A Bootstrap Comparison of Robust Regression Estimators

Jan Kalina[1], Patrik Janáček [2]

Abstract. The ordinary least squares estimator in linear regression is well known to be highly vulnerable to the presence of outliers in the data and available robust statistical estimators represent more preferable alternatives. It has been repeatedly recommended to use the least squares together with a robust estimator, where the latter is understood as a diagnostic tool for the former. In other words, only if the robust estimator yields a very different result, the user should investigate the dataset closer and search for explanations. For this purpose, a hypothesis test of equality of the means of two alternative linear regression estimators is proposed here based on nonparametric bootstrap. The performance of the test is presented on three real economic datasets with small samples. Robust estimates turn out not to be significantly different from non-robust estimates in the selected datasets. Still, robust estimation is beneficial in these datasets and the experiments illustrate one of possible ways of exploiting the bootstrap methodology in regression modeling. The bootstrap test could be easily extended to nonlinear regression models.

Keywords: linear regression, robust estimation, nonparametric bootstrap, bootstrap hypothesis testing

JEL Classification: C14
AMS Classification: 62F40

1 Introduction

As the linear regression represents the most fundamental model in current econometrics [5], it is crucial to estimate its parameters without being excessively influenced by the presence of outliers in the data [11]. Robust regression estimators started to become established alternatives to the least squares since late 1960s [18]. The practical analysis of economic data however lags behind the current trends in mathematical statistics, although new robust estimators have recently been developed and investigated. The robust estimates are more variable (less efficient) compared to the least squares for non-contaminated models and choosing the robust fit in every situation is not necessarily optimal. The robust procedures are either intended to replace the least squares as self-standing estimators, or they have the potential to accompany the least squares as a sort of diagnostic procedures [1]. In the second situation, the user does not have to decide whether the least squares estimator is reliable or whether the robust fit is preferable. In any case, users of robust statistics should also have the ambition to compare the performance of several methods and to decide for the method that is able to outperform other methods.

The approach based on understanding robust regression estimators as diagnostic tools for the least squares has been developed from the very dawn of robust estimation [6]. Such approach is still topical in current data analysis, as documented e.g. by the application of the least weighted squares estimator [24] in the study of [9], where robust analysis is presented primarily as a tool revealing non-robustness of a standard data analysis. An example of such a recent standing-alone methodology is the robust regression by means of the method of moments of [2], which is reliable under contaminated as well as non-contaminated models. In addition, robust estimators start to obtain their own diagnostic tools (cf. [25]). In general, if a robust estimator yields a (sufficiently) different result from the least squares, the user should investigate the dataset closer and search for explanations [17]; in this context, a formal hypothesis test would be very useful. Let us however proceed with formulating the test problem carefully.

In standard terminology, a coefficient estimate is a realization of an estimator obtained for the sample at hand. With the sample data in place, two different estimates are either different or not. Still, it may be useful to ask whether the expectations of two alternative estimates are equal or not. Thus, the question is how to perform a formal hypothesis test of equality of the means of two alternative estimators, especially for small sample sizes. If the two estimators are consistent, they are already asymptotically unbiased and any such test is redundant. Still, the test may be meaningful for situations with finite samples when assuming consistency is not desirable. In fact,

[1] The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Prague 8, Czech Republic & The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Prague 8, Czech Republic, kalina@cs.cas.cz
[2] The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Prague 8, Czech Republic, janacekpatrik@gmail.com

some of the robust estimators such as the least trimmed squares or least weighted squares require to assume a lengthy list of technical assumptions in order to achieve consistency; also, each of the estimators has its own set of specific assumptions [23, 24]. Not relying on the consistency has also the advantage that a possible extension of the test to nonlinear regression is straightforward; consistency properties remain unknown for some robust nonlinear regression esteimators (including common types of robust neural networks [20]). This motivates our aim to propose a test of equality of the means of two alternative linear regression estimators based on nonparametric bootstrap. We recall that nonparametric bootstrap represents a popular methodology for estimating variability (i.e. the covariance matrix) of various robust regression estimates [13].

Testing equality of the means of two regression estimates, which seems not be mentioned in recent robustness literature [6], was discussed as an important topic in financial applications in the recent paper [12]. There, a Hausman-type test was proposed to compare the mean of the MM-estimate with the mean of the least squares fit. We recall that the Hausman test based on the difference of the two-stage least squares and the ordinary least squares is based on the asymptotic covariance matrix of the difference; the test is meaningful in econometric models with endogenic variables, where the least squares estimator is not consistent [5]. Naturally, deriving an asymptotic test of H_0 requires to derive the asymptotic covariance matrix of $\hat{\beta}^A - \hat{\beta}^B$ and cannot be obtained only as a combination of two individual covariance matrices for $\hat{\beta}^A$ and $\hat{\beta}^B$. Nevertheless, the work [12] compared two estimators that are consistent and asymptotically normal, exploiting known formulas for their asymptotic covariance matrices. Because we do not want to assume asymptotic normality to be available, we resort to a nonparametric bootstrap procedure. In Section 2, a nonparametric bootstrap test of equality of two means of two alternative linear regression estimates is proposed. Its performance over three real economic datasets is presented in Section 3. Section 4 brings conclusions.

2 Comparing two regression estimates

Throughout the paper, we consider the standard linear regression model

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

which may be expressed in the matrix notation as $Y = X\beta + e$. Here, we consider p fixed regressors (predictors) available for the total number of n observations (measurements). In our notation, $\beta = (\beta_1, \ldots, \beta_p)^T$ is the vector of parameters, where the i-th row of X will be denoted as $X_i = (X_{i1}, \ldots, X_{ip})^T$ for $i = 1, \ldots, n$. We assume the errors to be independent identically distributed; such random sampling assumption justifies using nonparametric bootstrap in (1), and we need the assumptions of the so-called classical linear regression model [5], i.e. $\mathsf{E}\, e = 0$, $\mathsf{var}\, e = \sigma^2 \mathcal{I}$ for a $\sigma > 0$, $\mathsf{E}\, X^T e = 0$, and $h(X) = p$. While the consistency of the least squares is already ensured in the classical linear regression under these assumptions, we do not assume additional assumptions required to achieve consistency of robust estimators in (1).

Our aim is to test equality of two regression estimates of β in (1). These two estimates will be denoted as

$$\hat{\beta}^A = (\hat{\beta}_1^A, \ldots, \hat{\beta}_p^A)^T \quad \text{and} \quad \hat{\beta}^B = (\hat{\beta}_1^B, \ldots, \hat{\beta}_p^B)^T, \tag{2}$$

while we typically take one to be the least squares estimator and the other to be one of available robust estimators. We are interested in testing equality of the expectations of these two estimators. Formally, the null hypothesis can be expressed as

$$H_0: \ \mathsf{E}\hat{\beta}_j^A = \mathsf{E}\hat{\beta}_j^B \tag{3}$$

for any of the indexes $j = 1, \ldots, p$, against the corresponding two-sided alternative hypothesis. The true but infeasible expectations naturally depend on the distribution of errors, while we omit any distributional assumptions here, and on the particular contamination of the data (severity of contamination, type of outliers).

Principles of bootstrap (resampling) applicable to confidence intervals as well as hypothesis tests have been well known in econometrics and were thoroughly discussed e.g. in [4] and references presented therein. While the range of commonly used bootstrapping approaches is quite broad, we rely here on the nonparametric bootstrap, which is known to perform well in regression modeling tasks [4]. A confidence interval for $(\hat{\beta}^A - \hat{\beta}^B)$ is simply obtained from the bootstrap distribution for the random variable $(\hat{\beta}^A - \hat{\beta}^B)$. The approach for bootstrap-based testing of equality of the means of two alternative regression estimates is formally proposed in Algorithm 1. Algorithm 1 exploits the effective idea to select the bootstrap samples once for all j and only afterwards to construct the confidence intervals separately for every particular $j = 1, \ldots, p$. We recall that consistency of both $\hat{\beta}^A$ and $\hat{\beta}^B$ already ensures (3) to hold. The presented nonparametric bootstrap test is inspired by [8], where nonparametric bootstrap estimation for robust regression was discussed and presented with an algorithm.

3 Experiments

We consider three real publicly available datasets. These were selected as datasets, where it is meaningful to explain a continuous response by a linear regression model using several regressors. From the original data, we however keep only continuous regressors, omitting all discrete ones. Let us first describe these datasets, which all have an economic background and do not contain any missing values.

Algorithm 1 Nonparametric bootstrap test of H_0 (3) of equality of means of $\hat{\beta}_j^A$ and $\hat{\beta}_j^B$ for a given $j \in \{1, \ldots, p\}$

Input: Data rows $(X_{i1}, \ldots, X_{ip}, Y_i)$, $i = 1, \ldots, n$
Input: $j \in \{1, \ldots, p\}$
Input: $S > 0$
Output: Decision function of the test of H_0 in (3)
1: **for** $s = 1$ to S **do**
2: Generate n bootstrap samples

$$({}_{(s)}X_{j1}^*, \ldots, {}_{(s)}X_{jp}^*, {}_{(s)}Y_j^*), \quad j = 1, \ldots, n, \tag{4}$$

by sampling with replacement from $(X_{i1}, \ldots, X_{ip}, Y_i)$, $i = 1, \ldots, n$.
3: Consider a linear regression model in the form

$$_{(s)}Y_j^* = {}_{(s)}\gamma_0 + {}_{(s)}\gamma_{1(s)}X_{j1}^* + \cdots + {}_{(s)}\gamma_{p(s)}X_{jp}^* + {}_{(s)}v_j, \quad j = 1, \ldots, n, \tag{5}$$

with random errors ${}_{(s)}v_1, \ldots, {}_{(s)}v_n$.
4: Compute the estimator $\hat{\beta}_A$ of β in (5).
5: Compute the estimator $\hat{\beta}_B$ of β in (5).
6: $\hat{\theta}_s^j := (\hat{\beta}_j^A - \hat{\beta}_j^B)$.
7: **end for**
8: Arrange the values $\hat{\theta}_1^j, \ldots, \hat{\theta}_S^j$ in ascending order as

$$\hat{\theta}_{(1)} \leq \cdots \leq \hat{\theta}_{(S)}. \tag{6}$$

8: Construct the 95 % confidence interval as

$$\left[\hat{\theta}^j_{(\lfloor h \rfloor)}, \hat{\theta}^j_{(\lfloor n-h \rfloor)}\right], \tag{7}$$

where $h = \lfloor 0.025n \rfloor$ and $\lfloor x \rfloor$ denotes the greatest integer smaller or equal to x.
9: Reject H_0 (3) if and only if the confidence interval (7) does not cover 0.

The datasets represent important bench- marking data well known in robust statistics, as they contain outliers and robustness is known to be meaningful and beneficial for their modeling.

- Cirrhosis dataset with $n = 46$ and $p = 4$ available e.g. in [21]. The death rate from cirrhosis is considered in individual U.S. states. This response is explained by
  - X_1 = percentage of urban population,
  - X_2 = number of late births,
  - X_3 = wine consumption per capita, and
  - X_4 = consumption of hard liquor per capita (X_4).
- Education dataset with $n = 50$ and $p = 3$ contained e.g. in the package [22]. Per capita expenditures on public education are considered in the 50 U.S. states. This response is explained by
  - X_1 = number of residents in urban areas,
  - X_2 = per capita personal income, and
  - X_3 = percentage of individuals below 18 years of age in the population.
- Pasture dataset with $n = 67$ and $p = 3$ available e.g. in [21]. The rental price of pastures in different places in Minnesota is considered. This response is explained by
  - X_1 = rent of arable land,
  - X_2 = number of milk cows per square mile, and
  - X_3 = difference between pasturage and arable land.

3.1 Robust estimators

The least squares estimator, which is used as the reference estimator here for comparing with robust estimates, is computed by the function lm of R software. In the computations, we use the following robust estimators.

1. Least trimmed squares (LTS) [18]. In the computations, we use the function ltsReg of [22] with the trimming constant h = ⌊3n/4⌋. Properties of the LTS were derived in [23].
2. LTS-RLS, which denotes the LTS estimator accompanied by the reweighted version (reweighted least squares) described in [18]. We use again h = ⌊3n/4⌋.
3. MM-estimator with breakdown point equal to 0.5 and with efficiency equal to 0.95. Properties of MM-estimators were derived in [15]. For the computation, we use the function lmrob of [22].
4. LWS-lin, defined as the least weighted squares (LWS) estimator [24] with linearly decreasing weights [8]; properties of the LWS estimator bLWS of β were derived in [24].
5. LWS-log, defined with weights generated by the logistic function [8].
6. LWS-trim, defined with trimmed linear weights [8].
7. LWS-err, defined with weights exploiting the (so-called) error function [8].

Except for the LWS, which remains much less known in the econometric community, the considered robust estimators can be characterized as well established tools. We do not present results of S-estimators on the given data because of numerical instability of their implementation in [22].

3.2 Results

We perform all computations in R software [16] exploiting the robustbase package [22]. Point estimates of the differences $\hat{\beta} - \hat{\beta}^{LS}$ evaluated for the robust estimates of Section 3.1 are presented in Table 1. We do not present results for the intercept, because we understand the test to be meaningful only for the slopes. As revealed in the table, highly robust methods yield quite different results from non-robust methods. Particularly, LTS-RLS seems yield the estimates most different from those of the least squares, while MM-estimators and all versions of the LWS represent more or less a compromise between the least squares and LTS-RLS.

The table presents also nonparametric bootstrap confidence intervals for the differences $\hat{\beta} - \hat{\beta}^{LS}$ obtained always with $S = 1000$ bootstrap samples. Using the bootstrap confidence intervals for hypothesis testing, all obtained intervals cover the value 0. Thus, the bootstrap test yields no significant result on the usual level of 5 %. In other words, we do not find any significant difference between any two estimates. This is true in spite of the already mentioned differences among the point estimates corresponding to different estimation procedures.

The regression estimates that have narrower confidence intervals should be preferable for practical applications. In the cirrhosis dataset, the narrowest confidence intervals are those comparing the mean slopes of LTS and LWS-trim with the slopes of the least squares. In the education dataset, MM-estimator is the best and LWS-log remains only slightly behind. In the pasture dataset, LTS and MM-estimator have the narrowest confidence intervals and all versions of the LWS fall behind.

4 Conclusions

Robust estimators for the linear regression model have already established their position in the analysis of econometric data, although some promising estimators with a high breakdown point remain to be almost unknown to the econometric community. A bootstrap test of equality of the means of two regression estimates is developed in this paper based on a bootstrap confidence interval for the difference between the two estimators. We are particularly interested in the LWS estimator which can be characterized as an estimator with only rare applications; see [24] for the LWS in linear regression or [7] for the LWS in the location model. Still, the proposed bootstrap test can be used to compare any two robust estimators so that its usage is not limited to the LWS estimator.

The numerical study is performed here for three real economic datasets. Rejecting the null hypotheses would require more observations in our datasets, because of relatively large values of variances of individual regression estimates. Actually, significance remains achievable only for a very heavy contamination for $n < 70$, although point estimates subjectively seem very different from each other. This is an interesting result as such: on one hand, robust estimation is typically applied to handle small samples [10], but on the other hand, the variability of robust estimators (e.g. of the LWS estimator) has not been sufficiently investigated (see [8]). The results also reveal the difficulty of reliable regression modeling under small samples.

		Mean difference for the regressor			
Estimator	Intercept	X_1	X_2	X_3	X_4
Cirrhosis dataset ($n = 46$)					
LTS	17.8	0.617	−1.56	0.38	1.44
		[−1.233; 2.477]	[−3.30; 0.18]	[−2.49; 3.25]	[−0.48; 3.36]
LTS-RLS	−3.3	0.438	−1.36	0.04	1.38
		[−1.321; 2.209]	[−3.10; 0.38]	[−2.68; 2.76]	[−0.48; 3.24]
MM	1.9	0.053	−0.13	0.01	0.09
		[−2.095; 2.195]	[−2.22; 1.96]	[−2.94; 2.96]	[−1.90; 2.08]
LWS-lin	19.4	0.619	−1.54	−0.27	−0.03
		[−1.106; 2.344]	[−3.02; 0.34]	[−2.72; 2.25]	[−1.79; 1.73]
LWS-log	3.5	0.349	−0.46	−0.47	−1.38
		[−1.418; 2.113]	[−2.18; 1.26]	[−3.10; 2.16]	[−3.19; 0.43]
LWS-trim	0.4	0.297	−0.25	−0.53	−1.51
		[−1.303; 1.907]	[−1.77; 1.27]	[−2.89; 1.83]	[−3.32; 0.30]
LWS-err	5.0	0.196	−0.47	−0.20	0.23
		[−1.616; 2.015]	[−2.23; 1.29]	[−3.01; 2.61]	[−1.65; 2.11]
Education dataset ($n = 50$)					
LTS	264	0.126	−0.04	−0.56	-
		[−0.372; 0.624]	[−0.16; 0.08]	[−1.93; 0.81]	-
LTS-RLS	309	0.070	−0.02	−0.73	-
		[−0.380; 0.520]	[−0.12; 0.08]	[−1.83; 0.37]	-
MM	278	0.068	−0.02	−0.66	-
		[−0.232; 0.368]	[−0.09; 0.05]	[−1.36; 0.04]	-
LWS-lin	275	0.072	−0.03	−0.64	-
		[−0.278; 0.422]	[−0.11; 0.05]	[−1.44; 0.16]	-
LWS-log	322	0.066	−0.03	−0.78	-
		[−0.264; 0.396]	[−0.11; 0.05]	[−1.58; 0.02]	-
LWS-trim	345	0.075	−0.03	−0.80	-
		[−0.255; 0.405]	[−0.11; 0.05]	[−1.63; 0.03]	-
LWS-err	127	0.028	−0.01	−0.74	-
		[−0.342; 0.398]	[−0.10; 0.08]	[−1.64; 0.16]	-
Pasture dataset ($n = 67$)					
LTS	3.77	−0.102	−0.175	3.91	-
		[−0.278; 0.074]	[−0.478; 0.128]	[−0.80; 8.60]	-
LTS-RLS	4.67	−0.113	−0.104	1.60	-
		[−0.269; 0.043]	[−0.387; 0.179]	[−1.50; 4.70]	-
MM	1.57	−0.056	−0.018	2.03	-
		[−0.162; 0.050]	[−0.242; 0.206]	[−1.07; 5.13]	-
LWS-lin	1.77	−0.086	0.029	0.04	-
		[−0.196; 0.024]	[−0.201; 0.259]	[−3.06; 3.14]	-
LWS-log	3.77	−0.152	−0.048	−1.37	-
		[−1.102; 0.798]	[−0.252; 0.156]	[−4.17; 1.43]	-
LWS-trim	4.40	0.155	0.138	1.69	-
		[−0.805; 1.115]	[−0.072; 0.348]	[−1.21; 4.59]	-
LWS-err	0.77	0.036	0.008	−0.45	-
		[−0.084; 0.156]	[−0.232; 0.248]	[−3.65; 2.75]	-

Table 1 Results of the experiments over three datasets. For each dataset and estimate $\hat{\beta}$, point estimates of $\hat{\beta} - \hat{\beta}^{LS}$ are given, where $\hat{\beta}^{LS}$ is the least square estimate of the true β. Bootstrap confidence intervals for the estimated differences $\hat{\beta} - \hat{\beta}^{LS}$ are also given.

We can make a general conclusion that robust statistics has shifted since its origins in direction to self-standing efficient methods. A comparison of point estimates (without any hypothesis test) based entirely on a visual inspection of the presented tables was common in early books on robustness (such as [18]), but such approach becomes outdated and we recommend to consider point estimates to be always accompanied by bootstrap estimates of their variability. Such estimates were presented e.g. in [8] however only for data with $p = 1$. It is also necessary to mention that bootstrap as a computational technique, helpful in solving various practical questions (e.g. estimating the covariance matrices of regression estimators), has not been so much acknowledged in theoretical approaches to robust statistics [6].

Higher attention of robust statisticians should be paid to methods for high-dimensional data [3]. The robustbase package [22] for robust statistical methods contains mainly datasets that are even smaller than the datasets analyzed here. We intend to perform further computations on larger data as well as to extend the bootstrap test procedure to robust multivariate (or high-dimensional) estimators [14] or robust neural networks [20], for which there are no available results on consistency.

Acknowledgements

This work was supported by the Czech Science Foundation grant GA21-05325S (“Modern nonparametric methods in econometrics”). The authors would like to thank Barbora Peštová for discussion about the nonparametric bootstrap procedure. Jan Tichavský´ helped with the implementation and a reviewer provided helpful comments.

References

[1] Cerioli, A., Riani, M., Atkinson, A.C., & Corbellini, A. (2018). The power of monitoring: How to make the most of a contaminated multivariate sample. *Statistical Methods & Application*, 27, 559–587.
[2] Čížek, P. (2016). Generalized method of trimmed moments. *Inference*, 171, 63–78.
[3] Filzmoser, P. & Todorov, V. (2011). Review of robust multivariate statistical methods in high dimension. *Analytica Chimica Acta*, 705, 2–14.
[4] Godfrey, L. (2009). *Bootstrap tests for regression models*. London: Palgrave Macmillan.
[5] Greeene, W.H. (2018). *Econometric analysis*. 8th edn. Harlow: Pearson Education Limited.
[6] Jurečková, J., Picek, J. & Schindler, M. (2019). *Robust statistical methods with R*. 2nd edn. Boca Raton: CRC Press.
[7] Kalina, J. (2018). A robust pre-processing of BeadChip microarray images. *Biocybernetics and Biomedical Engineering*, 38, 556–563.
[8] Kalina, J., Tobišková, N. & Tichavský, J. (2019). A nonparametric bootstrap comparison of variances of robust regression estimators. In *Conference Proceedings MME 2019, 37th International Conference Mathematical Methods in Economics*. University of South Bohemia, České Budějovice, 168–173.
[9] Kalina, J., Vašaničová, P. & Litavcová, E. (2019). Regression quantiles under heteroscedasticity and multicollinearity: Analysis of travel and tourism competitiveness. *Ekonomický časopis/Journal of Economics*, 67, 69–85.
[10] Kitchenham, B., Madeyski, L., Budgen, D., Keung, J., Brereton, P., Charters, S., Gibbs, S. & Pohthong, A. (2017). Robust statistical methods for empirical software engineering. *Empirical Software Engineering*, 22, 579–630.
[11] [Kreinovich, V., Sriboonchitta, S. & Huynh, V.N., Eds. (2017). *Robustness in Econometrics*. Cham: Springer.
[12] Maravina, T.A., Martin, R.D. (2022). A Hausman type test for differences between least squares and robust time series factor model betas. *Journal of Mathematical Finance*, 12, 411–434.
[13] Maronna, R.A., Martin, R.D. & Yohai, V.J. (2019). *Robust statistics. Theory and methods*. 2nd edn. Hoboken: Wiley.
[14] Marozzi, M., Mukherjee, A. & Kalina, J. (2020). Interpoint distance tests for high-dimensional comparison studies. *Journal of Applied Statistics*, 47, 653–665.
[15] Omelka, M. & Salibiáan-Barrera, M. (2010). Uniform asymptotics for S- and MM-regression estimates. *Annals of the Institute of Statistical Mathematics*, 62, 897–927.
[16] R Core Team (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna. URL http://www.R-project.org/.
[17] Rousseeuw, P.J. & Hubert, M. (2018). Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery*, 8, e1236.
[18] Rousseeuw, P.J. & Leroy, A.M. (1987). *Robust regression and outlier detection*. New York: Wiley.

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

[20] Rusiecki, A. (2013). Robust learning algorithm based on LTA estimator. *Neurocomputing*, 120, 624–632.

[21] Späth, H. (1992). *Mathematical algorithms for linear regression*. San Diego: Academic Press.

[22] Todorov, V. & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32, 1–47.

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

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

[25] Víšek, J. Á. (2019). S-weighted instrumental variables. In Skiadas, C.H. & Bozeman, J.R., Eds.*: Data analysis and applications 1: Clustering and regression, modeling-estimating, forecasting and data mining*. Vol. 2. London: Wiley, 53–72.