J. Hlaváčová (Ed.): ITAT 2017 Proceedings, pp. 78–85
CEUR Workshop Proceedings Vol. 1885, ISSN 1613-0073, © 2017 J. Kalina, B. Peštová

Nonparametric Bootstrap Estimation for Implicitly Weighted Robust Regression

Jan Kalina, and Barbora Peštová

Institute of Computer Science CAS, Prague, Czech Republic
kalina@cs.cas.cz

Abstract: Implicitly weighted robust regression estimators for linear and nonlinear regression models include linear and nonlinear versions of the least trimmed squares and least weighted squares. After recalling known facts about these estimators, a nonparametric bootstrap procedure is proposed in this paper for estimates of their variances. These bootstrap estimates are elaborated for both the linear and nonlinear model. Practical contributions include several examples investigating the performance of the nonlinear least weighted squares estimator and comparing it with the classical least squares also by means of the variance estimates. Another theoretical novelty is a proposal of a two-stage version of the nonlinear least weighted squares estimator with adaptive (data-dependent) weights.

1 Introduction

Regression methodology has the aim to model (describe, estimate) a continuous variable (response) depending on one or more independent variables (regressors), which may be continuous and/or discrete. Such modelling finds applications in an enormously wide spectrum of applications and allows to predict values of an important variable, which is considered to play the role of the response, for particular values of regressors.

Standard estimation methods in various linear and nonlinear regression models are known to be too vulnerable (sensitive) to the presence of outlying measurements (outliers), which occur in real data for various reasons, e.g. measurement errors, different conditions or violations of assumptions of the model under consideration.

Therefore, numerous robust regression methods have been proposed since the development of robust statistical estimation in 1960s as diagnostic tools for classical methods. Some of them can be understood as reliable self-standing procedures tailor-made to suppress the effect of data contamination by various kinds of outliers [20, 6, 2, 13]. In the course of time, the breakdown point has become one of crucial measures of robustness, which can be interpreted as a high resistance (insensitivity) against outlying measurements in the data and one of crucial measures of robustness of statistical estimators. The finite-sample breakdown point is defined as the minimal fraction of data that can drive an estimator beyond all bounds when set to arbitrary values [13].

While M-estimators represent the most widely used robust statistical methods, they have been criticized for their low breakdown point in linear regression. Thus, other methods with a high value of the breakdown point (asymptotically up to $1/2$) are desirable, which are commonly denoted as highly robust.

This paper is devoted to the question of estimating the variance of highly robust implicitly weighted estimators in linear and nonlinear models, which has not been investigated in literature. After recalling the least weighted squares estimator in Section 2, a bootstrap estimate of its variance is described and illustrated in Section 3. The methodology is further generalized to nonlinear regression in Section 4, where also a two-stage version of the nonlinear least weighted squares estimator is proposed and investigated. Three examples comparing the performance of standard and robust methods in the nonlinear model are presented in a separate Section 5 and conclusions are summarized in Section 6.

2 Least Weighted Squares

This section recalls the least weighted squares estimator, which is one of promising tools estimating parameters of the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \quad i = 1, \ldots, n. \quad (1)$$

Here, $Y = (Y_1, \ldots, Y_n)^T$ denotes a continuous response, $(X_{1j}, \ldots, X_{nj})^T$ is the j-th regressor for $j = 1, \ldots, p$ and $e_1, \ldots, e_n$ are random errors of the model. The least squares estimator b_{LS} of $\beta = (\beta_0, \ldots, \beta_p)^T$ is well known to be highly vulnerable to the presence of outlying measurements in the data [13]. Therefore, numerous robust regression methods have been proposed as alternatives to the least squares.

The least weighted squares (LWS) estimator of β in the model (1) represents one of available robust estimators, which was proposed in [25]. It is based on implicit weighting of individual observations, down-weighting less reliable observations, which might be potential outliers with a high influence on the results. Thus, if suitable weights are used, it may reach a high breakdown point [12].

The weights are assigned to individual observations after an (implicitly given) permutation, which is determined only during the computation of the estimator. For a given estimate $b = (b_0, b_1, \ldots, b_p)^T \in \mathbb{R}^p$ of β, let a residual be defined as

$$u_i(b) = Y_i - b_0 - b_1 X_{i1} - \cdots - b_p X_{ip}, \quad i = 1, \ldots, n. \quad (2)$$

Squared residuals will be considered in ascending order

$$u^2_{(1)}(b) \le u^2_{(2)}(b) \le \cdots \le u^2_{(n)}(b). \tag{3}$$

The LWS estimator of β denoted as b_{LWS} is defined as

$$\arg\min \sum_{i=1}^{n} w_i u^2_{(i)}(b) \tag{4}$$

over $b \in \mathbb{R}^p$ for specified (given) magnitudes of nonnegative weights $w_1, w_2, \ldots, w_n$. We need to note that the knowledge of these magnitudes is crucial for the estimator to be properly defined (i.e. avoiding a circular definition) as (3) are fixed as well allowing to assign the weights directly to be non-increasing with respect to residuals (cf. [25, 24, 23]). An adaptation of the approximate algorithm of [20] may be used for computing the LWS algorithm.

Concerning the choice of weights, a general recommendation can be given to require the sequence $w_1, \ldots, w_n$ non-increasing with $\sum_{i=1}^n w_i = 1$. Some choices (from the simplest to the most complicated) include

- Linearly decreasing weights

$$w_i^{LD} = \frac{2(n-i+1)}{n(n+1)}, \quad i = 1, \ldots, n. \tag{5}$$

- Linearly decreasing weights for a true level of contamination $\varepsilon \cdot 100$ %. Let us assume $\varepsilon \in [0, 1/2)$ and $h = \lceil \varepsilon n \rceil$, where

$$\lceil x \rceil = \min\{n \in \mathbb{N};\ n \ge x\}. \tag{6}$$

  The weights equal

$$w_i = \begin{cases} (h-i+1)/h, & i \le h, \\ 0, & i > h. \end{cases} \tag{7}$$

- Data-dependent adaptive weights of [6]. With such weights, the LWS estimator attains a 100 % asymptotic efficiency of the least squares under Gaussian errors.

Weights according to (7) as well as the adaptive weights of [6] ensure a high breakdown point of the LWS estimator, which cannot be said about weights (5).

The least trimmed squares (LTS) estimator (e.g. [18]) denoted as b_{LTS} represents a special case of the LWS with weights equal either to zero or one. The LTS estimator depends on the value of the trimming constant h, requiring

$$w_{h+1} = \cdots = w_n = 0 \quad \text{and} \quad n/2 < h < n. \tag{8}$$

The advantages of the LWS compared to the LTS include sub-sample robustness, more delicate approach for dealing with moderately outlying values, robustness to heteroscedasticity [23], the possibility to derive diagnostic tools and to define a corresponding robust correlation coefficient or estimator of σ^2 [16, 17].

3 Nonparametric Bootstrap for the LWS

In this section, a new estimate of the variance for the least trimmed squares and least weighted squares estimators is proposed in Section 3.1. The proposal enables a systematic comparison of their estimation performances. An illustration on a real data set is presented in Section 3.2.

3.1 Bootstrap Variance Estimation for the LWS Estimator

The aim of this section is to apply resampling (bootstrap) techniques to estimate $\mathsf{var}\, b_{LTS}$ and $\mathsf{var}\, b_{LWS}$. The resulting bootstrap estimates are conceptually simple and can be computed for real data in a straightforward (although rather computationally demanding) way.

Let us first recall basic principles of bootstrap estimation (bootstrapping), which has found a big popularity in various statistical tasks. In general, bootstrap estimation exploits resampling with replacement. Incorporating the basic principles of bootstrapping, one may develop a great variety of resampling techniques that provide us with new possibilities of analyzing data. The range of bootstrap methods is rather large, including residual bootstrap, nonparametric bootstrap, semiparametric bootstrap, Bayesian bootstrap etc. Also the terminology is not used in a unique way. Unfortunately, not much can be said about properties of bootstrap estimates on a general level. Interesting comparisons of bootstrap procedures in a regression setup (but for the total least squares) were presented in [19], where some bootstrap approaches are valid but some are proven not to be consistent and thus not suitable.

In the specific task of estimating variability of regression estimators, the bootstrapping approach is very suitable. While the seminal work [3] described bootstrap estimation from a philosophical perspective, practical approaches to bootstrapping in linear regression were proposed by subsequent papers [9] or [8]. Other theoretical results were derived in [10, 11].

In this paper, we focus our attention to nonparametric bootstrap. However, a residual bootstrap may be a suitable alternative as well.

Under (1), we recall that

$$\mathsf{var}\, b_{LS} = \sigma^2 (X^T X)^{-1}. \tag{9}$$

An explicit formula for $\mathsf{var}\, b_{LWS}$ could be derived as analogy to [24]. Such result however remains impossible to be directly computed for real data, as it depends on

- Unknown magnitudes of $e_1, \ldots, e_n$,
- The asymptotic value of $(X^T X)/n$, which can be however hardly evaluated for a fixed sample size.

Such approach is also complicated because of the necessity to express the weights by means of a weight function and there is also a more restrictive assumption of normally distributed errors. Therefore, we take resort to a bootstrap estimate of $\mathrm{var}\, b_{LWS}$. Its computation is described by Algorithm 1, where the final estimate has the form of a bootstrap covariance matrix in step 5. An analogous procedure can be used for estimating $\mathrm{var}\, b_{LTS}$.

Algorithm 1 Nonparametric bootstrap for the LWS in linear regression.

Input: Data rows $(X_{i1},\ldots,X_{ip},Y_i)$, $i=1,\ldots,n$
Output: Empirical covariance matrix computed from individual estimates of $\hat{\gamma}_{LWS}$

1: Compute the least weighted squares estimator $\hat{\beta}_{LWS}$ of β in model (1)
2: **for** $r = 1$ to R **do** // repeat in order to obtain the empirical distribution
3: Generate n new bootstrap data rows

$$\left({}_{(r)}X^*_{j1},\ldots,{}_{(r)}X^*_{jp},{}_{(r)}Y^*_j\right), \quad j=1,\ldots,n, \tag{10}$$

by sampling with replacement from the original set of data rows $(X_{i1},\ldots,X_{ip},Y_i)$, $i=1,\ldots,n$
4: Consider a linear regression model in the form

$${}_{(r)}Y^*_j = {}_{(r)}\gamma_0 + {}_{(r)}\gamma_1\,{}_{(r)}X^*_{j1} + \cdots + {}_{(r)}\gamma_p\,{}_{(r)}X^*_{jp} + {}_{(r)}v_j \tag{11}$$

with $j=1,\ldots,n$ and random errors $v_1,\ldots,v_n$
5: Estimate ${}_{(r)}\gamma = \left({}_{(r)}\gamma_0,{}_{(r)}\gamma_1,\ldots,{}_{(r)}\gamma_p\right)^T$ in (11) by the LWS
6: Store the estimate from the previous step as ${}_{(r)}\hat{\gamma}_{LWS}$
7: **end for**
8: Compute the empirical covariance matrix from values ${}_{(r)}\hat{\gamma}_{LWS}$, $r=1,\ldots,R$

3.2 Example: Linear Model for Investment Data

The aim of the following example is to compare variance estimates of the LWS estimator with variances of other regression estimators. An investment data set that considers a regression of $n = 22$ yearly values of real gross private domestic investments in the USA in 10^9 USD against the GDP is used. We consider a linear model

$$Y_i = \beta_0 + \beta_1 X_i + e_i, \quad i=1,\ldots,n, \tag{12}$$

while the same data set was previously analyzed (considering another model) in [17].

The computations were performed in R software for the least squares, Huber's M-estimator (see [13]), LTS and LWS. Table 1 presents estimates of the intercept b_0 and slope b_1 together with packages of R software, which were used for the computation. The bootstrap procedure of Section 3.1 was used to find the covariance matrix of various robust regression estimators and the results are also presented in Table 1. There, the standard deviation of all estimates is denoted as s_0 for the intercept and s_1 for the slope.

For the least squares, the bootstrap estimates are very close to the exact result (9). The number of bootstrap repetitions within Algorithm 1 is chosen as 10000, which is

Estimator	b_0 (s_0)	b_1 (s_1)	R package
Least squares	−582 (108.9)	0.239 (0.016)	`base`
Huber's M-estimator	−576 (135.0)	0.238 (0.020)	`MASS`
LTS ($h=13$)	−252 (742.0)	0.185 (0.106)	`robustbase`
LWS (weights [6])	−465 (207.2)	0.221 (0.031)	own code

Table 1: Results of the example with investment data of Section 3.2. The classical and robust estimates of parameters β_0 and β_1 (without brackets) are accompanied by nonparametric bootstrap estimates of their standard deviation (underneath in brackets) denoted as s_0 and s_1, which were evaluated by the bootstrap procedure of Section 3.1.

sufficient for the asymptotics. Actually, we compared results obtained for 100 bootstrap samples and the results were very close. This is in accordance with our experience and a small number of bootstrap samples indeed seems to be sufficient if a single constant is estimated rather than the whole empirical distribution.

The smallest variance is obtained with the least squares estimator. Huber's M-estimator attains only a slightly higher variance. The loss of the LTS is remarkable. However, we point out at the closeness of the LWS result (computed with weights [6]) to the least squares or Huber's M-estimator compared to the very crude LTS result. Both the LTS and LWS are highly robust, but their performance for this data set without severe outliers reveals a great difference between them in terms of efficiency. The LWS cannot be the winner and must stay behind the least squares, but its retardation is only mild and its superiority against the LTS shows that the LWS (perceived as a generalization of the LTS) eliminates the main disadvantage of the LTS, namely its low efficiency for non-contaminated samples [7].

To the best of our knowledge, the superiority of the LWS compared to the LTS in terms of efficiency has never been presented in the literature. Our result indicates a possibly strong argument in favor of the efficiency of the LWS, at least for a single data set, while no theoretical result on the relative efficiency of LWS compared to LWS is available. Our result is empirical, which is obtained for a rather simplistic data set with only a single regressor, was obtained as an application of the nonparametric bootstrap estimation of the robust regression estimates proposed in Section 3.1.

4 Nonparametric Bootstrap in Robust Nonlinear Regression

In this section, the standard nonlinear regression model is recalled in Section 4.1 and the nonlinear least weighted squares estimator in Section 4.2. A two-stage version of the nonlinear least weighted squares is proposed in Section 4.3 as an extension of the methodology of Section 3.1. In addition, a two-stage version of the nonlinear least weighted squares estimator is proposed and theoretically investigated.

4.1 Nonlinear Regression Model

Let us consider the nonlinear regression model

$$Y_i = f(\beta_1 X_{i1} + \cdots + \beta_p X_{ip}) + e_i, \quad i = 1, \ldots, n, \tag{13}$$

where f is a given continuous nonlinear function, $Y = (Y_1, \ldots, Y_n)^T$ is a continuous response and $(X_{1j}, \ldots, X_{nj})^T$ is the j-th regressor for $j = 1, \ldots, p$. By $e_1, \ldots, e_n$ we denote the model's random errors. The classical estimator, which is the nonlinear least squares (NLS) estimator of β, is vulnerable to the presence of outliers in the data.

The nonlinear least trimmed squares (NLTS) estimator represents a natural extension of the LTS estimator to the nonlinear model (13). The breakdown point of the NLTS was derived already in [22], other properties were later investigated in [5]. The estimator may achieve a high robustness, if of course a suitable value of h is used reflecting the true contamination level in the data. An approximate algorithm may be obtained as an extension of the algorithm of [20]; however, it requires a tedious implementation and we are not aware of any implementation of NLTS in statistical software.

4.2 Nonlinear Least Weighted Squares

This section recalls the definition of the nonlinear least weighted squares (NLWS) estimator, which was proposed as our extension of the LWS estimator from the linear regression to the nonlinear model [14] and at the same time a weighted analogy of the NLTS estimator. So far, theoretical properties of the NLWS have not been derived and there is also no evidence in examples showing the robustness and efficiency of the estimator.

In the model (13), let

$$u_i(b) = Y_i - f(b_1 X_{i1} - \cdots - b_p X_{ip}), \quad i = 1, \ldots, n, \tag{14}$$

denote a residual corresponding to the i-th observation for a given estimator $b = (b_1, \ldots, b_p)^T \in \mathbb{R}^p$ of regression parameters $\beta = (\beta_1, \ldots, \beta_p)^T$. Let us now assume the magnitudes $w_1, w_2, \ldots, w_n$ of nonnegative weights to be given. The NLWS estimator of the parameters in the model (1) is defined as

$$\arg\min \sum_{i=1}^{n} w_i u_{(i)}^2(b), \tag{15}$$

where the argument of the minimum is computed over all possible values of $b = (b_1, \ldots, b_p)^T$ and the residuals are arranged as in (3).

The choice of weights has a determining influence on properties of the NLWS estimator. If it is allowed to have zero weights for the most outlying observations, then the estimator can be conjectured to be highly robust, which follows directly from the assignment of implicit weights to the observations in (15). Again, weights (7) (but not (5)) ensure a high breakdown point. The NLWS estimator with such weights is highly robust from the same reasons as the LWS estimator in the linear regression. The main reason for the robustness of the NLWS estimator is the construction of the estimator itself, just like for the LWS estimator in the linear regression.

An approximate algorithm for the optimization task (23) can be obtained as a straightforward adaptation of the LTS algorithm for the linear regression (cf. [20, 14]); nevertheless, its properties in this context have not been investigated. Empirical investigations will be performed on real data sets in Section 5.

4.3 Two-stage NLWS

In this section, a version of the NLWS estimator is proposed, which constructs data-dependent adaptive weights. The estimator has a two-stage structure and is inspired by a two-stage LWS estimator of [6].

Čížek [6] proved his two-stage estimator with quantile-based adaptive weights in the linear model to possess a high breakdown point and at the same time a 100 % asymptotic efficiency of the least squares under Gaussian errors. Further, he evaluated its relative efficiency to be high (over 85 %) compared to maximum likelihood estimators in a numerical study under various distributional models for samples of several tens of observations.

We propose a two-stage estimator denoted as 2S-NLWS which can be described as an improved version of the NLWS estimator which contains a construction of data-dependent adaptive weights. The model (13) is considered. The computation of the 2S-NLWS starts with an initial highly robust estimator $\hat{\beta}^0$ of β and proceeds to proposing values of the weights based on comparing the empirical distribution function of squared residuals with its theoretical counterpart assuming normality.

In the first stage, it is crucial to choose a suitable initial estimator, because it influences the properties of the resulting 2S-NLWS estimator. Therefore, it is recommendable to use a consistent estimator which is highly robust, i.e. NLTS with h between (say) $n/2$ and $3n/4$ or NLWS with weights (7). Residuals of the initial fit will be denoted as $u_1^0, \ldots, u_n^0$. We will need the notation $\left(G_n^0\right)^{-1}$ for the empirical quantile function computed from these residuals, F_χ^{-1} for the quantile function of χ_1^2 distribution and

$$b_n = \min\left\{\frac{m}{n};\ u_{(m)}^2 > 0\right\}. \tag{16}$$

In the second stage, the weights for the 2S-NLWS estimator are constructed. They are defined by means of a weight function $\tilde{w}(t)$ for $t \in [0,1]$, where

$$\tilde{w}(t) = \frac{F_\chi^{-1}\left(\max\{t, b_n\}\right)}{\left(G_n^0\right)^{-1}\left(\max\{t, b_n\}\right)}. \tag{17}$$

In other words, weights for a fixed number of observations n are given as

$$\tilde{w}(t) = \frac{F_\chi^{-1}(t)}{\left(G_n^0\right)^{-1}(t)} \quad \text{for} \quad t \in \left\{\frac{1}{2n}, \frac{3}{2n}, \ldots, \frac{2n-1}{2n}\right\}. \tag{18}$$

The computation of the 2S-NLWS estimator is straightforward. In a non-contaminated model, the 2S-NLWS estimator can be easily proven to have a full efficiency of the least squares, just like in the linear case [6].

Theorem 1. *Random vectors $X_1, \ldots, X_n$ are assumed to be independent identically distributed. Let $e_1, \ldots, e_n$ be independent identically distributed, independent on $X_1, \ldots, X_n$ and fulfilling $e_i \sim \mathsf{N}(0, \sigma^2)$ for each i. Let the initial estimator $\hat{\beta}^0$ be consistent with a corresponding consistent estimator of σ^2. Then it holds*

$$\tilde{w}(t) \xrightarrow{\mathsf{P}} \sigma^2 \quad \textit{for each } t \in (0,1), \tag{19}$$

where $\xrightarrow{\mathsf{P}}$ denotes the convergence in probability.

Proof. Analogy of [6]. □

Corollary 1. *Under the assumptions of Theorem 1, the 2S-NLWS estimator $\hat{\beta}_{2S-NLWS}$ fulfils*

$$\hat{\beta}_{2S-NLWS} \xrightarrow{\mathsf{P}} \beta. \tag{20}$$

Concerning other properties of the 2S-NLWS estimator, the breakdown point seems to require much more effort to be derived. Nevertheless, it remains clear that robustness properties of the 2S-NLWS are strongly influenced by those of the initial estimator.

5 Examples on Robust Estimation in Nonlinear Regression

This section presents three examples investigating the performance of various estimators in the nonlinear regression, especially focused on the soundness of the proposed methodology of Section 4. An example illustrating the performance of the NLWS estimator in both non-contaminated and contaminated data is presented in Section 5.1. The next example in Section 5.2 investigates the tightness of an approximate algorithm for computing the NLWS estimator. The final example presented in Section 5.3 exploits the nonparametric bootstrap estimation of the variance of various estimators in the nonlinear regression model.

Estimator	b_1	b_2
Original data set		
NLS	190.8	0.060
NLTS ($h = 18$)	191.6	0.061
NLWS (weights [6])	191.4	0.061
Contaminated data set		
NLS	193.9	0.067
NLTS ($h = 18$)	191.8	0.060
NLWS (weights [6])	191.6	0.061

Table 2: Results of example with real data of Section 5.1. Classical and robust estimators of β_1 and β_2 in the nonlinear regression model are computed in the original as well as contaminated version of the data set.

5.1 Example: Puromycin Data

This example has the aim to compare the performance of classical and robust estimators in the nonlinear model. This will be performed on a real data set without apparent outliers. To show the sensitivity of the NLS and on the other hand the robustness (resistance) of the NLTS and NLWS, the computations are repeated on a modified version of this data set, which contains one outlier.

A standard data set called Puromycin with $n = 23$ observations, which is available in the package `datasets` of `R` software, is considered. The reaction velocity (rate) Y is explained as a response of the substrate concentration X in the nonlinear regression model

$$Y_i = \frac{\beta_1 X_i}{\beta_2 + X_i} + e_i, \quad i = 1, \ldots, n, \tag{21}$$

where the aim is to estimate regression parameters β_1 and β_2.

The results of the least squares and NLWS estimators are shown in Table 2. The NLWS estimator turns out to perform reliably on a data set contaminated by outlying measurements as well as on data without such contamination. In addition, we verified the constant $R = 10\,000$ in Algorithm 1 to be more than sufficient in the nonlinear regression model and a moderate sample size.

Further, we also consider a contaminated data set, obtained by modifying the value of the observation in the Puromycin data set. Particularly, the substrate concentration (i.e. the regressor) of the first observation was modified from 0.02 to 0.05 to become the only outlier in the data set. This reveals the influence of a (local) change of one observation on the results and reveals the true advantage of the robust estimators. Robust estimates namely remain almost unchanged, while the contamination is revealed on the NLS estimator. In other words, the NLS starts to differ from the robust estimators, while all estimates were much more similar for the original data set.

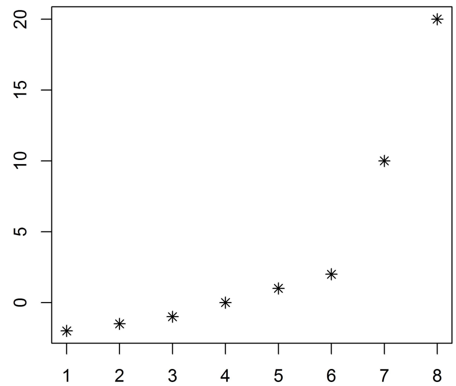

Figure 1: Data in the example of Section 5.2.

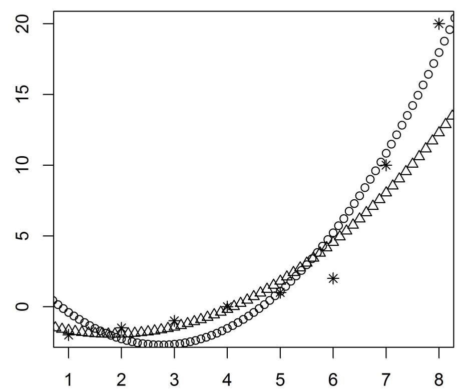

Figure 2: Results of the example of Section 5.2. The NLS (circles) and NLWS (triangles) in the example of Section 5.2.

5.2 Example: Simulated Data

The performance of the NLWS estimator will be now illustrated and investigated on a numerical example with simulated data. The data set consisting of 8 data points is shown in Figure 1. The nonlinear regression model is used in the form

$$Y_i = \beta_0 + \beta_1 (X_i - \beta_2)^2 + e_i, \quad i = 1, \ldots, n, \tag{22}$$

where $Y_1, \ldots, Y_n$ are values of the response, $X_1, \ldots, X_n$ values of the only regressor, β_0, β_1 and β_2 are regression parameters and $e_1, \ldots, e_n$ are random errors. Figure 2 shows fitted values corresponding to the NLS fit and also the NLWS fit with the linearly decreasing weights. The NLS fit has the tendency to fit well also influential data points. The robust fit better explains a subset of data points, while it considers data points corresponding to larger values of the regressor to be outliers.

	Loss function		
Estimator	(23)	(24)	(25)
NLS	23.44	2.47	1.21
NLTS ($h = 5$)	46.06	1.93	1.11
NLWS (weights (5))	70.94	6.82	0.67

Table 3: Results of the example with simulated data of Section 5.2. Values of various loss functions computed for the NLS, NLTS and NLWS estimators.

Table 3 gives values of loss functions

$$\sum_{i=1}^{n} u_i^2(b), \tag{23}$$

$$\sum_{i=1}^{h} u_{(i)}^2(b) \tag{24}$$

and

$$\sum_{i=1}^{n} w_i u_{(i)}^2(b) \tag{25}$$

corresponding to the NLS, NLTS and NLWS, respectively. These are evaluated for all of the three estimates.

The NLS estimator minimizes (23) as expected and thus can be expected to yield also a rather small value of (25). The NLWS estimator has a much larger value of (23) compared to the NLS fit. However, the algorithm used for computing the NLWS has found even a much smaller value of (25) than the NLS. On the whole, the results of Table 3 thus give a clear evidence in favor of the reliability of the algorithm for computing the NLWS estimator.

5.3 Example: Nonlinear Model for Investment Data

The bootstrap based variance estimation procedure described in Section 3.1 can be also utilized in the context of nonlinear regression models. The following example incorporates such an approach in order to compare the variances of the NLS, NLTS, and NLWS estimators.

The validity of the nonparametric bootstrap in linear regression model (i.e., Algorithm 3.1) is, however, going to be verified only via a simulation study. Theoretical justification of the nonparametric bootstrap procedure needs to be provided by a formal proof with properly stated assumptions. In general, bootstrapping should be used with caution, because the nonparametric bootstrap algorithm does not always provide a consistent estimate.

The same data set is used as in Section 3.2. This time, a nonlinear regression model

$$Y_i = \beta_1 (X_i - \bar{X})^2 + \beta_2 X_i + \beta_3 + e_i, \quad i = 1, \ldots, n, \tag{26}$$

is considered, where the centering of the regressor using its mean $\bar{X}$ is done for the sake of numerical stability.

Three robust estimators are computed together with bootstrap estimates of their variances. The results are shown in Table 4. The conclusions are analogous to those of the example in Section 3.2, namely the NLTS estimator loses its efficiency very much compared to the NLS. The NLS remains to be the most efficient, i.e. retains the smallest variance for the data set which does not contain severe outliers. Still, the NLWS loses relatively little compared to the NLS while it is able to outperform the NLTS

Estimator	β_1 (s_1)	β_2 (s_2)	β_3 (s_3)
NLS	$3.7 \cdot 10^{-5}$	0.22	98
	$(3.4 \cdot 10^{-5})$	(0.07)	(74)
NLTS	$4.1 \cdot 10^{-5}$	0.25	112
($h = 13$)	$(1.8 \cdot 10^{-4})$	(0.36)	(381)
NLWS	$4.0 \cdot 10^{-5}$	0.25	107
(weights (5))	$(6.7 \cdot 10^{-5})$	(0.13)	(140)

Table 4: Results of the example of Section 5.3. The classical and robust estimates of parameters β_1, β_2, β_3 computed for (26) are shown without brackets and accompanied by bootstrap estimates of their standard deviation denoted as s_1, s_2 and s_3 shown in brackets.

strongly. Thus, we can say that the NLWS estimator is able to combine the robustness with efficiency reasonably well, in comparison to the non-efficient (but much more renowned) NLTS estimator.

6 Conclusions

This paper investigates robust estimator for the linear and nonlinear regression methods and nonparametric bootstrap approaches to estimating its variance. Implicitly weighted estimators are considered, which include the least trimmed squares and least weighted squares (in linear and nonlinear versions).

After recalling the state of the art on implicitly weighted robust estimation in Section 2, a bootstrap method for estimating the variance of the LWS estimator is proposed in Section 3. A numerical example shows the LWS to have a much smaller variance compared to the more popular LTS estimator, which reveals another strong argument in favor of the LWS estimator and questions whether the LTS estimator deserves to be the most common highly robust regression estimator.

Considerations for the linear regression are further generalized to nonlinear regression. Thus, the main contribution can be found in Section 4 on the NLWS estimation, which has not been much investigated in the references so far. Our work is devoted to a bootstrap estimate suitable for estimating the variance of the NLWS estimator. In addition, a new version of the estimator is proposed, which computes data-dependent adaptive weights. Their construction allows to define the 2S-NLWS, which is improved compared to the basic NLWS in terms of efficiency.

Several examples reveal the suitability of the idea of the NLWS for modelling a nonlinear trend in the data. It also follows from the set of examples that an approximate algorithm, which is available for the NLWS, turns out to be reliable. The examples also give a warning that the NLWS estimator behaves in a rather intricate way and the estimator is much more complex compared to linear regression.

We also computed nonparametric bootstrap estimate of the variance of nonlinear estimators. These empirical results allow us to conclude that there seems a major advantage of the NLWS compared to the NLTS, although the NLS remains to be recommendable for data without severe outliers. The NLWS seems to be close to a reasonable combination of the efficiency (for normal regression errors) with high robustness (for models with contamination), which represents a dream of robust statisticians since the dawn of robust statistical inference.

The examples investigated in this paper lead us also to formulating the following disadvantages of robust estimators in nonlinear regression:

- They require various tuning constants with a difficult interpretation;
- Various robust methods yield rather different results;
- Computational intensity;
- Robustness only with respect to outliers but not to a misspecification of the model.

Important limitations of the robust nonlinear estimation include a non-robustness to small modifications of the nonlinear function f in the model (see [2]). Robust nonlinear estimators also require rather tedious proofs of their properties.

Some of the properties of robust estimators valid in the linear regression are not valid in the nonlinear model at all. As an example let us mention diagnostic tools, which can be derived for robust estimators in linear regression, but would be rather controversial in the nonlinear model [15]. The results on simulated data in Section 5.2 reveal that the residuals are far from homoscedasticity, even if the assumption of homoscedastic disturbances in the regression model is fulfilled. Thus, we find residuals to be unsuitable for making conclusions about the disturbances (random errors). While tests from the linear regression are no longer valid for the NLWS estimator, we do not recommend to use residuals even for a subjective diagnostics concerning the disturbances.

We intend to apply the robust regression methods of this paper within a future research in the area of metalearning, which aims at comparing the suitability of various machine learning methods for different data. Robustifying metalearning for regression methods is however not only a matter of using robust regression methods, but the process of metalearning itself suffers from instability [21] and a robust analogy of the whole process of metalearning is highly desirable to be performed in a complex and systematic way.

Acknowledgments

The project was supported by the project 17-01251S "Metalearning for extracting rules with numerical consequents" of the Czech Science Foundation.

References

[1] Agresti, A.: Analysis of ordinal categorical data. 2nd edn., Wiley, New York, 2010

[2] Baldauf, M., Silva, J.M.C.S.: On the use of robust regression in econometrics. Economic Letters **114** (2012) 124–127

[3] Bickel, P.J., Freedman, D.A.: Some asymptotic theory for the bootstrap. Annals of Statistics **9** (1981) 1196–1217

[4] Brazdil, P., Giraud-Carrier, C., Soares, C., Vilalta, E.: Metalearning: Applications to data mining. Springer, Berlin, 2009

[5] Čížek, P.: Least trimmed squares in nonlinear regression under dependence. Journal of Statistical Planning and Inference **136** (2006) 3967-–3988

[6] Čížek, P.: Semiparametrically weighted robust estimation of regression models. Computational Statistics & Data Analysis, **55** (2011) 774–788

[7] Čížek, P.: Reweighted least trimmed squares: An alternative to one-step estimators. Test **22** (2013) 514–533

[8] Efron, B., Tibshirani, R.J.: An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton, 1994

[9] Freedman, D.A., Peters, S.C.: Bootstrapping a regression equation: Some empirical results. Journal of the American Statistical Association **79** (1984) 97–106

[10] Godfrey, L.: Bootstrap tests for regression models. Palgrave Macmillan, London, 2009

[11] Hall, P., DiCiccio, T.J., Romano, J.P.: On smoothing and the bootstrap. Annals of Statistics **17** (1989) 692–704

[12] Jurczyk, T.: Ridge least weighted squares. Acta Universitatis Carolinae Mathematica et Physica **52** (2011) 15–26

[13] Jurečková, J., Sen, P.K., Picek, J.: Methodology in robust and nonparametric statistics. CRC Press, Boca Raton, 2012

[14] Kalina, J.: Some robust estimation tools for multivariate models. Proceedings International Days of Statistics and Economics MSED 2015, Melandrium, Slaný, 2015, 713–722.

[15] Kalina, J.: On robust information extraction from high-dimensional data. Serbian Journal of Management **9** (2014) 131–144

[16] Kalina, J.: Implicitly weighted methods in robust image analysis. Journal of Mathematical Imaging and Vision **44** (2012) 449–462

[17] Kalina, J.: Least weighted squares in econometric applications. Journal of Applied Mathematics, Statistics and Informatics **5** (2009) 115–125

[18] Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: On the least trimmed squares estimator. Algorithmica **69** (2014) 148–183

[19] Pešta, M.: Total least squares and bootstrapping with application in calibration. Statistics: A Journal of Theoretical and Applied Statistics **47** (2013) 966–991

[20] Rousseeuw, P.J., van Driessen, K.: Computing LTS regression for large data sets. Data Mining and Knowledge Discovery **12** (2006) 29–45

[21] Smith-Miles, K., Baatar, D., Wreford, B., Lewis, R.: Towards objective measures of algorithm performance across instance space. Computers and Operations Research **45** (2014) 12–24

[22] Stromberg A.J., Ruppert, D.: Breakdown in nonlinear regression. Journal of the American Statistical Association **87** (1992) 991–997

[23] Víšek J.Á.: Consistency of the least weighted squares under heteroscedasticity. Kybernetika **47** (2011) 179–206

[24] Víšek J.Á.: The least trimmed squares, Part III: Asymptotic normality. Kybernetika **42** (2006) 203–224

[25] Víšek J.Á.: Regression with high breakdown point. In Antoch J., Dohnal G. (Eds.): Proceedings of ROBUST 2000, Summer School of JČMF, JČMF and Czech Statistical Society, Prague, 2001, 324–356.