
Statistical learning for recommending (robust) nonlinear regression methods

J. KALINA AND J. TICHAŤAVSKÝ

Abstract

We are interested in comparing the performance of various nonlinear estimators of parameters of the standard nonlinear regression model. While the standard nonlinear least squares estimator is vulnerable to the presence of outlying measurements in the data, there exist several robust alternatives. However, it is not clear which estimator should be used for a given dataset and this question remains extremely difficult (or perhaps infeasible) to be answered theoretically. Metalearning represents a computationally intensive methodology for optimal selection of algorithms (or methods) and is used here to predict the most suitable nonlinear estimator for a particular dataset. The classification rule is learned over a training database of 24 publicly available datasets. The results of the primary learning give an interesting argument in favor of the nonlinear least weighted squares estimator, which turns out to be the most suitable one for the majority of datasets. The subsequent metalearning reveals that tests of normality and heteroscedasticity play a crucial role in finding the most suitable nonlinear estimator.

Mathematics Subject Classification 2010: 68T05, 62G35, 62J02, 68-04

General Terms: Statistical learning, Nonlinear regression, Robustness, Heteroscedasticity

Keywords: nonlinear least weighted squares, optimal method selection, optimization, computations

1. INTRODUCTION

The aim of regression modeling is to explain a continuous response variable based on one or more independent variables (regressors), where the latter may be continuous and/or discrete. This also allows to predict values of the response for individual values of regressors. The nonlinear regression model can be described as an important category of parametric regression models and will be considered in this paper.

The most traditional estimation tool for nonlinear regression, i.e. the nonlinear least squares estimator, is well known to be too vulnerable to the presence of outlying measurements (outliers) in the data [Seber and Wild 2003]. Therefore, robust estimation techniques for the nonlinear regression have been proposed and have become established e.g. in econometric applications [Riazoshams et al. 2010; Baldauf and Silva 2012]. The concept of breakdown point has become a fundamental

This work is supported by the Czech Research Council projects 17-07384S and 19-05704SS.

measure of (global) robustness suitable for nonlinear regression estimators [Stromberg and Ruppert 1992]. Nevertheless, it remains unknown to answer the question, which estimator should be used for a given dataset; this is perhaps impossible to be found theoretically across all possible datasets.

Metalearning is a methodology allowing to recommend the most suitable algorithm (or method) for a given dataset, based on information learned over a training database of datasets. It extracts (prior) knowledge from datasets applicable to a new dataset and therefore has become quite popular in recent computer science, optimization or data mining. Metalearning also starts to penetrate to economic applications. To give specific examples of metalearning in computational economics, metalearning was performed to compare economic genetic algorithms (especially with a modified mutation operator suitable for consumption models) [Riechmann 2001] or to compare various optimization tools (mainly hybrid algorithms) for business, economics and finance [Vasant 2012].

Metalearning has been recommended in the seminal paper [Smith-Miles et al. 2014] especially for those domains, in which a theoretical knowledge is too difficult to acquire. Finding the best nonlinear regression estimator is exactly a task, for which it remains very tedious (or perhaps infeasible) to derive any practical solution rigorously [Stromberg and Ruppert 1992]. Thus, metalearning represents a computational (however computationally intensive) approach able to offer some practical advice in this respect. For a discussion of advantages and limitations of metalearning, we refer to [Smith-Miles et al. 2014], but we hold the opinion that a conscientiously critical evaluation of metalearning is still missing. Based on our experience, metalearning in a habitual form is vulnerable to instability and sensitivity to data contamination by noise, outlying values (outliers), or presence of redundant variables.

In this paper, metalearning will be used in the context of nonlinear regression with the aim to predict the best method for particular datasets not contained in the training database. Section 2 recalls various estimators for the nonlinear regression model. Section 3 presents a numerical example revealing the severe bias of the nonlinear least trimmed squares under heteroscedasticity. Section 4 describes our metalearning study, the results of which are presented in Section 5. Finally, Section 6 concludes the paper.

2. ROBUST ESTIMATION IN NONLINEAR REGRESSION

Let us consider the standard nonlinear regression model

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

where f is a given continuous nonlinear function, Y_1, \dots, Y_n is a continuous response and $(X_{1j}, \dots, X_{nj})^T$ is the j -th regressor for $j = 1, \dots, p$. We use the notation e_1, \dots, e_n for random errors and X for the matrix with elements X_{ij} , where $i = 1, \dots, n$ and $j = 1, \dots, p$. The classical estimator, which is the nonlinear least squares (NLS) estimator of β , is vulnerable to the presence of outliers in the data [Maronna et al. 2006; Fasano et al. 2012]. Therefore, we recall several of its potential robust alternatives in this section. All of them will be also used later in our computations.

2.1. Nonlinear least trimmed squares

The nonlinear least trimmed squares (NLTS) estimator represents one of robust methods with a (possibly) high breakdown point [Stromberg and Ruppert 1992] and also an extension of the popular least trimmed squares from linear regression [Rousseeuw 1983; Hampel et al. 1986; Mount et al. 2014]. We denote by \mathbb{R} the set of real numbers and the residual corresponding to the i -th observation will be denoted as

$$u_i(b) = Y_i - f(b_1 X_{i1}, \dots, b_p X_{ip}), \quad i = 1, \dots, n, \quad (2)$$

for any (fixed) $b = (b_1, \dots, b_p)^T \in \mathbb{R}^p$. Let us arrange squared values in ascending order as

$$u_{(1)}^2(b) \leq \dots \leq u_{(n)}^2(b). \quad (3)$$

The user must specify a suitable value of the trimming constant h ($n/2 \leq h \leq n$). Then, the NLTS estimator b_{NLTS} of β is obtained as

$$b_{NLTS} = \arg \min_{b \in \mathbb{R}^p} \sum_{i=1}^h u_{(i)}^2(b). \quad (4)$$

While the choice of h should reflect the percentage of contaminated data, one usually takes 25 % of outliers in applications and thus h is commonly chosen as the integer part of $3n/4$. An alternative approach is to select h by repeating the evaluation of the NLTS with an increasing h (starting with $n/2$) up to the moment when the estimates of the model (especially the corresponding estimate of σ^2) abruptly and greatly change.

The computation of the NLTS estimator requires minimization over all $\binom{n}{h}$ h -subsets of $\{1, 2, \dots, n\}$, thus has combinatorial complexity and, depending on the relation between n and h , can become infeasible already for rather small numbers of observations n . Therefore, it is more convenient to use the FAST-LTS algorithm [Rousseeuw and van Driessen 2006], which approximates the LTS estimator in a computationally attractive way; it has become a standard tool for the LTS allowing very large sample sizes n .

2.2. Nonlinear least weighted squares

The nonlinear least weighted squares (NLWS) estimator represents an extension of the least weighted squares estimator from the linear regression and at the same time a weighted analogy of the NLTS estimator [Víšek 2011; Kalina 2014]. Let us assume the magnitudes w_1, \dots, w_n of nonnegative weights to be given. The NLWS estimator of the parameters in (1) is defined as

$$b_{NLWS} = \arg \min_{b \in \mathbb{R}^p} \sum_{i=1}^n w_i u_{(i)}^2(b), \quad (5)$$

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

The choice of weights clearly has a determining influence on properties of the estimator; if namely one assigns zero weights to outlying observations, then the estimator is ensured to be highly robust in terms of the breakdown point. The main reason for such robustness of the NLWS estimator is the construction of the estimator itself, just like for the LWS estimator in the linear regression. Various weighting schemes will be described in Section 2.4.

An approximate algorithm for the optimization task (5) can be obtained as a (weighted) adaptation of the NLTS algorithm [Rousseeuw and van Driessen 2006], which is presented here as Algorithm 1. To avoid any confusion, we denote coordinates of some vectors with the upper index there, e.g. the weights are denoted as $w = (w^1, \dots, w^n)^T$. It also exploits the weighted NLS estimator, which is a well described weighted analogue of the NLS [Seber and Wild 2003]. Concerning the choice of parameters, we use $J = 10000$ and $\varepsilon = 0.0001$ in all computations in this paper.

Algorithm 1 Nonlinear least weighted squares (NLWS) estimator of β in (1)

Input: X_1, \dots, X_n , where $X_i \in \mathbb{R}^p$ for each $i = 1, \dots, p$
Input: Y_1, \dots, Y_n
Input: Nonlinear function f
Input: $J > 0$
Input: $\varepsilon > 0$
Input: Magnitudes of weights $w^1 \geq \dots \geq w^n$
Output: Optimal permutation of w^1, \dots, w^n denoted as \tilde{w}

for $j = 1$ to J **do**
 $m_{j0} := +\infty$
 Select randomly p points
 Estimate β by $b_{j1} = (b_{j1}^1, \dots, b_{j1}^p)^T$ obtained as a NLS estimator using only the selected p points
 $k := 1$
 repeat
 $u_i(b_{jk}) := Y_i - f(b_{ik}^1 X_{i1}, \dots, b_{ik}^p X_{ip}), i = 1, \dots, n$
 Assign weights $w_{j,k+1} = (w_{j,k+1}^1, \dots, w_{j,k+1}^n)^T$ to individual observations based on (3)
 Estimate β by $b_{j,k+1} = (b_{j,k+1}^1, \dots, b_{j,k+1}^p)^T$ obtained as a weighted NLS with weights $w_{j,k+1}$
 $m_{j,k+1} := \sum_{i=1}^n w_{j,k+1}^i u_i^2(b_{j,k+1})$
 $k := k + 1$
 until $m_{jk} > m_{j,k-1} + \varepsilon$
 $\tilde{w}_j = (\tilde{w}_j^1, \dots, \tilde{w}_j^n)^T := w_{j,k-1}$
 $\tilde{b}_j = (\tilde{b}_j^1, \dots, \tilde{b}_j^n)^T := b_{j,k-1}$
end for

$$j^* := \arg \min_{j=1, \dots, J} \sum_{i=1}^n \tilde{w}_j^i u_i^2(\tilde{b}_j) \tag{6}$$

$b_{NLWS} :=$ weighted NLS estimate with weights

$$\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^n)^T := (\tilde{w}_{j^*}^1, \dots, \tilde{w}_{j^*}^n)^T$$

2.3. Nonlinear regression median

Regression quantiles represent a natural generalization of sample quantiles to the linear regression model. The estimator depends on a parameter α in the interval $(0,1)$, which corresponds to dividing the disturbances to $\alpha \cdot 100\%$ values below the regression quantile and the remaining $(1 - \alpha) \cdot 100\%$ values above the regression quantile. In general, regression quantiles represent an important tool of regression methodology, which is popular in economic applications. A natural extension of

Table I. Results of various nonlinear estimators in the example of Section 3.

Estimator	Estimates of		
	β_1	β_2	β_3
1 (NLS)	-4.5	1.1	-1.0
2 (NRM)	-4.7	1.5	-1.0
3 (NLTS)	-4.5	0.9	-1.0
4 (NLWS)	-4.6	1.1	-0.9
5 (NLWS)	-4.5	1.0	-1.0
6 (NLWS)	-4.6	0.9	-1.0
7 (NLWS)	-4.5	1.0	-0.9

regression quantiles to nonlinear regression is also available [Koenker and Park 1996], while the most important special case remains to be the nonlinear regression median (NRM) with $\alpha = 1/2$.

2.4. Estimators used in the computation

In the example of Section 3 as well as in the metalearning study (see Section 4), we use the following seven available estimators, which will be denoted as estimators 1, ..., 7:

- (1) Nonlinear least squares (NLS).
- (2) Nonlinear regression median (NRM).
- (3) NLTS with h equal to $\lfloor \frac{3n}{4} \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$.
- (4) NLWS with data-dependent adaptive weights of [Čížek 2011].
- (5) NLWS with linear weights

$$w_i = \frac{2(n+1-i)}{n(n+1)}, \quad i = 1, \dots, n. \quad (7)$$

- (6) NLWS with trimmed linear weights

$$w_i = \frac{h-i+1}{h} I[i \leq h], \quad i = 1, \dots, n, \quad (8)$$

where $I[\cdot]$ denotes an indicator function and h equals again to $\lfloor \frac{3n}{4} \rfloor$.

- (7) NLWS with weights generated by the (strictly decreasing) logistic function

$$w_i = \left(1 + \exp \left\{ \frac{i-n-1}{n} \right\} \right)^{-1}, \quad i = 1, \dots, n. \quad (9)$$

The habitual standardization of weights to $\sum_{i=1}^n w_i = 1$ is required in all cases.

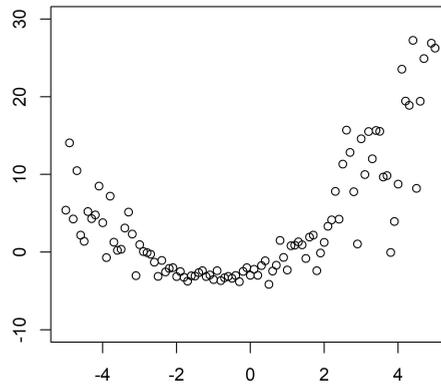


Fig. 1. Heteroscedastic data from the example of Section 3.

3. EXAMPLE: HETEROSCEDASTIC DATA

The effect of heteroscedasticity of errors on classical regression estimation, i.e. deterioration of the estimates or even their complete malfunction, seems to be still underestimated in statistical and econometric literature [Wooldridge 2001; Kalina et al. 2019]. The following example illustrates the performance various nonlinear estimators on artificial (randomly generated) heteroscedastic data shown in Figure 1. We generate the data from the model

$$Y_i = \beta_1 + \beta_2(X_i - \beta_3)^2 + e_i, \quad i = 1, \dots, n, \quad (10)$$

with $\beta_1 = -4.6$, $\beta_2 = 1$ and $\beta_3 = -1$. Estimates of β in this nonlinear model, i.e. a nonlinear function of a single regressor X_1, \dots, X_n , are shown in Table I. The NLTS ignores the heteroscedastic character of the data completely and chooses an unsuitable subset of data to play the role of the reliable majority of the data points. The NLS as well as NLWS estimates are able to find a more suitable regression estimate also for observations with a value of the regressors exceeding 2. Their residuals are much more symmetric around 0 compared to NLTS residuals.

It is mainly β_2 which may be affected by heteroscedasticity. The NLTS is a clear loser here as it ignores data generated with smallest (i.e. negative) residuals and considers them to be by outlying. On the other hand, the NLWS estimator seems to yield a reasonable result independently on the choice of weights so the particular choice of weights for the NLWS seems to be rather unimportant.

4. DESCRIPTION OF THE METALEARNING STUDY

Metalearning is a computational (machine learning) approach allowing to exploit information from previously observed datasets and to extend them to new datasets [Brazdil et al. 2009; Smith-Miles et al. 2014]. While it has become popular in optimization and to some extent in machine learning, it can be exploited for recommending the optimal method also in the context of statistical estimation. In this paper, we describe our metalearning study with the aim to compare various nonlinear regression estimators. The study allows us also to detect the most relevant criteria for determining the most suitable weights for the NLWS.

The primary learning task is to fit various nonlinear regression estimators for each of the given datasets. The best estimator is found using a specified characteristic of a goodness of fit. The subsequent metalearning part has the aim to learn a classification rule allowing to predict the best regression method for a new dataset not present in the training database. Its input data are only selected features of individual datasets together with the result of the primary learning, which typically has the form of the index of the best method for each of the training datasets. In general, the user of metalearning must specify a list of essential components (parameters). We will now describe our choices for the primary learning as well as the subsequent metalearning part of the task.

4.1. Primary learning

We use 24 real datasets previously used in [Kalina and Peřtová 2017]. In each of them, we consider the model

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \sum_{j=1}^p \beta_{p+j} (X_{ij} - \bar{X}_j)^2 + e_i, \quad i = 1, \dots, n, \quad (11)$$

which is a nonlinear model with the total number of $2p + 1$ regressors (i.e. nonlinear as a function of the original p regressors) and the variables are centered using the mean of the j -th variable \bar{X}_j (for $j = 1, \dots, p$) for the sake of numerical stability.

We use the most standard choice of the prediction measure, which is the (prediction) mean square error (MSE). We find the best method for each dataset using MSE in a standard leave-one-out cross validation, which represents a popular attempt for an independent validation [Hastie et al. 2008].

4.2. Metalearning

The output of Section 4.1 is used in the form of the factor variable (index, indicator) of the best method for each of the datasets together with a list of features computed for each dataset. In contrary to the linear case of [Kalina and Peřtová 2017], we need to consider features for the metalearning more carefully now in the nonlinear model. For example, there is no analogue of the coefficient of determination in the nonlinear model. Other features must be evaluated for the NLS fit. Thus we came to selecting the following set of 9 features.

- (1) The number of observations n .
- (2) The number of regressors p (excluding the intercept).
- (3) The ratio n/p .
- (4) Condition number of the matrix $(X^T X)^{-1}$.
- (5) Normality of residuals of the NLS, evaluated as the p -value of the Shapiro-Wilk test.
- (6) Skewness of residuals of the NLS.
- (7) Kurtosis of residuals of the NLS.
- (8) Heteroscedasticity of residuals of the NLS evaluated as the p -value of the White's test.
- (9) Test of linearity evaluated as the p -value of the test of $H_0 : \beta_{p+1} = \dots = \beta_{2p} = 0$ in (11) based on the NLS estimates.

For the subsequent metalearning task, which is a task of classification to 7 groups (i.e. finding the best among the 7 estimators of Section 2.4), we exploit various classification methods including support vector machines (SVM), k -nearest neighbors, or a classification tree. We use also several other less known methods including a regularized version of linear discriminant analysis (LDA) denoted as SCRDA of [Guo et al. 2007] or a robust version of LDA denoted as linear MWCD classification [Kalina 2012].

5. RESULTS

We used the R software for all the computations including necessary libraries (robustbase, quantreg, rda) for specific tasks (robust estimation, classification). Using MSE as a measure of prediction performance, a leave-one-out cross validation is used as a standard attempt for an independent validation. Table II shows the best

method of the primary learning for each of the datasets. There, the notation $1, \dots, 7$ according to list of methods in Section 2.4 is used. The ranks are presented here, so that method 4 (NLWS with data-dependent weights) is the best method for the Aircraft dataset, method 3 (NLTS) performs as the second best on this dataset etc. On the whole, the NLWS is the best estimator based on Table II, but no weighting scheme seems to be superior to the remaining choices.

Within the subsequent metalearning task, we computed various classifiers, while default settings of parameters was used for those computed using the R software. As the results presented in Table III indicate, the SVM classifier turns out to have the best performance. Because there are as many as 7 classes, the overall prediction ability is not very high (although not much lower compared to e.g. [Kalina and Peřtová 2017]). The most useful criteria for the choice of weights for the NLWS turn out to be heteroscedasticity and normality of the NLS residuals, which correspond to our intuition, because their heavy violation requires a robust approach, i.e. a highly robust estimation in terms of the breakdown point.

6. CONCLUSIONS

In this paper, we work with various estimators of parameters in the standard nonlinear regression model. We are especially interested in the NLWS estimator, which is a natural extension of the LWS estimator [Víšek 2011] to the nonlinear regression model. A novel algorithm for the NLWS is proposed in this paper. At the same time, an example with heteroscedastic data is presented in Section 3, which brings an argument in favor of the novel NLWS estimator compared to the previously investigated NLTS estimator. There, the NLWS estimator (basically with any weighting scheme) seems more suitable but at the cost of higher computational demands.

A metalearning study presented in this paper has the aim to construct a classification rule allowing to predict the most suitable nonlinear regression estimator for a particular dataset. For this purpose, we work with 24 real datasets with an economic background and performed a standard metalearning procedure. The NLWS seems to yield the best result for the majority of the datasets, while no weighting scheme is uniformly optimal over all datasets. The NLTS and the nonlinear regression median have a weaker performance due to a low efficiency. Thus, the concept of efficiency (and not only the robustness) seems to play a prominent role in the analysis of real data and it is the newly investigated NLWS estimator, which

Table II. Results of primary learning in a leave-one-out cross-validation study. Ranks corresponding to the mean prediction errors of various nonlinear regression estimators for each of the 24 datasets.

Dataset	1 NLS	2 NRM	3 NLTS	4 NLWS	5 NLWS	6 NLWS	7 NLWS
1 Aircraft	7	6	2	1	5	4	3
2 Ammonia	6	4	2	3	5	6	1
3 Auto MPG	3	5	1	4	7	6	2
4 Cirrhosis	2	1	5	6	3	7	4
5 Coleman	1	2	3	7	6	5	4
6 Delivery	7	3	6	1	2	4	5
7 Education	1	2	3	4	7	6	5
8 Electricity	2	1	4	5	3	7	6
9 Employment	4	2	3	7	1	5	6
10 Furniture 1	3	1	5	7	2	4	6
11 Furniture 2	7	4	3	6	1	2	5
12 GDP growth	2	1	7	4	5	6	3
13 Houseprices	6	4	1	3	5	7	2
14 Housing	3	1	2	6	7	4	5
15 Imports	5	2	4	6	7	1	3
16 Kootenay	1	2	5	3	7	6	4
17 Livestock	7	6	2	3	5	1	4
18 Machine	2	7	5	4	3	6	1
19 Murders	3	2	7	4	6	5	1
20 NOx	4	5	1	3	7	6	2
21 Octane	2	1	5	4	6	7	3
22 Pasture	7	2	6	3	5	1	4
23 Pension	7	4	3	1	2	5	6
24 Petrol	2	3	7	1	5	6	4

Table III. Results of metalearning evaluated as the ratio of correctly classified cases in a leave-one-out cross validation study.

Classification method	Prediction accuracy
Classification tree	0.33
<i>k</i> -nearest neighbor (<i>k</i> = 3)	0.46
LDA	0.54
SCRDA	0.54
Linear MWCD-classification	0.54
Multilayer perceptron	0.50
Logistic regression	0.50
SVM (linear)	0.54
SVM (Gaussian kernel)	0.58

seems able to combine robustness with efficiency very well.

ACKNOWLEDGEMENTS

The authors would like to thank Barbora Peřtová for acquiring the datasets and the reviewer for constructive advice.

REFERENCES

- BALDAUF, M. AND SILVA, J. 2012. On the use of robust regression in econometrics. *Economic Letters* 114, 124–127.
- BRAZDIL, P., GIRAUD-CARRIER, C., SOARES, C., AND VILALTA, E. 2009. *Metalearning: Applications to data mining*. Springer, Berlin.
- ČÍŽEK, P. 2011. Semiparametrically weighted robust estimation of regression models. *Computational Statistics & Data Analysis* 55, 774–788.
- FASANO, V., MARONNA, R., SUED, M., AND YOHAI, V. 2012. Continuity and differentiability of regression M-functionals. *Bernoulli* 18, 1284–1309.
- GUO, Y., HASTIE, T., AND TIBSHIRANI, R. 2007. Regularized discriminant analysis and its application in microarrays. *Biostatistics* 8, 86–100.
- HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J., AND STAHEL, W. A. 1986. *Robust statistics—The approach based on influence functions*. Wiley, New York.
- HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. 2008. *The elements of statistical learning*, 2nd ed. Springer, New York.
- KALINA, J. 2012. Highly robust statistical methods in medical image analysis. *Biocybernetics and Biomedical Engineering* 32, 3–16.
- KALINA, J. 2014. On robust information extraction from high-dimensional data. *Serbian Journal of Management* 9, 131–144.
- KALINA, J. AND PEŠTOVÁ, B. 2017. Robust regression estimators: A comparison of prediction performance. In *35th International Conference Mathematical Methods in Economics MME 2017*. University of Hradec Králové, Hradec Králové, 307–312.
- KALINA, J., VAŠANIČOVÁ, P., AND LITAVCOVÁ, E. 2019. Regression quantiles under heteroscedasticity and multicollinearity: Analysis of travel and tourism competitiveness. *Ekonomický časopis/Journal of Economics* 67, 69–85.
- KOENKER, R. AND PARK, B. 1996. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics* 71, 265–283.
- MARONNA, R., MARTIN, D., AND YOHAI, V. 2006. *Robust statistics: Theory and methods*. Wiley, New York.
- MOUNT, D., NETANYAHU, N., PIATKO, C., SILVERMAN, R., AND WU, A. 2014. On the least trimmed squares estimator. *Algorithmica* 69, 148–183.
- RIAZOSHAMS, H., MIDI, H., AND SHARIPOV, O. 2010. The performance of robust two-stage estimator in nonlinear regression with autocorrelated error. *Communications in Statistics – Simulation and Computation* 39, 1251–1268.
- RIECHMANN, T. 2001. *Learning in economics. Analysis and application of genetic algorithms*. Springer, Berlin.
- ROUSSEEUW, P. J. 1983. Multivariate estimation with high breakdown point. In *Proceedings of the 4th Pannonian Symposium on Mathematical Statistics*, W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds. D. Reidel, Dordrecht, 283–297.

-
- ROUSSEEUW, P. J. AND VAN DRIESSEN, K. 2006. Computing lts regression for large datasets. *Data Mining and Knowledge Discovery* 12, 29–45.
- SEBER, G. AND WILD, C. 2003. *Nonlinear regression*. Wiley, New York.
- SMITH-MILES, K., BAATAR, D., WREFORD, B., AND LEWIS, R. 2014. Towards objective measures of algorithm performance across instance space. *Computers & Operations Research* 45, 12–24.
- STROMBERG, A. AND RUPPERT, D. 1992. Breakdown in nonlinear regression. *Journal of the American Statistical Association* 87, 991–997.
- VASANT, P. M. 2012. *Meta-heuristics optimization algorithms in engineering, business, economics, and finance*. IGI Global, Hershey.
- VÍŠEK, J. Á. 2011. Consistency of the least weighted squares under heteroscedasticity. *Kybernetika* 47, 179–206.
- WOOLDRIDGE, J. M. 2001. *Econometric analysis of cross section and panel data*. MIT Press, Cambridge.

Jan Kalina

The Czech Academy of Sciences,

Institute of Computer Science,

Pod Vodárenskou věží 2, CZ-182 07 Praha 8, Czech Republic

&

The Czech Academy of Sciences,

Institute of Information Theory and Automation

Pod Vodárenskou věží 4, CZ-182 00 Praha 8, Czech Republic

e-mail: kalina@cs.cas.cz

Jan Tichavský

The Czech Academy of Sciences, Institute of Computer Science

Pod Vodárenskou věží 2

CZ-182 07 Praha 8

Czech Republic

e-mail: tichavsky@cs.cas.cz