ROBUST RATIO TYPE ESTIMATORS IN SIMPLE RANDOM SAMPLING USING HUBER M ESTIMATION

M. Subzar*[1] C., N. Bouza**, S. Maqbool*, T A Raja** and B. A. Para***
*Division of Agricultural Statistics, SKUAST-Kashmir (190025), India.
**Facultad de Matemática y Computación, Universidad de La Habana, Cuba
***Post Graduate Department of Statistics, University of Kashmir, Srinagar.

ABSTRACT
In survey sampling a very dominating problem is to obtain the better ratio estimators of the population mean and population variance. A very striking question, in every researcher's mind, is that if there are outliers in the data how to estimate the population mean and variance, as outliers mislead the results. So keeping the above problem under consideration, we propose some new modified ratio estimators using robust regression, which are robust against the outliers and give accurate results even in the presence of the outliers; also the properties are studied. It has been shown that the proposed class of estimators is more efficient than the existing classes of estimators. An empirical study has been carried out to examine the merits of the proposed class of estimators over others.

KEYWORDS: Quartiles, Non-Conventional location parameters, Median, M-Estimation, Bias, Mean Square Error, Efficiency.

MSC: 62D05, 62G35

RESUMEN
En encuestas por muestreo un problema muy dominante es el de obtener el mejor estimador de razón de la media y la varianza. Una pregunta muy inquietante, en la mente de todo investigador, es cuando hay observaciones aberrantes en la data como estimar los parámetros sin conllevar una mala interpretación de los resultados. Con esto en mente, proponemos una nueva clase de estimadores de razón modificados, usando una regresión robusta, cuyos resultados sean adecuados incluso ante la existencia de outliers; también se estudian sus propiedades. Se ha probado que la clase de estimadores propuesta es más eficiente que la de otros estimadores existentes. Un estudio empírico ha sido llevado a cabo para examinar los méritos de esta clase respecto a otras.

PALABRAS CLAVE: Cuartiles, Parámetros Non-Convencionales de Posición, Mediana, M-Estimación, Sesgo, Error Cuadrático Medio, Eficiencia

1. INTRODUCTION

Since from 1960, various theoretical efforts have been devoted to develop statistical procedures that are resistant to small deviations from the assumptions, i.e. robust with respect to outliers and stable with respect to small deviations from the assumed parametric model. In fact, it is well-known that classical optimum procedures behave quite poorly under slight violations of the strict model assumptions.

It is also well-known that to screen the data, to remove outliers and then to apply classical inferential procedures is not a simple and good way to proceed. First of all, in multivariate or highly structured data, it can be difficult to single out outliers or it can be even impossible to identify influential observations. Second, in place of rejecting an observation, it could be better to down-weight uncertain observation, although we may wish to reject completely wrong observations. Moreover, rejecting outliers reduces the sample size, could affect the distribution theory, and variance could be underestimated from the cleaned data. So even one outlying observation can destroy least squares estimation, resulting in parameter estimates that do not provide useful information for the majority of the data. So Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is and whether this should be thrown out. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. The properties of efficiency, breakdown point, and bounded influence are used to define the measure of robust technique performance in a theoretical sense. Efficiency can tell us how well a robust technique performs relative to least squares on clean data (without outliers). High efficiency is mostly desired on estimation. The breakdown point is a

[1] subzarstat@gmail.com

measure for stability of the estimator when the sample contains a large fraction of outliers (Hampel, 1975). It gives the minimum fraction of outliers which may produce an infinite bias. It is referred as the measure of global robustness in this sense. For example, least square has a breakdown point of $1/n$. This indicates that only one outlier can make the estimates useless. In contrast, some robust regression estimates attaches approximately 50% breakdown point, and it is called a high breakdown point in this case. Lastly, bounded influence is designed to counter the tendency of least squares to allow exterior X-space or high leverage points to exhibit greater influence, which can be especially important if these points are outliers. Robust regression estimators were first introduced by Huber (1973, 1981), and it is well known as M-regression estimator. Rousseeuw (1984) introduced the least median of squares (LMS) and the least trimmed squares (LTS) estimators. These estimators minimize the median and the trimmed mean of the squared residuals respectively. They are very high breakdown point estimator. The high breakdown point estimation has some drawbacks. First of all, computing any of these estimators exactly is impractical in all but small datasets, because they involve the combinatorial problem of determining how many cases are used. Therefore, they are based on resampling techniques and their solutions are determined randomly (Rousseeuw and Leroy, 1987), and then they can be even inconsistent. Second problem is their lower convergence rate. For example, LMS has the low convergence rate. It makes direct effect of efficiency of estimates. Third, they do not have bounded influence for X-space or high leverage points although they limitedly affected from response outlying observations. In the present study we make the use of M-estimation, introduced by Huber (1964).

2. EXISTING ESTIMATORS IN THE LITERATURE

Kadilar and Cingi (2004) suggested ratio type estimators for the population mean in the simple random sampling using some known auxiliary information on coefficient of kurtosis and coefficient of variation. They showed that their suggested estimators are more efficient than traditional ratio estimator in the estimation of the population mean. Kadilar & Cingi (2004) estimators are given by

$$\hat{\bar{Y}}_1 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}}\bar{X}, \qquad \hat{\bar{Y}}_2 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + C_x)}(\bar{X} + C_x), \qquad \hat{\bar{Y}}_3 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \beta_2)}(\bar{X} + \beta_2),$$

$$\hat{\bar{Y}}_4 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + C_x)}(\bar{X}\beta_2 + C_x), \qquad \hat{\bar{Y}}_5 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \beta_2)}(\bar{X}C_x + \beta_2).$$

Kadilar and Cingi (2006) developed some modified ratio estimators using known value of coefficient of correlation, kurtosis and coefficient of variation as follows:

$$\hat{\bar{Y}}_6 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \rho)}(\bar{X} + \rho), \qquad \hat{\bar{Y}}_7 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \rho)}(\bar{X}C_x + \rho),$$

$$\hat{\bar{Y}}_8 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\rho + C_x)}(\bar{X}\rho + C_x), \quad \hat{\bar{Y}}_9 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + \rho)}(\bar{X}\beta_2 + \rho), \hat{\bar{Y}}_{10} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\rho + \beta_2)}(\bar{X}\rho + \beta_2).$$

Where C_x and $\beta_2(x)$ are the population coefficient of variation and the population coefficient of the kurtosis, respectively, of the auxiliary variable; $\bar{y}$ and $\bar{x}$ are the sample means of the study variable and auxiliary variable, respectively and it is assumed that the population mean $\bar{X}$ of the auxiliary variable x is known. Here $b = \frac{s_{xy}}{s_x^2}$ is obtained by the LS method, where s_x^2 and s_y^2 are the sample variances of the auxiliary and the study variable, respectively and s_{xy} is the sample covariance between the auxiliary and the study variable.

MSE of the first estimator can be found using Taylor series method defined as

$$h(\bar{x}, \bar{y}) \cong h(\bar{X}, \bar{Y}) + \frac{\partial h(c,d)}{\partial c}\Big|_{\bar{X},\bar{Y}} (\bar{x} - \bar{X}) + \frac{\partial h(c,d)}{\partial d}\Big|_{\bar{X},\bar{Y}} (\bar{y} - \bar{Y}) \tag{2.1}$$

Where $h(\bar{x}, \bar{y}) = \hat{R}_{p1}$ and $h(\bar{X}, \bar{Y}) = R.$

As shown in Wolter (1985), (2.1) can be applied to the proposed estimator in order to obtain MSE equation as follows:

$$\hat{R}_{Kc1} - R \cong \frac{\partial((\bar{y} + b(\bar{X} - \bar{x}))/(\bar{x})}{\partial \bar{x}} |_{\bar{X},\bar{Y}} \ (\bar{x} - \bar{X}) + \frac{\partial((\bar{y} + b(\bar{X} - \bar{x}))/(\bar{x})}{\partial \bar{y}} |_{\bar{X},\bar{Y}} \ (\bar{y} - \bar{Y})$$

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

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

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

Where $B = \frac{s_{xy}}{s_x^2} = \frac{\rho s_x s_y}{s_x^2} = \frac{\rho s_y}{s_x}$. Note that we omit the difference of $(E(b) - B)$.

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

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

$$\cong \frac{(1-f)}{n} \left\{ \left(\frac{\bar{Y}^2}{(\bar{X})^2} + \frac{2B\bar{Y}}{(\bar{X})} + B^2 \right) S_x^2 - \left(\frac{2\bar{Y}}{(\bar{X})} + 2B \right) S_{xy} + S_y^2 \right\}$$

$$MSE(\bar{y}_{KC1}) \cong \frac{(1-f)}{n} (R_{KC1}^2 S_x^2 + 2BR_{KC1} S_x^2 + B^2 S_x^2 - 2R_{KC1} S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC1} = \frac{\bar{Y}}{\bar{X}}$$

Similarly the mean square error of the other estimators are given as

$$MSE(\hat{\bar{Y}}_{KC2}) \cong \frac{(1-f)}{n} (R_{KC2}^2 S_x^2 + 2BR_{KC2} S_x^2 + B^2 S_x^2 - 2R_{KC2} S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC2} = \frac{\bar{Y}}{\bar{X} + C_x}$$

$$MSE(\hat{\bar{Y}}_{KC3}) \cong \frac{(1-f)}{n} (R_{KC3}^2 S_x^2 + 2BR_{KC3} S_x^2 + B^2 S_x^2 - 2R_{KC3} S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC3} = \frac{\bar{Y}}{\bar{X} + \beta_2(x)}$$

$$MSE(\hat{\bar{Y}}_{KC4}) \cong \frac{(1-f)}{n} (R_{KC4}^2 S_x^2 + 2BR_{KC4} S_x^2 + B^2 S_x^2 - 2R_{KC4} S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC4} = \frac{\bar{Y}\beta_2(x)}{\bar{X}\beta_2(x) + C_x}$$

$$MSE(\hat{\bar{Y}}_{KC5}) \cong \frac{(1-f)}{n}(R_{KC5}^2 S_x^2 + 2BR_{KC5}S_x^2 + B^2 S_x^2 - 2R_{KC5}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC5} = \frac{\bar{Y}C_x}{\bar{X}C_x + \beta_2(x)}$$

$$MSE(\hat{\bar{Y}}_{KC6}) \cong \frac{(1-f)}{n}(R_{KC6}^2 S_x^2 + 2BR_{KC6}S_x^2 + B^2 S_x^2 - 2R_{KC6}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC6} = \frac{\bar{Y}}{\bar{X} + \rho}$$

$$MSE(\hat{\bar{Y}}_{KC7}) \cong \frac{(1-f)}{n}(R_{KC7}^2 S_x^2 + 2BR_{KC7}S_x^2 + B^2 S_x^2 - 2R_{KC7}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC7} = \frac{\bar{Y}C_x}{\bar{X}C_x + \rho}$$

$$MSE(\hat{\bar{Y}}_{KC8}) \cong \frac{(1-f)}{n}(R_{KC8}^2 S_x^2 + 2BR_{KC8}S_x^2 + B^2 S_x^2 - 2R_{KC8}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC8} = \frac{\bar{Y}\rho}{\bar{X}\rho + C_x}$$

$$MSE(\hat{\bar{Y}}_{KC9}) \cong \frac{(1-f)}{n}(R_{KC9}^2 S_x^2 + 2BR_{KC9}S_x^2 + B^2 S_x^2 - 2R_{KC9}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC9} = \frac{\bar{Y}\beta_2(x)}{\bar{X}\beta_2(x) + \rho}$$

$$MSE(\hat{\bar{Y}}_{KC10}) \cong \frac{(1-f)}{n}(R_{KC10}^2 S_x^2 + 2BR_{KC10}S_x^2 + B^2 S_x^2 - 2R_{KC10}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{KC10} = \frac{\bar{Y}\rho}{\bar{X}\rho + \beta_2(x)}$$

Yan and Tian (2010) proposed the following two modified ratio estimators using coefficient of skewness and kurtosis;

$$\hat{\bar{Y}}_{YT11} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \beta_1)}(\bar{X} + \beta_1), \qquad \hat{\bar{Y}}_{YT12} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_1 + \beta_2)}(\bar{X}\beta_1 + \beta_2)$$

The mean square error of the above estimators are given as

$$MSE(\hat{\bar{Y}}_{YT11}) \cong \frac{(1-f)}{n}(R_{YT11}^2 S_x^2 + 2BR_{YT11}S_x^2 + B^2 S_x^2 - 2R_{YT11}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{YT11} = \frac{\bar{Y}}{\bar{X} + \beta_1(x)}$$

The mean square error of the equation is given as

$$MSE(\hat{\bar{Y}}_{YT12}) \cong \frac{(1-f)}{n}(R_{YT12}^2 S_x^2 + 2BR_{YT12}S_x^2 + B^2 S_x^2 - 2R_{YT12}S_{xy} - 2BS_{xy} + S_y^2),$$

$$R_{YT12} = \frac{\bar{Y}\beta_1(x)}{\bar{X}\beta_1(x) + \beta_2(x)}$$

S_x^2 and S_y^2 are the population variance of the auxiliary variable and study variable. It is worthwhile pointing out that we take $E(b) = B$ in the above equations, where E represents the expected value.

3. MODIFIED RATIO ESTIMATORS USING ROBUST REGRESSION

As one of the serious drawback of the LS method is that it is sensitive to outliers and decreases the efficiency in estimating the population parameters. Keeping this in view we propose new modified ratio type estimators using robust regression in order to maintain the efficiency of parameters while estimating the population parameters as these proposed estimators are highly robust against the outliers.

$$\hat{\bar{Y}}_{p1} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}Q_1 + M_d)}(\bar{X}Q_1 + M_d),$$

The mean square error of the above estimator is given as

$$MSE(\hat{\bar{Y}}_{p1}) \cong \frac{(1-f)}{n}(R_{p1}^2 S_x^2 + 2B_{rob}R_{p1}S_x^2 + B_{rob}^2 S_x^2 - 2R_{p1}S_{xy} - 2B_{rob}S_{xy} + S_y^2),$$

$$R_{p1} = \frac{\bar{Y}Q_1}{\bar{X}Q_1 + M_d} \qquad \hat{\bar{Y}}_{p2} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}Q_2 + M_d)}(\bar{X}Q_2 + M_d),$$

The mean square error of the above estimator is given as

$$MSE(\hat{\bar{Y}}_{p2}) \cong \frac{(1-f)}{n}(R_{p2}^2 S_x^2 + 2B_{rob}R_{p2}S_x^2 + B_{rob}^2 S_x^2 - 2R_{p2}S_{xy} - 2B_{rob}S_{xy} + S_y^2),$$

$$R_{p2} = \frac{\bar{Y}Q_2}{\bar{X}Q_2 + M_d} \qquad \hat{\bar{Y}}_{p3} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}Q_3 + M_d)}(\bar{X}Q_3 + M_d),$$

The mean square error of the above estimator is given as

$$MSE(\hat{\bar{Y}}_{p3}) \cong \frac{(1-f)}{n}(R_{p3}^2 S_x^2 + 2B_{rob}R_{p3}S_x^2 + B_{rob}^2 S_x^2 - 2R_{p3}S_{xy} - 2B_{rob}S_{xy} + S_y^2),$$

$$R_{p3} = \frac{\bar{Y}Q_3}{\bar{X}Q_3 + M_d} \qquad \hat{\bar{Y}}_{p4} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}TM + M_d)}(\bar{X}TM + M_d),$$

The mean square error of the above estimator is given as

$$MSE(\hat{\bar{Y}}_{p4}) \cong \frac{(1-f)}{n}(R_{p4}^2 S_x^2 + 2B_{rob}R_{p4}S_x^2 + B_{rob}^2 S_x^2 - 2R_{p4}S_{xy} - 2B_{rob}S_{xy} + S_y^2),$$

$$R_{p4} = \frac{\bar{Y}TM}{\bar{X}TM + M_d} \qquad \hat{\bar{Y}}_{p5} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}MR + M_d)}(\bar{X}MR + M_d),$$

The mean square error of the above estimator is given as

$$MSE(\hat{\bar{Y}}_{p5}) \cong \frac{(1-f)}{n}(R_{p5}^2 S_x^2 + 2B_{rob}R_{p5}S_x^2 + B_{rob}^2 S_x^2 - 2R_{p5}S_{xy} - 2B_{rob}S_{xy} + S_y^2),$$

$$R_{p5} = \frac{\bar{Y}MR}{\bar{X}MR + M_d} \qquad \hat{\bar{Y}}_{p6} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}TM + M_d)}(\bar{X}TM + M_d),$$

The mean square error of the above estimator is given as

$$MSE(\hat{\bar{Y}}_{p6}) \cong \frac{(1-f)}{n}(R_{p6}^2 S_x^2 + 2B_{rob}R_{p6}S_x^2 + B_{rob}^2 S_x^2 - 2R_{p6}S_{xy} - 2B_{rob}S_{xy} + S_y^2),$$

$$R_{p6} = \frac{\bar{Y}HL}{\bar{X}HL + M_d}$$

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

The main advantage of Huber M-estimates over LS estimates is that they are not sensitive to outliers. Thus, when there are outliers in the data, M-estimation is more accurate than LS estimation. Huber M-estimates use a function $\rho(e)$ that is a compromise between e^2 and $|e|$, where e is the error term of the regression model $y = a + bx + e$, a being the constant of the model. The Huber $\rho(e)$ function has the form:

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

Where k is a tuning constant that controls the robustness of the estimators. Huber (1964) suggested $k = 1.5\hat{\sigma}$, where $\hat{\sigma}$ is an estimate of the standard deviation, σ of the population random errors. Details about constant k and M-estimators can be found in Candan (1995), Rousseeuw and Leroy (1987).

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

$$\sum_{i=1}^{n} \rho(y_i - a - bx_i)$$

With respect to a and b . The details for the minimization procedure can be found in Birkes and Dodge (1993).

We remark that the MSE equation of the proposed ratio estimators $\hat{\bar{Y}}_{pi} i = 1,2,\ldots,6$ is in the same form as the MSE equation existing estimators $\bar{Y}_i i = 1,2,\ldots,12$, but it is clear that B in MSE equations of existing estimators should be replaced by B_{rob}, whose value as obtained by Huber M-estimation

It is well known that since $E[\psi(e)] = 0$, where $\psi(e) = \rho'(e)$ and e has an identically independent distribution, we can easily assume that $E(b_{rob}) = B_{rob}$ as for b . 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 COMPARISON

In this section we have derived theoretically the efficiency comparison of the proposed estimators with the existing estimators by Kadilar and Cingi (2004), Kadilar and Cingi (2006) and Yan and Tian (2010)
We compare the MSE of the proposed estimators, with the MSE of the Existing ratio estimators.

$$MSE(\bar{Y}_{pi}) < MSE(\bar{Y}_{KCi}), pi = 1,2,\ldots,6. \qquad i = 1,2,\ldots,12$$

$$(2B_{rob}R_{pi}S_x^2 + B_{rob}S_x^2 - 2B_{rob}S_{xy}) < (2BR_iS_x^2 + BS_x^2 - 2BS_{xy}),$$

$$2R_{pi_{(i)}}, S_x^2(B_{rob} - B) - 2S_{xy}(B_{rob} - B) + S_x^2(B_{rob}^2 - B^2) < 0,$$

$$(B_{rob} - B)[2R_{pi_{(i)}}, S_x^2 - 2S_{xy} + S_x^2(B_{rob} + B) < 0,$$

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

$$2R_{pi_{(i)}}, S_x^2 - 2S_{xy} + S_x^2(B_{rob} + B) < 0,$$

$$(B_{rob} + B) < -2R_{pi_{(i)}} + 2\frac{S_{xy}}{S_x^2},$$

$$B_{rob} < B - 2R_{pi_{(i)}}.$$

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

$$B_{rob} > B - 2R_{pi_{(i)}}.$$

Consequently, we have the following conditions:

$$0 < B_{rob} - B < 2R_{pi_{(i)}} \quad (4.1)$$

or

$$-2R_{pi_{(i)}} < B_{rob} - B < 0. \quad (4.2)$$

When condition (4.1) or (4.2) is satisfied, the proposed estimators given in Section 3 are more efficient than the ratio estimator, given in section.2, respectively.

5. NUMERICAL ILLUSTRATION

We have taken the data from the book Theory and Analysis of Sample Survey Designs by Singh, D and Chaudhary, F. S. (1986) page 177, in which the data under wheat in 1971 and 1973 is given and in which area under wheat in the region was to be estimated during 1974 is denoted by Y (study variable) by using the data of cultivated area under wheat in 1971 is denoted by X (auxiliary variable)

Table 1. Characteristics of these populations.

Parameter	Population	Parameter	Population	Parameter	Population
N	34	S_x	150.5059	TM	162.25
n	20	C_x	0.7205	MR	284.5
$\bar{Y}$	856.4117	β_2	0.0978	HL	190
$\bar{X}$	208.8823	β_1	0.9782	Q_1	94.25
ρ	0.4491	QD	80.25	Q_2	150
S_y	733.1407	$Brob$	1.57	Q_3	254.75
C_y	0.8561	B	2.19	M_d	150

Table 2: The Statistical Analysis of the Estimators for these Populations

Estimators	**Constant**	**MSE**	**Estimators**	**Constant**	**MSE**
$\hat{\bar{Y}}_1$	4.100	16673.45	$\hat{\bar{Y}}_{10}$	4.096	16657.19
$\hat{\bar{Y}}_2$	4.086	16619.64	$\hat{\bar{Y}}_{11}$	4.081	16600.54
$\hat{\bar{Y}}_3$	4.098	16666.14	$\hat{\bar{Y}}_{12}$	4.098	16665.98
$\hat{\bar{Y}}_4$	3.960	16146.61	$\hat{\bar{Y}}_{p1}$	4.069	14373.69
$\hat{\bar{Y}}_5$	4.097	16663.31	$\hat{\bar{Y}}_{p2}$	4.080	14410.62
$\hat{\bar{Y}}_6$	4.091	16639.85	$\hat{\bar{Y}}_{p3}$	4.088	14436.48
$\hat{\bar{Y}}_7$	4.088	16626.87	$\hat{\bar{Y}}_{p4}$	4.082	14415.36
$\hat{\bar{Y}}_8$	4.069	16554.4	$\hat{\bar{Y}}_{p5}$	4.090	14445.47
$\hat{\bar{Y}}_9$	4.011	16338.65	$\hat{\bar{Y}}_{p6}$	4.085	14445.48

5.1. Percent Relative Efficiency of Modified Proposed estimators with Existing estimators

We obtain the MSE values of the traditional and proposed estimators as defined in Section (2) and Section (3), respectively, and using these values compute the relative efficiency for each proposed estimator in Section (3) with respect to the traditional estimators in Section (2) using the formulae:

$$RE(\hat{\bar{Y}}_{pi}) = \frac{MSE(\hat{\bar{Y}}_{pi})}{MSE(\hat{\bar{Y}}_{i})}; pi = 1,2,\ldots,6, \quad i = 1,2,\ldots,12. \tag{5.1}$$

Table 3. Theoretical results for the relative efficiencies of each proposed estimators with respect to the Traditional Estimators

	$\hat{\bar{Y}}_1$	$\hat{\bar{Y}}_2$	$\hat{\bar{Y}}_3$	$\hat{\bar{Y}}_4$	$\hat{\bar{Y}}_5$	$\hat{\bar{Y}}_6$	$\hat{\bar{Y}}_7$	$\hat{\bar{Y}}_8$	$\hat{\bar{Y}}_9$	$\hat{\bar{Y}}_{10}$	$\hat{\bar{Y}}_{11}$	$\hat{\bar{Y}}_{12}$
$\hat{\bar{Y}}_{p1}$	0.8621	0.8649	0.8624	0.8902	0.8626	0.8638	0.8645	0.8683	0.8797	0.8629	0.8659	0.8625
$\hat{\bar{Y}}_{p2}$	0.8643	0.8671	0.8647	0.8925	0.8648	0.8660	0.8667	0.8705	0.8820	0.8651	0.8681	0.8647
$\hat{\bar{Y}}_{p3}$	0.8658	0.8686	0.8662	0.8941	0.8664	0.8676	0.8683	0.8721	0.8836	0.8667	0.8696	0.8662
$\hat{\bar{Y}}_{p4}$	0.8646	0.8674	0.8649	0.8927	0.8650	0.8663	0.8669	0.8707	0.8822	0.8654	0.8683	0.8649
$\hat{\bar{Y}}_{p5}$	0.8660	0.8688	0.8664	0.8943	0.8665	0.8678	0.8684	0.8722	0.8838	0.8669	0.8698	0.8664
$\hat{\bar{Y}}_{p6}$	0.8650	0.8678	0.8654	0.8933	0.8656	0.8668	0.8675	0.8712	0.8828	0.8659	0.8688	0.8654

By this way, we compute relative efficiency values for each proposed estimator with the traditional estimator, as shown in Table 3. If the relative efficiency value obtained from (5.1.1) is smaller than 1, then it is apparent that the proposed estimator has a smaller MSE than the estimators presented in Kadilar and Cingi (2004, 2006) and Yan and Tian (2010). Therefore, from Table 3, we see that all the proposed ratio estimators using robust regression are more efficient than the traditional ratio estimators. However, this result is expected because the condition (4.2) is satisfied for all proposed estimators as follows:

$$-2R_{p1} \cong -2R_{p2} \cong -8,$$

$$B_{rob} - B \cong -0.62$$

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

5.2. Discussion

Analytical comparison shows that the proposed estimators using robust regression are more efficient than the existing estimators given by Kadilar and Cingi (2004), Kadilar and Cingi (2006) and Yan Tian (2010), when the conditions in equation 4.1 or 4.2 are satisfied.
Empirical results from Table 3 show that the proposed estimators using robust regression are more efficient than Kadilar and Cingi (2004), Kadilar and Cingi (2006) and Yan Tian (2010) for the population used in the present study. From Table 2 we observe that among all proposed estimators, $\hat{\bar{Y}}_{p1}$, $\hat{\bar{Y}}_{p2}$, and $\hat{\bar{Y}}_{p4}$ are the estimators which have least mean square error than other proposed estimators in the present study, but all the proposed estimators are more efficient than the existing ratio estimators.

6. CONCLUSIONS

Thus from the above study we analyze that our proposed estimators perform better than the existing estimators under study using the robust regression as their mean square error is much lower than the existing estimators.

Acknowlegmente: The development of this paper was benefited by the project A Cuban-Flemish Training and Research Program in Data Science and Big Data Analysis. The authors are grateful by the suggestions of two anonymous referees which allowed improving the first version of this paper.

RECEIVED: NOVEMBER, 2017
REVISED: MARCH, 2017

REFERENCES

[1] BIRKES, D. and DODGE, Y. (1993): **Alternative Methods of Regression**. John Wiley & Sons.
[2] CANDAN, M. (1995): Robust Estimators in Linear Regression Analysis. **Hacettepe University, Department of Statistics, Master Thesis (in Turkish)**.
[3] HAMPEL, F. R. (1975): Beyond location parameters: robust concepts and methods. **Proceedings of the 40th session of the ISI,** 46, 375-391.
[4] HUBER P. J. (1981): ***Robust Statistics*.** John Wiley & Sons, New York.
[5] HUBER, P. J. (1964): Robust Estimation of a Location Parameter. **Annals of Mathematical Statistics,** 35, 73-101.
[6] HUBER, P. J. (1973): Robust regression: Asymptotics, conjectures, and Monte Carlo. **Ann. Stat.,** 1, 799-821.
[7] KADILAR, C. and CINGI, H. (2004): Ratio estimators in simple random sampling. **Applied Mathematics and Computation,** 151, 893-902.
[8] KADILAR, C. and CINGI, H. (2006): An Improvement in Estimating the Population mean by using the Correlation Coefficient. **Hacettepe Journal of Mathematics and Statistics,** 35, 1, 103-109.
[9] ROUSSEEUW, P. J. (1984): Least median of squares regression. **Journal of the American Statistical Association,** 79, 871-880.
[10] ROUSSEEUW, P. J. and LEROY, M. A. (1987): ***Robust Regression and Outlier Detection.*** John Wiley & Sons, New York.
[11] SINGH, D. and CHAUDHARY, F. S. (1986): **Theory and Analysis of Sample Survey Designs.** 1 edn, New Age International Publisher, India.
[12] WOLTER, K. M. (1985): **Introduction to Variance Estimation.** Springer-Verlag.
[13] YAN, Z. and TIAN, B. (2010): Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable. **ICICA, Part II, CCIS,** 106, 103–110.