Pak. J. Statist.
2023 Vol. 39(3), 361-375

FAMILY OF NEW ROBUST RATIO-TYPE ESTIMATORS UNDER NON-NORMALITY FOLLOWING SYMMETRIC AND SKEWED DISTRIBUTIONS

Azaz Ahmed[1,2§] and **Muhammad Hanif**[2]
[1] Govt. Islamia College Railway Road, Lahore, Pakistan
Email: azazpu17@gmail.com
[2] National College of Business Administration & Economics
Lahore, Pakistan. Email: drhanif@ncbae.edu.pk
[§] Corresponding author Email: azazpu17@gmail.com

ABSTRACT

In case of wild observations or non-normality, using the usual ratio type estimator will be the worst. Using modified maximum likelihood estimators (MMLEs) Oral and Kadilar (2011) made robust ratio-type estimators to deal with the issue. A generalized least squares estimator (GLSE) based on order statistics can be more appropriate. In this paper, a family of proposed ratio-type estimators (PRTEs) using GLSE is advised. The situation is focused especially, if the errors follow the non-normal (symmetric or skewed) distribution. The performance of PRTEs is evaluated via simulation by the mean square error (MSE) and the relative efficiencies (RE).

KEYWORDS

Simple random sampling**,** Ratio type estimators, generalized least square estimators, Laplace distribution, Weibull distribution, robust estimators.

1. INTRODUCTION

In survey sampling, estimation of the characteristics of the population using sample information has been a demanding task for survey statisticians. The objective of sample selection is to attain fairly precise results regarding population parameters based on the sample. The simplest mean estimator of the population is the simple random sample (SRS) mean when no additional information is used. It is often the case that an auxiliary variable z positively related to the main variable of the study y is also available. In such circumstances, one of the most customary methods of estimation is the classical ratio method. The estimator of the population mean in this method is given as,

$$\bar{y}_r = \frac{\bar{y}}{\bar{z}}\bar{Z}\,, \qquad (1.1)$$

where $\bar{y}$ and $\bar{z}$ are the sample means of the study variable y and the auxiliary variable z, and $\bar{Z}$ is the population mean of the auxiliary variable z. The $\bar{y}_r$ is widely applied

© 2023 Pakistan Journal of Statistics

for estimating the population mean $\bar{Y}=\sum_{l=1}^{M} y_l / M$ when $\bar{Z}$ is known. The MSE of $\bar{y}_r$ is given by

$$\text{MSE}(\bar{y}_r) \cong \text{var}(\bar{y}) - 2R\,\text{cov}(\bar{y},\bar{z}) + R^2\,\text{var}(\bar{z}), \tag{1.2}$$

where $\text{var}(\bar{y})=\frac{1-f}{m}S_y^2$; $\text{var}(\bar{z})=\frac{1-f}{m}S_z^2$; $\text{cov}(\bar{y},\bar{z})=\frac{1-f}{m}S_{zy}$; $f=m/M$; m and M are the sample and population sizes, S_z^2 and S_y^2 are variances of the population of z and y, respectively; S_{zy} is the covariance of the population between z and y; $R=\bar{Y}/\bar{Z}$ is the population ratio. By utilizing available information about the population parameters like V_z (coefficient of variation), $\beta_2(z)$ (kurtosis) and ρ (population coefficient of correlation between z and y) into (1.1), efficiency may be enhanced (Singh & Tailor, 2003). Merging the estimators to form a general class specified by Ray and Singh (1981), Kadilar and Cingi (2004) recommended the subsequent different ratio-type estimators known as Kadilar-Cingi estimators (KCEs) in the SRS context:

$$\bar{y}_{KCj} = \frac{\bar{y}+\hat{\beta}_L(\bar{Z}-\bar{z})}{(\lambda_j \bar{z}+\gamma_j)}(\lambda_j \bar{Z}+\gamma_j), \text{ for } j=1,2,...,5 \tag{1.3}$$

where $\hat{\beta}_L = s_{zy}/s_z^2$ is the estimator of regression coefficient that is estimated by least square (LS) estimation method; $\lambda_1=1$, $\gamma_1=0$; $\lambda_2=1$, $\gamma_2=V_z$; $\lambda_3=1$, $\gamma_3=\beta_2(z)$; $\lambda_4=\beta_2(z)$, $\gamma_4=V_z$; $\lambda_5=V_z$, $\gamma_5=\beta_2(z)$. The MSE of (1.3) is obtained by applying the Taylor series approximation up to the first order as,

$$\text{MSE}(\bar{y}_{KCj}) \cong (1-f)/m(R_{KCj}^2 S_z^2 + 2\beta_L R_{KCj} S_z^2 + \beta_L^2 S_z^2 - 2R_{KCj} S_{zy} - 2\beta_L S_{zy} + S_y^2),$$

where $\beta_L = S_{zy}/S_z^2$ and $R_{KCj}=\frac{\lambda_j \bar{Y}}{\lambda_j \bar{Z}+\gamma_j}$ for $j=1,2,...,5.$

Although $\bar{y}_{KCj}$ are beneficial in estimating $\bar{Y}$ but they are somewhat responsive to outliers. However, it is essential to adjust the ratio estimators to do not a response to outliers or non-normality. In literature the ratio estimators are robustified in different attempts regarding the existence of outliers; see (Farrel & Barrera 2006; Oral & Oral 2011; Kumar & Chhaparwal 2017). Kadilar et al. (2007) robustified the KCEs by using Huber's M estimator. It was concluded that M estimation can enhance the efficiency of (1.3). Oral and Kadilar (2011) improved the efficiencies of KCEs using MMLE methodology in KCEs. Comparing MMLE to M estimators, Islam and Tiku 2004 proved that the MMLEs can provide better results when the errors follow a long-tailed symmetric (LTS) family. Moreover, Azaz et al. (2019) highlighted several problems with MMLE's weight function and suggested using the Generalized Least Squares Estimation (GLSE) instead of MMLE for the robustification process in SRS.

In this study, applying GLSE to an ordered sample by Lloyd (1952), the mean estimation of a non-normal population with auxiliary information is emphasized. To produce robust estimates the GLSE can be taken as an alternative to MMLEs. Because MMLEs can be proved ineffective if the maximum likelihood estimators (MLEs) are in explicit form, as in the case of Laplace distribution. In skewed distribution e.g. Weibull distribution, MMLEs do not exist when the shape parameter is less than 1. Whereas, Fisher information can only be defined if the shape parameter of Weibull distribution is greater than 2; have a look (Islam et. al, 2001). So, GLSE is introduced to deal with these issues.

Integrating the GLSE into (1.3) a family of PRTEs is formed not only to deal with non-normality but also to improve efficiency. Moreover, PRTEs are robust in the presence of outliers.

2. PROPOSED A FAMILY OF RATIO-TYPE ESTIMATORS UNDER NON-NORMALITY OF ERRORS

In the linear regression model

$$y_l = \beta z_l + e_l \,, 1 < l < m \tag{2.1}$$

let error term (e_l) is from a family of symmetric or skewed distributions, let $e_1, e_2, ..., e_m$ be a SRS from the parent distribution. Let $e_{(1)} \le e_{(2)} \le, ..., \le e_{(m)}$ be the order statistics of the random sample $e_1, e_2, ..., e_m$. Let $u_{(l)} = \frac{e_{(l)}}{\sigma}$ or $u_{(l)} = \frac{y_{[l]} - \beta z_{[l]}}{\sigma}$ be the standardized variate, $(y_{[l]}, z_{[l]})$ may be called concomitants of $u_{(l)}$, $1 < l < m$. Let $E(u_{(l)}) = \alpha_l$, $\mathrm{var}(u_{(l)}) = \omega_{ll}$ and $\mathrm{cov}(z_{(l)}, z_{(h)}) = \omega_{lh}$ for $l = 1, 2, ..., m$, $h = 1, 2, ..., m$. Further, let us denote $\mathrm{y} = (y_{(1)}, y_{(2)}, ..., y_{(m)})'$, $\alpha = (\alpha_{(1)}, \alpha_{(2)}, ..., \alpha_{(m)})'$, $\mathrm{z} = (z_{(1)}, z_{(2)}, ..., z_{(m)})'$ and $\Omega = [\omega_{lh}]$ for $l = 1, 2, ..., m$, $h = 1, 2, ..., m$. The best linear unbiased estimator (BLUE) of β and σ can be derived by minimizing the quadratic form as given below

$$Q(\Theta) = (\mathrm{y} - \mathrm{A}\Theta)'\Omega^{-1}(\mathrm{y} - \mathrm{A}\Theta) \text{, where } \mathrm{A} = [\mathrm{z} \vdots \alpha] \text{ and } \Theta = (\beta, \sigma)^T \tag{2.2}$$

minimizing the (2.2) the BLUE of β and σ are derived as under

$$\hat{\beta}_{GL} = \left[\frac{\alpha'\Omega^{-1}\alpha \mathrm{z}'\Omega^{-1} - \alpha'\Omega^{-1}\mathrm{z}\alpha'\Omega^{-1}}{\alpha'\Omega^{-1}\alpha \mathrm{z}'\Omega^{-1}\mathrm{z} - (\alpha'\Omega^{-1}\mathrm{z})^2} \right] \mathrm{y} \text{ and } \hat{\sigma}_{GL} = \left[\frac{\mathrm{z}'\Omega^{-1}\mathrm{z}\alpha'\Omega^{-1} - \mathrm{z}'\Omega^{-1}\alpha \mathrm{z}'\Omega^{-1}}{\alpha'\Omega^{-1}\alpha \mathrm{z}'\Omega^{-1}\mathrm{z} - (\alpha'\Omega^{-1}\mathrm{z})^2} \right] \mathrm{y} \tag{2.3}$$

The variance and covariance of these estimators may be found as,

$$\mathrm{var}(\hat{\beta}_{GL}) = \left[\frac{\alpha'\Omega^{-1}\alpha}{\alpha'\Omega^{-1}\alpha \, \mathrm{z}'\Omega^{-1}\mathrm{z} - (\alpha'\Omega^{-1}\mathrm{z})^2} \right] \sigma^2 ,$$

$$\text{var}(\hat{\sigma}_{GL}) = \left[\frac{z'\Omega^{-1}z}{\alpha'\Omega^{-1}\alpha\, z'\Omega^{-1}z - (\alpha'\Omega^{-1}z)^2}\right]\sigma^2,$$

and

$$\text{cov}(\hat{\beta}_{GL}, \hat{\sigma}_{GL}) = -\left[\frac{\alpha'\Omega^{-1}z'}{\alpha'\Omega^{-1}\alpha z'\Omega^{-1}z - (\alpha'\Omega^{-1}z)^2}\right]\sigma^2.$$

Consider the linear regression model (2.1), suppose that the error term (e_l) is from a family of Laplace distribution given by Farnoosh and Jafarpour (2005) as,

$$f(e): L(\theta,\sigma) = \frac{\theta}{2\sigma} Exp[-\theta|e|/\sigma]\text{, where } e = y - \beta z\text{,} \quad -\infty \le e \le \infty \qquad (2.4)$$

where σ is the scale parameter and θ is the sharply peaked parameter. The coefficient of kurtosis suggested by Farnoosh and Jafarpour (2005) for (2.4) is $E(\frac{e}{\sigma})^4 = 24/\theta^4$. The coefficient of kurtosis of (2.4) that we consider in this study is 24 for $\theta = 1$.

Let $u_{(l)} = \frac{y_{[l]} - \beta z_{[l]}}{\sigma}$ and $\theta = 1$, using the values of α and Ω tabulated by Govindarajulu (1966) for $L(1,1)$ and following Lloyd's method, we may calculate the values of the estimator $\hat{\beta}_{GL}$ from (2.3).

For the linear regression model (2.1) if the error term follows the LTS family

$$f(e): LTS(p,\sigma) = \frac{p}{\sigma\sqrt{k}\,\beta(1/2,\, p-1/2)}\left[1 + \frac{1}{k}\left(\frac{e}{\sigma}\right)^2\right]^{-p} \quad -\infty < e < \infty \qquad (2.5)$$

where the shape parameter is p, $k = 2p - 3$ and p should be greater than 2. The kurtosis of (2.5) is $\beta_2 = \frac{3}{(1-2/k)}$ for $k \to \infty$ then (2.5) follow to normal distribution. The expressions of likelihood equations for (2.5) are in the form of the complex functions $g(u_l) = u_l / \{1 + (1/k)u_l^2\}$, where $u_l = e_l/\sigma$, $1 < l < m$, and do not expressed in closed form. The robust MMLEs are attained as the likelihood equations are written in the form of the ordered variates $u_{(l)} = e_{(l)}/\sigma$, the functions $g(u_{(l)})$ are restored with their linear approximations and the resultant equations are evaluated for parameters. The solutions of the equations are explicit functions of the concomitant observations $(y_{[l]}, z_{[l]})$, $1 < l < m$ as,

$$\hat{\beta}_T = D + C\hat{\sigma} \text{ and } \hat{\sigma} = \frac{F + \sqrt{F^2 + 4mG}}{\sqrt{4m(m-2)}}, \qquad (2.6)$$

where

$$D=\sum_{l=1}^{m}\Delta_l z_{[l]}y_{[l]} \Big/ \sum_{l=1}^{m}\delta_l z_{[l]}^2,\ C=\sum_{l=1}^{m}\delta_l z_{[l]} \Big/ \sum_{l=1}^{m}\delta_l z_{[l]}^2, \tag{2.7}$$

$$G=(2p/k)\sum_{l=1}^{m}\Delta_l(y_{[l]}-Dz_{[l]}),\ F=(2p/k)\sum_{l=1}^{m}\delta_l(y_{[l]}-Dz_{[l]})^2,$$

$$\delta_l=(2/k)\alpha_l^3/[1+(1/k)\alpha_l^2]^2,\ \Delta_l=([1-(1/k)\alpha_l^2]/[1+(1/k)\alpha_l^2]^2.$$

The magnitudes of α and Ω are tabularized by Tiku and Kumra (1981) of the LTS family for $p\le 10$ Using of α and Ω, following Lloyd's method and MMLEs, we may calculate the values of estimators $\hat{\beta}_{GL}$ from (2.3) and $\hat{\beta}_T$ from (2.6), respectively.

Yet again for model (2.1), assume e_l tag on the family of skewed distributions, namely, Weibull distribution as

$$f(e):W(p,\sigma)=(p/\sigma)e^{p-1}\exp(-(e/\sigma)^p),\qquad 0<e<\infty \tag{2.8}$$

where p is the shape parameter. The (2.8) for $p<1$, $p=1$ and $p=2$ become reversed J-shaped, the exponential distribution and the Rayleigh distribution, respectively. Therefore, it is applied mostly in practical fields due to its elasticity for fitting data which follow the distributions away from symmetry. The behavior of skewness and kurtosis of (2.8) are given below.

p	1.5	2	2.5	3	4	6
$\alpha_1=\sqrt{\beta_1}$	1.064	0.631	0.358	0.168	-0.087	-0.158
β_2	4.365	3.246	2.858	2.705	2.752	2.538

The equations of ML from (2.8) may be written in the forms of the ordered variate $u_{(l)}$, $1<l<m$ and involve unyielding functions $g(u_{(l)})=u_{(l)}^{-1}$; $g(u_{(l)})=u_{(l)}^{p-1}$. The solutions of likelihood equations are obtained in a similar manner discussed for LTS as,

$$\hat{\beta}_T=D-C\hat{\sigma} \tag{2.9}$$

and

$$\hat{\sigma}=\frac{-B+\sqrt{B^2+4mQ}}{\sqrt{4m(m-2)}} \tag{2.10}$$

where

$$\delta_l=(p-1)\beta_{lo}+p\beta_l;\ \beta_{lo}=\alpha_l^{-2};\beta_l=(p-1)\alpha_l^{p-2};\ \alpha_l=E(U_{(l)}),$$

$$\Delta_l=(p-1)\pi_{lo}-p\pi_l;\ \pi_{lo}=2\alpha_l^{-1};\ \pi_l=(2-p)\alpha_l^{p-1},\ n=\sum_{l=1}^{m}\delta_l,\ \Delta=\sum_{l=1}^{m}\Delta_l,$$

$$D=\sum_{l=1}^{m}\delta_l z_{[l]}y_{[l]} \Big/ \sum_{l=1}^{m}\delta_l z_{[l]}^2,\ C=\sum_{l=1}^{m}\Delta_l z_{[l]}y_{[l]} \Big/ \sum_{l=1}^{m}\delta_l z_{[l]}^2, \tag{2.11}$$

$$B=\sum_{l=1}^{m}\Delta_l(y_{[l]}-Dz_{[l]}),\ Q=\sum_{l=1}^{m}\delta_l(y_{[l]}-Dz_{[l]})^2.$$

Note that $\delta_l > 0$ for all $p > 1$. The MML estimator $\hat{\sigma}$ is always real and positive. The MMLEs given in (2.9) and (2.10) by Islam, Tiku and Yildirim (2001) are extremely efficient and robust as compared to their competing LS estimators when the error term is from (2.8). The MMLE methodology for (2.8) provides an efficient estimator when the shape parameter is greater than 2. The Fisher information is also described only when the shape parameter is greater than 2. Therefore, it is necessary to employ an alternate method for the case $0 < p < 2$, such as GLS. Let $u_{(l)} = \dfrac{y_{[l]} - \beta z_{[l]}}{\sigma}$ be the standardized order statistics from (2.8), the GLS estimator of β and σ are derived on the same lines as given in (2.3)-(2.4), where the elements of Ω and α are regenerated using Lieblein's expression for (2.8) to nine decimal places for $n \le 15$ and $p = 1$, 2 and 3.5. We calculate the $\hat{\beta}_T$ and $\hat{\beta}_{GL}$ following (2.9) and (2.3) if the error term is from $W(p,\sigma)$.

Now using the generalized least square (GLS) estimation $\hat{\beta}_{GL}$ is obtained following (2.3) when the error is following $L(1,\sigma)$ or $LTS(p,\sigma)$ or $W(p,\sigma)$ and using MMLE, $\hat{\beta}_T$ is computed following (2.6) when the error is following $LTS(p,\sigma)$ and $\hat{\beta}_T$ is obtained following (2.9) when the error term follows $W(p,\sigma)$, to attain efficient estimators under non-normal distributions, proposed a family of ratio-type estimators as,

$$\bar{y}_{prjt} = \frac{\bar{y} + \hat{\phi}_t(\bar{Z} - \bar{z})}{\lambda_j \bar{z} + \gamma_j}(\lambda_j \bar{Z} + \gamma_j), \quad j = 1,2,...,5 \text{ and } t = 1,2 \tag{2.12}$$

where $\hat{\phi}_1 = \hat{\beta}_T$, $\hat{\phi}_2 = \hat{\beta}_{GL}$, λ_j and the values of γ_j are the same as given above. The MSEs of (2.12) may be evaluated in a similar way given in Kadilar and Cingi (2004) as,

$$MSE(\bar{y}_{prjt}) \cong (1-f)/m(R_{KCj}^2 S_z^2 + 2\phi_t R_{KCj} S_z^2 + \phi_t{}^2 S_z^2 - 2R_{KCj} S_{zy} - 2\phi_t S_{zy} + S_y^2),$$

where $\phi_1 = \beta_T$; $\phi_2 = \beta_{GL}$ are obtained over the whole population and R_{KCj} for $j = 1,2,...,5$ are the same as given above. It is important to note that Oral and Kadilar (2011) proposed ratio-type estimators using MMLE methodology are the special case of (2.12) when the error term is from $LTS(p,\sigma)$ and $\hat{\phi}_1 = \hat{\beta}_T$.

To set the conditions when the (2.12) are more efficient than the KCEs $\bar{y}_{KCj}$, we solve the inequalities

$$MSE(\bar{y}_{prjt}) < MSE(\bar{y}_{KCj}) \text{ for } j = 1,2,...,5,\ t = 1,2 \tag{2.13}$$

The solutions of (2.13) as,

$$-2R_{KCj} < \phi_t - \beta_L < 0, \text{ if } R_{KCj} > 0$$

$$0 < \phi_t - \beta_L < -2R_{KCj}, \text{ if } R_{KCj} < 0 \tag{2.14}$$

for $j=1,2,...,5, t=1,2$ respectively. Hence when condition (2.13) is fulfilled, the proposed estimators (2.14) are provided better results as compared to the KCEs given in (1.3).

Comment:

In practical life, there is a possibility that the magnitude of the shape parameter p in (2.5) and (2.8) may not be available in advance. In this situation, a Q-Q plot will be constructed by plotting the fractiles of the population in opposition to the ordered observations of the variable of interest. An approximate straight line points out the plausible distribution of error. Through the prescribed procedure, suitable values of p can be selected.

3. SIMULATIONS

In this section for the simulation study, we assume the model $y_l = \beta z_l + e_l$, in which e_l and z_l independently generated, and calculate y_l for $l=1,2,...,M$. Let errors $e_1, e_2, ..., e_M$ be random observation from a population from (2.4) or (2.5) or (2.8) and let Π_M denotes bivariate population consisting of $(x_1, y_1), (x_2, y_2), ..., (x_M, y_M)$. To calculate the MSEs of (2.12), we have to compute $\bar{y}_{prj1}$ and $\bar{y}_{prj2}$ ($j=1,2,...,5$) for all possible $S=\binom{M}{m}$ SRS of size m from Π_M, drawn from the population (2.5) or (2.8) and from (2.4), (2.5) and (2.8), respectively. Since S is large enough, therefore, conducting a Monte Carlo studies as, taking $M=100$ in each replication and z is from $U(0,1)$ and suppose $\beta=1$ with no loss of generality. To decide the parametric value of $\sigma_e^2=(1/12)\left[1/\rho^2-1\right]$ such that $\rho=0.70$. From the replicated population Π_{100} picked at random $N=25000$ of all the possible $\binom{100}{m}$ SRS of size $m=5,10$ and 15, which provide 25000 values of $\bar{y}_{prj1}$ and $\bar{y}_{prj2}$ ($j=1,2,...,5$). To compare the (2.12) in the form of efficiencies for a given m we determine the values of the MSEs,

$$MSE(\bar{y}_{prjt})=\sum_{k=1}^{N}(\bar{y}_{prjt}-\bar{Y})^2/N \ (t=1,2) \text{ and } MSE(\bar{y}_{KCj})=\sum_{k=1}^{N}(\bar{y}_{KCj}-\bar{Y})^2/N \quad (3.1)$$

where $\bar{Y}=\sum_{l=1}^{M} y_l/M$. We take $\sigma=1$ in each replication with no loss of generality. It is assumed that the condition (2.14) is satisfied for all populations. The values of the relative efficiencies may be computed as $E_{jt}=\left\{MSE(\bar{y}_{KCj})/MSE(\bar{y}_{prjt})\right\}$ $(j=1,2,...,5)$ and $(t=1,2)$. Where the MSEs are obtained from (2.8), and give the results in Table 1-3. Seeing as Table 1, a family of proposed ratio type estimators (2.12) is more efficient than the KCEs (1.3). All the E_{j2} values ($j = 1, 2, ..., 5$) slightly decrease for increasing the sample sizes.

In view of Table 2, it may be concluded that the proposed estimators $\bar{y}_{prj2}$, given in (2.12) for $t = 2$, are most efficient than the Oral and Kadilar estimators $\bar{y}_{prj1}$ and the KCEs under the LTS family for $p = 2.5$ (for conciseness, we consider the value of shape parameter 2.5 only).

From Table 3, it is clear that $p = 1$ the family of PRTEs (2.12) (based on MMLE methodology) does not exist but they exist for GLS. Since (2.12) $p = 1$ using GLS not only exist but also provides efficient results as compared to their competing estimators (1.3). Further, when the shape parameter p is increased from 1 the proposed estimators based on GLS are the most efficient among the proposed estimators (2.12) based on MMLE and the KCEs (1.3) in this study.

Table 1
Efficiencies under the Population when $e \sim L(1, \sigma)$

m	E_{12}	E_{22}	E_{32}	E_{42}	E_{52}
5	**1.466212**	**1.202601**	**1.144655**	**1.245044**	**1.128423**
10	1.248147	1.115067	1.070192	1.144685	1.056823
15	1.171657	1.08017	1.043782	1.102904	1.032572

Table 2
Efficiencies under the Population $e \sim LTS(p, \sigma)$

m		$\bar{y}_{pr1t}$	$\bar{y}_{pr2t}$	$\bar{y}_{pr3t}$	$\bar{y}_{pr4t}$	$\bar{y}_{pr5t}$
5	E_{j1}	1.024583	1.013687	1.010112	1.01604	1.009031
	E_{j2}	1.352614	1.177721	1.119948	1.217095	1.103137
10	E_{j1}	1.022215	1.013334	1.010329	1.015305	1.009427
	E_{j2}	1.225589	1.120399	1.084019	1.144168	1.073096
15	E_{j1}	1.017283	1.01062	1.008077	1.012227	1.007298
	E_{j2}	1.14668	1.074802	1.046809	1.09240	1.038205

Table 3
Efficiencies under the Population when $e \sim W(p,\sigma)$

	m		$\bar{y}_{pr1t}$	$\bar{y}_{pr2t}$	$\bar{y}_{pr3t}$	$\bar{y}_{pr4t}$	$\bar{y}_{pr5t}$
$p=1$	5	E_{j1}	-	-	-	-	-
		E_{j2}	3.883701	2.50833	1.688257	2.912046	1.43312
	10	E_{j1}	-	-	-	-	-
		E_{j2}	3.515879	2.599499	1.491376	3.148643	1.195488
	15	E_{j1}	-	-	-	-	-
		E_{j2}	3.802927	2.605158	1.37496	3.304044	1.08122
$p=2$	5	E_{j1}	1.347028	1.397311	1.41083	1.38561	1.409593
		E_{j2}	5.625179	4.629629	3.401895	4.998856	2.909491
	10	E_{j1}	1.268436	1.322325	1.334074	1.3078	1.329715
		E_{j2}	3.701511	3.626139	2.569918	3.903596	2.163619
	15	E_{j1}	1.25223	1.311403	1.322921	1.296032	1.317817
		E_{j2}	3.440089	3.392128	2.393679	3.656412	2.014284
$p=3.5$	5	E_{j1}	3.381481	3.565652	3.640328	3.461170	3.580654
		E_{j2}	5.521005	6.194745	5.463336	6.037743	4.854723
	10	E_{j1}	2.161697	2.503059	2.590530	2.404802	2.561859
		E_{j2}	3.551816	4.383657	3.849291	4.255842	3.423090
	15	E_{j1}	1.966566	2.259643	2.309175	2.181329	2.273262
		E_{j2}	3.18949	3.821161	3.284711	3.762029	2.918652

4. ROBUSTNESS OF A FAMILY OF PROPOSED RATIO-TYPE ESTIMATORS

More often outliers occur in sample observations. They may be unnoticed because nowadays much data is processed by computer without careful inspection or screening. Therefore, it is essential to employ a robust estimator that will not look for outliers but will reduce their effect if present. So, under the few data anomalies, the performance of (2.12) is assessed. In this section, taking $\sigma=1$, $\theta=1$ with no loss of generality and study the robustness properties of a family of PRTEs given in (2.12) as follows; firstly, it is assumed that z is from $U(0,1)$, and e_l is from $L(1,\sigma)$. We determine

1) True model $L(1,1)$
2) Dixon's outlier model; $m-m_o$ observations from $L(1,1)$ and m_o (we do not know which) from $L(1,4)$, where m_o is calculated from the formula $\left[\left|\frac{m}{10}+\frac{1}{2}\right|\right]$
3) Contaminated model; $0.90L(1,1)+0.10L(1,4)$

To recognize that model (1), may be considered as the true population model for comparisons purpose and the other models (2)-(3) are elected as its probable substitute. In Dixon's outlier model (2), we adopt the procedure to inject the outliers into each sample rather than the generated populations so that all the samples drawn from Π_{100} contain outliers. In estimating $\hat{\beta}_{GL}$, we calculate the (2.5) and use it in models (1)-(3).With the attention that all models have the same variance as that of y, standardized the generated $e_l^{'s}$ ($i=1,2,...,M$) in all models. The replicated MSEs of (2.12) for $t=2$ and their corresponding efficiencies E_{j2} for $j=1,2,...,5$, are given in Tables 3 and 4, respectively.

If e_l in the model (2.1) is from $LTS(2.5,1)$ and z is from $U(0,1)$ then the robustness properties of (2.12) following the below models

4) True model $LTS(2.5,1)$
5) Dixon's model; $m-m_o$ values from $LTS(2.5,1)$ and m_o (we do not know which) from $LTS(2.5,4)$ where $m_o=|1/2+m/10|$.
6) Contaminated model: $0.90LTS(2.5,1)+0.10L(2.5,4)$

For comparison purposes, we assume that model (4) is the population model, and models (5)-(6) are its alternatives. The replicated MSEs of the family of proposed ratio type estimators (2.12) and the efficiencies E_{jt} under model (4)-(6) are given in Table 5.

Further, assuming that e_l is follows an asymmetric family (2.8) and z is from $U(0,1)$. To determine our population model as

7) True model: $W(2.5,1)$
8) Dixon's model; $m-m_o$ values from $W(2.5,1)$ and m_o (we do not know which) from $W(2.5,4)$ where $m_o=|1/2+m/10|$.
9) Contamination model: $0.90W(2.5,1)+0.10W(1.3,1)$
10) Mis-specified model: $W(5,1)$.

Let model (7) is taken as the population model and all the models (8)-(10) are elected as its probable substitutes. The replicated MSEs of the proposed estimators are computed with the above-mentioned procedure and the efficiencies E_{jt} under model (7)-(10) are calculated and results are given in Table 6.

Table 4
Efficiencies under Models (1)-(3) for the Laplace Family

m		$\bar{y}_{pr1t}$	$\bar{y}_{pr2t}$	$\bar{y}_{pr3t}$	$\bar{y}_{pr4t}$	$\bar{y}_{pr5t}$
True Model (1)						
5	E_{j2}	1.391889	1.189135	1.129618	1.229872	1.11235
	$MSE(\bar{y}_{prj2})$	0.382723	0.245226	0.227687	0.26257	0.224337
10	E_{j2}	1.25747	1.100418	1.044556	1.136851	1.027878
	$MSE(\bar{y}_{prj2})$	0.092588	0.086054	0.08618	0.086799	0.086576
15	E_{j2}	1.215512	1.076187	1.022408	1.110328	1.006109
	$MSE(\bar{y}_{prj2})$	0.072633	0.069295	0.069848	0.069538	0.070278
Dixon Model (2)						
5	E_{j2}	1.236214	1.111619	1.079184	1.135602	1.070284
	$MSE(\bar{y}_{prj2})$	0.386443	0.272525	0.257082	0.286679	0.253687
10	E_{j2}	1.092396	1.041309	1.026710	1.051484	1.022478
	$MSE(\bar{y}_{prj2})$	0.167352	0.158763	0.156911	0.160344	0.156511
15	E_{j2}	1.049366	1.016025	1.004206	1.023659	1.000592
	$MSE(\bar{y}_{prj2})$	0.106298	0.104028	0.103713	0.104405	0.103696
Contaminated Model (3)						
5	E_{j2}	1.345613	1.189369	1.127896	1.229251	1.109553
	$MSE(\bar{y}_{prj2})$	0.379202	0.244599	0.227424	0.26185	0.224248
10	E_{j2}	1.29306	1.119954	1.058797	1.15995	1.040578
	$MSE(\bar{y}_{prj2})$	0.112496	0.102329	0.101945	0.103642	0.102285
15	E_{j2}	1.212375	1.076992	1.023788	1.110594	1.007617
	$MSE(\bar{y}_{prj2})$	0.072537	0.069143	0.069651	0.06941	0.070065

Table 4 shows that the (2.12) are more efficient than their corresponding KCEs (1.3) following the true model, and they sustain high efficiencies for the probable substitutes. In other words, the proposed estimators (2.12) are robust to plausible deviations from the true population.

Given Table 6, it may be analyzed that the (2.12) are remarkably efficient than the KCEs (1.3) under the true model (4), and are more efficient and robust under the models (7)-(10). Moreover, the proposed estimators (2.12) for $t = 2$ are the most efficient among the proposed estimators (2.12) for $t = 1$ and the KCEs (1.3).

Table 5
Efficiencies under Models (4)-(6) for LTS Family

m		$\bar{y}_{pr1t}$	$\bar{y}_{pr2t}$	$\bar{y}_{pr3t}$	$\bar{y}_{pr4t}$	$\bar{y}_{pr5t}$
True Model (4)						
5	E_{j1}	1.0256	1.01595	1.01214	1.01841	1.01098
	E_{j2}	1.4768	1.20923	1.14323	1.25639	1.12454
	$MSE(\bar{y}_{prj2})$	0.3628	0.24117	0.22428	0.25754	0.22100
10	E_{j1}	1.02697	1.01473	1.01026	1.01758	1.00890
	E_{j2}	1.30089	1.12062	1.05917	1.16117	1.04094
	$MSE(\bar{y}_{prj2})$	0.1157	0.10467	0.10414	0.10611	0.10444
15	E_{j1}	1.022465	1.012904	1.008839	1.015376	1.007564
	E_{j2}	1.231262	1.081996	1.024953	1.118417	1.007746
	$MSE(\bar{y}_{prj2})$	0.071799	0.068308	0.068873	0.068563	0.069313
Dixon Model (5)						
5	E_{j1}	1.02842	1.01156	1.00853	1.01395	1.00773
	E_{j2}	1.44316	1.14850	1.10421	1.18486	1.09278
	$MSE(\bar{y}_{prj2})$	0.38257	0.27169	0.25600	0.28596	0.25252
10	E_{j1}	1.00796	1.00362	1.00230	1.00453	1.00192
	E_{j2}	1.09007	1.04136	1.02646	1.05157	1.02211
	$MSE(\bar{y}_{prj2})$	0.16586	0.15747	0.15565	0.15902	0.15527
15	E_{j1}	1.005395	1.002648	1.001742	1.00326	1.001478
	E_{j2}	1.055003	1.0186	1.00571	1.026944	1.001776
	$MSE(\bar{y}_{prj2})$	0.104682	0.102416	0.10214	0.102779	0.102139
Contaminated Model (6)						
5	E_{j1}	1.0427	1.026045	1.019664	1.030099	1.017702
	E_{j2}	1.42599	1.19739	1.130847	1.243535	1.11171
	$MSE(\bar{y}_{prj2})$	0.330986	0.231848	0.21822	0.245456	0.215762
10	E_{j1}	1.035815	1.021229	1.015016	1.024991	1.013058
	E_{j2}	1.289697	1.115898	1.054223	1.156224	1.035858
	$MSE(\bar{y}_{prj2})$	0.114783	0.104372	0.10409	0.105674	0.104481
15	E_{j1}	1.029174	1.016603	1.011063	1.019916	1.00930
	E_{j2}	1.207157	1.075773	1.023081	1.108775	1.00697
	$MSE(\bar{y}_{prj2})$	0.073445	0.069935	0.070416	0.070224	0.070825

Table 6
Efficiencies under Models (7)-(10) for Weibull Distribution

m		$\bar{y}_{pr1t}$	$\bar{y}_{pr2t}$	$\bar{y}_{pr3t}$	$\bar{y}_{pr4t}$	$\bar{y}_{pr5t}$
True Model (7)						
5	E_{j1}	2.0869	2.0680	2.1027	2.0378	2.0951
	E_{j2}	5.9551	5.3028	4.1428	5.5129	3.5809
	$MSE(\bar{y}_{prj2})$	1.3838	0.3138	0.2410	0.4186	0.2382
10	E_{j1}	1.6010	1.7314	1.7614	1.6949	1.7499
	E_{j2}	3.6506	3.9711	3.0528	4.0848	2.6131
	$MSE(\bar{y}_{prj2})$	0.3957	0.1408	0.1196	0.1756	0.1216
15	E_{j1}	1.5254	1.6561	1.6812	1.6216	1.6685
	E_{j2}	3.3160	3.5930	2.7731	3.7068	2.3877
	$MSE(\bar{y}_{prj2})$	0.2474	0.0990	0.08602	0.1206	0.0876
Dixon Model (8)						
5	E_{j1}	2.1707	1.9884	1.9185	2.0075	1.8773
	E_{j2}	6.0530	4.2223	3.1179	4.7062	2.7275
	$MSE(\bar{y}_{prj2})$	1.2115	0.3387	0.2835	0.4181	0.2806
10	E_{j1}	1.5634	1.5367	1.4692	1.5548	1.4344
	E_{j2}	3.0773	2.5454	1.9753	2.7933	1.7740
	$MSE(\bar{y}_{prj2})$	0.3822	0.2002	0.1848	0.2246	0.1859
15	E_{j1}	1.4807	1.4882	1.4343	1.4973	1.4030
	E_{j2}	2.8154	2.4349	1.8970	2.6490	1.7012
	$MSE(\bar{y}_{prj2})$	0.2458	0.1367	0.1278	0.1522	0.1292
Contamination (9)						
5	E_{j1}	2.3932	1.9867	1.8908	2.0231	1.8414
	E_{j2}	8.2663	4.1191	2.9198	4.7319	2.5297
	$MSE(\bar{y}_{prj2})$	1.2115	1.1729	0.3568	0.3104	0.4309
10	E_{j1}	1.5956	1.5774	1.5103	1.5921	1.4732
	E_{j2}	3.3370	2.7169	2.0326	3.0155	1.7977
	$MSE(\bar{y}_{prj2})$	0.3106	0.3845	0.1938	0.1819	0.2176
15	E_{j1}	1.4689	1.4564	1.3941	1.4717	1.3618
	E_{j2}	2.7599	2.2736	1.7633	2.5022	1.5879
	$MSE(\bar{y}_{prj2})$	0.1847	0.2573	0.1539	0.1474	0.1680
Misspecification (10)						
5	E_{j1}	6.2277	6.0993	6.1209	5.8460	5.8355
	E_{j2}	6.4918	7.1599	7.2218	6.6447	6.6865
	$MSE(\bar{y}_{prj2})$	3.1854	0.6182	0.3664	0.9191	0.3363
10	E_{j1}	2.8121	3.4975	3.6383	3.2972	3.5432
	E_{j2}	3.4916	4.6365	4.4725	4.3445	4.1003
	$MSE(\bar{y}_{prj2})$	1.1079	0.3259	0.2184	0.4466	0.2060
15	E_{j1}	2.4634	2.9731	3.0299	2.8372	2.9445
	E_{j2}	3.1243	3.9251	3.6692	3.7605	3.3670
	$MSE(\bar{y}_{prj2})$	0.7470	0.2589	0.1851	0.3394	0.1763

5. CONCLUSIONS

Mostly in practical fields, error terms in simple linear regression models deviate from normality due to outliers. As a result, KCEs are seriously affected. To cope with it Oral and Kadilar incorporated MMLEs (using LTS family and generalized logistic distribution of error) into KCEs to enhance the efficiencies. In certain situations, the applicability of the MMLE methodology is restricted or ineffective e.g. If the error follows the Weibull distribution or Laplace distribution. In this study, the GLSEs (Lloyd, 1952) are planned to meet above mention situation. The GLS estimators are derived for the case if the error term follows the LTS family. By integrating the GLS estimator into KCEs, a family of ratio-type estimators is proposed, whose performance is better than the ratio-type estimators based on MMLEs suggested by Oral and Kadilar (2011) for small sample sizes and their classical counterpart advised by Kadilar and Cingi (2004). The results are extended for the case when the error term follows skewed distribution i.e. Weibull distribution. It is shown that the family of proposed ratio-type estimators is more efficient and robust than the ratio-type estimators proposed by Oral and Kadilar (2011) and Kadilar and Cingi (2004). By knowing the features of Laplace distribution discussed by Farnoosh and Jafarpour (2005), we further extended the results for the case when the error term is from Laplace distribution, which is widely used for robustness study, and concluded that PRTEs are more efficient and robust than the Kadilar and Cingi (2004) estimators.

REFERENCES

1. Ahmad, A., Sanaullah, A. and Hanif, M. (2019). Modified maximum likelihood integrated robust ratio estimator in simple random sampling. *Punjab University Journal of Mathematics*, 51(3), 133-145.
2. Farnoosh, R., and Jafarpour,H.(2005, December). About the shape parameters of the family of Laplace distribution functions. *In Proceedings of the 8th WSEAS International Conference on Applied Mathematics*, 86-89.
3. Farrella, J.P., and Barrera,M.S.(2006). A comparison of several robust estimators for a finite population mean. *Journal of Statistical Studies,* 26, 29-43.
4. Govindarajulu, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population. *Journal of the American Statistical Association*, 61(313), 248-258.
5. Islam, M.Q., Tiku, M.L. and Yildirim, F. (2001). Non-normal regression I, skew distributions. *Communications in Statistics Theory and Methods*, 30, 993-1020.
6. Islam, M.Q., and Tiku, M.L. (2004). Multiple linear regression model under non-normality. *Communications in Statistics Theory and Methods*, 33, 2443-2467.
7. Kadilar, C., Candan, M. and Çıngı, H. (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics,* 36(2) 181-188.
8. Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151(3), 893-902.
9. Kumar, S. and Chhaparwal, P. (2017). Robust exponential ratio and product type estimators for population mean using order statistics in simple random sampling. *International Journal of Ecological Economics and Statistic*, 38(3), 51-70.
10. Lloyd, E.H. (1952) Least-squares estimation of location and scale parameters using order statistics. *Biometrik*, 39(1/2) 88-95.

11. Oral, E. and Kadilar, C. (2011). Robust ratio-type estimators in simple random sampling. *Journal of the Korean Statistical Society*, 40(4) 457-467.
12. Oral, E. and Oral, E. (2011). A robust alternative to the ratio estimator under non-normality. *Statistics and Probability Letter*, 81(8) 930-936.
13. Ray, S.K. and Singh, R.K. (1981). Difference-cum-ratio type estimators. *Journal of Indian Statistical Association*, 19, 147-151.
14. Singh, H.P. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6, 555-560.
15. Tiku, M.L. and Kumra, S. (1985). Expected values and variances and covariances of order statistics for a family of symmetric distributions. *Selected tables in mathematical statistics, American Mathematical Society*, 8, 141-27.