REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 97–114
https://doi.org/10.57805/revstat.v21i1.398

Integrating Jackknife into the Theil–Sen Estimator in Multiple Linear Regression Model

Authors: Tolga Zaman ✉
– Department of Statistics, Faculty of Science, Çankırı Karatekin University, Çankırı 18100, Turkey
tolgazaman@karatekin.edu.tr

Kamil Alakuş
– Department of Statistics, Faculty of Science and Arts, Ondokuz Mayıs University, Samsun 55139, Turkey

Received: November 2020 Revised: October 2021 Accepted: October 2021

Abstract:

- In this study, we provide Theil–Sen parameter estimators, which are in multiple linear regression model based on a spatial median, to be examined by the jackknife method. To obtain the proposed estimator, apply the jackknife to a multivariate Theil–Sen estimator (MTSE) from Dang *et al.* estimators, who proved that the MTSE estimator is asymptotically normal. Robustness, efficiency, and non-normality of the proposed estimator is tested with simulation studies. As a result, the proposed estimator is shown to be robust, consistent, and more efficient in multiple linear regression models with arbitrary error distributions. Also, it is seen that the proposed estimator reduces the effects of outliers even more and gives more reliable results. So, it is clearly observed that the proposed estimator improves the outcome of the multivariate Theil–Sen estimator. In addition, we support with the aid of numerical examples to these results.

Keywords:

- *jackknife; robustness; efficiency; Theil–Sen estimator; multiple linear regression; spatial median.*

AMS Subject Classification:

- 62F40, 62G05, 62G35, 62H12.

✉ Corresponding author.

1. INTRODUCTION

Regression analysis; including the cause-result relationship examines the relationship between dependent (response) and independent (predictor) variables. Parametric regression analysis is based on certain assumptions. The most important of these assumptions, the mathematical form of the relationship between the dependent and independent variable is known in advance. The least squares method is proposed to be the most useful for solving such problems. The estimator of slope by this method is referred as the least squares estimator (LSE), which is the best linear estimator under the means of minimum variance if the variance of the error term is finite. However, LSE is vulnerable to gross errors and is also inefficient for distributions with heavy tails. In this case, in order to make better estimations, regression methods which allow the linearity assumption in the parametric regression to be stretched are needed. These methods are non-parametric regression models known as regression models. Nonparametric regression analysis, is the method that is successful for some of the assumptions used in case of failure in order to provide valid parametric regression methods. Several non-parametric methods were explored in the last century, such as the Theil–Sen estimator (TSE) [41]; [37], and various M-estimators [21]; [18]; [42].

We consider a multiple linear regression model,

$$Y = \beta X + \epsilon \tag{1.1}$$

where (X,Y) is observable but ϵ is not. β is an unknown parameter and ϵ has an unknown cdf, F. The mean of ϵ may not be zero. X and ϵ are independent. Let $(x_1, y_1), ..., (x_n, y_n)$ be independent random observations from the above model. In the literature, some researchers assumed that x_i s are random variables [3] and others assumed that the distribution of Y_i is

$$F_i(y) = F(y - \beta x_i), \qquad i = 1, 2, ..., n$$

where $x_i s$ $i = 1, 2, ..., n$ are known non-identical constant [37].

Since β is a slope parameter, under the assumption that all x_i s are distinct, Theil [41] proposed an estimator of β defined as

$$\hat{\beta} = med\Bigg[(s_{ij} \diagdown s_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad 1 \leq i \leq j = 1, ..., n)\Bigg] \tag{1.2}$$

where med stands for median. The estimator was referred to as Theil's estimator in literature. Theil's estimator was not defined if there exist ties among x_i s, and was extended by Sen [37] as

$$\hat{\beta}_n = med\Bigg[(s_{ij} \diagdown s_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad if x_i \neq x_j 1 \leq i \leq j = 1, ..., n)\Bigg]. \tag{1.3}$$

The new estimator was referred to as TSE.

Consider a multiple linear regression with p $\geq$ 1:

$$Y_j = \beta_0 + X_j^T \beta + \epsilon_j; \quad j = 1, 2, ..., n. \tag{1.4}$$

Following the above procedure, first, $\varphi = (\beta_0, \beta^T)^T$ can be found as the solution of equation (1.4)

$$Y_j - \beta_0 - X_j^T \beta = 0, \quad \boldsymbol{l}_{k+1} = \{j_1, ..., j_{k+1}\} \tag{1.5}$$

where, $\boldsymbol{l}_{k+1}$ is a $(k+1)$ subsets of $\{1, ..., n\}$. That is to say, if matrix $(k+1) \times (k+1)$ matrix $(X_l : l \in \boldsymbol{l}_{k+1})$, is invertible. This estimation is described with $\hat{\varphi}_{l_{k+1}}$ to emphasize dependency to $k+1$ observations. Later, natural expanding of TSE from a simple linear regression model to a multivariate regression model becomes multivariate median as following:

$$\hat{\varphi}_n = Mmed\{\hat{\varphi}_{l_{k+1}} : \forall \boldsymbol{l}_{k+1}\} \tag{1.6}$$

where, it should be pointed out that $\hat{\varphi}_{l_{k+1}}$ is at the same time the least squares estimator of φ based on $k+1$ observations $\{(X_j, Y_j) : j \in \boldsymbol{l}_{k+1}\}$. In this perspective, t different arbitrary combination of $\{(X_j, Y_j) : j \in \boldsymbol{l}_t\}$ can be chosen which means, here, $k+1 \leq t \leq n$ and it construct least squares estimator of $\hat{\varphi}_{\boldsymbol{l}_t}$. Then, multivariate TSE $\hat{\varphi}_n$ of the parameter φ is naturally will be multivariate median of all possible least square estimators and is described as below:

$$\hat{\varphi}_n = Mmed\{\hat{\varphi}_{l_t} : \forall \boldsymbol{l}_t\}. \tag{1.7}$$

Least squares estimation is as follows:

$$\hat{\varphi}_l = (X_{\boldsymbol{l}}^T X_{\boldsymbol{l}})^{-1} X_{\boldsymbol{l}}^T Y_{\boldsymbol{l}}. \tag{1.8}$$

In multivariate Theil–Sen estimator (MTSE), regression coefficients through the application of combination as $\binom{n}{t}$ is estimated with least squares method [10]. After every combination, spatial median belonging of obtained regression coefficient is computed. Regression coefficients belonging of MTSE are estimated by calculating as the spatial median obtained.

TSE is, with 0.293 breakdown point, robust and has a limited effect function and high asymptotic efficiency. For this reason, it is competes well with other slope estimators [37, 11, 44]. When we explore the literature for asymptotic characteristics of TSE, Sen [37] examined the asymptotic normality of estimation when cumulative probability function is continuous and showed that it is super-efficient for discrete error term. Even though most of its good characteristics are interpreted clearly and a lot of statisticians tried to expand on it [29, 50]. Since TSE is formulated only for a simple linear model, it is underdeveloped and rarely used. While for TSE to be expanded to a multiple linear model is obvious and attractive, this is technically hard and is a case which slows down the generalization and exploration processes. Oja and Niinimaa [29] generalized Theil–Sen estimation, which is in simple linear regression, to multiple linear regression using Oja [29] median. Oja's median is a special case of spatial median. These studies about Theil–Sen estimation are important for future studies. Peng *et al.* [32] established the asymptotic distribution and robust consistency of Theil–Sen estimation when cumulative probability function of the error term comes arbitrarily from both continuous and discrete distribution. Asymptotic distribution and robust consistency of Theil–Sen estimation can be examined as follows. In literature, there are various studies about TSE. See, e.g., regression estimation with Theil–Sen regression under the measurement errors, Fernandes and Leblanc [15]; inverse regression estimation with the help of Theil–Sen regression, Lavagnini *et al.* [23]; multivariate regional estimations using Theil–Sen estimator, Zhou and Serfling [50]; and asymptotic multiple linear regression estimation using Theil–Sen regression, Shen [38]. The Theil–Sen estimator has been widely acknowledged in several popular textbooks on nonparametric statistics and robust regression, see, e.g., Sprent [39], Hollander and Wolfe [19, 20], and Rousseeuw and Leroy [35], Wilcox [46]. It also has been extensively studied in the literature. Sen [37] and Wilcox [44] investigated its asymptotic relative efficiency to the least squares estimator. Akritas *et al.* [1] applied it to astronomy and Fernandes and Leblanc [15] to remote sensing. Wilcox [45] investigated some results on extensions and modifications of the Theil–Sen regression estimator. Wang [43] studied its asymptotic properties for model (1.1) with a random covariate. Wang [43] showed that TSE is strongly consistent, and obtain its asymptotic distribution, which may not be a normal distribution if F is not absolutely continuous. Many of its extensions can be found in the literature, for example, in censored data; for details, see, e.g., Akritas *et al.* [1], Jones [22], and Mount and Netanyahu [27]. Dang *et al.* [10] proposed the Theil–Sen estimators of parameters in a multiple linear regression model based on a multivariate median, generalizing the Theil–Sen estimator in a simple linear regression model. The sample mean of the bootstrap sample is known as the bagging estimator or smoothed bootstrap estimator. Empirically, bootstrapping with the bagging estimator often outperforms bootstrapping with the original estimator, especially when the asymptotic distribution is non-normal. See Breiman [5], Yang [48], and Efron [13]. See Büchlmann and Yu [6] and Friedman and Hall [16] for theory and references for the bagging estimator. Pelawa Watagoda and Olive [31] show that if $\sqrt{n}(T_n - \beta) \to N_p(0, \Sigma)$, then then under regularity conditions, $\sqrt{n}\left(\overline{T}^* - T_n\right) \to 0$, $\sqrt{n}(T_i^* - T_n) \to N_p(0, \Sigma)$ and $\sqrt{n}\left(\overline{T}^* - \beta\right) \to N_p(0, \Sigma)$. We are using a similar idea to bagging with the jackknife to produce the jackknife multivariate Theil–Sen estimator (JMTSE) estimator. In this paper, following Dang *et al.* [10], jackknife method which is one of the resampling methods is integrated in the multivariate Theil–Sen method (MTSE) and by doing this, a new estimator named jackknife multivariate Theil–Sen estimator (JMTSE) is offered. Robustness property of MTSE and is improved and became attractiveness for accomplishing well. In order to compare the proposed estimator with MTSE and LSE methods, various simulation studies are designed and results of multiple Theil–Sen estimation in multiple linear regression analysis are improved. Also, behaviors of these estimators are examined with two original data sets.

The remainder of the paper is organized as follows. In Section 2, we describe the properties of Theil–Sen estimator and spatial median. In Sections 3 and 4, we present jackknife method which is one of the resampling methods and we present some theoretical results of the jackknife method. In Section 5, we suggest a new estimator using jackknife method. In Section 6, we introduce the results for both simulations and real data set examples. In final section, we made conclusions about the obtained results.

2. STRONG CONSISTENCY AND ASYMPTOTIC NORMALITY PROPERTIES OF THEIL–SEN ESTIMATORS

In this section, it is stated the results on the strong consistency and the asymptotic distribution of TSE $\hat{\beta}_n(\varphi)$ under the assumption that X_is are random variables with $Var(X) > 0$.

Let ζ_0, $\left\{\varphi : \hat{\beta}_n(\varphi) = \beta\right\}$ (for all big n) be an event. That is to say, when $\varphi \in \zeta_0$ is $n > n_\varphi$ for each, $\hat{\beta}_n(\varphi) = \beta$ is so that there is a n_φ. The following theorem establishes an interesting property of the estimator

If F is continuous, then TSE $\hat{\beta}_n$is strongly consistent, that is $\hat{\beta}_n \to \beta$ [43].

In this section, we study the asymptotic distribution of the Theil–Sen estimator for both discontinuous and continuous error cdf F.

Firstly, F is assumed discontinuous. Then,

$$P\left(n^v\left(\hat{\beta}_n - \beta\right) \to 0\right) = 1, v \geq 0. \tag{2.1}$$

It has gotten the asymptotic behavior of $\hat{\beta}_n$ [43].

Now supposed that F is continuous. Denote the cdf of X_1 by G, the cdf of $X_1 - X_2$ by G_2 and the cdf of $\varepsilon_1 - \varepsilon_2$ by F_2. Then G_2 and F_2 are symmetric distribution function. Let

$$\mu(t) = \int [1 - 2F_2(xt)]dG_2(x) \tag{2.2}$$

and

$$\sigma^2 = \frac{1}{3}E\left[(1 - G(X_1) - G(X_1))^2\right]. \tag{2.3}$$

When G is continuous, $\sigma^2 = \frac{1}{9}$ ([43, 32]. Also, for further information about strong consistency and asymptotic distribution belonging of TSE estimation, please look into Wang [43]. Let's make statements about the spatial median.

Let W be a p-variate random vector with cdf F, $p > 1$. The spatial median (sm) of W minimizes the objective function:

$$D_F(w) = \frac{1}{n}\sum_{i=1}^{n}\{\|W_i - w\| - \|W_i\|\}; w \in R^d \tag{2.4}$$

where $\|.\|$ is the Euclidean form. Let $S(w) = w/\|w\|(w \neq 0)$ be the spatial sign function. The sample statistical spatial depth

$$D_F(w) = 1 - \left\|\frac{1}{n}\sum_{i=1}^{n} S(w - W_i)\right\|; w \in R^d. \tag{2.5}$$

The spatial median is the multivariate median defined by the spatial depth, which is any value that maximizes the sample depth,

$$\widehat{sm} = \arg\sup D_F(x); x \in R^d. \tag{2.6}$$

The estimate $\widehat{sm}$ is unique if the observations do not fall on a line. The spatial median has good efficiency properties. Möttönen *et al.* [28] for example calculated the asymptotic relative efficiencies. For the strong consistency and asymptotic normality of the spatial median (for see in detail information, Chaudhuri [9] and Bose [4].

3. JACKKNIFE METHOD

Jackknife method is defined as the method which minimizes sample error used to estimate population parameter. First definition of this method is made by Quenouille [33] and it is improved by Tukey using confidence interval approach [17]. Efron [12] contributed to the estimation of standard error and bias of the method. Martin and Roberts [25] proposed jackknife-after-bootstrap method which is developed in order to determine efficient observations. Jackknife method gives confidence intervals and decreases bias of estimation when known approaches are having hard time. Jackknife method is a resampling method in estimation of population parameters and developed in order to minimize sample error related to obtaining narrow confidence intervals. This method is considered as a statistical process that aims to reveal the relationship between variables in the data set in many fields that require parameter estimation [40]. Jackknife method does not consider the distribution belonging to variables in the data set and in this regard, it is known as a non-parametric statistical process. In parameter estimation process with this method, estimation is made by throwing out one observation in the sample each time. Thus, effect of deviated values is tried to be eliminate. The fundamental logic of the Jackknife method is to produce n different sample (sub-sample), each $(n-1)$ sized, by excluding each sample observation from the data set. The fundamental logic of the method bases on calculating sampling statistics from remaining observations through excluding an observation in data set. Thus, n different observations from n observations can be formed.

Let be $X = (x_1, x_2, ..., x_n)$ sample and $\hat{\theta} = s(X)$ our estimator. According to jackknife methods, when i. observation are excluded, new sample is defined as:

$$x_{(i)} = (x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n); i = 1, 2, ..., n. \tag{3.1}$$

Because of this, its estimator is also defined as:

$$\hat{\theta}_{(-i)} = s\big(x_{(i)}\big) \tag{3.2}$$

where $\hat{\theta}_{(.)}$ is the estimate of θ and calculated through the equation, $\hat{\theta}_{(.)} = \frac{\sum_{i=1}^{n} \hat{\theta}_{(-i)}}{n}$. Jackknife estimate of standard error is calculated as

$$\widehat{se}_{Jackk} = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{\theta}_{(-i)} - \hat{\theta}_{(.)}\right)^2}{n(n-1)}}. \tag{3.3}$$

Applying jackknife method in non-parametric regression can be thought as same logic. By excluding one observation from the current dependent and independent variables, non-parametric regression method is applied in current data. This process is repeated times of sample size.

The parameters of non-parametric regression methods using jackknife method are found as follows:

- Firstly, $(n-1)$ sized n different subsamples are formed by removing the observations from the data one by one.
- Regression coefficients belonging to non-parametric regression methods in interest are estimated for each formed subsample. This regression coefficients are called deleted slope coefficients and is indicated by $\beta_{i(-j)}, j = 1, 2, ..., n; i = 1, 2, ..., p$.
- Lastly, in order to obtain Jackknife estimator of intercept parameter value, mean of values $(y_j - \left(\hat{\beta}_{1(-j)}\right)x_1 + \left(\hat{\beta}_{2(-j)}\right)x_2 + ... + \left(\hat{\beta}_{p(-j)}\right)x_p)$ is estimated as estimate $\hat{\beta}_{0(-j)}$ as below:

$$\hat{\beta}_{0(-j)} = \frac{\sum_{j=1}^{n}\left(y_j - \left(\hat{\beta}_{1(-j)}\right)x_1 + \left(\hat{\beta}_{2(-j)}\right)x_2 + ... + \left(\hat{\beta}_{p(-j)}\right)x_p\right)}{n-1}. \tag{3.4}$$

- Mean value of these coefficients that obtained for each subsample are Jackknife estimators and expressed as below:

$$\hat{\beta}_i^{J*} = \frac{\sum_{j=1}^{n}\left(\hat{\beta}_{i(-j}\right)}{n}; i = 1, 2, ..., p; j = 1, 2, ..., n; \hat{\beta}_0^{J*} = \frac{\sum_{j=1}^{n}\hat{\beta}_{0(-j)}}{n}. \tag{3.5}$$

4. SOME PROPERTIES OF JACKKNIFE ESTIMATORS

In this section, the theoretical results of the jackknife estimation on unbiased, consistency, and asymptotic distribution are indicated.

For $j = 1, ..., n$ supposed that the point jackknife estimations of the parameters are $\hat{\boldsymbol{\beta}}^{J*}$ and determine the sampling distribution of these estimators. Here $\hat{\boldsymbol{\beta}}^{J*}$ is defined as in the following equation. Let us consider the $px1$ dimensional $\hat{\boldsymbol{\beta}}^{J*}$ jackknife estimator vector of $\boldsymbol{\beta}$ parameters:

$$\hat{\boldsymbol{\beta}}^{J*} = \frac{1}{n}\sum_{j=1}^{n}\hat{\boldsymbol{\beta}}_{(-j)}. \tag{4.1}$$

First, let's find the expected value of the estimator $\hat{\boldsymbol{\beta}}^{J*}$ in equation (4.1):

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}\sum_{j=1}^{n}E\left(\hat{\boldsymbol{\beta}}_{(-j)}\right), \tag{4.2}$$

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}E\left[\hat{\boldsymbol{\beta}}_{(-1)} + \hat{\boldsymbol{\beta}}_{(-2)} + \cdots + \hat{\boldsymbol{\beta}}_{(-n)}\right], \tag{4.3}$$

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}\left[E\left(\hat{\boldsymbol{\beta}}_{(-1)}\right) + E\left(\hat{\boldsymbol{\beta}}_{(-2)}\right) + \cdots + E\left(\hat{\boldsymbol{\beta}}_{(-n)}\right)\right], \tag{4.4}$$

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}[\boldsymbol{\beta} + \boldsymbol{\beta} + \cdots + \boldsymbol{\beta}] = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{\beta} = \frac{n\boldsymbol{\beta}}{n} = \boldsymbol{\beta}. \tag{4.5}$$

The result clearly shows that the estimate of $\hat{\boldsymbol{\beta}}^{J*}$ is unbiased for the parameter vector of $\boldsymbol{\beta}$ [7, 2].

Then, let's find the variance–covariance of the estimator $\hat{\boldsymbol{\beta}}^{J*}$ in equation (4.1). The jackknife variance–covariance estimator of $\hat{\boldsymbol{\beta}}^{J*}$ then can be written as follows:

$$
\begin{aligned}
(4.6)\qquad V(\hat{\boldsymbol{\beta}}^{J*}) &= S \\
&= \frac{1}{n}\sum_{j=1}^{n}(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*})'(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*}) \\
&= \frac{1}{n}\begin{bmatrix} \sum_{j=1}^{n}(\hat{\beta}_{1(-j)} - \hat{\beta}_1^{J*})^2 & \cdots & \sum_{j=1}^{n}(\hat{\beta}_{1(-j)} - \hat{\beta}_1^{J*})(\hat{\beta}_{p(-j)} - \hat{\beta}_p^{J*}) \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n}(\hat{\beta}_{1(-j)} - \hat{\beta}_1^{J*})(\hat{\beta}_{p(-j)} - \hat{\boldsymbol{\beta}}_p^{J*}) & \cdots & \sum_{j=1}^{n}(\hat{\beta}_{p(-j)} - \hat{\beta}_p^{J*})^2 \end{bmatrix}_{(\boldsymbol{p}\times\boldsymbol{p})} \\
&= \begin{bmatrix} V(\hat{\beta}_1^{J*}) & \cdots & \mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right) & \cdots & V\left(\hat{\beta}_p^{J*}\right) \end{bmatrix}.
\end{aligned}
$$

If the expected value of both sides is taken in equation (4.6), equation (4.7) is obtained.

$$
\begin{aligned}
(4.7)\qquad E\left[V\left(\hat{\boldsymbol{\beta}}^{J*}\right)\right] &= E(S) \\
&= \begin{bmatrix} E\left[V\left(\hat{\beta}_1^{J*}\right)\right] & \cdots & E\left[\mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right)\right] \\ \vdots & \ddots & \vdots \\ E\left[\mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right)\right] & \cdots & E\left[V\left(\hat{\beta}_p^{J*}\right)\right] \end{bmatrix}_{(\boldsymbol{p}\times\boldsymbol{p})} \\
&= \begin{bmatrix} V(\beta_1) & \cdots & \mathrm{Cov}(\beta_1, \beta_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(\beta_1, \beta_p) & \cdots & V(\beta_p) \end{bmatrix} = \boldsymbol{\Sigma}.
\end{aligned}
$$

As a result, the sampling distribution of $\hat{\boldsymbol{\beta}}^{J*}$ in jackknife estimator is obtained asymptotically $\hat{\boldsymbol{\beta}}^{J*} \sim N(\boldsymbol{\beta}; \boldsymbol{\Sigma})$.

The equation (4.7) clearly indicates that the estimate of $V\left(\hat{\boldsymbol{\beta}}^{J*}\right)$ is unbiased asymptotically for $\boldsymbol{\Sigma}$.

Finally, let's examine the consistency of the jackknife estimator of $\hat{\boldsymbol{\beta}}^{J*}$. Firstly the estimator has the variance $V\left(\hat{\boldsymbol{\beta}}^{J*}\right)$ and then we can write it as follows:

$$
(4.8)\qquad \lim_{n\to\infty} V(\hat{\boldsymbol{\beta}}^{J*}) = \lim_{n\to\infty} S = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*}\right)'\left(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*}\right)\right] \to 0.
$$

So, the equation (4.8) shows that the jackknife estimation $\hat{\boldsymbol{\beta}}^{J*}$ is consistent for the parameter vector $\boldsymbol{\beta}$. That is, $\hat{\boldsymbol{\beta}}^{J*} \to \boldsymbol{\beta}$ [36].

5. JACKKNIFE MULTIVARIATE THEIL–SEN ESTIMATOR (JMTSE) IN MULTIPLE LINEAR REGRESSION MODEL

JMTSE is a modification of the Theil–Sen estimator based on the jackknife method in multiple linear regression, a technique that narrows the confidence interval and reduces the effect of extreme values. In the analysis of proposed JMTSE method, subsamples, which are obtained through excluding each observation value sequentially from sample, are used instead of sample data. The algorithm steps for proposed JMTSE method is given below:

Step 1: Firstly, n different subsample, each $(n-1)$ sized, are formed by excluding each of n observation sequentially.

Step 2: Arbitrary number t is determined as $k+1 \leq t < n$. Here p indicates number of independent variable and n indicates sample size.

Step 3: Regression coefficients are calculated by applying LSE method to each possible $\binom{n-1}{t}$ combinations according to arbitrarily determined t value in order to estimate parameter estimation values, β_i, $(i = 1, 2, ..., p)$. If we express obtained regression coefficients as L_{ij}, regression parameter estimations are multivariate median of L_{ij} values. The multivariate median used here is spatial median. In other words, $\beta_{i(-j)} = Mmed(L_{ij})$ $(i = 1, 2, ..., p)$; $(j = 1, 2, ..., n)$. Since this process will be repeated for each subsample, n number of $\hat{\beta}_{i(-j)}$ will be calculated.

Step 4: To calculate $\hat{\beta}_{0(-j)}$ estimation, there are different alternatives. These are:

- **i)** If error term has a symmetric distribution around zero, $\hat{\beta}_{0(-j)}$ estimation is calculated by calculating each possible $\left(y_i - \hat{\beta}_{i(-j)}x_i\right)$. In other words, $\hat{\beta}_{0(-j)} = med\left(y_i - \hat{\beta}_{i(-j)}x_i\right)$.
- **ii)** $\hat{\beta}_{0(-j)}$ estimation can be calculated by averaging all possible $\left(y_i - \hat{\beta}_{i(-j)}x_i\right)$ values. That means $\hat{\beta}_{0(-j)} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{\beta}_{i(-j)}x_i\right)}{n-1}$.
- **iii)** Alternatively, (β_0, β_i) values can be estimated simultaneously with multivariate median in a less restricted situation. In other words, it can be calculated using $(\beta_0, \beta_i) = smed(\beta_{0(-j)}, \beta_{i(-j)})$.

Step 5: Means of all these coefficients for each sub-sample are proposed JMTSE estimations. In other words, $\hat{\beta}_0 = \frac{\sum_{j=1}^{n}\hat{\beta}_{0(-j)}}{n}$ ve $\hat{\beta}_i = \frac{\sum_{j=1}^{n}\hat{\beta}_{i(-j)}}{n}$.

There are certain things to consider while estimating with MTSE and proposed JMTSE methods. For example, if independent variables are categorical, then subsamples selected according to arbitrary t might be zero valued. In this case, estimation values with the LSE method can take values like (L_{ij}) $\pm\infty$. So, uncertainty may occur in this analysis. If L_{ij} values, which is calculated based on arbitrary t, is excluded from analysis, there would be data loss and this decreases the reliability of analysis. There are suggestions about this problem in simple linear regression models in Erilli and Alakuş [14]. If these suggestions are applied to JMTSE method:

1. β_i is calculated by replacing regression coefficients in (L_{ij}) value, which is found infinite, with maximum or minimum values in each possible L_{ij} values (in other words, $\max(L_{ij})$ instead of $+\infty$ and $\min(L_{ij})$ instead of $-\infty$) which is calculated according to arbitrary m.
2. The median of the data can be placed instead of regression coefficients in L_{ij}, which is found infinite, after infinite values are excluded from the data.
3. The trimmed mean values can be placed instead of regression coefficients in (L_{ij}), which is found infinite.

Using these methods, regression parameter estimations are calculated and regression coefficients for proposed JMTSE estimation method can be calculated considering the algorithm above. There is not a certain conclusion about which method gives the best result. It is advised to the researcher that all estimation results belonging to the all possible methods need to be obtained and compared according to data structure in order to obtain the best result.

In this article, the results of estimation of regression coefficients belonging to JMTSE method $(\beta_0; \beta_i, (i = 1, 2, ..., p))$, which is proposed in the simulation and real data applications, are obtained and interpreted simultaneously. In other words, $\left(\hat{\beta}_0, \hat{\beta}_i\right) = smed\left(\hat{\beta}_{0(-j)}, \hat{\beta}_{i(-j)}\right)$ [49].

6. RESULTS AND DISCUSSION

We conducted some numerical examples with Monte Carlo simulations and two real dataset examples. We measured the robustness and efficiency of the coefficients in the simulations. In real dataset examples we checked the prediction accuracies of the models and sparseness of the regression coefficients. The datasets are compatible for our aims which consist of heavy-tailed errors. All implementations were performed in R software [34].

6.1. Calculation and simulation studies

We performed Monte Carlo simulations study to evaluate the efficiency of proposed method. To compare the performance of proposed estimator, we employed two techniques ordinary least squares (LSE), multivariate Theil–Sen estimator (MTSE) [10]. Simulation design was constructed similar to Dang *et al.* [10].

In this section, various simulation studies are made with regards to robustness and efficiency in order to examine the behavior of proposed method. Some samples are produced from multiple linear regression model $Y_i = 2.5 + 3X_{1i} + 1.5X_{2i} + \varepsilon_i$, where, $X_{1i} \sim N(0, 1)$, $X_{2i} \sim U(0, 1)$ and ε_i are produced from different distributions with different purposes.

In this study, with the help of sub-samples obtained by excluding each observation values from the sample, we take a random sample of size t from the whole sample between

$k + 1 \leq t \leq n$ and calculate the LSE based this random sample. This process is repeated in such a way that it does not exceed the combination of $\binom{n}{t}$. Then, the spatial median of the obtained LSE estimators is calculated simultaneously. The mean of these coefficients calculated for each sub-sample are proposed JMTSE estimates. Breakdown point depends on the choice of t. The highest breakdown point is reached when t takes its minimal value $t = k + 1$. Therefore $t = 3$ was taken in this study.

Firstly, let us examine the robustness of proposed method. Sample sizes $n = 20, 40, 80$ are produced from multiple linear regression model $Y_i = 2.5 + 3X_{1i} + 1.5X_{2i} + \varepsilon_i$ with distribution $\varepsilon_i \sim N(0,1)$. Obtained data set is polluted with outliers (X_i, Y_i) of regression model $Y_i = -10 - 20X_{1i} - 25X_{2i} + \varepsilon_i$. Here, n_1 and n_2 are, respectively, count of good ones and count of bad ones (outliers). When we examine Table 1 that without outliers, all the LSE, MTSE and JMTSE performed well. However, with the presence of outliers, the LSE's method has completely deteriorated and become useless. While MTSE's and JMTSE's performed well until the ratio of outliers reaches 35–40%. But, it is seen that JMTSE's gives results closer to the real parameter value. Also, in the polluted data, LSE method have distant values to real regression coefficients. As pollution rate increased, it is observed that regression coefficients obtained by JMTSE is better compared to regression coefficients obtained by MTSE. In other words, it is shown that regression coefficients obtained through JMTSE method, which is proposed by hybridizing the resampling method, jackknife method, with MTSE estimation that is in literature, are closer to the real regression coefficients. Therefore, as a result of this simulation study, it is seen that proposed JMTSE method gives sufficient contribution to the literature.

Table 1: Robustness.

	True Parameter (2.5, 3, 1.5)		
	LSE	MTSE	JMTSE
$n = 20$	(2.480, 3.001, 1.521)	(2.488, 3.043, 1.523)	(2.469, 3.001, 1.509)
$n = 30$	(2.487, 3.004, 1.530)	(2.466, 3.023, 1.494)	(2.488, 3.017, 1.501)
$n = 40$	(1.894, 2.242, 1.084)	(2.792, 2.544, 0.417)	(2.226, 2.202, 0.875)
$n_1 = 19,\ n_2 = 1$	(2.264, 1.862, −0.436)	(2.420, 2.864, 1.550)	(2.462, 2.961, 1.482)
$n_1 = 18,\ n_2 = 2$	(0.855, 0.949, −0.219)	(2.180, 3.144, 1.370)	(2.563, 3.001, 1.221)
$n_1 = 16,\ n_2 = 4$	(−0.732, −2.403, −3.249)	(1.210, 2.363, −0.849)	(0.826, 1.183, 1.178)
$n_1 = 14,\ n_2 = 6$	(−0.935, −3.601, −6.696)	(0.912, 0.480, −2.213)	(1.633, 1.667, −1.413)
$n_1 = 12,\ n_2 = 8$	(−0.348, −0.736, −0.860)	(1.644, −0.225, 0.369)	(−0.187, −0.188, −0.394)

Robust estimator may lose efficiency. To investigate the efficiency, a simulation is conducted as follows. The values of mean square error belonging to LSE's, MTSE's and JMTSE's value $\hat{\beta}$ with sample size $n = 10, 20, 25, 30, 35, 40, 45, 50$ distribution $\varepsilon_i \sim N(0,1)$, and various outlier value ratios is calculated using $MSE = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\left(\hat{\beta}_i - \beta^{true}\right)^2$, where $\kappa = 1000$, $\beta^{true} = (2.5, 3, 1.5)$ and $\hat{\beta}_i$ is the estimate for i-th sample.

In Table 2, LSE method gave worse results compared to other methods when all cases of outlier ratios and sample sizes are considered. When sample size is fixed and outlier ratios are increased, experimental mean squared errors belonging to examined methods increased. For example, when sample size is taken as 20 and outlier ratios are 5%, 10%, and 20% respectively, experimental mean squared errors are 4.8272 when 5%; 15.7596 when 10% and 26.1113 when

20% respectively. Similarly, experimental mean squared errors belonging to MTSE method are 0.1373 when 5%, 0.2152 when 10% and 0.7689 when 20%. Finally, mean squared errors belonging to JMTSE method are 0.1381 when 5%, 0.148 when 10% and 0.7158 when 20%. That is to say, when examined methods are evaluated, it is seen that experimental mean squared errors belonging to estimation methods increases when sample size is fixed and outlier ratios are increased. When regression coefficients belonging to examined methods and outlier ratios are fixed, mean values for all sample sizes are calculated. In n=10,20,25,...,50 sample size and 5% pollution rate, MSE values belonging to estimations are calculated as 7.7903 for LSE, 0.6270 for MTSE and 0.3512 for JMTSE when outlier ratio is 5%; 13.3294 for LSE, 3.1737 for MTSE and 1.6983 for JMTSE when 10%; and 31.2867 for LSE, 5.1085 for MTSE and 4.4401 for JMTSE when 20%. If it is looked carefully, the in equation, $MSE_{JMTSE} < MSE_{MTSE} < MSE_{LSE}$ exists between MSE's belonging to estimation methods which are examined when outlier ratios are fixed. In this regard, when pollution rate is fixed and sample size is n=10,20,25,...,50, it is seen that JMTSE estimation method gives more efficient results. In a same way, when sample size is fixed and outlier ratio is increased, again, JMTSE estimation method gives results closer to real regression coefficients compared to LSE and MTSE estimation methods. In other words, it is seen that JMTSE method is more efficient in estimating real parameters in these cases. In summary, it is seen that results belonging to Theil–Sen method are improved with the addition of jackknife method. It is also observed that when the sample size increases, the values decrease in the mean square error of the methods.

Table 2: Efficiency comparisons.

Sample Size	Outlier ratio %								
	5%			10%			20%		
	LSE	MTSE	JMTSE	LSE	MTSE	JMTSE	LSE	MTSE	JMTSE
10	28.80	0.4422	0.5483	32.0605	0.5326	0.5677	66.5179	4.2187	5.2909
20	4.8272	0.1373	0.1381	15.7596	0.2152	0.1484	26.1113	0.7689	0.7158
25	8.145	3.85	1.754	12.0568	0.1455	0.1468	31.2967	1.2035	0.8902
30	4.8467	0.0697	0.0698	10.3119	4.5238	5.8034	27.5454	0.3196	0.3271
35	4.2601	0.0499	0.0499	12.5515	9.9076	0.3733	25.7719	9.1562	9.4418
40	3.2691	0.0509	0.0509	8.734	4.5144	1.8785	26.3822	6.2388	2.8074
45	4.3719	0.0477	0.0478	7.9876	5.4876	4.6048	22.2376	8.6498	7.3089
50	3.8026	0.3682	0.1505	7.1731	0.0631	0.0632	24.4306	10.3123	8.7384
Mean	7.7903	0.6270	0.3512	13.3294	3.1737	1.6983	31.2867	5.1085	4.4401

Considering multiple linear regression model $Y_i = 2.5 + 3X_{1i} + 1.5X_{2i} + \varepsilon_i$, for sample sizes $n = 20, 30, 40$ and $n = 50$ generate 1000 samples with errors from $\varepsilon_i \sim N(0,1)$ $\varepsilon_i \sim t(u)$ with two different degrees of freedoms (df) $u = 1, 3$. The prediction accuracies are evaluated with mean square error (MSE) as the following:

$$MSE = \frac{1}{\kappa} \sum_{i=1}^{\kappa} \left(\hat{\beta}_i - \beta^{true} \right)^2$$

where $\kappa = 1000$, $\beta^{true} = (2.5, 3, 1.5)$ and $\hat{\beta}_i$ is the estimate for i-th sample. As for relative efficiency (RE) of $\hat{\beta}$, it is obtained by dividing the MSE of the LSE by that of $\hat{\beta}$. In Tables 3 and 4, the values of MSE and RE are given.

The relative efficiencies and MSE values of the MTSE and JMTSE are computed with respect to LSE in Table 3. When we examine Table 3, under the Gaussian model, the finite sample RE values of MTSE and JMTS are about 12–58% and the 52–98% which are acceptable. However, it is found that RE values of the JMTSE is bigger than 1 for heavy tail distributions t with $df = 3$ and $df = 1$(Cauchy). Especially, under the Cauchy model, JMTSE is more efficient compared to LSE.

Table 3: MSE values and Relative efficiencies of the MTSE and JMTSE with respect to LSE for some continuous distributions.

		Normal			T3			T1 (Cauchy)		
		LSE	MTSE	JMTSE	LSE	MTSE	JMTSE	LSE	MTSE	JMTSE
$n=20$	MSE	0.072785	0.126266	0.10913	0.941811	0.533084	0.534346	802.2187	2.513192	2.364063
	RE	1	0.576442	0.666953	1	1.76672	1.76255	1	319.2031	339.339
$n=30$	MSE	0.047412	0.081497	0.069998	0.523894	0.372488	0.365735	169.8011	0.856768	0.802462
	RE	1	0.581769	0.677338	1	1.406473	1.432442	1	198.188	211.6002
$n=40$	MSE	0.035762	0.097724	0.068763	0.451233	0.52053	0.351836	163.7145	4.880104	3.658022
	RE	1	0.365944	0.520069	1	0.866873	1.28251	1	33.54733	44.75491
$n=50$	MSE	0.019061	0.156786	0.019393	0.319796	0.267127	0.254355	2.67E+07	3.51E−01	3.28E−01
	RE	1	0.121571	0.982851	1	1.197171	1.257285	1	76068376	81402439
Mean			0.411432	0.711803		1.309309	1.433697		19017232	20350759

The relative efficiencies and MSE values of the JMTSE are computed with respect to MTSE in Table 4. From the Table 4, it can be concluded that when the error comes from the heavily tail distributions t with $df = 3$, the JMTSE competes the MTSE, especially for Cauchy. That means, JMTSE is more efficient compared to MTSE.

Table 4: MSE values and Relative efficiencies of the JMTSE with respect to MTSE for some continuous distributions.

		T3		Cauchy	
		MTSE	JMTSE	MTSE	JMTSE
$n=20$	MSE	0.533084	0.534346	2.513192	2.364063
	RE	1	0.99764	1	1.063082
$n=30$	MSE	0.372488	0.365735	0.856768	0.802462
	RE	1	1.018464	1	1.067674
$n=40$	MSE	0.52053	0.351836	4.880104	3.658022
	RE	1	1.479468	1	1.334083
$n=50$	MSE	0.267127	0.254355	3.51E−01	3.28E−01
	RE	1	1.050214	1	1.070122
Mean			1.136446		1.13374

As a result, just like what simulation studies showed, it is found that proposed JMTSE method has more consistency than LSE and MTSE methods with regard to efficiency and robustness and results of MTSE method are improved.

6.2. Numerical illustrations

In this part we conducted some experiments on real datasets to evaluate the predictive performance of estimators. Meanwhile we presented the sparsity of the regression coefficients. For the application we used Coleman and Education expenditure data sets which are available in R software, ”MASS” and ”robustbase” packages [47, 24]. These datasets contain heavy-tailed errors so the real datasets are conformable for the computations. Coleman dataset contains 20 observations and 5 independent variables. This data set contains information on 20 schools from the Mid-Atlantic and New England states. The purpose is to predict the verbal mean test score [26]. Education expenditure data set consists of 50 observations and 3 independent variables. This data set is related with the education expenses of 50 states in the US. The aim is to predict the per capita expenditure for public education [8].

Predictive performance is measured by cross validation technique. The datasets are divided in two parts as test-train. Train sets contain 80% and test sets contain 20% of the datasets, respectively.

95% confidence intervals of the regression coefficients are given for LSE and JMTSE methods in Tables 5 and 6. From the Table 5, It is seen that independent variables, x_3, x_4 and x_5, are found significant for the model obtained through LSE method and all independent variables are found significant for the model obtained through proposed JMTSE method.

Table 5: Confidence intervals of the regression coefficients for the Coleman data set.

Coefficients	LSE		JMTSE	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Constant	16.743	59.624	35.175	37.619
x_1	−3.013	1.242	−0.675	−0.377
x_2	−0.004	0.159	0.058	0.086
x_3	0.540	0.835	0.654	0.684
x_4	0.323	1.619	0.881	0.975
x_5	−8.221	−1.629	−4.723	−4.342

In Table 6, independent variable x_2 is found significant for the model obtained through LSE method and independent variables, x_2 and x_3, are found significant for the model obtained through JMTSE method. As a result narrower confidence limits were estimated with the proposed estimator.

Table 6: Confidence intervals of the regression coefficients for the Education expenditure data set.

Coefficients	LSE		JMTSE	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Constant	−589.951	77.250	−346.571	−329.127
x_1	−0.113	0.134	−0.024	0.008
x_2	0.034	0.088	0.066	0.073
x_3	−0.142	1.625	0.879	0.954

The predictive performance of each approaches are given in Table 7. In experimental results, JMTSE performs better than LSE and MTSE in terms of prediction for new observations. It should be noted that the JMTSE proposed by using jackknife in MTSE is good for both data sets.

Table 7: Predictive performance results.

Methods	Coleman-MSE	Education E-MSE
LSE	16.423	4293.951
MTSE	15.209	3645.489
JMTSE	14.724	3564.373

7. CONCLUSION

Theil–Sen estimator is a point estimator of the slope parameter in the model and has many nice properties, including asymptotic normality. It has become a useful alternative solution for robust regression modelling with a high breakdown point and asymptotic efficiency. Although TSE has these many good properties, there are not many researches for Theil–Sen estimator in multiple linear regression methods. Jackknife method throwing an observation at a time from the sample which statistics calculates that as the number of individuals in the sample and the effect of extreme values can be defined as a method with relieving properties. The paper proposes a modification of the Theil–Sen estimator based on jackknife method. Simulation studies and real data applications is made in order to improve the results belonging to MTSE. Also, Jackknife estimations of parameters for Theil–Sen regression analysis, hypothesis test of parameters based on jackknife estimations and confidence intervals are examined.

According to simulations studies, robustness of methods investigated in the first phase of the simulation study is explored. When outliers do not exist in the data and distribution of error term is normal distribution, all of the LSE, MTSE and JMTSE estimation methods obtained results close to the real regression coefficients as a result of comparison with regard to robustness. Again, when the errors obtained from simulation study is normally distributed and the data is polluted with different ratios, LSE method got far away from real regression coefficients immediately and become unusable. However MTSE and JMTSE methods were able to endure until a certain point. It is seen that this ratio is around 35—40% for MTSE and JMTSE. Also JTSE is more capable of determining the real regression coefficients correctly. These findings exhibit the superiority of jackknife within MTSE in terms of variable selection. In the second phase of the simulation study, mean square error values are calculated for the methods which are investigated when errors are distributed normally but data has pollution with various ratios. As a result, it is clear that the proposed JMTSE method has a smaller mean square error than LSE and MTSE methods. In the third and fourth phase of the simulation study, no pollution is added to produced data. But when the error comes from the heavily tail distributions t with $df = 3$ and $df = 1$, it is found that JMTSE method, which is proposed with regards to sample sizes and arbitrary error term distributions, is more efficient than LSE and MTSE methods. Immediately after the simulation study, situations of methods in interest are investigated considering two original datasets. According to estimation results, it is found that JMTSE method gives more efficient results than MTSE and LSE methods. As a result, robustness and effectiveness of MTSE is improved using jackknife method.

It is clearly seen that the proposed JMTSE method works well when n and t are small. This shows that the proposed JMTSE method is computationally feasible and the method is to have useful outlier resistance.

Consequently, a new estimator named JMTSE is proposed by integrating Jackknife method to Theil–Sen method in multiple linear regression. It is observed that proposed method reduces the effects of outliers even more and gives more reliable results. According to obtained results, resampling methods like Jackknife method can be applied in non-parametric regression methods successfully. Moreover, we demonstrated the applicability of jackknife with MTSE and concluded the success with several numerical examples. We suggest using JMTSE when there are many predictors for further practical studies to accomplish model selection in the presence of heavy-tailed errors.

ACKNOWLEDGMENTS

The authors would like to thank the TÜBİTAK (The Scientific and Technological Research Council of Turkey) which financially supporting this study.

REFERENCES

[1] Akritas, M.; Murphy, S. and LaValley, M. (1995). The Theil–Sen estimator with doubly censored data and applications to astronomy, *J. Amer. Statist. Assoc.*, **90**(429), 170–177.

[2] Bhushan, S. and Misra, P.K. (2017). On an unbiased jack-knife regression type estimator, *Advances in Fuzzy Mathematics*, **12**(1), 147–155.

[3] Bickel, P.J. (1982). On adaptive estimation, *Annals of Statistics*, **10**, 647–671.

[4] Bose, A. (1998). Bahadur representation of M_m estimates, *Ann. Stat.*, **26**, 771–777.

[5] Breiman, L. (1996). Bagging predictors, *Machine Learning*, **24**, 123–140.

[6] Büchlmann, P. and Yu, B. (2002). Analyzing bagging, *The Annals of Statistics*, **30**, 927–961.

[7] Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press.

[8] Chatterje, S. and Price, B. (1977). *Regression Analysis by Example*, John Wiley and Sons, New York.

[9] Chaudhuri, P. (1996). Multivariate location estimation using extension of R estimates through U-statistics type approach, *Ann. Statist.*, **20**, 897–916.

[10] Dang, X.; Peng, H.; Wang, X. and Zhang, H. (2008). *Theil–Sen Estimators in a Multiple Linear Regression Model*, Olemiss.edu, The University of Mississippi.

[11] Dietz, E.J. (1987). A comparison of robust estimators in simple linear regression, *Comm. Statist., Part B Simulation Comput.*, **16**, 1209–1227.

[12] Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans* (Vol. 38), SIAM.

[13] Efron, B. (2014). Estimation and accuracy after model selection, (with discussion), *Journal of the American Statistical Association*, **109**, 991–1007.

[14] Erilli, N.A. and Alakus, K. (2016). Parameter estimation in Theil–Sen regression analysis with jackknife method, *Eurasian Academy of Sciences Eurasian Econometrics, Statistics and Empirical Economics Journal*, **5**, 28–41.

[15] Fernandes R. and Leblanc, S.G. (2005). 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*, **95**(3), 303–316.

[16] Friedman, J.H. and Hall, P. (2007). On bagging and nonlinear estimation, *Journal of Statistical Planning and Inference*, **137**(3), 669–683.

[17] Gray, H.L. and Schucany, W.R. (1972). *The Generalized Jackknife Statistic* (Vol. 1), M. Dekker.

[18] Hampel, F.R. (1971). A general qualitative definition of robustness, *The Annals of Mathematical Statistics*, 1887–1896.

[19] Hollander, M. and Wolfe, D.A. (1973). *Nonparametric Statistical Methods*, first ed., Wiley, New York, 205–206.

[20] Hollander, M. and Wolfe, D.A. (1999). *Nonparametric Statistical Methods*, second ed., Wiley, New York, 421–423.

[21] Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, 799–821.

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

[23] Lavagnini, I.; Badocco, D.; Pastore, P. and Magno, F. (2011). Theil–Sen nonparametric regression technique on univariate calibration, *Inverse Regression and Detection Limits Talanta*, **87**, 180–188.

[24] Maechler, M.; Rousseeuw, P.; Croux, C.; Todorov, V.; Ruckstuhl, A.; Salibian-Barrera, M.; Verbeke, T.; Koller, M.; Conceição, E.L.T. and Palma, M.A. (2016). *robustbase: Basic Robust Statistics R package version* 0.92–7. http://CRAN.R-project.org/package=robustbase.

[25] Martin, M.A. and Roberts, S. (2010). Jackknife-after-bootstrap regression influence diagnostics, *Journal of Nonparametric Statistics*, **22**(2), 257–269.

[26] Mosteller, F. and Tukey, J.W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, MA.

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

[28] Möttönen, J.; Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests, *Annals of Statistics*, **25**, 542–552.

[29] Oja, H. (1983). Descriptive statistics for multivariate distributions, *Statist. Probab. Lett.*, **1**, 327–332.

[30] Oja, H. and Niinimaa, A. (1984). *On robust estimation of regression coefficients*, Research Report, Department of Applied Mathematics and Statistics, University of Oulu, Finland.

[31] Pelawa Watagoda, L.C.R. and Olive, D.J. (2019). *Bootstrapping Multiple Linear Regression After Variable Selection*, online at http://lagrange.math.siu.edu/Olive/ppboottest.pdf.

[32] Peng, H.; Wang, S. and Wang, X. (2008). Consistency and asymptotic distribution of the Theil–Sen estimator, *Journal of Statistical Planning and Inference*, **138**(6), 1836–1850.

[33] Quenouille, M.H. (1949). *Approximate tests of correlation in time-series 3*. In: "Mathematical Proceedings of the Cambridge Philosophical Society", **45**(3), 483–484, Cambridge University Press.

[34] R Core Team (2017). R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

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

[36] Shao, J. and Wu, C.J. (1989). A general theory for jackknife variance estimation, *The Annals of Statistics*, 1176–1197.

[37] Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau, *J. Amer. Statist. Assoc.*, **63**, 1379–1389.

[38] Shen, G. (2009). Asymptotics of a Theil–Sen-type estimate in multiple linear regression, *Statistics & Probability Letters*, **79**(8), 1053–1064.

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

[40] Swingler, D.N. (1995). On the application of the jackknife to the estimation of the parameters of short multisinusoidal signals using the data-matrix formulation, *IEEE Transactions on Signal Processing*, **43**(9), 2135–2143.

[41] Theil, H. (1950). *A rank invariant method of linear and polynomial regression analysis, III.* In: "Proc. of Koninklijke Nederl. Akad. Wetensch., Series A", **53**, 1397–1412.

[42] Tukey, J.W. (1977). *Exploratory Data Analysis*, MA, Addison-Wesley.

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

[44] Wilcox, R.R. (1998). Simulations on the Theil–Sen regression estimator with right-censored data, *Statist. Probab. Lett.*, **39**, 43–47.

[45] Wilcox, R.R. (2004). Some results on extensions and modifications of the Theil–Sen regression estimator, *British Journal of Mathematical and Statistical Psychology*, **57l**, 265–280.

[46] Wilcox, R.R. (2017). *Introduction to Robust Estimation and Hypothesis Testing*, 4th Edition, San Diego, CA, Academic Press.

[47] Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*, Fourth Edition, Springer, New York, ISBN 0-387-95457-0.

[48] Yang, Y. (2003). Regression with multiple candidate models: selecting or mixing?, *Statistica Sinica*, **13**, 783–809.

[49] Zaman, T. (2017). *Contributions to the jackknife estimation and test problems in multiple linear regression analysis*, Doctoral Thesis, Ondokuz Mayıs University, Graduate School of Sciences, 468099.

[50] Zhou, W. and Serfling, R. (2008). Multivariate spatial U-quantiles: a Bahadur–Kiefer representation, a Theil–Sen estimator for multiple regression, and a robust dispersion estimator, *Journal of Statistical Planning and Inference*, **138**(6), 1660–1678.