THE LAD TEST OF A UNIT ROOT NULL WITH FAT TAILED INNOVATIONS

Valery Shardyko

Belarusian State University, Minsk, Belarus, e-mail: *vsh@it.org.by*

Abstract. This paper considers the least absolute deviations (LAD) test of a unit root null without drift when distributions of innovations are fat tailed. The limiting distributions of the test statistics are expressed in terms of functionals of a standard Brownian motion. Percentiles are tabulated. A Monte Carlo experiment indicates that the LAD tests dominate the Dickey-Fuller tests when innovations are fat tailed, except distributions with extremely heavy tails such as Cauchy.

1. Introduction

Economic time series often exhibit nonstationary behavior. In addition, a consensus exists, that distributions of first differences of many time series encountered, for example, in finance, are fat tailed. Unit root tests developed by Fuller (1976) and Dickey and Fuller (1979, 1981) and many alternatives are based on least squares (LS) procedure. There is a considerable drawback in applying these tests in small samples when innovations are fat tailed.

This paper proposes a procedure of unit root testing, which is robust in the situations of fat tailed distributions. The test is based on the least absolute deviations (LAD).

Section 2 gives the limit distributions of the new tests statistics developed here, while Section 3 presents the table of critical values. In Section 4, the empirical powers of the LAD unit root tests are compared to powers of the Dickey-Fuller tests. Section 5 concludes.

2. Least absolute deviations estimator and its limiting distribution

Suppose that the regression model is

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i, \qquad i = 1, \ldots, n,$$

where $y_i \in \Re^1$ is a dependent variable, $\mathbf{x}_i \in \Re^{m\times 1}$ is a vector of regressors, $\beta \in \Re^{m\times 1}$ is a vector of unknown parameters, and u_i is i.i.d. random error. The estimation of the parameters in LAD regression is done by minimizing the sum of the absolute values of the residuals

$$\hat{\beta}^{LAD} = \arg\min_{\beta\in\Re^{m\times 1}} \sum_{i=1}^{n} \left| y_i - \mathbf{x}_i^T \beta \right|.$$

The technique is important when the errors might be supposed to have fat tailed distributions. In particular, when the distribution of errors u_i is Laplace (i.e., double exponential), the LAD estimator of β is the maximum likelihood (ML) estimator.

The LAD estimator can be obtained as the solution to linear-programming (LP) problem. Though computationally cumbersome, the method is feasible. Tukey suggested to compute the LAD estimates by an iterative weighted least squares (WLS) method instead of LP.

In the framework of time series with unit root properties of the LAD estimator may be explored as follows. Suppose $\{y_t\}$ is a univariate AR(1) process with unit root without drift,

$$y_t = y_{t-1} + u_t, \tag{1}$$

where $y_0 = 0$ and innovations u_t are i.i.d. with mean zero and variance σ_u^2. Let $\{y_t\}_{t=1}^T$ be an observed time series of length T generated by the process (1).

Consider the following AR(1) regressions estimated by the LAD conditional on y_0:

$$y_t = \rho y_{t-1} + u_t, \tag{2}$$

$$y_t = \alpha + \rho y_{t-1} + u_t. \tag{3}$$

Mark them as Case 1 and Case 2, respectively, as in Hamilton (1994, Ch. 17). Then under the null hypothesis that $\rho = 1$ in Case 1 and $(\alpha, \rho)' = (0, 1)'$ in Case 2, the limiting distributions of the nonstudentized unit root statistics $T(\hat{\rho}^{LAD} - 1)$ are given in Theorem 1.

THEOREM 1 *Let time series $\{y_t\}$ is generated by (1). Define also $\psi(u) = \operatorname{sgn}(u)$. Then for the regression specification (2) estimated by LAD, as $T \to \infty$*

$$T(\hat{\rho}^{LAD} - 1) \xrightarrow{d} \frac{1}{2f(F^{-1}(\frac{1}{2}))} \frac{\sigma_{u\psi} \int_0^1 W_1(r)dW_1(r) + (\sigma_u^2\sigma_\psi^2 - \sigma_{u\psi}^2)^{\frac{1}{2}} \int_0^1 W_1(r)dW_2(r)}{\sigma_u^2 \int_0^1 [W_1(r)]^2 dr}. \tag{4}$$

For the regression specification (3) *estimated by LAD, as $T \to \infty$*

$$T(\hat{\rho}^{LAD} - 1) \xrightarrow{d} \frac{1}{2f(F^{-1}(\frac{1}{2}))} \frac{\sigma_{u\psi} \int_0^1 \underline{W}_1(r)dW_1(r) + (\sigma_u^2\sigma_\psi^2 - \sigma_{u\psi}^2)^{\frac{1}{2}} \int_0^1 \underline{W}_1(r)dW_2(r)}{\sigma_u^2 \int_0^1 [\underline{W}_1(r)]^2 dr}, \tag{5}$$

where $W_1(r)$ and $W_2(r)$ are standard Brownian motions, $\underline{W}_1(r) \equiv W_1(r) - \int_0^1 W_1(r)dr$ is a demeaned Brownian motion, $\sigma_u^2 \equiv Var\{u_t\}$, $\sigma_\psi^2 \equiv Var\{\psi(u_t)\}$, $\sigma_{u\psi} \equiv Cov\{u_t, \psi(u_t)\}$, $f(\cdot)$ and $F(\cdot)$ are the density and distribution functions of u_t, respectively, and the symbol $\xrightarrow{d}$ denotes convergence in distribution.

Proof follows immediately from Theorem 3.3, Corollary 3.2, and Remark 5.2 of Koenker and Xiao (2002). Thus, the details are omitted.

Now assume that the distribution of innovations u_t is Laplace with the density $f(u) = \frac{1}{2}\lambda \exp(-\lambda|u|)$, where λ is a parameter, such that the LAD estimator is also the ML estimator. Then noting that $Var\{\psi(u_t)\} = 1$, $Cov\{u_t, \psi(u_t)\} = (\frac{1}{2}Var\{u_t\})^{1/2}$, and $f(F^{-1}(\frac{1}{2})) = f(0) = (2\sigma_u^2)^{-1/2}$, Corollary 1 results as a by-product of Theorem 1.

COROLLARY 1 *Let time series $\{y_t\}$ is generated by (1) where $u_t \sim Laplace(\lambda)$. Then for the regression specification (2) estimated by LAD, as $T \to \infty$*

$$T(\hat{\rho}^{LAD} - 1) \xrightarrow{d} \frac{\int_0^1 W_1(r)dW_1(r) + \int_0^1 W_1(r)dW_2(r)}{2\int_0^1 [W_1(r)]^2 dr}. \tag{6}$$

For the regression specification (3) *estimated by LAD, as* $T \to \infty$

$$T\left(\hat{\rho}^{LAD} - 1\right) \xrightarrow{d} \frac{\int_0^1 \underline{W}_1(r)dW_1(r) + \int_0^1 \underline{W}_1(r)dW_2(r)}{2\int_0^1 \left[\underline{W}_1(r)\right]^2 dr}. \tag{7}$$

3. Empirical distribution of test statistics

An important point is that the asymptotic distributions of the test statistics given by (6) and (7) do not depend on any nuisance parameters. Therefore, a priori knowledge about a parameter of the distribution is not required for a unit root test.

Sample size, T	Probability that $T\left(\hat{\rho}^{LAD} - 1\right)$ is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
					Case 1			
50	–11.41	–8.65	–6.61	–4.65	0.88	1.24	1.60	2.08
100	–11.43	–8.67	–6.64	–4.68	0.84	1.18	1.50	1.93
250	–11.47	–8.71	–6.68	–4.72	0.82	1.15	1.45	1.84
500	–11.48	–8.73	–6.70	–4.74	0.82	1.13	1.43	1.81
∞	–11.50	–8.74	–6.72	–4.76	0.81	1.12	1.41	1.79
					Case 2			
50	–16.76	–13.63	–11.25	–8.86	–0.32	0.29	0.82	1.49
100	–16.60	–13.49	–11.15	–8.80	–0.41	0.18	0.67	1.27
250	–16.46	–13.40	–11.10	–8.79	–0.47	0.10	0.57	1.13
500	–16.40	–13.37	–11.08	–8.79	–0.50	0.07	0.54	1.08
∞	–16.35	–13.34	–11.07	–8.79	–0.52	0.04	0.50	1.03

Table 1 Critical values for the unit root test based on LAD estimates

Table 1 contains percentiles for the null distributions of the nonstudentized unit root statistics created using MacKinnon's (2000) response surfaces technique. Following MacKinnon, the test statistics for 22 various series length T ranging from 20 to 1200 were generated by Monte Carlo simulation. For every sample size 100 experiments were performed, each with 100,000 replications. In order to simplify the computational issues, Tukey's procedure was adopted.

4. Empirical power of tests

Table 2 gives information on the power of the tests for sample size $T = 100$ obtained through series of Monte Carlo sampling experiments, each with 10,000 replications.

Distribution	$\rho = 0.7$			$\rho = 0.8$			$\rho = 0.9$			$\rho = 0.95$			$\rho = 1.0$		
	α=0	α=.5	α=1	α=0	α=.5	α=1	α=0	α=.5	α=1	α=0	α=.5	α=1	α=0	α=.5	α=1
	Test statistics $T\left(\hat{\rho}^{OLS} - 1\right)$, Case 1 of the Dickey-Fuller unit root test														
Laplace	1.00	0.93	0.02	1.00	0.24	0.00	0.76	0.00	0.00	0.31	0.01	0.00	0.05	0.00	0.00
Cauchy	1.00	1.00	1.00	1.00	1.00	0.99	0.84	0.77	0.64	0.19	0.16	0.13	0.03	0.03	0.02

$t(2)$	1.00	1.00	0.94	1.00	0.94	0.46	0.78	0.31	0.06	0.28	0.06	0.01	0.04	0.01	0.00
$t(4)$	1.00	1.00	0.41	1.00	0.68	0.02	0.77	0.05	0.00	0.31	0.00	0.00	0.05	0.00	0.00
$t(6)$	1.00	0.99	0.17	1.00	0.52	0.00	0.77	0.01	0.00	0.31	0.00	0.00	0.05	0.00	0.00
$S(1.2)$	1.00	1.00	1.00	1.00	0.99	0.92	0.82	0.65	0.40	0.23	0.15	0.08	0.04	0.00	0.01
$S(1.6)$	1.00	1.00	0.94	1.00	0.93	0.44	0.78	0.31	0.07	0.28	0.06	0.01	0.05	0.01	0.00
Normal	1.00	0.95	0.01	1.00	0.24	0.00	0.76	0.00	0.00	0.31	0.00	0.00	0.05	0.00	0.00
Test statistics $T(\hat{\rho}^{LAD}-1)$, Case 1 of the LAD unit root test															
Laplace	1.00	0.94	0.10	1.00	0.37	0.00	0.89	0.00	0.00	0.34	0.01	0.00	0.04	0.00	0.00
Cauchy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.76	0.04	0.04	0.04	0.00	0.00	0.00
$t(2)$	1.00	1.00	0.94	1.00	0.97	0.53	0.93	0.35	0.03	0.28	0.03	0.00	0.02	0.00	0.00
$t(4)$	1.00	0.99	0.57	1.00	0.78	0.05	0.85	0.06	0.00	0.39	0.00	0.00	0.05	0.00	0.00
$t(6)$	1.00	0.97	0.40	1.00	0.65	0.02	0.82	0.04	0.00	0.41	0.00	0.00	0.07	0.00	0.00
$S(1.2)$	1.00	1.00	1.00	1.00	1.00	0.96	0.98	0.82	0.43	0.13	0.07	0.02	0.00	0.00	0.00
$S(1.6)$	1.00	1.00	0.94	1.00	0.96	0.55	0.89	0.39	0.07	0.31	0.04	0.00	0.03	0.00	0.00
Normal	1.00	0.91	0.18	0.99	0.45	0.01	0.79	0.02	0.00	0.44	0.00	0.00	0.09	0.00	0.00
Test statistics $T(\hat{\rho}^{OLS}-1)$, Case 2 of the Dickey-Fuller unit root test															
Laplace	1.00	1.00	1.00	0.96	0.96	0.96	0.47	0.42	0.31	0.20	0.06	0.00	0.05	0.00	0.00
Cauchy	1.00	1.00	1.00	0.98	0.98	0.98	0.35	0.35	0.34	0.10	0.10	0.09	0.05	0.05	0.03
$t(2)$	1.00	1.00	1.00	0.97	0.97	0.97	0.44	0.43	0.39	0.16	0.13	0.07	0.05	0.01	0.00
$t(4)$	1.00	1.00	1.00	0.96	0.96	0.96	0.47	0.44	0.37	0.19	0.11	0.02	0.05	0.00	0.00
$t(6)$	1.00	1.00	1.00	0.95	0.95	0.96	0.47	0.43	0.35	0.19	0.09	0.01	0.05	0.00	0.00
$S(1.2)$	1.00	1.00	1.00	0.98	0.98	0.98	0.39	0.38	0.38	0.12	0.12	0.10	0.06	0.00	0.02
$S(1.6)$	1.00	1.00	1.00	0.97	0.96	0.97	0.44	0.43	0.40	0.16	0.12	0.07	0.05	0.01	0.00
Normal	1.00	1.00	1.00	0.95	0.96	0.96	0.47	0.42	0.31	0.19	0.06	0.00	0.05	0.00	0.00
Test statistics $T(\hat{\rho}^{LAD}-1)$, Case 2 of the LAD unit root test															
Laplace	1.00	1.00	1.00	0.99	0.99	0.99	0.56	0.53	0.47	0.18	0.07	0.01	0.04	0.00	0.00
Cauchy	1.00	1.00	1.00	1.00	1.00	1.00	0.14	0.14	0.14	0.00	0.00	0.00	0.00	0.00	0.00
$t(2)$	1.00	1.00	1.00	1.00	1.00	1.00	0.49	0.48	0.46	0.10	0.07	0.03	0.01	0.00	0.00
$t(4)$	1.00	1.00	1.00	0.98	0.98	0.99	0.59	0.58	0.54	0.23	0.15	0.04	0.05	0.00	0.00
$t(6)$	1.00	1.00	1.00	0.97	0.97	0.98	0.62	0.59	0.55	0.28	0.17	0.04	0.07	0.00	0.00
$S(1.2)$	1.00	1.00	1.00	1.00	1.00	1.00	0.29	0.29	0.29	0.02	0.02	0.01	0.00	0.00	0.00
$S(1.6)$	1.00	1.00	1.00	0.99	0.99	0.99	0.51	0.50	0.48	0.15	0.11	0.06	0.03	0.00	0.00
Normal	1.00	1.00	1.00	0.96	0.96	0.96	0.63	0.60	0.54	0.34	0.18	0.03	0.11	0.00	0.00

Table 2 Empirical Power and Size of the unit root tests of nominal size 0.05, T = 100

Notes: $t(\cdot)$ states for Student's t distribution with degrees of freedom *df* given in parentheses, $S(\cdot)$ states for stable Paretian distribution with a characteristic exponent *alpha* in parentheses.

The data were generated from the model $y_t = \alpha + \rho y_{t-1} + u_t$ with α taking values 0, 0.5, and 1. When $\rho = 1$, y_t is a unit root process and the empirical rejection rate gives the empirical size of the tests. For the choice of alternatives, AR(1) with $\rho = 0.7, 0.8, 0.9, 0.95$ were considered.

For Case 1 when $\alpha = 0$ the LAD unit root test is slightly more powerful for all distributions except Cauchy and stable law with *alpha* = 1.2, where $\rho = 0.95$. There is little loss in accuracy with respect to the size of the test. However, the Dickey-Fuller test reports very accurate size.

For Case 2 when $\alpha \neq 0$ the LAD test again shows higher power than conventional Dickey-Fuller test in all cases except distributions with extremely heavy tails such as Cauchy and stable law with *alpha* = 1.2 where $\rho = 0.90$. For $\rho = 0.95$ both the Dickey-Fuller and the LAD test have little power. Thus, the predicted superiority of the LAD unit root test in the presence of fat tailed innovations is again confirmed.

5. Conclusions

The technique presented gives a test of a unit root without drift in univariate time series against stationary alternative when distribution of innovations is fat tailed. Though the distribution theory underlying this procedure is asymptotic, the critical values tabulated via extensive simulation for finite samples are provided.

The following general conclusion can be drawn from the Monte Carlo results: (1) the proposed LAD unit root test has higher power than the Dickey-Fuller test in the presence of fat tailed innovations, except distributions with extremely heavy tails such as Cauchy; (2) the LAD test has little size distortion in Case 1 and is conservative in Case 2. The puzzling result is that the LAD test demonstrates better performance even in the presence of Gaussian innovations.

The empirical power study suggests that the proposed LAD unit root test should be preferred to the conventional Dickey-Fuller test when distribution of innovations is moderate fat tailed.

References

[1] Dickey, D. A., and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-431.

[2] Dickey, D. A., and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, 49, 1057-1072.

[3] Fuller, W. A. (1976). *Introduction to Statistical Time Series*, New York: John Wiley and Sons.

[4] Hamilton, J. D. (1994). *Time Series Analysis*, Princeton: Princeton University Press.

[5] Koenker, R., and Z. Xiao (2002). Quantile autoregression, unit roots and asymmetric interest rate dynamics, preprint, Department of Economics, University of Illinois at Urbana-Champaign.

[6] MacKinnon, J. G. (2000). Computing numerical distribution functions in econometrics, *in* "High Performance Computing Systems and Applications" (A. Pollard, D. Mewhort, and D. Weaver, Eds.), pp. 455-470, Amsterdam: Kluwer.

[7] Madala, G. S. (1977). *Econometrics*, New York: McGraw-Hill.