Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 26 (2023), No. 4, pp. 823–832
DOI : 10.47974/JSMS-928

An efficient class of regression estimators using conventional measures

Muhammad Ijaz *
Syed Muhammad Asim
Bushra Haider
Department of Mathematics and Statistics
University of Haripur
Haripur
Pakistan

Samir Brahim
Division of Information and Computing Technology
College of Science and Engineering
Hamad Bin Khalifa University
Qatar

Tolga Zaman
Department of Statistics
Cankiri Karatekin University
Turkey

Abstract

The study suggests a class of regression type estimators for estimating the population mean when there exist either a positive or negative correlation between a study and auxiliary variable. The auxiliary attributes have been used in the form of the standard deviation, coefficient of kurtosis, mean deviation, quartile deviation, median and 10th decile. The bias and mean squared error has studied up to the first order of approximation. A comparison is made both theoretically and numerically with some existing regression estimators. It is observed that the suggested class of estimators provides more reliable and efficient results than others.

Subject Classification: *62D05.*

Keywords: *Standard deviation, Coefficient of kurtosis, Mean deviation, Quartile deviation, Median and 10th decile.*

**E-mail:* `m.ijaz@uoh.edu.pk`

1. Introduction

From the last few decades, it is usual practice of the researchers to improve the efficiency of the existing estimators by using the auxiliary information. The idea of the auxiliary information was first established by Cochran [1]. In an estimation of parameters, the knowledge of auxiliary variables improves the efficiency of the population parameters i.e. variance, mean, and coefficient of variation of the study variable y. When we are given the auxiliary information x, a lot of new modified ratio, product and regression type estimators have been presented by researchers. Among others, Kadilar and Cingi [2, 3], Yan and Tian [4], Subramani and Kumarapandiyan [5, 6, and 7], Jeelani et al. [8], and Subzar et al. [9]. The more recent contribution to the theory of sample surveys is Ünal and Kadilar [10], Pal et al. [11], Irfan et al. [12], Zamanzade and Mahdizadeh [13], Cekim and Kadilar [14], Yadav and Zaman [15].

Regression estimator has wide applications in many real world problems that may arise in business, economics, times series analysis, determinants of a particular disease, etc. In practice, the regression model involves unknown parameters which are to be estimated first. In this paper, we suggest a novel estimator that uses some known auxiliary information instead of the parameters. The proposed class of regression type estimators improves the estimation problem as compared to the existing estimators. The proposed class of estimators is compared with other estimator's i.e. Kadilar and Cingi [2, 3], Subramani and Kumarapandiyan [5, 6, and 7], Jeelani et al. [8], and Subzar et al. [9]. For the purpose of illustration, a real data set is considered to delineate the efficiency of the proposed over existing estimators.

2. Existing Estimators

Lets consider $U = (U_1, U_2, U_3, \ldots, U_N)$ be the finite population with a total of N units and let y_i and x_i be the specified value of a study variable Y and auxiliary variable X. If we take a random sample of size n from the population of U then $\bar{y}$ and $\bar{x}$ be the mean of the study and auxiliary variable. Other essential notations used in the manuscript are

S_x, S_y: Population standard deviation of the auxiliary and study variable

C_x, C_y: Coefficient of variation of the auxiliary and study variable

ρ_{xy} : Population Correlation coefficient between X and Y

$\beta_{2(x)}$: Kurtosis of the auxiliary variable

$\beta_{1(x)}$: Skewness of the auxiliary variable

Q.D: Quartile deviation of the auxiliary variable
M.D: Median of the auxiliary variable
D: Deciles of the auxiliary variable
MSE – Mean square error
b: regression coefficient
SRSWOR – Simple random sampling without replacement

Kadilar and Cingi [2] proposed an improved estimators utilizing population coefficient of variation and coefficient of kurtosis. The estimators are given as

$$\bar{y}_{lr1} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_2+C_x)}(\bar{X}\beta_2 + C_x)$$

$$\bar{y}_{lr2} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}C_x+\beta_2)}(\bar{X}C_x + \beta_2)$$

The mathematical expressions for the bias and the MSE are

$$Bias(\bar{y}_{lrj}) = \lambda \frac{S_x^2}{\bar{Y}} R_j^2 \tag{1}$$

$$MSE(\bar{y}_{lrj}) = \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \tag{2}$$

where j=1, 2 and $R_1 = \frac{\bar{Y}\beta_2}{\bar{X}\beta_2+C_x}$ and $R_2 = \frac{\bar{Y}C_x}{\bar{X}C_x+\beta_2}$.

Kadilar and Cingi [3] proposed another improved class of estimators by using the coefficient of correlation and coefficient of kurtosis. The estimators are given as

$$\bar{y}_{lr3} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\rho+\beta_2)}(\bar{X}\rho + \beta_2)$$

$$\bar{y}_{lr4} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_2+\rho)}(\bar{X}\beta_2 + \rho)$$

The expressions for the bias and the MSE are

$$Bias(\bar{y}_{lrj}) = \lambda \frac{S_x^2}{\bar{Y}} R_j^2 \tag{3}$$

$$MSE(\bar{y}_{lrj}) = \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \tag{4}$$

where j=3, 4 and $R_3 = \frac{\bar{Y}\rho}{\bar{X}\rho+\beta_2}$, $R_4 = \frac{\bar{Y}\beta_2}{\bar{X}\beta_2+\rho}$.

Yan and Tian [4] proposed estimators by utilizing the coefficient of skewness and kurtosis which are given by

$$\bar{y}_{lr5} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+\beta_1)}(\bar{X} + \beta_1)$$

$$\bar{y}_{lr6} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_1+\beta_2)}(\bar{X}\beta_1+\beta_2)$$

The expressions for the bias and the MSE of the above estimators are defined as

$$Bias(\bar{y}_{lrj}) = \lambda\frac{S_x^2}{\bar{Y}}R_j^2 \tag{5}$$

$$MSE(\bar{y}_{lrj}) = \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \tag{6}$$

where j=5, 6 and $R_5 = \frac{\bar{Y}}{\bar{X}+\beta_1}$ and $R_6 = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1+\beta_2}$

Subramani and Kumarapandiyan [5, 6, and 7] proposed efficient estimators by utilizing the population median. The estimator is given as

$$\bar{y}_{lr7} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}C_x+M_d)}(\bar{X}C_x+M_d)$$

The expressions for bias and the MSE of the above estimator are given as

$$Bias(\bar{y}_{lrj}) = \lambda\frac{S_x^2}{\bar{Y}}R_j^2 \tag{7}$$

$$MSE(\bar{y}_{lrj}) = \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \tag{8}$$

where j=7 and $R_7 = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1+M_d}$

Jeelani et al. [8] proposed improved the regression estimator by utilizing the coefficient of skewness and quartile deviation at a time. The estimator is

$$\bar{y}_{lr8} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_1+QD)}(\bar{X}\beta_1+QD)$$

The expressions for bias and the MSE of $\bar{y}_{lr10}$ is

$$Bias(\bar{y}_{lrj}) = \lambda\frac{S_x^2}{\bar{Y}}R_j^2 \tag{9}$$

$$MSE(\bar{y}_{lrj}) = \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \tag{10}$$

where, j=8 and $R_8 = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1+QD}$.

Subzar et al. [9] proposed regression estimators and improve its efficiency by using the population deciles and coefficient of skewness. Their proposed estimators are

$$\bar{y}_{lr9} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_1+D_2)}(\bar{X}\beta_1+D_2)$$

$$\bar{y}_{lr10} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_1+D_5)}(\bar{X}\beta_1 + D_5)$$

The expressions for bias and the MSE of $\bar{y}_{lr9}$ and $\bar{y}_{lr10}$ are respectively given by

$$Bias(\bar{y}_{lrj}) = \lambda \frac{S_x^2}{\bar{Y}} R_j^2 \tag{11}$$

$$MSE(\bar{y}_{lrj}) = \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \tag{12}$$

where j=9, 10 and $R_{09} = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1+D_2}$ and $R_{10} = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1+D_5}$.

3. Proposed Class of Regression Estimators

In this section, an improved class of regression type estimators is suggested for estimating the finite population mean $\bar{Y}$ by using the auxiliary information on some specific variable $G_{(x)}$. The proposed class is

$$\bar{y}_{blri} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_xG_{(x)})}(\bar{X} + S_x G_{(x)}), \quad i = 1,\ 2,\ 3,\ 4,\ 5$$

$G_{(x)}$ is to be considered $\beta_{2(x)}$, MD, QD, M_d, and D_{10}. The estimators are

$$\bar{y}_{blr1} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_x\beta_{2(x)})}(\bar{X} + S_x\beta_{2(x)})$$

$$\bar{y}_{blr2} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_xMD)}(\bar{X} + S_x MD)$$

$$\bar{y}_{blr3} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_xQD)}(\bar{X} + S_x QD)$$

$$\bar{y}_{blr4} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_xM_d)}(\bar{X} + S_x M_d)$$

$$\bar{y}_{blr5} = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+S_xD_{10})}(\bar{X} + S_x D_{10})$$

The bias and the MSE of the proposed estimators up to the first order of approximation is generally defined by

$$Bias(\bar{y}_{blri}) = \frac{1}{2}\lambda R_{bi}^2 \frac{S_x^2}{\bar{Y}} \tag{13}$$

$$MSE(\bar{y}_{blri}) = \lambda[R_{bi}^2 S_x^2 + S_y^2(1-\rho^2)] \tag{14}$$

where $R_{bi} = \frac{\bar{Y}}{\bar{X}+S_xG_{(x)}}$ and i=1,2,3,4,5

$R_{b1} = \frac{\bar{Y}}{\bar{X}+S_x\beta_{2(x)}}$, $R_{b2} = \frac{\bar{Y}}{\bar{X}+S_xMD}$, $R_{b3} = \frac{\bar{Y}}{\bar{X}+S_xQD}$ $R_{b4} = \frac{\bar{Y}}{\bar{X}+S_xM_d}$, $R_{b5} = \frac{\bar{Y}}{\bar{X}+S_xD_{10}}$.

4. Efficiency Comparison

In this section, the proposed class of estimators is compared both theoretically and numerically with existing estimators of the population mean under simple random sampling without replacement scheme. The proposed class of regression estimators will be more efficient then Kadilar and Cingi [2, 3], Yan and Tian [4], Subramani and Kumarapandiyan [5, 6 and 7], Jeelani et al. [8], and Subzar et al. [9] estimators if

$$
\begin{aligned}
MSE(\bar{y}_{blri}) &\leq MSE(\bar{y}_{lrj}) \\
\lambda[R_{bi}^2 S_x^2 + S_y^2(1-\rho^2)] &\leq \lambda[R_j^2 S_x^2 + S_y^2(1-\rho^2)] \\
R_{bi}^2 &\leq R_j^2 \\
R_{bi} &\leq R_j
\end{aligned}
\tag{15}
$$

where j=1,2,3,4,5,6,7,8,9,10 and i=1,2,3,4,5.

5. Empirical Study

For the purpose of numerical illustration, a real data set of Singh and Chaudhary [16] page 177 which is recently cited by [17] is considered. Descriptive Statistics about the population parameters is given below in Table 1. It is noted that the correlation between the study and auxiliary variable is 44% and with a sample of size n=20.

The percentage relative efficiency (PRE) of the proposed estimators are compared with the estimator $\bar{y}_{lr1}$ by using the formula

$$PRE = \frac{MSE(\bar{y}_{lr1})}{MSE(\bar{y}_{exisitng\ and\ proposed})} \times 100$$

Table 2 describes numerical illustrations of the theoretical conditions of the proposed (Column wise) vs existing estimators (Row wise) given in Eq. (15). We see that all the conditions are satisfied and thus the bias and the MSE of the proposed estimators will be less than the existing estimators. It has been concluded that the proposed estimators will perform better than others if these conditions are satisfied. The results of the bias and the MSE are given in Table 3.

Table 1
Descriptive statistics

N=34	n=20	$\bar{Y} = 856.4117$	$\bar{X} = 199.4412$	$\rho = 0.4453$	$S_y = 733.1407$
$S_x = 150.2150$	$C_y = 0.8561$	$C_x = 0.7531$	$\beta_1 = 1.1823$	$\beta_2 = 1.0445$	$QD = 89.375$
$MD = 71.5$	$M_d = 142.50$	$D_2 = 83.00$	$D_5 = 142.50$	$D_{10} = 144.481$	$\lambda = 0.0206$

Table 2

Justification of the efficiency conditions

$j \backslash i$	**1**	**2**	**3**	**4**	**5**
1	2.403<4.279	0.078<4.279	0.063<4.279	0.0396<4.279	0.0089<4.279
2	2.403<4.264	0.078<4.264	0.063<4.264	0.0396<4.264	0.0089<4.264
3	2.403<4.244	0.078<4.244	0.063<4.244	0.0396<4.244	0.0089<4.244
4	2.403<4.285	0.078<4.285	0.063<4.285	0.0396<4.285	0.0089<4.285
5	2.403<4.269	0.078<4.269	0.063<4.269	0.0396<4.269	0.0089<4.269
6	2.403<4.286	0.078<4.286	0.063<4.286	0.0396<4.286	0.0089<4.286
7	2.403<3.459	0.078<3.459	0.063<3.459	0.0396<3.459	0.0089<3.459
8	2.403<3.114	0.078<3.114	0.063<3.114	0.0396<3.114	0.0089<3.114
9	2.403<3.176	0.078<3.176	0.063<3.176	0.0396<3.176	0.0089<3.176
10	2.403<2.676	0.078<2.676	0.063<2.676	0.0396<2.676	0.0089<2.676

Table 3

Bias, MSE, and PRE of the proposed and existing estimators

Estimator	**Bias**	**MSE**	**PRE**
$\bar{y}_{lr1}$	9.938	17387.79	100
$\bar{y}_{lr2}$	9.870	17328.22	100.3438
$\bar{y}_{lr3}$	9.776	17249.129	100.8039
$\bar{y}_{lr4}$	9.966	17411.67	99.86285
$\bar{y}_{lr5}$	9.892	17348.06	100.229
$\bar{y}_{lr6}$	9.970	17415.66	99.83997
$\bar{y}_{lr7}$	6.494	14438.37	120.4277
$\bar{y}_{lr8}$	5.263	13383.71	129.9176
$\bar{y}_{lr9}$	5.475	13565.56	128.176
$\bar{y}_{lr10}$	3.887	12205.47	142.459
$\bar{y}_{blr1}$	**1.5676**	**11561.84**	**150.3895**
$\bar{y}_{blr2}$	**0.0017**	**8879.68**	**195.8155**
$\bar{y}_{blr3}$	**0.0011**	**8878.67**	**195.8378**
$\bar{y}_{blr4}$	**0.00043**	**8877.56**	**195.8623**
$\bar{y}_{blr5}$	**0.000022**	**8876.87**	**195.8775**

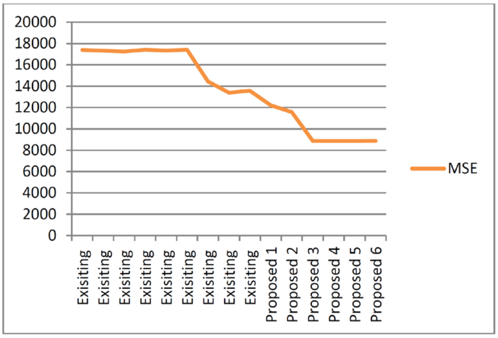

Figure 1
MSE of the proposed and existing estimators

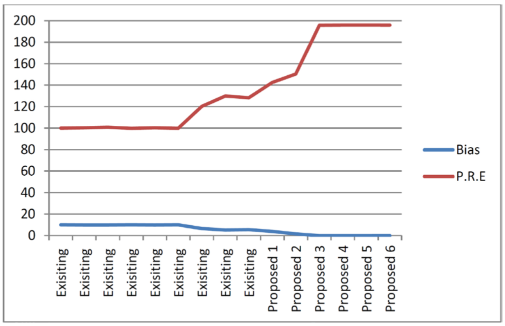

Figure 2
Bias and PRE of the proposed and existing estimators

Table 3 reflects the bias, and the MSE of the proposed and existing estimators. It has been noted that the proposed estimators have less bias and MSE than others. It is evident that the suggested estimators perform well as compared to the existing estimators.

Fig 1 defines the MSE of the proposed and existing estimators. It has been shown that the MSE of the proposed estimators is less than as compared to all other existing estimators. Fig 2 depicts the Bias and the PRE which clearly shows that the proposed estimators lead to a preferable result rather than the existing estimators.

6. Conclusion

In the present study, an improved class of regression type estimators of the finite population has been introduced on using the auxiliary information. The class of regression estimators has been proposed under simple random sampling without replacement scheme. The mathematical expressions for the bias and mean square error have been obtained for the proposed estimators up to the first order of approximation. A theoretical comparison of the proposed class of estimators has been made with existing estimators in terms of the MSE. Theoretical conditions are derived and it has been concluded that the proposed estimators leads to a significant results if these conditions are satisfied. A real data set is used to illustrate the theoretical conditions and the results are given in Table 2. The conditions are satisfied and hence the MSE and bias of the proposed estimators are smaller and PRE is larger for the proposed estimator than the existing ones. Thus, the proposed estimators work more perfectly than existing regression estimators of the population mean. Apart from the efficient results, the proposed estimators are also easy to understand and simple to use. Moreover, from the proposed regression estimators, the estimator $\bar{y}_{blr5}$ performs better than other four estimators both in terms of MSE and bias.

Reference

[1] Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*, *30*(2), 262-275.

[2] Kadilar, C., & Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, *151*(3), 893-902.

[3] Kadilar, C. & Cingi, H. (2006), 'An improvement in estimating the population mean by using the correlation coefficient', Hacettepe Journal of Mathematics and Statistics 35(1), 103–109

[4] Yan, Z. and Tian, B. (2010). Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable, ICICA 2010, Part II, CCIS 106, pp. 103–110

[5] Subramani, J. and Kumarapandiyan, G. (2012a). Estimation of Population Mean using Co-efficient of variation and Median of an Auxiliary Variable. *International Journal of Probability and Statistics,* (1), 111–118.

[6] Subramani, J. and Kumarapandiyan, G. (2012b). Estimation of population mean using known median and co-efficient of skewness. *American Journal of Mathematics and Statistics,* (2), 101-107.

[7] Subramani, J. and Kumarapandiyan, G. (2012c). Modified ratio estimators using known median and co-efficient of kurtosis. *American Journal of Mathematics and Statistics,* (2), 95–100.

[8] Jeelani, M. I. Maqbool, S. and Mir, S. A. (2013). Modified ratio estimators of population mean using linear combination of coefficient of skewness and quartile deviation. *International Journal of Modern Mathematical Sciences,* (6), 174–183.

[9] Subzar, M. A., Maqbool, S., Raja, T. A., Shabeer, M., & Lone, B. A. (2017). A Class of Improved Ratio Estimators for Population Mean using Conventional Location parameters. *Int. J. Modern Math. Sci*, *15*(2), 187-205.

[10] Ünal, C., & Kadilar, C. (2020). Exponential type estimator for the population mean in the presence of non-response. *Journal of Statistics and Management Systems*, *23*(3), 603-615.

[11] Pal, S.K., Singh, H.P. & Solanki, R.S. (2020). A new efficient class of estimators of finite population mean in simple random sampling. *Afr. Mat.* **31,** 595–607.

[12] Irfan, M., Javed, M., Bhatti, S. H., Raza, M. A., & Ahmad, T. (2020). Almost unbiased optimum estimators for population mean using dual auxiliary information. *Journal of King Saud University-Science*, 32(6), 2835-2844.

[13] Zamanzade, E., & Mahdizadeh, M. (2020). Using ranked set sampling with extreme ranks in estimating the population proportion. *Statistical Methods in Medical Research*, *29*(1), 165-177.

[14] Cekim, H. O., & Kadilar, C. (2020). In-type estimators for the population variance in stratified random sampling. *Communications in Statistics-Simulation and Computation*, *49*(7), 1665-1677.

[15] Yadav, S. K., & Zaman, T. (2021). Use of some conventional and non-conventional parameters for improving the efficiency of ratio-type estimators. *Journal of Statistics and Management Systems*, 24(5), 1077-1100.

[16] Singh, D., & Chaudhary, F. S. (1986). *Theory and analysis of sample survey designs*. John Wiley & Sons.

[17] Subzar, M., Maqbool, S., Verma, M. R., & Raja, T. A. (2017). An improved class of ratio estimators for estimating population mean using auxiliary information in survey sampling. *Journal of Reliability and Statistical Studies*, 10(2), 65-82.

Received August, 2021

Revised February, 2022