AUTOCORRELATED RESIDUALS OF ROBUST REGRESSION

Jan Kalina

Abstract

The work is devoted to the Durbin-Watson test for robust linear regression methods. First we explain consequences of the autocorrelation of residuals on estimating regression parameters. We propose an asymptotic version of the Durbin-Watson test for regression quantiles and trimmed least squares and derive an asymptotic approximation to the exact null distribution of the test statistic, exploiting the asymptotic representation for both regression estimators. Further, we consider the least weighted squares estimator, which is a highly robust estimator based on the idea to down-weight less reliable observations. We compare various versions of the Durbin-Watson test for the least weighted squares estimator. The asymptotic test is derived using two versions of the asymptotic representation. Finally, we investigate a weighted Durbin-Watson test using the weights determined by the least weighted squares estimator. The exact test is described and also an asymptotic approximation to the distribution of the weighted statistic under the null hypothesis is obtained.

Key words: linear regression, robust statistics, diagnostics, autocorrelation

JEL Code: C52, C14, C12

---

1 Model

In the whole paper, we consider the linear regression model

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

where $Y_1, \dots, Y_n$ are values of a continuous response variable and $e_1, \dots, e_n$ are random errors (disturbances). The task is to estimate the regression parameters $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$. We assume the data in the model (1) to be observed in equidistant time intervals. This work discusses the assumption of independence of the errors and its violation. In econometrics, this violation is most commonly formulated as the autocorrelation of the errors.

The violation of the independence of the errors may have severe negative consequences. Regression parameters β are not estimated efficiently. Denoting the least squares estimator of β by $b = (b_0, b_1, \ldots, b_p)^T$, the classical estimator of var b is biased. This disqualifies using classical hypothesis tests and confidence intervals for β as well as the value of the coefficient of determination R^2. Diagnostic tools checking the assumption of equality of variances of the disturbances can be based on residuals

$$u_i = Y_i - b_0 - b_1 X_{1i} - \cdots - b_p X_{pi}, \quad i = 1, \ldots, n. \tag{2}$$

This work is devoted to robust statistical procedures for estimating β in the model (1). Robust statistical methods have been developed as an alternative paradigm to data modeling tailor-made to suppress the effect of outliers. Modern robust methods can be used not only as a diagnostic tool for classical methods, but can also fully exploited as reliable self-standing procedures. M-estimators are the most widely used robust statistical methods (Jurečková and Picek, 2005). However, M-estimators in linear regression do not possess a high breakdown point (only $1/n$), which is a statistical measure of sensitivity against noise or outliers in the data (Rousseeuw and Leroy, 1987). Highly robust methods are defined as methods with a high breakdown point, which has become a crucial concept in robust statistics. Moreover, original M-estimators of regression parameters require a simultaneous estimation of the error variance, which entails losing computational simplicity.

2 Durbin-Watson test for least squares

The Durbin-Watson test is a classical test of the null hypothesis of homoscedasticity for the least squares estimator in the model (1). Econometricians most commonly consider the test against a one-sided alternative of a positive autocorrelation

$$H_1: e_i = \rho e_{i-1} + v_i, \;\; 0 < \rho < 1, \;\; i = 2, \ldots, n, \tag{3}$$

where ρ is the autocorrelation coefficient of the first order and $v_2, \ldots, v_n$ are independent random errors. Based on this notation, the null hypothesis of uncorrelated errors in the model (1) can be expressed as $H_0: \rho = 0$. Especially if $|\rho|$ is close to 1, the autocorrelation of the errors (3) may have severe negative consequences.

The test proposed by Durbin and Watson (1950) and later corrected by Durbin and Watson (1951) is based on the test statistic

$$d = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{i=2}^{n} u_i^2}, \tag{4}$$

which can be expressed as

$$d = \frac{u^T A u}{u^T u}. \tag{5}$$

The distribution of d depends not only on p and n, but also on the design matrix X. However, lower and upper bounds for critical values have been tabulated. The notation $d_L(\alpha)$ and $d_U(\alpha)$ is used for the lower and upper bound of the test with level α, respectively. The decision rule of the test can be described in the following way:

- Reject H_0, if $d < d_L(\alpha)$.
- Do not reject H_0, if $d > d_U(\alpha)$.
- No conclusion, if $d_L(\alpha) < d < d_U(\alpha)$.

The values of $d_L(\alpha)$ and $d_U(\alpha)$ have been tabulated; see Kmenta (1986) or other econometric textbooks.

There exist several approximations for the exact p-value for a large number of observations. However, every such approximation depends on the design matrix X. Durbin and Watson (1951) proposed to transform d to the interval $(0,1)$; let us denote this transformed d by d^*. Then the distribution of d^* can be approximated by a beta distribution. The authors warned that this approximation is very rough and can be recommended only if $n > 40$. In a later paper, Durbin and Watson (1971) admitted that their approximation is more accurate that they expected and their original caution was excessive. Anyway they recommend first to use tables with critical values and only for an inconclusive result to use their approximation.

There exists also an approximative Durbin-Watson test based on a large-sample normal approximation. This test can be used when the number of observations is so large that tables of $d_L(\alpha)$ and $d_U(\alpha)$ are not available. The idea is based on properties of the AR(1) process, which is an autoregressive process of the first order. The (theoretical) first-order auto-correlation coefficient in AR(1) model is equal to the coefficient ρ. Thus ρ can be estimated by the empirical first-order autocorrelation coefficient

$$\hat{\rho} = \frac{\sum_{i=2}^{n} u_i u_{i-1}}{\sum_{i=1}^{n} u_i^2}. \tag{6}$$

The test rejects H_0 if $\hat{\rho} > \frac{2}{n}$.

The Durbin-Watson test is a general test of misspecification of the model. Although it is sensitive to the autocorrelation of the disturbances, it can give a significant result also for a model with a missing variable. This can be for example the square of one of the independent variables in the model (1). A misspecification of the model can be revealed by a plot of residuals $(1, u_1), \dots, (n, u_n)$, which should be examined carefully before running the Durbin-Watson test.

If it is aimed to test H_0 against negative autocorrelation, the quantity $4 - d$ is usually used as a test statistic and then the decision rule is the same. A test against a two-sided alternative $H_1\colon \rho \neq 0$ is straightforward to be carried out, but not much used in econometric practice.

Cochrane and Orcutt (1949) proposed a method for estimating parameters in the model (1), which adjusts for the autocorrelation of the errors and can be recommended to be carried out if the Durbin-Watson test yields a significant result.

3 Durbin-Watson test for regression quantiles

Regression quantiles represent a natural generalization of samples quantiles to the linear regression model (1). Their theory was studied by Koenker (1995) and their asymptotic representation was derived by Jurečková and Sen (1996). The estimator depends on a fixed value of the 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. Here we describe the asymptotic Durbin-Watson test for regression quantiles, which is derived from the asymptotic representation. The proof of the theorems follows from the asymptotic considerations of Kalina (2007).

Theorem 1. Let the statistic d of the Durbin-Watson test be computed using the residuals of the regression quantile with a parameter α. Then, assuming normal errors $e \sim N(0, \sigma^2 I_n)$ in (1), the statistic d is asymptotically equivalent to

$$e^T MAMe / e^T Me \tag{7}$$

under H_0, where $M = I_n - X(X^TX)^{-1}X^T$ and I_n is a unit matrix of size nxn.

The test statistic (7) is exactly the Durbin-Watson statistic (5) for the least squares regression.

4 Durbin-Watson test for trimmed least squares

The trimmed least squares estimator (TLS) is a robust estimator in the linear regression model (1) based on regression quantiles. Its asymptotic representation derived by Jurečková and Sen (1996) allows to derive an asymptotic Durbin-Watson test for the TLS estimator.

The estimator depends on two parameters. These are fixed values α_1 and α_2 between 0 and 1. Let us denote by $\hat{\beta}(\alpha_1)$ and $\hat{\beta}(\alpha_2)$ the regression quantiles corresponding to the arguments α_1 and α_2. Let us assume $0 < \alpha_1 < \frac{1}{2} < \alpha_2 < 1$. To define the TLS estimator, let us introduce weights $w = (w_1, \dots, w_n)^T$, where w_i is equal to 1, if both conditions are fulfilled:

1. The fitted value of the i-th observation by the quantile regression with argument α_1 is smaller than Y_i;
2. The fitted value of the i-th observation by the quantile regression with argument α_2 is larger than Y_i.

Let W denote the diagonal matrix with diagonal elements $w_1, \dots, w_n$, where w_i is an indicator of the random event that the i-th observation lies between the values $\hat{\beta}(\alpha_1)$ and $\hat{\beta}(\alpha_2)$. The TLS estimator $b_{TLS}(\alpha_1, \alpha_2)$ in the model (1) is defined as

$$b_{TLS}(\alpha_1, \alpha_2) = (X^TWX)^{-1}X^TWY. \quad (8)$$

The asymptotic representation requires the assumption that the errors $e_1, \dots, e_n$ come from a continuous distribution with a density function symmetric around 0.

Theorem 2. Let the statistic d of the Durbin-Watson test be computed using the residuals of the TLS estimator with parameters α_1 and α_2. Then, assuming normal errors $e \sim N(0, \sigma^2 I_n)$ in (1), the statistic d is asymptotically equivalent to (7) under H_0.

5 Least weighted squares

A promising highly robust estimator in the linear regression context seems to be the least weighted squares estimator (Víšek, 2002). It does not remove outliers, but potential outliers are only down-weighted and not trimmed away completely.

We will now recall the least weighted squares (LWS) estimator. First, the user needs to specify a sequence of magnitudes of weights

$$w_1 \le w_2 \le \cdots \le w_n, \tag{9}$$

which are assigned to individual observations only after some permutation. Let us denote the squared residuals corresponding to a given estimator b of β as

$$u_{(1)}^2(b) \le u_{(2)}^2(b) \le \cdots \le u_{(n)}^2(b). \tag{10}$$

The LWS estimator b_{LWS} is defined as

$$\arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} w_i u_{(i)}^2(b). \tag{11}$$

The computation of the LWS estimator is intensive and an approximative algorithm can be obtained as a weighted version of the FAST-LTS algorithm proposed for the least trimmed squares (LTS) regression. The latter represents a special case of least weighted squares with weights equal to zero or one only and its ability to detect outliers was investigated by Hekimoglu et al. (2009). Heteroscedasticity tests for the LWS were studied by Kalina (2011). The LWS estimator has a small local sensitivity compared to the LTS, which suffers from high sensitivity to small deviations near the center of the data. Properties of the LWS estimator were investigated by Víšek (2011) or Kalina (2012).

6 Least weighted squares: asymptotic Durbin-Watson statistic

We recall the asymptotic Durbin-Watson test for the LWS estimator derived by Kalina (2007). Later in Sections 7 and 8, we will derive alternative approaches based on a weighted version of the Durbin-Watson test statistic. We use the notation $u^* = Y - Xb_{LWS}$ for the residuals of the LWS estimator.

Remark 1. The test statistic (4) is invariant with respect to σ^2.

We assume independent normal errors

$$e_i \sim N(0, \sigma^2), \quad i = 1, \dots, n. \tag{12}$$

Under the null hypothesis, the value of σ^2 is arbitrary in the spirit of Remark 1. We will use the asymptotic representation of the LWS estimator presented by Víšek (2011) and assume his technical Assumptions A. The weights $w_1, \dots, w_n$are assumed to fulfil (9). We introduce the notation

$$u^* = e - \frac{1}{\delta} X(X^T X)^{-1} \sum_{k=1}^{n} w_k \sum_{i=1}^{n} e_i X_i I \left\{ \Phi^{-2} \left(\frac{n+k}{2n} \right) \right\} + \frac{1}{\sqrt{n}} X\eta, \tag{13}$$

where coordinates of the remainder term $\eta = (\eta_1, \dots, \eta_p)^T$ are negligible in probability, I denotes an indicator, δ is a constant and Φ^{-1} is the quantile function of the standard normal distribution. We introduce the notation $\tau = \frac{1}{\sqrt{n}} X\eta$ and $\kappa = u^* - \tau$, where κ is known up to the value of the constant δ, while Kalina (2007) proposed its consistent estimator.

Theorem 3. Under H_0, the test statistic (4) is asymptotically equivalent in probability with

$$\frac{\kappa^T A \kappa}{\kappa^T \kappa} \tag{14}$$

under Assumptions A.

The proof is analogous with Kalina (2007).

Theorem 4. The test statistic (4) is asymptotically equivalent in probability with

$$e^T MAMe / e^T Me, \tag{15}$$

where $M = I - X(X^T X)^{-1} X^T$.

The proof is based on Theorem 3 using analogous steps as Kalina (2007).

Finally we explain how to compute the *p*-value or critical value for the asymptotic test, based on simulating from the null distribution of the test statistics (14) or (15).

Theorem 5. The approximative *p*-values of the Durbin-Watson test for the least weighted squares based on (10), or (11) respectively, assuming (12), defined as the probability

$$P\left[\frac{\kappa^T A \kappa}{\kappa^T \kappa} > \frac{u^{*T} A u^*}{u^{*T} u^*} \right], \tag{16}$$

respectively

$$P\left[\frac{E^T MAME}{E^T ME} > \frac{u^{*T} A u^*}{u^{*T} u^*}\right], \tag{17}$$

converge to the p-value of the exact test for the LWS residuals for $n \to \infty$ under Assumptions A, where $E_1, \dots, E_n$ are independent random variables with $N(0,1)$ distribution and $\kappa_1, \dots, \kappa_n$ are computed using the values $E_1, \dots, E_n$.

Remark 2. Without loss of generality, the unit variance of the random errors can be considered in Theorem 5, because the value of δ is scale-invariant, the vectors κ is scale-equivariant and the expressions (14) and (15) also scale-invariant.

7 Least weighted squares: exact weighted Durbin-Watson test

We study the null distribution of the weighted Durbin-Watson statistic for the least weighted squares conditioning on fixed weights. Let $w_1, \dots, w_n$ denote the resulting weights corresponding to individual observations obtained as the output of the LWS estimator, i.e. the weight w_i is assigned to the i-th observation for $i = 1, \dots, n$. Let W denote the diagonal matrix, which contains the weights $w_1, \dots, w_n$ on the main diagonal. The idea of the exact test is to express the LWS residuals $u^* = Y - Xb_{LWS}$ in the form $u^* = M_W e$, where the matrix M_W is defined as

$$M_W = I - X(X^T W X)^{-1} X^T W. \tag{18}$$

Theorem 6. The p-value of the Durbin-Watson test for the LWS estimator against the one-sided alternative of positive autocorrelation based on the weighted Durbin-Watson statistic

$$d_W = \frac{\sum_{i=2}^{n} (\sqrt{w_i} u_i^* - \sqrt{w_{i-1}} u_{i-1}^*)^2}{\sum_{i=2}^{n} w_i u_i^{*2}} = \frac{u^{*T} W^{1/2} A W^{1/2} u^*}{u^{*T} W u^*} \tag{19}$$

assuming

$$e \sim N(0, \sigma^2 W^{-1}) \tag{20}$$

is equal to the probability

$$P\left[\frac{\sum_{i=1}^{n-p} \gamma_i E_i^2}{\sum_{i=1}^{n-p} \lambda_i E_i^2} \le \frac{u^{*T} W^{1/2} A W^{1/2} u^*}{u^{*T} W u^*}\right], \tag{21}$$

where $E_1, \dots, E_{n-p}$ are independent random variables with $N(0,1)$ distribution, $\gamma_1, \dots, \gamma_{n-p}$ are positive eigenvalues of $M_W^T A M_W$, $\lambda_1, \dots, \lambda_{n-p}$ are positive eigenvalues of $M_W^T M_W$.

8 Least weighted squares: asymptotic weighted Durbin-Watson test

We use the asymptotic results described in Section 6 to obtain the asymptotic null distribution of the weighted Durbin-Watson statistic (19) computed from the LWS residuals. Again we use the notation $u^* = Y - Xb_{LWS}$.

Theorem 7. The approximative *p*-values of the Durbin-Watson test for the least weighted squares, assuming (20) and Assumptions A, defined as the probability

$$P\left[\frac{\kappa^T \sqrt{W} A \sqrt{W} \kappa}{\kappa^T W \kappa} > \frac{u^{*T} \sqrt{W} A \sqrt{W} u^*}{u^{*T} W u^*}\right], \tag{22}$$

or using an additional approximation as

$$P\left[\frac{E^T M \sqrt{W} A \sqrt{W} M E}{E^T M W M E} > \frac{u^{*T} \sqrt{W} A \sqrt{W} u^*}{u^{*T} W u^*}\right], \tag{23}$$

converge for $n \to \infty$ to the *p*-value of the exact weighted Durbin-Watson test for the LWS residuals under, where $E_1, \dots, E_n$ are independent random variables with $N(0,1)$ distribution and $\kappa_1, \dots, \kappa_n$ are computed from the values $E_1, \dots, E_n$, using $\kappa = u^* - \tau$ and (13).

Acknowledgement

The research was supported by the grant 13-01930S of the Granty Agency of the Czech Republic (Robust procedures for non-standard procedures, their diagnostics and implementation).

References

1. Cipra T.: *Ekonometrie*. Praha: SPN, 1984. In Czech.
2. Cochrane D., Orcutt G.H.: Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* 44, 32-61, 1949.
3. Durbin J., Watson G.S.: Testing for serial correlation in least squares regression I. *Biometrika* 37, 409-428, 1950.

4. Durbin J., Watson G.S.: Testing for serial correlation in least squares regression II. *Biometrika* 38, 159-178, 1951.
5. Durbin J., Watson G.S.: Testing for serial correlation in least squares regression III. *Biometrika* 58, 1-19, 1971.
6. 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 (6), 909-921, 2009.
7. Jurečková J., Picek J.: *Robust statistical methods with R.* Boca Raton: Chapman & Hall/CRC, 2005.
8. Jurečková J., Sen P.K.: *Robust statistical procedures: Asymptotics and interrelations.* New York: Wiley, 1996.
9. Kalina J.: On multivariate methods in robust econometrics. *Prague Economic Papers* 21 (1), 69-82, 2012.
10. Kalina J.: On heteroscedasticity in robust regression. *International Days of Statistics and Economics Conference Proceedings, September 22-23,2011,* Prague: University of Economics, 228-237, 2011. ISBN 978-80-86175-77-5.
11. Kalina J.: Asymptotic Durbin-Watson test for robust regression. *Bulletin of the International Statistical Institute* 62, 3406-3409, 2007.
12. Koenker R.: *Quantile regression*. Cambridge: Cambridge University Press, 2005.
13. Rousseeuw P.J., Leroy A.M.: *Robust regression and outlier detection.* New York: Wiley, 1987.
14. Víšek J.Á.: Consistency of the least weighted squares under heteroscedasticity. *Kybernetika* 47, 179-206, 2011.
15. Víšek J.Á.: The least weighted squares I. *Bulletin of the Czech Econometric Society* 9 (15), 31-56, 2002.

Contact

RNDr. Jan Kalina, Ph.D.

Institute of Computer Science of the Academy of Sciences of the Czech Republic

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

kalina@cs.cas.cz