Generalized Family of Efficient Estimators of Population Median Using Two-Phase Sampling Design

[1,]H.S. Jhajj , [2,] Harpreet Kaur   , [3,]Puneet Jhajj

ABSTRACT : *For estimating the population median of variable under study, a generalized family of efficient estimators has been proposed by using prior information of population parameters based upon auxiliary variables under two-phase simple random sampling design. The comparison of proposed family of estimators has been made with the existing ones with respect to their mean square errors and biases. It has been shown that efficient estimators can be obtained from the family under the given practical situations which will have smaller mean square error than the linear regression type estimator, Singh et al estimator (2006), Gupta et al (2008) estimator and Jhajj et al estimator (2014). Effort has been made to illustrate the results numerically as well as graphically by taking some empirical populations considered in the literature which also show that bias of efficient estimators obtained from the proposed family of estimators is smaller than other considered estimators.*

KEYWORDS: *Auxiliary variable; Bias; Mean square error; Median; Two-phase sampling.*

I.   INTRODUCTION

In survey sampling, statisticians have given more attention to the estimation of population mean, total, variance etc. but median is regarded as a more appropriate measure of location than mean when the distribution of variables such as income, expenditure etc. is highly skewed. In such situations, it is necessary to estimate median. Some authors such as Gross(1980), Kuk and Mak (1989), Singh et al (2001,2006), Jhajj et al (2013) etc. have considered the problem of estimating the median.

Sometimes in a trivariate distribution consisting of study variable Y and auxiliary variables   (X, Z), the correlation between the variables Y and Z is only because of their high correlation with the variable X but are not directly correlated to each other . For example:

1. In agriculture labour (say Z) and crop production (say Y) are highly correlated with the area under crop (say X) but not directly correlated to each other.
2. In any repetitive survey, the values of a variables of interest corresponding to both the last to last year (say Z) and current year (say Y) are highly correlated with the values of some variable corresponding to the last year (say X), whereas the values corresponding to the last to last year (Z) and the values corresponding to the current year (Y) are correlated with each other due to only their correlation with values of the some variable (X) corresponding to the last year.

Suppose the prior information about population median M_Z of variable Z is available where as the population median M_X of variable X is not known. Such unknown information is generally predicted by using the two- phase sampling design.

Let u consider a finite population U= (1, 2,… i ,...N). Let Y_i and X_i , Z_i be the values of study variable Y and auxiliary variables X, Z respectively on the ith unit of the population. Corresponding small letters indicate the value in the samples. Let M_Y and M_X, M_Z be the population medians of study variable and auxiliary variables respectively. Under two- phase sampling design, first phase sample of size n is selected using simple random sampling without replacement and observations on Y, X and Z are obtained on the sample units. Then second phase sample of size m under simple random sampling without replacement is drawn from the first phase sample and observations on the variables Y, X and Z are taken on selected units. Corresponding sample medians are denoted by $\hat{M}_Y$, $\hat{M}_X$ and $\hat{M}_Z$ for second phase sample while $\hat{M}'_Y$, $\hat{M}'_X$ and $\hat{M}'_Z$ are sample medians for the first phase sample.

Suppose that $y_{(1)}, y_{(2)},…,y_{(m)}$ are the values of variable Y on sample units in ascending order such   that $y_{(t)} \leq M_Y \leq y_{(t+1)}$ for some integral value of t. Let p = t/m be the proportion of Y values in the sample that are less than or equal to the value of median M_Y (an unknown population parameter). If $\hat{p}$ is an estimator of p ,

the sample median $\hat{M}_Y$ can be written in terms of quantiles as $\hat{Q}_Y(\hat{p})$ with $\hat{p} = 0.5$.

Let the correlation coefficients between the estimators in $(\hat{M}_X, \hat{M}_Y)$, $(\hat{M}_Y, \hat{M}_Z)$ and $(\hat{M}_X, \hat{M}_Z)$ are denoted by ρ_{xy}, ρ_{yz} and ρ_{xz} respectively which are defined as

$$\rho_{xy} = \rho_{(\hat{M}_X, \hat{M}_Y)} = 4\{P_{11}(x, y) - 1\}, \quad \text{where} \quad P_{11}(x, y) = P(X \le M_X \cap Y \le M_Y)$$

$$\rho_{yz} = \rho_{(\hat{M}_Y, \hat{M}_Z)} = 4\{P_{11}(y, z) - 1\}, \quad \text{where} \quad P_{11}(y, z) = P(X \le M_Y \cap Y \le M_Z)$$

$$\rho_{xz} = \rho_{(\hat{M}_X, \hat{M}_Z)} = 4\{P_{11}(x, z) - 1\}, \quad \text{where} \quad P_{11}(x, z) = P(X \le M_X \cap Y \le M_Z)$$

Assuming that as N→∞, the distribution of the trivariate variable (X, Y, Z) approaches a continuous distribution with marginal densities $f_X(x)$, $f_Y(y)$ and $f_Z(Z)$ of variables X, Y and Z respectively. This assumption holds in particular under a superpopulation model framework, treating the values of (X, Y, Z) in the population as a realization of N independent observations from a continuous distribution. Under these assumptions, Gross (1980) has shown that conventional sample median $\hat{M}_Y$ is consistent and asymptotically normal with median M_Y and variance

$$Var(\hat{M}_Y) = \frac{\left(\dfrac{1}{m} - \dfrac{1}{N}\right)}{4\{f(M_Y)\}^2} \tag{1.1}$$

For the case of trivariate distribution of X, Y and Z and using two phase sampling design, linear regression type estimator of population median M_Y is defined as

$$\hat{M}_{Ylr} = \hat{M}_Y + \alpha(\hat{M}'_X - \hat{M}_X) + \beta(M_Z - \hat{M}'_Z) \tag{1.2}$$

where α and β are constants.

Up to the first order of approximation, $MSE(\hat{M}_{Ylr})$ is minimized for

$$\alpha = \rho_{xy} \frac{f(M_X)}{f(M_Y)} \qquad \text{and} \qquad \beta = \rho_{yz} \frac{f(M_Z)}{f(M_Y)}$$

and its minimum is given by

$$MSE(\hat{M}_{Ylr}) = \frac{1}{4\{f(M_Y)\}^2}\left\{\left(\frac{1}{m} - \frac{1}{N}\right) - \left(\frac{1}{m} - \frac{1}{n}\right)\rho_{xy}^2 - \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yz}^2\right\} \tag{1.3}$$

Singh et al (2006) defined a ratio type estimator of median under the two phase sampling design as

$$\hat{M}_S = \hat{M}_Y \left(\frac{\hat{M}'_X}{\hat{M}_X}\right)^{\alpha_1} \left(\frac{M_Z}{\hat{M}'_Z}\right)^{\alpha_2} \left(\frac{M_Z}{\hat{M}_Z}\right)^{\alpha_3} \tag{1.4}$$

where $\alpha_i; i = 1, 2, 3$ are constants.

Up to first order of approximation, they minimized MSE ($\hat{M}_S$) for:

$$\alpha_1 = \frac{M_X f(M_X)}{M_Y f(M_Y)}\left(\frac{\rho_{xz}\rho_{yz} - \rho_{xy}}{\rho_{xz}^2 - 1}\right),$$

$$\alpha_2 = \frac{M_Z f(M_Z)\rho_{xz}}{M_Y f(M_Y)}\left(\frac{\rho_{xz}\rho_{yz} - \rho_{xy}}{\rho_{xz}^2 - 1}\right)$$

and
$$\alpha_3 = \frac{M_Z f(M_Z)}{M_Y f(M_Y)} \left(\frac{\rho_{xz}\rho_{xy} - \rho_{yz}}{\rho_{xz}^2 - 1} \right)$$

$$MSE\left(\hat{M}_S\right)_{\min} = \frac{1}{4\{f(M_Y)\}^2}\left[\left(\frac{1}{m}-\frac{1}{N}\right)-\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{yz}^2-\left(\frac{1}{m}-\frac{1}{n}\right)R_{y.xz}^2\right] \qquad (1.5)$$

where
$$R_{y.xz}^2 = \frac{\rho_{xy}^2 + \rho_{yz}^2 - 2\rho_{xy}\rho_{yz}\rho_{xz}}{1-\rho_{xz}^2}$$

Using the knowledge of range R_Z of variable Z along with its population median M_Z, Gupta et al (2008) defined the estimator of population median M_Y under the same sampling design considered by Singh et al (2006) as

$$\hat{M}_P = \hat{M}_Y \left(\frac{\hat{M}_X'}{\hat{M}_X} \right)^{\gamma_1} \left(\frac{M_Z + R_Z}{\hat{M}_Z' + R_Z} \right)^{\gamma_2} \left(\frac{M_Z + R_Z}{\hat{M}_Z + R_Z} \right)^{\gamma_3} \qquad (1.6)$$

where $\gamma_i (i = 1, 2, 3)$ are constants.

Up to the first order of approximation, they minimized MSE of $\hat{M}_P$ for

$$\gamma_1 = \frac{M_X f(M_X)}{M_Y f(M_Y)} \left(\frac{\rho_{xz}\rho_{yz} - \rho_{xy}}{\rho_{xz}^2 - 1} \right),$$

$$\gamma_2 = \frac{1}{\left(\dfrac{M_Z}{M_Z + R_Z} \right)} \frac{M_Z f(M_Z)\rho_{xz}}{M_Y f(M_Y)} \left(\frac{\rho_{xz}\rho_{yz} - \rho_{xy}}{\rho_{xz}^2 - 1} \right)$$

and,
$$\gamma_3 = \frac{1}{\left(\dfrac{M_Z}{M_Z + R_Z} \right)} \frac{M_Z f(M_Z)}{M_Y f(M_Y)} \left(\frac{\rho_{xz}\rho_{xy} - \rho_{yz}}{\rho_{xz}^2 - 1} \right)$$

The bias and MSE of optimum estimator of $\hat{M}_P$ obtained by them are:

$$Bias(\hat{M}_P) \cong \frac{M_Y}{8\{M_Y f(M_Y)\}^2 (1-\rho_{xz}^2)^2}$$

$$\times \left[\left(\frac{1}{m} - \frac{1}{n} \right) \left\{ \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right)^2 - 2\rho_{xy} \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right) \left(1-\rho_{xz}^2 \right) \right. \right.$$

$$\left. + \frac{M_Y f(M_Y)}{M_X f(M_X)} \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right) \left(1-\rho_{xz}^2 \right) \right\}$$

$$+ \left(\frac{1}{m} - \frac{1}{N} \right) \left\{ \left(\rho_{yz} - \rho_{xy}\rho_{xz} \right)^2 + 2\rho_{xz} \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right) \left(\rho_{yz} - \rho_{xy}\rho_{xz} \right) \right.$$

$$\left. + \frac{M_Y f(M_Y)}{M_Z f(M_Z)} \left(\frac{M_Z}{M_Z + R_Z} \right) \left(\rho_{yz} - \rho_{xy}\rho_{xz} \right) \left(1-\rho_{xz}^2 \right) \right\}$$

$$+ \left(\frac{1}{n} - \frac{1}{N} \right) \left\{ \rho_{xz}^2 \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right)^2 - 2\rho_{yz}\rho_{xz} \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right) \left(1-\rho_{xz}^2 \right) \right.$$

$$\left. \left. + \frac{M_Y f(M_Y)}{M_Z f(M_Z)} \left(\frac{M_Z}{M_Z + R_Z} \right) \rho_{xz} \left(\rho_{xy} - \rho_{yz}\rho_{xz} \right) \left(1-\rho_{xz}^2 \right) \right\} \right]$$

$$(1.7)$$

and

$$MSE\left(\hat{M}_P\right)_{\min} \cong \frac{1}{4\{f(M_Y)\}^2} \left[\left(\frac{1}{m} - \frac{1}{N} \right) - \left(\frac{1}{n} - \frac{1}{N} \right) \rho_{yz}^2 - \left(\frac{1}{m} - \frac{1}{n} \right) R_{y.xz}^2 \right] \qquad (1.8)$$

$$= \quad MSE\left(\hat{M}_S\right)_{\min}$$

Recently Jhajj et al (2014) defined an efficient family of estimators of median using the same information used by Singh et al (2006) under same sampling design as

$$\hat{M}_{YH} = \hat{M}_Y \left(\frac{M_Z}{\hat{M}_Z'} \right)^{\upsilon_1} \left(\frac{M_Z}{\hat{M}_Z} \right)^{\upsilon_2} e^{\left(\frac{\upsilon_3 \left(\hat{M}_X' - \hat{M}_X \right)}{\hat{M}_X' + \hat{M}_X} \right)}$$

$$(1.9)$$

here υ_1, υ_2 and υ_3 are constants.

Up to the first order of approximation, they minimized MSE of $\hat{M}_{YH}$ for

$$\upsilon_1 = \frac{M_Z f(M_Z) \rho_{xz}}{M_Y f(M_Y)} \left(\frac{\rho_{xz}\rho_{yz} - \rho_{xy}}{\rho_{xz}^2 - 1} \right), \qquad \upsilon_2 = \frac{M_Z f(M_Z)}{M_Y f(M_Y)} \left(\frac{\rho_{xy}\rho_{xz} - \rho_{yz}}{\rho_{xz}^2 - 1} \right)$$

and

$$\upsilon_3 = \frac{2M_X f(M_X)}{M_Y f(M_Y)} \left(\frac{\rho_{yz}\rho_{xz} - \rho_{xy}}{\rho_{xz}^2 - 1} \right)$$

The bias and MSE of optimum estimator of $\hat{M}_{YH}$ obtained by them are:

$$Bias(\hat{M}_{YH}) \cong \frac{M_Y}{8\{M_Y f(M_Y)\}^2 (1-\rho_{xz}^2)^2}$$

$$\times \left[\left(\frac{1}{m}-\frac{1}{n}\right)\left\{ \left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)^2 - 2\rho_{xy}\left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)\left(1-\rho_{xz}^2\right) \right.\right.$$

$$\left. + \frac{M_Y f(M_Y)}{M_X f(M_X)}\left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)\left(1-\rho_{xz}^2\right) \right\}$$

$$+ \left(\frac{1}{m}-\frac{1}{N}\right)\left\{ \left(\rho_{yz}-\rho_{xy}\rho_{xz}\right)^2 + 2\rho_{xz}\left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)\left(\rho_{yz}-\rho_{xy}\rho_{xz}\right) \right.$$

$$\left. - 2\rho_{yz}\left(\rho_{yz}-\rho_{xy}\rho_{xz}\right)\left(1-\rho_{xz}^2\right) + \frac{M_Y f(M_Y)}{M_Z f(M_Z)}\left(\rho_{yz}-\rho_{xy}\rho_{xz}\right)\left(1-\rho_{xz}^2\right) \right\}$$

$$+ \left(\frac{1}{n}-\frac{1}{N}\right)\left\{ \rho_{xz}^2\left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)^2 - 2\rho_{yz}\rho_{xz}\left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)\left(1-\rho_{xz}^2\right) \right.$$

$$\left.\left. + \frac{M_Y f(M_Y)}{M_Z f(M_Z)}\rho_{xz}\left(\rho_{xy}-\rho_{yz}\rho_{xz}\right)\left(1-\rho_{xz}^2\right) \right\} \right]$$

$$(1.10)$$

and

$$MSE\left(\hat{M}_{YH}\right)_{\min} \cong \frac{1}{4\{f(M_Y)\}^2}\left[\left(\frac{1}{m}-\frac{1}{N}\right) - \left(\frac{1}{n}-\frac{1}{N}\right)\rho_{yz}^2 - \left(\frac{1}{m}-\frac{1}{n}\right)R_{y.xz}^2 \right] \quad (1.11)$$

$$= MSE\left(\hat{M}_S\right)_{\min}$$

II.   PROPOSED ESTIMATOR AND ITS RESULTS

Assuming that M_Z is known for a trivariate distribution of variables X,Y and Z, in which variables Y and Z are highly correlated with variable X and correlated among themselves through X only. Then we proposed a family of estimators of population median M_Y under two phase sampling design described in Section 1 as

$$\hat{M}_{YHL\theta} = \left[\hat{M}_Y + \theta\left(\hat{M}'_Y - \hat{M}_Y\right)\right] \left[\frac{\hat{M}_X}{\hat{M}'_X}\right]^{\lambda_1} \left[\frac{M_Z}{\hat{M}'_Z}\right]^{\lambda_2} \exp\left[2\lambda_3\left(\frac{\hat{M}_Z - \hat{M}'_Z}{\hat{M}_Z + \hat{M}'_Z}\right)\right]$$

$$(2.1)$$

where λ_1 , λ_2 , λ_3 are constants and $\theta > 0$.

To find the bias and MSE of the estimator $\hat{M}_{YHL\theta}$, we define

$$\varepsilon_0 = \frac{\hat{M}_Y}{M_Y} - 1 \ , \quad \varepsilon_1 = \frac{\hat{M}'_Y}{M_Y} - 1, \quad \varepsilon_2 = \frac{\hat{M}_X}{M_X} - 1,$$

$$\varepsilon_3 = \frac{\hat{M}'_X}{M_X} - 1, \quad \varepsilon_4 = \frac{\hat{M}_Z}{M_Z} - 1, \quad \varepsilon_5 = \frac{\hat{M}'_Z}{M_Z} - 1$$

such that $E(\varepsilon_i) \cong 0$ for $i = 0, 1, 2, 3, 4, 5$

Assuming $\left|\varepsilon_i\right| < 1 \;\forall i$, we expand $\hat{M}_{YHL\theta}$ in terms of $\varepsilon's$ and retain the terms up to second degree of $\varepsilon's$ in the derivation of bias and mean square error for obtaining the expressions up to first order of approximation

$$Bias(\hat{M}_{YHL\theta}) = \frac{M_Y}{4}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left\{\frac{\lambda_2(\lambda_2+1)}{2}\left(M_Z f(M_Z)\right)^{-2} - \lambda_2\rho_{yz}\frac{\left(M_Z f(M_Z)\right)^{-1}}{\left(M_Y f(M_Y)\right)}\right\}\right.$$

$$+\left(\frac{1}{m}-\frac{1}{n}\right)\left\{\frac{\lambda_1(\lambda_1-1)}{2}\left(M_X f(M_X)\right)^{-2} + \frac{\lambda_3(\lambda_3-1)}{2}\left(M_Z f(M_Z)\right)^{-2}\right.$$

$$\left.\left.+\lambda_1\lambda_3\rho_{xz}\frac{\left(M_X f(M_X)\right)^{-1}}{\left(M_Z f(M_Z)\right)} + (1-\theta)\left(\lambda_1\rho_{xy}\frac{\left(M_X f(M_X)\right)^{-1}}{\left(M_Y f(M_Y)\right)} + \lambda_3\rho_{yz}\frac{\left(M_Z f(M_Z)\right)^{-1}}{\left(M_Y f(M_Y)\right)}\right)\right\}\right]$$

$$(2.2)$$

$$MSE(\hat{M}_{YHL\theta}) = \frac{M_Y^2}{4}\left[\left(\frac{1}{m}-\frac{1}{N}\right)\left(M_Y f(M_Y)\right)^{-2} + \left(\frac{1}{n}-\frac{1}{N}\right)\left\{\lambda_3^2\left(M_Z f(M_Z)\right)^{-2} - 2\lambda_3\rho_{yz}\frac{\left(M_Z f(M_Z)\right)^{-1}}{\left(M_Y f(M_Y)\right)}\right\}\right.$$

$$+\left(\frac{1}{m}-\frac{1}{n}\right)\left\{\left(\theta^2-2\theta\right)\left(M_Y f(M_Y)\right)^{-2} + \lambda_1^2\left(M_X f(M_X)\right)^{-2} + \lambda_2^2\left(M_Z f(M_Z)\right)^{-2}\right.$$

$$\left.\left.+2(1-\theta)\left(\lambda_1\rho_{xy}\frac{\left(M_X f(M_X)\right)^{-1}}{\left(M_Y f(M_Y)\right)} + \lambda_2\rho_{yz}\frac{\left(M_Z f(M_Z)\right)^{-1}}{\left(M_Y f(M_Y)\right)}\right) + 2\lambda_1\lambda_2\rho_{xz}\frac{\left(M_X f(M_X)\right)^{-1}}{\left(M_Z f(M_Z)\right)}\right\}\right]$$

$$(2.3)$$

For any fixed value of θ, $Bias\left(\hat{M}_{YHL\theta}\right)$ and $MSE\left(\hat{M}_{YHL\theta}\right)$ are minimized for

$$\lambda_1 = (1-\theta)\left(\frac{\rho_{yz}\rho_{xz}-\rho_{xy}}{1-\rho_{xz}^2}\right)\frac{M_X f(M_X)}{M_Y f(M_Y)}, \qquad \lambda_2 = \rho_{yz}\frac{M_Z f(M_Z)}{M_Y f(M_Y)}$$

and

$$\lambda_3 = (1-\theta)\left(\frac{\rho_{xy}\rho_{xz}-\rho_{yz}}{1-\rho_{xz}^2}\right)\frac{M_Z f(M_Z)}{M_Y f(M_Y)}$$

and their corresponding minimum values are given by

$$Bias(\hat{M}_{YHL\theta}) = \frac{1}{8M_Y\left(f(M_Y)\right)^2\left(1-\rho_{xz}^2\right)^2}\left[\left(\frac{1}{m}-\frac{1}{n}\right)\left\{(1-\theta)^2\left[\left(\rho_{yz}\rho_{xz}-\rho_{xy}\right)^2 - 2\rho_{xy}\left(\rho_{yz}\rho_{xz}-\rho_{xy}\right)\left(1-\rho_{xz}^2\right)\right]\right.\right.$$

$$+(1-\theta)\left[2\rho_{yz}^2\left(1-\rho_{xz}^2\right)^2 - \left(\frac{M_Y f(M_Y)}{M_X f(M_X)} - 2\rho_{xz}\rho_{yz}\right)\left(\rho_{yz}\rho_{xz}-\rho_{xy}\right)\left(1-\rho_{xz}^2\right)\right] + \rho_{yz}\left(1-\rho_{xz}^2\right)^2\left(\rho_{yz}-\frac{M_Y f(M_Y)}{M_Z f(M_Z)}\right)\right\}$$

$$\left.+\left(\frac{1}{n}-\frac{1}{N}\right)(1-\theta)\left\{(1-\theta)\left(\rho_{xy}\rho_{xz}-\rho_{yz}\right)^2 + \left(\frac{M_Y f(M_Y)}{M_Z f(M_Z)} - 2\rho_{yz}\right)\left(\rho_{xy}\rho_{xz}-\rho_{yz}\right)\left(1-\rho_{xz}^2\right)\right\}\right]$$

$$(2.4)$$

$$MSE(\hat{M}_{YHL\theta}) = \frac{1}{4\left(f(M_Y)\right)^2}\left[\left(\frac{1}{m}-\frac{1}{N}\right) - \left(\frac{1}{n}-\frac{1}{N}\right)\rho_{yz}^2 - \left(\frac{1}{m}-\frac{1}{n}\right)\left\{(1-\theta)^2 R_{y.xz}^2 - \left(\theta^2-2\theta\right)\right\}\right]$$

$$(2.5)$$

III. COMPARISON

We compare the proposed family of estimators with respect to its mean square error with Singh et al (2006) estimator as well as linear regression type estimator $\hat{M}_{Ylr}$ because it is generally considered best in literature under the same conditions.

Using the expressions (1.3) and (2.5), we have

$$MSE\left(\hat{M}_{Ylr}\right) - MSE\left(\hat{M}_{YHL\theta}\right) = \frac{1}{4\left(f\left(M_Y\right)\right)^2}\left(\frac{1}{m} - \frac{1}{n}\right)\left\{(1-\theta)^2\left(R_{y.xz}^2 - 1\right) + 1 - \rho_{xy}^2\right\}$$

$$> 0 \quad for \quad 1 - \sqrt{K} < \theta < 1 + \sqrt{K} \tag{3.1}$$

where $\quad K = \dfrac{1 - \rho_{xy}^2}{1 - R_{y.xz}^2}$

Using the expressions (1.5) and (2.5), we have

$$MSE\left(\hat{M}_S\right)_{\min} - MSE\left(\hat{M}_{YHL\theta}\right) = \frac{1}{4\left(f\left(M_Y\right)\right)^2}\left(\frac{1}{m} - \frac{1}{n}\right)\left(\theta^2 - 2\theta\right)\left(R_{y.xz}^2 - 1\right)$$

$$> 0 \quad for \quad 0 < \theta < 2 \tag{3.2}$$

From (3.1) and (3.2), we note that the proposed family of estimators is always better than the linear regression type estimator as well as Singh et al (2006) estimator for $0 < \theta < 2$.

Note: From the expressions of the biases obtained, we noted that no concrete conclusion can be obtained theoretically by comparing proposed estimator with the others with respect to their biases. So, the biases are compared numerically in Section 4.

IV. NUMERICAL ILLUSTRATION

For illustration of results numerically and graphically, we take the following data

Data 1 (Source: Singh, 2003). Y: the number of fish caught by marine recreational fishermen in 1995; X: The number of fish caught by marine recreational fishermen in 1994; Z: The number of fish caught by marine recreational fishermen in 1993.

N = 69,	$\rho_{xy} = 0.1505$,	$M_Y = 2068$,	$f\left(M_Y\right) = 0.00014$
n = 24,	$\rho_{yz} = 0.3166$,	$M_X = 2011$,	$f\left(M_X\right) = 0.00014$
m = 17,	$\rho_{xz} = 0.1431$,	$M_Z = 2307$,	$f\left(M_Z\right) = 0.00013$

Data 2 (Source: Aezel and Sounderpandian, 2004). Y: The U.S. exports to Singapore in billions of Singapore dollars; X: The money supply figures in billions of Singapore dollars; Z: The local supply in U.S. dollors.

N = 67,	$\rho_{xy} = 0.6624$,	$M_Y = 4.8$,	$f\left(M_Y\right) = 0.0763$
n = 23,	$\rho_{yz} = 0.8624$,	$M_X = 7.0$,	$f\left(M_X\right) = 0.0526$
m = 15,	$\rho_{xz} = 0.7592$,	$M_Z = 151$,	$f\left(M_Z\right) = 0.0024$

Data 3. (Source: MFA, 2004). Y: District-wise tomato production (tones) in 2003; X: District-wise tomato production (tones) in 2002; Z: District-wise tomato production (tones) in 2001.

N = 97,	$\rho_{xy} = 0.2096$,	$M_Y = 1242$,	$f\left(M_Y\right) = 0.00021$
n = 46,	$\rho_{yz} = 0.1233$,	$M_X = 1233$,	$f\left(M_X\right) = 0.00022$
m = 33,	$\rho_{xz} = 0.1496$,	$M_Z = 1207$,	$f\left(M_Z\right) = 0.00023$

For Data 1

Table 4.1 : **Bias and relative efficiency of estimators**

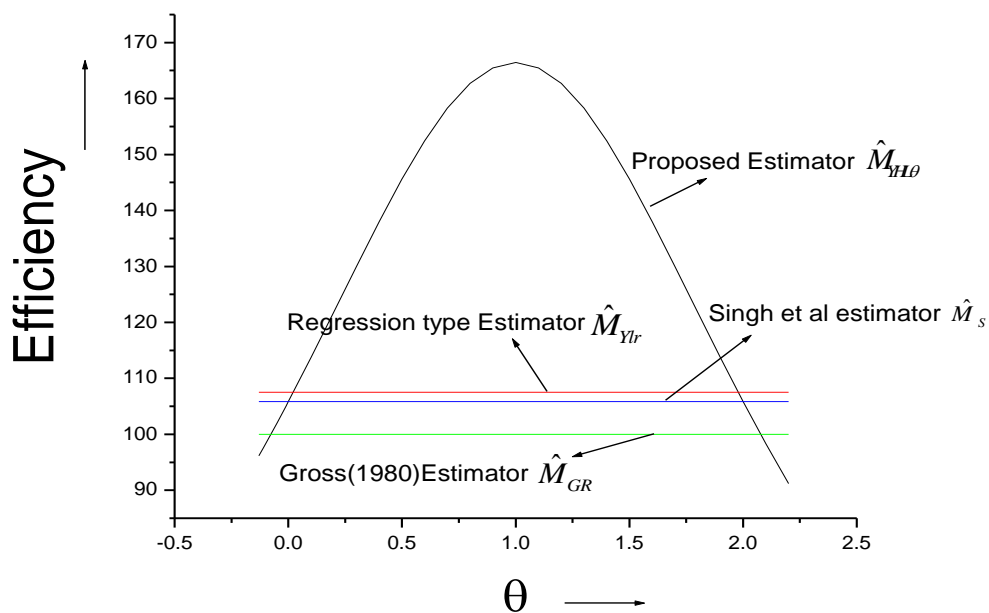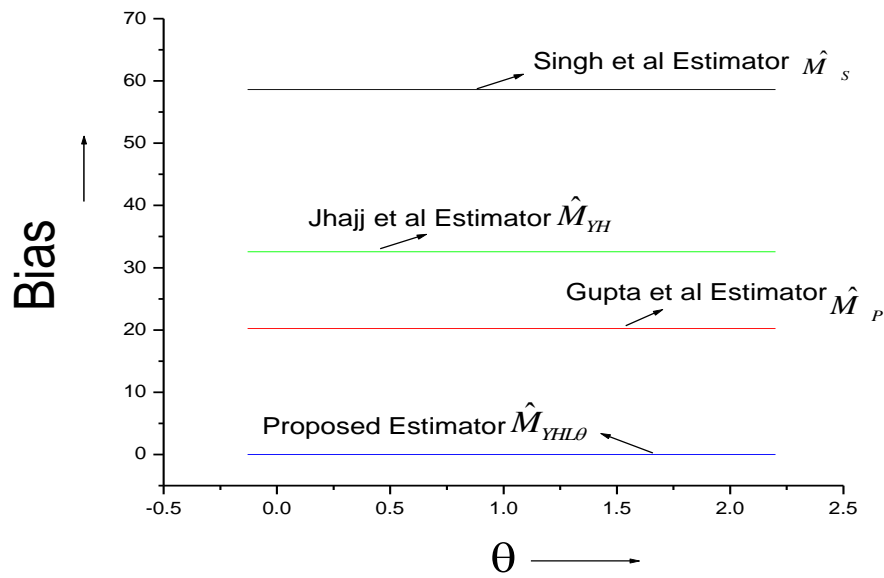| θ	$\lvert Bias \rvert$				Relative Efficiency of			
	$\hat{M}_S$	$\hat{M}_P$	$\hat{M}_{YH}$	$\hat{M}_{YHL\theta}$	$\hat{M}_{GR}$	$\hat{M}_{Ylr}$	$\hat{M}_S$	$\hat{M}_{YHL\theta}$
-0.13	58.62	20.22	32.57	2.65E-07	100	107.5	105.83	96.134
-0.08	58.62	20.22	32.57	2.27E-07	100	107.5	105.83	99.781
-0.03	58.62	20.22	32.57	1.91E-07	100	107.5	105.83	103.53
0	58.62	20.22	32.57	1.69E-07	100	107.5	105.83	105.83
0.3	58.62	20.22	32.57	1.8E-08	100	107.5	105.83	129.96
0.6	58.62	20.22	32.57	1.8E-08	100	107.5	105.83	152.46
0.9	58.62	20.22	32.57	2.6E-07	100	107.5	105.83	165.48
1.2	58.62	20.22	32.57	3.1E-07	100	107.5	105.83	162.71
1.5	58.62	20.22	32.57	3.1E-07	100	107.5	1053	145.59
1.8	58.62	20.22	32.57	2.7E-07	100	107.5	105.83	121.79
2.1	58.62	20.22	32.57	1.8E-07	100	107.5	105.83	98.309
2.4	58.62	20.22	32.57	5.1E-08	100	107.5	105.83	78.415

Figure 4.1

Figure 4.2

For Data 2

Table 4.2 :   **Bias and relative efficiency of estimators**

θ	$|Bias|$				Relative Efficiency of Estimators			
	$\hat{M}_S$	$\hat{M}_P$	$\hat{M}_{YH}$	$\hat{M}_{YHL\theta}$	$\hat{M}_{GR}$	$\hat{M}_{Ylr}$	$\hat{M}_S$	$\hat{M}_{YHL\theta}$
-1.7	0.37	0.28	0.026	0.023417	100	254.55	286.9	93.413
-1.2	0.37	0.28	0.026	0.017163	100	254.55	286.9	126.69
-0.7	0.37	0.28	0.026	0.011761	100	254.55	286.9	176.84
-0.2	0.37	0.28	0.026	0.00721	100	254.55	286.9	250.6
0	0.37	0.28	0.026	0.005628	100	254.55	286.9	286.91
0.4	0.37	0.28	0.026	0.002873	100	254.55	286.9	363.53
0.8	0.37	0.28	0.026	0.000662	100	254.55	286.9	419.55
1.2	0.37	0.28	0.026	0.001	100	254.55	286.9	419.55
1.6	0.37	0.28	0.026	0.00212	100	254.55	286.9	363.53
2	0.37	0.28	0.026	0.0027	100	254.55	286.9	286.91
2.4	0.37	0.28	0.026	0.00273	100	254.55	286.9	217.99
2.8	0.37	0.28	0.026	0.00222	100	254.55	286.9	165.11
3.2	0.37	0.28	0.026	0.00116	100	254.55	286.9	126.69
3.6	0.37	0.28	0.026	0.000446	100	254.55	286.9	99.041
4	0.37	0.28	0.026	0.002594	100	254.55	286.9	78.939

Figure 4.3
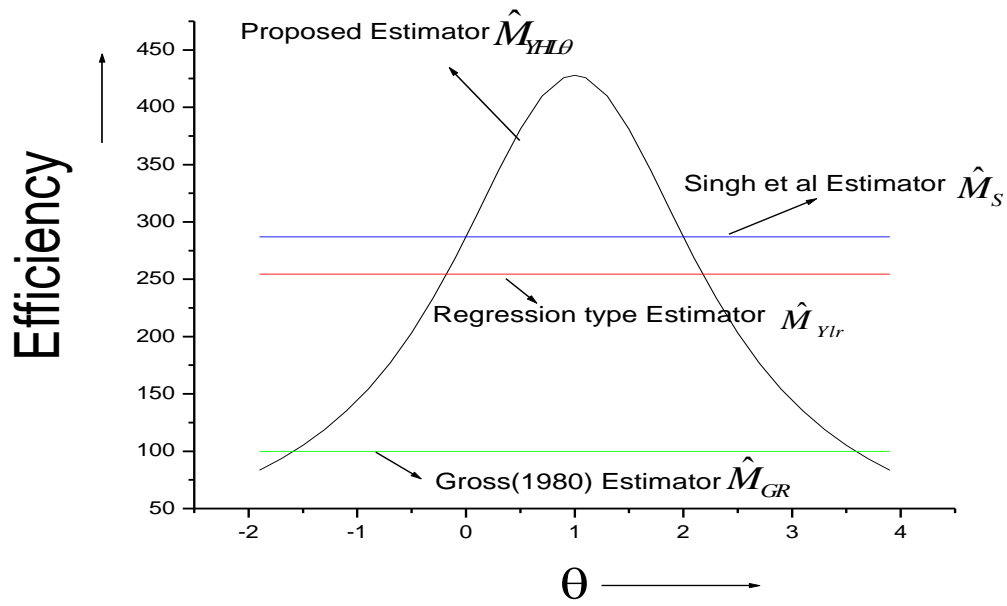
Figure 4.4
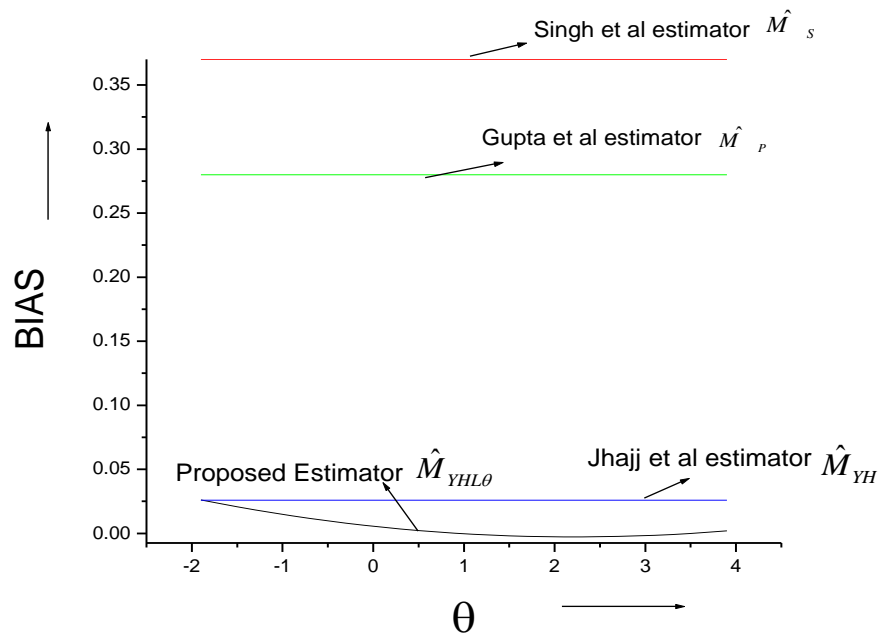
For Data 3

Table 4.3 :   **Bias and relative efficiency of estimators**

| θ	$\lvert Bias \rvert$				Relative Efficiency of Estimators			
	$\hat{M}_S$	$\hat{M}_P$	$\hat{M}_{YH}$	$\hat{M}_{YHL\theta}$	$\hat{M}_{GR}$	$\hat{M}_{Ylr}$	$\hat{M}_S$	$\hat{M}_{YHL\theta}$
-0.1	8.06	3.63	4.75	2.23E-07	100	102.8	103.7	95.277
0	8.06	3.63	4.75	1.77E-07	100	102.8	103.7	103.69
0.2	8.06	3.63	4.75	9.48E-08	100	102.8	103.7	122.2
0.4	8.06	3.63	4.75	2.44E-08	100	102.8	103.7	141.89
0.6	8.06	3.63	4.75	3.4E-08	100	102.8	103.7	160.36
0.8	8.06	3.63	4.75	8E-08	100	102.8	103.7	173.93
1	8.06	3.63	4.75	1.1E-07	100	102.8	103.7	178.99
1.2	8.06	3.63	4.75	1.4E-07	100	102.8	103.7	173.93
1.4	8.06	3.63	4.75	1.5E-07	100	102.8	103.7	160.36
1.6	8.06	3.63	4.75	1.5E-07	100	102.8	103.7	141.89
1.8	8.06	3.63	4.75	1.3E-07	100	102.8	103.7	122.2
2	8.06	3.63	4.75	1.1E-07	100	102.8	103.7	103.69
2.2	8.06	3.63	4.75	-6.8E-08	100	102.8	103.7	87.498

Figure 4.5
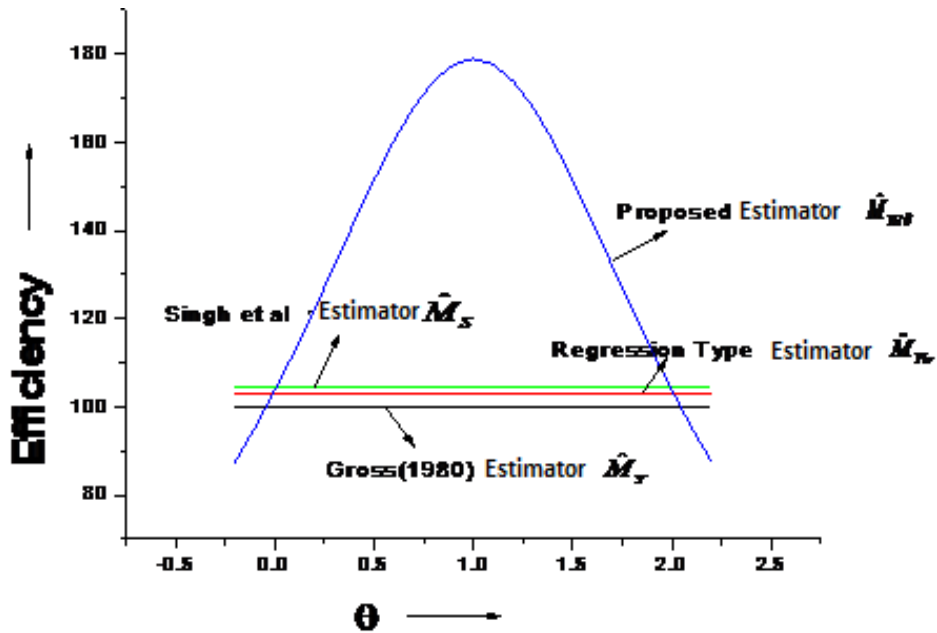
Figure 4.6
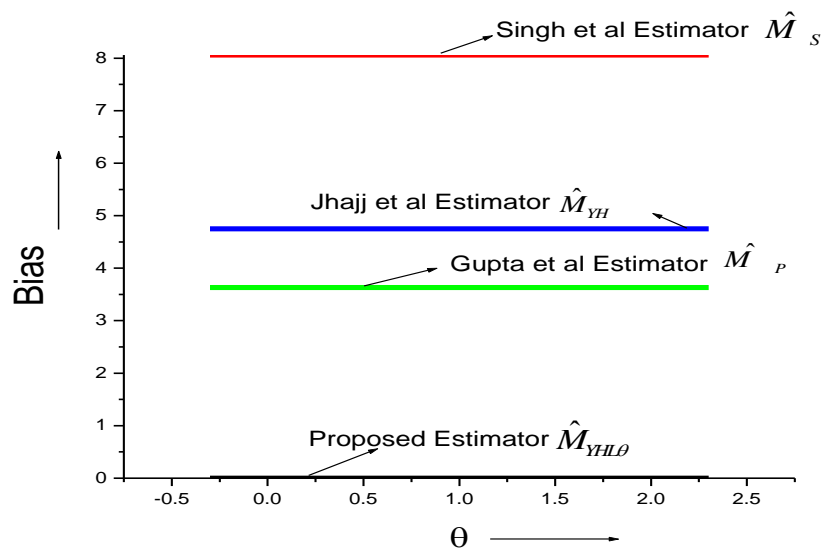
From tables (4.1), (4.2) and (4.3), we note that proposed family of estimators have smaller mean square error than Gross(1980) estimator, linear regression type estimator $\hat{M}_{Ylr}$, Singh et al (2006) estimator, Gupta et al (2008) estimator and Jhajj et al estimator (2014) as well as have smaller bias than Singh et al (2006) estimator, Gupta et al (2008) estimator and Jhajj et al estimator (2014) in all the three populations considered corresponding to $\theta \in (0,2)$. The graphical representation also supports the numerical results.

V. CONCLUSION

The range of variation of θ involved in the proposed family of estimators of population median under two-phase sampling design has been obtained theoretically under which its estimators are efficient than the existing ones. It has been found theoretically that proposed family of estimators is efficient than the linear regression type estimator $\hat{M}_{Ylr}$, Singh et al (2006) estimator Gupta et al (2008) estimator and Jhajj et al estimator (2014) for $\theta \in (0,2)$. Numerical results for all the three populations also show that optimum estimator of proposed family having smaller bias is efficient than them. Graphical representation also gives the same type of interpretation. Hence, we conclude that better estimators can be developed from the proposed family by choosing suitable values of $\theta \in (0,2)$. So, it is strongly recommended that proposed family of estimators should be used for getting the accurate value of population median.

REFERENCES

[1]. Azel A.D., Sounderpandian J. (2004).*Complete Business Statistics*. 5th ed. New York: McGraw Hill,
[2]. Gross S.T. (1980). Median estimation in sample *Proc. Surv. Res. Meth. Sect. Amer. Statist Ass.,*181-184.
[3]. Gupta S., Shabbir J. and Ahmad S.(2008). Estimation of median in two- phase sampling using two auxiliary variables. Communication in Statistics-Theory and Methods, 37(11), 1815- 1822.
[4]. Jhajj H.S.and Harpreet Kaur (2013),Generalized estimators of population median using auxiliary information, IJSER, 11(4), 378-389.
[5]. Jhajj H.S., Harpreet Kaur and Walia Gurjeet(2013), Efficient family of ratio-product type estimators of median, MASA, Vol 9, 2014, 277-282.
[6]. Jhajj H.S., Harpreet Kaur and Jhajj Puneet (2014), Efficient family of estimators of median using two phase sampling design, Communication in Statistics-Theory and Methods, Accepted.
[7]. Kuk Y.C.A. and Mak T.K. (1989). Median estimation in the presence of auxiliary information. *J.R Statist.Soc.* B, (2), 261-269.
[8]. MFA (2004). *Crops Area Production. Ministery of food and Agriculture*. Islamabad: Pakistan.
Singh S. (2003).*Advanced Sampling Theory with Applications*. Vol. 1 & 2. London Kluwer Academic Publishers.
[9]. Singh S. , Joarder A. H. and Tracy D. S.(2001), Median estimation using double sampling. *Austral. N.J. Statist.* 43(1) , 33-46.
[10]. Singh S., Singh H.P and Upadhyaya L.N. (2006). Chain ratio and regression type estimators for median estimation in survey sampling. *Statist. Pap*. 48, 23-46.