Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica

Jan Kalina
Some Diagnostic Tools in Robust Econometrics

Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica, Vol. 50 (2011), No. 2, 55--67

Persistent URL: http://dml.cz/dmlcz/141754

Acta Univ. Palacki. Olomuc., Fac. rer. nat.,
Mathematica **50**, 2 (2011) 55–67

Some Diagnostic Tools in Robust Econometrics*

Jan KALINA

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

Dedicated to Lubomír Kubáček on the occasion of his 80th birthday

(Received March 31, 2011)

Abstract

Highly robust statistical and econometric methods have been developed not only as a diagnostic tool for standard methods, but they can be also used as self-standing methods for valid inference. Therefore the robust methods need to be equipped by their own diagnostic tools. This paper describes diagnostics for robust estimation of parameters in two econometric models derived from the linear regression. Both methods are special cases of the generalized method of moments estimator based on implicit weighting of individual observations. This has the effect of down-weighting less reliable observations and ensures a high robustness and low sub-sample sensitivity of the methods. Firstly, for a robust regression method efficient under heteroscedasticity we derive the Durbin–Watson test of independence of random regression errors, which is based on the approximation to the exact null distribution of the test statistic. Secondly we study the asymptotic behavior of the Durbin–Watson test statistic for the weighted instrumental variables estimator, which is a robust analogy of the classical instrumental variables estimator.

Key words: robust regression, autocorrelated errors, heteroscedastic regression, instrumental variables, least weighted squares

2010 Mathematics Subject Classification: 62G35, 62J20, 62P20

1 Introduction

This paper is devoted to robust estimation in two recently proposed econometric models. One method is a modification of the linear regression model taking into account heteroscedasticity. The other method is the weighted instrumental variables estimator allowing to estimate parameters in the linear regression

*Supported by the grant 402/09/0557 of the Grant Agency of the Czech Republic.

model under the assumption that certain instruments are available, which are not a part of the explanatory equation but still can be exploited to improve the estimation. For the robust version of Cragg's (1983) approach to heteroscedastic regression we propose the Durbin–Watson test of independence of random regression errors against the alternative hypothesis of autocorrelation, which is computed as an exact test. For the instrumental weighted variables estimator we study an asymptotic approximation to the exact distribution of the Durbin–Watson test statistic.

Fitting linear regression to noisy data is a very common task since all real data are contaminated. The most usual estimation method in linear regression is the least squares method, which is very vulnerable with respect to outliers. Also it is suitable only for models with normally distributed random regression errors. Therefore highly robust statistical methods have been developed with the ability to detect outliers and the concept of the breakdown point has become a crucial robustness criterion, which is a statistical measure of global sensitivity against outliers in the data (see Jurečková and Picek, 2006). A popular estimator is the least trimmed squares estimator proposed by Rousseeuw and Leroy (1987), which has the maximal possible robustness in terms of the breakdown point, but suffers from local shift-sensitivity to small deviations in the center of data. Its weighted analogy is the least weighted squares regression proposed by Víšek (2001), which will be defined in Section 2. It down-weights less reliable observations and does not require to decide definitely if a particular observation is an outlier or not.

Robust estimation in econometric modifications of the basic linear regression model has been studied only recently. Such robust methods are considered reasonable which fulfill the requirements of accurate predictions, clear interpretation, high robustness in terms of the breakdown point, stability and low bias in parameter estimation or hypotheses testing. Sakata and White (2001) used S-estimation in econometric nonlinear regression. Ortelli and Trojani (2005) studied a robust version of the efficient method of moments, which is suitable for time series in a general context including latent nonlinear dynamics. Gagliardini *et al.* (2005) investigated test statistics based on the robust generalized method of moments estimation. Víšek (2005) proposed the robust generalized method of moments estimation and exploited the asymptotic theory of robust estimators developed by Jurečková and Sen (1996).

The econometric methods of the current paper are robustifications of two special cases of the generalized method of moments (GMM) estimator described by Hansen (1982). This is a general tool for statistical estimation given by orthogonality conditions and is defined in a very abstract way for a general parametric situation allowing for over-identification, in other words for situations with more conditions than parameters of the model. Wooldridge (2001) showed the connection to the classical method of moments, so that the GMM estimator can be defined by means of moment conditions. We consider such robust versions, which are based on implicit weighting of individual observations. This allows to down-weight less reliable observations.

Section 2 of this paper recalls the least weighted squares estimator. The remaining parts of the paper assume the data to be observed in equidistant time intervals. Section 3 recalls the Durbin–Watson test for the least squares. Although the test in the classical implementation is inconclusive for certain values of the test statistic below the lower and upper bounds for the critical value, we explain how to approximate the exact critical value for any value of the test statistic. Section 4 is devoted to robust regression efficient under heteroscedasticity, for which the Durbin–Watson test statistic is examined and its exact null distribution is approximated. Section 5 studies the weighted instrumental variables estimator based on the idea of the least weighted squares, for which we propose an approximation to the exact null distribution of the Durbin–Watson test statistic. Finally the conclusions of the paper are summarized in Section 6.

2 Least weighted squares regression

We consider the linear regression model

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

which can be rewritten in the standard matrix notation as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Here the i-th row of $\mathbf{X}$ is the vector $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T$ corresponding to the i-th observation for $i = 1, \ldots, n$.

The least weighted squares (LWS) regression estimates the regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ in the linear regression model (1). It is a highly robust regression method proposed by Víšek (2001). Its main idea of down-weighting less reliable data points is the basis for the definition of two robust econometric methods of Sections 4 and 5. The motivation for the LWS estimator is to down-weight less reliable data points. This does not require to specify if particular observations are outliers or not. Nevertheless Rousseeuw and Leroy (1987) documented that the outlier detection in the linear regression is a hard problem. Before the actual computation of the estimator only the magnitudes of nonnegative weights $w_1, w_2, \ldots, w_n$ must be specified. One of possible choices is to apply linearly decreasing weights

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

which fulfill $\sum_{i=1}^n w_i^L = 1$.

The weights are assigned to the data in an implicit way, namely after a permutation, which is determined automatically only during the computation based on the residuals

$$u_i(\mathbf{b}) = y_i - b_1 X_{i1} - \cdots - b_p X_{ip}, \quad i = 1, \ldots, n, \tag{3}$$

corresponding to the estimate $\mathbf{b} = (b_1, \ldots, b_p)^T \in \mathbb{R}^p$ of the parameter $\boldsymbol{\beta}$. Let us order the squared residuals

$$u_{(1)}^2(\mathbf{b}) \le u_{(2)}^2(\mathbf{b}) \le \cdots \le u_{(n)}^2(\mathbf{b}). \tag{4}$$

The LWS estimator of $\boldsymbol{\beta}$ is defined as

$$\arg\min \sum_{i=1}^{n} w_i u_{(i)}^2(\mathbf{b}) \tag{5}$$

over $\mathbf{b} \in \mathbb{R}^p$.

Čížek (2008) proposed an alternative two-stage procedure for computing data-dependent adaptive weights for the least weighted squares estimator and studied its robustness properties, which turn out to depend on the choice of the weights. The procedure allows to find automatically also the sizes of the weights based on comparing the empirical distribution function of squared residuals with the theoretical distribution function under normality. Such approach yields a high breakdown point combined with a high efficiency of the estimator. Concerning the computation of the LWS estimator, a weighted analogy of the approximative algorithm of Rousseeuw and van Driessen (2006) gives a tight approximation to the true value of the estimate.

The least trimmed squares (LTS) estimator proposed by Rousseeuw and Leroy (1987) represents a special case of the least weighted squares with weights equal either to zero or one. The LTS estimator depends on the value of the trimming constant h $(n/2 < h < n)$, while it is required $\sum_{i=1}^{n} w_i = h$. However the LTS estimator is based on a hard-rejection rule, which needs to determine if each particular observation is an outlier or not. The robustness of the LTS was inspected by Hekimoglu *et al.* (2009).

The advantages of the LWS compared to the LTS include the availability of diagnostic tools (Kalina, 2007), sub-sample robustness or more delicate approach for dealing with moderately outlying values. The LWS estimator does not necessarily solve the outlier detection, although the problem can be solved by comparing the outliers with a suitable robust estimate of the scale of the errors $\mathbf{e}$ using the result of Víšek (2010). Also the correlation coefficient based on the LWS outperforms that based on the LTS in the study of Kalina (2010) analyzing two-dimensional grey-scale images of faces for genetic applications.

3 Durbin–Watson test for least squares

In the linear regression model (1) we assume the data to be observed as a time series in equidistant time intervals. An intercept is not required in the model although it may be present. This work discusses the assumption of independence of the errors $\mathbf{e}$ and its violation. Autocorrelation of the errors $\mathbf{e}$ can lead to an inefficient estimator $\mathbf{b}$ of the regression parameters $\boldsymbol{\beta}$ and biased estimation of $\mathsf{var}\,\mathbf{b}$ and invalid confidence intervals and tests for $\boldsymbol{\beta}$. Also the value of the coefficient of determination R^2 is typically over-estimated if the disturbances are autocorrelated (see Greene, 2002). Therefore we study the Durbin–Watson test proposed in the papers by Durbin and Watson (1950, 1951) for independence of the errors $\mathbf{e}$ against the alternative hypothesis of their autocorrelation.

Let us use the notation $\boldsymbol{\mathcal{I}}_n$ for the unit matrix of size $n \times n$, let us define the matrix $\mathbf{M}$ by $\mathbf{M} = \boldsymbol{\mathcal{I}}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and the matrix $\mathbf{A}$ of size $n \times n$ by

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}. \tag{6}$$

Durbin and Watson (1950, 1951) proposed the test of autocorrelation of errors for model (1). The one-sided test considers the null hypothesis of independent errors $\mathbf{e}$ against the alternative of positive autocorrelation of the first order. Denoting the vector of residuals of the least squares regression by $\mathbf{u} = (u_1, \ldots, u_n)^T$, the one-sided test rejects H_0 for small values of the test statistic, which can be expressed as

$$d = \frac{\sum_{t=2}^{n}(u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2} = \frac{\mathbf{u}^T\mathbf{A}\mathbf{u}}{\mathbf{u}^T\mathbf{u}} = \frac{\mathbf{e}^T\mathbf{MAMe}}{\mathbf{e}^T\mathbf{Me}}. \tag{7}$$

There are three possibilities how to carry out the Durbin–Watson test as a diagnostic tool of the least squares estimator.

1. Compare the test statistic with the lower and upper bounds for the critical values. The tables were created by Durbin and Watson (1950, 1951), who assumed the normal distribution of the errors $\mathbf{e}$ and the presence of an intercept in the model (1). The skeleton decomposition of $\mathbf{M}$ allowed them to use the theorem of Poincaré to find such lower and upper bounds for the critical value which do not depend on $\mathbf{X}$. The test remains inconclusive, if the test statistic lies between the lower and upper bounds.

2. Farebrother (1980) proposed an approximative algorithm for the critical value, which assumes the intercept in the model (1).

3. The exact p-value or the exact critical value can be approximated (with an arbitrary precision) for any value of the test statistic and without the assumption of an intercept in the model (1). Just like Durbin and Watson (1950, 1951) it can be assumed that $\mathbf{e}$ come from the normal distribution with zero expectation. By repeated simulating of $\mathbf{E} = (E_1, \ldots, E_{n-p})^T \sim \mathsf{N}(0, \boldsymbol{\mathcal{I}}_{n-p})$, the exact p-value of the test against the one-sided alternative can be approximated by the empirical probability

$$\mathsf{P}\left[d \leq \frac{\mathbf{E}^T\mathbf{MAME}}{\mathbf{E}^T\mathbf{ME}}\right]. \tag{8}$$

   Because (7) is scale-invariant, the unit variance matrix of $\mathbf{E}$ is valid without loss of generality.

A numerical illustration of the third option is given by Kalina (2007).

4 Robust regression efficient under heteroscedasticity

Cragg (1983) proposed a modification of the least squares regression, which is efficient also under heteroscedasticity. Let us consider the linear regression model in the form (1). Even under heteroscedasticity the estimation using Cragg's transformation is reliable and it is possible to estimate $\boldsymbol{\beta}$ and its variance without testing whether the heteroscedasticity is present in the model (1). A robust version was proposed by Víšek (2005), which is based on the idea of down-weighting less reliable observations. Kalina (2011) suggested an alternative proposal, which is robust with respect to outliers both in the response and the regressors. This approach is studied in this section and we supplement the method with the Durbin–Watson test.

The idea of Cragg's approach is to use some auxiliary variables which could contribute to explaining the variability of the errors $\mathbf{e}$, typically squares of all independent variables from (1) and also products in the form $\mathbf{X}_i\mathbf{X}_j$ for $i \neq j$, where $i, j = 1, \ldots, n$. Therefore we work with the matrix $\mathbf{Q}$ consisting of all columns of $\mathbf{X}$ and the auxiliary variables as additional columns. The model (1) is transformed to

$$\mathbf{Q}^T\mathbf{Y} = \mathbf{Q}^T\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}^T\mathbf{e}, \tag{9}$$

where the regression parameters can be obtained using the least squares estimation. Let us estimate the variance matrix of the errors $\mathbf{e}$ by the diagonal matrix $\hat{\mathbf{S}}$ containing squares of residuals. The generalized least squares estimator of Aitken (1935) of the regression parameters $\boldsymbol{\beta}$ in (9) equals

$$\mathbf{b} = (\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{Y}^*, \tag{10}$$

where $\mathbf{X}^* = \mathbf{Q}^T\mathbf{X}$, $\mathbf{Y}^* = \mathbf{Q}^T\mathbf{Y}$ and $\mathbf{V}^* = \mathsf{var}\,(\mathbf{Q}^T\mathbf{e})$ can be approximated by $\mathbf{Q}^T\hat{\mathbf{S}}\mathbf{Q}$. Finally an estimator of $\mathsf{var}\,\mathbf{b}$ is obtained by

$$\widehat{\mathsf{var}\,\mathbf{b}} = \left(\mathbf{X}^T\mathbf{Q}(\mathbf{Q}^T\hat{\mathbf{S}}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{X}\right)^{-1}. \tag{11}$$

The robust version of Cragg's approach based on implicit weighting replaces the transformation (9) of the model (1) by

$$\mathbf{Q}^T\mathbf{W}\mathbf{Y} = \mathbf{Q}^T\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}^T\mathbf{W}\mathbf{e}, \tag{12}$$

where $\mathbf{W}$ is a weight matrix with weights determined by the least weighted squares in the original model (1). We assume the weights to be strictly positive in order for the following expressions to be correctly defined.

Definition 1 (Robust Cragg's estimator) The method can be described as a two-stage estimator:

1. The least weighted squares regression is used in the model (1). The matrices $\mathbf{W}$ and $\hat{\mathbf{S}}_1$ are obtained.

2. Computing $\mathbf{X}^* = \mathbf{Q}^T\mathbf{W}\mathbf{X}$ and $\mathbf{Y}^* = \mathbf{Q}^T\mathbf{W}\mathbf{Y}$, the estimator of $\boldsymbol{\beta}$ in (12) is obtained as (10) with $\mathbf{V}^*$ is replaced by $\mathsf{var}\,(\mathbf{Q}^T\mathbf{W}\mathbf{e})$, which is estimated by $\hat{\mathbf{V}}^* = \mathbf{Q}^T\mathbf{W}\hat{\mathbf{S}}_1\mathbf{W}\mathbf{Q}$. The matrix $\mathsf{var}\,\mathbf{b}$ is approximated by
$$\widehat{\mathsf{var}\,\mathbf{b}} = (\mathbf{X}^{*T}\hat{\mathbf{V}}^{*-1}\mathbf{X}^*)^{-1}. \tag{13}$$

For this robust version of Cragg's approach we derive the Durbin–Watson test. We assume the model (1) with normal errors and equal variances and the data are assumed to be observed in equidistant time intervals. We compute the Durbin–Watson statistic with weighted residuals of the weighted regression $\sqrt{v_1}u_1, \ldots, \sqrt{v_n}u_n$, where $v_1, \ldots, v_n$ are diagonal elements of $\hat{\mathbf{V}}^*$. The corresponding residuals will be denoted by $\mathbf{u}_{RC}$ to stress that they come from the robust Cragg's fit as
$$\mathbf{u}_{RC} = \mathbf{Y} - \mathbf{X}^*(\mathbf{X}^{*T}\hat{\mathbf{V}}^{*-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\hat{\mathbf{V}}^{*-1}\mathbf{Y}^* = \mathbf{M}_W\mathbf{Y} = \mathbf{M}_W\mathbf{e}, \tag{14}$$
where
$$\mathbf{M}_W = \boldsymbol{\mathcal{I}}_n - \mathbf{X}^*(\mathbf{X}^{*T}\hat{\mathbf{V}}^{*-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\hat{\mathbf{V}}^{*-1}. \tag{15}$$

Such test statistic has the form
$$\frac{\mathbf{u}_{RC}^T\hat{\mathbf{V}}^{*-1/2}\mathbf{A}\hat{\mathbf{V}}^{*-1/2}\mathbf{u}_{RC}}{\mathbf{u}_{RC}^T\hat{\mathbf{V}}^{*-1}\mathbf{u}_{RC}} \tag{16}$$
and the following theorem describes the null distribution of this statistic.

Theorem 1 *Let us denote positive eigenvalues of*
$$\hat{\mathbf{V}}^{*1/2}\mathbf{M}_W^T\hat{\mathbf{V}}^{*-1/2}\mathbf{A}\hat{\mathbf{V}}^{*-1/2}\mathbf{M}_W\hat{\mathbf{V}}^{*1/2} \tag{17}$$
by $\gamma_1, \ldots, \gamma_{n-p}$ *and positive eigenvalues of* $\mathbf{M}^*$ *by* $\lambda_1, \ldots, \lambda_{n-p}$. *Then the following equation holds in distribution*
$$\frac{\mathbf{u}_{RC}^T\hat{\mathbf{V}}^{*-1/2}\mathbf{A}\hat{\mathbf{V}}^{*-1/2}\mathbf{u}_{RC}}{\mathbf{u}_{RC}^T\hat{\mathbf{V}}^{*-1}\mathbf{u}_{RC}} \overset{\mathcal{D}}{=} \frac{\sum_{i=1}^{n-p}\gamma_i E_i^2}{\sum_{i=1}^{n-p}\lambda_i E_i^2} \tag{18}$$
for independent random variables $E_1, \ldots, E_{n-p}$ *with* $\mathsf{N}(0,1)$ *distribution, where* $\overset{\mathcal{D}}{=}$ *denotes equivalence in distribution.*

Proof The matrix (17) is symmetric and positive definite of rank $n-p$. The statistic (16) is scale-invariant. Let us apply the spectral decomposition (see Rao, 1973) on the matrices in both the numerator and the denominator of (16); the principle is to express any square matrix $\mathbf{Z}$ by the decomposition $\mathbf{Z} = \mathbf{Q}^T\boldsymbol{\Lambda}\mathbf{Q}^{-1}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix and the columns of $\mathbf{Q}$ are the eigenvectors of $\mathbf{Z}$. As a result of the decomposition and the general result of Kalina (2007), following the steps of Durbin and Watson (1950, 1951) we obtain the statement of the theorem. □

Theorem 1 can be used to approximate the exact p-value or the exact critical value with an arbitrary precision based on simulating $\mathbf{E} = (E_1, \ldots, E_{n-p})^T \sim \mathsf{N}(0, \boldsymbol{\mathcal{I}}_{n-p})$ in analogy to the least squares (Section 3). The approximative test is based on approximation to the *exact* null distribution of the test statistic, which depends on the weights and also the design matrix $\mathbf{X}$. The p-value of the Durbin–Watson test for the least weighted squares against the one-sided alternative of positive autocorrelation based on (16) assuming normal errors is equal to the probability

$$\mathsf{P}\left[\frac{\sum_{i=1}^{n-p} \gamma_i E_i^2}{\sum_{i=1}^{n-p} \lambda_i E_i^2} \leq \frac{\mathbf{u}_{RC}^T \hat{\mathbf{V}}^{*-1/2} \mathbf{A} \hat{\mathbf{V}}^{*-1/2} \mathbf{u}_{RC}}{\mathbf{u}_{RC}^T \hat{\mathbf{V}}^{*-1} \mathbf{u}_{RC}}\right] \tag{19}$$

with $E_1, \ldots, E_{n-p}, \gamma_1, \ldots, \gamma_{n-p}$ and $\lambda_1, \ldots, \lambda_{n-p}$ defined in Theorem 1.

The test does not require an intercept in the model and can be used for any value of the test statistic. Numerical examples show that 1000 simulations yield very reliable results with standard deviations less than 0.01. Similar results were observed by Kalina (2007) for the Durbin–Watson test for various robust regression methods. From the computational point of view there should be no problem with computing the eigenvalues also for larger data sets, because the QR decomposition (see Rao, 1973) can be exploited.

Table 1: Data.

Time	1	2	3	4	5	6	7	8	9	10
Regressor	6.2	8.1	10.3	12.1	14.1	16.4	18.2	20.1	22.3	24.1
Response	6.1	8.0	10.3	12.1	13.1	14.8	17.9	19.8	19.9	21.6

11	12	13	14	15	16	17	18	19	20
26.1	28.3	30.1	32.3	34.5	36.6	38.0	40.2	42.3	44.7
25.5	25.0	29.3	31.2	33.1	31.8	33.5	38.8	40.7	38.6

Example 1 We illustrate using the Durbin–Watson test as a diagnostic tool for the robust version of the Cragg's approach to heteroscedastic regression. We assume the time series of 20 measurements in equidistant time intervals. The data are described by Maddala (1988) as expenditures data and are examined also by Kalina (2011).

The least squares estimate of $\boldsymbol{\beta}$ is $\mathbf{b} = (0.847, 0.899)^T$ with standard errors $(0.703, 0.025)^T$, which are overestimated due to heteroscedasticity. The Durbin–Watson test statistic is equal to 2.06 and the p-value against the alternative hypothesis of positive autocorrelation equals 0.453. The LWS estimator with Čížek's (2008) adaptive weights estimates $\boldsymbol{\beta}$ by $\mathbf{b}_{LWS} = (0.691, 0.904)^T$. This regression line is very close to the least squares regression line. The asymptotic standard errors of $\mathbf{b}_{LWS}$ are $(0.704, 0.904)^T$.

We use Cragg's approach for the least squares with the square of the income as auxiliary variable contained in the matrix $\mathbf{Q}$. The regression parameters are estimated by $(0.628, 0.910)^T$ with standard errors $(0.298, 0.020)^T$. The estimate of $\boldsymbol{\beta}$ is very similar to the classical least squares, while there is reduction in the variability. The new estimate of $\boldsymbol{\beta}$ is therefore more accurate than the classical estimate, which is deceived by heteroscedasticity. The robust Cragg's approach with the square of the income as auxiliary variable and with data-adaptive weights gives the estimate of $\boldsymbol{\beta}$ equal to $(0.645, 0.906)^T$. The weights turn out to be strictly positive. The standard errors given by (13) are equal to $(0.047, 0.0003)^T$, where the improvement is remarkable compared to asymptotic variance for the least weighted squares. The Durbin–Watson statistic computed with the residuals of this robust Cragg's approach is equal to 2.74, which yields the p-value equal to 0.597. Therefore the test does not the reject the null hypothesis of independence of the errors $\mathbf{e}$ in the model (1).

5 Diagnostics for weighted instrumental variables estimator

The instrumental variables estimator is a popular estimation method in econometrics. In the model (1) it is assumed that the random regression errors $\mathbf{e}$ are not uncorrelated with independent variables, while there is a total number l $(l \geq p)$ of instrumental variables available. We start by recalling the classical instrumental variables estimator, then we focus on the robust version of Víšek (2006) called instrumental weighted variables estimator and derive the asymptotic Durbin–Watson test of independence of the errors $\mathbf{e}$.

In our notation $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{il})^T$, $i = 1, \ldots, n$, denotes the vector of values of the instruments corresponding to the $i-$th observation and the matrix $\mathbf{Z}$ contains these values using the notation $\mathbf{Z} = (Z_{ij})_{ij}$. As in Greene (2002) let us assume that $\mathbf{Z}^T\mathbf{e}/n \xrightarrow{P} \mathbf{0}$ and that $\mathbf{Z}^T\mathbf{X}/n$ converges in probability to a finite regular matrix. These properties explain the motivation for using the instruments. The instrumental variables estimator is defined in the following form as a two-stage estimator of $\boldsymbol{\beta}$ in the model (1), although there exist more general approaches, which allow to find the estimator as a solution of an optimization criterion incorporating orthogonality conditions; see for example Hansen (1982).

1. The projected regressor $\hat{\mathbf{X}}$ is computed as the projection
$$\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{X}$$
of the independent variables on the space of the instruments.

2. The regression parameters $\boldsymbol{\beta}$ are estimated by
$$\mathbf{b}_{IV} = (\hat{\mathbf{X}}^T\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^T\mathbf{Y},$$
which is the least squares estimator of the response against the projected variables $\hat{\mathbf{X}}$.

The fitted values of the response in the model (1) are computed as $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}_{IV}$ and the resulting estimator $\mathbf{b}_{IV}$ is used as estimator of $\boldsymbol{\beta}$ in the original model (1). The special case $l = p$ allows to obtain an explicit solution for estimating the parameters $\boldsymbol{\beta}$ as $\mathbf{b}_{IV} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$, which is equal to the solution obtained by the two-stage approach. However if there are less instruments than regressors, then it is often possible to include some of the regressors to the set of instruments so that the advantages of $l = p$ can be exploited.

We describe the Durbin–Watson test for the residuals

$$\mathbf{u}_{IV} = \mathbf{Y} - \mathbf{X}(\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$$

of the instrumental variables estimator for the special case $l = p$. Let us denote the mean of these residuals as $\bar{u}_{IV}$. In the following text we denote vectors of constants in bold face, such as $\bar{\mathbf{u}}_{IV}$. The statistic

$$(\mathbf{u}_{IV} - \bar{\mathbf{u}}_{IV})^T\mathbf{A}(\mathbf{u}_{IV} - \bar{\mathbf{u}}_{IV})/(\mathbf{u}_{IV} - \bar{\mathbf{u}}_{IV})^T(\mathbf{u}_{IV} - \bar{\mathbf{u}}_{IV}) \tag{20}$$

with the matrix $\mathbf{A}$ defined by (6) is equal to

$$\mathbf{e}^T\tilde{\mathbf{M}}^T\mathbf{A}\tilde{\mathbf{M}}\mathbf{e}/\mathbf{e}^T\tilde{\mathbf{M}}^T\tilde{\mathbf{M}}\mathbf{e}, \tag{21}$$

where $\tilde{\mathbf{M}} = \boldsymbol{\mathcal{I}}_n - \mathbf{X}(\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T$.

The instrumental variables estimator is highly sensitive to outliers, because it is based on the least squares estimation procedure. Víšek (2006) proposed a robust version of the instrumental variables estimator called *instrumental weighted variables* (IWV) estimator and examined its robustness properties, its consistency and asymptotic normality for the special case $l = p$ and also proposed an approximative algorithm for the computation of the IWV estimator. The estimator down-weights observations in the response, regressors and also instruments. Another approach to robustification of instrumental variables estimation based on S-estimators of multivariate location and scatter is described by Cohen–Freue (2011).

Now we propose the asymptotic Durbin–Watson test for the IWV estimator for data observed in equidistant time intervals. Let us denote the empirical distribution function of the absolute values of residuals for any fixed value of the estimator $\mathbf{b}$ of $\boldsymbol{\beta}$ by $F^n_{\mathbf{b}}(\cdot)$. Let us choose a nonincreasing continuous weight function $w\colon [0,1] \to [0,1]$, which determines the sizes of the weights. For a fixed n, the sizes of the weights w_i are obtained as $w_i = w\left(\frac{i-1}{n}\right) - w\left(\frac{i}{n}\right)$ for $i = 1, \dots, n$. The instrumental weighted variables estimator is defined by the system of equations

$$\sum_{i=1}^{n} w\left(F^n_{\mathbf{b}}(|u_i(\mathbf{b})|)\right)\mathbf{Z}_i(Y_i - b_1X_{i1} - \cdots - b_pX_{ip}) = 0. \tag{22}$$

We use the notation F_β for the distribution function of the errors $\mathbf{e}$ with the particular vector $\boldsymbol{\beta}$ of values of the regression parameters and let us further denote

$$\mathbf{Q} = \mathsf{E}\left[w(F_\beta(|e_1|))\mathbf{Z}_1\mathbf{X}_1^T\right]. \tag{23}$$

Víšek (2006) proved the asymptotic representation of $\mathbf{b}_{IWV}$ for $l = p$ under general assumptions, which will be further denoted as Assumptions $\mathcal{A}$, in the form

$$\sqrt{n}\,(\mathbf{b}_{IWV} - \boldsymbol{\beta}) = -\frac{1}{\sqrt{n}}\mathbf{Q}^{-1}\sum_{i=1}^{n} w(F_\beta(|e_i|))\mathbf{Z}_i e_i + \boldsymbol{\eta}, \tag{24}$$

where coordinates of $\boldsymbol{\eta}$ are of order $o_P(1)$. Based on the asymptotic representation for the IWV estimator, we can express the residuals $\mathbf{u}_{IWV}$ of the IWV estimator as

$$\mathbf{u}_{IWV} = \mathbf{Y} - \mathbf{X}\mathbf{b}_{IWV} = \mathbf{e} + \frac{1}{n}\mathbf{X}\mathbf{Q}^{-1}\sum_{i=1}^{n} w(F_\beta(|e_i|))\mathbf{Z}_i e_i + \frac{1}{\sqrt{n}}\mathbf{X}\boldsymbol{\eta} \tag{25}$$

and their mean will be denoted by $\bar{u}_{IWV}$. The asymptotic Durbin–Watson test for the IWV estimator is based on the following theorem, which uses residuals of the IWV estimator for the computation. The technical Assumptions $\mathcal{A}$ of Víšek (2006) are considered.

Theorem 2 *Assuming $l = p$, the test statistic*

$$d_{IWV} = (\mathbf{u}_{IWV} - \bar{u}_{IWV})^T\mathbf{A}(\mathbf{u}_{IWV} - \bar{u}_{IWV})/(\mathbf{u}_{IWV} - \bar{u}_{IWV})^T(\mathbf{u}_{IWV} - \bar{u}_{IWV}) \tag{26}$$

is asymptotically equivalent in probability with (21) under Assumptions $\mathcal{A}$.

Proof Modification of the result of Kalina (2007), by considering the more delicate approximation of (26) in the form

$$(\boldsymbol{\kappa} - \bar{\boldsymbol{\kappa}})^T\mathbf{A}(\boldsymbol{\kappa} - \bar{\boldsymbol{\kappa}})/(\boldsymbol{\kappa} - \bar{\boldsymbol{\kappa}})^T(\boldsymbol{\kappa} - \bar{\boldsymbol{\kappa}}) \tag{27}$$

with $\boldsymbol{\kappa} = \mathbf{e} + \frac{1}{n}\mathbf{X}\mathbf{Q}^{-1}\sum_{i=1}^{n} w(F_\beta(|e_i|))\mathbf{Z}_i e_i$. □

The computation of the Durbin–Watson test for the residuals of the instrumental weighted variables estimator can exploit Theorem 2, which gives an asymptotic approximation to the exact null distribution of the test statistic (26). For the computation we recommend to generate random vectors $\mathbf{E}$ 1000 times following normal distribution with zero expectation and unit variance matrix just like in Section 3. The asymptotic p-value of the test is obtained as the empirical probability

$$\mathsf{P}\left[d_{IWV} \le \frac{\mathbf{E}^T\tilde{\mathbf{M}}^T\mathbf{A}\tilde{\mathbf{M}}\mathbf{E}}{\mathbf{E}^T\tilde{\mathbf{M}}^T\tilde{\mathbf{M}}\mathbf{E}}\right] \tag{28}$$

comparing the test statistic (20) with the test statistic evaluated with random vectors $\mathbf{E}$.

6 Conclusion

This work fills the gap of diagnostic testing theory in robust econometrics. Robust methods were originally proposed as diagnostic tools for standard methods of statistics and econometrics. Modern robust methods attain a high efficiency, which makes them appropriate for self-standing using without being accompanied by the standard (non-robust) counterparts. Such robust methods require to be supplemented by their own diagnostic tools. The arguments in favor of the least weighted squares estimator were published by Víšek (2001) or Kalina (2007). Here we investigate the robustification of Cragg's approach to heteroscedastic regression and instrumental weighted variables. Both robust methods have not been supplemented by any diagnostic tools so far.

The Durbin–Watson test is a standard test of autocorrelation of errors in the linear regression model and this paper studies the classical Durbin–Watson test statistic for two robust econometric models, which are based on the idea of implicit weighting of the data allowing to down-weight less reliable data points. In both cases it turns out that the null distribution of the test statistic can be approximated by the exact null distribution in a classical case. 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 point out that there is only a limited possibility to apply diagnostic tests to robust methods based on trimming, such as the least trimmed squares estimator. Here it would be difficult to find arguments in favor of using all residuals in the test statistics, when some observations are trimmed away completely. Moreover certain methods, such as robust Cragg's estimator, cannot be applied at all because it requires strictly positive weights. Therefore the weighting in the least weighted squares is not just a technical amendment, but a crucial improvement over the popular least trimmed squares estimator, which clarifies the interpretation and permits to introduce essential diagnostic techniques. The Durbin–Watson test proposed for the econometric models of this paper is computationally elegant and can be evaluated as a natural extension of the classical Durbin–Watson test to new robust regression models.

Acknowledgments

The author is thankful to two anonymous referees for valuable comments and tips for improving the paper.

References

[1] Aitken, A. C.: *On least squares and linear combination of observations*. Proc. Roy. Statist. Soc. **55** (1935), 42–48.

[2] Čížek, P.: *Efficient robust estimation of time-series regression models*. Appl. Math. **53** (2008), 267–279.

[3] Cohen-Freue, G. V., Zamar, R. H.: *A robust instrumental variables estimator*. J. Roy. Statist. Soc. (2011), (submitted).

[4] Cragg, J. G.: *More efficient estimation in the presence of heteroscedasticity of unknown form*. Econometrica **51** (1938), 751–763.

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

[6] Durbin, J., Watson, G. S.: *Testing for serial correlation in least squares regression II*. Biometrika **38** (1951), 159–178.

[7] Farebrother, R. W.: *Pan's procedure for the tail probabilities of the Durbin-Watson statistic*. Appl. Stat. **29** (1980), 224–227.

[8] Gagliardini, P., Trojani, F., Urga, G.: *Robust GMM tests for structural breaks*. Journal of Econometrics **129** (2005), 139–182.

[9] Greene, W. H.: Econometric analysis. *Macmillan*, New York, 2002, Fifth edition.

[10] Hansen, L. P.: *Large samples properties of generalized method of moments estimators*. Econometrica **50** (1982), 1029–1054.

[11] Hekimoglu, S., Erenoglu, R. C., Kalina, J.: *Outlier detection by means of robust regression estimators for use in engineering science*. Journal of Zhejiang University Science A **10** (2009), 909–921.

[12] Jurečková, J., Picek, J.: Robust statistical methods with R. *Chapman & Hall/CRC*, Boca Raton, 2006.

[13] Jurečková, J., Sen, P. K.: Robust statistical procedures. Asymptotics and interrelations. *Wiley*, New York, 1996.

[14] Kalina, J.: *On multivariate methods in robust econometrics*. Prague economic papers 2011, (accepted, in print).

[15] Kalina, J.: *Robust image analysis of faces for genetic applications*. Eur. J. Biomed. Inf. **6**, 2 (2010), 6–13.

[16] Kalina, J.: *Asymptotic Durbin-Watson test for robust regression*. Bull. Int. Statist. Inst. **62** (2007), 3406–3409.

[17] Ortelli, C., Trojani, F.: *Robust efficient method of moments*. Journal of Econometrics **128** (2005), 69–97.

[18] Rao, C. R.: Linear methods of statistical induction and their applications. *Wiley*, New York, 1973, Second edition.

[19] Rousseuw, P. J., Leroy A. M.: Robust regression and outlier detection. *Wiley*, New York, 1987.

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

[21] Sakata, S., White, H.: *S-estimation of nonlinear regression models with dependent and heterogeneous observations*. Journal of Econometrics **103** (2001), 5–72.

[22] Víšek, J. Á.: *Robust error-term-scale estimate*. In: Antoch J., Hušková M., Sen P. K. (eds.) Nonparametrics and robustness in modern statistical inference and time series analysis, IMS Collections **7**, *Institute of Mathematical Statistics*, Beachwood, Ohio, 2010, 254–267.

[23] Víšek, J. Á.: *Instrumental weighted variables*. Austrian J. Statist. **35** (2006), 379–387.

[24] Víšek, J. Á.: *Robustifying generalized method of moments*. In: Kupka K. (ed.) Data analysis 2004/II, Progressive methods of statistical data analysis and modelling for research and technical practice, *Trilobyte Statistical Software*, Pardubice, 2005, 171–193.

[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*, 2001, 324–356.

[26] Wooldridge, J. M.: *Applications of generalized method of moments estimation*. J. Econ. Perspect. **15**, 4 (2001), 87–100.