Three contributions to robust regression diagnostics

J. KALINA

Abstract:

Robust regression methods have been developed not only as a diagnostic tool for standard least squares estimation in statistical and econometric applications, but can be also used as self-standing regression estimation procedures. Therefore, they need to be equipped by their own diagnostic tools. This paper is devoted to robust regression and presents three contributions to its diagnostic tools or estimating regression parameters under non-standard conditions. Firstly, we derive the Durbin-Watson test of independence of random regression errors for the regression median. The approach is based on the approximation to the exact null distribution of the test statistic. Secondly, we accompany the least trimmed squares estimator by a subjective criterion for selecting a suitable value of the trimming constant. Thirdly, we propose a robust version of the instrumental variables estimator. The new methods are illustrated on examples with real data and their advantages and limitations are discussed.

Mathematics Subject Classification 2000: 62G35, 62J20, 91B84, 62J05
General Terms: Robust regression, Robust econometrics, Hypothesis testing
Additional Key Words and Phrases: least trimmed squares, instrumental variables, diagnostic tools, autocorrelation

1. ROBUST REGRESSION

We consider the linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + e_i, \quad i = 1, \ldots, n. \tag{1}$$

Because the standard least squares (LS) estimator of $\beta = (\beta_0, \ldots, \beta_p)^T$ is highly vulnerable to the presence of outlying measurements in the data, numerous robust regression methods have been proposed as alternatives to the least squares. Nevertheless, they are sensitive to violations of the assumptions of the linear regression model, i.e. they require the random errors (disturbances) $e_1, \ldots, e_n$ to be independent, homoscedastic, and uncorrelated with the regressors [Jurečková et al. 2012]. Therefore, diagnostic tools for robust regression methods or specific estimation procedures for non-standard situations are highly desirable [Víšek 2004].

Regression median (or least absolute deviation, L_1-estimator) is a robust estimator

The work is supported by the Czech Science Foundation project No. 13-01930S and by the Neuron Fund for Support of Science.

10.1515/jamsi-2015-0013

of β obtained as

$$\arg\min \sum_{i=1}^{n} \left| Y_i - b_0 - b_1 X_{1i} - \cdots - b_p X_{pi} \right|, \tag{2}$$

where the minimization is considered over all estimators $\mathbf{b} = (b_0, \ldots, b_p)^T$ of β. It can be interpreted as the regression quantile with the parameter $\alpha = 0.5$ [Koenker 2005]. Thus, it divides the set of the disturbances to a half of values below the regression median and the remaining half of values above the regression median. Still, diagnostic tools are not available for the regression median. We use the asymptotic representation given by Jurečková and Sen [1996] to derive the asymptotic behavior of the Durbin-Watson test statistic computed with residuals of the regression median. The Durbin-Watson test for the regression median is presented in Section 2.

Section 3 will present a new method for selecting a suitable value of the trimming constant h for the least trimmed squares (LTS), which is one of highly robust regression estimators [Rousseeuw and Leroy 1987]. So far, there has been no criterion for a sophisticated choice of the number h of observations truly used for the computation, while the remaining $n-h$ observations are ignored.

The least weighted squares (LWS) estimator for the model (1) generalizes the LTS [Víšek 2011] based on implicit weighting of individual observations, down-weighting but not trimming away potential outliers. If the estimator is computed with the data-dependent adaptive weights of Čížek [2011], it attains a 100 % asymptotic efficiency of the least squares under Gaussian errors. At the same time, the estimator has a high breakdown point, i.e. a high resistance against noise or outliers in the data [Huber and Ronchetti 2009], and is robust to heteroscedasticity [Víšek 2011]. Just like other methods based on ranks have suitable properties in a variety of situations [Murakami 2014], the idea of the LWS, i.e. the implicit weights based on ranks of residuals, is successful in a variety of recent applications including a robust correlation coefficient [Kalina 2012] or robust dimensionality reduction [Kalina and Schlenker 2015]. We use the idea of implicit weights assigned to individual observations to define a robust instrumental variables estimator in Section 4.

Throughout the paper, the novel methods will be illustrated on an investment data set of the time series of real gross domestic product (GDP) and real gross private domestic investment (INVESTMENT) in the United States in the years from 1980 to 2001. Both variables are expressed as multiples of 10^9 of dollars. The data, which are adjusted for deflation of money for the sake of interpretation, come from the website www.stls.frb.org/fred of the U.S. Department of Commerce. The data were routinely

analyzed by means of robust regression by Kalina [2011].

2. DURBIN-WATSON TEST FOR REGRESSION MEDIAN

This section is devoted to the Durbin-Watson test for regression quantiles. The observations in (1) are assumed to be observed in equidistant time moments. The test considers the null hypothesis H_0 of autocorrelated errors $\mathbf{e} = (e_1, \ldots, e_n)^T$ in (1) against the one-sided alternative H_1 that the errors form an autoregressive process AR(1) with parameter $\rho > 0$. The Durbin-Watson test in the original form for least squares regression was proposed by Durbin and Watson [1950].

We assume the assumptions of Jurečková and Sen [1996] for the asymptotic representation for the regression median to be valid. Using the notation of above, let us further denote a unit matrix of size $n \times n$ by $\mathscr{I}_n$, let us define

$$\mathbf{M} = \mathscr{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{3}$$

and

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}, \tag{4}$$

where blank spaces represent zeros.

THEOREM 2.1. *Let us consider (1) and let us assume normally distributed errors*

$$\mathbf{e} \sim \mathsf{N}(0, \sigma^2 \mathscr{I}_n). \tag{5}$$

Let the statistic

$$\tilde{d} = \frac{\sum_{t=2}^{n} (\tilde{u}_t - \tilde{u}_{t-1})^2}{\sum_{t=1}^{n} \tilde{u}_t^2} = \frac{\tilde{\mathbf{u}}^T \mathbf{A} \tilde{\mathbf{u}}}{\tilde{\mathbf{u}}^T \tilde{\mathbf{u}}} \tag{6}$$

of the Durbin-Watson test be computed using residuals $\tilde{u}_1, \ldots, \tilde{u}_n$ of the regression median. Under H_0, the statistic d is asymptotically equivalent in probability to

$$\frac{\mathbf{e}^T \mathbf{M} \mathbf{A} \mathbf{M} \mathbf{e}}{\mathbf{e}^T \mathbf{M} \mathbf{e}}. \tag{7}$$

The proof follows from the asymptotic considerations of Kalina [2011]. The test statistic (7) is exactly the Durbin-Watson statistic for the least squares regression, which suggests how to compute the asymptotic test. The null distribution of d depends on $\mathbf{M}$ (and thus on $\mathbf{X}$). Upper and lower bounds for the critical value for the least squares are tabelated. They are asymptotically valid for the regression median. It is possible to approximate the exact p-value of the test for the regression median in the following way.

THEOREM 2.2. *The approximative p-value of the Durbin-Watson test statistic $\tilde{d}$ for the regression median based on the approximation of Theorem 2.1, assuming (5) in (1), defined as the probability*

$$P_{H_0}\left[\frac{\mathbf{E}^T\mathbf{MAME}}{\mathbf{E}^T\mathbf{ME}} < \tilde{d}\right], \tag{8}$$

converges to the p-value of the exact test for least squares for $n \to \infty$, where $\mathbf{E} = (E_1,\ldots,E_n)^T$ are independent random variables with $\mathsf{N}(0,1)$ distribution.

To illustrate the method, we consider the model

$$\text{INVESTMENT}_t = \beta_0 + \beta_1 \cdot \text{GDP}_t + e_t, \quad t = 1,\ldots,n = 22. \tag{9}$$

for the investment data of Section 1. We compute several robust regression estimators. Estimates of β_0 and β_1 are shown in Table 2. In order to verify the assumption of uncorrelated regression errors, the Durbin-Watson test statistic (6) is computed for the least squares as well as other estimators based on their residuals. Besides the regression median, we perform the computations with the least squares, LTS with the optimal h according to Section 3, LWS with linearly decreasing weights, and trimmed least squares (TLS), which is computed as the least squares regression with those observations which are in the "middle" part of the data between the first and third regression quartile. The critical value of the Durbin-Watson test is asymptotically valid, which was shown for the LTS by Víšek [2004] and for the LWS and TLS by Kalina [2014].

The critical value lies between tabelated lower and upper bound, i.e. in the interval [1.24; 1.43]. Based on 1000 simulation we compute the exact critical value (7) to be 1.24. The values of the test statistic are very similar for all regression estimators in the table and the Durbin-Watson test is highly significant in each case.

The asymptotic Durbin-Watson test turns out to be an important diagnostic tool,

	LS	LTS ($h = 19$)	LWS	TLS	Regression median
Intercept	−582	−371	−465	−526	−517
Slope	0.239	0.203	0.221	0.231	0.230
D-W statistic	0.418	0.351	0.408	0.420	0.418
Asymptotic critical value of the D-W test	1.24	1.24	1.24	1.24	1.24

Table I. Various regression estimates computed in the linear regression model for the investment data. Results of the Durbin-Watson test of Section 2.

Regression method	Est. of β_0	Est. of β_1
Least squares (i.e. LTS, $h = 22$)	-582	0.239
LTS, $h = 21$	−568	0.238
LTS, $h = 20$	−561	0.238
LTS, $h = 19$	−371	0.203
LTS, $h = 18$	−369	0.204
LTS, $h = 17$	−375	0.207
LTS, $h = 16$	−172	0.171
LTS, $h = 14$	−240	0.182
LTS, $h = 15$	−216	0.177
LTS, $h = 13$	−252	0.184

Table II. Various regression estimates computed in the linear regression model for the investment data. Selection of a suitable trimming constant for the LTS (Section 3).

which can be easily performed using the classical critical value or p-value of the least squares regression. On the other hand, these are valid only asymptotically and are derived under the assumption of normally distributed random errors in (1).

3. TRIMMING CONSTANT FOR THE LEAST TRIMMED

We illustrate our subjective method for selecting a suitable value of h for the LTS on the example with investment data with $n = 22$. The following code in R software introduces the data and shows the method for computing the LTS estimator for various values of h:

```
x=c(4900.9, 5021, 4919.3, 5132.3, 5505.2, 5717.1, 5912.4, 6113.3,
6368.4, 6591.8, 6707.9, 6676.4, 6880, 7062.6, 7347.7, 7543.8,
7813.2, 8159.5, 8508.9, 8856.5, 9224, 9333.8);
y=c(655.3, 715.6, 615.2, 673.7, 871.5, 863.4, 857.7, 879.3, 902.8,
936.5, 907.3, 829.5, 899.8, 977.9, 1107, 1140.6, 1242.7, 1393.3,
1558, 1660.1, 1772.9, 1630.8);
n=length(x);
```

```
library(robustbase);
FitLeastSquares=lm(y~x);
for (h in 13:22) FitLTS=ltsReg(x,y,alpha=h/n);
```

Table II presents estimates of β_0 and β_1 obtained with least squares and LTS. The common approach is to put h to be equal to

$$h = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+1}{2} \right\rfloor, \tag{10}$$

where $\lfloor a \rfloor$ denote the integer part of a [Rousseeuw and Leroy 1987]. Such choice ensures the highest possible breakdown point (asymptotically up to 50 %), but has the disadvantage of ignoring a too large number of observations. This explain why the LTS is habitually used only as a diagnostic tool, without being very efficient due to a loss of relevant information. However, the value of h can be improved to reflect the true level of data contamination.

Let us now describe our approach for the selection of such suitable h. The computation of LTS is performed for each h between (10) and n. The optimal (but unknown) value of h will be denoted as h^*. Values of $\mathbf{b}_{LTS}$ happen to be quite similar to each other for $h < h^*$. The same is true also for estimates of the variance σ^2 of the random regression errors. However, these values appear to be very different for values of h exceeding h^*, i.e. in the situation with the value of $(n-h)/n$ below the level of the contamination of the data. In our example, the value $h = 19$ seems to be a suitable estimate of h^*. The least squares estimator correspond to the LTS with $h = n = 22$. We can observe values of β_1 and β_0 to be very different from values obtained for all smaller values of h.

The results brings arguments in favor of the choice $h = 19$, while this subjective approach is not guaranteed to be helpful for each particular data set.

4. LEAST WEIGHTED INSTRUMENTAL VARIABLES

The instrumental variables (IV) estimation is a regression approach popular in econometrics, which is helpful in model (1) with errors correlated with the regressors, if there are some instrumental variables $\mathbf{Z}_i = (Z_{1i}, \ldots, Z_{li})^T$ available for the each of the observations with index $i = 1, \ldots, n$ [Greene 2011]. Here, l as the number of instruments must fulfil $l \geq p$. Because it is based on the least squares estimation, it is highly vulnerable with respect to outlying measurements. A robust instrumental variables estimator was proposed by Víšek [2006]. However, it requires exactly $l = p$. In this section, we present a very different approach to robust

instrumental variables, which can be computed also for the general case with $l \geq p$. Only in a very special case, this formula reduces to the weighted instrumental variables estimator of Víšek [2006].

We propose the least weighted instrumental variables (LWIV) estimator as a following two-stage procedure, where both stages ensure the robustness of the result by means of implicitly assigned weights to individual observations. The LWS estimator, which was described in Section 1, can be recommended again with the adaptive weighting scheme of Čížek [2011].

(1) The LWS regression is used in the linear regression $\mathbf{X} = \mathbf{Z}\gamma + \mathbf{v}$ with some parameters $\gamma = (\gamma_1, \ldots, \gamma_l)^T$ and random errors $\mathbf{v} = (v_1, \ldots, v_n)^T$. Let $\mathbf{W}_1$ denote the diagonal matrix containing the weights in the optimal permutation (i.e. determined by the LWS). The projections $\hat{\mathbf{X}}$ are obtained by

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\gamma} = \mathbf{Z}(\mathbf{Z}^T\mathbf{W}_1\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{W}_1\mathbf{X}). \tag{11}$$

(2) The LWS regression is used in the linear regression of $\mathbf{Y}$ against $\hat{\mathbf{X}}$. Denoting the diagonal matrix containing the weights in the correct permutation (optimal permutation determined by the LWS) by $\mathbf{W}_2$, the LWIV estimator of β is computed as

$$\mathbf{b} = \left(\hat{\mathbf{X}}^T\mathbf{W}_2\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}^T\mathbf{W}_2\mathbf{Y}. \tag{12}$$

Assuming $l = p$, the LWIV estimator coincides with the proposal of Víšek [2006]. For this special case, we have previously illustrated the performance of the method on real data [Kalina 2011]. Now we give another example for $l > p$, which reveals that the LWIV estimator must be used with care. The estimator does not always yield a clearly interpretable result even in simple situations, which is however not a limitation of the robust approach itself, but rather a controversy of the instrumental variables estimation in general.

As an example, we consider the model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1}, \quad t = 1, \ldots, n. \tag{13}$$

with the investment data of Section 1. Table 4 presents estimates of β in (13) for the least squares, LWS with data-adaptive weights, classical IV estimator, and LWIV estimator. Both the least squares and the LWS estimators correspond well to the data and their values are similar to each other. However, because the disturbances in the model are autocorrelated, it can be recommended to use X_{t-1} as instrument for Y_{t-1} and to retain X_t as an independent variable in the model. In other words, Y_{t-1} is

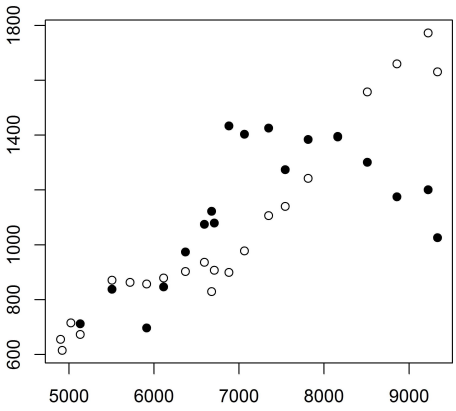

Fig. 1. Raw data ('○') and fitted values using the LWIV estimator ('•') in the example of Section 4.

	Estimate of		
Method	β_0	β_1	β_2
Least squares	−260	0.102	0.611
LWS with adaptive weights	−37.2	0.140	0.437
Instrumental variables	−1723	0.689	−1.939
LWIV	−1438	0.684	−2.211

Table III. Various regression estimates of β_0, β_1 and β_2 in the model (13) computed for the investment data. Comparison of various instrumental variables estimators (Section 4).

replaced by an instrument, while X_t can be interpreted as an instrument for itself. Thanks to a high correlation between the instrument and the regressor Y_{t-1} ($r = 0.958$), such instrument seems promising in terms of efficient estimation. Still, the IV estimate does not give a satisfactory fit due to the ill-conditioned design matrix of the linear regression due to multicollinearity.

Our proposal of a robust instrumental variables estimator allows both a simple computation and a clear interpretation. As robust regression brings improvement compared to least squares, the LWIV estimator brings improvement compared to the classical IV estimator in an analogous way. Possible limitations include the practical problem of finding suitable (not too weak) instruments. Further, we point out that it is necessary to be careful while using the instrumental variables estimator in the robust case. Nevertheless, the problematic results of our example are not a consequence of the robust approach, but rather of the instrumental variables estimation itself.

5. CONCLUSIONS

Numerous robust regression methods have been proposed since the development of robust statistical estimation in 1960s [Huber and Ronchetti 2009]. Nevertheless, each of the methods has been derived under its own assumptions and requires to be equipped by diagnostic tests or specific estimation procedures for non-standard situations. This paper fills the gap of such accompanying tools for highly robust regression, which can be useful particularly in econometric applications.

In Section 2, the asymptotic Durbin-Watson test is proposed for the regression median as a test of autocorrelation of random errors in the linear regression model. The null distribution of the test statistic computed for the regression median can be approximated by the exact null distribution in a classical case. The exact critical value for the least squares, which can be computed by simulations in statistical software, turns out to be also asymptotically valid for the regression median and the test is rejected if the test statistic is smaller than the critical value. The computation of the p-value or the critical value of the test can be performed by a numerical simulation in the same spirit as the classical Durbin-Watson test.

We accompany the LTS estimator by a subjective criterion for the selection of a suitable trimming constant in Section 3. It is found as an optimal value combining two contradictory requirements, namely a high robustness (in terms of breakdown point) and high efficiency for normal (non-contaminated) data.

In Section 4, the idea of the LWS regression is used to propose a new robust estimator in the instrumental variables model. The resulting LWIV estimator is based on the idea of implicit weighting of the data allowing to down-weight less reliable data points, which brings the advantage of a high efficiency in a model without contamination by outliers.

Examples on real data throughout this paper show advantages and at the same time limitations of our proposals. The common advantage of the procedures is their robustness to the presence of severe outliers in the data, while limitations of the methods have been discussed separately in each section of the paper. The examples also serve as a basis for interpreting the newly proposed methods.

REFERENCES

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

Durbin J., Watson G.S. (1950): Testing for serial correlation in least squares regression I. *Biometrika* **37**, 409 – 428.

Greene W.H. (2011): *Econometric analysis.* Prentice Hall, Upper Saddle River.

Huber P.J., Ronchetti E.M. (2009): *Robust statistics.* 2nd end. Wiley, New York.

Jurečková J., Sen P.K. (1996): *Robust statistical procedures: Asymptotics and interrelations.* Wiley, New York.

Kalina J. (2011): Some diagnostic tools in robust econometrics. *Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Mathematica* **50**, 55 – 67.

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

Kalina J., Vlčková (2014): Highly robust estimation of the autocorrelation coefficient. In Löster T., Pavelka T. (Eds.): *The 8th International Days of Statistics and Economics (MSED 2014)*, Melandrium, Slaný, pp. 588 – 597. 10 – 18.

Kalina J., Rensová D. (2015): How to reduce dimensionality of data: Robustness point of view. *Serbian Journal of Management* **10**, 131 – 140.

Koenker R. (2005): *Quantile regression.* Cambridge University Press, Cambridge.

Murakami H. (2014): Power comparison of multivariate Wilcoxon-type tests based on the Jurečková–Kalina's ranks of distances. *Communications in Statistics Simulation and Computation* **44**, 2176 – 2194.

Rousseeuw P.J., Leroy A.M. (1987): *Robust regression and outlier detection.* Wiley, New York.

Víšek J.Á. (2004): Durbin-Watson statistics in robust regression. *Probability and Mathematical Statistics* **23**, 435 – 483.

Víšek J.Á. (2006): Instrumental weighted variables. *Austrian Journal of Statistics* **35**, 379 – 387.

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

Jan Kalina
Institute of Computer Science of the Czech Academy of Sciences
Pod Vodárenskou věží 2
CZ-182 07 Praha 8
Czech Republic
e-mail: kalina@cs.cas.cz