

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(4): 117-124
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 17-06-2024
Accepted: 23-07-2024

Shrinath M Bijjargi
Department of Statistics,
Karnatak University, Dharwad,
Karnataka, India

Sharada V Bhat
Department of Statistics,
Karnatak University, Dharwad,
Karnataka, India

Quasi range dependent weighted classes of linear slope estimators

Shrinath M Bijjargi and Sharada V Bhat

DOI: <https://dx.doi.org/10.22271/maths.2024.v9.i4b.1782>

Abstract

The classes of adaptive estimators are proposed to the slope parameter of simple linear regression (SLR) model. These classes are based on the weighted sample slopes where in both slopes and their weights are depending on quasi ranges of predictor variables. The mean, variance of the proposed classes of estimators are obtained and their variances are minimized using numerical method. The performance of some members of these classes are discussed in terms of their relative efficiencies. A simulation study is also carried out to check their behaviour. An illustrative example is provided.

Keywords: Adaptive, predictor variable, quasi range, simple linear regression, slope parameter, weighted average

1. Introduction

Regression analysis serves as a statistical tool to explore the relationship between predictor (independent) variables and response (dependent) variable. It aims to envisage the value of the dependent variable based on independent variables. The SLR explores linear relationship between a single predictor variable and a response variable. It is established by the linear equation given by

$$y_i = \alpha + \beta x_i + e_i, 1 \leq i \leq n \quad (1)$$

where α is intercept parameter, β is slope parameter, y_i is response variable, x_i is predictor variable, e_i is independent and identically distributed (iid) random error from continuous distribution with distribution function $F(\cdot)$ having zero mean and finite variance σ^2 . The goal is to estimate α , β that best fit the observed data and make predictions based on this relationship. The slope parameter, β quantifies the average variation in y in response to a unit change in x .

Numerous estimators exist in the statistical literature to estimate β with least square (ls) estimator due to Legendre (1805) ^[10] and Gauss (1809) ^[9] being first of its kind. It is given by

$$\hat{\beta}_{ls} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2)$$

and its variance is given by

$$V(\hat{\beta}_{ls}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3)$$

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

Under the assumption of equal distances (say d) of x_i s, among the methods proposed by Bose (1938) ^[5] to estimate β , the method of differences at half range that outperforms all other estimators is given by

Corresponding Author:
Shrinath M Bijjargi
Department of Statistics,
Karnatak University, Dharwad,
Karnataka, India

$$\hat{\beta}_{eh} = \frac{\sum_{i=1}^m (y_{m+i} - y_i)}{m^2 d} \quad (4)$$

and its variance is

$$V(\hat{\beta}_{eh}) = \frac{16\sigma^2}{n^3 d^2} \quad (5)$$

Theil (1950) ^[15] proposed an estimator for β based on the median of pairwise slopes $b_i = \frac{y_j - y_i}{x_j - x_i}$, $1 \leq i < j \leq n$ resulting in $\frac{n(n-1)}{2}$ slopes. This method known as Theil's complete method is robust against outliers. Olkin and Yitzhaki (1992) ^[11] and Birkes and Dodge (1993) ^[4] discussed on $\hat{\beta}_{oy}$, incorporating weights into the slopes to refine this estimation process. By using the weights, $w_i = \frac{(x_j - x_i)^2}{\sum (x_j - x_i)^2}$, they showed that the weighted estimator $\hat{\beta} = \sum w_i b_i$ reduces to the least squares estimator.

Cliff and Billy (2017) ^[7] proposed an estimator for β using simple average method which is based on the average of successive slopes, $s_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$, $i = 2, \dots, n$. Prabowo et. al. (2020) ^[12] simplified this procedure and investigated its performance. Singthongchai et. al. (2021) ^[14] improved the simple average method to estimate β . Bhat and Bijjargi (2023) ^[1] generalized Bose's estimators along with some modified estimators when x_i 's have unequal distances. Among their estimators, $\hat{\beta}_{ud}$ based on quasi ranges has better performance than others and is equivalent to $\hat{\beta}_{uh}$ based on half range. The estimator $\hat{\beta}_{ud}$ is given by

$$\hat{\beta}_{ud} = \frac{\sum_{i=1}^m (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}}. \quad (6)$$

Here, $q_i = x_{(n-i)} - x_{(i+1)}$, $i = 1, 2, \dots, m-1$ is i^{th} quasi range among a set of observations (x_j, y_j) , $j = 1, \dots, n$ arranged in ascending order based on x_j , $x_{(j)}$ being j^{th} order statistic and y_j^* being the corresponding y value and $n = 2m$. Further, $q_0 = x_{(n)} - x_{(1)}$ is the range of n observations. Bhat and Bijjargi (2024a) ^[2] proposed adaptive $\hat{\beta}_{ud}$ estimators using various kinds of weights. Also, Bhat and Bijjargi (2024b) ^[3] developed two classes of estimators based on ratio of weighted average and weighted average of ratio of y observations to the quasi ranges of x observations. However, these weights do not account for magnitude of the x observations which can restrict the effect of original value of x .

In this paper, we propose the classes of adaptive estimators of β based on weighted average of quasi ranges and that of slopes obtained using quasi ranges with weights too depending on quasi ranges of predictor variables. The proposed classes of estimators are given in section 2. Their mean and variance are derived in section 3. The special cases of the proposed classes are discussed in section 4. Section 5 contains comparison among some members of these classes of estimators and are illustrated through examples in section 6. We record our conclusions based on observations in section 7.

2. Classes of adaptive linear slope estimators

The weighted average is a measure of central tendency. It takes into account the significance of each value by assigning different weights to them. The estimators based on weighted averages are desired in situations where different data points have varying degrees of importance. Hence, it is useful in

financial modeling or economic analysis, forecasting stock prices, epidemiology, market research or customer analytics and regression analysis to address the intricacies of heteroscedasticity, outliers and influential points.

The estimators due to Bhat and Bijjargi (2024a and 2024b) ^[2-3] are assigned with the weights that are independent of actual x values relying solely on the position of quasi range. The assignment of weights remains uniform across all datasets, irrespective of the magnitude of i^{th} quasi range which is again dependent on x values. This means that, regardless of whether the dataset comprises of high-magnitude or low-magnitude values, the weights assigned remain constant and lacks sensitivity to the magnitude of the observations.

On similar lines to the development of $\hat{\beta}_{oy}$ which has varying weights to account for data variability, the proposed method also assigns varying weights. However, unlike this method, that requires calculating $\frac{n(n-1)}{2}$ slopes and their weighted average, the proposed method simplifies the computation by requiring only $\frac{n}{2}$ slopes and their weighted average. This reduction in computational complexity makes the estimator more feasible with maintaining robustness and improvement in the overall performance. Additionally, $\hat{\beta}_{oy}$ considers all pairwise slopes which do not effectively distinguish between extreme and middle observations. But, incorporating the magnitude of distances obtained from quasi ranges as weights, the proposed classes of estimators aim to reflect the true variability and spread within the data distinguishing extreme and central observations. Hence, the proposed method enables a customized weighting system that prioritizes observations based on their positions in the dataset. This leads to a representative and accurate slope estimates.

In this section, we introduce two distinct classes of adaptive linear slope estimators viz. $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ for estimating β in SLR model. Both the classes are based on quasi ranges of x variable and their weights also consider the quasi ranges of x variable. The classes $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ are given by

$$\hat{\beta}_{1D_k} = \frac{\sum_{i=1}^m w_i^k (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m w_i^k q_{m-i}} = \frac{\sum_{i=1}^m w_i^k (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m w_i^{k+1}} \quad (7)$$

$$\hat{\beta}_{2D_k} = \sum_{i=1}^m w_i' b_i \quad (8)$$

where

$$w_i = (x_{(m+i)} - x_{(m-i+1)}) = q_{m-i}, b_i = \frac{y_{m+i}^* - y_{m-i+1}^*}{q_{m-i}} \text{ for } i =$$

$1, \dots, m$ are slopes, $w_i' = \frac{w_i^k}{\sum_{i=1}^m w_i^k}$, $\sum_{i=1}^m w_i' = 1$ and $k \in \mathbb{R}$ is known finite constant.

The weighting scheme in proposed classes of estimators is motivated by a desire to enhance efficiency, robustness and adaptability in estimation of β . The emphasis on extreme or middle observations is modulated by the value k , allowing for a flexible weighting scheme that adapts to the characteristics of the data. The weight assigned in both the classes is determined by the expression $w_i^k = q_{m-i}^k$. Since the x_i values are ordered, all the q_{m-i} are positive and increases as i goes from 1 to m . The exponent k is instrumental in defining the weighting scheme. For $k > 0$, the weights assigned to the quasi ranges increase with i . That is, q_{m-i} obtained from extreme x_i observations (those farther from the median) will have higher weights. For $k < 0$, the weights assigned to the quasi ranges decrease with i . That is, q_{m-i} derived from

observations closer to the median will have higher weights which enhances robustness against outliers. When predictor variables are equidistant,

$$q_{m-i} = x_{(m+i)} - x_{(m-i+1)} = (2i-1)d \quad \text{and} \quad \sum_{i=1}^m q_{m-i} = d \sum_{i=1}^m (2i-1).$$

3. Mean and variance of proposed classes of estimators

In this section, the mean and variance of proposed classes of estimators are obtained. The mean of class $\hat{\beta}_{1D_k}$ is given by

$$\begin{aligned} E(\hat{\beta}_{1D_k}) &= E\left(\frac{\sum_{i=1}^m w_i^k (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m w_i^k q_{m-i}}\right) \\ &= \frac{\sum_{i=1}^m w_i^k E(y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m w_i^k q_{m-i}} \\ &= \frac{\sum_{i=1}^m w_i^k (\alpha + \beta x_{(m+i)} - \alpha - \beta x_{(m-i+1)})}{\sum_{i=1}^m w_i^k q_{m-i}} \\ &= \beta \end{aligned}$$

and the mean of $\hat{\beta}_{2D_k}$ is given by

$$\begin{aligned} E(\hat{\beta}_{2D_k}) &= E(\sum_{i=1}^m w_i^k b_i) \\ &= E\left(\frac{\sum_{i=1}^m w_i^k b_i}{\sum_{i=1}^m w_i^k}\right) \\ &= \frac{\sum_{i=1}^m w_i^k \beta}{\sum_{i=1}^m w_i^k} \because E(b_i) = \beta \text{ for all } i = 1, 2, \dots, m. \\ &= \beta \end{aligned}$$

Both $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ admit unbiased estimators of β . The variance of $\hat{\beta}_{1D_k}$ is given by

$$\begin{aligned} V(\hat{\beta}_{1D_k}) &= V\left(\frac{\sum_{i=1}^m w_i^k (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m w_i^k q_{m-i}}\right) \\ &= \frac{\sum_{i=1}^m w_i^{2k} V(y_{m+i}^* - y_{m-i+1}^*)}{(\sum_{i=1}^m w_i^k q_{m-i})^2} \\ &= \frac{2\sigma^2 \sum_{i=1}^m w_i^{2k}}{(\sum_{i=1}^m w_i^k q_{m-i})^2} \\ &= \frac{2\sigma^2 \sum_{i=1}^m q_{m-i}^{2k}}{(\sum_{i=1}^m q_{m-i}^{k+1})^2} \because w_i = q_{m-i} \end{aligned} \quad (9)$$

When $x_{(i)}$ s are equidistant, $V(\hat{\beta}_{1D_k})$ can be written as,

$$V(\hat{\beta}_{1D_k}) = \frac{2\sigma^2 \sum_{i=1}^m ((2i-1)d)^{2k}}{(\sum_{i=1}^m ((2i-1)d)^{k+1})^2}. \quad (10)$$

Let $b_i = \frac{y_{m+i}^* - y_{m-i+1}^*}{q_{m-i}} = \beta + u_i$, where, $u_i = \frac{e_{m+i} - e_{m-i+1}}{q_{m-i}}$ is continuous random variable symmetric around zero. The variance of $\hat{\beta}_{2D_k}$ is given by

$$\begin{aligned} V(\hat{\beta}_{2D_k}) &= V\left(\frac{\sum_{i=1}^m w_i^k b_i}{\sum_{i=1}^m w_i^k}\right) \\ &= \frac{1}{(\sum_{i=1}^m w_i^k)^2} \left(\sum_{i=1}^m (w_i^k)^2 \frac{2\sigma^2}{(q_{m-i})^2}\right) \end{aligned}$$

$$\begin{aligned} &= \frac{2\sigma^2}{(\sum_{i=1}^m w_i^k)^2} \left(\sum_{i=1}^m \frac{(w_i^k)^2}{(q_{m-i})^2}\right) \\ &= \frac{2\sigma^2}{(\sum_{i=1}^m q_{m-i}^k)^2} \left(\sum_{i=1}^m \frac{(q_{m-i}^k)^2}{(q_{m-i})^2}\right) \because w_i = q_{m-i} \end{aligned} \quad (11)$$

When $x_{(i)}$'s are equidistant,

$$\begin{aligned} V(\hat{\beta}_{2D_k}) &= \frac{2\sigma^2}{(\sum_{i=1}^m ((2i-1)d)^k)^2} \left(\sum_{i=1}^m \frac{((2i-1)d)^{2k}}{((2i-1)d)^2}\right) \\ &= \frac{2\sigma^2}{(\sum_{i=1}^m ((2i-1)d)^k)^2} (\sum_{i=1}^m ((2i-1)d)^{2(k-1)}) \end{aligned} \quad (12)$$

Also, when $x_{(i)}$'s are evenly spaced, the least square estimator given in (2) reduces to

$$\hat{\beta}_{ls} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{d^2 \frac{n(n^2-1)}{12}} \quad (13)$$

and its variance is given by

$$V(\hat{\beta}_{ls}) = \frac{12\sigma^2}{d^2 n(n^2-1)}. \quad (14)$$

As variance expressions of two classes of proposed estimators are in complex forms, we obtain optimum k that yields minimum variances in (10) and (12) using a derivative-free optimization algorithm due to Brent (1971) [6]. This numerical method is widely known for its robustness and efficiency, often outperforming other optimization algorithms in terms of convergence speed and accuracy. The minimum variances of $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ through Brent's algorithm are obtained as per steps given below.

Step 1: Initialization

The objective function $f(k)$, representing either $V(\hat{\beta}_{1D_k})$ or $V(\hat{\beta}_{2D_k})$ is defined. Initial points $[a, b]$ such as $[-10, 10]$, $[-50, 50]$, $[0, 50]$, etc. where minimum $f(k)$ is expected to lie are taken to explore different regions of the search space. The tolerance level $\epsilon = 10^{-8}$ is considered to determine the accuracy of the solution. $f(a)$ and $f(b)$ are stored and an initialization of $c = a$ is taken to set the auxiliary point for interpolation initially to a . Then, $f(c) = f(a)$ set to store the function value at the auxiliary point c . The initial width of the interval: $d = b - a$ is calculated and $e = d$ is set to initialize an auxiliary variable for checking the convergence.

Step 2: Iteration process

While $|b - a| > \epsilon$ and $|f(b)| > \epsilon$, k is obtained using inverse quadratic interpolation or the secant method. If k is outside $[a, b]$ or not providing sufficient convergence then the bisection $k = \frac{a+b}{2}$ is performed and $f(k)$ is updated.

Step 3: Function evaluation and interval update

$f(k)$ is evaluated and the points are updated as per $f(k) < f(b)$ or $f(k) \geq f(b)$.

Step 4: Convergence check

The process is repeated until the width of an interval $|b - a| < \epsilon$, that is, the function value at the trial point becomes smaller than a predefined tolerance $|f(k)| < \epsilon$ and the maximum number of iterations N is reached.

Step 5: Output

Once the convergence is achieved, the value of k corresponding to the minimum of the function $f(k)$ is returned within the refined interval $[a, b]$. Thus, the value of k that minimizes $V(\hat{\beta}_{1D_k})$ and $V(\hat{\beta}_{2D_k})$ is achieved and we

observe that, the minimum value of $V(\hat{\beta}_{1D_k})$ is at $k = 1$ and $V(\hat{\beta}_{2D_k})$ is at $k = 2$.

In Figure 1, we plot $V(\hat{\beta}_{1D_k})$ and $V(\hat{\beta}_{2D_k})$ given in (10) and (12) for $d = 1, \sigma = 1$ and various values of m and k .

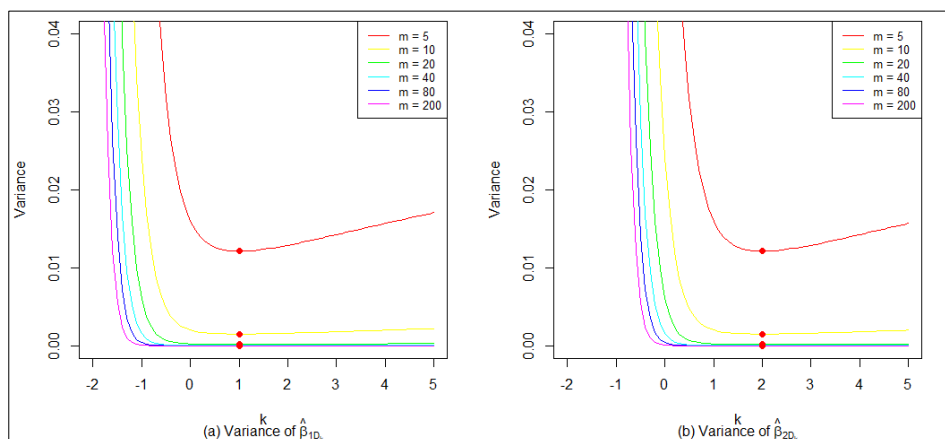


Fig 1: $V(\hat{\beta}_{1D_k})$ and $V(\hat{\beta}_{2D_k})$ for various values of m and k

The Figure 1 indicates that, the variance decreases as m increases. Also, the point marked on the line denotes the minimum value of the variance observed for $\hat{\beta}_{1D_1}$ and $\hat{\beta}_{2D_2}$.

Further, for $k = -2, -1, 0, 1, 2, 3$ and various values of n , we furnish the computations of $V(\hat{\beta}_{1D_k})$, $V(\hat{\beta}_{ls})$, $V(\hat{\beta}_{eh})$ and $V(\hat{\beta}_{2D_k})$ in Table 1 and Table 2 respectively.

Table 1: $V(\hat{\beta}_{1D_k})$, $V(\hat{\beta}_{ls})$, $V(\hat{\beta}_{eh})$ for various values of n and k

n	$V(\hat{\beta}_{1D_{-2}})$	$V(\hat{\beta}_{1D_{-1}})$	$V(\hat{\beta}_{1D_0})$	$V(\hat{\beta}_{1D_1})$	$V(\hat{\beta}_{1D_2})$	$V(\hat{\beta}_{1D_3})$	$V(\hat{\beta}_{ls})$	$V(\hat{\beta}_{eh})$
8	0.722065	0.146440	0.031250	0.023810	0.025267	0.027745	0.023810	0.031250
10	0.635174	0.094709	0.016000	0.012121	0.012887	0.014236	0.012121	0.016000
14	0.530860	0.048900	0.005831	0.004396	0.004681	0.005197	0.004396	0.005831
20	0.445927	0.024174	0.002000	0.001504	0.001603	0.001784	0.001504	0.002000
26	0.395791	0.014373	0.000910	0.000684	0.000729	0.000812	0.000684	0.000910
40	0.330041	0.006106	0.000250	0.000188	0.000200	0.000223	0.000188	0.000250
50	0.302237	0.003916	0.000128	0.000096	0.000102	0.000114	0.000096	0.000128
70	0.266510	0.002003	0.000047	0.000035	0.000037	0.000042	0.000035	0.000047
100	0.235137	0.000983	0.000016	0.000012	0.000013	0.000014	0.000012	0.000016
150	0.205760	0.000437	0.000005	0.000004	0.000004	0.000004	0.000004	0.000005
200	0.188132	0.000246	0.000002	0.000002	0.000002	0.000002	0.000002	0.000002

Table 2: $V(\hat{\beta}_{2D_k})$ for various values of n and k

n	$V(\hat{\beta}_{2D_{-2}})$	$V(\hat{\beta}_{2D_{-1}})$	$V(\hat{\beta}_{2D_0})$	$V(\hat{\beta}_{2D_1})$	$V(\hat{\beta}_{2D_2})$	$V(\hat{\beta}_{2D_3})$
8	1.459345	0.722065	0.146440	0.031250	0.023810	0.025267
10	1.429069	0.635174	0.094709	0.016000	0.012121	0.012887
14	1.395438	0.530860	0.048900	0.005831	0.004396	0.004681
20	1.370900	0.445927	0.024174	0.002000	0.001504	0.001603
26	1.357931	0.395791	0.014373	0.000910	0.000684	0.000729
40	1.343019	0.330041	0.006106	0.000250	0.000188	0.000200
50	1.337540	0.302237	0.003916	0.000128	0.000096	0.000102
70	1.331318	0.266510	0.002003	0.000047	0.000035	0.000037
100	1.326679	0.235137	0.000983	0.000016	0.000012	0.000013
150	1.323087	0.205760	0.000437	0.000005	0.000004	0.000004
200	1.321297	0.188132	0.000246	0.000002	0.000002	0.000002

From Table 1 and 2, it is observed that, the $V(\hat{\beta}_{1D_k})$ decreases upto $k = 1$, after which it increases for $k > 1$, whereas, the $V(\hat{\beta}_{2D_k})$ decreases upto $k = 2$, after which it increases for $k > 2$. Hence, among the classes of proposed estimators, $V(\hat{\beta}_{1D_k})$ and $V(\hat{\beta}_{2D_k})$ has minimum variance at $k = 1$ and $k = 2$ respectively. Also, $V(\hat{\beta}_{1D_0}) = V(\hat{\beta}_{eh})$, $V(\hat{\beta}_{2D_1}) = V(\hat{\beta}_{eh})$, $V(\hat{\beta}_{1D_1}) = V(\hat{\beta}_{ls})$ and $V(\hat{\beta}_{2D_2}) = V(\hat{\beta}_{ls})$. The

$V(\hat{\beta}_{1D_1})$, $V(\hat{\beta}_{1D_2})$, $V(\hat{\beta}_{1D_3})$, $V(\hat{\beta}_{2D_2})$, $V(\hat{\beta}_{2D_3})$ are less than $V(\hat{\beta}_{eh})$.

4. Some special cases of the classes of estimators

Results 1:

For $k = k' - 1$, $\hat{\beta}_{1D_k} = \hat{\beta}_{2D_{k'}}$. (15)

Proof: Let $\hat{\beta}_{1D_k} = \frac{\sum_{i=1}^m q_{m-i}^k (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}^k q_{m-i}}$

$$= \frac{\sum_{i=1}^m q_{m-i}^k (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}^{k+1}}$$

$$= \frac{\sum_{i=1}^m q_{m-i}^{k+1} b_i}{\sum_{i=1}^m q_{m-i}^{k+1}}$$

$$= \frac{\sum_{i=1}^m q_{m-i}^{k'} b_i}{\sum_{i=1}^m q_{m-i}^{k'}} = \hat{\beta}_{2D_{k'}}$$

Similarly, for $k = k' + 1$, $\hat{\beta}_{2D_k} = \hat{\beta}_{1D_{k'}}$.

Result 2: For $k = 0$, $\hat{\beta}_{1D_k} = \hat{\beta}_{ud}$ and for $k' = 1$, $\hat{\beta}_{2D_{k'}} = \hat{\beta}_{ud}$. When $x_{(i)}$'s are equidistant, $\hat{\beta}_{1D_0} = \hat{\beta}_{2D_1} = \hat{\beta}_{eh}$.

Proof: when $k = 0$, $\hat{\beta}_{1D_k} = \frac{\sum_{i=1}^m q_{m-i}^0 (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}^0 q_{m-i}}$

$$= \frac{\sum_{i=1}^m (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}} = \hat{\beta}_{ud} \quad (16)$$

Under equidistant $x_{(i)}$'s, $\sum_{i=1}^m q_{m-i} = m^2 d$.

Therefore, $\hat{\beta}_{1D_0} = \frac{\sum_{i=1}^m (y_{m+i}^* - y_{m-i+1}^*)}{m^2 d}$

$$= \frac{\sum_{i=1}^m (y_{m+i}^* - y_i^*)}{m^2 d} = \hat{\beta}_{eh} \quad (17)$$

On similar lines, $\hat{\beta}_{2D_{k'}} = \hat{\beta}_{ud}$ and under equidistant $x_{(i)}$'s, $\hat{\beta}_{2D_1} = \hat{\beta}_{eh}$.

Result 3: For $k = 1$, $k' = 2$ and equidistant $x_{(i)}$'s, $V(\hat{\beta}_{1D_k}) = V(\hat{\beta}_{2D_{k'}}) = V(\hat{\beta}_{ls})$.

Proof: Let $k = 1$, $\hat{\beta}_{1D_1} = \frac{\sum_{i=1}^m q_{m-i} (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}^2}$

and its variance is given by

$$V(\hat{\beta}_{1D_1}) = \frac{2\sigma^2 \sum_{i=1}^m q_{m-i}^2}{(\sum_{i=1}^m q_{m-i}^2)^2}$$

$$= \frac{2\sigma^2}{\sum_{i=1}^m q_{m-i}^2} \quad (18)$$

When $x_{(i)}$'s are equidistant, $\hat{\beta}_{1D_1}$ reduces to

$$\hat{\beta}_{1D_1} = \frac{\sum_{i=1}^m (2i-1)(y_{m+i}^* - y_{m-i+1}^*)}{d \sum_{i=1}^m (2i-1)^2}$$

and its variance is given by

$$V(\hat{\beta}_{1D_1}) = \frac{2\sigma^2 \sum_{i=1}^m (2i-1)^2}{d^2 [\sum_{i=1}^m (2i-1)^2]^2}$$

$$= \frac{2\sigma^2}{d^2 [\sum_{i=1}^m (2i-1)^2]}$$

$$= \frac{6\sigma^2}{d^2 m(2m+1)(2m-1)}$$

$$= \frac{12\sigma^2}{d^2 n(n^2-1)} = V(\hat{\beta}_{ls}) \quad (19)$$

Proceeding on similar lines, we get $V(\hat{\beta}_{2D_2}) = V(\hat{\beta}_{ls})$.

Result 4: For $k = -1$ and $k' = 0$, $\hat{\beta}_{1D_k} = \hat{\beta}_{2D_{k'}} = \bar{b}$.

Proof: Let $k = -1$, $\hat{\beta}_{1D_{-1}} = \frac{\sum_{i=1}^m q_{m-i}^{-1} (y_{m+i}^* - y_{m-i+1}^*)}{\sum_{i=1}^m q_{m-i}^{-1} q_{m-i}}$

$$= \frac{\sum_{i=1}^m \frac{1}{q_{m-i}} (y_{m+i}^* - y_{m-i+1}^*)}{m}$$

$$= \frac{\sum_{i=1}^m b_i}{m} = \bar{b} \quad (20)$$

On similar lines, we notice that, $\hat{\beta}_{2D_0} = \bar{b}$.

5. Performance of the proposed classes of estimators

In this section, we study the performance of the classes of estimators using relative efficiency (RE) and simulated data. The RE serves as a quantitative indicator to assess the comparative precision of estimators. In the context of two unbiased estimators, T_1 and T_2 , $RE(T_1, T_2)$ is given by

$$RE(T_1, T_2) = \frac{Var(T_2)}{Var(T_1)} \quad (21)$$

The $RE(T_1, T_2) > 1$ indicates that estimator T_1 offers better precision relative to estimator T_2 . From table 1, it is evident that, $RE(\hat{\beta}_{1D_{k-1}}, \hat{\beta}_{2D_k}) = 1$. Hence, $RE(\hat{\beta}_{1D_1}, \hat{\beta}_{2D_2}) = 1$. For various values of $n \leq 200$, $\hat{\beta}_{1D_1}$ performs better than $\hat{\beta}_{1D_0}$ by 25% to 34%. It is also better than $\hat{\beta}_{1D_2}$ by 6% to 8% for $n < 100$. Moreover, $\hat{\beta}_{1D_1}$ shows 17% to 18% higher performance over $\hat{\beta}_{1D_3}$ for $n < 100$. For $n \leq 200$, $\hat{\beta}_{2D_2}$ shows 17% to 19% higher efficiency than $\hat{\beta}_{2D_0}$. It exhibits about 2% to 3% higher efficiency than $\hat{\beta}_{2D_1}$. Moreover, $\hat{\beta}_{2D_2}$ is approximately 7% to 9% more efficient than $\hat{\beta}_{2D_3}$. For $n > 200$, $\hat{\beta}_{2D_2}$ and $\hat{\beta}_{2D_0}$, $\hat{\beta}_{2D_1}$, $\hat{\beta}_{2D_3}$ are equally efficient.

To evaluate the robustness and effectiveness of the proposed classes of estimators, we generate the simulated data. The random samples of various size n is generated for independent variable x , incorporating various percentages of outliers to emulate real scenarios. The true values of α and β are fixed at 2 and 3, respectively, while errors are introduced from a normal distribution with mean 0 and variance 1. Employing the simple linear regression model with these predetermined values, corresponding y observations are computed for different sample sizes and percentages of outliers in x . Subsequently, the proposed estimator is applied to the generated data to estimate the slope parameter β . In this study, we analyze only the performance of $\hat{\beta}_{2D_k}$ which is equal to $\hat{\beta}_{1D_{k-1}}$. By fixing the true value of β at 3, we analyze how closely the estimated β values match this true value. The results are tabulated for various values of n , k and percentage of outliers introduced for x observations in Table 3.

Table 3: The values of $\hat{\beta}_{2D_k}$ and $\hat{\beta}_{ls}$ for various values of n , k and % outliers.

% outliers	$\hat{\beta}_{2D_{-2}}$	$\hat{\beta}_{2D_{-1.5}}$	$\hat{\beta}_{2D_{-1}}$	$\hat{\beta}_{2D_{-0.5}}$	$\hat{\beta}_{2D_0}$	$\hat{\beta}_{2D_{0.5}}$	$\hat{\beta}_{2D_1}$	$\hat{\beta}_{2D_{1.5}}$	$\hat{\beta}_{2D_2}$	$\hat{\beta}_{ls}$
$n = 10$										
0	3.9109	3.7933	3.6138	3.3977	3.2049	3.0768	3.0106	2.9842	2.9789	2.9789
2	2.6584	2.6853	2.7377	2.8158	2.9023	2.9759	3.0284	3.0623	3.0833	3.0833
5	3.2759	3.2535	3.2154	3.1631	3.1077	3.0609	3.0263	3.0018	2.9844	2.9844
8	3.4432	3.2776	3.0185	2.7060	2.4401	2.2869	2.2357	2.2464	2.2873	2.2963
10	2.8451	2.8173	2.7672	2.6963	2.6207	2.5545	2.4982	2.4472	2.3989	2.3410
20	3.0888	3.0130	2.8993	2.7472	2.5712	2.3969	2.2468	2.1315	2.0503	1.9351
30	3.3103	3.1707	2.9120	2.5318	2.1458	1.8925	1.7844	1.7605	1.7706	1.7342
$n = 30$										
0	2.7635	2.8058	2.8847	2.9691	3.0141	3.0234	3.0198	3.0153	3.0131	3.0131
2	3.5915	3.4226	3.2047	3.0124	2.8933	2.8128	2.7349	2.6499	2.5590	2.4691
5	2.6095	2.7359	2.8886	2.9519	2.8836	2.7616	2.6387	2.5264	2.4236	2.3610
8	4.0037	3.9807	3.8149	3.4771	3.1125	2.8369	2.6431	2.4969	2.3782	2.2598
10	3.3326	3.1378	2.8893	2.6839	2.5768	2.5183	2.4690	2.4221	2.3785	2.3889
20	5.2781	4.9390	4.2359	3.3499	2.7164	2.3747	2.1842	2.0677	1.9940	1.9045
30	4.1365	3.6328	3.5742	3.1424	2.5779	2.1688	1.9242	1.7823	1.6984	1.6781
$n = 50$										
0	2.3020	2.3478	2.4874	2.6972	2.8607	2.9480	2.9911	3.0142	3.0284	3.0284
2	4.5683	4.3414	3.5342	2.9499	2.7865	2.8159	2.8757	2.9258	2.9629	2.9604
5	3.7836	3.6373	3.3753	3.0921	2.9201	2.8307	2.7700	2.7178	2.6694	2.6730
8	2.7771	2.8557	2.9251	2.8925	2.7683	2.6384	2.5301	2.4394	2.3596	2.3136
10	3.6829	3.4392	2.9971	2.5614	2.3744	2.3461	2.3455	2.3302	2.2978	2.2827
20	3.6909	3.5514	3.1709	2.6519	2.3111	2.1475	2.0535	1.9833	1.9255	1.8416
30	3.8097	3.6610	3.2949	2.7371	2.2229	1.9080	1.7604	1.7046	1.6917	1.6718
$n = 100$										
0	3.6959	3.4788	3.2548	3.1244	3.0816	3.0692	3.0662	3.0660	3.0659	3.0659
2	2.4233	2.4972	2.6510	2.8191	2.9097	2.9273	2.9113	2.8846	2.8558	2.8515
5	3.6369	3.5965	3.3735	3.0422	2.8378	2.7383	2.6682	2.6014	2.5323	2.4825
8	5.3227	4.4097	4.0164	2.9174	2.5642	2.5069	2.4845	2.4489	2.4001	2.3562
10	4.3192	3.9878	3.2961	2.6488	2.4407	2.4178	2.3989	2.3581	2.3038	2.2630
20	6.2222	5.2531	4.3979	3.3491	3.0091	2.2601	2.0528	1.9768	1.9308	1.8480
30	1.9724	2.4161	3.0588	3.0520	2.4679	2.0099	1.7953	1.7225	1.7096	1.6736

The results of the simulation study reflects the robustness of the proposed estimators compared to ls estimator across varying data conditions. We observe that, for error percentages greater than 2%, $\hat{\beta}_{ls}$ exhibits diminishing performance. In contrast, the proposed class of estimators yields more reliable estimates of β , showing its resilience to outliers. Further, when $k = 2$, the estimators have better performance in the absence of outliers and is equivalent to $\hat{\beta}_{ls}$. However, as the percentage of outliers in x increase, estimators with decreasing values of k (i.e., placing more emphasis on central observations) surpass the estimator with $k = 2$ in performance. This indicates that emphasizing central observations improves robustness and accuracy in estimating β in the presence of outliers.

While exploring different values of k ranging from -2 to 2, we notice that, a trend where decreasing k values lead to increased robustness of the estimator. However, selecting excessively low k values might overemphasize central observations and introduce inaccuracies. If a certain value of k (say k_1) achieves the desired level of robustness or yields a well-fitted regression line, then we need not explore for the values of $k < k_1$.

6. Illustration

In this section, the proposed estimators are illustrated through an example. We compute $\hat{\beta}$ and variance of various members of the proposed classes of estimators. To fit the SLR model given in (1), $\hat{\alpha}$ is given by

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \quad (22)$$

Using various estimators, the regression lines are fitted.

Example: The Pilot-Plant data originally given by Daniel and Wood (1971)^[8] and later studied by Rousseeuw and Leroy (2005)^[13] is considered. The dataset explains acid content determined through titration (y_i) and organic acid content determined through extraction and weighing (x_i). Initially, we employ the dataset as in Daniel and Wood (1971)^[8] which consists of 20 observations without any outliers. Subsequently, Rousseeuw and Leroy (2005)^[13] discussed the scenario where the x -value of the 6th observation might have been wrongly recorded as 370 instead of 37, thus producing an outlier. We take this point of view and utilize the modified dataset containing the outlier in our analysis. By considering both original dataset and modified dataset with the outlier, we illustrate the performance of some members of the proposed classes of estimators.

Table 4: Pilot-Plant data due to Daniel and Wood (1971)^[8]

Observation (i)	Extraction (x)	Titration (y)	Observation (i)	Extraction (x)	Titration (y)
1	123	76	11	138	82
2	109	70	12	105	68
3	62	55	13	159	88
4	104	71	14	75	58
5	57	55	15	88	64
6	37	48	16	164	88
7	44	50	17	169	89
8	100	66	18	167	88
9	16	41	19	149	84
10	28	43	20	167	88

Here, $\hat{\beta}_{ls} = 0.321600$ and $V(\hat{\beta}_{ls}) = 0.0000204\sigma^2$. The computed values of $\hat{\beta}_{1D_k}$, $\hat{\beta}_{2D_k}$ and their variances are given in Table 5 and regression lines are plotted in Figure 2. When an outlier is induced, the computed values of $\hat{\beta}_{1D_k}$, $\hat{\beta}_{2D_k}$ and

their variance is furnished in Table 6 and regression lines are plotted in Figure 3. Also, $\hat{\beta}_{ls} = 0.080710$ with variance $V(\hat{\beta}_{ls}) = 0.000009\sigma^2$.

Table 5: Value of $\hat{\beta}_{1D_k}$, $\hat{\beta}_{2D_k}$ and their variances for original dataset

k	Value of $\hat{\beta}_{1D_k}$	$V(\hat{\beta}_{1D_k})$	Value of $\hat{\beta}_{2D_k}$	$V(\hat{\beta}_{2D_k})$
0	0.3218117	$0.0000284\sigma^2$	0.008694	$0.0202773\sigma^2$
0.5	0.3223964	$0.0000212\sigma^2$	0.291622	$0.0003437\sigma^2$
1	0.3208606	$0.0000206\sigma^2$	0.321812	$0.0000284\sigma^2$
1.5	0.3195962	$0.0000208\sigma^2$	0.322396	$0.0000212\sigma^2$
2	0.3186570	$0.0000214\sigma^2$	0.320861	$0.0000206\sigma^2$

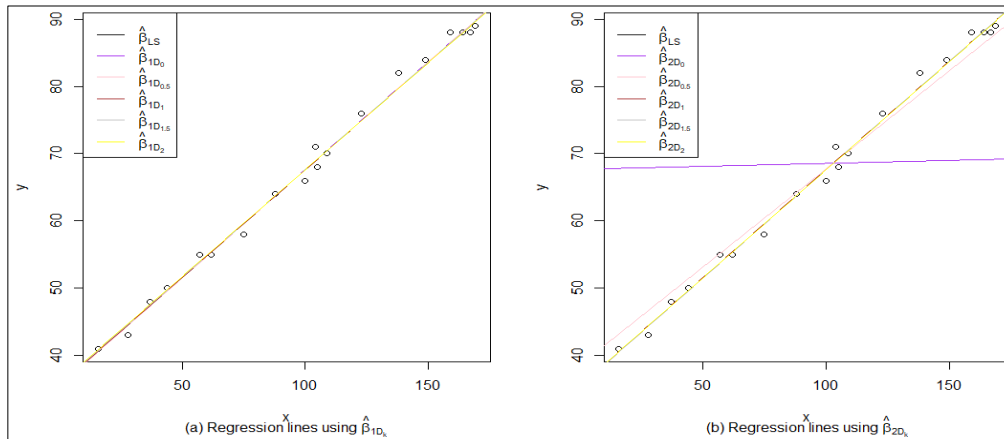


Fig 2: The fitted regression lines using $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$

Table 6: Value of $\hat{\beta}_{1D_k}$, $\hat{\beta}_{2D_k}$ and their variances for the dataset with an outlier

k	Value of $\hat{\beta}_{1D_k}$	$V(\hat{\beta}_{1D_k})$	Value of $\hat{\beta}_{2D_k}$	$V(\hat{\beta}_{2D_k})$
-2	0.432124	$0.0503852\sigma^2$	0.487463	$0.1098416\sigma^2$
-1.5	0.371866	$0.0131396\sigma^2$	0.470254	$0.0891342\sigma^2$
-1	0.314771	$0.0013335\sigma^2$	0.432124	$0.0503851\sigma^2$
-0.5	0.268465	$0.0000934\sigma^2$	0.371866	$0.0131396\sigma^2$
0	0.222008	$0.0000186\sigma^2$	0.314771	$0.0013335\sigma^2$
0.5	0.173504	$0.0000114\sigma^2$	0.268465	$0.0000934\sigma^2$
1	0.128359	$0.0000102\sigma^2$	0.222008	$0.0000186\sigma^2$
1.5	0.091787	$0.0000109\sigma^2$	0.173504	$0.0000114\sigma^2$
2	0.065398	$0.0000122\sigma^2$	0.128359	$0.0000102\sigma^2$

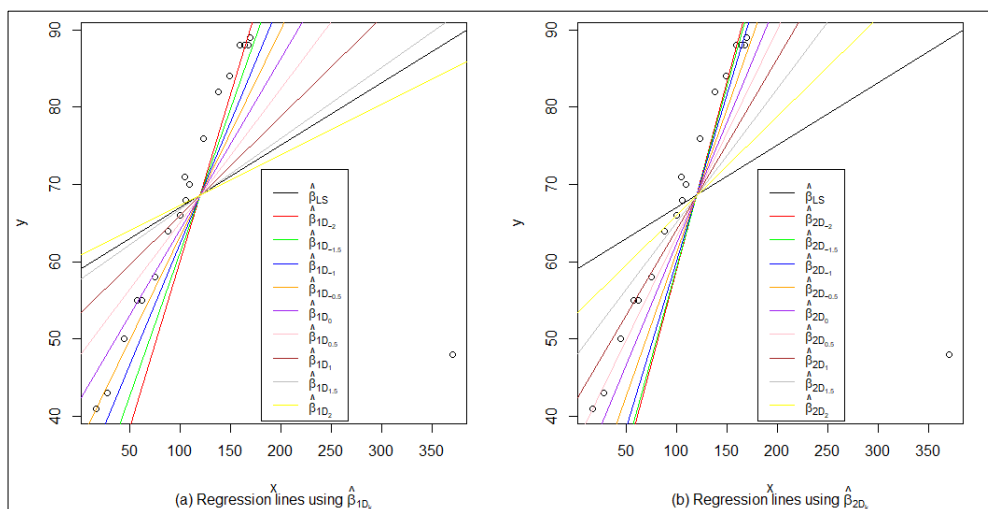


Fig 3: Fitted regression lines using $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ in presence of an outlier

From Table 5 and Figure 2, it is observed that, $\hat{\beta}_{1D_k}$, $0 \leq k \leq 2$, $\hat{\beta}_{2D_k}$, $1 \leq k \leq 2$, $\hat{\beta}_{ls}$ have near values to the slope with

overlapping regression lines and $\hat{\beta}_{2D_0}$ results in poor fit. As $k > 0$ effectively captures data patterns, further exploration

has limited benefits. After introducing an outlier in the dataset and plotting regression lines based on $\hat{\beta}_{1D_k}, \hat{\beta}_{2D_k}, -2 \leq k \leq 2$, from Table 6 and Figure 3, it is seen that, there is a notable good fit of regression lines to the data. Specifically, for $k \leq 0$, the estimator exhibits increased robustness against outlier leading to improved accuracy in estimating the slope parameter β .

7. Conclusions

- Two classes of estimators $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ which utilize the weights based on quasi ranges of predictor variable are proposed for slope parameter, β in SLR model.
- One of the classes is based on ratio of sum of the weighted deviations of corresponding y observations to sum of the weighted quasi ranges of predictor variable (x) and other is weighted average of slopes obtained using quasi ranges.
- The proposed classes utilize the weights depending on quasi ranges of x emphasizing the magnitude of the observations.
- The classes of estimators admit unbiased and consistent estimators for β .
- The proposed weighting scheme effectively distinguishes between extreme and central observations. Hence, allows for a customized weighting system that prioritizes observations based on their positions within the dataset.
- The exponent k in weighting scheme introduces flexibility in intensifying the customization.
- The optimal value of k (say k^*) for the proposed classes of estimators obtained using Brent's method by minimizing their variances are $k^* = 1$ and $k^* = 2$ for $\hat{\beta}_{1D_k}$ and $\hat{\beta}_{2D_k}$ respectively. Also, $\hat{\beta}_{1D_{k-1}} = \hat{\beta}_{2D_k}$.
- $\hat{\beta}_{1D_0}, \hat{\beta}_{2D_1}$ reduces to the estimator due to Bhat and Bijjargi (2023) ^[1]. Under equidistant x_i 's they reduce to the estimator due to Bose (1938) ^[5].
- The variances of $\hat{\beta}_{1D_1}$ and $\hat{\beta}_{2D_2}$ are equal to the variance of least square estimator when x_i 's have equal distance.
- For $k > 0$, both the classes of estimators have heavier weights to extreme observations and for $k < 0$, they have heavier weights to middle observations.
- The members of classes of estimators show equal performance for large samples, whereas, $\hat{\beta}_{1D_1}$ and $\hat{\beta}_{2D_2}$ are better than other members of their respective classes for smaller samples.

8. References

1. Bhat SV, Bijjargi SM. Nonparametric approach to estimation in linear regression. *International Journal of Current Research*. 2023;15(04):24419-24424.
2. Bhat SV, Bijjargi SM. Simple linear slope estimators based on sample quasi ranges. To appear in *Statistics and Applications*; c2024a.
3. Bhat SV, Bijjargi SM. Classes of slope estimators based on quasi ranges in simple linear regression. *International Journal of Statistics and Systems*. 2024b;19(1):1-11.
4. Birkes D, Dodge Y. *Alternative Methods of Regression*. John Wiley and Sons, New York; c1993.
5. Bose SS. Relative efficiencies of regression coefficients estimated by method of finite differences. *Sankhya*. 1938;3:339-346.
6. Brent RP. An algorithm with guaranteed convergence for finding a zero of a function. *The computer journal*. 1971;14(4):422-425.

7. Cliff KR, Billy KM. Estimation of the parameters of a linear regression system using the simple averaging method. *Global Journal of Pure and Applied Mathematics*. 2017;13(11):7749-7758.
8. Daniel C, Wood E, Gorman, J. *Fitting equations to data*. John Wiley and Sons. Inc. USA; c1971.
9. Gauss CF. *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium* [Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections]. Little, Brown and Company, Boston; c1809.
10. Legendre AM. *Nouvelles méthodes pour la détermination des orbites des comètes* [New Methods for the Determination of the Orbits of Comets] (in French). Paris: F. Didot; c1805.
11. Olkin I, Yitzhaki S. Gini regression analysis. *International Statistical Review/Revue Internationale de Statistique*; c1992 .p. 185-196.
12. Prabowo A, Sugandha A, Tripena A, Mamat M, Sukono, Budiono R. A New Method to Estimate Parameters in the Simple Regression Linear Equation. *Mathematics and Statistics*. 2020;8(2):75-81.
13. Rousseeuw PJ, Leroy AM. *Robust regression and outlier detection*. John Wiley & sons; c2005.
14. Singthongchai J, Thongmual N, Nitisuk N. An Improved Simple Averaging Approach for Estimating Parameters in Simple Linear Regression Model. *Mathematics and Statistics*. 2021;9(6):939-946.
15. Theil H. A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae*. 1950;12(85):85-173.