com

In this paper we present the estimators in literature for estimating the finite population mean in simple random sampling and their mean square equations in the section 2. We propose an improved ratio type estimators using robust regression with their mean square equations in section 3. The efficiency comparisons between the estimators in the literature and the improved estimators, based on the mean square error equations are given in section 4. The findings of the study are given in section 5, simulation study in section 6 and finally conclusion is given in last section.

2. ESTIMATORS IN LITERATURE

Kadilar and Cingi [10] proposed the following estimators by using the auxiliary information of coefficient of variation and coefficient of kurtosis for estimating the finite population mean in simple random sampling and the estimators are given below:

$$\hat{\bar{Y}}_{N(KC1)} = \frac{\bar{y}_n + b(\bar{X}_N - \bar{x}_n)}{\bar{x}_n}\bar{X}_N,$$

$$\hat{\bar{Y}}_{N(KC2)} = \frac{\bar{y}_n + b(\bar{X}_N - \bar{x}_n)}{\{\bar{x}_n + C_{x_n}\}}\{\bar{X}_n + C_{x_n}\},$$

$$\hat{\bar{Y}}_{N(KC3)} = \frac{\bar{y}_n + b(\bar{X}_N - \bar{x}_n)}{\{\bar{x}_n + \beta_2(x_n)\}}\{\bar{X}_N + \beta_2(x_n)\},$$

$$\hat{\bar{Y}}_{N(KC4)} = \frac{\bar{y}_n + b(\bar{X}_N - \bar{x}_n)}{\{\bar{x}_n\beta_2(x_n) + C_{x_n}\}}\{\bar{X}_N\beta_2(x_n) + C_{x_n}\},$$

$$\hat{\bar{Y}}_{N(KC5)} = \frac{\bar{y}_n + b(\bar{X}_N - \bar{x}_n)}{\{\bar{x}_n C_{x_n} + \beta_2(x_n)\}}\{\bar{X}_N C_{x_n} + \beta_2(x_n)\}, \quad (2.1)$$

Where C_{x_n} and $\beta_2(x_n)$ are the population coefficient of variation and the population coefficient of the kurtosis, respectively, of the auxiliary variable; $\bar{y}_n$ and $\bar{x}_n$ are the sample means of the study variable and auxiliary variable, respectively and it is assumed that the population mean $\bar{X}_N$ of the auxiliary variable x_n is known. Here $b = \frac{s_{x_n y_n}}{s^2_{x_n}}$ is obtained by the Least Square method, where $s^2_{x_n}$ and $s^2_{y_n}$ are the sample variances of the auxiliary and the study variable, respectively and $s_{x_n y_n}$ is the sample covariance between the auxiliary and the study variable.

The mean square error equations of the estimators $\bar{Y}_{N(kci)}$, $i = 1,2,3,4,5$ can be found using a Taylor series expansion upto first degree approximation, and is as follows:

$$\hat{\bar{Y}}_{N(KC1)} = \frac{\bar{y}_n + b(\bar{X}_N - \bar{x}_n)}{\bar{x}_n}\bar{X}_N = \hat{R}_1\bar{X}_N$$

For this estimator the mean square error equation is obtained as

$$h(\bar{x}_n, \bar{y}_n) \cong h(\bar{X}_N, \bar{Y}_N) + \frac{\partial h(c,d)}{\partial c}\Big|_{\bar{X}_N,\bar{Y}_N}(\bar{x}_n - \bar{X}_N) + \frac{\partial h(c,d)}{\partial d}\Big|_{\bar{X}_N,\bar{Y}_N}(\bar{y}_n - \bar{Y}_N) \quad (2.2)$$

Where $h(\bar{x}_n, \bar{y}_n) = \hat{R}_{KC1}$ and $h(\bar{X}_N, \bar{Y}_N) = R.$

As shown in Wolter [18], (2.2) can be applied to the proposed estimator in order to obtain MSE equation as follows:

$$\hat{R}_{KC1} - R \cong \frac{\partial((\bar{y}_n + b(\bar{X}_N - \bar{x}_n))/(\bar{x}_n)}{\partial \bar{x}_n}\Big|_{\bar{X}_N,\bar{Y}_N}(\bar{x}_n - \bar{X}_N) + \frac{\partial((\bar{y}_n + b(\bar{X}_N - \bar{x}_n))/(\bar{x}_n)}{\partial \bar{y}_n}\Big|_{\bar{X}_N,\bar{Y}_N}(\bar{y}_n - \bar{Y}_N)$$

$$\cong -\left(\frac{\bar{y}_n}{\bar{x}_n^2} + \frac{b\bar{X}_N}{\bar{x}_n^2}\right)\Big|_{\bar{X}_N,\bar{Y}_N}(\bar{x}_n - \bar{X}_N) + \frac{1}{\bar{x}_n}\Big|_{\bar{X}_N,\bar{Y}_N}(\bar{y}_n - \bar{Y}_N)$$

$$E(\hat{R}_{KC1} - R)^2 \cong \frac{(\bar{Y}_N + B\bar{X}_N)^2}{\bar{X}_N^4}V(\bar{x}_n) - \frac{2(\bar{Y}_N + B\bar{X}_N)}{\bar{X}_N^3}Cov(\bar{x}_n, \bar{y}_n) + \frac{1}{\bar{X}_N^2}V(\bar{y}_n)$$

$$\cong \frac{1}{\bar{X}_N^2}\left\{\frac{(\bar{Y}_N + B\bar{X}_N)^2}{\bar{X}_N^2}V(\bar{x}_n) - \frac{2(\bar{Y}_N + B\bar{X}_N)}{\bar{X}_N}Cov(\bar{x}_n, \bar{y}_n) + V(\bar{y}_n)\right\}$$

Where $B = \frac{s_{x_n y_n}}{s^2_{x_n}} = \frac{\rho s_{x_n} s_{y_n}}{s^2_{x_n}} = \frac{\rho s_{y_n}}{s_{x_n}}.$

Note that we omit the difference of $(E(b) - B)$.

$$MSE(\bar{Y}_{KC1}) = \bar{X}_N^2\, E(\hat{R}_{KC1} - R)^2 \cong \frac{(\bar{Y}_N + B\bar{X}_N)^2}{\bar{X}_N^2}V(\bar{x}_n) - \frac{2(\bar{Y}_N + B\bar{X}_N)}{\bar{X}_N}Cov(\bar{x}_n, \bar{y}_n) + V(\bar{y}_n)$$

$$\cong \frac{\bar{Y}_N^2 + 2B\bar{X}_N\bar{Y}_N + B^2\bar{X}_N^2}{\bar{X}_N^2}V(\bar{x}_n) - \frac{2\bar{Y}_N + 2B\bar{X}_N}{\bar{X}_N}Cov(\bar{x}_n, \bar{y}_n) + V(\bar{y}_n)$$

$$\cong \frac{(1-f)}{n}\left\{\left(\frac{\bar{Y}_N^2}{\bar{X}_N^2} + \frac{2B\bar{Y}_N}{\bar{X}_N} + B^2\right)S^2_{x_n} - \left(\frac{2\bar{Y}_N}{\bar{X}_N} + 2B\right)S_{x_n y_n} + S^2_{y_n}\right\}$$

$$\cong \frac{(1-f)}{n}(R^2_{KC1}S^2_x + 2BR_{KC1}S^2_{x_n} + B^2S^2_{x_n} - 2R_{KC1}S_{x_n y_n} - 2BS_{x_n y_n} + S^2_{y_n})$$

Similarly for the other estimators given by Kadilar and Cingi mentioned above, mean square error equations can be obtained as

$$MSE(\overline{Y}_{N(KCi)}) \cong \frac{1-f}{n}(R_{KCi}^2 S_{x_n}^2 + 2BR_{KCi} S_{x_n}^2 + BS_{x_n}^2 - 2R_{KCi} S_{x_n y_n} - 2BS_{x_n y_n} + S_{y_n}^2)$$

$$i = 2,3,4,5 \qquad (2.3)$$

Where $B = \frac{s_{x_n y_n}}{s_{x_n}^2}$ is obtained by the Least Square method; $f = \frac{n}{N}$; n is the sample size; N is the population size;

$$R_{kc1} = \frac{\overline{Y}_N}{\overline{X}_N}, \ R_{kc2} = \frac{\overline{Y}_N}{(\overline{X}_N + C_{x_n})}, \ R_{kc3} = \frac{\overline{Y}_N}{(\overline{X}_N + \beta_2(x_n))},$$

$$R_{kc4} = \frac{\overline{Y}_N \beta_2(x_n)}{(\overline{X}_N \beta_2(x_n) + C_{x_n})} \quad \text{and} \quad R_{kc5} = \frac{\overline{Y}_N C_{x_n}}{(\overline{X}_N C_{x_n} + \beta_2(x_n))}$$

are the population ratios; $S_{x_n}^2$ and $S_{y_n}^2$ are the population variance of the auxiliary variable and study variable. We take $E(b) = B$ in (2.3), where E represents the expected value.

3. IMPROVED RATIO ESTIMATORS USING ROBUST REGRESSION

In this section we propose two improved ratio type estimators for estimating the finite population mean using the robust regression, to the data which have outliers, and the improved ratio estimators are given as below;

$$\hat{\overline{Y}}_{N(SSTM1)} = \frac{\overline{y}_n + b_{rob}(\overline{X}_N - \overline{x}_n)}{(\overline{x}_n M_d + \rho)}(\overline{X}_N M_d + \rho),$$

$$\hat{\overline{Y}}_{N(SSTM2)} = \frac{\overline{y}_n + b_{rob}(\overline{X}_N - \overline{x}_n)}{(\overline{x}_n M_d + C_{x_n})}(\overline{X}_N M_d + C_{x_n}), \qquad (3.1)$$

The Mean Square expression for the above estimators is obtained as

$$h(\overline{x}_n, \overline{y}_n) \cong h(\overline{X}_N, \overline{Y}_N) + \frac{\partial h(c,d)}{\partial c}\Big|_{\overline{X}_N, \overline{Y}_N} (\overline{x}_n - \overline{X}_N) + \frac{\partial h(c,d)}{\partial d}\Big|_{\overline{X}_N, \overline{Y}_N} (\overline{y}_n - \overline{Y}_N) \qquad (3.2)$$

Where $h(\overline{x}_n, \overline{y}_n) = \hat{R}_{KC1}$ and $h(\overline{X}_N, \overline{Y}_N) = R.$

As shown in Wolter [18], (3.2) can be applied to the proposed estimator in order to obtain MSE equation as follows:

$$\hat{R}_{SSTM1} - R \cong \frac{\partial((\overline{y}_n + b_{rob}(\overline{X}_N - \overline{x}_n))/(\overline{x}_n Md + \rho)}{\partial(\overline{x}_n)}\Big|_{\overline{X}_N, \overline{Y}_N} (\overline{x}_n - \overline{X}_N) + \frac{\partial((\overline{y}_n + b_{rob}(\overline{X}_N - \overline{x}_n))/(\overline{x}_n Md + \rho)}{\partial(\overline{y}_n)}\Big|_{\overline{X}_N, \overline{Y}_N} (\overline{y}_n - \overline{Y}_N)$$

$$\cong -\left(\frac{\overline{y}_n Md}{(\overline{x}_n Md + \rho)^2} + \frac{b_{rob}(\overline{X}_N Md + \rho)}{(\overline{x}_n Md + \rho)^2}\right)\Big|_{\overline{X}_N, \overline{Y}_N} (\overline{x}_n - \overline{X}_N) + \frac{1}{(\overline{x}_n Md + \rho)}\Big|_{\overline{X}_N, \overline{Y}_N} (\overline{y}_n - \overline{Y}_N)$$

$$E(\hat{R}_{SSTNM1} - R)^2 \cong \frac{(\overline{Y}_N Md + B_{rob}(\overline{X}_N Md + \rho))^2}{(\overline{X}_N Md + \rho)^4} V(\overline{x}_n) - \frac{2(\overline{Y}_N Md + B_{rob}(\overline{X}_N Md + \rho))}{(\overline{X}_N Md + \rho)^3} Cov(\overline{x}_n, \overline{y}_n) + \frac{1}{(\overline{X}_N Md + \rho)^2} V(\overline{y}_n)$$

$$\cong \frac{1}{(\overline{X}_N Md + \rho)^2}\left\{\frac{(\overline{Y}_N Md + B_{rob}(\overline{X}_N Md + \rho))^2}{(\overline{X}_N Md + \rho)^2} V(\overline{x}_n) - \frac{2(\overline{Y}_N Md + B_{rob}(\overline{X}_N Md + \rho))}{(\overline{X}_N Md + \rho)} Cov(\overline{x}_n, \overline{y}_n) + V(\overline{y}_n)\right\}$$

Where b_{rob} is obtained by Huber M-estimates in robust regression.

As we know that Huber M-estimates here used over least square method as they are not affected by the extreme values in the data. So using here M-estimation is that it gives more accurate results than the least square method of estimation in the presence of extreme values in the data. Huber M-estimates use a function $\rho(u)$ that is a compromise between u^2 and $|u|$, where u is the error term of the regression model $y = a + bx + u$,a being the constant of the model. The Huber $\rho(u)$ function has the form:

$$\rho(u) = \begin{cases} u^2 & -t \le e \le t \\ 2t|u| - t^2 & u < -t \ or \ t < u \end{cases}$$

Where t is a tuning constant that controls the robustness of the estimators. Huber suggested $t = 1.5\hat{\sigma}_u$, where $\hat{\sigma}_u$ is an estimate of the standard deviation σ_u of the population random errors. Details about constant t and M-estimators can be found in Candan [2], Rousseeuw and Leroy [14].

The value of the regression coefficient, b_{rob} is obtained by minimizing

$$\sum_{i=1}^{n}\rho(y_{n(i)}-a-bx_{n(i)})$$

With respect to a and b. The details for the minimization procedure can be found in Birkes and Dodge [1].

$$MSE(\bar{Y}_{N(SSTM1)})=(\bar{X}_N Md+\rho)^2\, E(\hat{R}_{SSTM1}-R)^2$$

$$\cong\frac{(\bar{Y}_N Md+B_{rob}(\bar{X}_N Md+\rho))^2}{(\bar{X}_N Md+\rho)^2}V(\bar{x}_n)-\frac{2(\bar{Y}_N Md+B_{rob}(\bar{X}_N Md+\rho))}{(\bar{X}_N Md+\rho)}Cov(\bar{x}_n,\bar{y}_n)+V(\bar{y}_n)$$

$$\cong\frac{\bar{Y}_N^2 Md+2B_{rob}(\bar{X}_N Md+\rho)\bar{Y}_N+B_{rob}^2(\bar{X}_N Md+\rho)^2}{(\bar{X}_N Md+\rho)^2}V(\bar{x}_n)-\frac{2\bar{Y}_N Md+2B_{rob}(\bar{X}_N Md+\rho)}{(\bar{X}_N Md+\rho)}$$

$$Cov(\bar{x}_n,\bar{y}_n)+V(\bar{y}_n)$$

$$\cong\frac{(1-f)}{n}\left\{\left(\frac{\bar{Y}_N^2 Md}{(\bar{X}_N Md+\rho)^2}+\frac{2B_{rob}\bar{Y}_N Md}{(\bar{X}_N Md+\rho)}+B_{rob}^2\right)S_{x_n}^2-\left(\frac{2\bar{Y}_N Md}{(\bar{X}_N Md+\rho)}+2B_{rob}\right)S_{x_n y_n}+S_{y_n}^2\right\}$$

$$\cong\frac{(1-f)}{n}(R_{SSTM1}^2 S_x^2+2B_{rob}R_{SSTM1}S_{x_n}^2+B_{rob}^2 S_{x_n}^2-2R_{SSTM1}S_{x_n y_n}-2B_{rob}S_{x_n y_n}+S_{y_n}^2) \qquad (3.3)$$

$$MSE(\bar{Y}_{N(SSMT2)})\cong\frac{1-f}{n}(R_{SSTM2}^2 S_{x_n}^2+2B_{rob}R_{SSTM2}S_{x_n}^2+B_{rob}S_{x_n}^2-2R_{SSTM2}S_{x_n y_n}-2B_{rob}S_{x_n y_n}+S_{y_n}^2) \qquad (3.4)$$

We remark that the Mean Square Error equation of the proposed ratio estimators $\hat{\bar{Y}}_{N(SSTM1)}$, and $\hat{\bar{Y}}_{N(SSTM2)}$ is in the same form as the Mean Square Error equation in (2.2), but it is clear that B in (2.2) should be replaced by B_{rob}, whose value as obtained by Huber M-estimation.

It is well known that since $E[\phi(u)]=0$, where $\phi(u)=\rho'(u)$ and u has an identically independent distribution, we can easily assume that $E(b_{rob})=B_{rob}$ in (3.2), as for in (2.3). We would like to remark that the value of B_{rob} is computed as b_{rob}, but the population data is used for B_{rob}.

4. EFFICIENCY COMPARISONS

In this section we compare the MSE of the proposed estimators, given in (3.2), with the MSE of the ratio estimators, given in (2.2).

$$MSE(\bar{Y}_{N(SSMTi)})<MSE(\bar{Y}_{N(kci)}),\ SSMT\ i=1,2.$$

$$KC\ i=1,2,...,5,$$

$$(2B_{rob}R_{SSTMi}S_{x_n}^2+B_{rob}S_{x_n}^2-2B_{rob}S_{x_n y_n})<(2BR_{kci}S_{x_n}^2+BS_{x_n}^2-2BS_{x_n y_n}),$$

$$2R_{SSTMikci},\ S_{x_n}^2(B_{rob}-B)-2S_{x_n y_n}(B_{rob}-B)+S_{x_n}^2(B_{rob}^2-B^2)<0,$$

$$(B_{rob}-B)[2R_{SSTMikci},\ S_{x_n}^2-2S_{x_n y_n}+S_{x_n}^2(B_{rob}+B)<0,$$

For $B_{rob}-B>0$, that is $B_{rob}>B$:

$$2R_{SSTMikci},\ S_{x_n}^2-2S_{x_n y_n}S_{x_n}^2(B_{rob}+B)<0,$$

$$(B_{rob}+B)<-2R_{SSTMikci}+2\frac{S_{x_n y_n}}{S_{x_n}^2},$$

$$B_{rob}<B-2R_{SSTMikci}.$$

Similarly, for $B_{rob}-B<0$, that is $B_{rob}<B$:

$$B_{rob}>B-2R_{SSTMikci}.$$

Consequently, we have the following conditions:

$$0<B_{rob}-B<2R_{SSTMikci} \qquad (4.1)$$

$$\text{or } -2R_{SSTMikci}<B_{rob}-B<0. \qquad (4.2)$$

When condition (4.1) or (4.2) is satisfied, the proposed estimators given in (3.1), are more efficient than the ratio estimator, given in (2.1), respectively.

5. NUMERICAL ILLUSTRATION

For Numerical illustration we have used the data concern the level of apple production (in tons) as the study variable and number of apple trees (1 unit =100 trees) as the auxiliary variable in 106 villages of the Aegean Region in Turkey in 1999 (Source: Institute of Statistics, Republic of Turkey). The statistics of the population are given in Table 1.

Hence by using the simple random sampling we take the sample of size $n=30$. We would like to point out that the efficiency comparisons of the estimators does not depend on the sample size n, as it is not involved in the efficiency conditions (4.1) or (4.2), as shown in Section 4. Note that the correlation, ρ, between the auxiliary and study variable is 0.86 for this data set.

Table 1. Statistical analysis of the population

Parameters	Values	Parameters	Values	Parameters	Values	Parameters	Values
N	106	$\bar{Y}_N$	2212.59	C_{x_n}	2.10	B_{rob}	5.02
n	30	$\bar{X}_N$	274.22	M_d	7297.500	B	17.21
ρ	0.86	C_{y_n}	5.22	$\beta_2(x_n)$	34.57	R_{SSTM1}	8.06866
S_{y_n}	11551.53	S_{x_n}	574.61	$S_{x_n y_n}$	5681761.76	R_{SST2}	8.06865

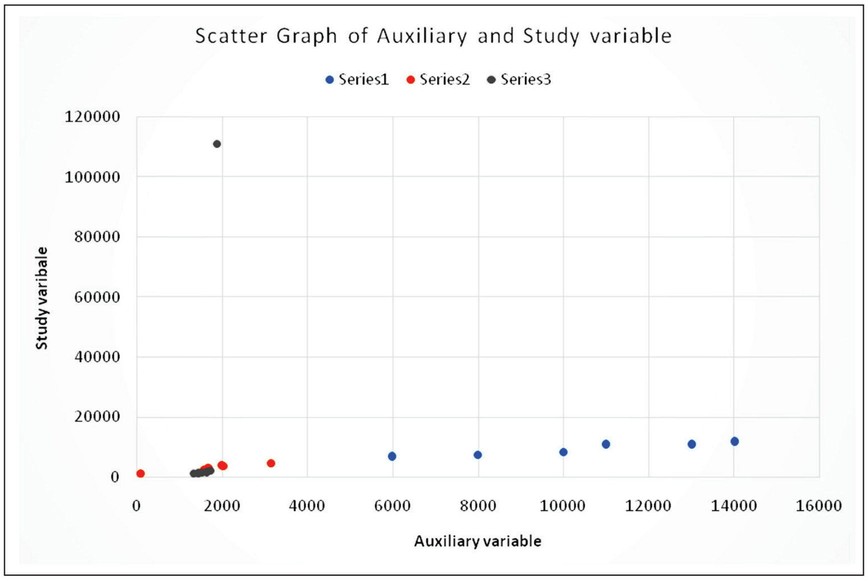

Table 2. Mean square error (MSE) of proposed estimators and existing estimators in literature

MSE	$\hat{\bar{Y}}_{KC1}$	$\hat{\bar{Y}}_{KC2}$	$\hat{\bar{Y}}_{KC3}$	$\hat{\bar{Y}}_{KC4}$	$\hat{\bar{Y}}_{KC5}$	$\hat{\bar{Y}}_{SSTM1}$	$\hat{\bar{Y}}_{SSTM2}$
	1370287.16	1362379.47	1261539.55	1369932.64	1313604.06	986606.75	987085.088

Table 3. Relative efficiencies of each proposed estimators with respect to the estimators in literature

Relative Efficiencies	$\hat{\bar{Y}}_{KC1}$	$\hat{\bar{Y}}_{KC2}$	$\hat{\bar{Y}}_{KC3}$	$\hat{\bar{Y}}_{KC4}$	$\hat{\bar{Y}}_{KC5}$
$\hat{\bar{Y}}_{SSTM1}$	0.7200	0.7241	0.7820	0.7201	0.7510
$\hat{\bar{Y}}_{SSTM2}$	0.7204	0.7245	0.7824	0.7205	0.7514

We obtain the Mean Square Error values of the estimators in literature and proposed estimators as defined in Section 2 and Section 3, respectively given in Table 2 and using these values we compute the relative efficiency for each proposed estimator in (3.1) with respect to the estimators in literature in (2.1) given in Table 3, using the formulae given as under;

Relative efficiency

$$(\hat{\bar{Y}}_{SSTMi}) = \frac{MSE(\hat{\bar{Y}}_{SSTMi})}{MSE(\hat{\bar{Y}}_{KCi})};$$

$$SSTMi = 1, 2 \quad \text{and} \quad KCi = 1, 2, ..., 5. \tag{5.1}$$

Thus from the above table we asses that the proposed estimators are efficient than the estimators in literature as it is clear from the condition (5.1) that whose value comes out to be less than 1, then the mean square error of the proposed estimators is less than the estimators in literature. As we see from the table that our proposed estimators are efficient than the estimators in literature by using the robust regression and this result is to be expected as the condition (4.2) is satisfied for all the proposed estimators given as

$$-2R_{SSTM1} \cong -2R_{SSTM2} \cong -16,$$

$$B_{rob} - B \cong -12$$

Thus the condition: $-2R_{SSTMi} < B_{rob} - B < 0$ is satisfied.

6. SIMULATION STUDY

For performing the simulation study, we perform the following steps and were coded in R- program and summarize the simulation procedures used to find the MSE of an estimator, say for $\hat{\bar{Y}}_{N(KCi)}$ $i = 1, 2, .., 5.$ and for $\hat{\bar{Y}}_{SSTMi}$ $i = 1, 2.$

Step 1: We select 2500 samples of size n (where $n = 20$, $n = 30$, $n = 40$, $n = 50$) of shape = 1.25 and scale = 1.25 from the generated Gamma distribution using SRSWOR.

Step 2: We use the data from 2500 samples in Step 1 to obtain the value of $\hat{\bar{Y}}$. Thus, we find 2500 samples values of $\hat{\bar{Y}}$ samples for each n.

Step 3: For each n, the MSE of $\hat{\hat{\bar{Y}}}$ is computed by

$$MSE\left(\hat{\hat{\bar{Y}}}\right)=\frac{1}{2500}\sum_{i=1}^{2500}\left(\hat{\hat{\bar{Y}}}-\bar{Y}\right)^{2},$$

Where $\bar{Y}$ is the population mean of the study variable.

In this simulation study, we take sample size n = 20, 30, 40, 50 the values of the MSE ratios of the proposed estimators with respect to estimators in literature for each n are given in table 4. These values are computed using (5.1). From Table 4, we can conclude that all proposed estimators are more efficient than the estimators on literature for all samples sizes. These simulation results support the theoretical findings in table 2 and table 3. We would also like to point out that the values of relative efficiencies of the proposed estimators with respect to the estimators in literature in Table 4 would decrease dramatically, in other words, the efficiencies of the proposed estimators would increase significantly, if there were more outliers in data.

Table 4. Simulation results for the relative efficiencies of each proposed estimator with respect to the traditional estimators for different sample sizes

Sample Size	$RE\left(\bar{y}_{pri}\right)$	$\hat{\hat{\bar{Y}}}_{KC1}$	$\hat{\hat{\bar{Y}}}_{KC2}$	$\hat{\hat{\bar{Y}}}_{KC3}$	$\hat{\hat{\bar{Y}}}_{KC4}$	$\hat{\hat{\bar{Y}}}_{KC5}$
$n=20$	$\bar{y}_{pr1}$	0.6365	0.6416	0.6878	0.6043	0.6554
	$\bar{y}_{pr2}$	0.6578	0.6712	0.6789	0.6598	0.6876
$n=30$	$\bar{y}_{pr1}$	0.6202	0.6245	0.6623	0.6012	0.6432
	$\bar{y}_{pr2}$	0.6431	0.6630	0.6521	0.6545	0.6670
$n=40$	$\bar{y}_{pr1}$	0.6713	0.6940	0.6848	0.7047	0.6893
	$-_{pr}$	0.6717	0.6733	0.6757	0.6789	0.6913
$n=50$	$\bar{y}_{pr1}$	0.6925	0.6528	0.6817	0.6712	0.6691
	$\bar{y}_{pr2}$	0.6711	0.6851	0.6981	0.6981	0.6813

7. CONCLUSION

Thus from the above study both theoretical and simulation study, we conclude that our proposed estimators perform better than the estimators in the literature even when there are extreme values in the data as we have used the robust regression in our proposed estimators as this robust regression have increased the efficiency of our linear regression ratio type estimators obtained by Huber M-estimation. Therefore our proposed estimators preferred over the estimators in literature whenever there are extreme values in the data and should be used in future for practical applications.

ACKNOWLEDGEMENT

The Authors are highly thankful to worthy referees and Chief Editor for appreciated recommendations on original form of the document.

REFERENCES

Birkes, D. and Dodge, Y., (1993). *Alternative Methods of Regression*. John Wiley & Sons.

Candan, M., (1995). *Robust Estimators in Linear Regression Analysis*. Hacettepe University, Department of Statistics, Master Thesis (in Turkish)

Chambers R.L., (1986). Outlier robust finite population estimation, *J. Amer. Statist. Assoc.,* **81,** 1063-1069.

Chatterjee, S. and Price, B., (1991). *Regression Analysis by Example*. Wiley, Second Edition.

Cochran, W.G., (1977). *Sampling Techniques*. John Wiley and Sons, New-York.

Hampel, F.R., (1975). Beyond location parameters: robust concepts and methods. *Proceedings of the 40th Session of the ISI*, **46,** 375-391.

Hampel, F.R., Ronchetti, E. M., Rousseeww, P.J. and Stahel, W.A., (1981). *Robust Statistics* (John Huber, P. Robust Statistics, Wiley, New York).

Kadilar, C. and Cingi, H., (2003). Estimator of a population mean using two auxiliary variables in simple random sampling. *International Mathematics Journal,* **5,** 357-360,.

Kadilar, C. and Cingi, H., (2003). Ratio estimators in stratified random sampling, *Biometrical Journal,* **45,** 218-225.

Kadilar, C. and Cingi, H., (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation,* **151,** 893-902.

Kadilar, C. and Cingi, H., (2005). A new estimator using two auxiliary variable. *Applied Mathematics and Computation,* **162,** 901-908.

Kadilar, C. and Cingi, H., (2005). A new ratio estimator in stratified random sampling. *Communication in Statistics: Theory and Methods,* **34,** 597-602.

Kadilar, C., Candan, M. and Cingi, H., (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics*, **36(2),** 181-188.

Rousseeuw, P. and Leroy, A., (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Singh, H.P. and Kakran, M.S., (1993). A modified ratio estimator using known coefficient of kurtosis of an auxiliary character, unpublished.

Sisodia, B.V.S. and Dwivedi, V.K., (1981). A modified ratio estimator using coefficient of variation of auxiliary variable, *J. Ind. Soc. Agril. Statist.*, **33,** 13-18,

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

Wolter K.M., (1985). *Introduction to Variance Estimation*, Springer-Verlag.