International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.04, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

ROBUST AND EFFICIENT RATIO ESTIMATORS FOR FINITE POPULATION MEAN IN SURVEY SAMPLING USING NON- CONVENTIONAL MEASURES OF DISPERSION

***Mir Subzar[1], S. Maqbool[2] , T A Raja[3]**

[1,2,3] *Division of Agricultural Statistics, SKUAST-Kashmir (India)*

ABSTRACT

A new Robust approach to estimate the finite population mean when the population under survey is skewed by using Gini's mean difference, Downton's method and Probability weighted moment as auxiliary variables for the improvement of estimators. Where as usual estimation method like least square method does not provide us valid information and precise estimates. Thus the technique of Huber M-Estimation has been employed to these robust estimators to obtain valid and precise estimates under such situations. Theoretical and numerical illustration is given to seek the efficiency of estimators using robust regression over the estimators without using robust regression.

Keywords - Efficiency, Ratio Estimators, Robust regression, Supplementary Information.

I.INTRODUCTION

From the last few decades, researcher's interest takes place towards this, that 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.

So dealing with this type of situation when there is unusual data, it is first to screen the data. If outliers are present then either to remove and then apply classical inferential procedure is not simple and good way to proceed. In multivariate or highly structured data, it can be difficult to single out outliers or it can be even impossible to identify influential observations. If the influential observation is rejected or discarded from the data, can reduce the sample size, which can affect the distribution theory and variance could be underestimated from the cleaned data. Thus even one outlying observation can destroy least squares estimation and does not provide us useful information for the majority of data. Keeping the above mentioned problem in view we have utilized a new approach known as Robust regression which was first introduced by [6], [7], and it is well known as M-regression estimator. However, the outlier problem, which is the presence of extreme values in data, generally decreases the efficiency, since classical estimators are sensitive to these extreme values [9]. So utilization of robust regression provide us valid information and the primary purpose is to fit a model which represents the information in the majority of the data. So in the present study we have adapted robust regression

to the ratio estimators using the auxiliary information of non-conventional measures of dispersion such as Gini's mean difference, Downton's method and Probability weighted moment.

II.EXISTING ESTIMATORS IN LITERATURE

In this section we discuss the estimators which [3] Proposed for estimating the finite population mean in simple random sampling. The estimators introduced by [3] are given as under:

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

$$\hat{\bar{Y}}_2 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + D)}(\bar{X} + D),$$

$$\hat{\bar{Y}}_3 = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + S_{pw})}(\bar{X} + S_{pw})$$

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

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

(1)

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

As shown in [10], (1) can be applied to the proposed estimator in order to obtain MSE equation as follows:

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

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

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

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

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 $(E(b) - B).$

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

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

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

$$MSE(\bar{y}_1) \cong \frac{(1-f)}{n}(R_1^2 S_x^2 + 2BR_1 S_x^2 + B^2 S_x^2 - 2R_1 S_{xy} - 2BS_{xy} + S_y^2), \text{ where } R_1 = \frac{\bar{Y}}{\bar{X}+G}$$

Similarly the MSE of the another estimators can be obtained as

$$MSE(\hat{\bar{Y}}_2) \cong \frac{(1-f)}{n}(R_2^2 S_x^2 + 2BR_2 S_x^2 + B^2 S_x^2 - 2R_2 S_{xy} - 2BS_{xy} + S_y^2), \text{ where } R_2 = \frac{\bar{Y}}{\bar{X}+D}$$

$$MSE(\hat{\bar{Y}}_3) \cong \frac{(1-f)}{n}(R_3^2 S_x^2 + 2BR_3 S_x^2 + B^2 S_x^2 - 2R_3 S_{xy} - 2BS_{xy} + S_y^2), \text{ where } R_3 = \frac{\bar{Y}}{\bar{X}+S_{pw}}$$

Where $\bar{y}$ 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. $G = \frac{4}{N-1}\sum_{i=1}^{N}\left(\frac{2i-N-1}{2N}\right)X_{(i)}$ is Gini's Mean Difference, $D = \frac{2\sqrt{\pi}}{N(N-1)}\sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)X_{(i)}$ is Downton's Method and $S_{pw} = \frac{\sqrt{\pi}}{N^2}\sum_{i=1}^{N}(2i-N-1)X_{(i)}$ Probability Weighted moments.

III.IMPROVED ESTIMATORS USING ROBUST REGRESSION

In this section we keep the above cited problem of outliers in the data in view, so in this paper we have adapted the robust regression to the above mentioned estimators as the above mentioned estimators are obtained by linear least square estimation which can give inaccurate estimates in case of presence of influential observations. Thus the improved and robust estimators are given as under:

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

$$\hat{\bar{Y}}_{p2} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}+D)}(\bar{X}+D),$$

$$\hat{\bar{Y}}_{p3} = \frac{\bar{y} + b_{rob}(\bar{X} - \bar{x})}{(\bar{x}+S_{pw})}(\bar{X}+S_{pw})$$

The mean square error of the above estimators are given as under

$$MSE(\bar{y}_{p1}) \cong \frac{(1-f)}{n}(R_{p1}^2 S_x^2 + 2B_{rob}R_1 S_x^2 + B_{rob}^2 S_x^2 - 2R_{p1}S_{xy} - 2B_{rob}S_{xy} + S_y^2), \text{ where } R_1 = \frac{\bar{Y}}{\bar{X}+G}$$

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

where $R_2 = \frac{\bar{Y}}{\bar{X}+D}$

$$MSE(\bar{y}_{p3}) \cong \frac{(1-f)}{n}(R_{p3}^2 S_x^2 + 2B_{rob}R_3 S_x^2 + B_{rob}^2 S_x^2 - 2R_{p3}S_{xy} - 2B_{rob}S_{xy} + S_y^2), \quad \text{where}$$

$$R_3 = \frac{\bar{Y}}{\bar{X}+S_{pw}}$$

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. [5] 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 [4], [8].

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 [1].

We remark that the MSE equation of the proposed ratio estimators $\hat{\bar{Y}}_{Pj}$ $j = 1,2,3.$ is in the same form as the MSE equation given in section (2), but it is clear that B in equations given in section (2), 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 $E(b_{rob}) = B_{rob}$ in section (3), as for b in section (2). 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}.

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.04, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

IV.EFFICIENCY COMPARISON

In this section we have derived theoretically the efficiency comparison of the proposed estimators with the existing estimators by [3]. We compare the MSE of the proposed estimators, with the MSE of the Existing ratio estimators.

$$MSE(\bar{Y}_{pj}) < MSE(\bar{Y}_i),\ \ j = 1,2,3. \qquad i = 1,2,3.$$

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

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

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

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

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

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

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

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

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

Consequently, we have the following conditions:

$$0 < B_{rob} - B < 2R_{pj(i)}$$

(2)

or

$$-2R_{pj(i)} < B_{rob} - B < 0.$$

(3)

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

V.NUMERICAL ILLUSTRATION

For numerical illustration we have taken the data from the book Theory and Analysis of Sample Survey Designs by [2] 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). The Characteristics of the population is given in Table 1 and statistical analysis is given in Table 2.

Table 1. Characteristics of the population.

Parameter	Population	Parameter	Population
N	34	S_x	150.5059
n	20	C_x	0.7205
$\bar{Y}$	856.4117	β_2	0.0978
$\bar{X}$	208.8823	β_1	0.9782
ρ	0.4491	G	155.446
S_y	733.1407	D	140.891
C_y	0.8561	S_{pw}	199.961
B	2.19	$Brob$	1.57

Table 2: The statistical analysis of the estimators for the population

Estimators	Constant	MSE	Estimators	Constant	MSE
$\hat{\bar{Y}}_1$	2.3507	11415.84	$\hat{\bar{Y}}_{p1}$	2.3507	10234.44
$\hat{\bar{Y}}_2$	2.4485	11634.98	$\hat{\bar{Y}}_{p2}$	2.4485	10397.01
$\hat{\bar{Y}}_3$	2.0947	10884.69	$\hat{\bar{Y}}_{p3}$	2.0947	9851.30

VI.CONCLUSION

Thus from the above tables we reveal that our suggested estimators using robust regression perform better than the estimators without using Robust regression whenever there are influential observations in the data. Hence we strongly recommend that our estimators preferred over existing estimators for use in practical applications under such situations.

REFERENCES

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

[2] D. Singh and F. S. Chaudhary, *Theory and Analysis of Sample Survey Designs* (1 edn, New Age International Publisher, India, 1986).

[3] M. Abid, N. Abbas, R. A. K. Sherwani, and H. F. Nazir, Improved Ratio estimators for the population mean using non-conventional measures of dispersion, *Pakistan Journal of Statistics and operation research, 12(2)*, 2016, 353-367.

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

[5] P. J. Huber, Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics, 35*, 1964, 73-101.

[6] P. J. Huber, Robust regression: Asymptotics, conjectures, and Monte Carlo, *Ann. Stat., 1,* 1973, 799-821.

[7] P. J. Huber, *Robust Statistics* (John Wiley & Sons, New York, 1981).

[8] P. J. Rousseeuw and M. A. Leroy, *Robust Regression and Outlier Detection* (John Wiley & Sons, New York, 1987).

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

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