HIGHLY ROBUST ESTIMATION OF THE AUTOCORRELATION COEFFICIENT

Jan Kalina, Katarína Vlčková

Abstract

The classical autocorrelation coefficient estimator in the time series context is very sensitive to the presence of outlying measurements in the data. This paper proposes several new robust estimators of the autocorrelation coefficient. First, we consider an autoregressive process of the first order AR(1) to be observed. Robust estimators of the autocorrelation coefficient are proposed in a straightforward way based on robust regression.

Further, we consider the task of robust estimation of the autocorrelation coefficient of residuals of linear regression. The task is connected to verifying the assumption of independence of residuals and robust estimators of the autocorrelation coefficient are defined based on the Durbin-Watson test statistic for robust regression. The main result is obtained for the implicitly weighted autocorrelation coefficient with small weights assigned to outlying measurements. This estimator is based on the least weighted squares regression and we exploit its asympotic properties to derive an asymptotic test that the autocorrelation coefficient is equal to 0. Finally, we illustrate different estimators on real economic data, which reveal the advantage of the approach based on the least weighted squares regression. The estimator turns out to be resistant against the presence of outlying measurements.

Key words: time series, autoregressive process, linear regression, robust econometrics

JEL Code: C14, C13, C22

---

1 Introduction

The stationary autoregressive process of the first order has been described as a popular model for economic time series in a vast number of econometric references. Its model for a time series $X_1, X_2, \ldots$ is commonly denoted as AR(1) and has the form

$$X_t = \rho X_{t-1} + e_t, \tag{1}$$

where the parameter $\rho \in (-1,1)$ is the (population) autocorrelation coefficient of the first order, shortly autocorrelation coefficient (Chatfield, 2004). However, we observe only a sequence $X_1, \ldots, X_n$ in practice.

The classical estimate of ρ in the context of an AR(1) process is equal to

$$\hat{\rho} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_t^2}. \tag{2}$$

In the statistical theory of time series, it is a common task to test the null hypothesis

$$H_0: \rho = 0. \tag{3}$$

The asymptotic one-sided test has been described to reject the null hypothesis if and only if

$$\hat{\rho} > -\frac{1}{n} + 1.96\sqrt{\frac{n-1}{n(n+2)}}. \tag{4}$$

This approximative test was derived for normally distributed errors (Chatfield, 2004) and can be additionally approximated by

$$\hat{\rho} > \frac{2}{\sqrt{n}}. \tag{5}$$

The autocorrelation coefficient $\hat{\rho}$ is too sensitive to the presence of outlying measurements (outliers) in the data and is biased for a contaminated normal distribution, i.e. for a time series with the majority of data points with normally distributed random errors and a minority of data points following a distribution with a much higher variance. There have been various attempts to estimate the autocorrelation coefficient in a robust way (Shevlyakov & Smirnov, 2011).

This paper has the following structure. Section 2 proposes a highly robust estimator of the autocorrelation coefficient for an AR(1) process, which is based on the least weighted squares regression estimator. Sections 3 and 4 are devoted to robust estimation of the autocorrelation coefficient for a sequence of regression residuals. The main result is obtained in Section 4 for the implicitly weighted autocorrelation coefficient with small weights assigned to outlying measurements. This estimator is based on the least weighted squares regression and we study its asymptotic approximation. The robust methods are compared with the classical estimator in an analysis of real economic data in Section 5, where only the estimator of Section 4 confirms its robustness under the data contamination by an outlying measurement.

2 Autocorrelation coefficient based on robust regression

This section recalls the least weighted squares regression and proposes a new estimator of the autocorrelation coefficient based on this highly robust regression estimator. In general, one of possible ways for defining a robust autocorrelation coefficient of an AR(1) process is to exploit the following general form of (Shevlyakov & Smirnov, 2011). Let $\hat{\beta}_1$ denote a robust estimator of the slope in the regression of $X_2, \ldots, X_n$ against $X_1, \ldots, X_{n-1}$. Let $\hat{\beta}_2$ denote a robust estimator of the slope in the regression of $X_1, \dots, X_{n-1}$ against $X_2, \dots, X_n$. Then, the population autocorrelation coefficient can be estimated by

$$\rho^* = \sqrt{\hat{\beta}_1 \hat{\beta}_2}. \tag{6}$$

The estimator (7) inherits the robustness properties of the robust regression estimator used to obtain the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$. This straightforward way allows to define a robust autocorrelation coefficient based on regression quantile (Koenker, 2005), trimmed least squares (Jurečková & Sen, 1996), or least weighted squares (Víšek, 2002) estimators. In practice, the autocorrelation coefficient is often used in the context of linear regression to verify the assumption of independence of the random regression errors. Therefore, we investigate an alternative approach for defining a tailor-made robust estimator of the autocorrelation coefficient of regression residuals.

3 Autocorrelation coefficient of regression residuals

This section proposes a robust estimator of the autocorrelation coefficient tailor-made for regression residuals. Such approach is suitable e.g. for a robust alternative to the Cochrane-Orcutt transformation (Cochrane & Orcutt, 1949).

In our previous work (Kalina, 2013), we investigated the Durbin-Watson statistic (Durbin & Watson, 1950) computed for residuals of regression quantiles or trimmed least squares estimators. Let us consider the linear regression model

$$Y_t = \beta_0 + \beta_1 X_t + e_t, \quad t = 1, \dots, n. \tag{7}$$

The classical Durbin-Watson test statistic fulfills

$$d = \frac{\sum_{t=2}^{n} (u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2} \doteq 2(1 - \hat{\rho}), \tag{8}$$

where $u_1, \dots, u_n$ are residuals of the least squares estimator.

Let us now consider the Durbin-Watson test statistic computed using residuals of the regression quantile estimator of β in the linear regression with some parameter α in the interval $(0,1)$. This statistic will be denoted by d_{RQ}. We apply (8) to obtain a robust estimator of the autocorrelation coefficient in the form

$$\hat{\rho}_{RQ} = 1 - \frac{d_{RQ}}{2}. \tag{9}$$

It holds from the asymptotic equivalence of d_{RQ} and d (Kalina, 2012) that the estimator $\hat{\rho}_{RQ}$ is a consistent estimator of the population autocorrelation coefficient ρ.

The autocorrelation coefficient for the residuals of the trimmed least squares estimator denoted as $\hat{\rho}_{TLS}$ can be defined in an analogous way. By considering the asymptotic behaviour of the Durbin-Watson test statistic computed from the LTS residuals (Kalina, 2012), we conclude that $\hat{\rho}_{TLS}$ is a consistent estimator of the population autocorrelation coefficient ρ.

4 Implicitly weighted autocorrelation coefficient of regression residuals

The robust autocorrelation coefficient of regression residuals based on the least weighted squares regression can be defined in an analogous way as in Section 3 following the general idea (6) of (Shevlyakov & Smirnov, 2011). Nevertheless, the procedure would not be very suitable for the least weighted squares estimator, because it ignores the optimal weights corresponding to individual observations. Therefore, we will rather consider two different versions of the weighted autocorrelation coefficient and their properties. Particularly, we will derive a robust estimator from the exact hypothesis test and also its asymptotic version based on the asymptotic representation of the least weighted squares estimator (Víšek, 2011).

The least weighted squares (LWS) estimator was proposed by (Víšek, 2002) for the linear regression with several independent variables as a robust estimator of the regression parameters with a high breakdown point. The estimator does not include the outlier detection intrinsically, while the potential outliers are only down-weighted and not trimmed away completely. This estimator is based on implicit weighting of individual observations and turns out to possess a high breakdown point, which is a statistical measure of sensitivity against noise or outliers in the data. The estimator may use linearly decreasing weights or adaptive data-dependent weights proposed by (Čížek, 2011), allowing to combine robustness and efficiency in a very appealing way. The LWS estimator attains a 100 % asymptotic efficiency of the least squares under Gaussian errors (Čížek, 2011). It is robust to heteroscedasticity (Víšek, 2011) and diagnostic tools for checking its assumptions are available (Kalina, 2012).

The computation of the LWS estimator is intensive and an approximative algorithm can be obtained as a weighted version of the algorithm proposed for the least trimmed squares (LTS) regression (Rousseeuw & van Driessen, 2006), which represents a special case of least weighted squares with weights equal to zero or one only. While the LTS estimator is very reliable for outlier detection (Hekimoglu, Erenoglu & Kalina, 2009), it suffers from a high sensitivity to small deviations near the center of the data. On the other hand, the advantage of the LWS estimator is a small local sensitivity.

Let us consider the LWS regression in the model (7). The residuals of the LWS regression will be denoted by $u_1, \dots, u_n$ and the optimal weights found by the LWS estimator will be denoted as $w_1, \dots, w_n$. We will consider weighted residuals with fixed weights $w_1, \dots, w_n$ in the form

$$u_t^* = \sqrt{w_t}\, u_t, \quad t = 1, \dots, n. \tag{10}$$

The first definition of the LWS-autocorrelation coefficient is

$$\hat{\rho}_{LWS} = \frac{\sum_{t=2}^{n} \sqrt{w_t w_{t-1}}\, u_t u_{t-1}}{\sum_{t=1}^{n} w_t u_t^2} \tag{11}$$

equal to the weighted autocorrelation coefficient (Choi, Kim, Feng, Lee & Jung, 2012) with fixed weights, while the optimal weights by the LWS regression estimator are used. Let us now recall the Durbin-Watson statistic d_{LWS} (Kalina, 2013) computed from the residuals of the LWS regression. There holds a connection

$$\hat{\rho}_{LWS} = 1 - \frac{d_{LWS}}{2}, \tag{12}$$

which allows to conclude that $\hat{\rho}_{LWS}$ is a consistent estimator of the population autocorrelation coefficient ρ. It inherits the robustness properties of the LWS regression estimator described above, particularly the high breakdown point.

Now we use the asymptotic theory to derive another version of the LWS-based autocorrelation coefficient, which is asymptotically equivalent to (12). This reasoning will allow us to study asymptotic properties of $\hat{\rho}_{LWS}$, which remain unknown for the robust estimators of Section 2. The asymptotic approximation of d_{LWS} enables us to conclude that $\hat{\rho}_{LWS}$ is asymptotically equivalent to

$$\rho_{LWS}^* = 1 - \frac{\kappa^T A \kappa}{2\kappa^T \kappa} = 1 - \frac{\sum_{t=2}^{n} \kappa_t \kappa_{t-1}}{2\sum_{t=1}^{n} \kappa_t^2}, \tag{13}$$

where κ is obtained from the asymptotic approximation to the LWS residuals and can be estimated easily (Kalina, 2013). This is a more realistic setting compared to the approach of (11), which considers the weights found by the LWS to be fixed. We can formulate the following theorem.

Theorem 1. Under the assumption of normally distributed random errors in the regression model, the estimator ρ_{LWS}^* computed from the residuals of the LWS regression with the adaptive weights of (Čížek, 2011) converges in distribution to a random variable with normal distribution

$$N\left(-\frac{1}{n}, \frac{n-1}{n(n+2)}\right). \tag{14}$$

Thanks to the 100 % efficiency of the LWS estimator with adaptive weights, also the asymptotic variance of ρ_{LWS}^{*} is equal to the asymptotic variance of the classical autocorrelation coefficient computed for the residuals of the least squares regression.

5 Example: Investment data

This section illustrates the robustness properties of several versions of the autocorrelation coefficient for regression residuals, including the robust ones proposed in this paper.

We consider the linear regression of real gross private domenstic investments in the USA in 10^9 USD against the GDP. The data retrieved from the internet (www.stls.frb.org/fred) are yearly measurements from the years 1980-2001 and are shown in Figure 1. To show the harmful effect of outliers in the data, we consider also a contaminated data set shown in Figure 2, which is different from the original data only in the value of one observation. The residuals of the LWS regression are shown in Figure 3. The figure clearly shows the autocorrelation structure in the residuals.

We computed the estimates of the regression parameters using the least squares, L1-estimator (a special case of regression quantiles), and LWS estimator (with linearly decreasing weights). Clearly, the random errors in the regression are not independent and the AR(1) model seems suitable. Thus, it is important to estimate the parameter ρ. Various versions of the autocorrelation coefficient and Durbin-Watson statistic were computed based on residuals of different robust regression estimators of the regression parameters. For the LWS regression, we use the weighted version of the autocorrelation coefficient (11).

The results of the computations are shown in Table 1. Comparing the original and the contaminated data sets, both the least squares estimator and the L1-estimator are heavily influenced by the presence of the outlying measurement. The estimation of the autocorrelation coefficient based on these two regression estimators is also heavily influenced by the only outlier in the data. While the L1-estimator is known to be robust with respect to outliers in the response, it is revealed to be very non-robust with respect to a single outlier in the independent variable. Only the weighted autocorrelation coefficient based on the least weighted squares turns out to be resistant against the presence of the outlying measurement.

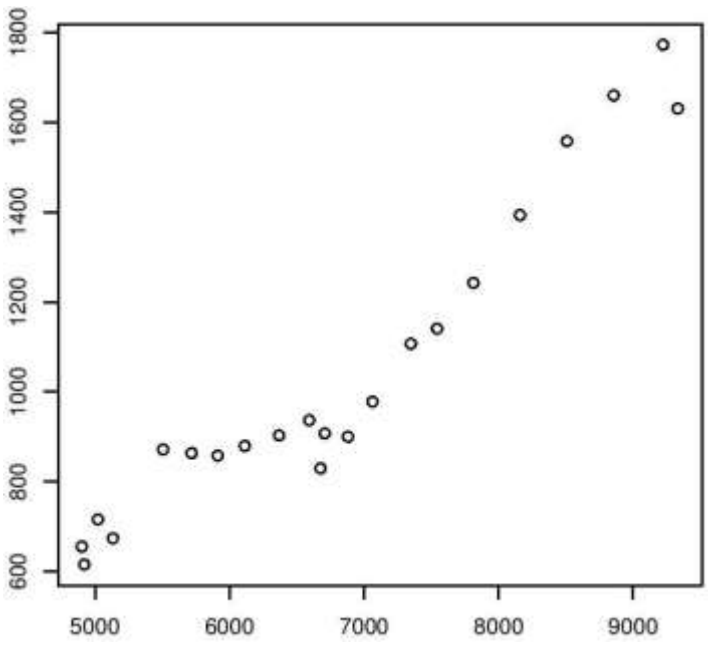
To summarize, this paper extends the theoretical results on the Durbin-Watson test statistic for robust regression (Kalina, 2012) and defines a robust autocorrelation coefficient based on robust regression. While methods based on ranks of observations have appealing properties in variety of situations (Jurečková & Kalina, 2012), our work is the first approach to autocorrelation based on implicit weighting based on ranks of residuals in a suitable model. The numerical example in Section 5 computes the autocorrelation coefficient from regression residuals in both real and contaminated data. The sensitivity of the classical autocorrelation coefficient as well as the robustness of the LWS-autocorrelation coefficient is revealed.

Tab. 1: Results of the analysis of the investment data in Section 5

	Least squares	L1-estimator	LWS estimator
Original data			
Intercept	-582.0	-516.5	-509.2
Slope	0.239	0.230	0.224
Autocorrelation coefficient	0.791	0.791	0.797
Durbin-Watson statistic	0.418	0.418	0.407
Contaminated data			
Intercept	-306.6	-401.9	-465.3
Slope	0.193	0.210	0.221
Autocorrelation coefficient	-0.080	0.555	0.804
Durbin-Watson statistic	2.159	0.889	0.392

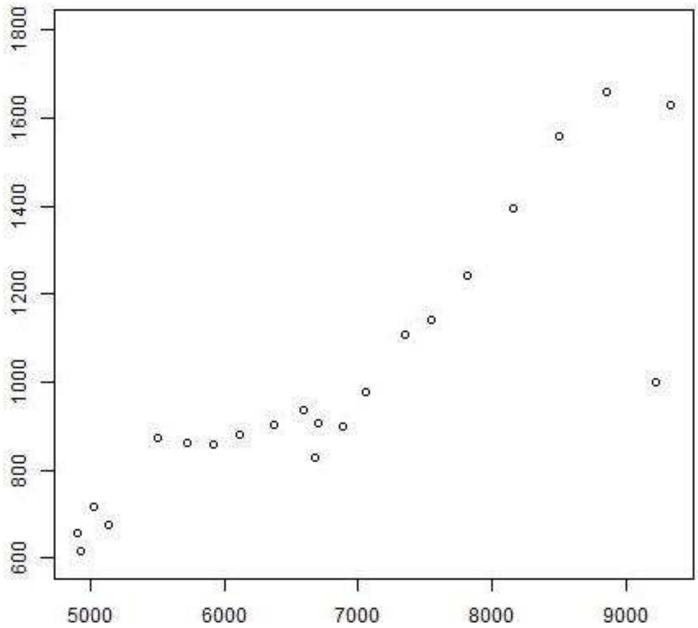
Source: own computation.

Fig. 1: Investment data in the example of Section 5



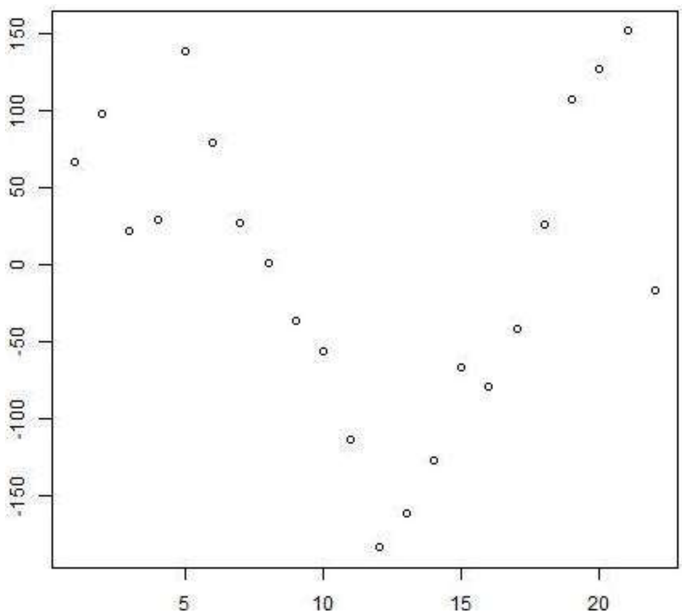
Source: www.stls.frb.org/fred.

Fig. 2: Contaminated investment data



Source: own modification of data of Fig. 1.

Fig. 3: Residuals of the least weighted squares in the regression of real private domestic investments against GDP in the contaminated investment data set



Source: own computation

Acknowledgment

The research was financially supported by the Neuron Fund for Support of Science.

References

Chatfield, C. (2004). *The analysis of time series. An introduction*. 6th ed. Boca Raton: Chapman & Hall/CRC.

Choi, J., Kim, M., Feng, L., Lee, C., & Jung, H. (2012). A new weighted correlation coefficient method to evaluate reconstructed brain electrical sources. *Journal of Applied Mathematics, 2012,* Article ID 251295, 1-11.

Čížek, P. (2011). Semiparametrically weighted robust estimation of regression models. *Computational Statistics & Data Analysis, 55*(1), 774-788.

Cochrane, D., & Orcutt, G.H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association, 44*, 32-61.

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

Hekimoglu, S., Erenoglu, R.C., & Kalina J. (2009). Outlier detection by means of robust regression estimators in use in engineering science. *Journal of Zhejiang University Science A, 10*(6), 909-921.

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

Jurečková, J., & Kalina, J. (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli, 18*(1), 229-251.

Kalina, J. (2012). On multivariate methods in robust econometrics. *Prague Economic Papers, 21*(1), 69-82.

Kalina, J. (2013). *Autocorrelated residuals of robust regression.* In Löster, T., & Pavelka, T. (Eds.), *The 7-th International Days of Statistics and Economics* (pp. 551-560). Slaný: Melandrium.

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

Rousseeuw, P.J., & van Driessen, K. (2006): Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery, 12*(1), 29-45.

Shevlyakov, G., & Smirnov, P. (2011). Robust estimation of the correlation coefficient: An attempt of survey. *Austrian Journal of Statistics, 40*(1&2), 147-156.

Víšek, J.Á. (2002). The least weighted squares I. *Bulletin of the Czech Econometric Society, 9* (15), 31-56.

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

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

Bc. Katarína Vlčková

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

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

vlckova@cs.cas.cz