Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 2, 2024 (New Series), pp 29–45
http://www.ssca.org.in/journal

Simple Linear Slope Estimators Based On Sample Quasi Ranges

Sharada V. Bhat and Shrinath M. Bijjargi
Department of Statistics,
Karnatak University, Dharwad - 580 003, India.

Received: 01 October 2023; Revised: 25 October 2023; Accepted: 28 October 2023

Abstract

The slope parameter in simple linear regression measures the change in mean of distribution of response variable for unit change in predictor variable. Some estimators based on sample quasi ranges of predictor variables are proposed. The mean and variance of the proposed estimators are derived. The relative efficiencies among the proposed estimators are obtained. Also, these estimators are compared with the estimators available in the literature. Few datasets are considered to illustrate the fitting of simple linear regression using proposed estimators and comparing their performances.

Key words: Simple linear regression; Sample quasi ranges; Unbiased estimators; Variance; Slope parameter; Relative efficiency.

AMS Subject Classifications: 62G05, 62J05

1. Introduction

Regression analysis helps in understanding the nature and strength of the relationship among two or more variables. Linear regression model is helpful in modeling the relationship among response variable (y) and the predictor variable (x). This model is used by economists to relate variables such as consumption, savings with income; by environmental scientists to relate environmental factors such as temperature, pollution levels with ecosystem or public health; psychologists to relate human behavior with mental health and stress levels with academic performances, *etc.* In addition, it is used in various domains of studies like finance, marketing, real estate, pharmaceuticals, clinical trials, national development, education and many others. The least square estimator is widely used in linear regression to estimate the slope parameter. The literature reveals that the method of least squares was due to Legendre (1805). Gauss (1809) claimed that he had been using the procedure since 1795. Harter (1974), Stigler (1986) and Hald (1998) noticed that, "Euler (1749) and Mayer (1750) independently developed a method known as method of averages" for fitting a linear equation to observed data. Their method deals with arranging the predictor variables in descending order and grouping them into as many numbers of existing parameters.

Corresponding Author: Sharada V. Bhat
Email: bhat_sharada@yahoo.com

Bose (1938) proposed three estimators based on method of successive differences, method of differences at half range and method of range as alternative to least square estimator for slope parameter in simple linear regression, when predictor variables are equidistant. Wald (1940) observed that the efficiency of slope estimator will be maximum when x_i's are arranged in ascending order. Nair and Shrivastava (1942) generalized procedure of Bose (1938) to method of group averages to improve the relative efficiency of the estimators.

Liu and Preve (2016) proposed estimators to slope parameter in simple linear regression based on robust measures of location, *viz.* median and trimmed mean. The focus is on the case where predictor variable is assumed to be stochastic, having symmetric stable distribution and error having distribution either symmetric stable or a normal mixture. Cliff and Billy (2017) developed simple averaging method based on the average of successive slopes. Prabowo *et al.* (2020) simplified this method and investigated its performance. Singthongchai *et al.* (2021) developed improved simple averaging method replacing median in the place of mean in the method due to Cliff and Billy (2017). Jlibene *et al.* (2021) studied the least square estimator when the error has uniform distribution. Yao *et al.* (2021) proposed best linear unbiased estimators using moving extremes ranked set sampling.

Bhat and Bijjargi (2023) proposed estimation procedures generalizing the methods due to Bose (1938) including some adaptive estimators, in the presence of unequal distances among predictor variables. Among the methods proposed, the method of differences among ordered predictors lying equally on either side of the half range outperforms all other estimators. Basically, this estimator is based on quasi ranges. The immediate quest that arises is, whether the estimator is improved by taking some weights to quasi ranges. To investigate this fact, we develop few estimators based on different types of weights to quasi ranges.

In this paper, we propose some estimators for slope parameter of simple linear regression model based on sample quasi ranges given in Govindarajulu (2007). Suppose x_i , $1 \leq i \leq n$ are arranged in ascending order of magnitude, $x_{(i)}$ is the i^{th} order statistic, then, for $n = 2m$, the j^{th} quasi range, $j = 1,\ 2,\ \cdots,\ m-1$ is defined as the range of $(n-2j)$ sample values. Suppose q_j is the j^{th} quasi range, then q_j is given by $q_j = x_{(n-j)} - x_{(j+1)}$. We observe that, $q_0 = x_{(n)} - x_{(1)}$ is the range of n observations. Mosteller (2006) and Harter (1959) used quasi ranges to estimate population standard deviation.

The proposed estimators are given in section 2, their mean and variance are derived in section 3 and their performance using relative efficiency is investigated in section 4. The simple linear regression using proposed estimators along with their performances are illustrated through examples in section 5. Section 6 contains conclusions.

2. Estimators based on sample quasi ranges

Consider the simple linear regression model,

$$y_i = \alpha + \beta x_i + e_i\ , \qquad\qquad 1 \leq i \leq n \tag{1}$$

where, y_i is response variable, x_i is predictor variable, e_i is independent and identically distributed random error from distribution with zero mean and finite variance σ^2. Here, α is intercept parameter and β is slope parameter to be estimated from the data to explore the linear relation between x_i and y_i. The slope parameter β represents the change in mean of distribution of y for unit change in x. The least square estimator of β is given by

$$\widehat{\beta}_{ul} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}, \tag{2}$$

where, $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$.

Among the methods proposed by Bose (1938), the estimator $\widehat{\beta}_{eh}$ obtained by method of differences at half range outperforms other estimators and is given by

$$\widehat{\beta}_{eh} = \frac{\sum_{i=1}^{m} (y_{m+i} - y_i)}{m^2 d}, \tag{3}$$

where d is distance among ordered x_i.

In case of unequal distances among predictor variables, estimator due to Bhat and Bijjargi (2023) based on method of distances among ordered observations lying equally on either side of half range outperforms other proposed estimators and is given by

$$\widehat{\beta}_{ud} = \frac{\sum_{i=1}^{m} \left(y^*_{m+i} - y^*_{m-i+1}\right)}{\sum_{i=1}^{m} \left(x_{(m+i)} - x_{(m-i+1)}\right)}. \tag{4}$$

Here, y_i^* is y observation corresponding to $x_{(i)}$, i^{th} order statistic. $\widehat{\beta}_{ud}$ reduces to the method of differences at half range given by

$$\widehat{\beta}_{uh} = \frac{\sum_{i=1}^{m} \left(y^*_{m+i} - y^*_{i}\right)}{\sum_{i=1}^{m} \left(x_{(m+i)} - x_{(i)}\right)}. \tag{5}$$

Also, when distances among ordered predictor variables are equal, $\widehat{\beta}_{ud} = \widehat{\beta}_{uh}$ reduces to $\widehat{\beta}_{eh}$.

We propose estimators $\widehat{\beta}_k,\ k = 1,\ 2,\ \cdots,\ 6$ using quasi ranges respectively based on the weights $w_k,\ k = 1,\ 2,\ \cdots,\ 6$. Representing arbitrary weight by a_{ki}, $k = 1,\ \cdots,\ 6$, $i = 1,\ \cdots,\ m$, w_1 is given by $a_{1i} = \frac{1}{m-i+1}$, w_2 by $a_{2i} = \frac{1}{i}$, w_3 by $a_{3i} = \frac{m-i+1}{\sum_{i=1}^{m} m-i+1}$, $a_{4i} = \frac{i}{\sum_{i=1}^{m} i}$, $a_{5i} = m - i + 1$ and $a_{6i} = i$. We see that, a_{1i}, a_{4i} and a_{6i} relatively assign heavier weights to quasi range with extreme order statistics, where as, a_{2i}, a_{3i} and a_{5i} assign lower weights. That is, a_{1i}, a_{4i} and a_{6i} assign highest weight to q_0, relatively lesser weight to q_1, q_2, $\cdots$ and q_{m-1}. Similarly, a_{2i}, a_{3i} and a_{5i} assign lowest weight to q_0, relatively heavier weights to q_1, q_2, $\cdots$ and q_{m-1}. As efficiency and robustness are vital to estimators, the motivation for assigning various weights to the quasi ranges is to develop adaptive estimators in terms of efficiency and robustness. In the presence of several estimators, researcher seeks efficient estimator that closely estimates the parameter, whereas, robust estimator is sought to estimate the parameter sensibly in the presence of outliers in the data.

The weights employed to propose various estimators are given in detail in Table 1.

Table 1: Weights for proposition of estimators

range/quasi ranges	y values	w_1	w_2	w_3	w_4	w_5	w_6
$q_0 = x_{(n)} - x_{(1)}$	$y_n^* - y_1^*$	1	$\frac{1}{m}$	$\frac{1}{\sum_{i=1}^m (m-i+1)}$	$\frac{m}{\sum_{i=1}^m i}$	1	m
$q_1 = x_{(n-1)} - x_{(2)}$	$y_{n-1}^* - y_2^*$	$\frac{1}{2}$	$\frac{1}{m-1}$	$\frac{2}{\sum_{i=1}^m (m-i+1)}$	$\frac{m-1}{\sum_{i=1}^m i}$	2	$m-1$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
$q_{m-2} = x_{(m+2)} - x_{(m-1)}$	$y_{m+2}^* - y_{m-1}^*$	$\frac{1}{m-1}$	$\frac{1}{2}$	$\frac{m-1}{\sum_{i=1}^m (m-i+1)}$	$\frac{2}{\sum_{i=1}^m i}$	$m-1$	2
$q_{m-1} = x_{(m+1)} - x_{(m)}$	$y_{m+1}^* - y_m^*$	$\frac{1}{m}$	1	$\frac{m}{\sum_{i=1}^m (m-i+1)}$	$\frac{1}{\sum_{i=1}^m i}$	m	1

The proposed estimators are given by

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^m \left(\frac{y_{m+i}^* - y_{m-i+1}^*}{m-i+1}\right)}{\sum_{i=1}^m \left(\frac{x_{(m+i)} - x_{(m-i+1)}}{m-i+1}\right)}, \tag{6}$$

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^m \left(\frac{y_{m+i}^* - y_{m-i+1}^*}{i}\right)}{\sum_{i=1}^m \left(\frac{x_{(m+i)} - x_{(m-i+1))}}{i}\right)}, \tag{7}$$

$$\widehat{\beta}_3 = \frac{\sum_{i=1}^m \frac{(m-i+1)\left(y_{m+i}^* - y_{m-i+1}^*\right)}{\sum_{i=1}^m (m-i+1)}}{\sum_{i=1}^m \frac{(m-i+1)\left(x_{(m+i)} - x_{(m-i+1)}\right)}{\sum_{i=1}^m (m-i+1)}}, \tag{8}$$

$$\widehat{\beta}_4 = \frac{\sum_{i=1}^m \frac{i\left(y_{m+i}^* - y_{m-i+1}^*\right)}{\sum_{i=1}^m i}}{\sum_{i=1}^m \frac{i\left(x_{(m+i)} - x_{(m-i+1)}\right)}{\sum_{i=1}^m i}}, \tag{9}$$

$$\widehat{\beta}_5 = \frac{\sum_{i=1}^m (m-i+1)\left(y_{m+i}^* - y_{m-i+1}^*\right)}{\sum_{i=1}^m (m-i+1)\left(x_{(m+i)} - x_{(m-i+1)}\right)} = \widehat{\beta}_3 \tag{10}$$

and

$$\widehat{\beta}_6 = \frac{\sum_{i=1}^m i\left(y_{m+i}^* - y_{m-i+1}^*\right)}{\sum_{i=1}^m i\left(x_{(m+i)} - x_{(m-i+1)}\right)} = \widehat{\beta}_4. \tag{11}$$

To obtain these estimators under the situation that the predictor variables are equidistant, $d\sum_{i=1}^m (2i-1)$ is substituted in the place of $\sum_{i=1}^m \left(x_{(m+i)} - x_{(m-i+1)}\right)$.

For odd number of sample sizes, i.e. $n = 2m + 1$, the middle pair of observation, $(y^*_{m+1}, x_{(m+1)})$ is not considered. When the distances among $x_{(i)}$'s are unequal and the weights are equal, the estimators $\widehat{\beta}_k,\ k = 1,\ 2,\ 3,\ 4$ reduce to $\widehat{\beta}_{ud}$ and to $\widehat{\beta}_{eh}$ when distances are equal.

3. Mean and variance of the proposed estimators

In this section, the mean of the proposed estimators and their variances are obtained. The mean of $\widehat{\beta}_1$ is given by

$$\begin{aligned}
E\left(\widehat{\beta}_1\right) &= E\left(\frac{\sum_{i=1}^m \left(\frac{y^*_{m+i}-y^*_{m-i+1}}{m-i+1}\right)}{\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)}\right)\\
&= \frac{1}{\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)} E\left(\sum_{i=1}^m \left(\frac{y^*_{m+i}-y^*_{m-i+1}}{m-i+1}\right)\right)\\
&= \frac{1}{\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)} \sum_{i=1}^m \frac{1}{m-i+1} E\left(y^*_{m+i}-y^*_{m-i+1}\right)\\
&= \frac{1}{\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)} \sum_{i=1}^m \frac{\beta}{m-i+1} \left(x_{(m+i)}-x_{(m-i+1)}\right)\\
&= \frac{1}{\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)} \sum_{i=1}^m \beta \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)
\end{aligned}$$

$$E\left(\widehat{\beta}_1\right) = \beta. \tag{12}$$

Hence, $\widehat{\beta}_1$ is an unbiased estimator of β.

The variance of $\widehat{\beta}_1$ is given by

$$\begin{aligned}
V\left(\widehat{\beta}_1\right) &= V\left(\frac{\sum_{i=1}^m \left(\frac{y^*_{m+i}-y^*_{m-i+1}}{m-i+1}\right)}{\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)}\right)\\
&= \frac{1}{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2} V\left(\sum_{i=1}^m \left(\frac{y^*_{m+i}-y^*_{m-i+1}}{m-i+1}\right)\right)\\
&= \frac{1}{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2} \left(\sum_{i=1}^m \frac{1}{(m-i+1)^2}\ V(y^*_{m+i}-y^*_{m-i+1})\right)\\
&= \frac{1}{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2} \left(\sum_{i=1}^m \frac{1}{(m-i+1)^2}\ 2\sigma^2\right)
\end{aligned}$$

$$V\left(\widehat{\beta}_1\right) = \frac{2\sigma^2 \sum_{i=1}^m \frac{1}{(m-i+1)^2}}{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2} \tag{13}$$

Under equidistant $x_{(i)}$'s,

$$V\left(\widehat{\beta}_1\right) = \frac{2\sigma^2 \sum_{i=1}^m \frac{1}{(m-i+1)^2}}{\left[d\ \sum_{i=1}^m \left(\frac{2i-1}{m-i+1}\right)\right]^2} \tag{14}$$

Similarly, we observe that, all the proposed estimators are unbiased estimators of β and they have different variances. The variances of $\widehat{\beta}_k$, $k = 1, 2, 3, 4$. for equal and unequal distances among $x_{(i)}$'s are furnished in Table 2.

Table 2: Variance of $\widehat{\beta}_k$, $k = 1, 2, 3, 4$

Estimator	$V\left(\widehat{\beta}_k\right)$ under unequal distances	$V\left(\widehat{\beta}_k\right)$ under equal distances
$\widehat{\beta}_1$	$\frac{2\sigma^2 \sum_{i=1}^{m} \frac{1}{(m-i+1)^2}}{\left[\sum_{i=1}^{m}\left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2}$	$\frac{2\sigma^2 \sum_{i=1}^{m} \frac{1}{(m-i+1)^2}}{\left[d \sum_{i=1}^{m}\left(\frac{2i-1}{m-i+1}\right)\right]^2}$
$\widehat{\beta}_2$	$\frac{2\sigma^2 \sum_{i=1}^{m} \frac{1}{i^2}}{\left[\sum_{i=1}^{m}\left(\frac{x_{(m+i)}-x_{(m-i+1)}}{i}\right)\right]^2}$	$\frac{2\sigma^2 \sum_{i=1}^{m} \frac{1}{i^2}}{\left[d \sum_{i=1}^{m}\left(\frac{2i-1}{i}\right)\right]^2}$
$\widehat{\beta}_3$	$\frac{2\ \sigma^2 \sum_{i=1}^{m}(m-i+1)^2}{\left[\sum_{i=1}^{m}(m-i+1)\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2}$	$\frac{48\ \sigma^2}{d^2 n(n+1)(n+2)}$
$\widehat{\beta}_4$	$\frac{2\ \sigma^2 \sum_{i=1}^{m} i^2}{\left[\sum_{i=1}^{m} i\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2}$	$\frac{48\ (n+1)\sigma^2}{d^2 n(n+2)(2n-1)^2}$

When the distances among $x_{(i)}$'s are equal, the least square estimator given in (2) reduces to

$$\widehat{\beta}_{el} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{d^2\frac{n(n^2-1)}{12}} \tag{15}$$

and its variance given by

$$V\left(\widehat{\beta}_{el}\right) = \frac{12\ \sigma^2}{d^2 n\left(n^2-1\right)} \tag{16}$$

When $d = 1$, $\sigma = 1$, $V(\widehat{\beta}_k)$ is computed from Table 2 for various values of n and are given in Table 3 and plotted in Figure 1.

Table 3: $V\left(\widehat{\beta}_k\right)$, $k = 1, 2, 3, 4$ for various values of n

n	$V\left(\widehat{\beta}_4\right)$	$V\left(\widehat{\beta}_1\right)$	$V\left(\widehat{\beta}_3\right)$	$V\left(\widehat{\beta}_2\right)$	$V\left(\widehat{\beta}_{el}\right)$	$V\left(\widehat{\beta}_{eh}\right)$
6	0.057851	0.058299	0.142857	0.15680	0.057143	0.074074
8	0.024000	0.024638	0.066667	0.081333	0.023810	0.031250
10	0.012188	0.012810	0.036364	0.049158	0.012121	0.016000
14	0.004409	0.004879	0.014286	0.023236	0.004396	0.005831
18	0.002068	0.002409	0.007018	0.013380	0.002064	0.002743
22	0.001131	0.001384	0.003953	0.008650	0.001129	0.001503
26	0.000684	0.000877	0.002442	0.006033	0.000684	0.000910
30	0.000445	0.000595	0.001613	0.004440	0.000445	0.000593
40	0.000188	0.000276	0.000697	0.002409	0.000188	0.000250
50	0.000096	0.000154	0.000362	0.001506	0.000096	0.000128
70	0.000035	0.000064	0.000134	0.000746	0.000035	0.000047
100	0.000012	0.000026	0.000047	0.000356	0.000012	0.000016

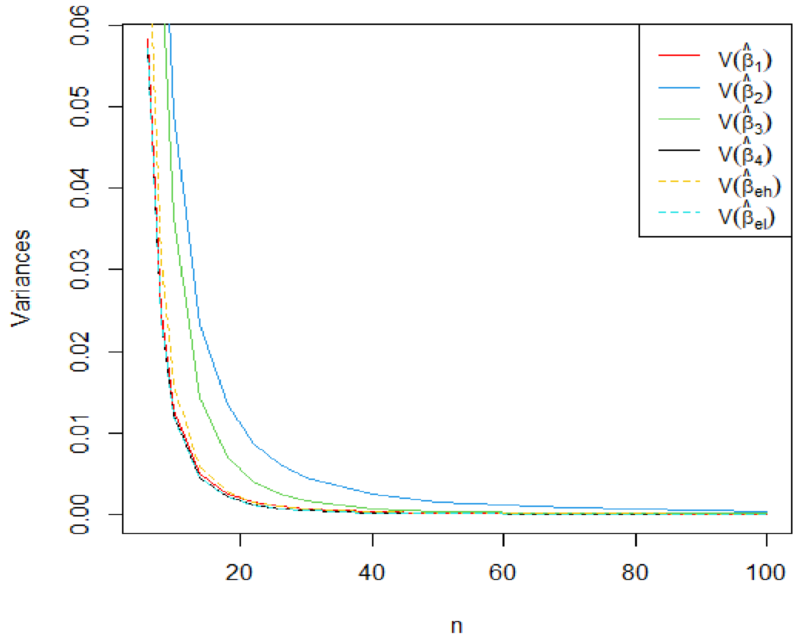


Figure 1: Variance of various slope estimators for different values of n

From Table 3 and Figure 1, we observe that, for all n, $V\left(\widehat{\beta}_4\right) < V\left(\widehat{\beta}_1\right) < V\left(\widehat{\beta}_3\right) < V\left(\widehat{\beta}_2\right)$. Among the proposed estimators, $\widehat{\beta}_4$ has minimum variance and for $n > 22$, $V\left(\widehat{\beta}_4\right)$ is equal to $V(\widehat{\beta}_{el})$. The $V\left(\widehat{\beta}_4\right)$ is less than $V\left(\widehat{\beta}_{eh}\right)$ and for $n \leq 30$, $V\left(\widehat{\beta}_1\right)$ is less than $V\left(\widehat{\beta}_{eh}\right)$. Also, for increasing value of n, $V\left(\widehat{\beta}_k\right)$, $k = 1, 2, 3, 4$ is decreasing.

4. Performance of the proposed estimators

In this section, we study the performance of proposed estimators using relative efficiencies. The relative efficiency (RE) of two estimators, namely, A and B is given by

$$RE(A, B) = \frac{V(B)}{V(A)}. \tag{17}$$

We conclude that, A is better than B in terms of its performance if $RE(A, B) > 1$. The RE among proposed estimators for both cases where in predictor variables have unequal distance and equal distance are derived and given in Table 4. A comparison among $\widehat{\beta}_k$, $k = 1,\ 2,\ 3,\ 4$ is carried out in Table 5 in terms of computed values of RE for various values of n when $x_{(i)}$'s are equidistant. Using Table 5, RE of $\widehat{\beta}_4$ with respect to (wrt) $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, RE of $\widehat{\beta}_1$ wrt $\widehat{\beta}_2$, $\widehat{\beta}_3$ and RE of $\widehat{\beta}_3$ wrt $\widehat{\beta}_2$ are given in Figure 2.

From Table 5 and Figure 2, it is observed that, $RE\left(\widehat{\beta}_4, \widehat{\beta}_2\right) > RE(\widehat{\beta}_4, \widehat{\beta}_3) > RE(\widehat{\beta}_4, \widehat{\beta}_1)$ and $RE\left(\widehat{\beta}_1, \widehat{\beta}_2\right) > RE\left(\widehat{\beta}_1, \widehat{\beta}_3\right)$. Hence, $\widehat{\beta}_4$ is performing better than all other proposed estimators, $\widehat{\beta}_1$ performs better than $\widehat{\beta}_2$, $\widehat{\beta}_3$ and $\widehat{\beta}_3$ outperforms $\widehat{\beta}_2$. Also, RE of $\widehat{\beta}_4$ wrt $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, RE of $\widehat{\beta}_1$ wrt $\widehat{\beta}_2$ and RE of $\widehat{\beta}_3$ wrt $\widehat{\beta}_2$ increases for increasing values of n, whereas, RE of $\widehat{\beta}_1$ wrt $\widehat{\beta}_3$ decreases for $n > 14$. As $\widehat{\beta}_4$ outperforms $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$, we compute $RE(\widehat{\beta}_4, \widehat{\beta}_{el})$

Table 4: RE among proposed $\widehat{\beta}_k$, $k = 1,\ 2,\ 3,\ 4$

	For unequal distances among $x_{(i)}$'s	For equal distances among $x_{(i)}$'s
$RE(\widehat{\beta}_1, \widehat{\beta}_2)$	$\dfrac{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2}{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{i}\right)\right]^2}$	$\dfrac{\left[\sum_{i=1}^m \left(\frac{2i-1}{m-i+1}\right)\right]^2}{\left[\sum_{i=1}^m \left(\frac{2i-1}{i}\right)\right]^2}$
$RE(\widehat{\beta}_1, \widehat{\beta}_3)$	$\dfrac{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2 \sum_{i=1}^m (m-i+1)^2}{\left[\sum_{i=1}^m (m-i+1)\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2 \sum_{i=1}^m \frac{1}{(m-i+1)^2}}$	$\dfrac{48\ \left[\sum_{i=1}^m \left(\frac{2i-1}{m-i+1}\right)\right]^2}{n(n+1)(n+2)\sum_{i=1}^m \frac{1}{(m-i+1)^2}}$
$RE(\widehat{\beta}_1, \widehat{\beta}_4)$	$\dfrac{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{m-i+1}\right)\right]^2 \sum_{i=1}^m i^2}{\left[\sum_{i=1}^m i\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2 \sum_{i=1}^m \frac{1}{(m-i+1)^2}}$	$\dfrac{24(n+1)\left[\sum_{i=1}^m \left(\frac{2i-1}{m-i+1}\right)\right]^2}{n(n+2)(2n-1)^2\sum_{i=1}^m \frac{1}{(m-i+1)^2}}$
$RE(\widehat{\beta}_2, \widehat{\beta}_3)$	$\dfrac{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{i}\right)\right]^2 \sum_{i=1}^m (m-i+1)^2}{\left[\sum_{i=1}^m (m-i+1)\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2 \sum_{i=1}^m \frac{1}{i^2}}$	$\dfrac{24\left[\sum_{i=1}^m \left(\frac{2i-1}{i}\right)\right]^2}{n(n+1)(n+2)\sum_{i=1}^m \frac{1}{i^2}}$
$RE(\widehat{\beta}_2, \widehat{\beta}_4)$	$\dfrac{\left[\sum_{i=1}^m \left(\frac{x_{(m+i)}-x_{(m-i+1)}}{i}\right)\right]^2 \sum_{i=1}^m i^2}{\left[\sum_{i=1}^m i\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2 \sum_{i=1}^m \frac{1}{i^2}}$	$\dfrac{24(n+1)\left[\sum_{i=1}^m \left(\frac{2i-1}{i}\right)\right]^2}{n(n+2)(2n-1)^2\sum_{i=1}^m \frac{1}{i^2}}$
$RE(\widehat{\beta}_3, \widehat{\beta}_4)$	$\dfrac{\left[\sum_{i=1}^m (m-i+1)\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2}{\left[\sum_{i=1}^m i\left(x_{(m+i)}-x_{(m-i+1)}\right)\right]^2}$	$\dfrac{(n+1)^2}{(2n-1)^2}$

Table 5: RE among proposed estimators for various n

n	$RE(\widehat{\beta}_4, \widehat{\beta}_1)$	$RE(\widehat{\beta}_4, \widehat{\beta}_2)$	$RE(\widehat{\beta}_4, \widehat{\beta}_3)$	$RE(\widehat{\beta}_1, \widehat{\beta}_2)$	$RE(\widehat{\beta}_1, \widehat{\beta}_3)$	$RE(\widehat{\beta}_3, \widehat{\beta}_2)$
6	1.007729	2.710394	2.469380	2.689600	2.450440	1.097598
8	1.026578	3.388911	2.777778	3.301130	2.705850	1.220003
10	1.050994	4.033234	2.983472	3.837540	2.838730	1.351845
14	1.106672	5.270092	3.240021	4.762070	2.927710	1.626545
18	1.165107	6.469979	3.393396	5.553050	2.912470	1.906650
22	1.223496	7.649354	3.495281	6.251810	2.856780	2.188423
26	1.280902	8.815233	3.567861	6.881760	2.785450	2.470600
30	1.337024	9.971084	3.622270	7.457780	2.709190	2.752773
40	1.471540	12.835320	3.712642	8.722070	2.523000	3.456978
50	1.598491	15.673980	3.768181	9.804740	2.357310	4.159215
70	1.834458	21.303790	3.832739	11.613810	2.089310	5.558644
100	2.155869	29.691210	3.882138	13.772790	1.800680	7.648769

and $RE(\widehat{\beta}_4, \widehat{\beta}_{eh})$ for various values of n and furnish in Table 6.

From Table 6, we notice that, $RE(\widehat{\beta}_4, \widehat{\beta}_{eh}) > 1$, increases as n increases, stabilizes at 1.3333 and $RE(\widehat{\beta}_4, \widehat{\beta}_{el}) \cong 1$ for increasing values of n.

5. Illustration

In this section, we illustrate the performance of $\widehat{\beta}_k$, $k = 1,\ 2,\ 3,\ 4$ through some examples considered in literature. We compute $\widehat{\beta}_*$ and its variance, where $\widehat{\beta}_*$ is any estimator of β. Also, we compute $RE(\widehat{\beta}_4, \widehat{\beta}_*)$. To fit the simple linear regression model given in (1),

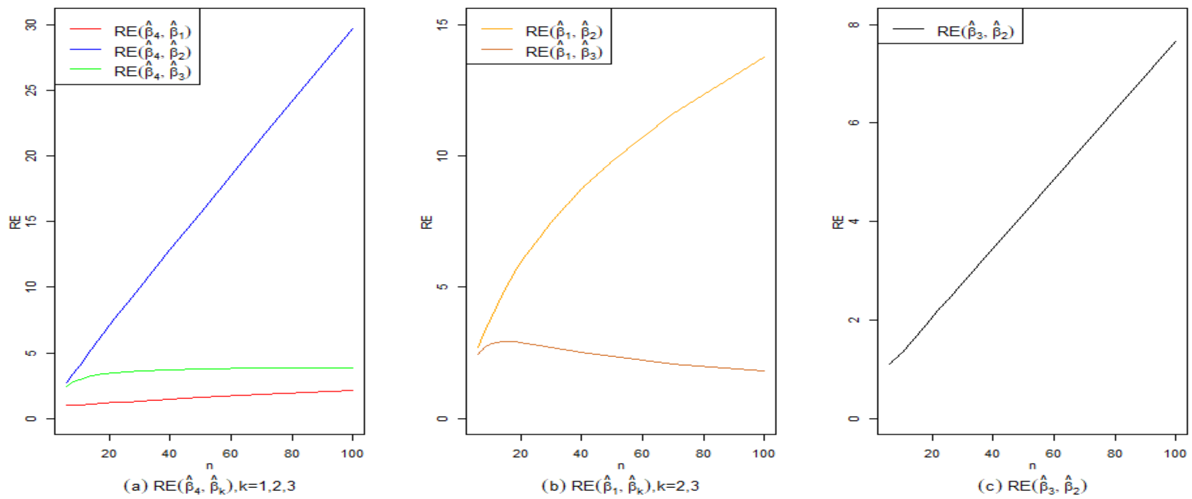


Figure 2: RE among proposed estimators

Table 6: RE of $\widehat{\boldsymbol{\beta}}_4$ wrt $\widehat{\boldsymbol{\beta}}_{el}$ and $\widehat{\boldsymbol{\beta}}_{eh}$

n	$RE(\widehat{\beta}_4, \widehat{\beta}_{el})$	$RE(\widehat{\beta}_4, \widehat{\beta}_{eh})$
6	0.987755	1.280423
8	0.992063	1.302083
10	0.994490	1.312727
14	0.996923	1.322449
18	0.998045	1.326619
22	0.998650	1.328782
26	0.999012	1.330046
30	0.999246	1.330848
40	0.999565	1.331921
50	0.999718	1.332424
70	0.999853	1.332866
100	0.999927	1.333103

the intercept parameter α is estimated using various $\widehat{\beta}_*$,

$$\widehat{\alpha}_* = \overline{y} - \widehat{\beta}_* \overline{x} \tag{18}$$

and

$$\widehat{\alpha}'_* = \widetilde{y} - \widehat{\beta}_* \widetilde{x}, \tag{19}$$

where $\widetilde{x}$, $\widetilde{y}$ are median of x and y values respectively. Using various estimators, the regression lines are fitted.

Example 1: The data due to Anscombe (1973) taken from R software consists of four datasets known as Anscombe's quartet. Here, we consider the data of third quartet given in Table 7.

Using equation (2), (5), (6), (7), (8) and (9), $\widehat{\beta}_{ul}, \widehat{\beta}_{uh}$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$ and $\widehat{\beta}_4$ and their variances are computed. The relative efficiency of $\widehat{\beta}_4$ wrt other estimators are computed.

Table 7: Third quartet due to Anscombe (1973)

x	10	8	13	9	11	14	6	4	12	7	5
y	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73

Table 8: Computed values of $\widehat{\beta}_*$, V($\widehat{\beta}_*$) and $RE(\widehat{\beta}_4, \widehat{\beta}_*)$ for example 1

$\widehat{\beta}_*$	Value of $\widehat{\beta}_*$	V($\widehat{\beta}_*$) (in 10^{-2})	$RE(\widehat{\beta}_4, \widehat{\beta}_*)$
$\widehat{\beta}_{ul}$	0.49972	0.9090 σ^2	1
$\widehat{\beta}_{uh}$	0.48700	1.1111 σ^2	1.22222
$\widehat{\beta}_1$	0.46727	0.9660 σ^2	1.06353
$\widehat{\beta}_2$	0.45175	2.9272 σ^2	3.21994
$\widehat{\beta}_3$	0.46700	2.2448 σ^2	2.46939
$\widehat{\beta}_4$	0.49972	0.9090 σ^2	-

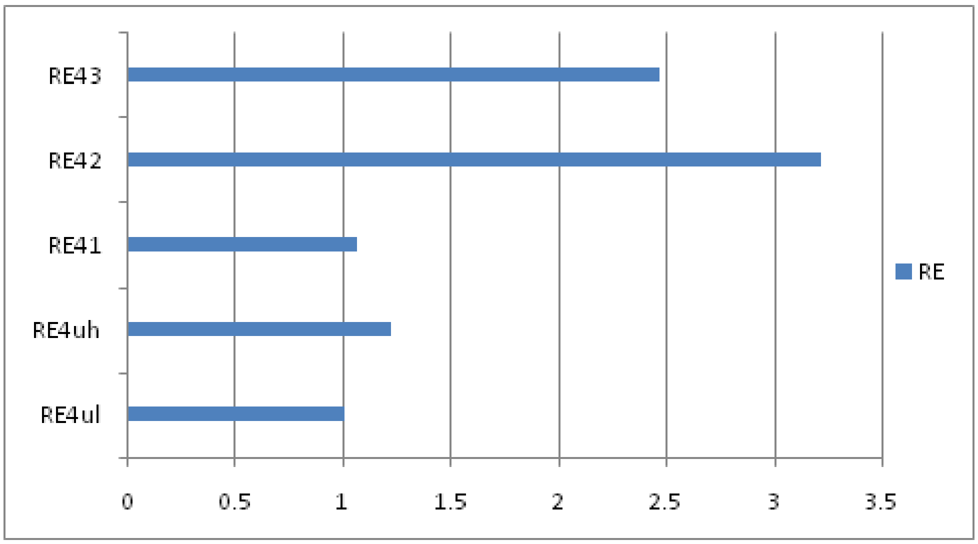
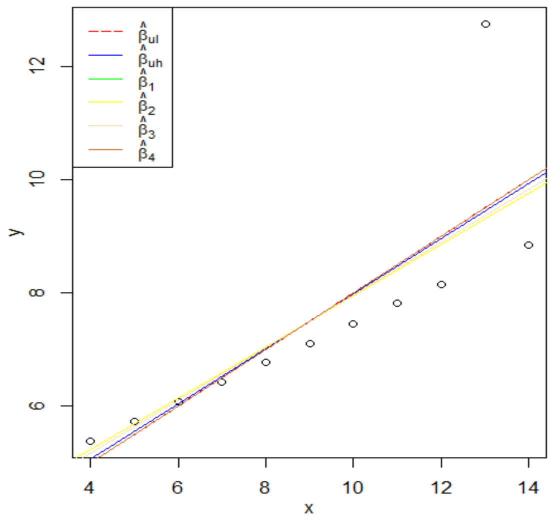


Figure 3: RE of $\widehat{\beta}_4$ wrt $\widehat{\beta}_*$

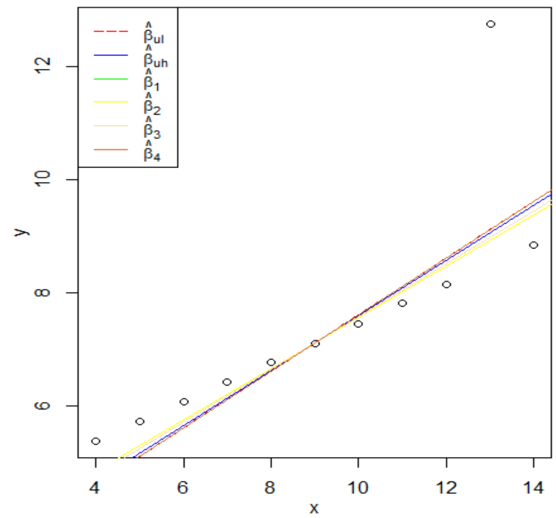
From Figure 3 and computed V($\widehat{\beta}_*$) given in Table 8, it is observed that, performance of $\widehat{\beta}_4$ and $\widehat{\beta}_{ul}$ are equivalent. Also, $\widehat{\beta}_4$ and $\widehat{\beta}_1$ are better than $\widehat{\beta}_{uh}$. From Figure 4(a) and 4(b), it is observed that, all the regression lines fitted using various $\widehat{\beta}_*$ show slight change in their slopes. In Figure 4(b), as α is estimated using $\widehat{\alpha}'_*$, we see a shift in the intercept and the outlier present in the data has not influenced the regression lines where as the influence of outlier observation is evident in Figure 4(a).

Example 2: This example is due to Montgomery *et al.* (2021) and is given in Table 9. The dataset explains, the shear strength ($Y_{\rm i}$) of bond between two types of propellant used to manufacture a rocket motor and age in weeks ($X_{\rm i}$) of the batch of propellant.

From Figure 5 and Table 10, it is observed that, the performance of $\widehat{\beta}_4$ and $\widehat{\beta}_{ul}$ is almost identical. Also, $\widehat{\beta}_1$ and $\widehat{\beta}_4$ are performing better than $\widehat{\beta}_{uh}$. From Figure (6), we observe that, various regression lines fitted using $\widehat{\alpha}_*$ differ in their intercepts than those fitted using $\widehat{\alpha}'_*$. In both cases $\widehat{\beta}_4$ and $\widehat{\beta}_{ul}$ are the lines of best fit.



(a) Regression line with $\hat{\alpha}_*$



(b) Regression line with $\hat{\alpha}'_*$

Figure 4: The fitted regression lines using $\widehat{\alpha}_*$ and $\widehat{\alpha}'_*$

Table 9: Data due to Montgomery *et al.* (2021)

Sr. no.	y	x	Sr. no.	y	x
1	2158.7	15.5	11	2165.2	13
2	1678.15	23.75	12	2399.55	3.75
3	2316	8	13	1779.8	25
4	2061.3	17	14	2336.75	9.75
5	2207.5	5.5	15	1765.3	22
6	1708.3	19	16	2353.5	18
7	1784.7	24	17	2414.4	6
8	2575	2.5	18	2200.5	12.5
9	2357.9	7.5	19	2654.2	2
10	2256.7	11	20	1753.7	21.5

Table 10: Computed values of $\widehat{\beta}_*$, V($\widehat{\beta}_*$) and $RE(\widehat{\beta}_4, \widehat{\beta}_*)$ for example 2

$\widehat{\beta}_*$	Value of $\widehat{\beta}_*$	V($\widehat{\beta}_*$) (in 10^{-2})	$RE(\widehat{\beta}_4, \widehat{\beta}_*)$
$\widehat{\beta}_{ul}$	-35.9	0.09037 σ^2	0.994988
$\widehat{\beta}_{uh}$	-34.62457	0.11788 σ^2	1.297978
$\widehat{\beta}_1$	-36.31487	0.11260 σ^2	1.240828
$\widehat{\beta}_2$	-33.28090	0.63480 σ^2	6.989736
$\widehat{\beta}_3$	-32.74453	0.29373 σ^2	3.234026
$\widehat{\beta}_4$	-35.6700	0.09082 σ^2	-

Example 3: The dataset studied by Graybill and Iyer (1994) is considered. The variable y is average systolic blood pressure (BP) at 8 A.M. over two weeks and x is age of individuals ranging 21 to 70 years. The dataset is given in Table 11.

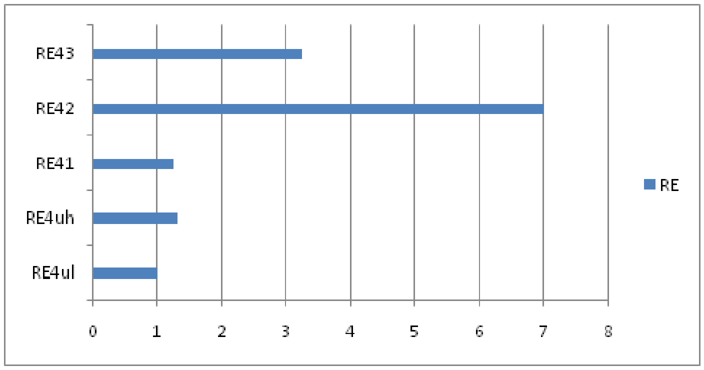


Figure 5: RE of $\widehat{\beta}_4$ wrt $\widehat{\beta}_*$

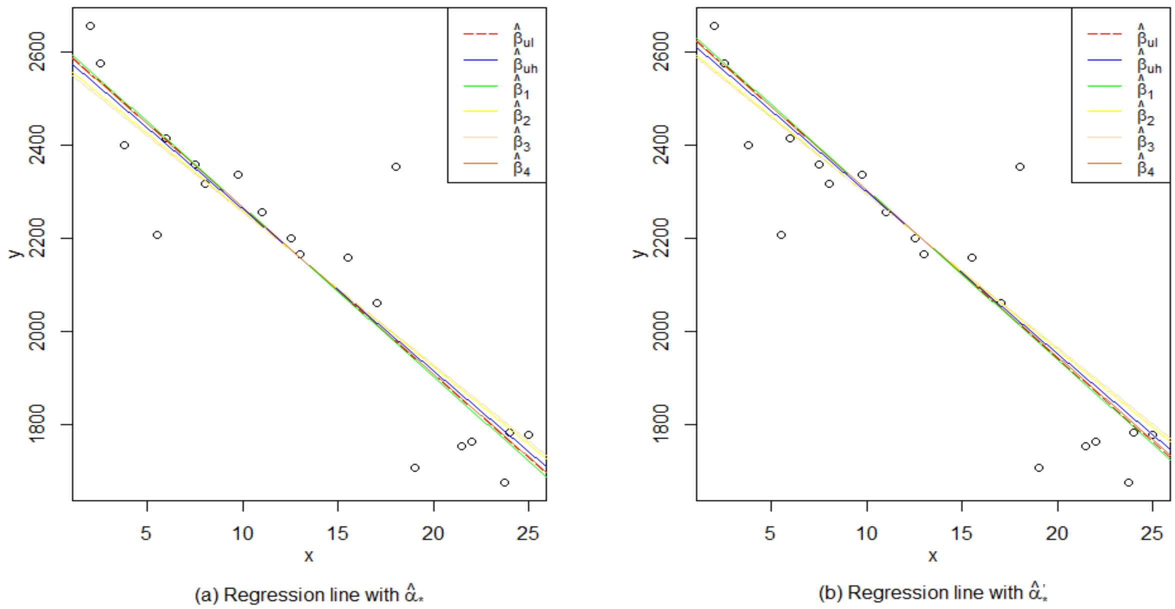


Figure 6: The fitted regression lines using $\widehat{\alpha}_*$ and $\widehat{\alpha}_*'$

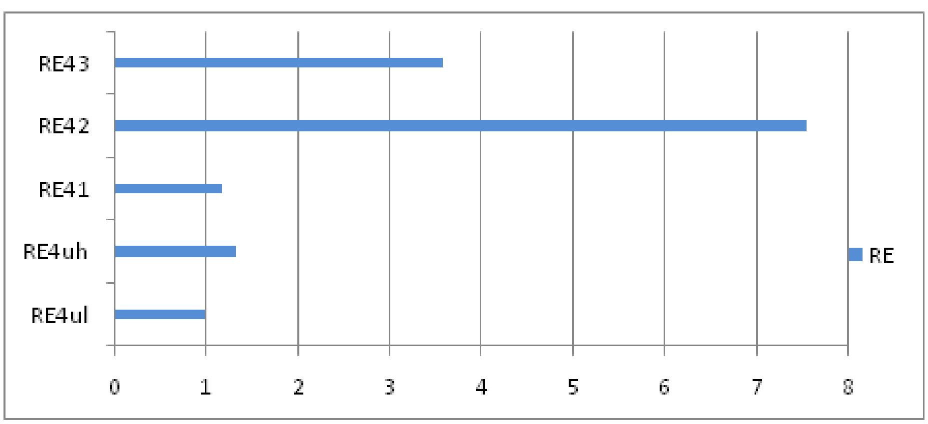


Figure 7: RE of $\widehat{\beta}_4$ wrt $\widehat{\beta}_*$

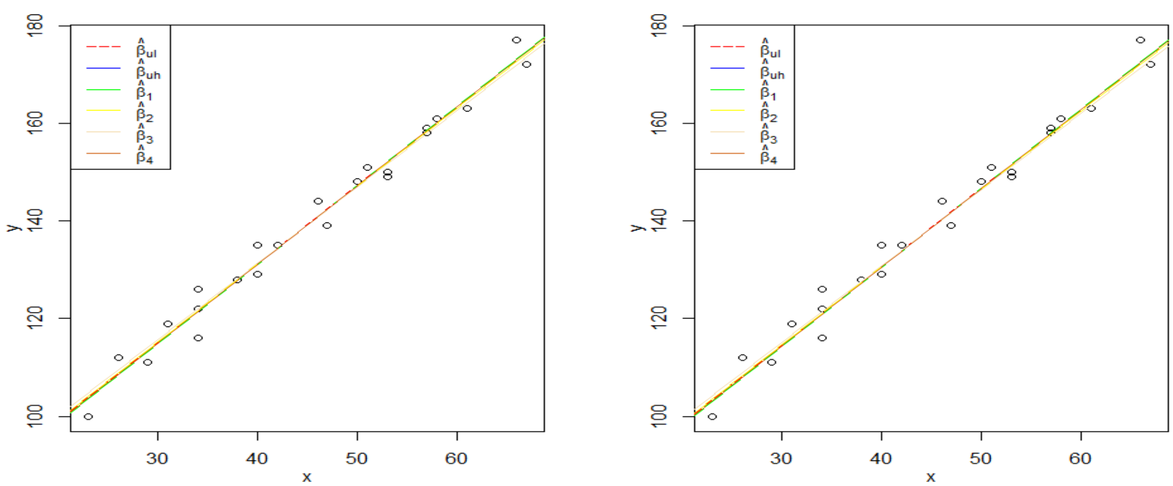
From computed values of $\widehat{\beta}_*$, V$(\widehat{\beta}_*)$ given in Table 12 and Figure 7, it is observed that, all the values of $\widehat{\beta}_*$ are nearly same. $\widehat{\beta}_4$ performs better than $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, $\widehat{\beta}_{uh}$ and is

Table 11: Data due to Graybill and Iyer (1994)

Sr. no.	x	y	Sr. no.	x	y
1	34	116	13	40	135
2	26	112	14	34	126
3	51	151	15	67	172
4	58	161	16	23	100
5	34	122	17	47	139
6	40	129	18	42	135
7	31	119	19	61	163
8	57	158	20	38	128
9	56	144	21	57	159
10	53	150	22	66	177
11	29	111	23	42	135
12	50	148	24	53	149

Table 12: Computed values of $\widehat{\beta}_*$, V($\widehat{\beta}_*$) and $RE(\widehat{\beta}_4, \widehat{\beta}_*)$ for example 3

$\widehat{\beta}_*$	Value of $\widehat{\beta}_*$	V($\widehat{\beta}_*$) (in 10^{-2})	$RE(\widehat{\beta}_4, \widehat{\beta}_*)$
$\widehat{\beta}_{ul}$	1.60900	0.02771 σ^2	0.98823
$\widehat{\beta}_{uh}$	1.59288	0.03749 σ^2	1.33702
$\widehat{\beta}_1$	1.61958	0.03296 σ^2	1.17546
$\widehat{\beta}_2$	1.59735	0.21211 σ^2	7.56455
$\widehat{\beta}_3$	1.56162	0.10074 σ^2	3.59272
$\widehat{\beta}_4$	1.60938	0.02804 σ^2	-



(a) Regression line with $\hat{\alpha}_*$ (b) Regression line with $\hat{\alpha}'_*$

Figure 8: The fitted regression lines using $\widehat{\alpha}_*$ and $\widehat{\alpha}'_*$

almost equivalent to $\widehat{\beta}_{ul}$. Also, $\widehat{\beta}_1$ performs better than $\widehat{\beta}_{uh}$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$. From Figure 8(a) and 8(b), we observe that, all the regression lines plotted using various $\widehat{\beta}_*$, $\widehat{\alpha}_*$ and $\widehat{\alpha}'_*$ are identical.

Example 4: The data is taken from from `nseindia.com` and `bseindia.com`. It explains daily closing price (x) of index NIFTY50 from National Stock Exchange (NSE) and daily closing price (y) of index SENSEX50 from Bombay Stock Exchange (BSE). The data consists of 988 observations of 4 years from 2017 to 2020. Here, we furnish the values of estimators, their variances and relative efficiencies along with fitting of regression lines using various estimators.

Table 13: Computed values of $\widehat{\beta}_*$, V($\widehat{\beta}_*$) and $RE(\widehat{\beta}_4, \widehat{\beta}_*)$ for example 4

$\widehat{\beta}_*$	Value of $\widehat{\beta}_*$	V($\widehat{\beta}_*$) (in 10^{-10})	$RE(\widehat{\beta}_4, \widehat{\beta}_*)$
$\widehat{\beta}_{ul}$	1.05800	8.593147 σ^2	0.94391
$\widehat{\beta}_{uh}$	1.06240	13.97219 σ^2	1.53477
$\widehat{\beta}_1$	1.00837	38.82940 σ^2	4.26520
$\widehat{\beta}_2$	1.09918	3754.035 σ^2	412.3605
$\widehat{\beta}_3$	1.06689	57.23892 σ^2	6.28739
$\widehat{\beta}_4$	1.06061	9.10377 σ^2	-

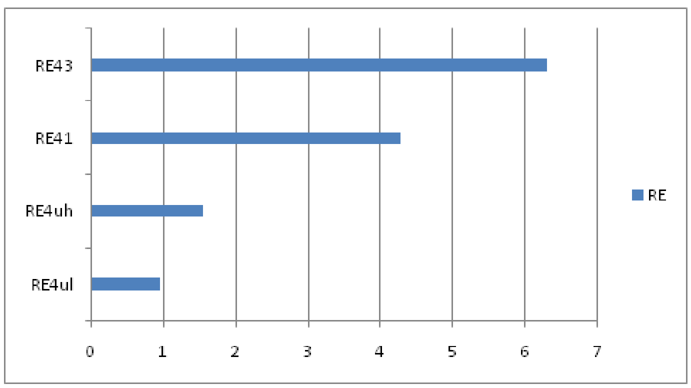
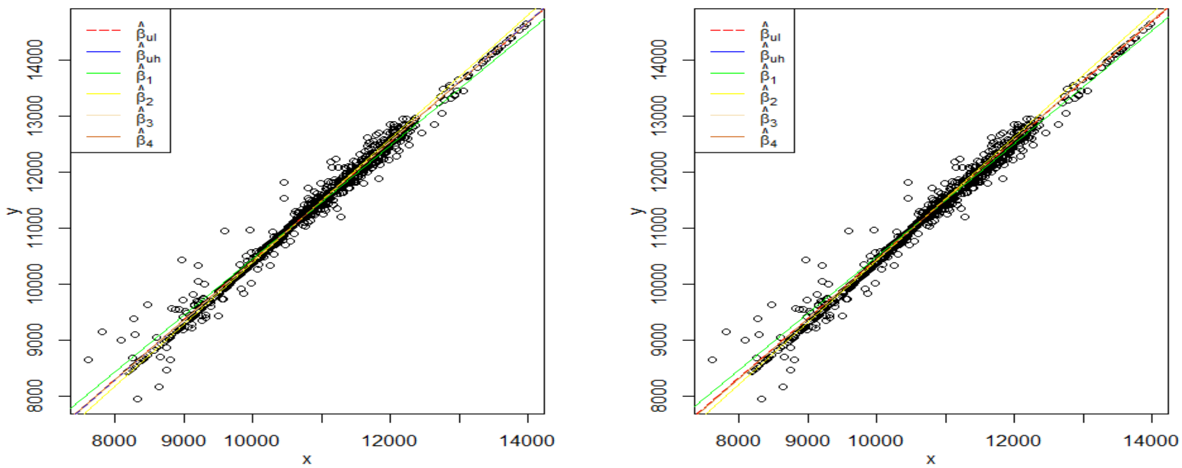


Figure 9: RE of $\widehat{\beta}_4$ wrt $\widehat{\beta}_*$

From Table 13, the computed values of $\widehat{\beta}_*$ and $V(\widehat{\beta}_*)$, $RE(\widehat{\beta}_4, \widehat{\beta}_2)$ is too high to record in Figure 9. The proposed estimator, $\widehat{\beta}_4$ performs better than $\widehat{\beta}_1$, $\widehat{\beta}_3$ and $\widehat{\beta}_{uh}$. From Figure 10, we notice that all the regression lines fitted either using $\widehat{\alpha}_*$ or $\widehat{\alpha}'_*$ seem to be the same as number of observations are very large.

6. Conclusions

- Some estimators based on quasi ranges are proposed for slope parameter of simple linear regression model, $y_i = \alpha + \beta x_i + e_i, \quad i = 1,\ 2,\ \cdots,\ n$.
- Among the proposed estimators, *viz.* $\widehat{\beta}_k,\ k = 1,\ 2,\ \cdots,\ 6$ based on weighted sample quasi ranges, $\widehat{\beta}_5$ reduces to $\widehat{\beta}_3$ and $\widehat{\beta}_6$ reduces to $\widehat{\beta}_4$.
- When equal weights are assigned to each quasi range, all the proposed estimators reduce to $\widehat{\beta}_{ud}$.



(a) Regression line with $\hat{\alpha}_*$ (b) Regression line with $\hat{\alpha}'_*$

Figure 10: The fitted regression lines using $\widehat{\alpha}_*$ and $\widehat{\alpha}'_*$

- For equal weights and equidistant $x_{(i)}$'s, all the proposed estimators reduce to $\widehat{\beta}_{eh}$, due to Bose (1938).

- All the proposed estimators are unbiased estimators of slope parameter β.

- The variance of proposed estimators is decreasing with the increasing values of n.

- Among the estimators proposed, $\widehat{\beta}_4$ outperforms $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$; $\widehat{\beta}_1$ outperforms $\widehat{\beta}_2$, $\widehat{\beta}_3$ and $\widehat{\beta}_3$ outperforms $\widehat{\beta}_2$.

- $RE\left(\widehat{\beta}_4, \widehat{\beta}_1\right)$, $RE(\widehat{\beta}_4, \widehat{\beta}_2)$, $RE(\widehat{\beta}_4, \widehat{\beta}_3)$, $RE(\widehat{\beta}_1, \widehat{\beta}_2)$ and $RE(\widehat{\beta}_3, \widehat{\beta}_2)$ increase as n increases, but $RE(\widehat{\beta}_1, \widehat{\beta}_3)$ increases upto $n = 14$ and decreases for $n > 14$.

- $\widehat{\beta}_4$ outperforms $\widehat{\beta}_{uh}$, due to Bhat and Bijjargi (2023) and its performance is equivalent to least square estimate $\widehat{\beta}_{ul}$.

- As a_{4i} and a_{1i} assign relatively heavier weights to quasi ranges with extreme order statistics, the estimators $\widehat{\beta}_4$ based on a_{4i} and $\widehat{\beta}_1$ based on a_{1i} are relatively more efficient than other estimators.

- $\widehat{\beta}_2$ based on a_{2i} and $\widehat{\beta}_3$ based on a_{3i} exhibit robustness to outliers if present in the data, since a_{2i} and a_{3i} assign lower weights to quasi ranges with extreme order statistics.

Acknowledgements

Authors thank the unknown referee for his useful comments and second author acknowledges Karnatak University, Dharwad for the award of University Research Scholarship.

References

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, **27**, 17–21.

Bhat, S. V. and Bijjargi, S. M. (2023). Nonparametric approach to estimation in linear regression. *International Journal of Current Research*, **15**, 24419–24424.

Bose, S. S. (1938). Relative efficiencies of regression coefficients estimated by the method of finite differences. *Sankhyā: The Indian Journal of Statistics*, **3**, 339–346.

Cliff, K. R. and Billy, K. M. (2017). Estimation of the parameters of a linear regression system using the simple averaging method. *Global Journal of Pure and Applied Mathematics*, **13**, 7749–7758.

Euler, L. (1749). Recherches sur la question des inégalités du mouvement de saturne et de jupiter. *Pièce qui ont remporté le prix de l'académie royale des sciences*, **3**, 1–123.

Gauss, C. F. (1809). Theoria motus corporum coelestum. *Werke*, **3**, 1–123.

Govindarajulu, Z. (2007). *Nonparametric Inference*. World Scientific.

Graybill, F. A. and Iyer, H. K. (1994). *Regression Analysis: Concepts and Applications*. Duxbury Press.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley Interscience.

Harter, H. L. (1959). The use of sample quasi-ranges in estimating population standard deviation. *The Annals of Mathematical Statistics*, **30**, 980–999.

Harter, W. L. (1974). The method of least squares and some alternatives: Part i. *International Statistical Review/Revue Internationale de Statistique*, **42**, 147–174.

Jlibene, M., Taoufik, S., and Benjelloun, S. (2021). *Analysis of Least square estimator for simple Linear Regression with a uniform distribution error*. arXiv preprint arXiv:2111.04200 –arxiv.org.

Legendre, A. M. (1805). *Mémoire sur les opérations trigonométriques: dont les résultats dépendent de la figure de la terre*. F. Didot.

Liu, X. and Preve, D. (2016). Measure of location-based estimators in simple linear regression. *Journal of Statistical Computation and Simulation*, **86**, 1771–1784.

Mayer, J. T. (1750). Abhandlung über die umwälzung des monds und seine axe. *Kosmographische Nachrichten un Sammlungen*, **1**, 52–183.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Mosteller, F. (2006). *On Some Useful "Inefficient" Statistics*. Springer.

Nair, K. and Shrivastava, M. (1942). On a simple method of curve fitting. *Sankhyā: The Indian Journal of Statistics*, **6**, 121–132.

Prabowo, A., Sugandha, A., Tripena, A., Mamat, M., Firman, S., and Budiono, R. (2020). A new method to estimate parameters in the simple regression linear equation. *Mathematics and Statistics*, **8**, 75–81.

Singthongchai, j., Thongmual, N., and Nitisuk, N. (2021). An improved simple averaging approach for estimating parameters in simple linear regression model. *Mathematics and Statistics*, **9**, 939–946.

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, **11**, 284–300.

Yao, D.-S., Chen, W.-X., and Long, C.-X. (2021). Parametric estimation for the simple linear regression model under moving extremes ranked set sampling design. *Applied Mathematics-A Journal of Chinese Universities*, **36**, 269–277.