العنوان:	Statistical Inference for LAD Regression With Autocorrelated Errors
المصدر:	المجلة المصرية للدراسات التجارية
الناشر:	جامعة المنصورة - كلية التجارة
المؤلف الرئيسي:	Abd Alsallam, Moawad Al Fallah
المجلد/العدد:	مج31, ع1
محكمة:	نعم
التاريخ الميلادي:	2007
الصفحات:	1 - 10
رقم MD:	659892
نوع المحتوى:	بحوث ومقالات
قواعد المعلومات:	EcoLink
مواضيع:	الاستدلال الإحصائي، الإحصاء الرياضي
رابط:	http://search.mandumah.com/Record/659892

Statistical Inference for LAD Regression with Autocorrelated Errors

Moawad El-Fallah Abd El-Sallam[*]

Abstract:

The method of least absolute deviation (LAD) provides a robust alternative to least squares, particularly when the disturbances follow distributions that are non-normal and subject to outliers. While inference in least squares estimation is well understood, inferential procedures in the context of least absolute deviation estimation have not been studied as extensively particularly in the presence of auto correlation. In this paper we study two alternative significance test procedures in least absolute deviation regression, along with two approaches used to correct for serial correlation. The study is based on a Monte Carlo simulation, and comparisons are made based on observed significance levels.

* **Department of Statistics & Mathematics and Insurance, Faculty of Commerce, Zagazig university.**

Keywords: Least absolute deviation regression; Robust regression; Autocorrelation; Simulation.

1. Introduction:

In regression analysis, the Ordinary Least Squares (OLS) method yields parameter estimates that are unbiased and have minimum variance when the disturbances are independent and identically normally distributed. However, in the presence of non-normal errors, the performance of OLS can be not optimum, especially if the errors follow a distribution that tends to produce outliers. Thus much research has been aimed at developing estimation approaches that are robust to such outlier-producing error distributions. Least absolute deviation (LAD) method has emerged as one of the most commonly employed techniques for robust regression. LAD estimates are affected less strongly by extreme observations, relative to OLS estimates. However, less is understood about the behavior of LAD estimates, particularly for small samples, and the process of inference is less straightforward. So, inference in LAD estimation is an active area of research. In this respect, Koenker and Bassett (1982) suggested the Wald, Likelihood ratio (LR), and Lagrange multiplier (LM) tests when using LAD estimation. These approaches can be used to test for coefficient significance in the regression model. Dielman and Pfaffenberger (1990) studied inference for regression using LAD estimation when disturbances are independent but not necessarily normal.

On the other hand, although LAD estimation has been suggested as an alternative to Least Squares regression, it is much less used, and thus can be viewed as a non-traditional technique. In addition, autocorrelation correction procedures in LAD regression have seen little use in practice. These procedures have not been fully studied and the inference techniques appropriate for LAD regressions after correcting for autocorrelation have only recently been developed. Thus, the use of these autocorrelation corrections can be viewed as nontraditional. In this paper we present results of a simulation study addressing questions of inference for regression using LAD estimation in the presence of serial correlation. The performances of various tests and corrections for autocorrelation are compared, based on observed significance levels. We concentrate on model performance in small samples, due to the practical importance of smaller sample sizes, particularly for application in business and economics. The emphasis of this study is on the performance of hypothesis tests about the regression Coefficients in contrast with the earlier papers that dealt with estimation.

The paper is organized as follows. The linear regression model with autocorrelation and the LAD estimation along with the existing two corrections for serial correlation are introduced in section (2). Issues of inference are discussed in section (3), including descriptions of the test procedures and a review of the applicable literature. The simulation study is described in section (4), and the results are discussed in section (5). Section (6) concludes with some suggestions of areas for future research.

2.The Model and Correction for Serial Correlation:

Consider the following simple regression model:

$$y_i = \beta_0 + \beta_i x_i + u_i,$$

$$u_i = \rho u_{i-1} + \varepsilon_i, \quad i = 1, \ldots\ldots, n \quad (1)$$

where, y_i and x_i are the i^{th} observations on the dependent and explanatory variables, respectively, and u_i is a random error for the i^{th} observation and may be subject to autocorrelation. The ε_i represent disturbance components that are assumed to be independent and identically distributed, but not necessarily normal. The parameters β_0 and β_i are unknown and must be estimated. The parameter ρ is the autocorrelation coefficient, with $|\rho| < 1$.

The LAD criterion chooses the estimates of β_0 and β_1 that minimize the sum of the absolute residuals. The use of this criterion, rather than the minimization of the sum of the squared residuals used in OLS estimation, provides robustness against outliers, and is particularly useful when ε_i are generated by a fat-tailed distribution. LAD estimation can be formulated as a linear programming problem or iteratively reweighted least squares algorithm, see Morgenthaler (1992).

Using matrix notation, the model can be written as:

$$Y = X\beta + U, \qquad (2)$$

where:

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ X=\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix},\ U=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix},\ \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad (3)$$

The problem of serial correlation has been investigated extensively in the context of OLS, and numerous approaches have been proposed for correction; see for example Cochrane and Orcutt (1949), Banerjee et al., (1993), Gujarati (1995) and Ramanathan (2002).

Two procedures, both two-stage and based on a generalized Least Squares approach, are commonly employed to correct for autocorrelation in the Least Squares regression context. These are the Prais-Winsten (PW) and Cochrane Orcutt (CO) procedures. Both procedures transform the data using the autocorrelation coefficient, ρ, after which the transformed data are used in estimation. The procedures differ in their treatments of the first observation, (y_1, x_1). The PW transformation matrix is:

$$M_1 = \begin{bmatrix} \sqrt{1-\rho^2} & 0 \ldots\ldots & 0 & 0 \\ -\rho & 1 \ldots\ldots & 0 & 0 \\ & \ldots\ldots & & \\ 0 & 0 \ldots\ldots & \rho & 1 \end{bmatrix} \qquad (4)$$

Pre-multiplying equation (2) by M_1 yields:

$$M_1 Y = M_1 X \beta + M_1 U, \qquad (5)$$

or

$$Y^* = X^* \beta + \varepsilon, \qquad (6)$$

where Y^* contains the transformed dependent variable values and X^* is the matrix of transformed independent variable values, so

$$Y^* = \begin{bmatrix} \sqrt{1-\rho^2}\, y_1 & y_2 - \rho y_1 \ldots\ldots y_n - \rho y_{n-1} \end{bmatrix} \qquad (7)$$

and

$$X^* = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\, x_1 \\ 1-\rho & x_2 - \rho\, x_1 \\ \vdots & \vdots \\ 1-\rho & x_n - \rho x_{n-1} \end{bmatrix} \qquad (8)$$

In equation (6), ε is the vector of serially uncorrelated ε_i errors. The PW approach may be effective in the LAD context as well as in OLS.

The Co transformation matrix is the (n-1) x 1 matrix obtained by removing the first row of the M_1 transformation matrix. Coursey and Nyquist (1983) investigated the performance of the CO correction with LAD estimation for disturbances from the symmetric stable family.

The use of CO transformation means that (n-1) observations, rather than n, are used to estimate the model. In the CO transformation, the first observation is omitted, whereas, it is transformed and included in the estimation with the PW transformation. Asymptotically, the loss of this single observation is probably of minimal concern. However, for small samples omitting the first observation has been shown to result in a least squares estimator inferior to that obtained when the first observation is retained and transformed. In all cases, the use of either correction approach requires that the correlation coefficient, ρ, be estimated from sample data. In this case, we estimate ρ by applying LAD estimation to the following equation:

$$\hat{u}_i = \rho \hat{u}_{i-1} + \epsilon_i \qquad (9)$$

where the $\hat{u}_i$ are the residuals from the LAD fit to equation (1). It will be shown in section (4) that, the PW correction approach is more effective following pre-testing for autocorrelation, in the context of LAD regression.

3.Testing for LAD Regression:

An important form of inference in regression is the significance testing for coefficients. This remains an underdeveloped area in LAD regression. Koenker and Bassett (1982) developed the Wald and likeLihood ratio (LR) test statistics for use in significance testing in LAD regression, and they showed that the two test statistics have identical asymptotic chi- square distributions, with the degrees of freedom equal to the number of coefficients included in the test, (e.g.,1, for testing Ho: $\beta_0 = 0$). In this paper, the Wald and LR testing approaches are considered.

The Wald test statistics in the general regression case is given by $\hat{\beta}\ \hat{\beta}\ \lambda^2 D$, where $\hat{\beta}$ is the vector of LAD estimates for the coefficients included in the test; D is the appropriate block of the $(X' X)^{-1}$ matrix to be used in the test, and λ represents a scale parameter, such that $\lambda = 1/(2\ f(m))$, where f (m) is the p.d.f of the disturbance distribution evaluated at the median.

The LR test statistic is 2 $(SAR_R - SAR_U) / \lambda$, where SAR_R is the sum of the absolute residuals in the restricted model, and SAR_U is the sum of the absolute residuals in the unrestricted, or full model. The scale parameter, λ, is identical to that in the Wald test statistic.

Both the Wald and the LR test statistics require the estimation of the scale parameter λ. The estimator of λ used in this paper is based on that suggested by Mckean and Schrader (1984) as: $\hat{\lambda}=n\left[e_{(n-K+1)}-e_{(K)}\right]/\left[2Z_{\alpha/2}\right]$, where $k=(n+1)/2-Z_{\alpha/2}(n/4)^{1/2}$, and the $e_{(.)}$ are ordered residuals from the LAD fitted model. Mckean and Schrader determined-using Monte Carlo simulation-that the estimator of λ offers the best performance
when $\alpha = 0.05$

For additional studies of inference in LAD regression, Dielman and Pfaffennberger (1990) examined the small sample performance of the Wald and LR tests for simple LAD regression using independent disturbances, and considered two different bootstrap approaches to hypothesis testing for LAD regression coefficients. However, the bootstrap procedure performed well, but is quite computationally intensive, and was applied in cases when the disturbances were independent.

4. Monte Carlo Simulation:

In this section, we are interested in studying the performances of the two "standard" procedures for testing the null hypothesis that the slope coefficient, β_1, is equal to zero. The model is that shown as equation (1), and we consider the Wald and LR tests along with the PW and CO approaches to correcting for serial correlation.

(4.1) Design of the Experiment.

The experimental design for the Monte Carlo simulation consists of the following factors:

Sample size: we consider a sample size of n=20 throughout the experiment many applications of practical interest involve data histories of approximately this length (e.g. 5 years of quarterly financial data). The sample size of 20 observations is small enough to give a reliance on asymptotic results, so the simulation approach is useful for studying the small sample behaviors of the models.

LAD estimation studied by Dielman and Pfaffendberger (1990) indicated that, model behavior is relatively stable for sample sizes over n = 40. Behavior for n = 20 and n =30 were relatively consistent, while reducing the sample size much below 20 yielded notably different results. Therefore, the use of n =

20 represents an effort to study small-sample results.

Coefficient values: the intercept, β_0 , is set to 0 throughout the experiment. This causes no loss of generality, based on the results of Andrews (1986). The slope coefficient varies, with $\beta_1 = 0, 0.2, 0.4, 0.6, 0.8, 1.00$. Results with $\beta_1 = 0$ are used to study significance levels, while the full range of β_1 values is used to study the power performances of the tests.

Autocorrelation: we use $\rho = 0, 0.2, 0.4, 0.6, 0.8, 0.95$. This range permits evaluations of the effect of several autocorrelations on the performances of the tests. Only, we consider positive autocorrelation in this study, because it is encountered more in practical applications, particularly in business and economic data.

Disturbance distributions: Four different distributional forms for the ε_i disturbances are considered, to permit an investigation of model performance in a broad range of circumstances.

The distributional forms are:

**Normal, with mean 0 and variance 1; i.e N (0, 1).

**Contaminated normal, where ε_i are drawn from a N (0, 1) distribution with probability 0.85 and from a N (0, 25) distribution with probability 0.15.

**Laplace (double exponential), with mean 0 and variance 2.

**Cauchy, with median 0 and scale parameter 1.

The contaminated normal (CN), Laplace and Cauchy are all "fat-tailed" distributions, which tend to produce outliers. (It is interesting to note that LAD is the maximum likelihood estimator for egression with Laplace-distributed, independent errors).

Once the ε_i values are generated, the u_i values are created as $u_i = \rho u_{i-1} + \varepsilon_i$, where $u_0 = \varepsilon_0 / (1-\rho^2)$, and ε_0 is an initial draw from the disturbance distribution.

Explanatory variable:

The independent variable x_i, is generated as $x_i = ax_{i-1} + v_i$, with $a = 0, 0.4, 0.8$, and $v_i \sim N(0, 2)$.

We note that, when $a = 0$, the explanatory variable values are drawn from a normally distributed random variable. While, if a assumes the values of 0.4 or 0.8, the explanatory variable is an autoregression with a normal error term. The patterns of the explanatory variable generated in these ways are encountered with practical time series applications. Thus, these various patterns are used to enhance the generalizability of the results.

Once generated, the explanatory variable values are held fixed throughout the experiment. For each factor combination in this design (value of β_1, autocorrelation level, disturbance distribution, and explanatory variable type), 1500 Monte Carlo trials are used, and the resulting parameter estimates are recorded. All random numbers are generated using IMSL subroutines, and the explanatory variable values are generated independently of the disturbances.

(4.2) The results:

Based on the design of the simulation study, we can study the effects of the two corrections for autocorrelation and the two tests for coefficient significance. The simulation results are compared based on the observed significance levels. In this respect, the hypothesis tests are performed at the 5% level of significance. Therefore, when H_0 is true, we expect to reject it in approximately 5% of the 1500 replications of each pattern of the experiment. Tables 1 and 2 show the observed significance levls for the sets of 4500 replications formed by combining the results from the three types of explanatory variables. The results represent the percentage of trails in which the null hypothesis, H_0: $\beta_1 = 0$, is rejected in favor of the two tailed alternative when H_0 is, in fact, true. These percentages are estimates of α, the probability of a type 1 error. For all of the correction / test combinations, the estimated α, increases with the degree o f a utocorrelation, and the positive effect of correcting for severe autocorrelation is clear. Generally, it is important to correct for autocorrelation when $\rho > 0.2$, and the importance tends to increase as the value of ρ increases.

Table (1): Observed significance levels: Wald test

ρ	Cauchy			Laplace		
	None	PW	CO	None	PW	CO
0	0.069	0.071	0.041	0.069	0.082	**0.049**
0.2	0.134	0.113	0.058	0.124	0.133	0.068
0.4	0.175	0.144	0.064	0.163	0.159	0.084
0.6	0.262	0.163	0.084	0.223	0.212	0.111
0.8	0.378	0.189	0.115	0.314	0.243	0.155
0.95	0.481	0.224	0.179	0.451	0.266	0.199
ρ	Normal			Contaminated Normal		
	None	PW	CO	None	PW	CO
0	0.087	0.101	0.062	0.071	0.092	**0.048**
0.2	0.139	0.143	0.072	0.131	0.133	0.063
0.4	0.154	0.148	0.083	0.152	0.142	0.079
0.6	0.243	0.226	0.144	0.244	0.199	0.103
0.8	0.341	0.249	0.161	0.351	0.213	0.149
0.95	0.421	0.273	0.194	0.461	0.226	0.203

* Bold values represent observed significance levels that do not differ from the nominal 5% with 95 % confidence

Table (2): Observed significance levels: LR test

ρ	Cauchy			Laplace		
	None	PW	CO	None	PW	CO
0	**0.083**	0.073	**0.081**	0.059	0.091	0.071
0.2	0.114	**0.094**	0.068	0.101	0.113	0.079
0.4	0.173	0.153	0.089	0.156	0.143	0.105
0.6	0.241	0.173	0.113	0.019	0.185	0.139
0.8	0.351	0.199	0.163	0.283	0.195	0.177
0.95	0.443	0.243	0.202	0.396	0.263	0.246
ρ	Normal			Contaminated Normal		
	None	PW	CO	None	PW	CO
0	0.062	0.102	0.079	0.061	0.095	0.074
0.2	0.113	0.118	0.098	0.099	0.101	0.076
0.4	0.163	0.154	0.121	0.165	0.134	0.098
0.6	0.221	0.181	0.142	0.234	0.141	0.124
0.8	0.312	0.214	0.163	0.335	0.163	0.167
0.95	0.413	0.273	0.196	0.421	0.245	0.248

* Bold values represent observed significance levels that do not differ from the nominal 5% with 95 % confidence

The results of Tables (1) and (2) show that, the CO correction yields observed significance levels that are closer to the nominal 5% than those from the PW correction for the Wald and LR tests. Overall, the CO / Wald combination seems to perform better than any other correction / test combination. In addition, it is interesting to note that the uncorrelated Wald test has very high observed levels of significance when P = 0. However, the CO / Wald combination actually has observed levels of significance closer to the nominal level when ρ = 0 than does the uncorrelated Wald test.

Generally, it be noted that, the rejection rates under the null hypothesis are quite high for all of the tests examined. This is may be due to two possible reasons. First, consider the fact that the asymptotic chi-square critical values are used in assessing the observed significance levels. It may be that the sample size of 20 is not large enough to justify the reliance on the asymptotic distribution. Second, the CO and PW corrections are based on estimates of the true autocorrelation coefficient, ρ.

5. Conclusion:

In this paper, using Monte Carlo simulation, we compare the performances of two procedures for testing the significance of the slope coefficient in small-sample LAD simple regression: the Wald and Likelihood Ratio test statistics. The Wald and LR tests employ an estimate of the scale parameter proposed by Mckean and Schrader (1984).

In addition to the inferential approaches we consider two corrections for serial correlation: analogues to the Cochrane-Orcutt (CO) and Prais Winsten (PW) approaches widely employed in the Least Squares context. The various approaches for correction and inference are compared based on observed significance levels. The simulation results indicates that correction for autocorrelated errors is important for larger ρ, although correction clearly does not remove the full effect of the serial correlation. The CO approach generally yields better results than the PW approach for inference, although the reverse seems to be true for model fit. Thus, based on the level of significance, the CO / Wald combination appears to be preferred.

Finally, the results of this paper suggest several areas for future research, which should lead to a more complete understanding of inference in least absolute deviation regression. This study has considered the case of simple regression and a single sample size. Interesting extensions would include investigation the sensitivity of the results to sample size and the extension to multiple regression.

References

Andrews, D.W.K (1986): "A note on the unbiasedness of feasible GLS, quasi –maximum likelihood, robust, adaptive and spectral estimators of the linear model" ***Econometrica***, 54, 687-698

Banerjee, A., Dolado, J., Galbraith, J. W., and Hendry, D.F. (1993): "Co-Integration, Error correlation and the Econometric Analysis of Non-Stationary Data" ***Oxford University Press***.

Bassett, G., and Koenker, R. (1978): "Asymptomatic theory of least absolute error regression" ***Journal of the American statistical Association***, 73, 618-622.

Cade, B.S., and Richards, J.D. (1996): "Permutation tests for least absolute deviation regression" ***Biometrics***, 52, 886-902.

Cochrane, D., and Orcutt, G.H. (1949): "Application of least squares regression relationship containing autocorrelated error terms" ***Journal of the American statistical Association***, 44, 32-61.

Coursey, D. and Nyquist, H. (1983): "On least absolute error estimation of linear regression models with dependent stable residuals" ***Review of Economics and statistics,*** 65, 687-692

Dielman, T.E., and Pfaffenberger, R. (1990): "Tests of linear hypothesis in LAV regression" ***Communications in Statistics - Simulation and Computation***, 19, 1179-1199.

Greene, W.H. (2003): "Econometric Analysis 5th ed., ***Prentice Hall***.

Gujarati, D.N. (1995): "Basic Econometrics" ***McGraw-Hill***.

Koenker, R., and Bassett, G. (1982): "Tests of linear hypothesis and L_1 estimation" ***Econometrica***, 50, 1577-1583.

Mckean, J., and Schrader, R. (1984): "A comparison of methods for studentizing the sample median" ***Communication in Statistics - Simulation and Computation***, 13, 751-773.

Morgenthaler, S. (1992): "Least absolute deviation fits for mgeneralized linear models" ***Biometrika***, 79, 747 – 754

Zhang, J., and Boos, D.D. (1994): "Adjusted power estimates in Monte Carlo experiments" ***Communication in Statistics - Simulation and Computation***, 23, 165-173