International Journal of Innovative Technology and Interdisciplinary Sciences
www.IJITIS.org
ISSN:2613-7305
Volume 5, Issue 2, pp. 892-906, 2022
DOI: https://doi.org/10.15157/IJITIS.2022.5.2.892-906
Received October 21, 2021; Accepted March 16, 2022

Weighted Trimean as a Regressor in the Estimate of Theil-Sen Regression

Necati Alp Erilli *1

[1]Department of Econometrics, Sivas Cumhuriyet University, Sivas, Turkey

* aerilli@cumhuriyet.edu.tr

ABSTRACT

The most used method in nonparametric regression analysis is the Theil-Sen approach. With this method, all coefficient estimations are made with the median parameter as opposed to parametric methods. The most important criticism in computations with the median parameter is that the impact of extreme values does not participate in calculations. In this study, it was proposed to use the trimean parameter by weighting, which more effectively adds the effect of outliers to the average account in Theil-Sen regression analysis. In applications with 5 data sets, Theil-Sen calculations with weighted trimean were found to be more successful than calculations with the median parameter. Thus, in cases where the outliers are too high or directly affect the data, it can be said that the use of weighted trimean will yield more effective results.

Keywords: Theil-Sen Regression Analysis; weighted trimean; non-parametric regression analysis; trimean

1. INTRODUCTION

Regression analysis is perhaps the furthest widely used numerical techniques in statistical or mathematical forecasting studies. The aim of regression analysis is to investigate the connection between the addicted and detached variable(s), which is estimated to have a cause-effect relationship between them, to explain the assumed relationship between variables functionally and to define this relationship with a model. Some assumptions need to be provided for these relationships to produce successful results. Providing these assumptions in real-life applications is often difficult. The researcher can use nonparametric statistical techniques based on more flexible calculations in such cases. Nonparametric regression is an important sub-category of regression analysis in which the estimator does not take a prearranged form however, it is structured according to the information obtained from the data. Perhaps the most common of these techniques is the Theil-Sen method. Here, it is used the median-parameter instead of arithmetic mean and that gives consistent results. The Theil–Sen estimator is powerful estimator with a maximum cut-off point of around 0.293 and has a restricted impact function. It equates properly with the OLS estimator in low-sample size performance [1] and is vying in terms of MSE with different slope estimators [2].

The univariate Theil–Sen estimator possess eternal multivariate supplementations and has been widely adopted in many popular works on alternative statistics area. See, e.g., [3-7]. In research works [1] and [8] it is shown asymptotic notional performance to the least-squares estimator. At [9] applied it to astronomy, [10] for calibration and [11] to remote sensing. In [12] studied asymptotic properties for Theil-Sen estimator with a random covariate where [13] and [14] also studied with censored data. At [15] provided the robust conformity and asymptotic dispersion of the Theil–Sen estimator in simple linear regression models with tolerant error dispersions. In [16] it has been introduced a method for the estimation of regression parameters under data that have equal values. Furthermore, [17] introduced Theil-Sen estimators in multiple linear regression analysis. In [18] is also studied Theil-Sen type models in multiple regression analysis based on Oja median parameter. Finally, [19] suggested using Trimean-parameter, which is one of the members of central tendency, alternative to the median-parameter in all Theil-Sen calculations.

In this study, it was proposed to use weighted trimean alternative to the median parameter in Theil-Sen regression analysis. This change is recommended for cases where the median parameter does not reflect the effect of outliers in the model. With weighting, it is aimed to calculate the arithmetic contributions of each observation value to the model directly. Calculations with the proposed method were tested on 5 different data set and it can be said that the results were more consistent than calculations with the median.

2. MATERIALS AND METHOD

Theil-Sen regression analysis was first proposed by [20] and the procedure is firstly known as Theil's Method. After [8] highlighted the method with the relationship to Kendall's tau, it is named as the Theil-Sen method. In literature, it's also named as Theil–Kendall as well. Theil proposed estimating a regression line and its slope as the median of all line slopes connecting each pair of points with different independent values. For a pair (x_i, y_i) and (x_j, y_j) the relevant slope can be calculated as $S_{ij} = (y_j - y_i)/(x_j - x_i)$. So, there must be $n(n+1)/2$ slopes for any data. The $\hat{\beta}_1$ statistic, which is the estimator of the parameter β_1 in simple regression analysis, is calculated as the median of the slope values: $\hat{\beta}_1 = Median(S_{ij})$. Theil suggested for the estimation of the intercept as $\hat{\beta}_0 = Median\left(y_i\right) - \hat{\beta}_1 Median\left(x_i\right)$ [20].

To show the slope parameter $\hat{\beta}_1$, which is estimated from the Theil-Sen method, the following two different methods have been developed to estimate the intercept parameter β_0 for the model.

2.1 Optimum and Hodges-Lehmann Method

Let's define $d_i = y_i - \hat{\beta}_1 x_i$ calculated for all observations where $\hat{\beta}_1$ is calculated with the Theil-Sen method. $\hat{\beta}_0$ is calculated as the median parameter of all d_i; d_i $\left(\hat{\beta}_0 = Median\{d_i\}\right)$. The optimum approach does not require the assumption of symmetrically distributed d_i. It is better suited especially for data with outliers.

Hodges-Lehmann method to predict $\hat{\beta}_0$ is defined as the mean value of d_i $\left(\hat{\beta}_0 = Mean\{d_i\}\right)$. Hodges-Lehmann method cannot give powerful against data with outliers [21, 22].

2.2 Weighted Median

The median value of a sequence of observations is acquired by putting in order the numbers by smallest to largest, then placing them in ascending order and choosing the value in the middle of the sorted list. The weighted median parameter of a list of numbers x_i with heights w_i is obtained as given:

Foremost, the x_i numbers are put in ascending order from smallest to largest. By changing the indicators, we can collocate the data to be $x_1 \le x_2 \le \cdots \le x_n$. The weights w_i should be nonnegative and should add to 1. The weights can calculate with the formula $w_{ij} = |x_i - x_j| / \sum |x_i - x_j|$. Here, we must find the index *k* with given in equation (1):

$$\begin{aligned} w_1 + w_2 + \cdots + w_{k-1} &< 0.5 \\ w_1 + w_2 + \cdots + w_{k-1} + w_k &> 0.5 \end{aligned} \tag{1}$$

In Equation 1, x_k described as the weighted-median. Sometimes in the case of $w_1 + w_2 + \cdots + w_{k-1} = 0.5$ such an index of *k* is defined that it is then $(x_{k-1} + x_k)/2$ defined as the weighted median.

The slope of the line passing through the point (x_0, y_0) minimizing the sum of absolute deviations from the data observations can be defined as the weighted-median of the slopes $b_i = (y_i - y_0)/(x_i - x_0)$ for the routes between data points (x_i, y_i) and the given points (x_0, y_0), with each weight comparative to the *x*-distance $|x_i - x_0|$ between the two points [4].

2.3 Trimean

The trimean (TM) is a measure of a probability distribution's location defined as a weighted average of the distribution's median and its two quartiles, see equation (2).

$$TM = \frac{Q_1 + 2 \times Median + Q_3}{4} \tag{2}$$

After Tukey has given this formula's name with a set of techniques in his book it is sometimes called Tukey's Trimean [23].

2.4 Weighted Trimean

The weighted trimean calculation can be done as like the weighted median calculation. The weights of the trimean parameters described above, i.e. the weighted quartiles, are found similarly to the median parameter. After putting the numbers x_i in increasing order, the weights w_i should be nonnegative and should add to 1. The first and third quartile values are calculated by finding j and t indices that provide one of the following conditions:

$$\begin{aligned} &w_1 + w_2 + \cdots + w_{j-1} < 0.25 \\ &w_1 + w_2 + \cdots + w_{j-1} + w_j > 0.25 \end{aligned} \tag{3}$$

and

$$\begin{aligned} &w_1 + w_2 + \cdots + w_{t-1} < 0.75 \\ &w_1 + w_2 + \cdots + w_{t-1} + w_t > 0.75 \end{aligned} \tag{4}$$

Then x_j is the weighted first quartile $(W.Q_1)$ and x_t is the weighted third quartile $(W.Q_3)$. Like weighted median, if there is an index j or t such that $w_1 + w_2 + \cdots + w_{j-1} = 0.25$ or $w_1 + w_2 + \cdots + w_{t-1} = 0.75$; then $(x_{j-1} + x_j)/2$ or $(x_{t-1} + x_t)/2$ can define as the weighted quartile. It can be arranged the weighted trimean formula as given in equation (5):

$$WTM = \frac{Weighted\,Q_1 + \left(2 \times Weighted\ Median\right) + Weighted\,Q_3}{4} \tag{5}$$

2.5 Significance test of slope parameter

To test $H_0 : \hat{\beta}_1 = 0$ in Theil-Sen regression, we can use the test statistics given in equation (6) and (7) respectively [4]:

$$|t| = \frac{|U|}{SD(U)} \tag{6}$$

Where

$$U = \sum \left[rank(y_i) - \frac{n+1}{2} \right] x_i \text{ and } SD(U) = \sqrt{\frac{n(n+1)}{12} \sum (x_i - \bar{x})^2} \tag{7}$$

The approximate p-value of the test is calculated to be Prob$\left[|Z| \geq |t|\right]$, where Z is a random variable having a standard normal distribution.

2.6 Proposed Method

In this study, new parameter prediction method for Theil-Sen regression analysis have been proposed for both the slope and the intercept parameter. In the Theil-Sen method, the slope parameter was calculated while the median of the calculated slope values was taken from observation binaries. Here slope parameters were also estimated using trimean, weighted median and weighted trimean instead of the median. Similarly, when calculating the intercept parameter, trimean and weighted trimean were used instead of Theil's proposed median calculation. Finally, trimean and weighted trimean were used in addition to arithmetic mean and median use in the d_i calculation obtained from all observations. Thus, 4 slope parameters and 22 intercept parameters were estimated for each data. The calculation methods proposed in the application are summarized in Table.1.

Table 1. Model calculation methods used in applications

Slope	Intercept
With Median	Theil-Sen with Median
	Mean of d_i
	Median of d_i
	Trimean of d_i
With Weighted Trimean	Theil-Sen with Median
	Theil-Sen with W. Median
	Mean of d_i
	Median of d_i
	W. Median of d_i
	Trimean of d_i
With Trimean	Theil-Sen with Median
	Theil-Sen with Trimean
	Mean of d_i
	Median of d_i
	Trimean of d_i
With Weighted Trimean	Theil-Sen with Median
	Theil-Sen with Trimean
	Theil-Sen with W. Trimean
	Mean of d_i
	Median of d_i
	Trimean of d_i
	W. Trimean of d_i

Ordinary Least Squares (OLS), M-Estimator, MM-Estimator, S-Estimator, Least Median Square (LMS) and Least Trimmed Squares (LTS) methods applied to the same data to compare the power of the generated models, and all the results obtained were compared with the information criteria Akaike, Schwarz, and Mean Absolute Percentage Error (MAPE) value, see equation (8):

$$
\begin{aligned}
AIC &= 2k - 2\ln(L) \\
BIC &= k\ln n - 2\ln(L) \\
MAPE &= \left(\frac{1}{n} \sum \frac{|Actual - Forecast|}{|Actual|} \right) \times 100
\end{aligned}
\tag{8}
$$

AIC and BIC criteria are important comparison criteria in regression analysis - although other comparison criteria are used in different types of studies [24, 25] - the MAPE criterion is more important for us here. MAPE is an asymmetric value and states higher errors if the estimate is more than the actual and lower errors when the estimate is less than the actual. This can be explained by looking at the outliers: a forecast of zero can never be off by more than 100%, but there is no limitation to the errors on the superior side [26].

3. APPLICATION

In the application part, 5 different data were used to compare the proposed method. In addition, 10%, 20%, and 30% outliers were created in blood pressure data (for dependent and independent variable separately), an outlier observation was created by changing the last observation value of forearm data and the strength of the proposed method was tested on these data sets. The scatter plots of the data sets used in the application is given in Figure.1.

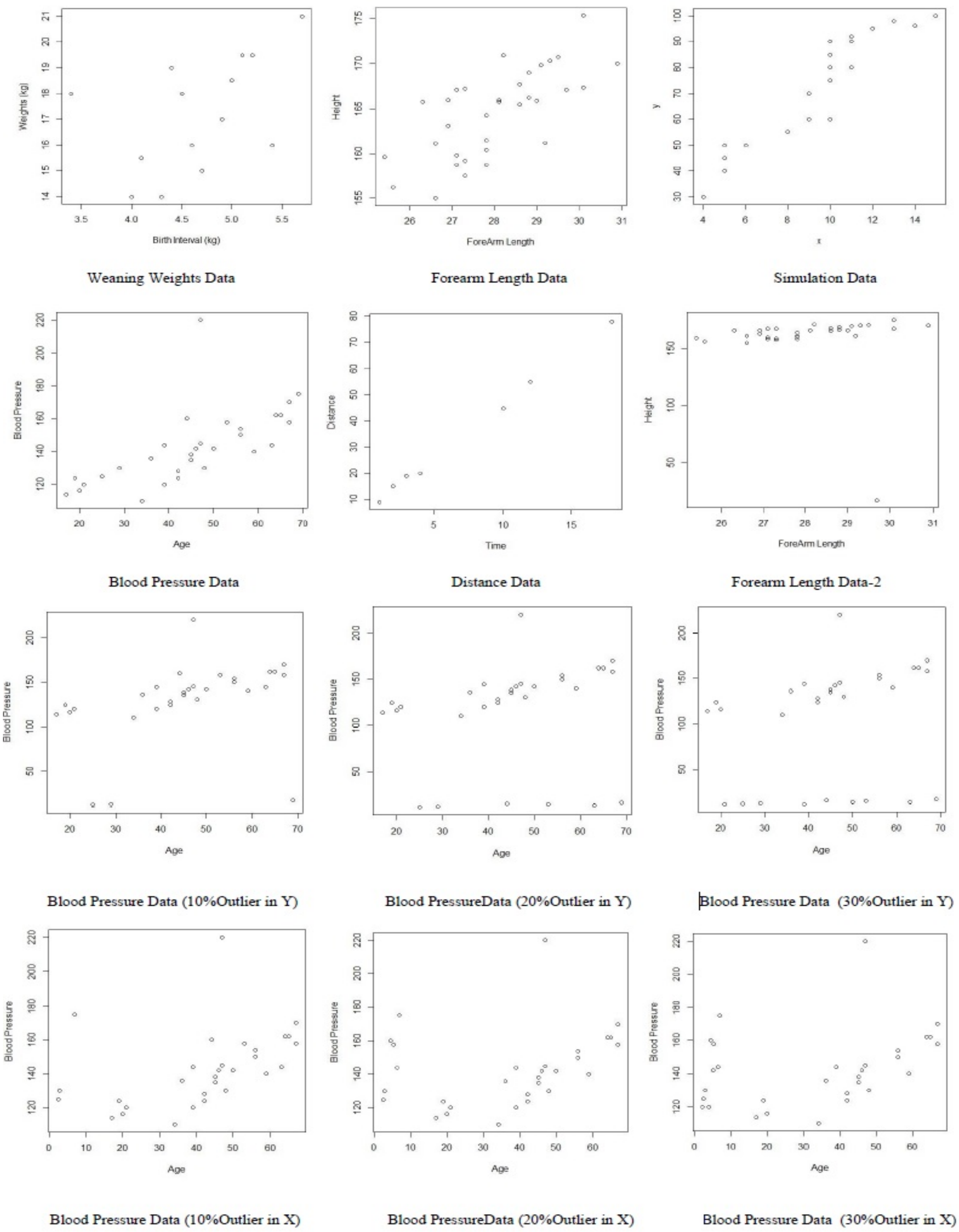

Figure 1. Scatter Plot of all data sets used in application

The short definitions of all data sets are given at Table. 2.

Table 2. Short definitions of data sets

Data	Definitions	Sample Size	Data Source
Forearm Length	The heights (Y) in cm and forearm lengths (X) in cm of 33 black female applicants	33	[4]
Simulation	Dependent (Y) and Independent (X)	20	[16]
Distance	Distance (Y) represents the *ith* distance at time (X) t_i	7	[8], [27]
Weaning Weights	Birth intervals in kg. (X) and weaning weights in kg. (Y) of ships	14	[28]
Blood Pressure	The systolic blood pressure (Y) and age (X)	30	[29]

Parameter estimates for each data set according to the 22 model structures given in Table.1 were made according to the Theil-Sen method both classical (with median and weighted median) and proposed form (with trimean and weighted trimean). In addition, parameter estimates were obtained separately with the 6 methods (OLS, S-Est., M-Est., MM-Est, LTS, LMS) mentioned above.

4. RESULTS

Parameter estimation and model selection criteria values obtained according to all data sets are given with Tables 3 until 9 respectively.

Table 3. Forearm Data Parameter Results

	Method	β_0	β_1	AIC	BIC	MAPE
With Median	Theil-Sen with Median	91,44065	2,674797	2,85391	2,944607	1,960853
	Mean of d_i	89,604484	2,674797	2,606352	2,69705	1,798419
	Median of d_i	89,94065	2,674797	2,615723	2,706421	1,79545
	Trimean of d_i	89,737602	2,674797	2,607828	2,698525	1,797243
With Weighted Median	Theil-Sen with Median	91,44065	2,682927	2,910263	3,000961	2,009203
	Theil-Sen with W. Median	91,44065	2,682927	2,910263	3,000961	2,009203
	Mean of d_i	89,376644	2,682927	2,606574	2,697271	1,799091
	Median of d_i	89,714634	2,682927	2,616044	2,706741	1,796107
	W. Median of d_i	89,714634	2,682927	2,616044	2,706741	1,796107
	Trimean of d_i	89,508537	2,682927	2,608021	2,698719	1,797926
With Trimean	Theil-Sen with Median	90,11041	2,722647	2,857698	2,948396	1,963387
	Theil-Sen with Trimean	88,820079	2,722647	2,633361	2,724058	1,810557
	Mean of d_i	88,263513	2,722647	2,60792	2,698617	1,802376
	Median of d_i	88,61041	2,722647	2,61788	2,708578	1,799313
	Trimean of d_i	88,389417	2,722647	2,609238	2,699935	1,801265
	Theil-Sen with Median	92,206158	2,647261	2,851958	2,942656	1,959395
	Theil-Sen with Trimean	90,92525	2,647261	2,630564	2,721261	1,808493
	Theil-Sen with W.	90,92525	2,647261	2,630564	2,721261	1,808493

With Weighted Trimean	Trimean					
	Mean of d_i	90,376166	2,647261	2,60574	2,696438	1,796141
	Median of d_i	90,68287	2,647261	2,613552	2,704249	1,793433
	Trimean of d_i	90,501791	2,647261	2,607055	2,697753	1,795032
	W. Trimean of d_i	90,460524	2,647261	2,606334	2,697031	1,795396
	OLS	92,2093	2,5818	***2,60514***	***2,695837***	1,790739
	M Est	92,23464	2,581122	2,605142	2,695839	1,790627
	S Est	97,25878	2,406018	2,610707	2,701405	***1,781087***
	LMS	-1,361111	5,888889	3,561604	3,652302	2,5287
	LTS	97,2143	2,4286	2,649528	2,740225	1,814522
	MM Est	92,2093	2,5818	2,60514	2,695837	1,790739

Table 4. Blood Pressure Data Parameter Results

Method		β_0	β_1	AIC	BIC	MAPE
With Median	Theil-Sen with Median	95,5	1	5,780839	5,874253	6,39413
	Mean of d_i	97,4	1	5,768028	5,861441	6,355294
	Median of d_i	97	1	5,768599	5,862012	6,321016
	Trimean of d_i	96,25	1	5,77274	5,866153	6,331468
With Weighted Median	Theil-Sen with Median	97,833333	0,948718	5,780303	5,873716	6,282484
	Theil-Sen with W. Median	96,461538	0,948718	5,804849	5,898262	6,416658
	Mean of d_i	99,71453	0,948718	5,767739	5,861152	6,339352
	Median of d_i	99,217949	0,948718	5,768619	5,862032	6,317232
	W. Median of d_i	100,33333	0,948718	5,769106	5,862519	6,381161
	Trimean of d_i	98,400641	0,948718	5,773887	5,867301	***6,280827***
With Trimean	Theil-Sen with Median	93,79375	1,0375	5,783883	5,877296	6,509715
	Theil-Sen with Trimean	93,527344	1,0375	5,78771	5,881123	6,541245
	Mean of d_i	95,7075	1,0375	5,770924	5,864337	6,428302
	Median of d_i	95,425	1,0375	5,771208	5,864621	6,416189
	Trimean of d_i	94,726563	1,0375	5,774345	5,867758	6,427263
With Weighted Trimean	Theil-Sen with Median	97,425	0,957692	5,780094	5,873507	6,296993
	Theil-Sen with Trimean	97,19351	0,957692	5,783357	5,87677	6,311087
	Theil-Sen with W. Trimean	94,375962	0,957692	5,850889	5,944302	6,738688
	Mean of d_i	99,309487	0,957692	5,767482	5,860896	6,339261
	Median of d_i	98,848077	0,957692	5,768243	5,861656	6,318708
	Trimean of d_i	98,033413	0,957692	5,773285	5,866698	6,28242
	W. Trimean of d_i	99,853846	0,957692	5,768541	5,861954	6,367782
	OLS	98,71472	0,97087	***5,767342***	***5,860755***	6,339127
	M Est	97,96348	0,935373	5,787942	5,881355	6,289953
	S Est	99,8314	0,956852	5,768338	5,861751	6,36193
	LMS	103,5333	0,8444	5,782974	5,876387	6,371104
	LTS	98,9118	0,9084	5,794699	5,888112	6,293745
	MM Est	97,71829	0,938997	5,789116	5,88253	6,300701

Table 5. Simulation Data Parameter Results

	Method	β_0	β_1	AIC	BIC	MAPE
With Median	Theil-Sen with Median	10,83333	6,666667	4,257661	4,357234	9,464635
	Mean of d_i	9,3833333	6,666667	4,220633	4,320206	9,16802
	Median of d_i	10	6,666667	4,227433	4,327006	9,172657
	Trimean of d_i	9,6458333	6,666667	4,221869	4,321442	9,169994
With Weighted Median	Theil-Sen with Median	11,25	6,625	4,257826	4,3574	9,512695
	Theil-Sen with W. Median	18,75	6,625	5,114296	5,21387	15,89248
	Mean of d_i	9,775	6,625	4,219492	4,319065	9,176606
	Median of d_i	10,3125	6,625	4,224668	4,324241	9,193147
	W. Median of d_i	11,875	6,625	4,295724	4,395297	9,73168
	Trimean of d_i	9,9765625	6,625	4,220221	4,319795	9,178121
With Trimean	Theil-Sen with Median	9,167	6,8333	4,26269	4,362263	9,272438
	Theil-Sen with Trimean	9,0419875	6,8333	4,257136	4,35671	9,228637
	Mean of d_i	7,81698	6,8333	4,230835	4,330408	***9,133685***
	Median of d_i	8,75025	6,8333	4,246185	4,345758	9,176409
	Trimean of d_i	8,2606938	6,8333	4,234325	4,333898	9,137021
With Weighted Trimean	Theil-Sen with Median	11,875	6,5625	4,259142	4,358716	9,584784
	Theil-Sen with Trimean	11,648438	6,5625	4,248136	4,34771	9,505402
	Theil-Sen with W. Trimean	13,984375	6,5625	4,430548	4,530122	10,8555
	Mean of d_i	10,3625	6,5625	***4,218848***	***4,318421***	9,189484
	Median of d_i	10,78125	6,5625	4,221994	4,321568	9,223882
	Trimean of d_i	10,519531	6,5625	4,219291	4,318864	9,190665
	W. Trimean of d_i	15,703125	6,5625	4,632711	4,732284	12,18105
	OLS	10,3638	6,5624	***4,218848***	***4,318421***	9,189507
	M Est	8,01292	6,779232	4,228282	4,327856	9,142478
	S Est	7,600644	6,806752	4,232467	4,332041	9,135653
	LMS	3	7,66667	4,5292	4,628774	9,746567
	LTS	9,8939	6,6201	4,21949	4,319064	9,178163
	MM Est	8,13464	6,758158	4,22782	4,327393	9,146246

Table 6. Weaning Weights Data Parameter Results

	Method	β_0	β_1	AIC	BIC	MAPE
With Median	Theil-Sen with Median	5,875	2,5	1,609254	1,700548	9,90708
	Mean of d_i	5,4615385	2,5	1,583683	1,674977	9,641627
	Median of d_i	5,25	2,5	1,606279	1,697573	9,505814
	Trimean of d_i	5,3125	2,5	1,597149	1,688443	9,545941
With Weighted Median	Theil-Sen with Median	10,346154	1,538462	1,550779	1,642073	10,41799
	Theil-Sen with W. Median	10,769231	1,538462	1,667667	1,758961	10,68962
	Mean of d_i	9,9142012	1,538462	1,528164	1,619458	10,14067
	Median of d_i	9,4615385	1,538462	1,615832	1,707126	9,850045
	W. Median of d_i	11,076923	1,538462	1,795785	1,887079	11,09506
	Trimean of d_i	9,5673077	1,538462	1,586079	1,677372	9,917951

With Trimean	Theil-Sen with Median	5,68125	2,541667	1,615554	1,706848	9,884941
	Theil-Sen with Trimean	5,4927083	2,541667	1,592575	1,683869	9,763892
	Mean of d_i	5,2685897	2,541667	1,589972	1,681266	9,620002
	Median of d_i	5,0791667	2,541667	1,608848	1,700142	***9,498388***
	Trimean of d_i	5,1260417	2,541667	1,602421	1,693715	9,528483
With Weighted Trimean	Theil-Sen with Median	8,6918269	1,894231	1,551967	1,643261	10,22895
	Theil-Sen with Trimean	8,5194712	1,894231	1,529703	1,620996	10,1183
	Theil-Sen with W. Trimean	8,7997596	1,894231	1,574037	1,665331	10,29825
	Mean of d_i	8,266716	1,894231	1,527655	1,618948	9,956022
	Median of d_i	7,7336538	1,894231	1,637998	1,729292	9,613782
	Trimean of d_i	7,8850962	1,894231	1,592342	1,683636	9,711012
	W. Trimean of d_i	9,1103365	1,894231	1,668122	1,759416	10,56057
	OLS	9,2288	1,7121	***1,520663***	***1,611957***	10,12674
	M Est	8,831151	1,797491	1,521399	1,612693	10,081
	S Est	-2,928626	4,232271	2,043818	2,135112	9,630911
	LMS	-6,75	5	2,313628	2,404921	10,8251
	LTS	-0,6911	3,8084	3,883532	3,974826	33,84806
	MM Est	-2,7485	4,1976	2,030871	2,122165	9,614567

Table 7. Distance Data Parameter Results

Method		β_0	β_1	AIC	BIC	MAPE
With Median	Theil-Sen with Median	4	4	1,428879	1,413425	4,081533
	Mean of d_i	5,8571429	4	0,113595	0,098141	2,61496
	Median of d_i	6	4	0,129596	0,114142	2,648667
	Trimean of d_i	6	4	0,129596	0,114142	2,648667
With Weighted Median	Theil-Sen with Median	4	4	1,428879	1,413425	4,081533
	Theil-Sen with W. Median	7	4	0,822743	0,807289	3,067766
	Mean of d_i	5,8571429	4	0,113595	0,098141	2,61496
	Median of d_i	6	4	0,129596	0,114142	2,648667
	W. Median of d_i	6,5574519	4	0,441171	0,425717	2,882294
	Trimean of d_i	6	4	0,129596	0,114142	2,648667
With Trimean	Theil-Sen with Median	4	4	1,428879	1,413425	4,081533
	Theil-Sen with Trimean	7	4	0,822743	0,807289	3,067766
	Mean of d_i	5,8571429	4	0,113595	0,098141	2,61496
	Median of d_i	6	4	0,129596	0,114142	2,648667
	Trimean of d_i	6	4	0,129596	0,114142	2,648667
With Weighted Trimean	Theil-Sen with Median	3,9238095	4,019048	1,377907	1,362453	4,089835
	Theil-Sen with Trimean	5,147619	4,019048	0,334863	0,319409	***2,419407***
	Theil-Sen with W. Trimean	-0,247619	4,019048	3,485823	3,470368	14,03076
	Mean of d_i	5,7210884	4,019048	0,101266	0,085812	2,566428
	Median of d_i	5,6571429	4,019048	0,104533	0,089078	2,539628

	Trimean of d_i	5,7666667	4,019048	0,102927	0,087473	2,58553
	W. Trimean of d_i	6,5574519	4,019048	0,545754	0,5303	2,916946
	OLS	5,704626	4,021352	***0,10112***	***0,085666***	2,563721
	M Est	5,704626	4,021352	***0,10112***	***0,085666***	2,563721
	S Est	6,225942	3,992537	0,198714	0,18326	2,729782
	LMS	7,6667	3	4,061658	4,046204	7,085495
	LTS	7,125	3,9375	0,6513	0,635845	3,006451
	MM Est	5,7046	4,0214	0,10112	0,085666	2,563797

Table 8. Forearm Data with Outlier Parameter Results

Method		β_0	β_1	**AIC**	**BIC**	**MAPE**
With Median	Theil-Sen with Median	95,93421	2,513158	6,717272	6,80797	29,78193
	Mean of d_i	89,57429	2,513158	6,660446	6,751143	29,58168
	Median of d_i	94,25	2,513158	6,69156	6,782258	29,33255
	Trimean of d_i	94,18882	2,513158	6,690764	6,781461	29,32311
With Weighted Median	Theil-Sen with Median	96,60889	2,488889	6,71668	6,807377	29,77565
	Theil-Sen with W. Median	96,30889	2,488889	6,71158	6,802277	29,66534
	Mean of d_i	90,25441	2,488889	6,659919	6,750616	29,57435
	Median of d_i	94,89556	2,488889	6,690598	6,781296	29,31857
	W. Median of d_i	95,10889	2,488889	6,693436	6,784133	29,35951
	Trimean of d_i	94,855	2,488889	6,690072	6,78077	29,31232
With Trimean	Theil-Sen with Median	95,47479	2,529684	6,717677	6,808374	29,78642
	Theil-Sen with Trimean	94,05858	2,529684	6,695566	6,786263	29,39278
	Mean of d_i	89,11117	2,529684	6,660806	6,751503	29,58668
	Median of d_i	93,81041	2,529684	6,692219	6,782916	29,34206
	Trimean of d_i	93,73518	2,529684	6,691236	6,781933	29,33046
With Weighted Trimean	Theil-Sen with Median	98,87428	2,4074	6,714709	6,805407	29,75586
	Theil-Sen with Trimean	98,87428	2,4074	6,714709	6,805407	29,75586
	Theil-Sen with W. Trimean	97,27336	2,4074	6,690141	6,780838	29,31199
	Mean of d_i	92,53808	2,4074	6,658169	6,748867	29,54972
	Median of d_i	97,06316	2,4074	6,687406	6,778103	29,27165
	Trimean of d_i	97,09187	2,4074	6,687772	6,77847	29,27716
	W. Trimean of d_i	97,26595	2,4074	6,690042	6,78074	29,31057
	OLS	214,8408	-1,95676	***6,610594***	***6,701291***	***29,24901***
	M Est	90,43813	2,647577	6,69385	6,784548	29,37348
	S Est	91,1261	2,629785	6,695965	6,786663	29,39636
	LMS	-1,36111	5,888889	6,774331	6,865029	30,83539
	LTS	89,7697	2,6885	6,701229	6,791926	29,47653
	MM Est	90,7107	2,6373	6,693428	6,784126	29,36744

Table 9. MAPE Results for Blood Pressure Data with Outliers

Method		MAPE Results					
		10% Outlier added for dependent variable	20% Outlier added for dependent variable	30% Outlier added for dependent variable	10% Outlier added for independent variable	20% Outlier added for independent variable	30% Outlier added for independent variable
With Median	Theil-Sen with Median	90,011964	167,18701	247,72648	***8,36351234***	10,09043303	9,699071086
	Mean of d_i	85,799721	152,44328	206,73265	9,02343751	10,90764041	10,46759306
	Median of d_i	90,989571	173,5558	262,65671	8,367091274	10,06199915	9,725130722
	Trimean of d_i	89,521598	171,73784	219,60047	8,398201712	10,02816284	9,715246283
With Weighted Median	Theil-Sen with Median	90,142288	167,14282	247,72648	8,435132229	10,06508756	9,878245984
	Theil-Sen with W. Median	90,518145	175,67283	206,73265	8,422650529	***9,9738842***	9,791228585
	Mean of d_i	85,961161	152,34967	206,73265	8,96170457	10,56878409	10,44552516
	Median of d_i	90,892153	174,01833	262,65671	8,450830651	9,979131185	9,90253354
	W. Median of d_i	93,37771	178,21479	277,2059	8,635538227	10,32604327	10,12991036
	Trimean of d_i	89,540339	172,03913	219,60047	8,439552706	9,997818947	9,979039344
With Trimean	Theil-Sen with Median	89,561329	167,15552	246,79008	8,36719276	10,087444	9,701505939
	Theil-Sen with Trimean	88,622905	165,15626	205,26019	8,381995548	10,41496696	10,16295343
	Mean of d_i	***85,17595***	152,37658	205,89369	9,029051369	10,87442478	10,55597492
	Median of d_i	89,541725	173,97509	258,90975	8,371852674	10,06340974	***9,6933922***
	Trimean of d_i	88,96768	172,03246	218,31859	8,404334841	10,02932298	9,717454197
With Weighted Trimean	Theil-Sen with Median	89,945993	167,12188	244,90174	8,444526677	10,07486466	9,93196711
	Theil-Sen with Trimean	89,092217	165,13369	***203,2056***	8,449112336	10,2723477	10,18544636
	Theil-Sen with W. Trimean	89,147611	172,62751	258,80529	8,453859289	9,975492264	9,831925972
	Mean of d_i	85,717998	152,3053	204,20185	8,9603464	10,57289488	10,47977282
	Median of d_i	90,856118	173,66012	254,81754	8,458527016	9,982304487	9,966368591
	Trimean of d_i	89,47429	171,76569	217,21412	8,448865397	10,0064451	10,03585577
	W. Trimean of d_i	92,417431	178,34936	276,95901	8,589038456	10,48685782	10,14300173
	OLS	85,687843	***151,7372***	206,6479	9,028318453	10,7496849	10,50390427
	M Est	90,52445	174,25551	266,95822	8,392491025	10,16382495	10,01283878
	S Est	91,867893	176,95347	269,02092	8,52788841	10,83557705	9,991691735
	LMS	94,103197	181,82479	261,09828	8,745206573	10,9154672	10,49225217
	LTS	92,819919	166,01738	219,42656	8,382941437	9,962196631	9,724423279
	MM Est	90,499687	174,05126	266,59564	8,395536512	10,84162949	9,72226608

According to the estimation results made for the 5 data sets defined above, the models which include slope parameter calculation with weighted trimean and intercept parameter calculation with mean were found to give the best model criteria among all Theil-Sen estimations. The best-estimated model in the distance data used in Sen (1968) was obtained by weighted trimean. The results obtained with trimean were found to be close to one percent of OLS or other best model results, even in cases where the best results were not.

In the Blood Pressure and Forearm Data where outliers were created also showed that the conditions were the same. The best AIC, BIC and MAPE values belong to OLS, while the second-best results were determined for weighted trimean.

When we look at the results of the analysis in general, it is seen that the MAPE values obtained from OLS generally have the best value. However, a decrease in the MAPE values obtained from OLS is observed as the rate of outliers in the data increases. For the results obtained from Theil-Sen method, MAPE values obtained from trimean and weighted trimean can be said to be more successful than others.

Calculations were made according to the proposed method with 12 data, including 5 original forms and 7 with outliers' value-added form. Apart from parameter calculations, model significance values are also calculated. In Table 10, the model significance values for Theil-Sen and OLS are given for all data groups.

Table 10. Significance test results for all data sets

Data	OLS	Theil-Sen
Forearm Length	*p=0.000**	*t=2.9524**
Blood Pressure	*p=0.000**	*t=4.3603**
Simulation	*p=0.000**	*t=4.1381**
Weaning Weights	*p=0.084*	*t=1.8014*
Distance	*p=0.000**	*t=2.3265**
Forearm Length - 1 Outlier Added	*p=0.586*	*t=3.2944**
Blood Pressure %10 Outlier Added (Added in Y)	*p=0.056*	*t=3.5583**
Blood Pressure %20 Outlier Added (Added in Y)	*p=0.426*	*t=2.6783**
Blood Pressure %30 Outlier Added (Added in Y)	*p=0.229*	*t=2.7501**
Blood Pressure %10 Outlier Added (Added in X)	*p=0.017**	*t=2.8917**
Blood Pressure %20 Outlier Added (Added in X)	*p=0.142*	*t=1.5725*
Blood Pressure %30 Outlier Added (Added in X)	*p=0.072*	*t=1.8249*

For the results in Table 10, it is seen that Theil-Sen with weighted trimean results is significant in the outliers in the dependent variable. Similarly, the number of significant models appears to be higher in models estimated with weighted trimean.

5. CONCLUSION

Non-parametric statistical methods are frequently preferred because they are not based on many assumptions. Non-parametric regression analysis is an alternative method in prediction areas. Theil-Sen regression method is one of the most used methods among all these. In this process, parameter estimations are calculated with the help of median parameter of slope values among all observation values.

In all computations made with the median parameter, the influence of the agglomeration in the data cannot be included in the estimated model. With this paper, it was proposed to use weighted trimean instead of median to solve this problem. Thus, it is planned to add the effect of agglomeration to the measure of central tendency directly. In application part 5 data sets and 12 model structures were applied to try the calculation method proposed with weighted trimean parameter. To make comparisons, the results obtained from OLS, S-Est., M-Est., MM-Est., LTS and LMS methods were compared with the model selection criteria. When we look at the results obtained with Trimean in general, it was seen that the Theil-Sen method gave the best results in its calculations. Calculations with OLS have been shown to give the best results in many models, but 7 of 12 models have been calculated as meaningless. This number was found to be 2 in 12 in calculations made with weighted trimean.

To look for the results in general, it is clearly seen that the parameter calculations made with weighted trimean give more successful results according to the model selection criteria and the model significances compared to the methods studied above. It is recommended to use calculations with weighted trimean in other non-parametric regression analysis methods.

CONFLICT OF INTERESTS

The authors would like to confirm that there is no conflict of interests associated with this publication and there is no financial fund for this work that can affect the research outcomes.

REFERENCES

[1] Wilcox R.R. Simulations on the Theil–Sen regression estimator with right-censored data. *Statistics & Probability Letters,* 1998; 39; 43-47.

[2] Dietz E.J. A comparison of robust estimators in simple linear regression. *Comm. Statist., Part B Simulation Comput*., 1987; 16; 1209-1227.

[3] Sprent P. (1993) Applied Nonparametric Statistical Methods. second ed. CRC Press, New York, USA.

[4] Birkes D. and Dodge Y. (1993) Alternative Methods of Regression, John Wiley & Sons Inc., NY, USA.

[5] Hollander M. and Wolfe D.A. (1999) Nonparametric Statistical Methods, second ed., Wiley, New York, USA.

[6] Rousseeuw P.J. and Leroy A.M. (2003) Robust Regression and Outlier Detection. Wiley, New York, USA.

[7] Hardle W. (1994) Applied Nonparametric Regression, Cambridge University Press, London, UK.

[8] Sen P.K. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association,* 1968; 63; 1379–1389.

[9] Akritas M.G., Murphy, S.A. and LaValley M.P. The Theil–Sen estimator with doubly censored data and applications to astronomy. *J. Amer. Statist. Assoc.,* 1995; 90; 170-177.

[10] Lavagnini I., Badocco D., Pastore P. and Magno F. Theil–Sen nonparametric regression technique on univariate calibration, inverse regression and detection limits. *Talanta*, 2011; 87; 180-188.

[11] Fernandes R. and Leblanc S.G. Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. *Remote Sensing of Environment*, 2005; 95; 303-316.

[12] Wang X.Q. Asymptotics of the Theil–Sen estimator in simple linear regression models with a random covariate. *Nonparametric Statist.,* 2005; 17; 107-120.

[13] Jones M.P. A class of semi-parametric regressions for the accelerated failure time model. *Biometrika*, 1997; 84; 73–84.

[14] Mount D.M., Netanyahu, N.S. Efficient randomized algorithms for robust estimation of circular arcs and aligned ellipses. *Comput. Geometry: Theory Appl.,* 2001; 19; 1–33.

[15] Peng H., Wang, S. and Wang, X. Consistency and asymptotic distribution of the Theil–Sen estimator. *Journal of Statistical Planning and Inference*, 2008; 138; 1836-1850

[16] Erilli N.A. and Alakus K. Non-Parametric Regression Estimation for Data with Equal Values. *European Scientific Journal*, 2014; (10)4; 70-82.

[17] Dang X., Peng H., Wang X. and Zhang, H. Theil-Sen Estimators in a Multiple Linear Regression Model. *Olemiss Edu*. Available online. http://home.olemiss.edu/~xdang/papers (Accessed on 1 January 2019).

[18] Shen G. Asymptotics of a Theil-type estimate in multiple linear regression, *Statistics & Probability Letters,* 2009; 79; 1053-1064.

[19] Erilli A.N. Use of Trimean in Theil-Sen Regression Analysis. *Bulletin of Economic Theory and Analysis*, 2021; 6(1); 15-26.

[20] Theil H.A. Rank invariant method of linear and polynomial regression analysis. *III. Nederl.Akad. Wetensch.Proc.,* 1950; 53; 1397-1412.

[21] Hodges J.L. and Lehmann E.L. Estimates of location based on rank tests. *Ann. Math. Statistics*, 1963; 34; 598-611.

[22] Lehmann E.L. and Dabrera H.J.M. (1975) Nonparametrics: Statistical Methods Based on Ranks. SF, USA.

[23] Tukey J.W. (1977) Exploratory Data Analysis, Reading, Addison-Wesley, USA.

[24] Minh V.T., Moezzi R., Dhoska K., Pumwa, J. Model Predictive for Autonomous Vehicle Tracking. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 2021; 4(1); 560-603.

[25] Durmuş B., İşçi Güneri Ö. Investigation of Factors Affecting Immunotherapy Treatment Results by Binary Logistic Regression and Classification Analysis. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 2020; 3(3); 467-473.

[26] Armstrong J.S. (1985) Long-range forecasting: From crystal ball to computer. John Wiley & Sons Inc., Canada.

[27] Graybill F. (1961) Introduction to linear statistical models, Volume 1. McGraw-Hill Book Company, New York, USA.

[28] Topal M. *A* Study on nonparametric regression methods. *Ph.D. Thesis*. Ataturk University, Graduate School of Natural and Applied Sciences / Department of Animal Science, Turkey, 1999.

[29] Spaeth H. (1991) Mathematical Algorithms for Linear Regression. London: Academic Press, UK.