UTILIZATION OF DIFFERENT ROBUST REGRESSION TECHNIQUES FOR ESTIMATION OF FINITE POPULATION MEAN IN SRSWOR IN CASE OF PRESENCE OF OUTLIERS THROUGH RATIO METHOD OF ESTIMATION

Mir Subzar*, Carlos. N. Bouza** and Amer Ibrahim Al-Omari***
*Division of Agricultural Statistics SKUAST-K, Shalimar (190025), India
**Facultad de Matemática y Computación, Universidad de La Habana, Cuba
***Department of Mathematics, Faculty of Science, Al al-Bayt University, Mafraq 25113, Jordan.

ABSTRACT
Ratio type estimators are extensively used in sampling theory in order to get precise estimates of the population parameters by taking the advantage of positive (high) correlation between study and auxiliary variable than usual sample mean estimator. In this study we encountered with the problem of presence of outliers in the data and using of traditional methods usually decreases the efficiency in estimating the population parameters as these methods are sensitive to outliers. So in the present study we adapt the various robust regression techniques such as LTS, LMS, LAD, Huber M, Hampel M, Tukey M and Huber MM estimation to the ratio estimators which were suggested by Abid *et al.* (2016) by incorporated ancillary information using OLS method and also adapt Huber M-estimation to above estimators. Theoretically, we obtain the mean square error (MSE) for these estimators. We compared MSE values of the proposed estimators with MSE values based on Huber M which was proposed by Kadilar *et al.* (2007) and OLS methods. From this comparison we observe that our proposed estimators give more efficient results than both Huber M and OLS approach. These theoretical results are supported with the aid of a numerical example.

KEYWORDS: ratio type estimators, robust regression methods, ancillary information, simple random sampling, efficiency.

MSC:62D05

RESUMEN
Estimadores del tipo razón son usados extensamente en la teoría del muestreo, para obtener estimados más precisos de la media de la población que los obtenidos usando la media muestral, aprovechando la correlación positiva (alta) entre la variable de estudio y la auxiliar. En este estudio tratamos el problema de la presencia de outliers en la data; usando los métodos tradicionales la eficiencia decrece usualmente en la estimación de los parámetros poblacionales pues estos son sensitivos a los outliers. Así que en el presente estudio adaptamos varios estimadores robustos de la regresión como LTS, LMS, LAD, Huber M, Hampel M, Tukey M y Huber MM-estimación para estimadores de razón, sugeridos por Abid *et al.* (2016) incorporando información auxiliar proveída por el método OLS, y también adaptando la M-estimación de Huber a esos métodos.. Teóricamente, obtenemos el error cuadrático medio (MSE) de esos estimadores. Comparamos los valores de los MSE de las propuestas con los basados en M- Huber, propuestos por Kadilar *et al.* (2007), y los métodos OLS . En sus comparaciones observamos que nuestra propuesta provee más eficientes estimadores que los obtenidos por el enfoque M- Huber y OLS. Estos resultados teóricos son ilustrados usando un ejemplo numérico

PALABRAS CLAVE: estimadores de tipo razón, métodos de regresión robusta, información auxiliar, muestreo simple aleatorio, eficiencia.

1. INTRODUCTION

In survey sampling it is always advantageous that the use of ancillary information increases the precision while estimating the population parameters. In simple random sampling, when the correlation between study and auxiliary variable exist and that is positive (high) then ratio method of estimation ids generally used, as this method give more precise results in such situation. However in survey sampling one big issue comes there, that is when there is a presence of outliers in the data. In that situation OLS method does not yield precise result. In that situation then in 2007 Kadilar *et al*. adapted the Huber M-Estimation technique to the estimators obtained by OLS method in order to reduce the negative effects of outliers. However in the present study our main focus is to reduce the negative effect of outliers in survey sampling by busing the other Robust Regression technique other than Huber M-Estimation technique.

The remainder of this paper is organized as follows. In section 2, we introduce robust regression techniques which we used in the present study. In section 3 we offered the traditional ratio estimators in srswor given by Abid et al (2016) with their MSE equations. In section 4 we offered the above mentioned ratio estimators but adapted the Huber M estimation technique and other different robust regression techniques. In section 5 we compare the efficiencies of our suggested estimators and efficiencies of the estimators obtained by OLS. Numerical example is provided in section 6 and finally arriving at the conclusion from these results.

2. ROBUST REGRESSION TECHNIQUES

As outlier exists in the data set, using OLS method do not yield reliable results as they are very sensitive to outliers. So keeping in mind this serious problem, there are lot of methods which are not sensitive to outliers known as Robust Regression methods and are introduced in this section.

2.1 Least Absolute Deviations Method (LAD)

Suggested by Boschovich in 1757 and improved by Edgeworth in 1877. Lad regression is the first step for robust regression methods (Nadia and Mohammad 2013): LAD is a method which minimizes the sum of absolute error and is described as follows

$$\min \sum_{i=1}^{n} |\varepsilon_i| \quad \text{i.e., (Minimize } |e| \text{ instead of } e^2 \qquad (2.1.1)$$

This method was developed to decrease outliers in the direction of y in OLS. Outliers in the direction of y has very little effect on LAD method. But it is sensitive for outlier in the direction of x just like in OLS. Actually breaking point of LAD is low. This ratio is $1/n$ and $o\left(\frac{1}{n}\right)$.

2.2 Least Median of Squares Method (LMS)

This method was suggested by Rousseeuw and improved by Rousseeuw and Leroy (1987): This method is an alternative bounded influence method. This method rather than minimize the sum of the least squares function, this model minimizes the median of the squared residuals i.e., E_i^2, so LMS is very robust with respect to outliers both in terms of X and Y values. i.e.,

$$\min median\left(\varepsilon_i^2\right)$$
(2.2.1)

2.3 Least Trimmed Squares Method (LTS)

These estimators can have a breakdown point 50%, i.e., half the data can be influential in the OLS sense before the LTS estimator is seriously affected. Least trimmed squares essentially proceeds with OLS after eliminating the most extreme positive or negative residuals. Least trimmed squares orders the squared residuals from smallest to largest: $\left(E^2\right)_{(1)}, \left(E^2\right)_{(2)}, \ldots, \left(E^2\right)_{(n)}$ and then it calculates b that minimizes the sum of only the smaller half of the residuals

$$\sum_{i=1}^{m} \left(E^2\right)_{(i)} \qquad (2.3.1)$$

Where $m = [n/2] + 1$; the square bracket indicates rounding down to the nearest integer.

2.4 M-Estimation

Huber, Hampel and Tukey-M methods used commonly in the literature are analyzed in this section. M Estimators are developed as an alternative to OLS for the situations where error terms do not satisfy normal distribution assumption for the universe (Ergrol; 2006):

2.4.1 Huber-M Estimation function:

Huber (1973) suggested an estimator class known as M-Estimators. This method is actually a good compromise between the efficiency of the least squares and the robustness of the least absolute values estimators is the Huber objective function. At the center of the distribution the Huber function behaves like the OLS function, but at the extremes it behaves like the LAV function.

$$rH\left(E\right)=\begin{cases}\frac{1}{2}E^{2} & for\ E\leq K\\ K\left|E\right|-\frac{1}{2}K^{2} & for\ E>K \qquad K\ is\ small\end{cases} \tag{2.4.1.1}$$

The influence function is determined by taking the derivative

$$\psi H(E)=\begin{cases}k & for\ E>K\\ E & for\ E\leq K\\ -K & for\ E<-K\end{cases} \tag{2.4.1.2}$$

The tuning constant K defines the center and tails. This tuning constant is expressed as a multiple of the scale (the spread) of Y , $K=cS$, where S is the measure of the scale of Y (i.e., the spread):
We could use the standard deviation as a measure of scale, but it is more influenced by extreme observations than is the mean, so instead, we use the median absolute deviation.

$$MAD=median\left|Y_i-\hat{\mu}\right|. \tag{2.4.1.3}$$

The median of Y serves as an initial estimate of $\hat{\mu}$, thus allowing us to define $SMAD/0.6745$, which ensures that S estimates σ when the population is normal. So using $k=1.345(1.345/.6745)$ is about 2 produces 95% efficiency relative to the sample mean when the population is normal and gives substantial resistance to outliers when it is not. A smaller k gives more resistance.

2.4.2 Hampel M Estimation function

The function suggested by Hampel (1971) is as follows,

$$\rho(y)=\begin{cases}\frac{y^2}{2} & ,0<|y|<a\\ a|y|-\frac{y^2}{2} & ,a<|y|\leq b\\ \frac{-a}{2(c-b)}(c-y)^2+\frac{a}{2}(b+c-a) & ,b<|y|\leq c\\ \frac{a}{2}(b+c-a) & ,c<|y|\end{cases} \tag{2.4.2.1}$$

Where $a=1.7,\ b=3.4$ and $c=8.5$.

2.4.3 Tukey M estimation function:

This function was suggested by Tukey in (1977) and is as under,

$$\rho(y)=\begin{cases}\frac{1}{6}\left(1-\left(1-\left(\frac{y}{k}\right)^2\right)^3\right) & ,\ |y|\leq k\\ \frac{1}{6} & ,\ |y|>k\end{cases} \tag{2.4.3.1}$$

Where $k=5$ or $k=6$

M regression method is sensitive to outliers in the direction of x. That is it is under the effect of x observation values.

2.4.4 Huber MM estimation method

It is suggested by Yohai (1987) as a method whose statistical efficiency is high and has a break down point. The algorithm of MM estimation method is described as follows;

Step 1: A starting estimation with high breakdown point (0.5 if possible) is chosen.

Step 2: Outliers are calculated as $e_i(T_O) = y_i - T_0 x_i,\ \ 1 \le i \le n$

Where T_0 is starting estimation. Under $\frac{b}{\alpha} = 0.5$ constraints, b is calculated as below;

$$\left(\frac{1}{n}\right)\sum_{i=1}^{n}\rho\left(\frac{e_i(\beta)}{s_n}\right) = b \tag{2.4.4.1}$$

Where s_n is M scale estimation and it is calculated as $s_n = s(e(T_0))$.. And using ρ_0 which satisfies the assumptions given in Yohai (1987), it is represented as $\alpha = \max \rho_0(u)$. For more detailed information, Yohai (1987) study can be reviewed.

3 EXISTING RATIO ESTIMATORS USING OLS METHOD:

In this section we mention the ratio estimators suggested by Abid et al. (2016) by utilizing the auxiliary information of correlation coefficient, coefficient of variation with Gini's mean difference, Downton's method and probability the weighted moment method of the auxiliary variable. The suggested estimators are given as

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

$$\hat{\bar{Y}}_4 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+D)}(\bar{X}+D),\quad \hat{\bar{Y}}_5 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\rho+D)}(\bar{X}\rho+D),\quad \hat{\bar{Y}}_6 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}C_x+D)}(\bar{X}C_x+D),$$

$$\hat{\bar{Y}}_7 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_{pw})}(\bar{X}+S_{pw}),\quad \hat{\bar{Y}}_8 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\rho+S_{pw})}(\bar{X}\rho+S_{pw}),$$

$$\hat{\bar{Y}}_9 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}C_x+S_{pw})}(\bar{X}C_x+S_{pw}),$$

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

$$h(\bar{x},\bar{y}) \cong h(\bar{X},\bar{Y}) + \frac{\partial h(c,d)}{\partial c}\Big|_{\bar{X},\bar{Y}}(\bar{x}-\bar{X}) + \frac{\partial h(c,d)}{\partial d}\Big|_{\bar{X},\bar{Y}}(\bar{y}-\bar{Y}) \tag{3.1}$$

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

As shown in Wolter (1985), (3.1) can be applied to the proposed estimator in order to obtain MSE equation as follows:

$$\hat{R}_1 - R \cong \frac{\partial((\bar{y}+b(\bar{X}-\bar{x}))/(\bar{x}+G)}{\partial\bar{x}+G}\Big|_{\bar{X},\bar{Y}}(\bar{x}-\bar{X}) + \frac{\partial((\bar{y}+b(\bar{X}-\bar{x}))/(\bar{x})}{\partial\bar{y}+G}\Big|_{\bar{X},\bar{Y}}(\bar{y}-\bar{Y})$$

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

$$E(\hat{R}_1-R)^2 \cong \frac{(\bar{Y}+B(\bar{X}+G))^2}{(\bar{X}+G)^4}V(\bar{x})-\frac{2(\bar{Y}+B(\bar{X}+G))}{(\bar{X}+G)^3}Cov(\bar{x},\bar{y})+\frac{1}{(\bar{X}+G)^2}V(\bar{y})$$

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

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 $(E(b)-B)$.

$$MSE(\bar{y}_1)=(\bar{X}+G)^2E(\hat{R}_1-R)^2 \cong \frac{(\bar{Y}+B(\bar{X}+G))^2}{(\bar{X}+G)^2}V(\bar{x})-\frac{2(\bar{Y}+B(\bar{X}+G))}{(\bar{X}+G)}Cov(\bar{x},\bar{y})+V(\bar{y})$$

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

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

$$MSE(\bar{y}_1)\cong \frac{(1-f)}{n}(R_1^2S_x^2+2BR_1S_x^2+B^2S_x^2-2R_1S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_1=\frac{\bar{Y}}{\bar{X}+G}$$

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

$$MSE(\hat{\bar{Y}}_2)\cong \frac{(1-f)}{n}(R_2^2S_x^2+2BR_2S_x^2+B^2S_x^2-2R_2S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_2=\frac{\bar{Y}\rho}{\bar{X}\rho+G}$$

$$MSE(\hat{\bar{Y}}_3)\cong \frac{(1-f)}{n}(R_3^2S_x^2+2BR_3S_x^2+B^2S_x^2-2R_3S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_3=\frac{\bar{Y}C_x}{\bar{X}C_x+G}$$

$$MSE(\hat{\bar{Y}}_4)\cong \frac{(1-f)}{n}(R_4^2S_x^2+2BR_4S_x^2+B^2S_x^2-2R_4S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_4=\frac{\bar{Y}}{\bar{X}+D}$$

$$MSE(\hat{\bar{Y}}_5)\cong \frac{(1-f)}{n}(R_5^2S_x^2+2BR_5S_x^2+B^2S_x^2-2R_5S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_5=\frac{\bar{Y}\rho}{\bar{X}\rho+D}$$

$$MSE(\hat{\bar{Y}}_6)\cong \frac{(1-f)}{n}(R_6^2S_x^2+2BR_6S_x^2+B^2S_x^2-2R_6S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_6=\frac{\bar{Y}C_x}{\bar{X}C_x+D}$$

$$MSE(\hat{\bar{Y}}_7)\cong \frac{(1-f)}{n}(R_7^2S_x^2+2BR_7S_x^2+B^2S_x^2-2R_7S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_7=\frac{\bar{Y}}{\bar{X}+S_{pw}}$$

$$MSE(\hat{\bar{Y}}_8)\cong \frac{(1-f)}{n}(R_8^2S_x^2+2BR_8S_x^2+B^2S_x^2-2R_8S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_8=\frac{\bar{Y}\rho}{\bar{X}\rho+S_{pw}}$$

$$MSE(\hat{\bar{Y}}_9)\cong \frac{(1-f)}{n}(R_9^2S_x^2+2BR_9S_x^2+B^2S_x^2-2R_9S_{xy}-2BS_{xy}+S_y^2),\ \text{where } R_9=\frac{\bar{Y}C_x}{\bar{X}C_x+S_{pw}}$$

4. SUGGESTED ESTIMATORS USING DIFFERENT ROBUST REGRESSION TECHNIQUES

In this section we adapt the different robust regression techniques to the ratio estimators mentioned in the section 3 as the above mentioned estimators are sensitive to the extreme values in the data, so the main aim of this present study is reduce the lacuna of sensitivity to the extreme values in order to get precise results even when the extreme values are present. The suggested estimators using different robust regression techniques such as LTS, LMS, LAD Huber M, Hampel M, Tukey M and Huber MM estimation are given as under

$$\hat{\bar{Y}}_{p1} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x} + G)}(\bar{X} + G), \qquad \hat{\bar{Y}}_{p2} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x}\rho + G)}(\bar{X}\rho + G),$$

$$\hat{\bar{Y}}_{p3} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x}C_x + G)}(\bar{X}C_x + G), \qquad \hat{\bar{Y}}_{p4} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x} + D)}(\bar{X} + D),$$

$$\hat{\bar{Y}}_{p5} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x}\rho + D)}(\bar{X}\rho + D), \qquad \hat{\bar{Y}}_{p6} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x}C_x + D)}(\bar{X}C_x + D),$$

$$\hat{\bar{Y}}_{p7} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x} + S_{pw})}(\bar{X} + S_{pw}), \qquad \hat{\bar{Y}}_{p8} = \frac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x}\rho + S_{pw})}(\bar{X}\rho + S_{pw}),$$

$\hat{\bar{Y}}_{p9} = \dfrac{\bar{y} + b_{(rob)i}(\bar{X} - \bar{x})}{(\bar{x}C_x + S_{pw})}(\bar{X}C_x + S_{pw}),$ where $b_{(rob)i}$ = LTS, LMS, LAD, Huber M, Hampel M, Tukey M and Huber MM estimation.

The Mean square error and the related constant of the above estimators are respectively given as

$$MSE(\bar{y}_{p1}) \cong \frac{(1-f)}{n}(R_1^2 S_x^2 + 2B_{(rob)i}R_1 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_1 S_{xy} - 2B_{(rob)i}S_{xy} + S_y^2),$$

where $R_1 = \dfrac{\bar{Y}}{\bar{X} + G}$

$$MSE(\hat{\bar{Y}}_{p2}) \cong \frac{(1-f)}{n}(R_2^2 S_x^2 + 2B_{(rob)i}R_2 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_2 S_{xy} - 2B_{(rob)i}S_{xy} + S_y^2),$$

where $R_2 = \dfrac{\bar{Y}\rho}{\bar{X}\rho + G}$

$$MSE(\hat{\bar{Y}}_{p3}) \cong \frac{(1-f)}{n}(R_3^2 S_x^2 + 2B_{(rob)i}R_3 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_3 S_{xy} - 2B_{(rob)i}S_{xy} + S_y^2),$$

where $R_3 = \dfrac{\bar{Y}C_x}{\bar{X}C_x + G}$

$$MSE(\hat{\bar{Y}}_{p4}) \cong \frac{(1-f)}{n}(R_4^2 S_x^2 + 2B_{(rob)i}R_4 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_4 S_{xy} - 2B_{(rob)i}S_{xy} + S_y^2),$$

where $R_4 = \dfrac{\bar{Y}}{\bar{X} + D}$

$$MSE(\hat{\bar{Y}}_{p5}) \cong \frac{(1-f)}{n}(R_5^2 S_x^2 + 2B_{(rob)i}R_5 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_5 S_{xy} - 2B_{(rob)i}S_{xy} + S_y^2),$$

where $R_5 = \dfrac{\bar{Y}\rho}{\bar{X}\rho + D}$

$$MSE(\hat{\bar{Y}}_{p6}) \cong \frac{(1-f)}{n}(R_6^2 S_x^2 + 2B_{(rob)i}R_6 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_6 S_{xy} - 2B_{(rob)i}S_{xy} + S_y^2),$$

where $R_6 = \frac{\bar{Y}C_x}{\bar{X}C_x + D}$

$$MSE(\hat{\bar{Y}}_{p7}) \cong \frac{(1-f)}{n}(R_7^2 S_x^2 + 2B_{(rob)i} R_7 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_7 S_{xy} - 2B_{(rob)i} S_{xy} + S_y^2),$$

where $R_7 = \frac{\bar{Y}}{\bar{X} + S_{pw}}$

$$MSE(\hat{\bar{Y}}_{p8}) \cong \frac{(1-f)}{n}(R_8^2 S_x^2 + 2B_{(rob)i} R_8 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_8 S_{xy} - 2B_{(rob)i} S_{xy} + S_y^2),$$

where $R_8 = \frac{\bar{Y}\rho}{\bar{X}\rho + S_{pw}}$

$$MSE(\hat{\bar{Y}}_{p9}) \cong \frac{(1-f)}{n}(R_9^2 S_x^2 + 2B_{(rob)i} R_9 S_x^2 + B_{(rob)i}^2 S_x^2 - 2R_9 S_{xy} - 2B_{(rob)i} S_{xy} + S_y^2),$$

where $R_9 = \frac{\bar{Y}C_x}{\bar{X}C_x + S_{pw}}$

Where $b_{(rob)i}$ = LTS, LMS, LAD, Huber M, Hampel M, Tukey M and Huber MM estimation. The above suggested estimators are 63 and the robust regressions are adapted one by one.

5. EFFICIENCY COMPARISONS

In this section we have derived the theoretical efficiency comparisons of the ratio estimators using OLS method with ratio estimators in which different robust regression techniques has been adapted.

$$MSE(\bar{Y}_{pj}) < MSE(\bar{Y}_k), pj = 1,2,....,63. \qquad k = 1,2,...,9$$

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

where $B_{(rob)i}$ indicated the robust regression techniques which are adapted to the ratio estimators mentioned in section 3, the techniques are: LTM, LMS, LAD, Huber M, Hampel M, Tukey M and Huber MM.

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

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

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

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

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

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

Similarly, for $B_{(rob)i} - B < 0,$ that is $B_{(rob)i} < B: B_{(rob)i} > B - 2R_{pi_{(i)}}.$

Consequently, we have the following conditions:

$$0 < B_{(rob)i} - B < 2R_{pi_{(i)}} \tag{5.1}$$

or

$$-2R_{pi_{(i)}} < B_{(rob)i} - B < 0. \tag{5.2}$$

When condition (5.1) or (5.2) is satisfied, the proposed estimators given in Section 3 are more efficient than the ratio estimator, given in section.2, respectively.

6. 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	Parameter	Population
N	34	D_1	70.3	S_y	733.1407	D_6	227.2
n	20	D_2	76.8	B	2.19	D_7	250.4
$\bar{Y}$	856.4117	D_3	108.2	$Brob$	1.57	D_8	335.6
$\bar{X}$	208.8823	D_4	129.4	S_x	150.5059	D_9	436.1
ρ	0.4491	D_5	150.0	M_d	150	D_{10}	564.0

Table 2: The Statistical Analysis (MSE) of the Estimators for the Populations

Estimators	Constant	OLS	LTS	LMS	LAD	Huber M	Tukey M	Hampel M	Huber MM
1	2.3507	11415.84	10092.74	10069.86	10237.67	10234.44	10125.13	10137.58	10146.95
2	1.5431	10881.01	9790.14	9772.16	9904.97	9233.25	9175.92	9182.34	9187.19
3	2.0168	10735.21	9633.18	9614.97	9749.46	9746.85	9659.02	9668.98	9676.48
4	2.4485	11634.98	10247.10	10222.85	10400.42	10397.01	10281.40	10294.58	10304.50
5	1.6387	10089.82	9238.18	9225.26	9322.02	9320.11	9256.62	9263.75	9269.14
6	2.1176	10929.66	9760.92	9741.30	9885.85	9883.05	9788.74	9799.45	9807.51
7	2.0947	10884.69	9731.09	9711.79	9854.06	9851.30	9758.47	9769.00	9776.94
8	1.3092	9636.10	9002.68	8994.36	9058.24	9056.95	9014.66	9019.34	9022.88
9	1.7607	10283.36	9350.98	9336.35	9445.28	9443.15	9371.81	9379.85	9385.93

Table 3: % RE of the estimators using OLS with the estimators Using Different Robust Regression Techniques except Huber M Estimation

Estimators	OLS/LTS	OLS/LMS	OLS/LAD	OLS/Hampel M	OLS/Tukey M	OLS/Huber MM
1	113.109	113.366	111.508	112.609	112.748	112.505
2	111.143	111.347	109.854	118.499	118.582	118.437
3	111.440	111.651	110.111	111.027	111.142	110.941
4	113.544	113.813	111.870	113.020	113.165	112.912
5	109.219	109.372	108.236	108.917	109.001	108.854
6	111.974	112.199	110.559	111.533	111.655	111.442
7	111.855	112.077	110.459	111.421	111.541	111.330
8	107.036	107.135	106.379	106.838	106.894	106.796
9	109.971	110.143	108.873	109.632	109.727	109.561

Table 4: % RE of the estimators using Huber M estimation used by Kadilar in (2007) with the estimators Using Different Robust Regression Techniques

Estimators	Huber M/LTS	Huber M /LMS	Huber M /LAD	Huber M /Hampel M	Huber M /Tukey M	Huber M /Huber MM
1	101.404	101.634	99.968	101.080	100.955	100.862
2	94.312	94.485	93.218	100.625	100.554	100.501
3	101.180	101.372	99.973	100.909	100.805	100.727
4	101.463	101.704	99.967	101.124	100.995	100.898
5	100.887	101.028	99.980	100.686	100.608	100.550
6	101.251	101.455	99.972	100.963	100.853	100.770
7	101.235	101.436	99.972	100.951	100.842	100.761
8	100.603	100.696	99.986	100.469	100.417	100.378
9	100.986	101.144	99.977	100.761	100.675	100.610

7. DISCUSSION

From the numerical study given in the section 5 which are represented in table 2, 3 and 4 respectively. In table 2 we provide the MSE values of the estimators using OLS method and the other different robust regression techniques, from this table we conclude that using different robust regression techniques have lower MSE than the estimators using OLS. Table 3 reveals that in which %relative efficiency is calculated between the estimators using OLS and other robust regression techniques except Huber M-Estimation and finally we came to the conclusion that our suggested estimators in which different robust regression techniques are adapted perform better than the estimators using OLS in case of the outliers present in the data. In Table 4 in which %relative efficiency is calculated between the estimators using Huber M estimation and Different Robust Regression Techniques. This table reveals that almost all the robust regression techniques are more efficient than the Huber M Estimation technique which was adapted by Kadilar et al. (2007), but the Least Absolute Deviation Estimation method does not perform better than the Huber M estimation in case of presence of outliers in the data.

8. CONCLUSION

Thus from the above study we conclude that our suggested estimators in which different Robust Regression techniques are adapted perform better than the estimators in which the OLS and Huber M-Estimation is adapted. Hence we strongly recommend our suggested estimators in case of the presence of the outliers over the estimators in which the OLS and Huber M Estimation is adapted.

RECEIVED: FEBRUARY, 2019.
REVISED: APRIL, 2019.

REFERENCES

[1] ABID, M., SHERWANI, R. A. K., ABBAS, N. and NAWAZ, T. (2016): Some improved modified ratio estimators based on decile mean of an auxiliary variable. **Pakistan Journal of Statistic and Operation Research, 12,** 787-797.
[2] ERGUL, B. (2006): **Robust regression and applications.** Master thesis, Eski sehir Osmangazi University. 184031.
[3] HAMPEL, F. R. (1971): A general qualitative definition of robustness. **The Annals of Mathematical Statistics**, 42 , 1887–96.
[4] HUBER, P. J. (1973): Robust regression: Asymptotics, conjectures and Monte Carlo. **The Annals of Statistics**, 1, 799–821.
[5] KADILAR, C., M. CANDAN, and H. CINGI. (2007): Ratio estimators using robust regression. **Hacettepe Journal of Mathematics and Statistics,** 36, 181–88.

[6] NADIA, H., and MOHAMMAD, A. A. (2013): Model of robust regression with parametric and nonparametric methods. **Mathematical Theory and Modeling**, 3, 27-39.
[7] ROUSSEEUW, P. J. and A. M. LEROY. (1987): **Robust regression and outlier detection**. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.
[8] TUKEY, J. W. (1977): **Exploratory data analysis**. Addison-Wesley, MA.
[9] WOLTER, K. M. (1985): **Introduction to Variance Estimation**. Springer-Verlag, Berlin.
[10] YOHAI, V. J. (1987): High breakdown-point and high efficiency robust estimates for regression. **The Annals of Statistics**, 15, 642–656.