International Journal of Statistics and Systems
ISSN 0973-2675 Volume 19, Number 1 (2024), pp. 1-11

https://dx.doi.org/10.37622/IJSS/19.1.2024.1-11

Classes of Slope Estimators Based On Quasi Ranges in Simple Linear Regression

***Sharada V. Bhat and Shrinath M. Bijjargi**

Department of Statistics, Karnatak University, Dharwad - 580003, India.

Abstract

Two classes of estimators for slope parameter in simple linear regression model are proposed. One of the classes is based on weighted average of the quasi ranges of the predictor variable while the other class is based on weighted average of slopes obtained from quasi ranges of predictor variable. The mean and variance of these classes of estimators are derived. The optimum weights are obtained. The performance of the proposed classes of estimators is analyzed. The feasibility of some members of the classes are illustrated through an example.

Keywords: quasi range, simple linear regression, slope parameter, weighted average, relative efficiency, predictor variable.

1. INTRODUCTION

In the realm of linear regression analysis, the estimation of parameters holds paramount importance, influencing the predictive accuracy and interpretability of the model. Parameters in linear regression model encapsulate vital information about the relationship between predictor variables and the response variable, providing insights into the underlying trends and patterns in the data. Accurate estimation of these parameters is essential for making informed decisions, formulating effective strategies and drawing reliable inferences from the regression model.

The simple linear regression (SLR) serves as a foundational framework within linear regression analysis, particularly when exploring relationship between a single predictor variable and a response variable. In this model, the relationship between the predictor and response variable is represented by a straight line, facilitating intuitive interpretation and analysis. The slope parameter plays an important role in characterizing the relationship between the predictor and response variables. It serves as a fundamental indicator of the rate of change in the response variable with respect to (wrt) changes in the predictor variable.

Various estimators have been developed in the literature for estimating parameters in SLR model, each with its own merits and limitations. Among these, the least squares estimator due to Legendre (1805) and Gauss (1809) stands as earliest and popular method. It has efficiency but is sensitive to outliers and influential observations. It prioritizes minimizing squared error deviations without explicit consideration for the robustness in estimation process. Hence, alternative methods have been developed to address its limitations and accommodate diverse data characteristics.

Bose (1938) introduced three different methods to estimate the slope parameter, considering various kinds of distances viz. successive differences, differences at half range and range among predictor variables, assuming that they are evenly spaced. Theil (1950), Gore and Rao (1982) and Rao (1982) proposed procedures based on median of slopes obtained from half ranges. Bhat and Bijjargi (2023) extended Bose's methods including estimators based on quasi ranges under unequal distances among predictor variables. The estimator based on quasi ranges is found to be equivalent to estimator based on half ranges and outperformed all other estimators based on various kinds of distances. Further, Bhat and Bijjargi (2024) proposed few more estimators applying various arbitrary weights to the quasi ranges to enhance accuracy.

Suppose $x_1, x_2, \ldots x_n$ represent n observations of predictor variable and $m = n/2$. Denoting $x_{(i)}$ as the i^{th} order statistic, the quasi range q_i, is given by $q_i = x_{(n-i)} - x_{(i+1)}$, $i = 1, 2, \ldots, m-1$. Here, $q_0 = x_{(n)} - x_{(1)}$ is the range of n observations. The choice of quasi ranges is justified by their inherent advantages. Quasi ranges offer a structured approach that distinguishes between extreme and middle observations more effectively. This distinction allows for a customized weighting scheme, enabling the prioritization of the relative positions of observations within the dataset, thus enhancing the robustness and accuracy.

In this paper, we introduce flexible classes of estimators based on weighted averages, offering enhanced adaptability in estimating slope parameter. We propose two classes of estimators based on quasi ranges of predictor variables. One class of estimators is based on weighted average of quasi ranges of x_i variables and the other is based on weighted average of slopes obtained from quasi ranges of x_i variables. The weighted average is a measure that assigns different weights to each data point in a dataset based on their relative importance. This method allows for the incorporation of varying degrees of significance for different data points, providing a more nuanced representation of central tendency.

The mathematical formulation of the proposed classes of estimators is given in section 2, followed by the derivation of their mean and variance in section 3. In section 4, the performance of the proposed classes of estimators is investigated. Section 5 illustrates the application of the proposed estimators through example and section 6 provides the conclusions drawn from the study.

2. Mathematical Formulation of Proposed Classes of Estimators

The SLR model is given by

$$y_i = \alpha + \beta x_i + e_i\,, \quad 1 \le i \le n\,, \qquad \textbf{(1)}$$

where, y_i is response variable, α is the intercept parameter, β is the slope parameter and 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.

We propose two classes of estimators based on quasi ranges of predictor variables. Let y_i^* be the y value corresponding to $x_{(i)}$. Representing the two classes by $1C_k$ and $2C_k$,

$$\hat{\beta}_{1C_k} = \frac{\sum_{i=1}^{m} i^k (y^*_{m+i} - y^*_{m-i+1})}{\sum_{i=1}^{m} i^k q_{m-i}} \qquad \textbf{(2)}$$

is the ratio of weighted averages of deviations of corresponding y values wrt quasi ranges of x values and

$$\hat{\beta}_{2C_k} = \frac{\sum_{i=1}^{m} i^k b_i}{\sum_{i=1}^{m} i^k}\,,\ b_i = \frac{y^*_{m+i} - y^*_{m-i+1}}{q_{m-i}}\,,\ i = 1, \cdots, m \qquad \textbf{(3)}$$

is the weighted average of slopes b_i, which is ratio of deviation of corresponding y values wrt quasi ranges of x values.

The weights, represented by i^k, $k \in \mathbb{R}$ is known finite constant, are chosen strategically to assign lower (higher) weights to quasi ranges derived from middle observations and higher (lower) weights to those obtained from extreme values for $k > 0$ ($k < 0$). The motivation behind choosing the weighting scheme i^k lies in its ability to adaptively adjust the influence of observations based on their distances. For $k > 0$, heavier weights are assigned to observations farther away from the median, leading to increased influence of extreme values in x observations. This scenario is particularly useful when extreme observations are preferred. Conversely, for $k < 0$, heavier weights are assigned to observations closer to the median, thereby reducing the impact of outliers and enhancing robustness against extreme values in the predictor variable. By incorporating the exponent k into the weighting scheme, the proposed classes offer a comprehensive and effective approach to parameter estimation, catering to diverse datasets and analytical requirements.

When $k = 0$, $\hat{\beta}_{1C_0}$ is an estimator proposed by Bhat and Bijjargi (2023) and also under equidistant $x_i s$, it is an estimator due to Bose (1938) based on half ranges. Furthermore, for $k = 1$ and -1, $\hat{\beta}_{1C_1}$ and $\hat{\beta}_{1C_{-1}}$ simplifies to estimators developed by Bhat and Bijjargi (2024). Also, $\hat{\beta}_{2C_0}$ is simple average of b_i.

3. Mean and Variance of Proposed Classes of Estimators

In this section, we obtain mean and variance of the proposed classes of estimators.

The mean of $\hat{\beta}_{1C_k}$ is given by

$$E\left(\hat{\beta}_{1C_k}\right) = E\left(\frac{\sum_{i=1}^{m} i^k (y^*_{m+i} - y^*_{m-i+1})}{\sum_{i=1}^{m} i^k q_{m-i}}\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{m} i^k q_{m-i}\right)} E\left(\sum_{i=1}^{m} i^k (y^*_{m+i} - y^*_{m-i+1})\right)$$

$$= \frac{\sum_{i=1}^{m} i^k \left(\alpha + \beta x_{(m+i)} - \alpha - \beta x_{(m-i+1)}\right)}{\left(\sum_{i=1}^{m} i^k q_{m-i}\right)}$$

$$= \beta \qquad \textbf{(4)}$$

Similarly, the mean of $\hat{\beta}_{2C_k}$ is given by

$$E\left(\hat{\beta}_{2C_k}\right) = E\left(\frac{\sum_{i=1}^{m} i^k b_i}{\sum_{i=1}^{m} i^k}\right)$$

$$= \left(\frac{\sum_{i=1}^{m} i^k E(b_i)}{\sum_{i=1}^{m} i^k}\right)$$

$$= \beta \qquad \because E(b_i) = \beta \qquad \textbf{(5)}$$

Both the classes of estimators admit unbiased estimators of β.

The variance of $\hat{\beta}_{1C_k}$ is given by

$$V\left(\hat{\beta}_{1C_k}\right) = V\left(\frac{\sum_{i=1}^{m} i^k (y^*_{m+i} - y^*_{m-i+1})}{\sum_{i=1}^{m} i^k q_{m-i}}\right)$$

$$= \frac{\sum_{i=1}^{m} i^{2k} V(y^*_{m+i} - y^*_{m-i+1})}{\left(\sum_{i=1}^{m} i^k q_{m-i}\right)^2}$$

$$= \frac{2\sigma^2 \sum_{i=1}^{m} i^{2k}}{\left(\sum_{i=1}^{m} i^k q_{m-i}\right)^2} \qquad \textbf{(6)}$$

and variance of $\hat{\beta}_{2C_k}$ is given by

$$V\left(\hat{\beta}_{2C_k}\right) = V\left(\frac{\sum_{i=1}^{m} i^k b_i}{\sum_{i=1}^{m} i^k}\right)$$

$$= \frac{\sum_{i=1}^{m} i^{2k} V(b_i)}{\left(\sum_{i=1}^{m} i^k\right)^2}$$

$$= \frac{\sum_{i=1}^{m} i^{2k} \frac{2\sigma^2}{(q_{m-i})^2}}{\left(\sum_{i=1}^{m} i^k\right)^2}$$

$$= \frac{2\sigma^2 \sum_{i=1}^{m} \left(\frac{i^{2k}}{(q_{m-i})^2}\right)}{\left(\sum_{i=1}^{m} i^k\right)^2}. \qquad \textbf{(7)}$$

The $\hat{\beta}_{1C_k}$ and $\hat{\beta}_{2C_k}$ are consistent estimators of β.

When $x_{(i)}s$ are equidistant with distances being d, then

$$q_{m-i} = x_{(m+i)} - x_{(m-i+1)}$$
$$= \left(x_{(m)} + id\right) - \left(x_{(m)} - (i-1)d\right)$$
$$= id + (i-1)d$$
$$= (2i-1)d \quad \textbf{(8)}$$

and $= \sum_{i=1}^{m} q_{m-i} = \sum_{i=1}^{m}\left(x_{(m+i)} - x_{(m-i+1)}\right)$ **(9)**

Under equidistant $x_{(i)}s$, the variance expression given in (6) and (7) can be written as

$$V\left(\hat{\beta}_{1C_k}\right) = \frac{2\sigma^2 \sum_{i=1}^{m} i^{2k}}{\left(\sum_{i=1}^{m} i^k (2i-1)d\right)^2} = \frac{2\sigma^2 \sum_{i=1}^{m} i^{2k}}{d^2\left(\sum_{i=1}^{m} i^k (2i-1)\right)^2} \quad \textbf{(10)}$$

$$V\left(\hat{\beta}_{2C_k}\right) = \frac{2\sigma^2 \sum_{i=1}^{m} \frac{i^{2k}}{((2i-1)d)^2}}{\left(\sum_{i=1}^{m} i^k\right)^2} = \frac{2\sigma^2}{d^2\left(\sum_{i=1}^{m} i^k\right)^2} \sum_{i=1}^{m} \frac{i^{2k}}{(2i-1)^2} \quad \textbf{(11)}$$

The optimization of the variance of two classes of estimators wrt k, provides an optimal estimator of β. To identify the value of k that minimizes complex variance expressions in (10) and (11), we employ a numerical optimization algorithm, viz. Brent's method. It iteratively refines search intervals to converge to the minimum of a function. Initialization involves selecting a possible interval containing the minimum, followed by iterative updates and interpolation to approximate the location of the minimum within the interval. Convergence is achieved by refining the interval until specified criterion is met.

After employing Brent's method, we determine the optimal k values for both the classes of estimators for various values of n, which are summarized in the Table 1.

Table 1: Optimum values of k for various values of n

n	Optimum k	
	$\widehat{\boldsymbol{\beta}}_{1C_k}$	$\widehat{\boldsymbol{\beta}}_{2C_k}$
6	1.377037	2.773202
8	1.281407	2.588176
10	1.225069	2.476232
20	1.112888	2.243209
30	1.075333	2.161740
50	1.045197	2.095716
70	1.032267	2.067579
100	1.022572	2.046747
200	1.011272	2.022963

These results indicate that, as n increases the optimal values of k tend to 1 for $\hat{\beta}_{1C_k}$ and towards 2 for $\hat{\beta}_{2C_k}$. This suggests a tendency towards assigning relatively lower weights to middle observations when compared to extreme values. However, negative values of k would assign higher weights to middle observations. Thus, to explore the weighting adjustment, we consider $-2 \leq k \leq 3$. The variances of $\hat{\beta}_{1C_k}$ and $\hat{\beta}_{2C_k}$ are computed for various values of n, k and are furnished in Table 2 and Table 3.

Table 2: $V(\hat{\beta}_{1C_k})$ for various values of n and k

n	$V(\hat{\beta}_{1C_{-2}})$	$V(\hat{\beta}_{1C_{-1}})$	$V(\hat{\beta}_{1C_0})$	$V(\hat{\beta}_{1C_1})$	$V(\hat{\beta}_{1C_{1.5}})$	$V(\hat{\beta}_{1C_2})$	$V(\hat{\beta}_{1C_{2.5}})$	$V(\hat{\beta}_{1C_3})$
6	0.404413	0.156800	0.074074	0.057851	0.057241	0.058264	0.060005	0.062031
8	0.286736	0.081333	0.031250	0.024000	0.023906	0.024498	0.025404	0.026456
10	0.224397	0.049158	0.016000	0.012188	0.012196	0.012549	0.013065	0.013662
20	0.116597	0.010636	0.002000	0.001506	0.001522	0.001579	0.001656	0.001743
30	0.084671	0.004440	0.000593	0.000445	0.000451	0.000470	0.000493	0.000521
50	0.059606	0.001506	0.000128	0.000096	0.000098	0.000102	0.000107	0.000113
70	0.048557	0.000746	0.000047	0.000035	0.000036	0.000037	0.000039	0.000041
100	0.039817	0.000356	0.000016	0.000012	0.000012	0.000013	0.000013	0.000014
200	0.028339	0.000086	0.000002	0.000002	0.000002	0.000002	0.000002	0.000002

Table 3: $V(\hat{\beta}_{2C_k})$ for various values of n and k

n	$V(\hat{\beta}_{2C_{-2}})$	$V(\hat{\beta}_{2C_{-1}})$	$V(\hat{\beta}_{2C_0})$	$V(\hat{\beta}_{2C_1})$	$V(\hat{\beta}_{2C_{1.5}})$	$V(\hat{\beta}_{2C_2})$	$V(\hat{\beta}_{2C_{2.5}})$	$V(\hat{\beta}_{2C_3})$
6	1.087580	0.614215	0.255802	0.100247	0.072907	0.061406	0.057696	0.057517
8	0.994260	0.476236	0.146440	0.042620	0.029499	0.024983	0.023907	0.024173
10	0.940676	0.396651	0.094709	0.021685	0.014628	0.012534	0.012158	0.012396
20	0.839002	0.241170	0.024174	0.002558	0.001690	0.001518	0.001515	0.001567
30	0.806752	0.187916	0.010818	0.000725	0.000487	0.000447	0.000450	0.000467
50	0.781546	0.142094	0.003916	0.000149	0.000103	0.000096	0.000097	0.000101
70	0.770906	0.120327	0.002003	0.000053	0.000037	0.000035	0.000036	0.000037
100	0.762989	0.102215	0.000983	0.000018	0.000013	0.000012	0.000012	0.000013
200	0.753823	0.076893	0.000246	0.000002	0.000002	0.000002	0.000002	0.000002

From Table 2 and 3, it is observed that, the variance of members of proposed classes decreases as n increases. For $n > 8$, it is seen that, $\hat{\beta}_{1C_1}$ and $\hat{\beta}_{2C_2}$ have minimum variance respectively from Table 2 and Table 3 in support of the findings from Table 1.

4. Performance of Proposed Classes of Estimators

In this section, we evaluate the performance of the proposed classes of estimators using relative efficiency (RE) which is a statistical measure used to compare the performance of two estimators in terms of their precision. It quantifies how well one estimator performs relative to another by comparing their variances and is given by

$$RE(A,B) = \frac{Var(B)}{Var(A)}. \qquad \textbf{(12)}$$

where, A and B are any two estimators. The RE of $\hat{\beta}_{2C_k}$ wrt $\hat{\beta}_{1C_k}$ is given by

$$RE\left(\hat{\beta}_{2C_k}, \hat{\beta}_{1C_k}\right) = \frac{\left(\sum_{i=1}^{m} i^k\right)^2 \left(\sum_{i=1}^{m} i^{2k}\right)}{\sum_{i=1}^{m}\left(\frac{i^{2k}}{(q_{m-i})^2}\right)\left(\sum_{i=1}^{m} i^k q_{m-i}\right)^2}. \quad \textbf{(13)}$$

When $x_{(i)}s$ are equidistant, $RE\left(\hat{\beta}_{2C_k}, \hat{\beta}_{1C_k}\right) = \frac{\left(\sum_{i=1}^{m} i^k\right)^2 \left(\sum_{i=1}^{m} i^{2k}\right)}{\sum_{i=1}^{m}\left(\frac{i^{2k}}{(2i-1)^2}\right)\left(\sum_{i=1}^{m} i^k (2i-1)\right)^2}$ **(14)**

Using (14), the computed values of $RE\left(\hat{\beta}_{2C_k}, \hat{\beta}_{1C_k}\right)$ for various n and k are given in Table 4.

Table 4: $RE\left(\hat{\beta}_{2C_k}, \hat{\beta}_{1C_k}\right)$for various values of n and k

k / n	-2	-1	0	1	1.5	2	2.5	3
6	0.371846	0.255285	0.289575	0.577088	0.785122	0.948834	1.040015	1.078483
8	0.288391	0.170783	0.213398	0.563123	0.810393	0.980605	1.062646	1.094482
10	0.238548	0.123933	0.168938	0.562052	0.833751	1.001185	1.074596	1.102138
20	0.138971	0.044102	0.082732	0.588839	0.900539	1.040203	1.092710	1.112530
30	0.104953	0.023627	0.054778	0.613913	0.927803	1.050929	1.096776	1.114484
50	0.076267	0.010596	0.032688	0.646174	0.950109	1.058192	1.099322	1.115495
70	0.062986	0.006197	0.023294	0.665490	0.959408	1.060909	1.100229	1.115776
100	0.052185	0.003487	0.016277	0.683342	0.966113	1.062787	1.100839	1.115926
200	0.037594	0.001121	0.008122	0.709749	0.973466	1.064812	1.101478	1.116035

From Table 4, it is noticed that, $RE\left(\hat{\beta}_{2C_k}, \hat{\beta}_{1C_k}\right) > 1$ for $k = 2$, $n \geq 10$ and $k > 2$. Also, $RE\left(\hat{\beta}_{2C_2}, \hat{\beta}_{1C_1}\right) \approx 1$. It is known that,

$$V\left(\hat{\beta}_{LS}\right) = \frac{12\,\sigma^2}{d^2 n(n^2-1)} \quad \textbf{(15)}$$

where, LS is a least square estimator. It is observed that, $RE\left(\hat{\beta}_{1C_1}, \hat{\beta}_{LS}\right) \approx 1$ and $RE\left(\hat{\beta}_{2C_2}, \hat{\beta}_{LS}\right) \approx 1$.

5. Simulation Study

In this section, a simulation study is carried out to investigate the performances of proposed classes in the presence and absence of outliers. The errors, e_i from $N(0,1)$ distribution and x_i samples of size n, i.e., $i = 1, \ldots, n$ are generated. The y values are computed using (1) with generated values, $\alpha = 1$ and $\beta = 2$. After computing y values, certain percentages of outliers are introduced into the x observations. This is achieved by replacing a specified percentage of the x values randomly to be significantly different from the rest. Once the outliers are introduced, β is estimated using both the proposed classes of estimators with $k \in (-2,2)$ and the LS estimator. These estimates represent the estimated relationship of y with x based on the observed data, accounting for the presence of outliers. The objective is to assess how closely these estimated β matches with true value of β. The estimated values of $\hat{\beta}_{1C_k}$, $\hat{\beta}_{2C_k}$ and $\hat{\beta}_{LS}$ for various values of n, k and % outliers are furnished in Table 5.

Table 5: Values of estimators $\hat{\beta}_{1C_k}$, $\hat{\beta}_{2C_k}$ and $\hat{\beta}_{LS}$ for various values of n, k

% Outliers	$\hat{\beta}_{1C_{-2}}$	$\hat{\beta}_{1C_{-1}}$	$\hat{\beta}_{1C_0}$	$\hat{\beta}_{1C_1}$	$\hat{\beta}_{1C_2}$	$\hat{\beta}_{2C_{-2}}$	$\hat{\beta}_{2C_{-1}}$	$\hat{\beta}_{2C_0}$	$\hat{\beta}_{2C_1}$	$\hat{\beta}_{2C_2}$	$\hat{\beta}_{LS}$
$n = 10$											
0	2.4988	2.2647	2.1515	2.1096	2.0931	2.9281	2.6364	2.3495	2.1845	2.1198	2.1032
2	1.3859	1.7139	1.9147	2.0211	2.0810	0.8804	1.2239	1.6020	1.8626	1.9994	2.0372
5	2.0868	2.0317	2.0554	2.0985	2.1325	2.3192	2.1613	2.0612	2.0564	2.0927	2.1050
ss8	1.9258	1.6599	1.4172	1.2274	1.0858	2.2264	2.0816	1.8501	1.6049	1.4003	0.8954
10	1.9783	1.8016	1.6053	1.4459	1.3285	2.0838	2.0227	1.8854	1.7124	1.5557	1.1784
20	1.7168	1.4327	1.2261	1.1043	1.0393	2.0929	1.9379	1.6739	1.4105	1.2250	0.8785
30	1.5731	1.0437	0.7563	0.6457	0.6136	2.3452	1.8574	1.2880	0.8946	0.7107	0.5425
$n = 30$											
0	3.2294	2.2944	2.0406	1.9911	1.9779	5.2471	3.7099	2.4510	2.0663	1.9947	1.9886
2	1.9609	1.8539	1.7014	1.5723	1.4662	1.8953	1.9572	1.9007	1.7704	1.6546	1.3740
5	2.1859	1.7647	1.5310	1.3734	1.2493	2.9381	2.3932	1.8897	1.6521	1.5097	1.0679
8	2.1316	1.7187	1.4982	1.3369	1.2031	3.0693	2.4085	1.8622	1.6162	1.4634	1.0574
10	2.1715	1.7498	1.4193	1.2284	1.1072	2.3514	2.2010	1.8565	1.5746	1.3876	0.8903
20	1.7935	1.1967	0.9015	0.7793	0.7347	2.5811	2.1044	1.4496	1.0397	0.8454	0.6189
30	1.4953	0.9660	0.7810	0.7446	0.7593	3.0453	2.0924	1.2095	0.8597	0.7751	0.6047
$n = 50$											
0	2.2213	2.1305	2.0104	1.9860	1.9823	0.8365	1.9299	2.1152	2.0144	1.9869	1.9853
2	2.0509	1.9220	1.8202	1.7158	1.6224	2.5252	2.1556	1.9569	1.8558	1.7608	1.6184
5	2.5772	1.8863	1.6707	1.5689	1.4868	3.6448	2.8290	2.0170	1.7712	1.6860	1.2677
8	2.0274	1.6973	1.4258	1.2338	1.0911	2.3867	2.1076	1.7786	1.5229	1.3386	0.9898
10	1.9327	1.6027	1.2938	1.0717	0.9157	2.0263	1.9727	1.7097	1.4222	1.1963	0.8263
20	1.9089	1.2377	0.9631	0.8431	0.7841	2.5662	1.9127	1.3204	1.0478	0.9031	0.6411
30	1.7763	0.9683	0.6584	0.5968	0.6105	2.5840	1.9740	1.0825	0.6917	0.6034	0.4957
$n = 100$											
0	3.3993	2.1985	2.0041	2.0085	2.0194	4.0537	3.5175	2.2578	2.0091	2.0085	2.0086
2	2.4114	1.9853	1.8734	1.7929	1.7209	4.0719	2.7219	2.0426	1.9186	1.8528	1.6385
5	2.2637	1.7104	1.5293	1.4120	1.3061	2.4285	2.3435	1.8053	1.6287	1.5322	1.1272
8	1.8024	1.5538	1.4343	1.3009	1.1920	4.8154	2.3776	1.6802	1.5786	1.4585	0.9793
10	2.6292	1.7532	1.3698	1.1761	1.0485	1.9663	2.6042	1.9263	1.5218	1.3188	0.8747
20	1.8025	1.2311	0.9557	0.8338	0.7876	2.9097	1.9048	1.3591	1.0615	0.8983	0.6431
30	2.0852	0.8412	0.5950	0.6052	0.6611	3.2496	2.4356	1.0418	0.6375	0.6012	0.5188

The simulation results given in Table 5, shows that, the proposed class of estimators yield accuracy in β values when compared with LS estimator. As error percentages in the x exceed 2%, the LS estimator struggles, while proposed classes consistently provide more accurate estimates of the slope parameter. We also notice that, adjusting the weights, particularly with decreasing values of k, enhances the robustness of the estimator ensuring more reliable estimates of the slope parameter. However, it is important to be cautious that assigning excess low value to k might discount the importance of extreme observations. Once we find a k value that provides the desired level of robustness or fits the regression line well, there is no need to consider lower values. Therefore, striking a balance between robustness and the risk of introducing inaccuracies is crucial when determining the finest value of k.

6. Illustration

In this section, we provide an example due to Montgomery et al. (2003) to illustrate the application of the proposed classes of estimators. It is a data consisting of 19 observations on first-unit satellite cost (y) and the weight of the electronics suite (x) from the US Air Force. The dataset is given by

Observation	Cost ($K)	Weight (lb)
1	2449	90.6*
2	2248	87.8*
3	3545	38.6
4	794	28.6
5	1619	28.9
6	2079	23.3
7	918	21.1
8	1231	17.5
9	3641	27.6
10	4314	39.2
11	2628	34.9
12	3989	46.6
13	2308	80.9*
14	376	14.6
15	5428	48.1
16	2786	38.1
17	2497	73.2*
18	5551	40.8
19	5208	44.6

* indicates outliers

To estimate α for the SLR model specified in (1), we utilize various members from the proposed classes of estimators and obtain $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. The LS estimate is given by $\hat{\beta}_{LS} = 13.88$ and $V(\hat{\beta}_{LS}) = 0.000103\,\sigma^2$. The values of $\hat{\beta}$ of some members of $1C_k$, $2C_k$ and their variance are given in Table 6.

Table 6: Computed values of $\hat{\beta}_{1C_k}$, $\hat{\beta}_{2C_k}$ and their variances

k	$\hat{\beta}_{1C_k}$	$V(\hat{\beta}_{1C_k})$	$\hat{\beta}_{2C_k}$	$V(\hat{\beta}_{2C_k})$
-2	275.966930	$0.018187\,\sigma^2$	1009.371100	$0.698733\,\sigma^2$
-1.5	176.756840	$0.004589\,\sigma^2$	836.019200	$0.430613\,\sigma^2$
-1	114.818710	$0.001140\,\sigma^2$	630.748400	$0.208517\,\sigma^2$
-0.5	78.065820	$0.000358\,\sigma^2$	430.528100	$0.076220\,\sigma^2$
0	56.494330	$0.000179\,\sigma^2$	272.349800	$0.021377\,\sigma^2$
0.5	43.669960	$0.000132\,\sigma^2$	167.841790	$0.004979\,\sigma^2$
1	35.875670	$0.000118\,\sigma^2$	106.350600	$0.001128\,\sigma^2$
1.5	31.035650	$0.000114\,\sigma^2$	71.951630	$0.000336\,\sigma^2$
2	27.983030	$0.000115\,\sigma^2$	52.713020	$0.000174\,\sigma^2$

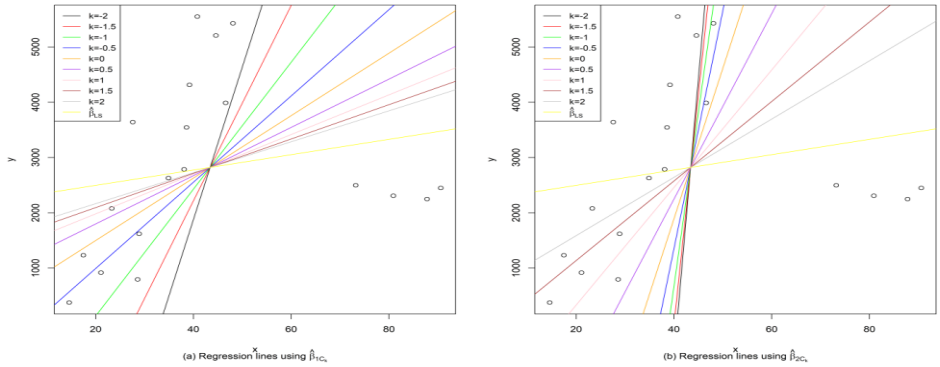

Figure 1: Fitted regression lines

From Table 6 and Figure 1, it is observed that, as k decreases, the impact of outliers diminishes on the proposed classes of estimators and exhibits a better fit to the data, indicating increased resilience against outliers.

CONCLUSIONS

In this study, we propose two classes of estimators for the slope parameter β in SLR. The classes are based on weighted averages of quasi ranges of the predictor variables, that is, one of the classes, $\hat{\beta}_{1C_k}$ is ratio of weighted deviations of corresponding y observations to the weighted quasi ranges of x observations and $\hat{\beta}_{2C_k}$ is weighted average of slopes based on quasi ranges of x observations. The proposed classes are unbiased and consistent estimators of β. The classes admit flexibility in being sensitive to outliers as well as resistant to outliers depending on the value of k. The variance of the proposed classes of estimators is minimized wrt k using Brent's method. $\hat{\beta}_{1C_k}$ admits optimal estimator when $k = 1$, whereas, $\hat{\beta}_{2C_k}$ admits when $k = 2$. $\hat{\beta}_{2C_k}$ is better than $\hat{\beta}_{1C_k}$ for $k > 2$. Also, for optimal values of k, $RE(\hat{\beta}_{2C_k}, \hat{\beta}_{1C_k}) \approx 1$, $RE(\hat{\beta}_{1C_k}, \hat{\beta}_{LS}) \approx 1$ and $RE(\hat{\beta}_{2C_k}, \hat{\beta}_{LS}) \approx 1$. Based on the simulation study, for $k < 0$, the robustness of the estimators from both the classes increases as less weight is assigned to extreme observations, while for $k > 0$, the efficiency improves. The proposed classes are useful to practitioners as they can be used under different situations just by choosing an appropriate value of k.

REFERENCES

[1] Bhat, S. V. and Bijjargi S. M. (2023). Nonparametric approach to estimation in linear regression. *International journal of current research*, 15, (04), 24419-24424.

[2] Bhat, S. V. and Bijjargi S. M. (2024). Simple linear slope estimators based on sample quasi ranges. To appear in *Statistics and Applications*.

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

[4] Gauss, C. F. (1809). Theoria motus corporum coelestum. *Werke*.

[5] Gore, A. P., and Rao, K. M. (1982). Nonparametric tests for slope in linear regression problems. *Biometrical Journal*, *24*(3), 229-237.

[6] Legendre, A. M. (1805). *Mémoire sur les opérations trigonométriques: dont les résultats dépendent de la figure de la terre* (No. 1). F. Didot.

[7] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2003). *Introduction to linear regression analysis*. John Wiley & Sons.

[8] Rao, M. K. (1982). Contribution to nonparametric inference in regression and other linear models. PhD thesis, Savitribai Phule Pune University, Pune, India.

[9] Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae*, *12*(85), 173.