Scientia Iranica E (2022) 29(3), 1705–1715

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

Quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information

U. Shahzad[a,b,*], M. Hanif [b], I. Sajjad[c], and M.M. Anas[d]

a. *Department of Mathematics and Statistics, International Islamic University, Islamabad, 46000, Pakistan.*
b. *Department of Mathematics and Statistics - PMAS-Arid Agriculture University, Rawalpindi, 46300, Pakistan.*
c. *Department of Lahore Business School - University of Lahore, Islamabad, 46000, Pakistan.*
d. *School of Science, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, P.R. China.*

Received 13 September 2019; received in revised form 29 May 2020; accepted 21 September 2020

KEYWORDS
Quantile regression;
OLS regression;
Ratio-type estimators;
Simple random sampling;
Two stage sampling.

Abstract. Traditional Ordinary Least Square (OLS) regression is commonly utilized to develop regression-ratio-type estimators through traditional measurement of location. Abid et al. [Abid, M., Abbas, N., Zafar Nazir, H., et al. "Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters", *Revista Colombiana de Estadistica*, 39(1), pp. 63–79 (2016b)], extended this idea and developed regression-ratio-type estimators based on traditional and non-traditional measures of location. In this article, the quantile regression with traditional and non-traditional measures of location is utilized and a class of ratio type mean estimators is proposed. The theoretical Mean Square Error (MSE) expressions are also derived. The work is also extended to two-phase sampling (partial information). The relationship between the proposed and existing groups of estimators is shown by considering real data collections originating from different sources. The discoveries are empowering and prevalent execution of the proposed group of the estimators is witnessed and documented throughout the article.

1. Introduction

As a commonly known fact, the era of information is here. This fact not only highlights the volume and pace of information but also underlines the necessity of its accurate flow. The aforementioned issue is linked to its true urge for data collection. This consideration enables profiling the surroundings precisely which, in turn, helps make an optimal decision. To date, sampling theory and methods are still considered the core of the literature of multidisciplinary research. Utility of surveys and sampling methods in applied research are clearly observed in the surveys, e.g., Eurostat on European Union Labor Force giving periodical statistics of labor participation. Other examples depicting the significance of sampling methods include: (a) the market profiling-segmentation survey providing significant evidence to encourage market share investigation, and (b) national co-morbidity survey evaluating the substance disorder and anxiety levels.

The most philosophical objective of sampling methods is the valuation of the occurrence of patterns or attributes present in inhabitants or population under

*. *Corresponding author. Tel.: 00 92 33 15995587*
E-mail addresses: usman.stat@yahoo.com (U. Shahzad); mhpuno@hotmail.com (M. Hanif); irsasajjad@yahoo.com (I. Sajjad); anas.uaar@gmail.com (M.M. Anas)

doi: 10.24200/sci.2020.54423.3744

study. Auxiliary information is widely used to ensure greater precision in sample estimates. Laplace in 1814 gave a justification for significance of practicing auxiliary information in an easy but succinct manner. For instance, he found that the birth register providing the conditions of the residents could help estimate the population of a large country without holding a census of its residents. However, this is only possible if the ratio of masses to annual birth is known. In addition, McClellan et al. [1] investigated the distinction of separations of heart assault injured individual's home to closest cardiovascular catheterization medical clinic and the separation to the closest emergency clinic of any kind, as auxiliary information to help the estimation of impact of catheterization on patients health.

Ranked Set Sampling (RSS) method is an alternative method for the condition in which the abundance of auxiliary information is available. This method was presented by McIntyre [2] and was far more cost-efficient than simple random sampling method. In this respect, see Adel Rastkhiz et al. [3], Zamanzade and Vock [4], Zamanzade and Wang [5], Zamanzade and Mahdizadeh [6], Zamanzade and Wang [7], and Mahdizadeh and Zamanzade [8] for more information about this.

A significant scope of literature on ratio-type estimators for mean estimation is available. Koyuncu [9], Shahzad et al. [10], and Hanif and Shahzad [11] developed some classes of estimators utilizing supplementary information under a simple random sampling scheme. However, in case of positive correlation, the traditional regression-ratio-type estimator functions quite better in estimating the values of population parameters (see, [12–15]). Note that the traditional regression-type-ratio estimators function based on conventional regression coefficient, also known as OLS regression coefficient. However, the OLS estimate may not function well in case the data are contaminated by outliers. For solving this issue, there are some modifications available in the literature (see, [17–21]. Abid et al. [21] introduced the notion of utilizing non-traditional measures of location for the first time and enhanced population mean estimation. They found the enhanced estimates of population mean utilizing non-traditional measures. The current study is done based on the extension of the research study of Abid et al. [21] by incorporating quantile regression.

Outliers have a negative impact on the accuracy of the estimators obtained by OLS technique. The existence of outliers is overlooked in the OLS model assumption [22]. They may change the regression parameters to become smaller or larger than the parameters evaluated when the outliers in the data were excluded [23,24]. It is recommended that the outliers be barred from the examination if they were, from exhaustive examinations, proved to be non-legitimate perceptions. Yet, when the outliers are substantial, they can give new bits of knowledge about the nature of the data [24]. It is implied that a measurable procedure is required that would capture the outliers in this investigation, but is less impacted by their quality. An elective procedure enjoying a greater ability to settle a few issues referenced before is called quantile regression. It was established by Boscovich in the eighteenth century, even before the possibility of developing least squares regression estimators. Previously, utilization of quantile regression was restricted to financial matters or natural investigations; however, now, it is used in almost all the fields of social sciences as a data analysis device. In light of preceding lines, we are introducing a quantile regression coefficient rather than OLS in ratio-type mean estimators by extending the notion proposed by Abid et al. [21].

Inspired by the above documented developments, the current study proposes a new family of estimators for estimating population mean using a more scrupulous use of auxiliary variable. The objective is met by utilizing the quantile regression as well as traditional and non-traditional measures of location of auxiliary variable. The applicability of the scheme is further demonstrated in simple and two-stage random sampling frameworks applied to three diverse data sets derived from various fields of inquiries. Moreover, comparative investigations between Abid et al. [21] and the proposed family are keenly conducted by means of numerical evaluations. The rest of this paper is arranged in seven major parts. Section 2 presents a review of estimators constructed in the study conducted by Abid et al. [21]. Section 3 proposes a class of quantile regression-ratio-type estimators. Section 4 provides a two-stage sampling scheme version of the reviewed and proposed estimators. Section 5 shows numerical illustration and general findings. Section 6 gives the final remarks.

2. A review of Abid et al. estimators

Abid et al. [21] incorporated non-traditional measures of location into traditional measures of location for mean estimation. They used mid-range of X, denoted by MR, as the first non-traditional measure of location. Hodges-Lehmann estimator, denoted by HL, is the second non-traditional measure of location. The mentioned authors also used the weighted average of the population median and two extreme quartiles, known as Tri-mean, denoted by TM, as the third non-traditional measure of location. According to Abid et al. [21], these measures are highly sensitive in the absence of normality and in the presence of outliers. They introduced the following class of OLS regression-ratio-type estimators utilizing traditional and non-

Table 1. Family members of Abid et al. [21].

Under complete information estimators	(A_1, A_2)	Under partial information estimators
$\bar{y}_{ab_1}$	$(1, TM)$	$\bar{y}'_{ab_1}$
$\bar{y}_{ab_2}$	(C_x, TM)	$\bar{y}'_{ab_2}$
$\bar{y}_{ab_3}$	(ρ, TM)	$\bar{y}'_{ab_3}$
$\bar{y}_{ab_4}$	$(1, MR)$	$\bar{y}'_{ab_4}$
$\bar{y}_{ab_5}$	(C_x, MR)	$\bar{y}'_{ab_5}$
$\bar{y}_{ab_6}$	(ρ, MR)	$\bar{y}'_{ab_6}$
$\bar{y}_{ab_7}$	$(1, HL)$	$\bar{y}'_{ab_7}$
$\bar{y}_{ab_8}$	(C_x, HL)	$\bar{y}'_{ab_8}$
$\bar{y}_{ab_9}$	(ρ, HL)	$\bar{y}'_{ab_9}$

traditional measures of location as given by:

$$\bar{y}_{ab_i} = \frac{\bar{y} + \hat{\beta}_{(ols)}(\bar{X} - \bar{x})}{(A_1\bar{x} + A_2)} \, (A_1\bar{X} + A_2)$$

$$\text{for} \quad i = 1, 2, ..., 9,$$

where $(\bar{X}, \bar{Y})$ represent the population means and $(\bar{x}, \bar{y})$ are the sample means when a simple random sample of size n is drawn from the population. Further, A_1 and A_2 are either (0,1) or some known population measures, namely HL (the Hodges Lemon), MR (the Mid-Range), TM (Tri Mean), C_x (coefficient of variation), ρ (coefficient of correlation), and $\hat{\beta}_{(ols)}$ (OLS regression coefficient). The family members of $\bar{y}_{ab_i}$ are provided in Table 1. The MSE of $\bar{y}_{ab_i}$ is given below:

$$MSE(\bar{y}_{ab_i}) = \theta\Big[S_y^2 + u_i^2 S_x^2 + 2\beta_{(ols)} u_i S_x^2 + \beta_{(ols)}^2 S_x^2 - 2u_i S_{xy} - 2\beta_{(ols)} S_{xy}\Big]$$

$$\text{for } i = 1, 2, ..., 9, \tag{1}$$

where $u_i = \dfrac{A_1\bar{Y}}{A_1\bar{X} + A_2}$ and $\theta = (\dfrac{1-f}{n})$ for $i = 1, 2, ..., 9$. Further, S_y^2 and S_x^2 are the unbiased variances, and S_{xy} is the co-variance of Y and X.

3. Proposed class of quantile regression-ratio-type estimators

Outliers are the observations in a data set that appear to be inconsistent with the rest of that data set. The presence of outliers significantly affects mean estimation, which is one of the most important measures of central tendency. Mean estimators using OLS regression coefficient are the most ideal choices for the estimation of population mean, i.e., $(\bar{Y})$. However, outliers may have a significant impact on the traditional regression coefficient calculated using OLS tool. Hence, the estimate of population mean $(\bar{y})$, based on OLS indicates poor performance. One of the solutions is to utilize quantile regression. It can be used as a robust approach in case data are non-normal and contaminated with outliers, because it is not sensitive to outliers [22–25]. Quantile regression is similar to customary OLS regression in a sense that both of them explore connections between endogenous and exogenous variables. The primary distinction is that OLS regression chooses parameter esteems that have the least squared deviation from the regression line as the parameter estimates, while quantile regression selects parameter esteems that have the least absolute deviation from the regression line as the parameter estimates. Thus, this study proposes a group of quantile regression-ratio-type estimators by extending the idea proposed by Abid et al. [21] as given below:

$$\bar{y}_{p_i} = \frac{\bar{y} + \hat{\beta}_{(q)}(\bar{X} - \bar{x})}{(A_1\bar{x} + A_2)}(A_1\bar{X} + A_2)$$

$$\text{for} \quad i = 1, 2, ..., 9,$$

with:

$$\hat{\beta}_{(q)} = argmin_{\beta\epsilon R^p} \ \rho_q(v) \sum_{i=1}^{n} (y_i - \langle x_i, \beta \rangle),$$

where $\rho_q(v)$ is a continuous piecewise linear function (or asymmetric absolute loss function) for quantile $q\epsilon$ (0,1), nondifferentiable at $v = 0$. Note that all the notations of $\bar{y}_{p_i}$ have usual meanings as discussed in the previous section. However, $\hat{\beta}_{(q)}$ is the quantile regression coefficient for $p = 2$ variables. For a deeper understanding of quantile regression, interested readers may refer to Koenker and Hallock [27]. The family members of $\bar{y}_{p_i}$ are provided in Table 2. The MSE of the proposed family of estimators is given below:

$$MSE(\bar{y}_{p_i}) = \theta[S_y^2 + u_i^2 S_x^2 + 2\beta_{(q)} u_i S_x^2 + \beta_{(q)}^2 S_x^2 - 2u_i S_{xy} - 2\beta_{(q)} S_{xy}]$$

$$\text{for} \quad i = 1, 2, ..., 9. \tag{2}$$

It is worth mentioning that we are using $q^{15\text{th}} = 0.15$, $q^{25\text{th}} = 0.25$, and $q^{35\text{th}} = 0.35$ quantiles for the purpose of the current article. Based on the consequences of the numerical study conducted in Section 5, utilizing the quantile regression coefficients based on these referenced quantiles enhances the efficiency of

Table 2. Family members of the proposed class.

Under complete information estimators	(A_1, A_2)	Under partial information estimators
$\bar{y}_{p_1}$	$(1, TM)$	$\bar{y}'_{p_1}$
$\bar{y}_{p_2}$	(C_x, TM)	$\bar{y}'_{p_2}$
$\bar{y}_{p_3}$	(ρ, TM)	$\bar{y}'_{p_3}$
$\bar{y}_{p_4}$	$(1, MR)$	$\bar{y}'_{p_4}$
$\bar{y}_{p_5}$	(C_x, MR)	$\bar{y}'_{p_5}$
$\bar{y}_{p_6}$	(ρ, MR)	$\bar{y}'_{p_6}$
$\bar{y}_{p_7}$	$(1, HL)$	$\bar{y}'_{p_7}$
$\bar{y}_{p_8}$	(C_x, HL)	$\bar{y}'_{p_8}$
$\bar{y}_{p_9}$	(ρ, HL)	$\bar{y}'_{p_9}$

the proposed estimators incredibly. Note that with these three referenced quantiles at hand, the proposed class contains twenty-seven members. For the sake of brevity and convenience, twenty-seven members of the proposed class along with their MSE values are provided in a compact form as follows:

$$\bar{y}_{p_i} = \begin{cases} \frac{\bar{y}+\hat{\beta}_{(0.15)}(\bar{X}-\bar{x})}{(A_1\bar{x}+A_2)}(A_1\bar{X}+A_2) \\ \quad \text{for } (q)=(0.15)=1,2,...,9 \\ \\ \frac{\bar{y}+\hat{\beta}_{(0.25)}(\bar{X}-\bar{x})}{(A_1\bar{x}+A_2)}(A_1\bar{X}+A_2) \\ \quad \text{for } (q)=(0.25)=1,2,...,9 \\ \\ \frac{\bar{y}+\hat{\beta}_{(0.35)}(\bar{X}-\bar{x})}{(A_1\bar{x}+A_2)}(A_1\bar{X}+A_2) \\ \quad \text{for } (q)=(0.35)=1,2,...,9 \end{cases} \tag{3}$$

$$MSE(\bar{y}_{p_i}) = \begin{cases} \theta[S_y^2+u_i^2S_x^2+2\beta_{(0.15)}u_iS_x^2+\beta_{(0.15)}^2S_x^2 \\ \quad -2u_iS_{xy}-2\beta_{(0.15)}S_{xy}] \\ \\ \theta[S_y^2+u_i^2S_x^2+2\beta_{(0.25)}u_iS_x^2+\beta_{(0.25)}^2S_x^2 \\ \quad -2u_iS_{xy}-2\beta_{(0.25)}S_{xy}] \\ \\ \theta[S_y^2+u_i^2S_x^2+2\beta_{(0.35)}u_iS_x^2+\beta_{(0.35)}^2S_x^2 \\ \quad -2u_iS_{xy}-2\beta_{(0.35)}S_{xy}] \end{cases} \tag{4}$$

4. Two-stage sampling scheme (partial information)

One can use the two-stage sampling exactly when the information on the population mean of supplementary variables is not available. Neyman [28] was the pioneer who gave the possibility of this sampling scheme in assessing the population parameters. Two-stage sampling is economically insightful and more straightforward, as well. This sampling plan is used to get the data about supplementary variables productively by choosing a more noteworthy sample from the first stage and to moderate the size sample at the second stage. Sukhatme [29] used a two-stage sampling plan to propose a general ratio-type estimator. Cochran [30] contributed to the discussion on two-stage sampling through a more comprehensive study.

Under a two-stage sampling plan, the first stage sample of size n_a units from the population of size N is selected with the assistance of SRSWOR plan. Then, the second stage sample, i.e., n_b, is selected as a subsample of the first stage sample, i.e., n_a.

4.1. Reviewed and proposed estimators in the two-stage sampling scheme

This sub-section presents Abid et al. [21] and the proposed family of estimators under two-stage sampling scheme as given by:

$$\bar{y}'_{ab_i} = \frac{\bar{y}_b+\hat{\beta}_{(ols)}(\bar{x}_a-\bar{x}_b)}{(A_1\bar{x}_b+A_2)}(A_1\bar{x}_a+A_2)$$

$$\text{for} \quad i=1,2,...,9,$$

$$\bar{y}'_{p_i} = \frac{\bar{y}_b+\hat{\beta}_{(q)}(\bar{x}_a-\bar{x}_b)}{(A_1\bar{x}_b+A_2)}(A_1\bar{x}_a+A_2)$$

$$\text{for} \quad i=1,2,...,9,$$

where $(\bar{x}_b, \bar{y}_b)$ represent the sample means at the second stage and $\bar{x}_a$ is the sample mean at first stage.

Further, A_1 and A_2 have the same meanings as described in the previous section. The family members of $\bar{y}'_{ab_i}$ and $\bar{y}'_{p_i}$ are available in Tables 1 and 2, respectively.

Utilizing Taylor series method, we are obtaining MSE for $\bar{y}'_{ab_i}$ as follows:

$$MSE(\bar{y}'_{ab_i}) = d_1\Sigma d_1', \tag{5}$$

where:

$$d_1 = \left[\frac{\delta h(\bar{y}_b,\bar{x}_a,\bar{x}_b)}{\delta\bar{y}_b}|_{\bar{Y},\bar{X}} \ \frac{\delta h(\bar{y}_b,\bar{x}_a,\bar{x}_b)}{\delta\bar{x}_a}|_{\bar{Y},\bar{X}} \ \frac{\delta h(\bar{y}_b,\bar{x}_a,\bar{x}_b)}{\delta\bar{x}_b}|_{\bar{Y},\bar{X}}\right],$$

$$d_1 = \left[1(u_i+\beta_{(ols)}) - (u_i+\beta_{(ols)})\right].$$

$$\sum = \begin{bmatrix} V(\bar{y}_b) & Cov(\bar{y}_b,\bar{x}_a) & Cov(\bar{y}_b,\bar{x}_b) \\ Cov(\bar{x}_a,\bar{y}_b) & V(\bar{x}_a) & Cov(\bar{x}_b,\bar{x}_a) \\ Cov(\bar{x}_b,\bar{y}_b) & Cov(\bar{x}_b,\bar{x}_a) & V(\bar{x}_b) \end{bmatrix}$$

where:

$$V(\bar{y}_b) = \lambda_2 S_y^2, \quad V(\bar{x}_a) = \lambda_1 S_x^2, \quad V(\bar{x}_b) = \lambda_2 S_y^2,$$

$$Cov(\bar{y}_b, \bar{x}_a) = Cov(\bar{x}_a, \bar{y}_b) = \lambda_1 S_{yx},$$

$$Cov(\bar{y}_b, \bar{x}_b) = Cov(\bar{x}_b, \bar{y}_b) = \lambda_2 S_{yx},$$

$$Cov(\bar{x}_a, \bar{x}_b) = Cov(\bar{x}_b, \bar{x}_a) = \lambda_1 S_x^2.$$

Utilizing these defined notations of variances and co-variances, and hence substituting the values of d_1 and Σ in Eq. (5), MSE expressions of $\bar{y}'_{ab_i}$ are as follows:

$$MSE(\bar{y}'_{ab_i}) = \lambda_2 S_y^2 + (\lambda_2 - \lambda_1)[(u_i + \beta_{(ols)})^2 S_x^2 - 2(u_i + \beta_{(ols)})S_{yx}],$$

$$\text{for} \quad i = 1, 2, ..., 9.$$

Replacing $\beta_{(ols)}$ with $\beta_{(q)}$, we can easily find the MSE of the proposed class of estimators as follows:

$$MSE(\bar{y}'_{p_i}) = \lambda_2 S_y^2 + (\lambda_2 - \lambda_1)[(u_i + \beta_{(q)})^2 S_x^2 - 2(u_i + \beta_{(q)})S_{yx}],$$

where $\lambda_1 = \left(\frac{1}{n_a} - \frac{1}{N}\right)$ and $\lambda_2 = \left(\frac{1}{n_b} - \frac{1}{N}\right)$.

The twenty seven family members of proposed class with their MSE in a compact form, are given under partial information as follows:

$$\bar{y}'_{p_i} = \begin{cases} \frac{\bar{y}_b + \hat{\beta}_{(0.15)}(\bar{x}_a - \bar{x}_b)}{(A_1\bar{x}_b + A_2)}(A_1\bar{x}_a + A_2) & \text{for } (q) = (0.15); \\ \quad i = 1, 2, ..., 9 & \\ \frac{\bar{y}_b + \hat{\beta}_{(0.25)}(\bar{x}_a - \bar{x}_b)}{(A_1\bar{x}_b + A_2)}(A_1\bar{x}_a + A_2) & \text{for } (q) = (0.25); \\ \quad i = 1, 2, ..., 9 & \\ \frac{\bar{y}_b + \hat{\beta}_{(0.35)}(\bar{x}_a - \bar{x}_b)}{(A_1\bar{x}_b + A_2)}(A_1\bar{x}_a + A_2) & \text{for } (q) = (0.35); \\ \quad i = 1, 2, ..., 9. & \end{cases} \tag{6}$$

$$MSE(\bar{y}'_{p_i}) = \begin{cases} \lambda_2 S_y^2 + (\lambda_2 - \lambda_1) \\ \quad [(u_i + \beta_{(0.15)})^2 S_x^2 - 2(u_i + \beta_{(0.15)})S_{yx}] \\ \lambda_2 S_y^2 + (\lambda_2 - \lambda_1) \\ \quad [(u_i + \beta_{(0.25)})^2 S_x^2 - 2(u_i + \beta_{(0.25)})S_{yx}] \\ \lambda_2 S_y^2 + (\lambda_2 - \lambda_1) \\ \quad [(u_i + \beta_{(0.35)})^2 S_x^2 - 2(u_i + \beta_{(0.35)})S_{yx}] \end{cases} \tag{7}$$

5. Numerical illustration

In this section, the performance of the proposed and existing estimators is assessed using three real life data sets as given below:

- **Population-1 (Pop-1):** The data set of Singh is used [31]. In this data set, "amount of non-real estate farm loans during 1977" is taken as the auxiliary variable (X), while "amount of real estate farm loans during 1977" is taken as the study variable (Y). Further, $N = 50$, $n = n_a = 20$, and $n_b = 16$.
- **Population-2 (Pop-2):** "UScereals" data set is used. It describes 65 commonly available breakfast cereals in the USA based on the information available on the mandatory food label on the packet. The measurements are normalized here to a serving size of one American cup. The data are derived from ASA Statistical Graphics Exposition and used by Venables and Ripley [32]. The data set contains a number of variables: "grams of fibre in one portion" as the auxiliary variable (X), while "grams of potassium" as the study variable (Y). Further, $N = 65, n = n_a = 20$, and $n_b = 15$.
- **Population-3 (Pop-3):** Third population is also derived from the "UScereals" data set. In this case, "grams of sodium in one portion" is taken as the auxiliary variable (X), while "number of calories" is taken as the study variable (Y). Further, $N = 65, n = n_a = 20$, and $n_b = 15$.

The remaining characteristics of all the three populations are provided in Table 3. Further, histogram, box-plot, and scatter-plot are drawn for all the three referenced populations; see Figures 1–9. These figures show non-normality and presence of outliers, hence suitable for utilization of non-conventional measures of location similar to the research application done by Abid et al. [21] and for quantile regression, too.

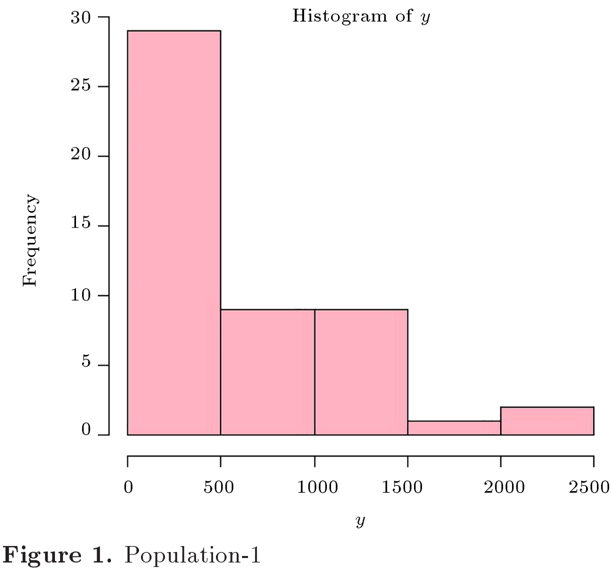

Figure 1. Population-1

Table 3. Characteristics of populations.

(Pop-1)	$\bar{Y} = 555.4345$	$\bar{X} = 878.1624$	$HL = 628.0575$
	$S_y = 584.826$	$S_x = 1084.678$	$MR = 1964.483$
	$\rho = 0.8038$	$S_{xy} = 509910.4$	$TM = 536.4088$
	$C_x = 1.235168$	$\beta_{(ols)} = 0.4334034$	$\beta_{(0.25)} = 0.3542911$
	$C_y = 1.052916$	$\beta_{(0.15)} = 0.3447828$	$\beta_{(0.35)} = 0.3570439$
(Pop-2)	$\bar{Y} = 159.1197$	$\bar{X} = 3.870844$	$HL = 2.5$
	$S_y = 180.2886$	$S_x = 6.133404$	$MR = 15.15151$
	$\rho = 0.9638$	$S_{xy} = 1065.827$	$TM = 2.119403$
	$C_x = 1.584513$	$\beta_{(ols)} = 28.3324$	$\beta_{(0.25)} = 27.175$
	$C_y = 1.133037$	$\beta_{(0.15)} = 22.77778$	$\beta_{(0.35)} = 27.72727$
(Pop-3)	$\bar{Y} = 149.4083$	$\bar{X} = 237.8384$	$HL = 235.1244$
	$S_y = 62.41187$	$S_x = 130.6296$	$MR = 393.9394$
	$\rho = 0.5286$	$S_{xy} = 4310.041$	$TM = 233.5$
	$C_x = 0.549237$	$\beta_{(ols)} = 0.2525795$	$\beta_{(0.25)} = 0.1737113$
	$C_y = 0.4177271$	$\beta_{(0.15)} = 0.198324$	$\beta_{(0.35)} = 0.1578947$

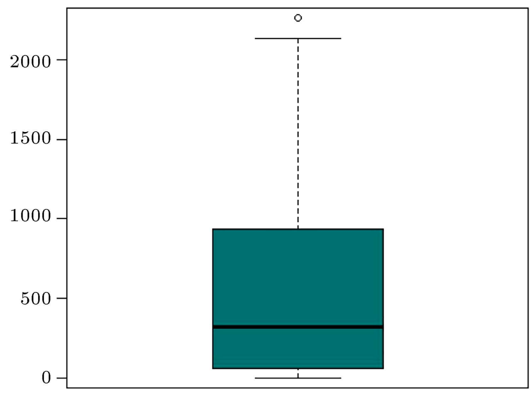

Figure 2. Box plot of population-1

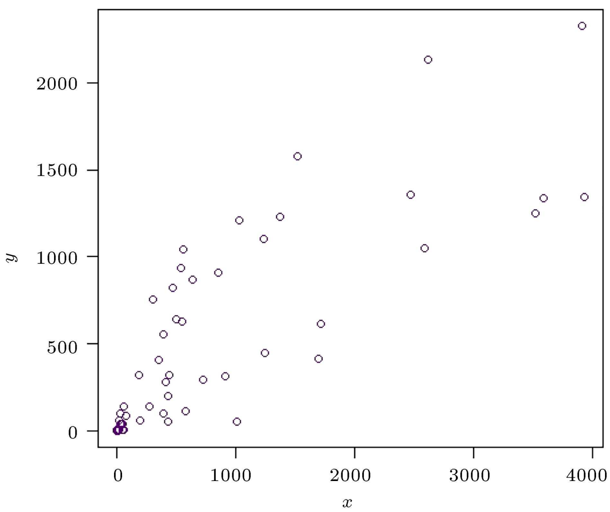

Figure 3. Scatter plot of population-1

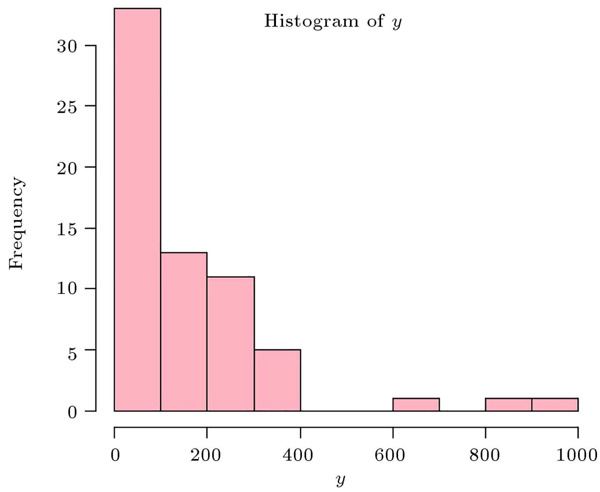

Figure 4. Population-2

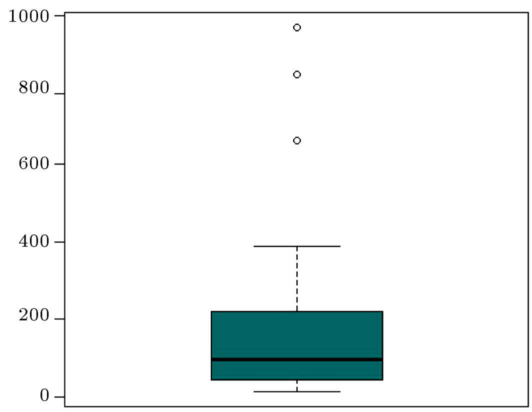

Figure 5. Box plot of population-2

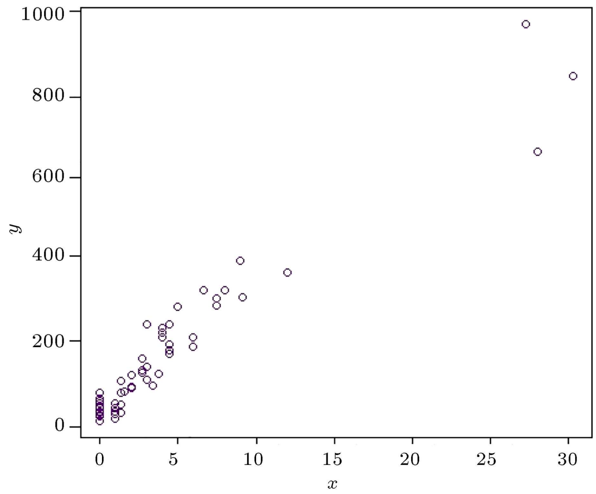

Figure 6. Scatter plot of population-2

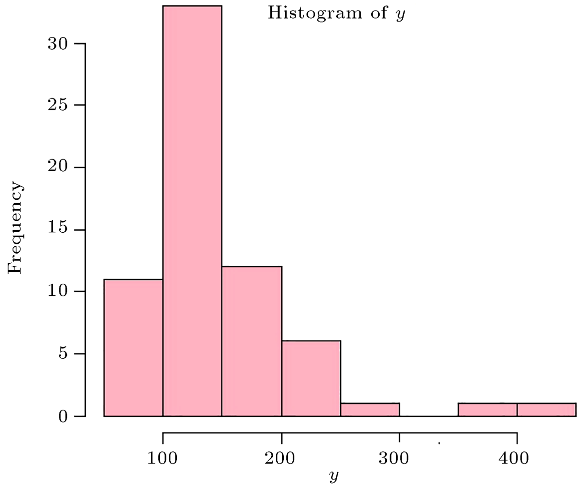

Figure 7. Population-3

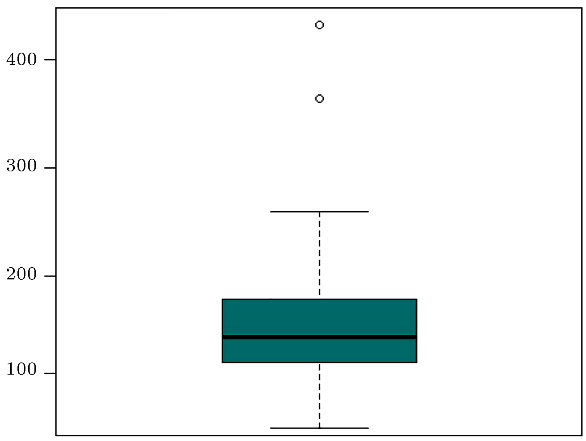

Figure 8. Box plot of population-3

Each existing estimator is based on $\beta_{(ols)}$. Each proposed estimator is based on $\beta_{(0.15)}$, $\beta_{(0.25)}$, and $\beta_{(0.35)}$. Therefore, results of each existing estimator with their corresponding proposed estimators, say $\hat{\theta}$, are provided on the same row, including $\hat{\theta} = (\beta_{(ols)}, \beta_{(0.15)}, \beta_{(0.25)}, \beta_{(0.35)})$, in Tables 4–6, under complete information. Similarly, Tables 4–6 show PRE of existing and proposed estimators, say $\hat{\theta}'$, under partial information. Further, Figures 10–12 presents MSE of $(\hat{\theta}, \hat{\theta}_0)$ graphically.

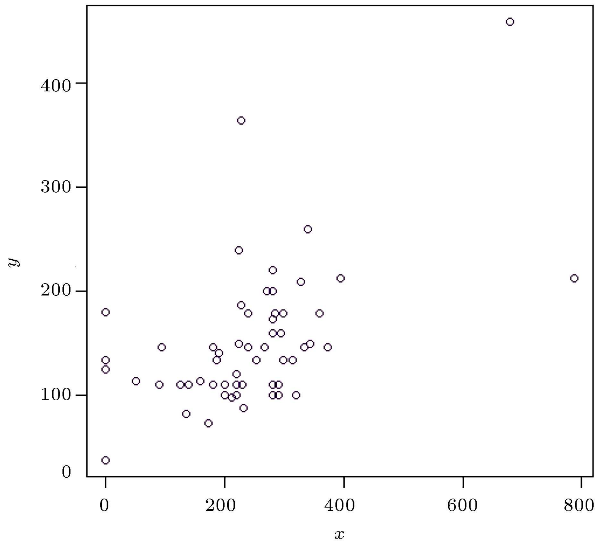

Figure 9. Scatter plot of population-3

5.1. Results

The consequences of the numerical study are given in Tables 4-6 under complete and partial information settings. From Tables 4-6, it is clear that every one of the estimators $\bar{y}_{pi}$, with $i = 1, \ldots, 9$, considered in the proposed class, beats their counterparts in the existing class, i.e., $\bar{y}_{abi}$. Also, it is clear that the proposed estimators are more productive than the existing estimators in case of partial information. From Figures 10-12, the proposed estimators have smaller MSE values than their corresponding ones. Hence, in both cases, the proficiency increase yielded by the new estimators is striking in Populations 1-3.

6. Final remarks

Beginning from some ongoing contributions of Abid et al. [1] concerning mean estimation under design/configuration-based sampling, this study constructed a class of quantile regression-ratio-type estimators for the population mean whenever the data is non-normal and contaminated with outliers. The proposed class beats existing ones. In case of complete information where population mean of auxiliary variable is known, the MSE expressions were determined theoretically. Further, the existing and proposed classes were introduced to the case of partial information. In the article, three real-life populations were considered. The numerical computations well underscored the predominance of the proposed class in both complete and partial information settings, at

Table 4. PRE of the proposed and existing estimators for Pop-1.

	Under complete information					Under partial information			
$\hat{\boldsymbol{\theta}}$	$\beta_{(ols)}$	$\beta_{(0.15)}$	$\beta_{(0.25)}$	$\beta_{(0.35)}$	$\hat{\boldsymbol{\theta}}'$	$\beta_{(ols)}$	$\beta_{(0.15)}$	$\beta_{(0.25)}$	$\beta_{(0.35)}$
$\hat{\theta}_1$	3769.87	4961.64	4816.81	4775.65	$\hat{\theta}'_1$	2435.90	2604.31	2587.30	2582.32
$\hat{\theta}_2$	3436.58	4510.95	4379.39	4342.06	$\hat{\theta}'_2$	2373.92	2548.62	2530.73	2525.50
$\hat{\theta}_3$	4176.24	5500.01	5341.16	5295.90	$\hat{\theta}'_3$	2501.43	2661.29	2645.43	2640.77
$\hat{\theta}_4$	6870.27	8480.31	8325.77	8269.79	$\hat{\theta}'_4$	2772.78	2864.23	2856.77	2854.50
$\hat{\theta}_5$	6313.21	7977.74	7805.28	7754.63	$\hat{\theta}'_5$	2732.24	2839.06	2829.79	2827.01
$\hat{\theta}_6$	7407.26	8886.77	8759.47	8720.64	$\hat{\theta}'_6$	2807.01	2882.79	2877.13	2875.38
$\hat{\theta}_7$	4056.99	5343.50	5188.49	5144.36	$\hat{\theta}'_7$	2483.21	2645.66	2629.46	2624.70
$\hat{\theta}_8$	3680.05	4840.89	4699.50	4659.34	$\hat{\theta}'_8$	2420.02	2590.18	2572.93	2567.88
$\hat{\theta}_9$	4509.32	5929.52	5761.35	5713.30	$\hat{\theta}'_9$	2548.41	2700.73	2685.85	2681.47

Table 5. PRE of the proposed and existing estimators for Pop-2.

	Under complete information					Under partial information			
$\hat{\boldsymbol{\theta}}$	$\beta_{(ols)}$	$\beta_{(0.15)}$	$\beta_{(0.25)}$	$\beta_{(0.35)}$	$\hat{\boldsymbol{\theta}}'$	$\beta_{(ols)}$	$\beta_{(0.15)}$	$\beta_{(0.25)}$	$\beta_{(0.35)}$
$\hat{\theta}_1$	3254.74	4965.69	3531.74	3395.46	$\hat{\theta}'_1$	2023.94	2256.74	2072.61	2049.36
$\hat{\theta}_2$	2509.54	3638.07	2697.71	2605.44	$\hat{\theta}'_2$	1858.68	2089.86	1906.10	1883.40
$\hat{\theta}_3$	3334.54	5113.07	3621.65	3480.35	$\hat{\theta}'_3$	2038.54	2271.03	2087.25	2063.99
$\hat{\theta}_4$	19012.92	36065.05	22038.47	20539.70	$\hat{\theta}'_4$	2691.95	2781.61	2717.38	2705.66
$\hat{\theta}_5$	12379.90	24741.64	14219.13	13301.50	$\hat{\theta}'_5$	2597.09	735.12	2631.47	2615.39
$\hat{\theta}_6$	19610.21	36757.77	22728.23	21185.38	$\hat{\theta}'_6$	2697.55	2783.55	2722.82	2710.90
$\hat{\theta}_7$	3643.28	5692.21	3970.58	3809.31	$\hat{\theta}'_7$	2090.69	2321.58	2139.42	2116.18
$\hat{\theta}_8$	2730.46	4022.14	2943.84	2839.09	$\hat{\theta}'_8$	1913.91	2146.56	1961.90	1938.94
$\hat{\theta}_9$	3741.34	5879.02	4081.73	3913.95	$\hat{\theta}'_9$	2105.94	2336.16	2154.64	2131.47

Table 6. PRE of the proposed and existing estimators for Pop-3.

	Under complete information					Under partial information			
$\hat{\boldsymbol{\theta}}$	$\beta_{(ols)}$	$\beta_{(0.15)}$	$\beta_{(0.25)}$	$\beta_{(0.35)}$	$\hat{\boldsymbol{\theta}}'$	$\beta_{(ols)}$	$\beta_{(0.15)}$	$\beta_{(0.25)}$	$\beta_{(0.35)}$
$\hat{\theta}_1$	2488.90	2824.16	2981.56	3083.08	$\hat{\theta}'_1$	1853.20	1935.58	1969.90	1990.75
$\hat{\theta}_2$	3063.47	3403.58	3546.67	3632.31	$\hat{\theta}'_2$	1986.79	2050.79	2075.07	2088.95
$\hat{\theta}_3$	3098.65	3436.36	3577.02	3660.63	$\hat{\theta}'_3$	1993.86	2056.48	2080.05	2093.43
$\hat{\theta}_4$	2992.03	3335.88	3483.27	3572.64	$\hat{\theta}'_4$	1972.09	2038.78	2064.49	2079.33
$\hat{\theta}_5$	3490.14	3770.11	3867.94	3918.58	$\hat{\theta}'_5$	2065.65	2110.31	2124.79	2132.08
$\hat{\theta}_6$	3515.64	3789.35	3883.19	3930.94	$\hat{\theta}'_6$	2069.93	2113.20	2127.01	2133.83
$\hat{\theta}_7$	2495.39	2831.08	2988.55	3090.05	$\hat{\theta}'_7$	1854.93	1937.14	1971.36	1992.15
$\hat{\theta}_8$	3069.89	3409.59	3552.25	3637.53	$\hat{\theta}'_8$	1988.09	2051.83	2075.99	2089.78
$\hat{\theta}_9$	3104.98	3442.22	3582.43	3665.65	$\hat{\theta}'_9$	1995.12	2057.49	2080.93	2094.22

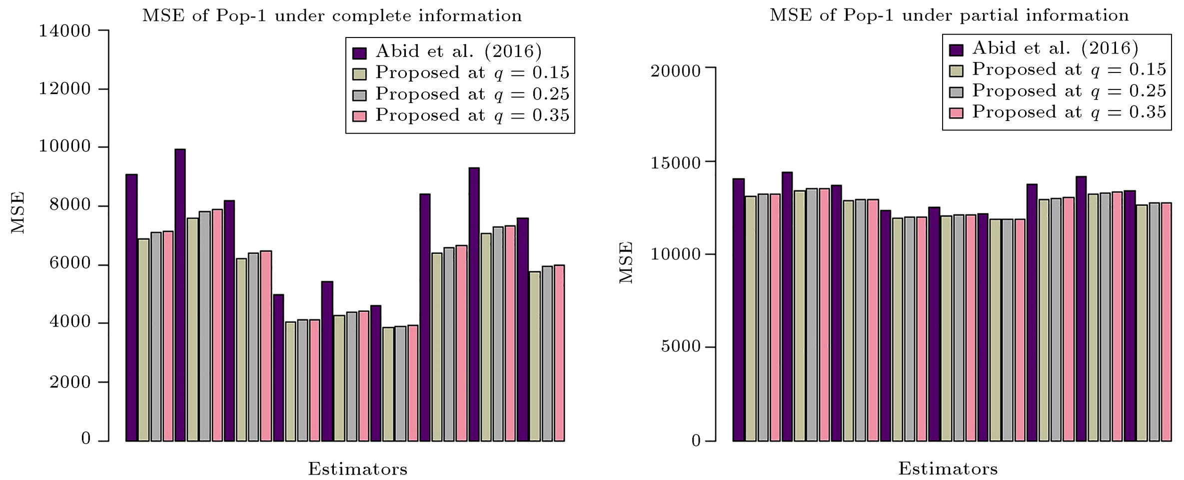

Figure 10. MSE of population-1

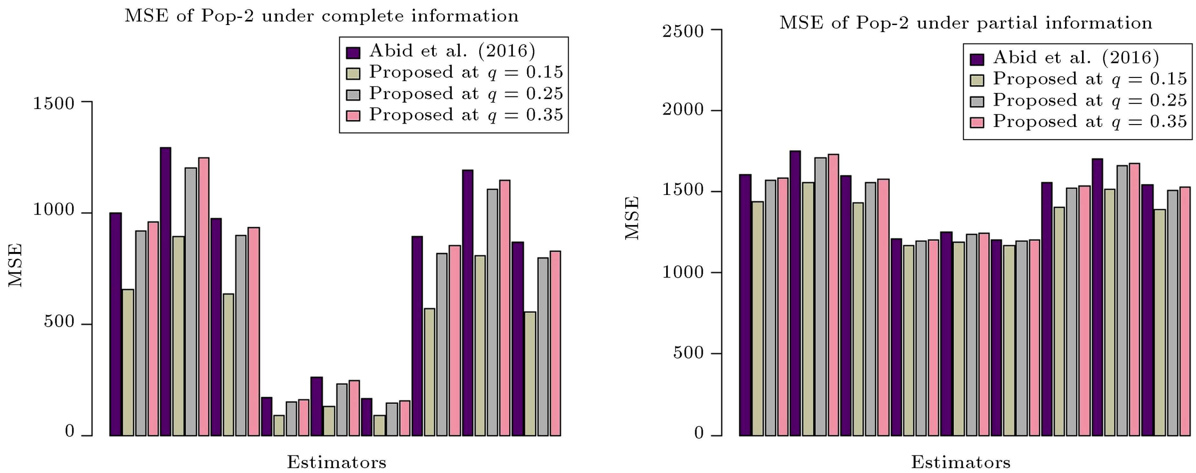

Figure 11. MSE of population-2

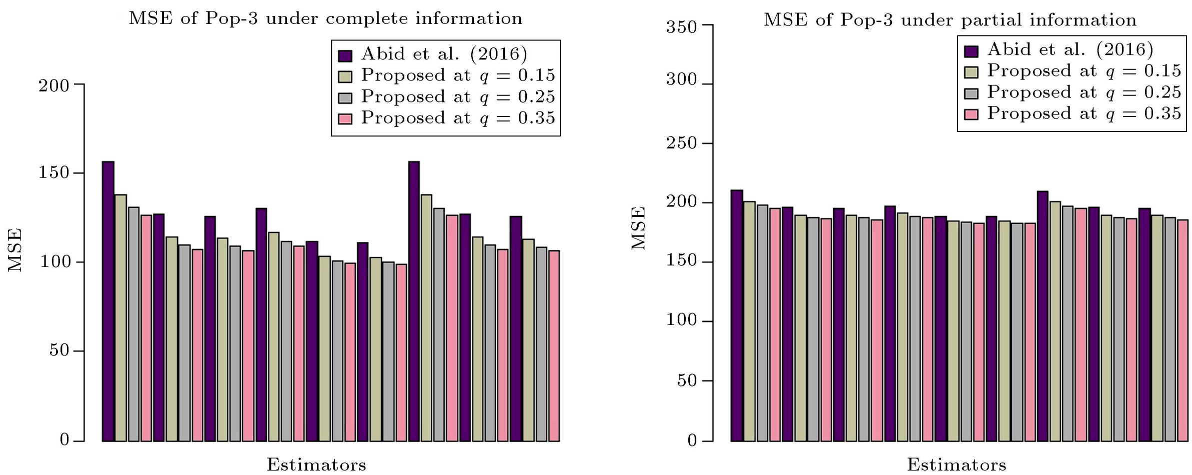

Figure 12. MSE of population-3

least for the considered experimental circumstances in the article. Hence, survey practitioners should use the proposed class since it might expand the odds of getting progressively proficient evaluations of obscure objective parameters.

References

1. McClellan, M., McNeil, B.J. and Newhouse, J.P. "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: Analysis

using instrumental variables", *Jama*, **272**(11), pp. 859–866 (1994).

2. McIntyre, G.A. "A method of unbiased selective sampling using ranked sets", *Australian Journal of Agricultural Research*, **3**(4), pp. 385–390 (1952).

3. Adel Rastkhiz, S.E., MobiniDehkordi, A., Yadollahi Farsi, J., et al. "A new approach to evaluating entrepreneurial opportunities", *Journal of Small Business and Enterprise Development* (2019) In Press.

4. Zamanzade, E. and Vock, M. "Variance estimation in ranked set sampling using a concomitant variable", *Statistics and Probability Letters*, **105**, pp. 1–5 (2015).

5. Zamanzade, E. and Wang, X. "Estimation of population proportion for judgment post-stratification", *Computational Statistics and Data Analysis*, **112**, pp. 257–269 (2017).

6. Zamanzade, E. and Mahdizadeh, M. "A more efficient proportion estimator in ranked set sampling", *Statistics and Probability Letters*, **129**, pp. 28–33 (2017).

7. Zamanzade, E. and Wang, X. "Proportion estimation in ranked set sampling in the presence of tie information", *Computational Statistics*, **33**(3), pp. 1349–1366 (2018).

8. Mahdizadeh, M. and Zamanzade, E. "Efficient body fat estimation using multistage pair ranked set sampling", *Statistical Methods in Medical Research*, **28**(1), pp. 223–234 (2019).

9. Koyuncu, N. "Efficient estimators of population mean using auxiliary attributes", *Applied Mathematics and Computation*, **218**, pp. 10900–10905 (2012).

10. Shahzad, U., Perri P.F., and Hanif, M. "A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information", *Communications in Statistics-Simulation and Computation*, **48**(9), pp. 2566–2585 (2019).

11. Hanif, M. and Shahzad, U. "Estimaton of population variance using kernel matrix", *Journal of Statistics and Management Systems*, **22**(3), pp. 563–586 (2019).

12. Abid, M., Abbas, N., and Riaz, M. "Improved modified ratio estimators of population mean based on deciles", *Chiang Mai Journal of Science*, **43**(1), pp. 1311–1323 (2016a).

13. Irfan, M., Javed, M., Abid, M., et al. "Improved ratio type estimators of population mean based on median of a study variable and an auxiliary variable", *Hacettepe Journal of Mathematics and Statistics*, **47**(3), pp. 659–673 (2018).

14. Naz, F., Abid, M., Nawaz, T., et al. "Enhancing the efficiency of the ratio-type estimators of population variance with a blend of information on robust location measures", *Scientia Iranica*, **27**(4), pp. 2040–2056 (2020).

15. Abid, M., Ahmed, S., Tahir, M., et al. "Improved ratio estimators of variance based on robust measures", *Scientia Iranica*, **26**(4), pp. 2484–2494 (2019).

16. Zaman, T., Saglam, V., Sagir, M., et al. "Investigation of some estimators via Taylor series approach and an application", *American Journal of Theoretical and Applied Statistics*, **3**(5), pp. 141–147 (2014).

17. Zaman, T. and Bulut, H. "Modified ratio estimators using robust regression methods", *Communications in Statistics-Theory and Methods*, **48**(8), pp. 2039–2048 (2019).

18. Zaman, T. and Bulut, H. "Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling", *Communications in Statistics-Theory and Methods*, (2019) In Press.

19. Zaman, T. "Improvement of modified ratio estimators using robust regression methods", *Applied Mathematics and Computation*, **348**, pp. 627–631 (2019).

20. Ali, N., Ahmad, I., Hanif, M., and Shahzad, U. "Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information", *Communications in Statistics-Theory and Methods*, **50**(4), pp. 979–992 (2021).

21. Abid, M., Abbas, N., Zafar Nazir, H., et al. "Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters", *Revista Colombiana de Estadistica*, **39**(1), pp. 63–79 (2016b).

22. Hao, L., Naiman, D.Q., and Naiman, D.Q., *Quantile Regression*, Sage (2007).

23. Moore, D.S. and Kirkland, S., *The Basic Practice of Statistics*, **2**, New York: WH Freeman (2007).

24. Pedhazur, E.J., *Multiple Regression in Behavioral Research: Explanation and Prediction*, Thompson Learning Inc: New York (1997).

25. Koenker, R., it Quantile Regression, Cambridge University Press, New York (2005).

26. Koenker, R. and Bassett Jr, G. "Regression quantiles", *Econometrica: Journal of the Econometric Society*, pp. 33–50 (1978).

27. Koenker, R. and Hallock, K.F. "Quantile regression", *Journal of Economic Perspectives*, **15**(4), pp. 143–156 (2001).

28. Neyman, J. "Contribution to the theory of sampling human populations", *Journal of the American Statistical Association*, **33**(201), pp. 101–116 (1938).

29. Sukhatme, B.V. "Some ratio-type estimators in two-phase sampling", *Journal of the American Statistical Association*, **57**(299), pp. 628–632 (1962).

30. Cochran, W.G., *Sampling Techniques*, 3rd Ed., John Wiley and Sons, New York (1977).

31. Singh, S., *Advanced Sampling Theory with Applications. How Michael Selected Amy*, Kluwer Academic Publishers: Dordrecht (2003).

32. Venables, W.N. and Ripley, B.D., *Modern Applied Statistics with S-PLUS*, 3rd Ed., Springer (1999).

Biographies

Usman Shahzad obtained his MSc degree in Statistics from International Islamic University, Islamabad, Pakistan. He completed his MPhil in Statistics from Pir Mehr Ali Shah Arid Agriculture University, Rawalpinid, Pakistan. Currently, he is studying his PhD in Statistics from International Islamic University, Islamabad, Pakistan. He served as a lecturer at Pir Mehr Ali Shah Arid Agriculture University, Rawalpinid, Pakistan. He has published more than 25 research papers in research journals. His research interests include survey sampling, extreme value theory, stochastic process, probability, and non-parametric statistics.

Muhammad Hanif obtained his MSc degree in Statistics from University of Agriculture, Faisalabad, Pakistan. He completed his MPhil in Statistics from Government College University, Lahore, Pakistan. He completed his PhD degree in Statistics from Zhejiang University, Hangzhou, China. Currently, he is an Associate Professor and serving as the Chairman of the Department of Mathematics and Statistics at Pir Mehr Ali Shah Arid Agriculture University, Rawalpinid, Pakistan. He has published more than 50 research papers in research journals. His main research interests include stochastic process, probability and mathematical statistics, survey sampling, and non-parametric estimation.

Irsa Sajjad completed her MSc and MPhil in Statistics from Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan. She is serving as a lecturer at the Department of Lahore Business School-University of Lahore, Islamabad, Pakistan. Her research interests include sampling, stochastic process, methods, and probability.

Malik Muhammad Anas completed his MSc and MPhil in Statistics from Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan. Currently, he is studying his PhD in Statistics from Nanjing University of Science and Technology, Nanjing, China. He served as a lecturer at Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan. His research interests include non-parametric regression, stochastic process, kernel smoothing, and sampling methods.