Communications for Statistical Applications and Methods
2024, Vol. 31, No. 3, 291–308

https://doi.org/10.29220/CSAM.2024.31.3.291
Print ISSN 2287-7843 / Online ISSN 2383-4757

Some efficient ratio-type exponential estimators using the Robust regression's Huber M-estimation function

Vinay Kumar Yadav[1,a b], Shakti Prasad[c]

[a]Department of Basic and Applied Science, National Institute of Technology Arunachal Pradesh, India;
[b]Department of Mathematics, School of Computational and Applied Sciences, Brainware University, India;
[c]Department of Mathematics, National Institute of Technology Jamshedpur, India

Abstract

The current article discusses ratio type exponential estimators for estimating the mean of a finite population in sample surveys. The estimators uses robust regression's Huber M-estimation function, and their bias as well as mean squared error expressions are derived. It was campared with Kadilar, Candan, and Cingi (Hacet J Math Stat, 36, 181–188, 2007) estimators. The circumstances under which the suggested estimators perform better than competing estimators are discussed. Five different population datasets with a well recognized outlier have been widely used in numerical and simulation-based research. These thorough studies seek to provide strong proof to back up our claims by carefully assessing and validating the theoretical results reported in our study. The estimators that have been proposed are intended to significantly improve both the efficiency and accuracy of estimating the mean of a finite population. As a result, the results that are obtained from statistical analyses will be more reliable and precise.

Keywords: ratio type exponential estimator, mean squared error (MSE), Huber M function, Robust regression, auxiliary variable, percent relative efficiency

1. Introduction

In statistics, estimation is the method of estimating an unknown population parameter utilising sample data. Determining the population mean using a sample of data is a frequent example of this. A single value called a point estimator, which is used for estimating the population parameter, is one method of estimation. Unreliable or biassed estimates may result from point estimators' sensitivity to outliers or other odd observations in the data.

A ratio estimator, that employs the ratio of two variables that are present in the sample data to determine the ratio of the associated population parameters, is one approach to overcoming this issue. Whenever the correlation coefficient between the variable being studied and an auxiliary variable is positive, it indicates that the two variables typically vary together, and this relationship may be used to increase estimate accuracy. By modifying the estimate of the study variable with the auxiliary variable, a more precise and effective estimate of the population mean may be achieved.

In order to improve the performance of the ratio estimator even further, information about the auxiliary variable may be integrated into the estimating process. For simplicity, the estimator may

[1]Corresponding author: Department of Basic & Applied Science, National Institute of Technology, Arunachal Pradesh, Jote, Papum pare-791113, India. E-mail: vkyadavbhu@gmail.com

Published 31 May 2024 / journal homepage: http://csam.or.kr

be modified based on the information about the auxiliary variable's unpredictability in relation to the mean considering the coefficient of variation, and the estimator could be modified based on the auxiliary variable's distributional shape considering the coefficient of kurtosis. Now, the ratio estimator may be tailored to the unique properties of the data by taking into consideration these other variables, which can result in even more precise and effective estimations of the unknown population parameter. Overall, ratio estimators are considered an excellent method for determining population characteristics from sample observations, and they may be particularly useful when there is a strong positive correlation between the study and auxiliary variables. Notably, it's important to exercise precautions when utilising any estimating technique and to take into account any potential restrictions and underlying presumptions that might be present.

The potential benefits of incorporating data on the auxiliary variable to improve the performance of ratio estimators have been discovered by several statisticians.

A number of studies made use of this approach, for example Kadilar and Cingi (2004), Kadilar *et al.* (2007), Noor-ul-Amin *et al.* (2016, 2018, and 2022), Prasad (2020), Zaman (2020, 2021), and Zaman and Kadilar (2021a, 2021b). By using additional information from the auxiliary variables, these investigations are able to increase the ratio estimators' precision and accuracy, which has significant consequences for the validity and precision of statistical research. A cutting-edge and well-liked technique for handling outliers in datasets is the use of robust regression algorithms. These techniques were created to reduce the impact of outliers on regression analysis while still taking into account a significant portion of the data points. There are several strategies to deal with outliers, including changing the data, using Winsorization, and locating and erasing outliers. The unique features of the data and the objectives of the study influence the methodology used.

The L_1 criteria for this purpose was initially proposed by Edgworth (1887), according to sources. The most widely employed M-estimator is afterwards suggested by Huber (1973). When compared to the LS estimator, the Huber M-estimator has the additional advantage of not being as sensitive to outliers. When it comes to handling data outliers, Huber M-estimation is a more reliable approach than LS estimation. As a result, it is frequently used in circumstances wherein outliers might be present. The Huber M-estimator makes use of the function $\rho(\epsilon)$, which strikes a balance between ϵ^2 and $|\epsilon|$. Here, ϵ refers to the error term from the linear regression model $y = a + bx + \epsilon$, and an is the model constant. The following are the parameters that the Huber $\rho(\epsilon)$ function accepts:

$$\rho(\epsilon) = \begin{cases} \epsilon^2; & -l \leq \epsilon \leq l, \\ l|\epsilon| - l^2; & \epsilon < -l \quad \text{or} \quad \epsilon < l. \end{cases} \tag{1.1}$$

The tuning constant l is a parameter that affects the level of robustness of the estimator used in statistical analysis. By adjusting the value of l, one can control the sensitivity of the estimator to outliers and other unusual observations in the data. A larger value of l makes the estimator more robust, while a smaller value of l makes the estimator more sensitive to outliers. The appropriate value of l depends on the specific characteristics of the data and the goals of the analysis, and may need to be determined through experimentation or other means.

The formula $l = 1.5\hat{\sigma}$, was developed by Huber (1981), where $\hat{\sigma}$ is an estimate of the SD σ of the population's random errors. For additional details on constant l and M-estimators, see Rousseeuw and Leroy (1987). The regression coefficient, which is calculated by minimising, is $\hat{\beta}_{hm}$.

$$\sum_{i}^{n} \rho\left(y_i - a - bx_i\right) \tag{1.2}$$

with respect to a and b.

Huber (1981) developed the M-estimators method inside the robust regression framework to deal with outliers. By developing ratio estimators that incorporate Huber's M-estimator, which has been shown to be more successful in delivering accurate and reliable results in the presence of outliers, Kadilar *et al.* (2007) furthered this method. Essentially, these methods enable the removal of outlier effects, improving the accuracy and robustness of regression models. In order to lessen the negative impacts of the outlier data, we explore employing the Huber M-estimation function in this paper's ratio type exponential estimators.

To study more about robust regression see quantreg (Koenker, 2009) package in R-Software (2021), Zaman and Bulut (2019, 2021), Zaman *et al.* (2021, 2022), and Bulut and Zaman (2022).

In Section 2, we consider the existing estimators using robust regression. In Section 3, we discuss new ratio type exponential estimators based on the Huber M-estimation function, as well as their MSEs. Section 4 provides efficiency comparisons of the existing and considered estimators based on the expression of MSEs. Sections 5 and 6 present the results of the numerical illustration and simulation study, respectively. In the final section, we draw a conclusion based on these results.

2. Existing ratio estimators

Kadilar *et al.* (2007) explored ratio estimators $\bar{y}_{cki}$, (i = 1, 2, 3, 4, 5) for estimating finite population mean $\bar{Y}$ using robust regression is given as

$$\bar{y}_{ck1} = \frac{\bar{y} + \hat{\beta}_{hm}(\bar{X} - \bar{x})}{\bar{x}} \bar{X}. \tag{2.1}$$

$$\bar{y}_{ck2} = \frac{\bar{y} + \hat{\beta}_{hm}(\bar{X} - \bar{x})}{\bar{x} + C_x} \left(\bar{X} + C_x\right). \tag{2.2}$$

$$\bar{y}_{ck3} = \frac{\bar{y} + \hat{\beta}_{hm}(\bar{X} - \bar{x})}{\bar{x} + \beta_2(x)} \left[\bar{X} + \beta_2(x)\right]. \tag{2.3}$$

$$\bar{y}_{ck4} = \frac{\bar{y} + \hat{\beta}_{hm}(\bar{X} - \bar{x})}{\bar{x}\beta_2(x) + C_x} \left[\bar{X}\beta_2(x) + C_x\right]. \tag{2.4}$$

$$\bar{y}_{ck5} = \frac{\bar{y} + \hat{\beta}_{hm}(\bar{X} - \bar{x})}{\bar{x}C_x + \beta_2(x)} \left[\bar{X}C_x + \beta_2(x)\right], \tag{2.5}$$

where C_x and $\beta_2(x)$ are the auxiliary variable's population coefficients of variation and kurtosis, respectively; $\bar{y}$ and $\bar{x}$ are the study and auxiliary variable's sample means, respectively, and it is assumed that the population mean $\bar{X}$ is known. In robust regression, Huber M-estimation function are utilised to calculate $\hat{\beta}_{hm}$.

Using a first-degree-approximation expansion, the MSEs of the estimators (1)–(5) can be calculated as follows:

$$\mathrm{MSE}\left(\bar{y}_{cki}\right) = \frac{1-f}{n} \left(R_{cki}^2 S_x^2 + 2\beta_{hm} R_{cki} S_x^2 + \beta_{hm}^2 S_x^2 - 2R_{cki} S_{xy} - 2\beta_{hm} S_{xy} + S_y^2\right), \tag{2.6}$$

where $i = 1, 2, 3, 4, 5.$; $f = n/N$; n is size of sample; N is the size of population;

$R_{ck1} = \bar{Y}/\bar{X}$, $R_{ck2} = \bar{Y}/(\bar{X} + C_x)$, $R_{ck3} = \bar{Y}/(\bar{X} + \beta_2(x))$, $R_{ck4} = \bar{Y}\beta_2(x)/(\bar{X}\beta_2(x) + C_x)$, $R_{ck5} = \bar{Y}C_x/(\bar{X}C_x + \beta_2(x))$ are the population ratios; the variances of the study and auxiliary variables are S_y^2 and S_x^2, respectively, while the covariance between the study and auxiliary variable is S_{xy}.

3. Mathematical formulation of suggested ratio type exponential estimators

In recent years, the use of robust statistical approaches in sampling studies and finite population mean estimates has received a lot of attention. The necessity to improve the effectiveness and accuracy of predicting the population mean in the context of outliers served as the primary motivation for establishing the development and study of the ratio type exponential estimators discussed in this article. Sample surveys are an important method for deriving conclusions regarding finite populations. However, when outliers are present in the data, which might have a disproportionate impact on traditional estimators, their dependability may be compromised. Robust estimating strategies are now being investigated as a result of this constraint.

The suggested estimators are based on the Huber M-estimation function of the robust regression. The objective was to develop estimators that are capable of handling the disruptive impacts of outliers while providing more accurate estimations of the population mean is what contributed to this conclusion. We intend to reduce the possible biases and inefficiencies associated with conventional estimators in the context of unusual data points by utilising the adaptive characteristics of Huber M-estimation.

In this paper, we formulate mathematical equations for the bias and the mean squared error of the suggested estimators, which enable a thorough evaluation of their performance. We contrast these estimators' performance with that of those developed by Kadilar, Candan, and Cingi (Hacet J Math Stat, 36, 181–188, 2007) in order to assess their efficacy. To determine the scenarios in which the recommended estimators perform better than current techniques, a comparison study is necessary. Furthermore, real-world application is essential, thus we undertake numerical and simulation-based investigations employing five population datasets, each of which contains an outlier. These empirical studies support our theoretical conclusions and offer perceptions on how the suggested estimators actually work in real-world situations.

We presented ratio type exponential estimators employing the Huber (1981) M function, inspired by the work of Kadilar *et al.* (2007) and Prasad (2020). The suggested estimators can produce effective results even when there are outliers. The following estimators are recommended for estimating the population mean:

$$\bar{y}_{sv1} = \left[\bar{y} + \hat{\beta}_{hm}\left(\bar{X} - \bar{x}\right)\right] \exp\left[\frac{\left(\bar{X} - \bar{x}\right)}{\left(\bar{X} + \bar{x}\right)}\right]. \tag{3.1}$$

$$\bar{y}_{sv2} = \left[\bar{y} + \hat{\beta}_{hm}\left(\bar{X} - \bar{x}\right)\right] \exp\left[\frac{\left(\bar{X} - \bar{x}\right)}{\left(\bar{X} + \bar{x}\right) + 2C_x}\right]. \tag{3.2}$$

$$\bar{y}_{sv3} = \left[\bar{y} + \hat{\beta}_{hm}\left(\bar{X} - \bar{x}\right)\right] \exp\left[\frac{\left(\bar{X} - \bar{x}\right)}{\left(\bar{X} + \bar{x}\right) + \beta_2(x)}\right]. \tag{3.3}$$

$$\bar{y}_{sv4} = \left[\bar{y} + \hat{\beta}_{hm}\left(\bar{X} - \bar{x}\right)\right] \exp\left[\frac{\beta_2(x)\left(\bar{X} - \bar{x}\right)}{\beta_2(x)\left(\bar{X} + \bar{x}\right) + 2C_x}\right]. \tag{3.4}$$

$$\bar{y}_{sv5}=\left[\bar{y}+\hat{\beta}_{hm}\left(\bar{X}-\bar{x}\right)\right]\exp\left[\frac{C_x\left(\bar{X}-\bar{x}\right)}{C_x\left(\bar{X}+\bar{x}\right)+2\beta_2(x)}\right]. \tag{3.5}$$

To calculate the mean square error (MSE) of the suggested estimators $\bar{y}_{sv1}$, $\bar{y}_{sv2}$, $\bar{y}_{sv3}$, $\bar{y}_{sv4}$ and $\bar{y}_{sv5}$, up to the first order of large approximations using the following transformations:

$\bar{y}=\bar{Y}(1+\epsilon_0)$, and $\bar{x}=\bar{X}(1+\epsilon_1)$ such that $E(\epsilon_j)=0$, $|\epsilon_j|<1\forall\ j=0,1,2,3.$, $E(\epsilon_0^2)=((1/n)-(1/N))C_y^2$, $E(\epsilon_1^2)=((1/n)-(1/N))C_x^2$, $E(\epsilon_0\epsilon_1)=((1/n)-(1/N))\rho_{yx}C_yC_x$.

We would like to point out that the population data are used to calculate $\hat{\beta}_{hm}$. Using the above transformations, express the equations "(3.1), (3.2), (3.3), (3.4) and (3.5) " in terms of $\epsilon's$, we get

$$\bar{y}_{sv1}=\left\{\bar{Y}(1+\epsilon_0)-\bar{X}\beta_{hm}\epsilon_1(1+\epsilon_2)(1+\epsilon_3)^{-1}\right\}\exp\left[-\frac{1}{2}\epsilon_1\left(1+\frac{1}{2}\epsilon_1\right)^{-1}\right]. \tag{3.6}$$

$$\bar{y}_{sv2}=\left\{\bar{Y}(1+\epsilon_0)-\bar{X}\beta_{hm}\epsilon_1(1+\epsilon_2)(1+\epsilon_3)^{-1}\right\}\exp\left[-\frac{1}{2}\Phi_{sv2}\epsilon_1\left(1+\frac{1}{2}\Phi_{sv2}\epsilon_1\right)^{-1}\right]. \tag{3.7}$$

$$\bar{y}_{sv3}=\left\{\bar{Y}(1+\epsilon_0)-\bar{X}\beta_{hm}\epsilon_1(1+\epsilon_2)(1+\epsilon_3)^{-1}\right\}\exp\left[-\frac{1}{2}\Phi_{sv3}\epsilon_1\left(1+\frac{1}{2}\Phi_{sv3}\epsilon_1\right)^{-1}\right]. \tag{3.8}$$

$$\bar{y}_{sv4}=\left\{\bar{Y}(1+\epsilon_0)-\bar{X}\beta_{hm}\epsilon_1(1+\epsilon_2)(1+\epsilon_3)^{-1}\right\}\exp\left[-\frac{1}{2}\Phi_{sv4}\epsilon_1\left(1+\frac{1}{2}\Phi_{sv4}\epsilon_1\right)^{-1}\right]. \tag{3.9}$$

$$\bar{y}_{sv5}=\left\{\bar{Y}(1+\epsilon_0)-\bar{X}\beta_{hm}\epsilon_1(1+\epsilon_2)(1+\epsilon_3)^{-1}\right\}\exp\left[-\frac{1}{2}\Phi_{sv5}\epsilon_1\left(1+\frac{1}{2}\Phi_{sv5}\epsilon_1\right)^{-1}\right], \tag{3.10}$$

where $\Phi_{sv2}=\bar{X}/(\bar{X}+C_x)$, $\Phi_{sv3}=\bar{X}/(\bar{X}+\beta_2(x))$, $\Phi_{sv4}=\bar{X}\beta_2(x)/(\bar{X}\beta_2(x)+C_x)$, $\Phi_{sv5}=\bar{X}C_x/(\bar{X}C_x+\beta_2(x))$.

Extending the right side of "(3.6), (3.7), (3.8), (3.9) and (3.10)", multiplying and ignoring the terms of $\epsilon's$ with power higher than 2, we have

$$\bar{y}_{sv1}-\bar{Y}\cong\bar{Y}\left[\epsilon_0-\frac{1}{2}\epsilon_1+\frac{3}{8}\epsilon_1^2-\frac{1}{2}\epsilon_0\epsilon_1-\frac{\bar{X}\beta_{hm}}{\bar{Y}}\left(\epsilon_1-\frac{1}{2}\epsilon_1^2+\epsilon_1\epsilon_2-\epsilon_1\epsilon_3\right)\right]. \tag{3.11}$$

$$\bar{y}_{sv2}-\bar{Y}\cong\bar{Y}\left[\epsilon_0-\frac{1}{2}\Phi_{sv2}\epsilon_1+\frac{3}{8}\Phi_{sv2}^2\epsilon_1^2-\frac{1}{2}\Phi_{sv2}\epsilon_0\epsilon_1-\frac{\bar{X}\beta_{hm}}{\bar{Y}}\left(\epsilon_1-\frac{1}{2}\Phi_{sv2}\epsilon_1^2+\epsilon_1\epsilon_2-\epsilon_1\epsilon_3\right)\right]. \tag{3.12}$$

$$\bar{y}_{sv3}-\bar{Y}\cong\bar{Y}\left[\epsilon_0-\frac{1}{2}\Phi_{sv3}\epsilon_1+\frac{3}{8}\Phi_{sv3}^2\epsilon_1^2-\frac{1}{2}\Phi_{sv3}\epsilon_0\epsilon_1-\frac{\bar{X}\beta_{hm}}{\bar{Y}}\left(e_1-\frac{1}{2}\Phi_{sv3}\epsilon_1^2+\epsilon_1\epsilon_2-\epsilon_1\epsilon_3\right)\right]. \tag{3.13}$$

$$\bar{y}_{sv4}-\bar{Y}\cong\bar{Y}\left[\epsilon_0-\frac{1}{2}\Phi_{sv4}\epsilon_1+\frac{3}{8}\Phi_{sv4}^2\epsilon_1^2-\frac{1}{2}\Phi_{sv4}\epsilon_0\epsilon_1-\frac{\bar{X}\beta_{hm}}{\bar{Y}}\left(\epsilon_1-\frac{1}{2}\Phi_{sv4}\epsilon_1^2+\epsilon_1\epsilon_2-\epsilon_1\epsilon_3\right)\right]. \tag{3.14}$$

Table 1: Parameters of five natural population data sets

A	B	C	D	E
UScereals (Ripley *et al.*, 2013)	Singh (pp: 1111) (2003)	Murthy (pp: 399) (1967)	Murthy (pp: 288) (1967)	Engel (Koenker and Bassett, 1982)
$N = 65$	$N = 50$	$N = 34$	$N = 80$	$N = 235$
$n = 20$	$n = 20$	$n = 20$	$n = 20$	$n = 20$
$\bar{Y} = 149.4083$	$\bar{Y} = 555.4345$	$\bar{Y} = 199.4412$	$\bar{Y} = 5182.637$	$\bar{Y} = 624.1501$
$\bar{X} = 237.8384$	$\bar{X} = 878.1624$	$\bar{X} = 208.8824$	$\bar{X} = 1126.463$	$\bar{X} = 982.473$
$\rho = 0.5286552$	$\rho = 0.8038341$	$\rho = 0.9800867$	$\rho = 0.9413055$	$\rho = 0.9112434$
$C_y = 0.4177271$	$C_y = 1.052916$	$C_y = 0.7531797$	$C_y = 0.3541939$	$C_y = 0.4429335$
$C_x = 0.549239$	$C_x = 1.235168$	$C_x = 0.7205298$	$C_x = 0.7506772$	$C_x = 0.5284938$
$\beta_{hm} = 0.1928509$	$\beta_{hm} = 0.4123359$	$\beta_{hm} = 0.9537324$	$\beta_{hm} = 1.989718$	$\beta_{hm} = 0.5368326$
$\beta_2(x) = 8.191083$	$\beta_2(x) = 4.617048$	$\beta_2(x) = 2.912272$	$\beta_2(x) = 2.866433$	$\beta_2(x) = 17.63426$
$S_x = 130.6296$	$S_x = 1084.678$	$S_x = 150.506$	$S_x = 845.6097$	$S_x = 519.2309$
$S_y = 62.41187$	$S_y = 584.826$	$S_y = 150.215$	$S_y = 1835.659$	$S_y = 276.457$
$S_{xy} = 4310.041$	$S_{xy} = 509910.4$	$S_{xy} = 22158.05$	$S_{xy} = 1461142$	$S_{xy} = 130804.4$

$$\bar{y}_{sv5} - \bar{Y} \cong \bar{Y}\left[\epsilon_0 - \frac{1}{2}\Phi_{sv5}\epsilon_1 + \frac{3}{8}\Phi_{sv5}^2\epsilon_1^2 - \frac{1}{2}\Phi_{sv5}\epsilon_0\epsilon_1 - \frac{\bar{X}\beta_{hm}}{\bar{Y}}\left(\epsilon_1 - \frac{1}{2}\Phi_{sv5}\epsilon_1^2 + \epsilon_1\epsilon_2 - \epsilon_1\epsilon_3\right)\right]. \quad (3.15)$$

Squaring "(3.11), (3.12), (3.13), (3.14) and (3.15)" both sides, and discarding the terms of $\epsilon's$ having power of bigger than 2, we get

$$\left[\bar{y}_{sv1} - \bar{Y}\right]^2 = \bar{Y}^2\left[\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{2} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)^2 - 2\epsilon_0\epsilon_1\left(\frac{1}{2} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)\right]. \quad (3.16)$$

$$\left[\bar{y}_{sv2} - \bar{Y}\right]^2 = \bar{Y}^2\left[\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{2}\Phi_{sv2} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)^2 - 2\epsilon_0\epsilon_1\left(\frac{1}{2}\Phi_{sv2} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)\right]. \quad (3.17)$$

$$\left[\bar{y}_{sv3} - \bar{Y}\right]^2 = \bar{Y}^2\left[\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{2}\Phi_{sv3} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)^2 - 2\epsilon_0\epsilon_1\left(\frac{1}{2}\Phi_{sv3} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)\right]. \quad (3.18)$$

$$\left[\bar{y}_{sv4} - \bar{Y}\right]^2 = \bar{Y}^2\left[\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{2}\Phi_{sv4} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)^2 - 2\epsilon_0\epsilon_1\left(\frac{1}{2}\Phi_{sv4} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)\right]. \quad (3.19)$$

$$\left[\bar{y}_{sv5} - \bar{Y}\right]^2 = \bar{Y}^2\left[\epsilon_0^2 + \epsilon_1^2\left(\frac{1}{2}\Phi_{sv5} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)^2 - 2\epsilon_0\epsilon_1\left(\frac{1}{2}\Phi_{sv5} + \frac{\bar{X}\beta_{hm}}{\bar{Y}}\right)\right]. \quad (3.20)$$

Taking the expectation of both sides of the equations "(3.16)–(3.20)", we obtain the MSEs of the considered estimators $\bar{y}_{sv1}$, $\bar{y}_{sv2}$, $\bar{y}_{sv3}$, $\bar{y}_{sv4}$ and $\bar{y}_{sv5}$, up to the first order of large approximations as

$$\mathrm{MSE}\left(\bar{y}_{sv1}\right) = \frac{1-f}{n}\left[S_y^2 + \frac{1}{4}R_{sv1}^2 S_x^2 + \beta_{hm}^2 S_x^2 - R_{sv1}S_{yx} + R_{sv1}\beta_{hm}S_x^2 - 2\beta_{hm}Syx\right]. \quad (3.21)$$

$$\mathrm{MSE}\,(\bar{y}_{sv2}) = \frac{1-f}{n}\left[S_y^2 + \frac{1}{4}R_{sv2}^2 S_x^2 + \beta_{hm}^2 S_x^2 - R_{sv2}S_{yx} + R_{sv2}\beta_{hm}S_x^2 - 2\beta_{hm}S\,yx\right]. \tag{3.22}$$

$$\mathrm{MSE}\,(\bar{y}_{sv3}) = \frac{1-f}{n}\left[S_y^2 + \frac{1}{4}R_{sv3}^2 S_x^2 + \beta_{hm}^2 S_x^2 - R_{sv3}S_{yx} + R_{sv3}\beta_{hm}S_x^2 - 2\beta_{hm}S\,yx\right]. \tag{3.23}$$

$$\mathrm{MSE}\,(\bar{y}_{sv4}) = \frac{1-f}{n}\left[S_y^2 + \frac{1}{4}R_{sv4}^2 S_x^2 + \beta_{hm}^2 S_x^2 - R_{sv4}S_{yx} + R_{sv4}\beta_{hm}S_x^2 - 2\beta_{hm}S\,yx\right]. \tag{3.24}$$

$$\mathrm{MSE}\,(\bar{y}_{sv5}) = \frac{1-f}{n}\left[S_y^2 + \frac{1}{4}R_{sv5}^2 S_x^2 + \beta_{hm}^2 S_x^2 - R_{sv5}S_{yx} + R_{sv5}\beta_{hm}S_x^2 - 2\beta_{hm}S\,yx\right]. \tag{3.25}$$

$R_{sv1} = \bar{Y}/\bar{X}$, $R_{sv2} = \bar{Y}/(\bar{X} + C_x)$, $R_{sv3} = \bar{Y}/(\bar{X} + \beta_2(x))$, $R_{sv4} = \bar{Y}\beta_2(x)/(\bar{X}\beta_2(x) + C_x)$, $R_{sv5} = \bar{Y}C_x/(\bar{X}C_x + \beta_2(x))$.

4. Theoretical efficiency comparison

In this section, the efficiency criteria for proposed estimators $\bar{y}_{svi}$ ($i = 1, 2, \ldots, 5$) have been determined algebraically according to the Kadilar *et al.* (2007) estimators.

$\mathrm{MSE}(\bar{y}_{cki}) - \mathrm{MSE}(\bar{y}_{svi}) = ((1-f)/n)[(3/4)R_{svi}^2 S_x^2 + \beta_{hm}R_{svi}S_x^2 - R_{svi}S_{xy}] > 0$, if

$$R_{svi} > 0, (i = 1, 2, \ldots, 5). \tag{4.1}$$

According to the equations (4.1), the proposed estimators $\bar{y}_{svi}$ (where $i = 1, 2, 3, 4, 5$) are more dominant than that of the Kadilar *et al.* (2007) estimators as long as the conditions (4.1) fulfilled.

5. Numerical illustration

The considered estimators are compared to the existing estimators in the literature in this section. To compare the behaviour of the suggested estimators to other existing estimators, five different types of natural population data sets (shown in Table 1) were used. Since real population data sets include outliers, we take them all into account.

We visually detected outliers in the datasets A, B, C, D, E and displayed them in Figure 1. These data points are considered outliers because they considerably diverge from the overall trend of the dataset, which suggests that they may have an impact on parameter estimation. We used the Huber *M* robust regression method, which has become known for its proficiency in handling various kinds of outliers, to mitigate this effect. Since they are created to give more robust and trustworthy parameter estimates in the case of outliers, we anticipate that our proposed estimators will outperform those found in the literature.

Data set - A : is considered from "UScereals" (Ripley *et al.*, 2013) from the "MASS" package in R-Software (2021) where,

Y = the weight of the calories in grams.

X = the weight in grams of the sodium.

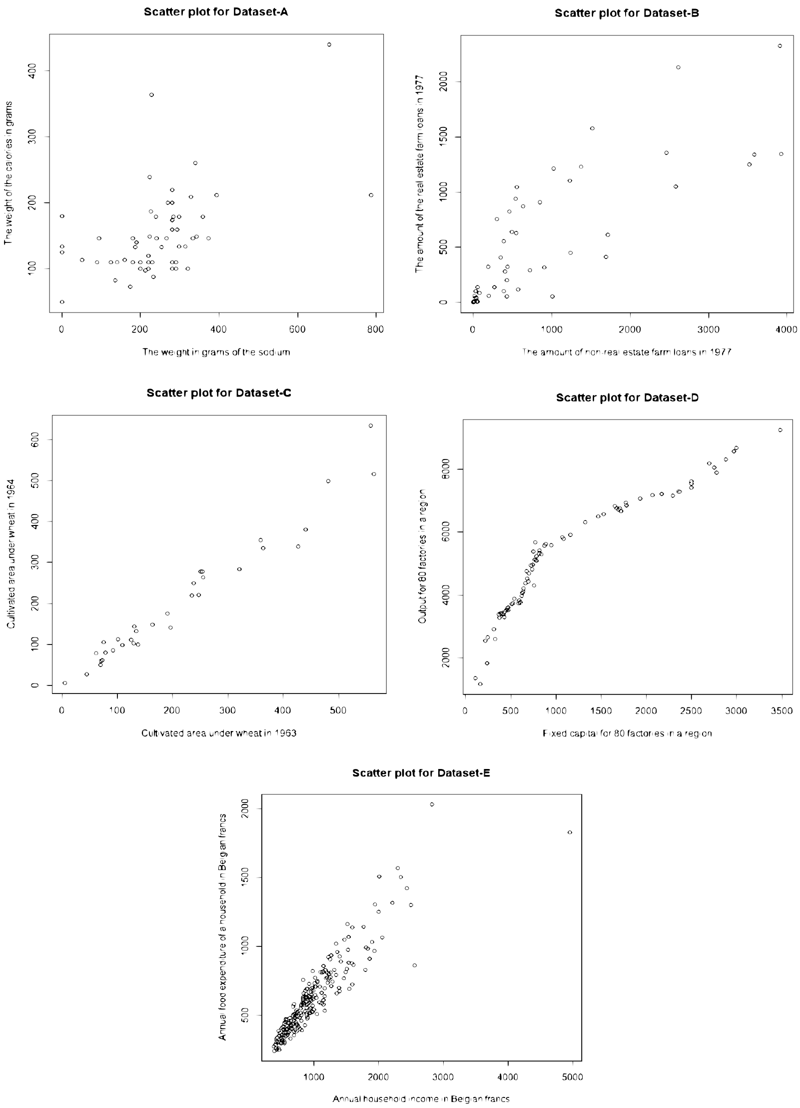

Figure 1: *Scatter plots across various datasets.*

Outlier description: We uses scatter plots to visually recognise outliers in this dataset, concentrating on cases where the weight of sodium and the weight of calories differed significantly from the majority of data points. Outliers in this dataset are mostly high-sodium and high-calorie cereals.

Data set - B : is considered from the "Singh" (2003) (Page no: 1111), where,

Y = Considered as the amount of the real estate farm loans taken out in 1977.

X = Considered as the amount of non-real estate farm loans taken out in 1977.

Outlier description: Using scatter plots, outliers in this dataset were identified, revealing multiple points of data with substantially higher values for non-real estate farm loans as well as real estate farm loans in 1977.

Data set - C : is taken from "Murthy" (1967) (Page no: 399) where,

Y = The region's cultivated area was under wheat in 1964.
X = The region's cultivated area was under wheat in 1963.

Outlier description: This dataset's outliers were identified using scatter plots, which showed situations when the cultivated area under wheat in 1964 diverged significantly from the normal range.

Data set - D : is taken from the from "Murthy" (1967) (Page no: 288), where,

Y = Output data for 80 factories in a region.
X = Fixed capital for 80 factories in a region.

Outlier description: In this dataset, we utilised scatter plots to visually evaluate data points where fixed capital and output data for factories in an area had unusual patterns, and then we used those data points to identify outliers.

Data set - E : is picked from the "Engel data set" (Koenker and Bassett, 1982) from the "quantreg" package in R-Software (2021) where,

Y = Annual food expenditure of a household in Belgian francs.
X = Annual household income in Belgian francs.

Outlier description: In order to visually identify outliers, scatter plots were used, concentrating on families with unusually high yearly incomes or annual food expenditures in Belgian francs.

We went into discussion about how these visually distinguished outliers affected our study and the estimating techniques we used. We also want to highlight that we used the Huber M robust regression method in addition to visual outlier identification. The adaptability of Huber M regression to different kinds of outliers and extreme values is well recognised. In comparison to conventional least squares regression, this technique reduces the effect of outliers on parameter estimation and produces results that are more accurate. It has been designed especially for datasets having differed forms of outliers, including high leverage, influential, and extreme values, and it is developed to manage outliers well. Our dedication to tackling outlier difficulties and producing robust estimates in the presence of multiple outlier types is demonstrated by the adoption of Huber M robust regression in our research. This improves the transparency of our research and gives a clearer picture of any possible difficulties brought on by outliers in the data sets under study. The RE(%) of the suggested estimators $\bar{y}_{svi}$, $(i = 1, 2, \ldots, 5$) with respect to the Kadilar *et al.* (2007) estimators are given as:

$$\text{RE(Existing Estimators, Proposed Estimators)} = \frac{\text{MSE(Existing Estimators)}}{\text{MSE(Proposed Estimators)}} \times 100.$$

In Tables 2–5, where we show the relative efficiency (RE(%)) results, the performance of the estimators is compared and evaluated. These tables provide insightful information about the performance of our suggested estimators in comparison to the existing estimators. The RE(%) values above 100 in

Table 2: Percent relative efficiencies of the suggested estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators

Data sets	Estimators	$\bar{y}_{ck1}$	$\bar{y}_{ck2}$	$\bar{y}_{ck3}$	$\bar{y}_{ck4}$	$\bar{y}_{ck5}$
	$\bar{y}_{sv1}$	212.7720	212.0549	202.5874	212.6842	194.9832
	$\bar{y}_{sv2}$	213.1139	212.3956	202.9129	213.0259	195.2965
A	$\bar{y}_{sv3}$	217.7221	216.9884	207.3006	217.6323	199.5195
	$\bar{y}_{sv4}$	212.8138	212.0966	202.6272	212.7260	195.0216
	$\bar{y}_{sv5}$	221.5548	220.8081	210.9498	221.4634	203.0318
	$\bar{y}_{sv1}$	250.9038	250.3324	248.7804	250.7798	249.1821
	$\bar{y}_{sv2}$	251.2503	250.6781	249.1240	251.1261	249.5262
B	$\bar{y}_{sv3}$	252.1963	251.6219	250.0620	252.0717	250.4657
	$\bar{y}_{sv4}$	250.9788	250.4073	248.8549	250.8549	249.2567
	$\bar{y}_{sv5}$	251.9507	251.3770	249.8185	251.8263	250.2219
	$\bar{y}_{sv1}$	370.1677	367.6736	360.2442	369.3084	356.5053
	$\bar{y}_{sv2}$	372.4287	369.9194	362.4446	371.5642	358.6828
C	$\bar{y}_{sv3}$	379.3293	376.7735	369.1601	378.4487	365.3287
	$\bar{y}_{sv4}$	370.9437	368.4444	360.9993	370.0825	357.2526
	$\bar{y}_{sv5}$	382.8989	380.3190	372.6340	382.0100	368.7665
	$\bar{y}_{sv1}$	379.8456	379.3468	377.9464	379.6715	377.3188
	$\bar{y}_{sv2}$	380.3142	379.8148	378.4127	380.1399	377.7843
D	$\bar{y}_{sv3}$	381.6361	381.1350	379.7280	381.4612	379.0975
	$\bar{y}_{sv4}$	380.0090	379.5100	378.1091	379.8348	377.4812
	$\bar{y}_{sv}$	382.2315	381.7296	380.3204	382.0563	379.6889
	$\bar{y}_{sv1}$	281.8757	281.6214	273.6011	281.8612	266.5705
	$\bar{y}_{sv2}$	282.0685	281.8140	273.7882	282.0541	266.7529
E	$\bar{y}_{sv3}$	288.2916	288.0315	279.8286	288.2768	272.6380
	$\bar{y}_{sv4}$	281.8866	281.6323	273.6117	281.8722	266.5809
	$\bar{y}_{sv5}$	293.9832	293.7180	285.3532	293.9681	278.0206

these tables, in specific, indicate an important benefit for the suggested estimators.
This result highlights a key finding: Our suggested estimators frequently beat their competitors in terms of mean squared error. This result occurs when the percent relative efficiencies surpass 100. This is an excellent illustration of how well our estimators perform when applied to the simulated datasets in terms of prediction accuracy and precision. These findings highlight the usefulness in real-world applications and enhanced efficiency of our suggested estimators, highlighting their potential as useful tools for statistical estimation tasks.

6. Simulation studies

To find the RE(%)of the suggested estimators, we will conduct a simulation study that is carried out by considering the "Engel Data Set" (Koenker and Bassett, 1982) presented in Table 1. This data set contains data on income and food expenditure for 235 working-class Belgian households. To load this data, load the quantreg library, and then enter the command data (engel) in R programming.

We carry out the procedures listed below for carrying out the simulation study, which were coded in R-program (2021), and we describe the simulation processes taken into account to determine the MSEs of the suggested estimators $\bar{y}_{svi}$ ($i = 1, 2, \ldots, 5$) and traditional estimator $\bar{y}_{cki}$, ($i = 1, 2, \ldots, 5$)

Step 1 : Select the 5,000, 10,000 and 1,00,000 samples of the different size n (where $n = 20, n = 30, n = 40$ and $n = 50$) using the "Engel Data Set" (Koenker and Bassett, 1982) that are mentioned in R program using the SRS without repalcement techniques.

Step 2 : After that we will considered the data from 5,000, 10,000 and 1,00,000 samples to find the

Table 3: Relative efficiencies (%) of the considered estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators for simulation studies for 5,000 iterations

Sample Sizes	Estimators	$\bar{y}_{ck1}$	$\bar{y}_{ck2}$	$\bar{y}_{ck3}$	$\bar{y}_{ck4}$	$\bar{y}_{ck5}$
	$\bar{y}_{sv1}$	287.4186	287.1696	284.8613	287.3576	282.4920
	$\bar{y}_{sv2}$	287.6115	287.3623	285.0525	287.5505	282.6816
$n = 20$	$\bar{y}_{sv3}$	289.4916	289.2408	286.9159	289.4302	284.5295
	$\bar{y}_{sv4}$	287.4619	287.2128	284.9042	287.4009	282.5346
	$\bar{y}_{sv5}$	291.1673	290.9150	288.5766	291.1055	286.1765
	$\bar{y}_{sv1}$	281.6029	281.3562	278.3832	281.5545	275.7349
	$\bar{y}_{sv2}$	281.7920	281.5452	278.5702	281.7435	275.9200
$n = 30$	$\bar{y}_{sv3}$	284.2157	283.9668	280.9662	284.1669	278.2933
	$\bar{y}_{sv4}$	281.6370	281.3903	278.4170	281.5886	275.7683
	$\bar{y}_{sv5}$	286.1736	285.9230	282.9017	286.1244	280.2103
	$\bar{y}_{sv1}$	280.9477	280.7004	277.2045	280.9056	274.2236
	$\bar{y}_{sv2}$	281.1367	280.8892	277.3909	281.0946	274.4081
$n = 40$	$\bar{y}_{sv3}$	283.9790	283.7289	280.1953	283.9364	277.1823
	$\bar{y}_{sv4}$	280.9777	280.7303	277.2340	280.9356	274.2529
	$\bar{y}_{sv5}$	286.2625	286.0104	282.4484	286.2196	279.4112
	$\bar{y}_{sv1}$	284.7560	284.5029	280.5511	284.7162	277.1985
	$\bar{y}_{sv2}$	284.9496	284.6964	280.7418	284.9098	277.3870
$n = 50$	$\bar{y}_{sv3}$	288.1667	287.9106	283.9114	288.1265	280.5187
	$\bar{y}_{sv4}$	284.7844	284.5314	280.5791	284.7447	277.2262
	$\bar{y}_{sv5}$	290.7960	290.5376	286.5019	290.7554	283.0782

value of the $\hat{\hat{Y}}$. Now, we have the 5,000, 10,000 and 1,00,000 values of $\hat{\hat{Y}}$ from the 5,000, 10,000 and 1,00,000 samples for each sample n.

Step 3 : The mean squared error of $\hat{\hat{Y}}$ is computed for each n by

$$\mathrm{MSE}\left(\hat{\hat{Y}}\right) = \frac{1}{5000}\sum_{i=1}^{5000}\left(\hat{\hat{Y}} - \bar{Y}\right)^2, \tag{6.1}$$

where $\bar{Y}$ is population mean of the study variable.

The MSE ratio of the investigated estimators to the current estimators for each sample size (n) is computed to estimate the relative efficiency. All of the suggested estimators clearly outperform the current ones across all sample sizes, demonstrating their greater effectiveness when compared to conventional estimators. The simulation results support this observation, demonstrating the accuracy of our theoretical results.

It is important to highlight that the suggested estimators' efficiency significantly increases when compared to the current estimators. To put it another way, the suggested estimators show considerably higher efficiency in situations where outliers were more likely to occur in the data. The Tables 3–5 provide a brief overview of the results of the simulation after multiple iterations.

7. Analysis of numerical illustration and simulation study

From the Tables 1–3, the following interpretation can be found:

1. We present descriptions of five real-world data sets in the Table 1 to demonstrate the applications of our research.

2. From the Table 2

Table 4: Relative efficiencies (%) of the considered estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators for simulation studies for 10,000 iterations

Sample Sizes	Estimators	$\bar{y}_{ck1}$	$\bar{y}_{ck2}$	$\bar{y}_{ck3}$	$\bar{y}_{ck4}$	$\bar{y}_{ck5}$
	$\bar{y}_{sv1}$	283.7811	283.5358	281.2569	283.7217	278.9507
	$\bar{y}_{sv2}$	283.9702	283.7247	281.4443	283.9108	279.1366
$n = 20$	$\bar{y}_{sv3}$	285.8136	285.5665	283.2714	285.7538	280.9487
	$\bar{y}_{sv4}$	283.8230	283.5776	281.2984	283.7636	278.9919
	$\bar{y}_{sv5}$	287.4332	287.1847	284.8765	287.3731	282.5407
	$\bar{y}_{sv1}$	281.8515	281.6043	278.6550	281.8028	276.0324
	$\bar{y}_{sv2}$	282.0411	281.7937	278.8425	281.9923	276.2181
$n = 30$	$\bar{y}_{sv3}$	284.4471	284.1975	281.2211	284.3978	278.5743
	$\bar{y}_{sv4}$	281.8860	281.6387	278.6891	281.8372	276.0661
	$\bar{y}_{sv5}$	286.3941	286.1429	283.1461	286.3446	280.4812
	$\bar{y}_{sv1}$	282.4314	282.1826	278.6957	282.3882	275.6903
	$\bar{y}_{sv2}$	282.6213	282.3723	278.8831	282.5781	275.8757
$n = 40$	$\bar{y}_{sv3}$	285.4612	285.2097	281.6854	285.4175	278.6478
	$\bar{y}_{sv4}$	282.4620	282.2132	278.7259	282.4189	275.7203
	$\bar{y}_{sv5}$	287.7615	287.5080	283.9553	287.7175	280.8932
	$\bar{y}_{sv1}$	284.7022	284.4496	280.5371	284.6623	277.2014
	$\bar{y}_{sv2}$	284.8955	284.6427	280.7276	284.8556	277.3896
$n = 50$	$\bar{y}_{sv3}$	288.0771	287.8214	283.8626	288.0367	280.4873
	$\bar{y}_{sv4}$	284.7308	284.4782	280.5653	284.6909	277.2293
	$\bar{y}_{sv5}$	290.6930	290.4350	286.4403	290.6522	283.0343

(I) For Data Set-A, the PREs of the estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 194.9832% to 221.5548% for the Population size of 65 and the sample size of 20.

(II) For data set B, with Population size 50 and sample size 20, the PREs of the estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 249.12% to 252.19%.

(III) For data set C, the PREs of the estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 356.50% to 382.89% for the Population size of 34 and the sample size of 20.

(IV) For data set D, with Population size 80 and sample size 20, the PREs of the estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 377.31% to 382.23%.

(V) For data set E, the PREs of the estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 266.57% to 293.98% for the Population size of 235 and the sample size of 20.

3. From the Table 3

(I) For sample size of 20 and the population size of 5000, the PREs of the suggested estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 282.49% to 291.16%.

(II) For sample size of 30 and the population size of 5000, the PREs of the suggested estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 275.73% to 286.17%.

(III) For sample size of 40 and the population size of 5000, the PREs of the suggested estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 274.22% to 286.26%.

(IV) For sample size of 50 and the population size of 5000, the PREs of the suggested estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators remain between 277.19% to 290.79%.

4. Similar results we will get from the Tables 4–5.

Table 5: Relative efficiencies (%) of the considered estimators $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$ over the existing estimators for simulation studies for 1,00,000 iterations

Sample Sizes	Estimators	$\bar{y}_{ck1}$	$\bar{y}_{ck2}$	$\bar{y}_{ck3}$	$\bar{y}_{ck4}$	$\bar{y}_{ck5}$
	$\bar{y}_{sv1}$	283.8537	283.6076	281.3037	283.7942	278.9635
	$\bar{y}_{sv2}$	284.0422	283.7959	281.4905	283.9826	279.1487
$n = 20$	$\bar{y}_{sv3}$	285.8925	285.6447	283.3242	285.8326	280.9672
	$\bar{y}_{sv4}$	283.8955	283.6494	281.3451	283.8360	279.0046
	$\bar{y}_{sv5}$	287.5312	287.2819	284.9482	287.4709	282.5776
	$\bar{y}_{sv1}$	283.6235	283.3746	280.3902	283.5742	277.6850
	$\bar{y}_{sv2}$	283.8144	283.5653	280.5789	283.7651	277.8719
$n = 30$	$\bar{y}_{sv3}$	286.2491	285.9979	282.9859	286.1994	280.2556
	$\bar{y}_{sv4}$	283.6584	283.4095	280.4247	283.6091	277.7192
	$\bar{y}_{sv5}$	288.2599	288.0069	284.9737	288.2098	282.2243
	$\bar{y}_{sv1}$	282.6203	282.3717	278.8912	282.5771	275.8742
	$\bar{y}_{sv2}$	282.8102	282.5614	279.0786	282.7670	276.0596
$n = 40$	$\bar{y}_{sv3}$	285.6450	285.3937	281.8760	285.6013	278.8267
	$\bar{y}_{sv4}$	282.6510	282.4024	278.9215	282.6078	275.9042
	$\bar{y}_{sv5}$	287.9584	287.7051	284.1589	287.9144	281.0849
	$\bar{y}_{sv1}$	282.9058	282.6569	278.8137	282.8660	275.5154
	$\bar{y}_{sv2}$	283.0956	282.8465	279.0007	283.0558	275.7003
$n = 50$	$\bar{y}_{sv3}$	286.2084	285.9566	282.0685	286.1682	278.7318
	$\bar{y}_{sv4}$	282.9342	282.6853	278.8417	282.8945	275.5432
	$\bar{y}_{sv5}$	288.7843	288.5303	284.6072	288.7438	281.2404

We introduced and compared our estimators, designated as $\bar{y}_{svi}$ $(i = 1, 2, \ldots, 5)$, to current approaches in our research of real-world datasets. Our estimators frequently beat the alternatives in a variety of situations with different population sizes and sample sizes, demonstrating their efficiency in predicting population characteristics. Notably, all of our suggested relative efficiency (RE%) values were greater than 100%, showing clearly that our estimators are more efficient than existing estimators. We strongly advise survey practitioners to use these estimators since they offer excellent and trustworthy estimations for a variety of survey applications.

8. Conclusions

In the presence of outliers, using traditional statistical approaches for data analysis might lead to incorrect outcomes. Robust regression approaches have been used to enhance methods for predicting the population mean in order to solve this problem. This article presents an innovative approach for analysing sample survey data, concentrating on the development of exponential estimators of the ratio type using the Huber M-function. The study compares these newly suggested estimators' mean square errors (MSEs) to those of the estimators previously proposed by Kadlar, Candan, and Cingi in 2007. The study offers a thorough review that includes both theoretical derivations and actual implementations. The results of the study constantly show that the newly suggested estimators work better than their competitors, producing lower MSEs under different circumstances. The aforementioned findings are supported by rigorous numerical illustrations and simulation studies, that intentionally take into account the existence of outliers in the data. As a result, the estimators based on the robust regression techniques given in this article outperformed the estimators from Kadilar *et al.* (2007) across both real-world data sets and simulated scenarios. The positive findings of this study not only confirm the effectiveness of the suggested estimators but also open the door for future efforts to broaden the applicability of estimators across various sampling techniques. This article advances data analysis methods, especially in situations where outliers are a challenge. It also holds significant potential for applications in a variety of industries, including business, economics, and agriculture, which will eventually encourage intelligent policy development. Future research will be conducted to improve and diversify the estimator toolbox for sample surveys.

Data Availability Statement All the relevant data information is avaiable within the manuscript and code is given in Appendix.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgement

Authors are thankful to the Editor-in-Chief and learned referees for their inspiring and fruitful suggestions. The authors are also thankful to the National Institute of Technology Arunachal Pradesh, Jote, for providing the necessary infrastructure for the completion of the present work.

Appendix A: R - Code

```
# Load necessary libraries
library(quantreg)

# Load the ‘engel’ dataset
data(engel)

# Define the variable (food expenditure) and (income)
Y <- engel$foodexp
X <- engel$income

# Calculate the means variables
Ybar <- mean(Y)
Xbar <- mean(X)

# Specify the number of bootstrap replications (5000, 10000, 100000)
B <- 100000

# Set the sample size ‘n’ (20, 30, 40, 50)
n <- 50

# Define ‘N’ as the length of the ’income’ vector
N <- length(engel$income)

# Initialize vectors to store results for different estimators
T1.1 <- numeric(B)
T1.2 <- numeric(B)
T1.3 <- numeric(B)
T1.4 <- numeric(B)
T1.5 <- numeric(B)
P1.1 <- numeric(B)
P1.2 <- numeric(B)
P1.3 <- numeric(B)
P1.4 <- numeric(B)
P1.5 <- numeric(B)
```

```
# Loop for bootstrap replications
for (K in 1:B) {

# Randomly sample data without replacement to create a bootstrap sample
swor <- sample(N, size = n, replace = FALSE)
y <- engel$foodexp[swor]
x <- engel$income[swor]
ybar <- mean(y)
xbar <- mean(x)
Cx <- sd(x) / mean(x)
B2 <- kurtosis(x)


library(MASS)
Br1 <- rlm(y ~ x)
Br <- Br1$coefficients[2]

# Calculate various estimators
T1.1[K] <- ((ybar + Br * (Xbar - xbar)) / xbar) * Xbar
T1.2[K] <- ((ybar + Br * (Xbar - xbar)) / (xbar + Cx)) * (Xbar + Cx)
T1.3[K] <- ((ybar + Br * (Xbar - xbar)) / (xbar + B2)) * (Xbar + B2)
T1.4[K] <- ((ybar + Br * (Xbar - xbar)) / (xbar * B2 + Cx)) * (Xbar * B2 + Cx)
T1.5[K] <- ((ybar + Br * (Xbar - xbar)) / (xbar * Cx + B2)) * (Xbar * Cx + B2)

P1.1[K] <- (ybar + Br * (Xbar - xbar)) * exp((Xbar - xbar) / (Xbar + xbar))
P1.2[K] <- (ybar + Br * (Xbar - xbar)) * exp((Xbar - xbar) / ((Xbar + xbar)
+ 2 * Cx))
P1.3[K] <- (ybar + Br * (Xbar - xbar)) * exp((Xbar - xbar) / ((Xbar + xbar)
+ 2 * B2))
P1.4[K] <- (ybar + Br * (Xbar - xbar)) * exp((B2 * (Xbar - xbar)) / (B2 * (Xbar
+ xbar) + 2 * Cx))
P1.5[K] <- (ybar + Br * (Xbar - xbar)) * exp((Cx * (Xbar - xbar)) / (Cx * (Xbar
+ xbar) + 2 * B2))
}

# Calculate Mean Squared Errors (MSE) for each estimator
MSEY1 <- mean((T1.1 - mean(T1.1))^2)
MSEY2 <- mean((T1.2 - mean(T1.2))^2)
MSEY3 <- mean((T1.3 - mean(T1.3))^2)
MSEY4 <- mean((T1.4 - mean(T1.4))^2)
MSEY5 <- mean((T1.5 - mean(T1.5))^2)

d <- data.frame(MSEY1, MSEY2, MSEY3, MSEY4, MSEY5)

MSEYpr1 <- mean((P1.1 - mean(P1.1))^2)
MSEYpr2 <- mean((P1.2 - mean(P1.2))^2)
MSEYpr3 <- mean((P1.3 - mean(P1.3))^2)
MSEYpr4 <- mean((P1.4 - mean(P1.4))^2)
MSEYpr5 <- mean((P1.5 - mean(P1.5))^2)
```

```
d1 <- data.frame(MSEYpr1, MSEYpr2, MSEYpr3, MSEYpr4, MSEYpr5)

# Relative Efficiency
da1 <- c(MSEY1 / MSEYpr1, MSEY2 / MSEYpr1, MSEY3 / MSEYpr1, MSEY4 / MSEYpr1, MSEY5
/ MSEYpr1)
da2 <- c(MSEY1 / MSEYpr2, MSEY2 / MSEYpr2, MSEY3 / MSEYpr2, MSEY4 / MSEYpr2, MSEY5
/ MSEYpr2)
da3 <- c(MSEY1 / MSEYpr3, MSEY2 / MSEYpr3, MSEY3 / MSEYpr3, MSEY4 / MSEYpr3, MSEY5
/ MSEYpr3)
da4 <- c(MSEY1 / MSEYpr4, MSEY2 / MSEYpr4, MSEY3 / MSEYpr4, MSEY4 / MSEYpr4, MSEY5
/ MSEYpr4)
da5 <- c(MSEY1 / MSEYpr5, MSEY2 / MSEYpr5, MSEY3 / MSEYpr5, MSEY4 / MSEYpr5, MSEY5
/ MSEYpr5)

da <- c(da1, da2, da3, da4, da5)
RE_matrix <- matrix(da, ncol = 5, nrow = 5, byrow = T) * 100

# Display or export results
d  # MSE results for T1 estimators
d1  # MSE results for P1 estimators
RE_matrix  # Relative Efficiency matrix
```

References

Bulut H and Zaman T (2022). An improved class of robust ratio estimators by using the minimum covariance determinant estimation, *Communications in Statistics-Simulation and Computation*, **51**, 2457–2463.

Edgeworth FY (1887). Observations relating to several quantities, *Hermathena*, **6**, 279–285.

Huber PJ (1973). Robust regression asymptotics, conjectures and monte carlo, *Annals of Statistics*, **1**, 799–821.

Huber P (1981). *Robust Statistics*, Wiley, New York.

Kadilar C, Candan M, and Cingi H (2007). Ratio estimation using robust regression, *Hacettepe Journal of Mathematics and Statistics*, **36**, 181–188.

Kadilar C and Cingi H (2004). Ratio estimators in simple random sampling, *Applied Mathematics and Computation*, **151**, 893–902.

Koenker R (2009). Quantreg: Quantile regression, Available from: http://CRAN. R-project. org/-package= quantreg.

Koenker R and Bassett G (1982). Robust tests of heteroscedasticity based on regression quantiles, *Econometrica*, **50**, 43–61.

Murthy MN (1967). *Sampling: Theory and Methods*, Statistical Pub. Society.

Noor-ul-Amin M, Shahbaz MQ, and Kadilar C (2016). Ratio estimators for population mean using robust regression in double sampling, *Gazi University Journal of Science*, **29**, 793–798.

Noor-ul-Amin M, Asghar S, and Sanaullah A (2018). Redescending M-estimator for robust regression, *Journal of Reliability and Statistical Studies*, **11**, 69–80.

Noor-ul-Amin M, Asghar S, and Sanaullah A (2022). Ratio estimators in the presence of outliers using redescending M-estimator, *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, **92**, 65–70.

Prasad S (2020). Some linear regression type ratio exponential estimators for estimating the population mean based on quartile deviation and deciles, *Statistics in Transition, New Series*, **21**, 85–98.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rousseeuw P and Leroy A (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, and Ripley MB (2013). Package ‘MASS’, *Cran r*, **538**, 113–120.

Singh S (2003). *Advanced Sampling Theory With Applications: How Michael ‘Selected’ Amy (Vol. 2)*, Kluwer Academic Publishers.

Zaman T (2020). Generalized exponential estimators for the finite population mean, *Statistics in Transition New Series*, **21**, 159–168.

Zaman T (2021). An efficient exponential estimator of the mean under stratified random sampling, *Mathematical Population Studies*, **28**, 104–121.

Zaman T and Kadilar C (2021a). Exponential ratio and product type estimators of the mean in stratified two-phase sampling, *AIMS Mathematics*, **6**, 4265–4279.

Zaman T and Kadilar C (2021b). New class of exponential estimators for finite population mean in two-phase sampling, *Communications in Statistics-Theory and Methods*, **50**, 874–889.

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

Zaman T and Bulut H (2023). An efficient family of robust-type estimators for the population variance in simple and stratified random sampling, *Communications in Statistics-Theory and Methods*, **52**, 2610–2624.

Zaman T, Dünder E, Audu A, Alilah DA, Shahzad U, and Hanif M (2021). Robust regression-ratio-type estimators of the mean utilizing two auxiliary variables: A simulation study, *Mathematical Problems in Engineering*, **2021**, 1–9.

Zaman T, Bulut H, and Yadav SK (2022). Robust ratio-type estimators for finite population mean in simple random sampling: A simulation study, *Concurrency and Computation: Practice and Experience*, **34**, e7273.

Received June 16, 2023; Revised September 22, 2023; Accepted December 16, 2023