International Journal of Mathematics and Statistics Studies
Vol.2, No.2, pp.1-12, June 2014
Published by European Centre for Research Training and Development UK (www.ea-journals.org)

A ROBUST LEAST TRIMMEDD SQUARES FOR AUTOCORRELATED RESIDUALS

Moawad El-Fallah Abd El-Salam
Department of Statistics & Mathematics and Insurance,
Faculty of Commerce, Zagazig University, Zagazig, Egypt,

ABSTRACT: *It is well known that, the classical Durbin-Watson test is the most commonly used regression technique for detecting autocorrelation. However, this test is affected by outliers. Therefore, we cannot both detect the autocorrelation of disturbances and remedy the harm caused by this phenomenon. In this paper, we conjecture how to robustify the Durbin-Watson test for detecting the autocorrelation problem. A description of the least trimmed squares regression and least weighted squares follows . Thus, we can to robustify the Durbin-Watson test for residuals of the least trimmed squares regression. An example with real data supports the practical character of this paper.*

KEYWORDS: Durbin-Watson test; Cochrane-Orcutt transformation; Least Trimmed Squares; Reweighted Least Squares; Durbin-Watson test in robust regression.

INTRODUCTION

Consider the following classical linear regression model as :

$$Y_t = \beta_1 + \beta_2 X_{t1} + \ldots, + \beta_p X_{tp} + e_t \quad , \qquad t = 1, 2, \ldots, n, \quad (1)$$

Which can be rewritten in the usual matrix notation as $Y = X\beta + e$.

The design matrix X consists of p columns, where the first one is a vector (1,1, …, 1)' corresponding to the intercept. From the usual assumption $\text{Var}(c) = \sigma^2 I$, it follows (among others) that the disturbances $e_1, \ldots, e_n$ are independent. Let us assume that the data are observed in equidistant time moments, which is a crucial requirement .

We will model the violation of independence of disturbances by an AR(1) process, namely:

$$e_t = \rho\, e_{t-1} + \nu_t \quad , \qquad t = 2, \ldots, n \quad , \qquad (2)$$

with a parameter $\rho, -1 < \rho < 1$, and uncorrelated random variables $\nu_2, \ldots, \nu_n$ with zero mean and variance σ_v^2 , $0 < \sigma^2 < \infty$. This model is not only simple, but it also turns out that it works well in many practical situations. Judge et al. (1985) describes tests and estimation procedures for models with more complicated behaviour of disturbances .

Let us denote by b to the ordinary least squares estimator of the parameter β. As we are never able to observe disturbances work with residuals , which can be viewed as an estimate of disturbances. This is the reason, why the topic of this paper is often called autocorrelated residuals, when in fact the disturbances are autocorrelated. We hope this cause no confusion. Let us also note that, the independence of disturbances does not the independence of residuals, however, in the model with independent disturbances, the independence of residuals can be expected to be violated only slightly.

The consequences of autocorrelated disturbances can be severe, especially when $|\rho|$ is close to 1. The estimator b loses its efficiency. Worse than that, the classical estimator of Var (b) is on longer unbiased. Therefore, we cannot trust confidence intervals and tests for β. Next, the outline of this paper is as follows. In section (2), we describe the Durbin-Watson test and the Cochrane-Orcutt transformation, so that we can both detect the autocorrelation of disturbances and remedy the harm caused by this phenomenon. Section (3) is devoted to the describtion of the least trimmed squares and least weighted squares, and we will show how to robustify the Durbin-Watson test for residuals of the least trimmed squares regression. An example with real data supports the practical character of this paper is presented in section (4) . Section (5) concludes.

Durbin-Watson Test and Cochrane-Orcutt Transformation :

Durbin-Watson Test:

It is practical to know to test that the disturbances are autocorrelated. Econometricians usually consider the test of independence of disturbances against an alternative of a positive autocorrelation, which means to test :

$$H_0 : \rho = 0 \text{ against}$$
$$H_1 : \text{moodel (2) holds with } \; 0 < \rho < 1.$$

The test proposed by Durbin and Watson (1950) has become a classical test of H_0 against H_1. The test statistic is defined by :

$$d = \frac{\sum_{t=2}^{n} (u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2} \quad , \qquad (3)$$

where $u_1, \ldots, u_n$ are the residuals of the least squares regression. 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 can be found, which are tabulated in most econometric textbooks. The notation $d_L(\alpha)$ and $d_u(\alpha)$ is used for the lower and upper bound of the test with level α. Then the decision rule is the following:

$$d \le d_L(\alpha) \rightarrow \text{reject } H_0$$

$$d > d_u(\alpha) \rightarrow \text{do not reject } H_0$$

$$d_L(\alpha) < d \le d_u(\alpha) \rightarrow \text{no conclusion}$$

There exist several approximations for the exact p value, which 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 with expectation $E(d^*)$ and variance Var (d^*). The econometricians warn that this approximation is very rough and can be recommended only if $n > 40$. In a later paper, Durbin and Watson (1971) admit that their approximation is more accurate than 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. It is good to know that some statistical packages (Such as S-plus) compute this beta approximation of the Durbin-Watson test.

In most textbooks, values of $d_L(\alpha)$ *and* $d_u(\alpha)$ are tabulated for n ranging form 15 to 100. For smaller than 15, Pindyck and Rubinfeld (1991) recommend not to use the Durbin-Watson test at all. But for example, Kmenta (1986) gives tables for $d_L(\alpha)$ and $d_u(\alpha)$ beginning with $n = 6$. It would be fair to admit that the Durbin-Watson test for such small samples is very rough. There exists also an approximate test of

$H_0 : \rho = 0$ against $H_1 : \rho > 0$, 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 that in an AR(1) process, the (theoretical) first – order autocorrelation coefficient is equal to the coefficient ρ. In this respect, ρ can be estimated by the empirical first-order autocorrelation coefficient as :

$$\hat{\rho} = \frac{\sum_{t=2}^{n} u_t \, u_{t-1}}{\sum_{t=1}^{n} u_t^2} \qquad (4)$$

In this case, we reject $H_0 : \rho = 0$ if $\hat{\rho} > 2/\pi$. In addition, 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 $(1, u_1), (2, u_2), \ldots, (n, u_n)$, which should be examined carefully before running the Durbin-Watson test.

Our final notes comment the use of the Durbin-Watson test in different situations. For a test of negative autocorrelation, $4 - d$ is usually used as a test statistic and then the decision rule is the same. A test against a two-sided alternative $H_1 : \rho \neq 0$ is simple to carry out , but not much used in practice .

Cochrane-Ocrutt Transformation :

Cochrane and Ocrutt (1949) proposed an iterative estimation procedure, which is a remedy technique against the autocorrelation of disturbances. When the Durbin-Watson test in the model (1) is significant, the following process should be used.

Estimate ρ by means of $\hat{\rho}$ in(4) and estimate the parameter $(\beta_1^*, \ldots, \beta_\rho^*)'$ using ordinary least squares by the following transformed model :

$$y_t - \hat{\rho} y_{t-1} = \beta_1^*(1 - \hat{\rho}) + \beta_2^*(X_{t2} - \hat{\rho} X_{t-1,2}) + \ldots + \beta_\rho^*(X_{t\rho} - \hat{\rho} X_{t-1,\rho}) + e_t - \hat{\rho} e_{t-1}$$

Where, $t = 2, \ldots, n$

In this respect, a new estimate of β is obtained as :

$$b^{(1)} = (b_1^{(1)},\dots,b_\rho^{(1)})' = (\frac{b_1^*}{1-\hat{\rho}}, b_2^*,\dots,b_\rho^*)'$$

Then calculate the following residuals as :

$$u^{(1)} = (u_1^{(1)}, \dots, u_\rho^{(1)}) = Y - Xb^{(1)}, \qquad (5)$$

Thus, we estimate ρ by putting $u^{(1)}$ to the formula (4) and repeat the same steps. This method converges very fast, usually two iterations are sufficient.

The idea behind this method is to transform the regression model so that the disturbances are no more autocorrelated. This can be checked by carrying out the Durbin-Watson test with residuals $u^{(1)}$ from the transformed model. Only if this test is not significant, it is advised to use the results of the transformation. Let us stress that this does not happen always. There can be two reasons, why the transformation does not work. The disturbances can follow a more complicated model than AR(1). But even for disturbances truly coming from the AR(1) model, the transformation need not work, because $\hat{\rho}$ as an estimate of ρ is usually recommended only for $n > 50$.Otherwise the Cohrane - Orcutt transformation is only rough.

We should be aware of the so-called R^2-syndrome. The value of the coefficient of determination R^2 is typically overestimated if the disturbances are autocorrelated. This is another reason for using the Cochrane-Orcutt transformation.

Autocorrelated Residuals of Robust Regression:

Least Trimmed Squares Regression:

The least trimmed squares regression (LTS) is one of robust regression methods with a high break-down point. Before defining the LTS estimator b_{LTS}, Let us denote the real line by R and the residual corresponding to the t-th observation by :

$$u_t^{(b)} = y_t - b_1 - b_2X_{t2} - \dots - b_pX_{tp} \quad \text{for any} \quad b = (b_1,\dots,bp)' \in R^p,$$

$$t = 1, \dots, n \qquad (6)$$

International Journal of Mathematics and Statistics Studies
Vol.2, No.2, pp.1-12, June 2014
Published by European Centre for Research Training and Development UK (www.ea-journals.org)

Let us order the squared residuals as :

$$u^2_{(1)}(b) \le u^2_{(2)}(b) \le \ldots \le u^2_{(n)}(b). \qquad (7)$$

For an integer h, which must satisfy $n/2 \le h \le n$, the LTS estimator is defined by :

$$b_{LTS} = \arg\min_{b \in R^p} \sum_{t=1}^{h} u^2_{(t)}(b) \qquad (8)$$

The constant h is called a trimming constant. Obviously b_{LTS} ignores n-h observations, but it is not known before the computation, which observations should be discarded. Their deletion is defined in an implicit way. Thus, the estimator can be expressed using an indicator function as :

$$b_{LTS} = \arg\min_{b \in R^p} \sum_{t=1}^{n} u^2_{(t)}(b).I[u^2_t(b) \le u^2_{(h)}(b)]. \qquad (9)$$

In practice, the LTS estimate can be computed by some statistical Packages, for example S-plus.

The results of the LTS depend heavily on the choice of h. For h=n, the LTS coincides with the classical least squares regression, which has asymptotic breakdown point $\in^* = 0$. On the other hand, the LTS attains its maximum breakdown point for

$$h = [\frac{n}{2}] + [\frac{p+1}{2}] , \qquad (10)$$

where [a] stands for the integer part of a. In this situation, the asymptotic breakdown point attains its maximum possible value, which is 50 %.

The value of h should reflect the level of contaminancy of the data. It can be recommended to fit the LTS for every value of h between $h = [\frac{n}{2}] + [\frac{p+1}{2}]$ and n. Let us denote the proper value of h by h^*. Values of b_{LTS} are similar to each other for such different values of h, which are less or equal to h^*. The same property is true for estimator of σ^2. However, this stability breaks when h exceeds h^*. This is the situation when the ratio of discarded observations $(n-h)/n$ is lower than contaminacy level. Of course such search of h^* is rather subjective and requires also some experience.

Reweighted Least Squares:

The reweighted least squares regression (RLS) has been proposed by Rousseeuw and Leroy (1987). They recommended it as a one-step improvement of the LTS, which keeps a high breakdown point and attains a higher efficiency. The method itself is a classical weighted least squares regression, where the weights are assigned to observations according to the results of the least trimmed squares fit. We will describe the RLS regression and consider its properties.

Let us denote residuals of the LTS regression by $u_1^*, \ldots, u_n^*$. Rousseeuw and Leroy (1987) proposed to define the weights by means of an indicator function as :

$$w_t = I\left[\left|\frac{u_t^*}{\sigma^*}\right| \le 2.5\right] , \tag{11}$$

where σ^* is the robust estimate of the standard deviation of disturbances in the model (1).

There are also other ways used to define the weights. In general, the weight matrix $W = \text{diag}\,(w_1, \ldots, w_n)$ must be specified before computing the following RLS estimator.

$$b_{RLS} = (X'WX)^{-1} X'WY \tag{12}$$

This formula for the classical weighted least squared is equivalent with :

$$b_{RLS} = \arg\min{}_{b}\,{}_{b\in R^p} \sum_{t==1}^{n} W_t\, u_t^2(b) \tag{13}$$

The idea of Rousseeuw and leroy was to use the RLS always after an LTS fit. Their approach was to fit the LTS with $h = [\frac{n}{2}] + [\frac{p+1}{2}]$, which is not (in general) efficient. The next step, which is the RLS , takes into account observations with positive weights. Let us say there are h_{RLS} such observations. Rousseeuw and leroy would carry out least squares with these fixed h_{RLS} observations. On the other hand, our approach described above searches for a suitable h and then the LTS looks for a minimal value across all possible h- trimming observations. Therefore our approach (the LTS with a proper value of the trimming constant) is more efficient than the RLS and we do not recommend to use the RLS.

Least Weighted Squares :

We present very briefly the least weighted squares regression (LWS), which is proposed a robust method with high breakdown point. Its properties are still under research, as well as the question of choosing optimal weights; see Yohai (1987) for details. The definition of the estimator

$$b_{LWS} = \operatorname{argmin}_{b \in R^p} \sum_{t=1}^{n} w_t \sum_{i=1}^{n} u_i^2(b) I[u_i^2(b) \le u_{(t)}^2(b)] \qquad (14)$$

requires a specification of weights $w_1, w_2, \ldots, w_n$, which have to be positive and satisfy $\sum_{t=1}^{n} w_t = 1$. The formula (14) can be transformed to:

$$b_{LWS} = \operatorname{argmin}_{b \in R^p} \sum_{i=1}^{n} u_i^2(b) \sum_{t=1}^{n} w_t I[u_i^2(b) \le u_{(t)}^2(b)], \qquad (15)$$

which can be denoted as :

$$b_{LWS} = \operatorname{argmin}_{b \in R^p} \sum_{i=1}^{n} \tilde{w}_t u_{(i)}^2(b) \; . \qquad (16)$$

One of weights $\tilde{w}_1, \ldots, \tilde{w}_n$ is assigned to each observation, but after some permutation, which is not known before the computation .

Obviously, the least trimmed squares (as well as the least squares) is a special case of least weighted squares.

Autocorrelated Residuals of LTS Regression:

Although robust regression methods usually serve as diagnostic tools for ordinary least squares, they themselves need to have diagnostic tools and modifications. One of such tools, which is still to be constructed, is a test of autocorrelation of LTS and LWS residuals.

Let us recall the notation $u_1^*, \ldots, u_n^*$ for the residuals of the LTS regression. It turns out that the following Durbin-Watson test statistic

$$d = \frac{\sum_{t=2}^{n} (u_t^* - u_{t-1}^*)^2}{\sum_{t=1}^{n} u_t^{*2}} \qquad , \qquad (17)$$

has the same asymptotic behaviour as the classical test statistic computed with the lest squares residuals. Moreover, the Durbin-Watson test can be used in the same way as for the least squares for a moderate sample size, especially when there is only a small dimention of the model (p is small).

If the trimming constant h is properly selected, then the n-h observations ignored in equation (8) are outliers and their residuals differ a lot from residuals corresponding to " good " observations. Why to use them for calculation, such as for the Durbin-Watson test ?

On the other hand , a complete omission of outlier would be wrong, because the data from a time series. Deletion of an observation and shifting all the remaining observations a time unit ahead makes impossible to study the autocorrelation structure of the original time series.

The following idea may be better. An outlier detected by the LTS can be replaced by such value, which corresponds to the behaviour of the majority of the data (non-outliers). In other words, the time series of LTS residuals can be smoothened to replace outliers while keeping the " good " data. Then the Durbin-Watson test, estimation of ρ and the Cochrane-Orcutt transformation can be carried out. The principal idea, how to replace the outlying residuals, is still under research.

For the LWS, the situation is more complicated. A modification of Durbin-Watson test statistic should take into account different weights of individual observations.

Illustrative Example :

In this section, we demonstrate the practical use of the Durbin-Watson test and the Cochrane-Orcutt transformation on an example with real data. The package S-Plus was used for the calculations. We also calculate the (exact) value of the LTS estimate. Table (1) lists values of real gross domestic product (GDP) and real gross private domestic investment (INVEST) in the United States in the years from 1986 to 2007. Both variables are expressed in billions of dollars. The data are copied from the website www.Stls.frb.org/fred, and originally come from the U.S/ Department of Commerce .

Table (1) : Investment data .

Year	1986	1987	1988	1989	1990	1991	1992	1993
GDP	4900.9	5021	4919.3	5132.3	5505.2	5717.1	5912.4	6113.3

INVEST	655.3	715.6	615.2	673.7	871.5	863.4	857.7	879.3
Year	1994	1995	1996	1997	1998	1999	2000	2001
GDP	6368.4	6591.8	6707.9	6676.4	6880	7062.6	7347.7	7543.8
INVEST	902.8	936.5	907.3	829.5	899.8	977.9	1107	1140.6
Year	2002	2003	2004	2005	2006	2007		
GDP	7813.2	8159.5	8508.9	8856.5	9224	9333.8		
INVEST	1242.7	1393.3	1558	1660.1	1772.9	1630.8		

Because the values are not nominal, but adjusted for deviation of money, the following model

$$INVEST_t = \beta_o + \beta_1.GDP_t + e_t, \qquad t = 1, \ldots, n \qquad (18)$$

has its reasonable economic interpretation; see Greene (1993) for a similar investment equation. Table (2) presents two iterations of the Cochrane-Orcutt transformation with results of the ordinary least squares (OLS), and least trimmed squares (LTS) should be self-explaining.

Table (2) : Results of the example .

	OLS	Cochrane-Orcutt first iteration	Cochrane –Orcutt second iteration	LTS
Intercept	- 582	- 980.5	- 962.4	-371.4
GDP	0.239	0.282	0.286	0.203
$\hat{\rho}$	0.779	0.790	0.792	0.807
R^2	92.1 %	74.3 %	73.4 %	91.0 %

By the original model (18), the Durbin-Watson test (with OLS residuals) is significant, because $d = 0.418 < d_L(0.05) = 1.24$. In the transformed model, the value $d = 1.31$ lies in the inconclusive region, $d_L(0.05) = 1.22 < d \le d_u(0.05) = 1.42$. Thus, we believe that the autocorrelation has been (almost) removed and the results are satisfactory.

This example demonstrates the importance of using the Cochrane-Orcutt transformation. Not only there is a big difference in the estimates of the parameters, but also the true value of the coefficient of determination R^2 (about 73%) is substantially over-estimated in the original model (18). It is also worth noting that there is only a slight difference between results of the first and second iteration of the Cochrane-Orcutt transformation. This does not hold for the intercept, which is well known to be very sensitive to slight changes of the slope. By the way, we did not forget to divide the estimate of the intercept in the transformed models by $1-\hat{\rho}$.

A (subjective) search for a proper value of the trimming constant for the LTS hints to choose $h=19$. It turns out that the years(2004),(2005) and (2006) are ignored. Just for comparison, the RLS approach (not subjective at all) gives exactly the same results. The value of $\hat{\rho}$ for the LTS is influenced by outliers, as well the Durbin-Watson statistic computed with the LTS residuals, which equals $d=0.338$. Both these values do not have an interpretation and it is not known yet how to robustify them in this context.

CONCLUSION

In this paper, we aimed to explore the issue of whether one should use OLS residuals or residuals from a robust regression method (TLS) as a basis for autocorrelated test . We conjecture how to robustify the Durbin-Watson test for residuals of the least trimmed squares regression, in order to detect the autocorrelation problem.

REFERENCES

- Cochrane D. and G.H.Orcutt (1949) " Application of least squares regression to relationships Containing autocorrelated error terms" J. Amer. Statist. Assoc. 44, pp. 32-61.
- Durbin J. and G.S. Watson (1950. 1951, 1971) " Testing for Serial Correlation in Least squares regression " I, II, III, Biometrika, 37, 38, 58, pp. 409-428, 159-178, 1-19.
- Green W.H.(1993) " Econometric Analysis " Macmillan, New York.
- Hample, F.R., Ronchetti, E.M., Rousseeuw, P.J. and W.A. Stahel (1986) " Robust Statistics, The Approach Based on Influence Functions " John wiley & Sons, New York.
- He, D.G. and S.L. Portnoy (1990) " Breakdown robustness of tests " J. Amer. Statist. Assoc. 85, pp. 446-452.
- Holland, P. and R. Welsch (1977) " Robust regression using , Iteratively reweighed least squares " Commun. Statist. Theor. Meth. b, pp. 813-827.
- Judge G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, and T.C. lee (1985) " The Theory and practices of Econometrics " Wiley , New York.

- Rousseeuw, P.J. and A.M. Leroy (1987) " Robust regression and outlier detection " Wiley, New York.
- Yohai, V.J. (1987) " High breakdown point and high efficiency robust estimates for regression " Ann. Statist. 15, pp. 642-656.