

# Bootstrap tests for unit roots based on LAD estimation

Marta Moreno<sup>1</sup>, Juan Romo<sup>\*,2</sup>



*Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid 26,  
28903 Getafe, Madrid, Spain*

Received 11 May 1998; accepted 2 April 1999

## Abstract

In this paper we propose a new bootstrap test for unit roots in first-order autoregressive models based on least absolute deviation (LAD) estimators. It is well known that the behaviour of this estimator when the distribution is heavy tailed is very good compared with least-squares estimation. The innovations distribution dependence of the LAD asymptotic law is overcome using bootstrap, which automatically approaches the target distribution. Our strategy avoids the usual problem of estimating the variance matrix and the density at zero, and makes also unnecessary the construction of distribution free statistics through linear combinations with the least-squares estimator. We provide the bootstrap functional limit theory necessary to prove the asymptotic validity of the procedure. Moreover, a large simulation study shows that our test has very good power behaviour compared with others proposed in the literature.

*MSC:* 62E20; 62E25; 62M10

*Keywords:* Autoregressive process; Bootstrap; Least absolute deviation; Unit root

## 1. Introduction

Let  $\{X_t\}$ ,  $t = 1, 2, \dots$  be a first-order autoregressive process defined by

$$X_t = \beta X_{t-1} + u_t, \quad X_0 = 0,$$

where  $\{u_t\}$  is a sequence of independent and identically distributed random variables with  $E(u_t) = 0$  and  $\text{Var}(u_t) = \sigma_1^2 < \infty$ . We are interested in testing the null hypothesis

$$H_0: \beta = 1$$

---

\* Corresponding author. Tel.: +34-916-249-805; fax: +34-916-249-847.

*E-mail address:* romo@est-econ.uc3m.es (J. Romo)

<sup>1</sup> Research partially supported by DGICYT PB93-0232.

<sup>2</sup> Research partially supported by DGICYT PB93-0232 and HCM R N CHRX-CT94-0514 CONTRACT(EUROPEAN UNION).

The best-known unit root test was proposed by Dickey and Fuller (1979) and is based on the least-squares estimator of the parameter of the autoregressive process. Basawa et al. (1991a, b) showed that the naive bootstrap fails in this context and presented a modified sequential bootstrap which correctly approaches the limit distribution of the least-squares estimator. Ferretti and Romo (1996) considered a new approach resampling from the estimated residuals and proved its validity. Our main interest in this paper is to propose and study an alternative test based on a different estimator, the least absolute deviation estimator (LAD)

$$\hat{\beta}_{\text{LAD}} = \arg \min_{\beta \in \mathbb{R}} \left[ n^{-1} \sum_{t=1}^n |X_t - \beta X_{t-1}| \right]. \quad (1)$$

The literature about LAD-estimation is quite extensive. Basset and Koenker (1978) found the asymptotic distribution of the least absolute deviation estimator of the regression parameter with fixed regressors and stated that it is more efficient than the least-squares estimator for any error distribution for which the median is more efficient than the mean. Also in regression parameter estimation, Pollard (1991) proved that the LAD estimator is asymptotically normal both for fixed and stochastic regressors. As a particular case, he obtained the asymptotic distribution of the LAD estimator of the parameter of a stationary first-order autoregressive process with finite variance errors, and he showed that, in the case of Cauchy errors, the rate of convergence is of order  $n$ .

Knight (1989, 1991) studied the LAD estimator asymptotic distribution in an integrated first order autoregressive process for independent innovations with infinite variance. He compared it with the least-squares estimator and found that, in the case of infinite variance, the LAD estimator has a faster rate of convergence.

The problem of robust estimation of the parameters in autoregressive models is considered by Koul (1991) in the context of generalized M-estimators. The asymptotic normality of the LAD estimator follows as a consequence of a more general result contained in the paper. Also, Phillips (1991) considered the problem of LAD estimation but from a different point of view. His main interest was in adapting the classical approach to this kind of non-regular problems using generalized functions. In an heuristic way he established Knight's (1989) results using this approach. For a complete and deep treatment of this subject see Bloomfield and Steiger (1983).

Herce (1996) considers the case of finite variance even with non-independent errors. In particular, when the errors are independent he establishes that, under certain assumptions, when  $\beta = 1$ ,

$$n(\hat{\beta}_{\text{LAD}} - 1) \xrightarrow{w} \frac{\int_0^1 B_1 dB_2}{2f(0) \int_0^1 B_1^2},$$

where  $B(s) = (B_1(s), B_2(s))'$  is a bivariate Brownian motion that satisfies  $B(s) = \Sigma^{1/2} W(s)$ , where  $W(s) = (W_1(s), W_2(s))'$  is a standard Wiener process with sample paths in  $C[0, 1]^2$

and independent components, and  $\Sigma$  is the covariance matrix of  $(u_t, \text{sign}(u_t))$ , i.e.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \quad (2)$$

where  $\sigma_1^2 = E[u_t^2]$ ,  $\sigma_{12} = E[u_t \text{sign}(u_t)] = E[|u_t|]$  and  $\sigma_2^2 = E[(\text{sign}(u_t))^2] = 1$ .

In this paper we will show that the bootstrap version of  $\hat{\beta}_{\text{LAD}}$  converges to the same limit under adequate modifications to the general bootstrap framework. The aim is to find a unit root test appropriate for series where the errors have finite variance but heavy tails. This is particularly interesting in financial series where the assumption of normal distribution of the innovations is sometimes too strong. The simulations at the end of the paper show that the test we propose has a very good behaviour in the cases studied. Several distributions are considered and our robust unit root test seem to behave very well compared with others proposed in the literature in all the cases; as expected, it works specially well when the distribution is heavy tailed and symmetric.

In Section 2 the main results of this paper are presented: a bootstrap invariance principle is proved and the main theorem shows that the procedure correctly replicates the behaviour of the LAD-estimator. Section 3 shows how to use these results to test the existence of a unit root and presents simulation work to analyze its behaviour in finite samples. Finally, Section 4 concludes the paper.

## 2. Bootstrap least absolute deviation estimator

In this section we study the asymptotic behaviour of the bootstrap least absolute deviation estimator when the innovations are independent and identically distributed and they have both zero mean and median. We will propose a bootstrap scheme that provides a valid procedure in this framework. The assumptions needed are the following:

**Assumption 1.**  $\{u_t\}$  is a sequence of independent and identically distributed random variables with  $E(u_t) = 0$ ,  $E(\text{sign}(u_t)) = 0$ ,  $\text{var}(u_t) = \sigma_1^2 < \infty$  (this could be considered an almost symmetry condition).

**Assumption 2.** The density  $f$  of the errors  $\{u_t\}$  is continuous and positive at 0.

These assumptions are those in Hecce (1996) for independent errors. We present now the resampling method we propose. A complete and general review of the bootstrap literature can be found in Efron and Tibshirani (1993). The bootstrap for regression and autoregression models is also considered in González Manteiga et al. (1994). Our scheme is the following:

(i) Estimate  $\hat{\beta}_{\text{LAD}}$  as in (1) and obtain the sequence of residuals  $\varepsilon_t = X_t - \hat{\beta}_{\text{LAD}} X_{t-1}$ ,  $t = 1, \dots, n$ .

(ii) Denote by  $\hat{F}_n$  the empirical distribution function based on  $\{\varepsilon_t: t = 1, \dots, n\} \cup \{-\varepsilon_t: t = 1, \dots, n\}$  and take a random sample  $\{\varepsilon_{n,t}^*: t = 1, \dots, n\}$  from  $\hat{F}_n$ ; so that, the

random variables  $\{\varepsilon_{n,t}^*: t = 1, \dots, n\}$  are independent and identically distributed with distribution  $\hat{F}_n$ , conditionally on  $(X_1, \dots, X_n)$ .

(iii) The bootstrap sample  $\{X_{n,t}^*: t = 1, \dots, n\}$  is then recursively obtained from the model under  $H_0$  as  $X_{n,t}^* = X_{n,t-1}^* + \varepsilon_{n,t}^*$  for  $t = 1, \dots, n$  and  $X_{n,0}^* = 0$ .

Let  $F^s$  be the symmetrized distribution function  $F^s(x) = \frac{1}{2}[1 + F(x) - F(-x)]$ , where  $F(\cdot)$  is the distribution function of the errors. Note that in step (ii) we are resampling from a distribution that mimics the behaviour of  $F^s$  instead of  $F$ ; the reason to do this is that it is a usual condition in this kind of problems to consider innovations with mean and median equal to zero. We will see in the proof of the main theorem that this fact does not affect the asymptotic distribution of the statistic even if the distribution  $F$  is not symmetric.

The bootstrap LAD estimator is then defined as

$$\hat{\beta}_{\text{LAD}}^* = \arg \min_{\beta \in \mathbb{R}} \left[ n^{-1} \sum_{t=1}^n |X_{n,t}^* - \beta X_{n,t-1}^*| \right]. \quad (3)$$

The goal is to show that in this case the asymptotic distribution of  $n(\hat{\beta}_{\text{LAD}}^* - 1)$  is the same as that obtained by Hecce (1996) for the LAD estimators. To do this we need a bootstrap invariance principle. We will start by proving the lemmas needed to establish this invariance principle in Proposition 2.1.

We will use the following notation. Let  $a_{n,t}^* = \text{sign}(\varepsilon_{n,t}^*)$  and

$$\hat{\Sigma} = \begin{bmatrix} \text{Var}^*(\varepsilon_{n,t}^*) & \text{Cov}^*(\varepsilon_{n,t}^*, a_{n,t}^*) \\ \text{Cov}^*(\varepsilon_{n,t}^*, a_{n,t}^*) & \text{Var}^*(a_{n,t}^*) \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_1^{*2} & \hat{\sigma}_{12}^* \\ \hat{\sigma}_{12}^* & 1 \end{bmatrix},$$

where  $\hat{\sigma}_1^{*2} = \text{var}^*(\varepsilon_{n,t}^*) = (2n)^{-1} \sum (\varepsilon_t^2 + (-\varepsilon_t)^2) = n^{-1} \sum \varepsilon_t^2$ , and  $\hat{\sigma}_{12}^* = E^*(\varepsilon_{n,t}^* a_{n,t}^*) = E^*(|\varepsilon_{n,t}^*|) = (2n)^{-1} \sum (|\varepsilon_t| + |-\varepsilon_t|) = n^{-1} \sum |\varepsilon_t|$  (note that  $\hat{\sigma}_2^{*2} = \text{var}^*(a_{n,t}^*) = 1$ ). Hereafter,  $P^*$ ,  $E^*$ ,  $\text{Var}^*$  and  $\text{Cov}^*$  will denote, respectively, the bootstrap probability, expectation, variance and covariance conditionally on the sample  $(X_1, \dots, X_n)$ .

If we define

$$\hat{\Sigma}^{1/2} = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \quad (4)$$

so that

$$\begin{aligned} S_{11}^2 + S_{12}^2 &= \hat{\sigma}_1^{*2}, \\ S_{11}S_{12} + S_{22}S_{12} &= \hat{\sigma}_{12}^*, \\ S_{12}^2 + S_{22}^2 &= 1, \end{aligned}$$

we have that

$$\hat{\Sigma}^{-1/2} = \frac{1}{\Delta} \begin{bmatrix} S_{22} & -S_{12} \\ -S_{12} & S_{11} \end{bmatrix},$$

where  $\Delta = S_{11}S_{22} - S_{12}^2$ . Now,

$$\hat{\Sigma}^{-1/2} n^{-1/2} \sum_{t=1}^{[ns]} (\varepsilon_{n,t}^*, a_{n,t}^*)' = \begin{bmatrix} n^{-1/2} \Delta^{-1} \sum_{t=1}^{[ns]} (S_{22} \varepsilon_{n,t}^* - S_{12} a_{n,t}^*) \\ n^{-1/2} \Delta^{-1} \sum_{t=1}^{[ns]} (-S_{12} \varepsilon_{n,t}^* + S_{11} a_{n,t}^*) \end{bmatrix}.$$

Finally, let us call

$$\begin{aligned} \omega_{n,t}^{*(1)} &= S_{22} \varepsilon_{n,t}^* - S_{12} a_{n,t}^*, \\ \omega_{n,t}^{*(2)} &= -S_{12} \varepsilon_{n,t}^* + S_{11} a_{n,t}^* \end{aligned} \quad (5)$$

and define

$$S_{n,k}^{*(i)} = \sum_{t=1}^k \omega_{n,t}^{*(i)} \quad \text{for } i = 1, 2.$$

From this, we construct continuous-time processes by linear interpolation:

$$\begin{aligned} W_n^{*(i)}(s) &= \Delta^{-1} n^{-1/2} S_{n,[ns]}^{*(i)} + (ns - [ns]) \Delta^{-1} n^{-1/2} \omega_{n,[ns]+1}^{*(i)} \\ &\quad \text{for } s \in [0, 1] \quad \text{and } i = 1, 2. \end{aligned}$$

The first step is to prove the weak convergence of the finite-dimensional distributions for almost all samples  $(X_1, \dots, X_n)$ .

**Lemma 2.1.** *Conditionally on  $(X_1, \dots, X_n)$  and for almost all sample paths  $(X_1, X_2, \dots)$ , it holds that  $(W_n^{*(i)}(s_1), \dots, W_n^{*(i)}(s_d)) \xrightarrow{w} (W_i(s_1), \dots, W_i(s_d))$  as  $n$  tends to infinity, for all  $s_1, \dots, s_d \in [0, 1]^d$  and for  $i = 1, 2$ .*

**Proof.** It is enough to show that, for all  $r, s \in [0, 1]$  and for  $i = 1, 2$ ,  $(W_n^{*(i)}(r), W_n^{*(i)}(s)) \xrightarrow{w} (W_i(r), W_i(s))$  almost surely. Now, conditionally on  $(X_1, \dots, X_n)$ , it is easy to show that

$$\left\| (W_n^{*(i)}(r), W_n^{*(i)}(s)) - \Delta^{-1} n^{-1/2} \left( \sum_{t=1}^{[nr]} \omega_{n,t}^{*(i)}, \sum_{t=1}^{[ns]} \omega_{n,t}^{*(i)} \right) \right\|_{\infty} \xrightarrow{P^*} 0 \quad \text{a.s.}$$

and it suffices to prove that

$$\Delta^{-1} n^{-1/2} \left( \sum_{t=1}^{[nr]} \omega_{n,t}^{*(i)}, \sum_{t=1}^{[ns]} \omega_{n,t}^{*(i)} \right) \xrightarrow{w} (W_i(r), W_i(s)) \quad \text{a.s.}$$

This is equivalent to show that, if  $r < s$ ,  $\Delta^{-1} n^{-1/2} (\sum_{t=1}^{[nr]} \omega_{n,t}^{*(i)}, \sum_{t=[nr]+1}^{[ns]} \omega_{n,t}^{*(i)})$  converges weakly to  $(W_i(r), W_i(s) - W_i(r))$  almost surely. But the components in the left-hand side are conditionally independent random variables with zero mean and variance one and the result follows by the bootstrap central limit theorem for triangular arrays obtained using Lindeberg's condition.  $\square$

Once we have weak convergence of finite-dimensional distributions, we need to prove the tightness of the sequence; to establish it, it is enough to prove the conditions included in the next lemma (see, e.g., Karatzas and Shreve, 1991, pp. 68–71).

**Lemma 2.2.** For  $i = 1, 2$  it holds that:

(i) For all  $\eta > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \delta^{-1} P^* \left\{ \max_{1 \leq j \leq [n\delta] + 1} |S_{n,k}^{*(i)}| > \eta \Delta n^{1/2} \right\} = 0$$

conditionally on  $(X_1, \dots, X_n)$  and for almost all sample paths  $(X_1, X_2, \dots)$ .

(ii) For all  $\eta > 0$  and  $T > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} P^* \left\{ \max_{\substack{1 \leq j \leq [n\delta] + 1 \\ 0 \leq k \leq [nT] + 1}} |S_{n,j+k}^{*(i)} - S_{n,k}^{*(i)}| > \eta \Delta n^{1/2} \right\} = 0$$

conditionally on  $(X_1, \dots, X_n)$  and for almost all sample paths  $(X_1, X_2, \dots)$ .

**Proof.** The proof of this lemma is similar to that in Karatzas and Shreve (1991) and will not be included here.

Now, we are ready to establish the bootstrap invariance principle in probability; this means that the distance between the law of the bootstrap statistic and the asymptotic distribution tends to zero in probability for any distance metrizing weak convergence.

**Proposition 2.1.** For the sequences  $\{e_{n,t}^*: t = 1, \dots, n\}$  and  $\{a_{n,t}^*: t = 1, \dots, n\}$  defined above, we have that

$$n^{-1/2} \sum_{t=1}^{[ns]} (e_{n,t}^*, a_{n,t}^*)' \xrightarrow{w} B(s) \quad \text{in probability,}$$

where  $B(s) = (B_1(s), B_2(s))' = \Sigma^{1/2} (W_1(s), W_2(s))'$ ,  $W_1(s)$  and  $W_2(s)$  are independent standard Wiener process and  $\Sigma$  is the covariance matrix in (2).

**Proof.** We have established in the previous lemmas that, for  $i = 1, 2$ ,

$$\Delta^{-1} n^{-1/2} S_{n,[ns]}^{*(i)} \xrightarrow{w} W_i(s),$$

the joint weak convergence

$$\Delta^{-1} n^{-1/2} \begin{pmatrix} S_{n,[ns]}^{*(1)} \\ S_{n,[ns]}^{*(2)} \end{pmatrix} \rightarrow \begin{pmatrix} W_1(s) \\ W_2(s) \end{pmatrix}$$

is obtained from  $\text{Cov}^*(S_{n,[ns]}^{*(1)}, S_{n,[ns]}^{*(2)}) = 0$  and from the independence of the errors. Now, given that the limit processes are Gaussian we have that they are independent and the joint convergence follows.

Now, using a bootstrap version of Slutsky's theorem, we get that

$$n^{-1/2} \sum_{t=1}^{[ns]} (e_{n,t}^*, a_{n,t}^*)' \xrightarrow{w} \Sigma^{1/2} (W_1(s), W_2(s))' = B(s) \quad \text{in probability.} \quad \square$$

The next proposition contains the remaining results that will be needed for the proof of the main theorem in this section.

**Proposition 2.2.** *The following results hold:*

- (i)  $n^{-3/2} \sum_{t=1}^n X_{n,t-1}^* \xrightarrow{w} \int_0^1 B_1(s) ds$  in probability.
- (ii)  $n^{-2} \sum_{t=1}^n X_{n,t-1}^{*2} \xrightarrow{w} \int_0^1 B_1^2(s) ds$  in probability.
- (iii)  $n^{-1} \sum_{t=1}^{n-1} a_{n,t+1}^* X_{n,t}^* \xrightarrow{w} \int_0^1 B_2(s) dB_2(s)$  in probability.

**Proof.** To make the notation simpler, let

$$U_n^*(s) = \sum_{j=1}^{[ns]} \varepsilon_{n,j}^* \quad \text{and} \quad V_n^*(s) = \sum_{j=1}^{[ns]} \text{sign}(\varepsilon_{n,j}^*) = \sum_{j=1}^{[ns]} a_{n,j}^*.$$

(i) In Proposition 2.1 we proved that  $n^{-1/2} U_n^*(s) \xrightarrow{w} B_1(s)$  in probability. Moreover,  $U_n^*(s)$  is a jump process such that

$$\int_{(t-1)/n}^{t/n} U_n^*(s) ds = n^{-1} U_n^* \left( \frac{t-1}{n} \right).$$

On the other hand,

$$\begin{aligned} n^{-3/2} \sum_{t=1}^n X_{n,t-1}^* &= n^{-1/2} \sum_{t=1}^n n^{-1} U_n^* \left( \frac{t-1}{n} \right) = n^{-1/2} \sum_{t=1}^n \int_{(t-1)/n}^{t/n} U_n^*(s) ds \\ &= \int_0^1 n^{-1/2} U_n^*(s) ds \end{aligned}$$

and  $\int_0^1 n^{-1/2} U_n^*(s) ds$  converges weakly to  $\int_0^1 B_1(s) ds$  in probability by the continuous mapping theorem. This proves (i).

(ii) The proof of this part is similar to that of part (i) for the process  $U_n^{*2}(s) = (\sum_{j=1}^{[ns]} \varepsilon_{n,j}^*)^2$ .

(iii) Using the notation we have introduced we can rewrite  $n^{-1} \sum_{t=1}^{n-1} a_{n,t+1}^* X_{n,t}^*$  as

$$G_n^* = n^{-1} \sum_{t=1}^{n-1} \left( V_n^* \left( \frac{t+1}{n} \right) - V_n^* \left( \frac{t}{n} \right) \right) U_n^* \left( \frac{t}{n} \right)$$

and we have to prove that  $G_n^* \xrightarrow{w} \int_0^1 B_1(s) dB_2(s)$  in probability.

Using the fact that  $B_1$  and  $B_2$  have continuous paths, and by the Skorokhod representation theorem (see, e.g., Dudley, 1989, p. 324) there exist a probability space  $\Omega$  and random elements  $U^n, V^n$  in  $D[0, 1]$  such that

$$(U^n, V^n) =_{\mathcal{L}} n^{-1/2} (U_n^*, V_n^*),$$

that is, the distribution of  $(U^n, V^n)$  is the same as that of  $n^{-1/2} (U_n^*, V_n^*)$  conditionally on the sample  $(X_1, \dots, X_n)$ , and

$$\|(U^n, V^n) - (B_1, B_2)\|_{\infty} \rightarrow 0 \quad \text{a.s.} \tag{6}$$

Let

$$G^n = \sum_{t=1}^{n-1} U^n \left( \frac{t}{n} \right) \left( V_n \left( \frac{t+1}{n} \right) - V_n \left( \frac{t}{n} \right) \right).$$

Then  $G^n$  and  $G_n^*$  have the same distribution. In order to show the weak convergence of  $G_n^*$  to  $\int_0^1 B_1(s) dB_2(s)$  in probability, it is sufficient to show that  $G^n \xrightarrow{P} \int_0^1 B_1(s) dB_2(s)$ .

By (6) and Egorov's theorem, given  $\varepsilon > 0$ , there is an event  $\Omega_\varepsilon \subset \Omega$  such that  $P(\Omega_\varepsilon) \geq 1 - \varepsilon$  and

$$\sup\{ |(U^n(\omega), V^n(\omega)) - (B_1(\omega), B_2(\omega))|_\infty : \omega \in \Omega_\varepsilon \} = \delta_n \rightarrow 0. \quad (7)$$

Note that  $\delta_n$  is a sequence of constants. We can choose integers  $N(n) \uparrow \infty$  such that

$$N(n)\delta_n^2 \rightarrow 0 \quad \text{and} \quad N(n)/n \rightarrow 0.$$

For each  $n$ , we can further choose a partition  $\{t_0, \dots, t_{N(n)}\}$  of  $[0, 1]$  such that

$$0 = t_0 < t_1(n) = \frac{n_1}{n} < t_2(n) = \frac{n_2}{n} < \dots < t_{N(n)} = \frac{n_{N(n)}}{n} = 1$$

with

$$\max\{|t_{i+1} - t_i| : 0 \leq i \leq N(n) - 1\} = o(1).$$

We first claim that

$$G^n = \sum_{k=1}^{N(n)} U^n(t_{k-1})(V^n(t_k) - V^n(t_{k-1})) + o_P(1). \quad (8)$$

Let

$$\begin{aligned} J_n &= G^n - \sum_{k=1}^{N(n)} U^n(t_{k-1})(V^n(t_k) - V^n(t_{k-1})) \\ &= \sum_{k=1}^{N(n)} \left[ \sum_{i=n_{k-1}}^{n_k-1} \left( U^n \left( \frac{i}{n} \right) - U^n(t_{k-1}) \right) \left( V^n \left( \frac{i+1}{n} \right) - V^n \left( \frac{i}{n} \right) \right) \right]. \end{aligned}$$

Now,

$$\begin{aligned} E(J_n^2) &= E \left[ \sum_{k=1}^{N(n)} \sum_{i=n_{k-1}}^{n_k-1} \left( U^n \left( \frac{i}{n} \right) - U^n(t_{k-1}) \right) \left( V^n \left( \frac{i+1}{n} \right) - V^n \left( \frac{i}{n} \right) \right) \right]^2 \\ &= n^{-2} E^* \left[ \sum_{k=1}^{N(n)} \sum_{i=n_{k-1}}^{n_k-1} \left( U_n^* \left( \frac{i}{n} \right) - U_n^*(t_{k-1}) \right) \left( V_n^* \left( \frac{i+1}{n} \right) - V_n^* \left( \frac{i}{n} \right) \right) \right]^2 \\ &= n^{-2} E^* \left[ \sum_{k=1}^{N(n)} \Phi_k^* \right]^2, \end{aligned}$$

where we have written

$$\Phi_k^* = \sum_{i=n_{k-1}}^{n_k-1} \left( U_n^* \left( \frac{i}{n} \right) - U_n^*(t_{k-1}) \right) \left( V_n^* \left( \frac{i+1}{n} \right) - V_n^* \left( \frac{i}{n} \right) \right).$$



Using the fact that  $(\varepsilon_{n,j}^*, a_{n,j}^*)'$  is a sequence of independent random variables it can be proved that

$$E^* \left[ \sum_{k=1}^{N(n)} \Phi_k^* \right]^2 = \sum_{k=1}^{N(n)} E^* [\Phi_k^{*2}] \leq \hat{\sigma}_1^2 (n_k - n_{k-1})^2.$$

Putting everything together,

$$\begin{aligned} E(J_n^2) &= n^{-2} E^* \left[ \sum_{k=1}^{N(n)} \Phi_k^* \right]^2 \leq \sum_{k=1}^{N(n)} \hat{\sigma}_1^2 \left( \frac{n_k}{n} - \frac{n_{k-1}}{n} \right)^2 \\ &\leq \hat{\sigma}_1^2 \max_{1 \leq k \leq N(n)} (t_k - t_{k-1}) \rightarrow 0 \end{aligned}$$

and Markov's inequality gives (8).

The next step is to show that

$$I_{\Omega_\varepsilon} \sum_{k=1}^{N(n)} U^n(t_{k-1}) (V^n(t_k) - V^n(t_{k-1})) = I_{\Omega_\varepsilon} \sum_{k=1}^{N(n)} B_1(t_{k-1}) (V^n(t_k) - V^n(t_{k-1})) + o_p(1),$$

i.e., that in  $\Omega_\varepsilon$  we can change  $U^n(t_{k-1})$  by  $B_1(t_{k-1})$  and the error goes to zero when the sample size goes to infinity. By Cauchy-Schwarz's inequality and (7),

$$\begin{aligned} &\left| \sum_{k=1}^{N(n)} [U^n(t_{k-1}) - B_1(t_{k-1})] I_{\Omega_\varepsilon} (V^n(t_k) - V^n(t_{k-1})) \right|^2 \\ &\leq \sum_{k=1}^{N(n)} (U^n(t_{k-1}) - B_1(t_{k-1}))^2 I_{\Omega_\varepsilon} \sum_{k=1}^{N(n)} (V^n(t_k) - V^n(t_{k-1}))^2 \\ &\leq N(n) \delta_n^2 \sum_{k=1}^{N(n)} (V^n(t_k) - V^n(t_{k-1}))^2. \end{aligned}$$

The expectation of this last expression is

$$\begin{aligned} E \left[ N(n) \delta_n^2 \sum_{k=1}^{N(n)} (V^n(t_k) - V^n(t_{k-1}))^2 \right] &= N(n) \delta_n^2 \sum_{k=1}^{N(n)} E[(V^n(t_k) - V^n(t_{k-1}))^2] \\ &= N(n) \delta_n^2 \sum_{k=1}^{N(n)} (t_k - t_{k-1}) \\ &= N(n) \delta_n^2 \rightarrow 0 \end{aligned}$$

and this proves what we wanted.

Now, it is straightforward that

$$\begin{aligned} \sum_{k=1}^{N(n)} B_1(t_{k-1}) (V^n(t_k) - V^n(t_{k-1})) &= - \sum_{k=1}^{N(n)} V^n(t_k) (B_1(t_k) - B_1(t_{k-1})) \\ &\quad + B_1(1) V^n(1). \end{aligned}$$

By a similar argument we can replace  $V^n(t_k)$  by  $B_2(t_k)$  and obtain

$$I_{\Omega_e} \sum_{k=1}^{N(n)} V^n(t_k) (B_1(t_k) - B_1(t_{k-1})) = I_{\Omega_e} \sum_{k=1}^{N(n)} B_2(t_k) (B_1(t_k) - B_1(t_{k-1})) + o_p(1),$$

so, using basic algebra

$$\begin{aligned} & I_{\Omega_e} \sum_{k=1}^{N(n)} B_1(t_{k-1}) (V^n(t_k) - V^n(t_{k-1})) \\ &= I_{\Omega_e} \sum_{k=1}^{N(n)} B_1(t_{k-1}) (B_2(t_k) - B_2(t_{k-1})) + o_p(1) \\ &= I_{\Omega_e} \int_0^1 B_1(t) dB_2(t) + o_p(1), \end{aligned}$$

where the last identity is given by the fact that

$$\begin{aligned} & E \left[ \sum_{k=1}^{N(n)} B_1(t_{k-1}) (B_2(t_k) - B_2(t_{k-1})) - \int_0^1 B_1(t) dB_2(t) \right]^2 \\ &= \sum_{k=1}^{N(n)} E \left[ \left( \int_{t_{k-1}}^{t_k} (B_1(t_{k-1}) - B_1(t)) dB_2(t) \right)^2 \right] \\ &\quad + \sum_{k=1}^{N(n)} \sum_{j \neq k} E \left[ \int_{t_{k-1}}^{t_k} (B_1(t_{k-1}) - B_1(t)) dB_2(t) \right. \\ &\quad \left. \times \int_{t_{j-1}}^{t_j} (B_1(t_{j-1}) - B_1(s)) dB_2(s) \right]. \end{aligned}$$

The second part of the right-hand side of the equality is clearly zero because  $B_1$  has independent increments. Therefore, we have that

$$\begin{aligned} & E \left[ \sum_{k=1}^{N(n)} B_1(t_{k-1}) (B_2(t_k) - B_2(t_{k-1})) - \int_0^1 B_1(t) dB_2(t) \right]^2 \\ &= \sum_{k=1}^{N(n)} \frac{\sigma_1^2}{2} (t_k - t_{k-1})^2 \\ &\leq \frac{\sigma_1^2}{2} \max_{1 \leq k \leq N(n)} (t_k - t_{k-1}) = o(1). \quad \square \end{aligned}$$

Following Knight (1991) and Hecce (1996) we will prove the convergence of  $n(\hat{\beta}_{\text{LAD}}^* - 1)$  using the convexity theorem in Knight (1989). The theorem says that if the finite-dimensional distributions of a sequence of convex stochastic processes  $Z_n(\cdot)$  converge weakly to those of  $Z(\cdot)$ , and  $Z(\cdot)$  has a unique minimum, then the minimizer of  $Z_n(\cdot)$  converges in distribution to the minimizer of  $Z(\cdot)$ .

The value we have to find is

$$\hat{\beta}_{\text{LAD}}^* = \arg \min_{\beta \in \mathbb{R}} \sum_{t=1}^n |X_{n,t}^* - \beta X_{n,t-1}^*|.$$

Defining  $\theta = n(\beta - 1)$  we can restate this as the problem of finding the minimum of

$$Z_n^*(\theta) = \sum_{t=1}^n (|\varepsilon_{n,t}^* - \theta n^{-1} X_{n,t-1}^*| - |\varepsilon_{n,t}^*|).$$

Moreover, if  $\hat{\theta}^*$  is a minimizer of  $Z_n^*(\theta)$ , then  $\hat{\theta}^* = n(\hat{\beta}_{\text{LAD}}^* - 1)$ ; so we have to prove that

$$Z_n^*(\cdot) \xrightarrow{w} Z(\theta) = -\theta \int_0^1 B_1(s) dB_2(s) + \theta^2 f(0) \int_0^1 B_1^2(s) ds \quad \text{in probability,}$$

because in this case

$$\hat{\theta}^* = n(\hat{\beta}_{\text{LAD}}^* - 1) \xrightarrow{w} \frac{\int_0^1 B_1(s) dB_2(s)}{2f(0) \int_0^1 B_1^2(s) ds} \quad \text{in probability,}$$

and this is exactly what we need to prove the validity of our bootstrap procedure. To use the convexity theorem, we have to prove the following lemma.

**Lemma 2.3.** *For  $Z_n^*(\theta)$  and  $Z(\theta)$  as defined above,*

$$Z_n^*(\theta) \xrightarrow{w} Z(\theta) \quad \text{in probability.}$$

**Proof.** It is easy to see that

$$\begin{aligned} Z_n^*(\theta) &= -\sum_{t=1}^n \text{sign}(\varepsilon_{n,t}^*) \theta n^{-1} X_{n,t-1}^* \\ &\quad + 2 \sum_{t=1}^n (\theta n^{-1} X_{n,t-1}^* - \varepsilon_{n,t}^*) [I(0 < \varepsilon_{n,t}^* < \theta n^{-1} X_{n,t-1}^*) \\ &\quad - I(\theta n^{-1} X_{n,t-1}^* < \varepsilon_{n,t}^* < 0)]. \end{aligned}$$

Proposition 2.2 gives us the asymptotic behavior of the first term:

$$-\sum_{t=1}^n \text{sign}(\varepsilon_{n,t}^*) \theta n^{-1} X_{n,t-1}^* \xrightarrow{w} -\theta \int_0^1 B_1 dB_2 \quad \text{in probability.}$$

Thus, we have to show that

$$\begin{aligned} &2 \sum_{t=1}^n (\theta n^{-1} X_{n,t-1}^* - \varepsilon_{n,t}^*) [I(0 < \varepsilon_{n,t}^* < \theta n^{-1} X_{n,t-1}^*) - I(\theta n^{-1} X_{n,t-1}^* < \varepsilon_{n,t}^* < 0)] \\ &\quad \xrightarrow{w} \theta^2 f(0) \int_0^1 B_1^2 \end{aligned}$$

in probability. We will do it in two steps. First, we will prove the weak convergence of

$$2 \sum_{t=1}^n (\theta n^{-1} X_{n,t-1}^* - \varepsilon_{n,t}^*) I(0 < \varepsilon_{n,t}^* < \theta n^{-1} X_{n,t-1}^*)$$

to

$$\theta^2 f(0) \int_0^1 B_1^2(s) I(0 < \theta B_1(s)) \, ds.$$

We will use the following definitions:

$$V_{nm}^*(\theta) = (\theta n^{-1} X_{n,t-1}^* - \varepsilon_{n,t}^*) I(0 < \varepsilon_{n,t}^* < \theta n^{-1} X_{n,t-1}^*) I(\theta n^{-1/2} X_{n,t-1}^* \leq m),$$

$$V_{nm}^*(\theta) = \sum_{t=1}^n V_{nm,t}^*,$$

$$\mu_{nm}^*(\theta) = E^*(V_{nm}^* | \varepsilon_{n,t-1}^*, \dots)$$

and

$$\mu_{nm}^*(\theta) = \sum_{t=1}^n \mu_{nm,t}^*(\theta).$$

Notice that the sequence  $\{V_{nm}^*(\theta) - \mu_{nm}^*(\theta)\}_{t=1}^n$  is a martingale difference sequence.

We will prove now that

$$\mu_{nm}^*(\theta) \xrightarrow{w} \frac{\theta^2 f(0)}{2} \int_0^1 B_1^2(s) I(0 < \theta B_1(s) \leq m) \, ds \quad \text{in probability.}$$

Let us define  $A_{n,t}(\theta) = \theta n^{-1} X_{n,t-1}^* I(0 < \theta n^{-1/2} X_{n,t-1}^* \leq m)$  and observe that

$$V_{nm}^*(\theta) = (A_{n,t}(\theta) - \varepsilon_{n,t}^*) I(0 < \varepsilon_{n,t}^* < A_{n,t}(\theta)),$$

so that

$$\begin{aligned} \mu_{nm}^*(\theta) &= \sum_{t=1}^n E^*[(A_{n,t}(\theta) - \varepsilon_{n,t}^*) I(0 < \varepsilon_{n,t}^* < A_{n,t}(\theta)) | \varepsilon_{n,t-1}^*, \dots] \\ &= \sum_{t=1}^n \int_0^{A_{n,t}(\theta)} (A_{n,t}(\theta) - x) \, d\hat{F}_n(x) \\ &= \sum_{t=1}^n \int_0^{A_{n,t}(\theta)} (\hat{F}_n(s) - \hat{F}_n(0)) \, ds. \end{aligned}$$

Changing the variable  $s = v/n^\alpha$ ,  $ds = dv/n^\alpha$  and since  $s \in (0, A_{n,t}(\theta))$  implies  $v \in (0, n^\alpha A_{n,t}(\theta))$ , we have that

$$\mu_{nm}^*(\theta) = \sum_{t=1}^n \int_0^{n^\alpha A_{n,t}(\theta)} (\hat{F}_n(v/n^\alpha) - \hat{F}_n(0)) \frac{dv}{n^{2\alpha}}$$

and

$$\left(\frac{v}{n^\alpha}\right)^{-1} (\hat{F}_n(v/n^\alpha) - \hat{F}_n(0)) = f(0) + o_P(1) \quad \text{if } \alpha < \frac{1}{2}.$$

So

$$\begin{aligned}
\mu_{nm}^*(\theta) &= \sum_{t=1}^n \int_0^{n^x A_{n,t}(\theta)} \frac{v}{n^{2x}} (f(0) + o_p(1)) dv \\
&= \frac{1}{2} \sum_{t=1}^n (f(0) + o_p(1)) A_{n,t}^2(\theta) \\
&= \frac{1}{2} (f(0) + o_p(1)) \theta^2 n^{-2} \sum_{t=1}^n X_{n,t-1}^{*2} I(0 < \theta n^{-1/2} X_{n,t-1}^* \leq m).
\end{aligned}$$

Using Propositions 2.1 and 2.2, part (ii), we have that

$$\mu_{nm}^*(\theta) \xrightarrow{w} \mu_m^*(\theta) = \frac{1}{2} f(0) \theta^2 \int_0^1 B_1^2(s) I(0 < \theta B_1(s) \leq m) ds$$

in probability as  $n$  tends to infinity. On the other hand,

$$\lim_{m \rightarrow \infty} \mu_m^*(\theta) = \frac{1}{2} f(0) \theta^2 \int_0^1 B_1^2(s) I(0 < \theta B_1(s)) ds.$$

By the Asymptotic Equivalence Lemma (see, e.g., White, 1984, p. 63), the limiting distribution of  $V_{nm}^*(\theta)$  is the same as that for  $\mu_{nm}^*(\theta)$  provided that  $V_{nm}^*(\theta) - \mu_{nm}^*(\theta) \xrightarrow{P} 0$ . But  $V_{nm}^*(\theta) - \mu_{nm}^*(\theta) = \sum_{t=1}^n (V_{tm}^*(\theta) - \mu_{tm}^*(\theta))$  and  $\{V_{tm}^*(\theta) - \mu_{tm}^*(\theta)\}$  is a martingale difference sequence; using the corresponding convergence result for them (see, e.g., Pollard, 1984, p. 171), we just have to show that, when  $n \rightarrow \infty$

$$\sum_{t=1}^n E^*(V_{tm}^{*2}(\theta) | \varepsilon_{n,t-1}^*, \dots) \rightarrow 0 \quad \text{in probability,}$$

but this can be proved using that  $A_{n,t}(\theta) \leq mn^{-1/2}$ .

We have shown the following two weak convergences:

$$V_{nm}^*(\theta) \rightarrow \mu_m^*(\theta) \quad \text{in probability, as } n \rightarrow \infty,$$

$$\mu_m^*(\theta) \rightarrow \mu^*(\theta) \quad \text{in probability, as } m \rightarrow \infty.$$

Now we have to deal with the error due to truncation, i.e., if we define

$$V_n^*(\theta) = (\theta n^{-1} X_{n,t-1}^* - \varepsilon_{n,t}^*) I(0 < \varepsilon_{n,t}^* < \theta n^{-1} X_{n,t-1}^*),$$

we have to show that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P\{|V_n^*(\theta) - V_{nm}^*(\theta)| \geq \varepsilon\} = 0$$

for each positive  $\varepsilon$ , because this implies that  $V_{nm}^*(\theta) \xrightarrow{w} \mu^*(\theta)$  in probability. Now

$$\begin{aligned}
&P\{|V_n^*(\theta) - V_{nm}^*(\theta)| \geq \varepsilon\} \\
&\leq P\left\{\sum_{t=1}^n (\theta n^{-1} X_{n,t-1}^* - \varepsilon_{n,t}^*) I(0 < \varepsilon_{n,t}^* < \theta n^{-1} X_{n,t-1}^*) I(\theta n^{-1/2} X_{n,t-1}^* > m) \geq \varepsilon\right\}
\end{aligned}$$

$$\begin{aligned} &\leq P \left\{ \sum_{t=1}^n I(\theta n^{-1/2} X_{n,t-1}^* > m) \geq 0 \right\} \\ &= P \left\{ \max_{1 \leq t \leq n} \theta n^{-1/2} X_{n,t-1}^* > m \right\}. \end{aligned}$$

So,

$$\begin{aligned} &\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P \{ |V_n^*(\theta) - V_{nm}^*(\theta)| \geq \varepsilon \} \\ &\leq \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \max_{1 \leq t \leq n} \theta n^{-1/2} X_{n,t-1}^* > m \right\} \\ &\leq \lim_{m \rightarrow \infty} P \left\{ \sup_{s \leq 1} \theta B(s) > m \right\} = 0. \end{aligned}$$

The term  $I(\theta n^{-1} X_{n,t-1}^* < \varepsilon_{n,t}^* < 0)$  is handled similarly, and therefore

$$Z_n^*(\theta) \xrightarrow{w} -\theta \int_0^1 B_1 dB_2 + \theta^2 f(0) \int_0^1 B_1^2 \quad \text{in probability.} \quad \square$$

Now we are ready to state the main theorem:

**Theorem 2.1.** For  $\hat{\beta}_{\text{LAD}}^*$  defined as in (3),

$$\hat{\theta}^* = n(\hat{\beta}_{\text{LAD}}^* - 1) \xrightarrow{w} \frac{\int_0^1 B_1(s) dB_2(s)}{2f(0) \int_0^1 B_1^2(s) ds}$$

in probability.

**Proof.** Now the proof is straightforward. Note that  $Z_n^*(\theta)$  and  $Z(\theta)$  are respectively minimized at  $\hat{\theta}^* = n(\hat{\beta}_{\text{LAD}}^* - 1)$  and  $\int_0^1 B_1 dB_2 / 2f(0) \int_0^1 B_1^2$ ; using the convexity theorem of Knight (1989), we have that

$$n(\hat{\beta}_{\text{LAD}}^* - 1) \xrightarrow{w} \frac{\int_0^1 B_1 dB_2}{2f(0) \int_0^1 B_1^2} \quad \text{in probability.} \quad \square$$

**Remark.** Hecce (1996) states that even when the median of the innovations is not zero, the statistic converges; but in this case the convergence is at rate  $n^{1/2}$  and the limiting distribution is different. The same result can be proved using bootstrap. The only difference in the resampling algorithm is that in step (ii) we should resample from the empirical distribution of the residuals centered in mean ( $\hat{F}_n$  would be, in this case, the empirical distribution of  $\{\hat{\varepsilon}_t\} = \{\varepsilon_t\} - n^{-1} \sum_{t=1}^n \varepsilon_t$ ). The asymptotic distribution of the statistic is

$$n^{1/2}(\hat{\beta}_{\text{LAD}}^* - 1) \xrightarrow{w} \frac{m \int_0^1 B_1(s) ds}{2f(0) \int_0^1 B_1^2(s) ds},$$

where  $m$  is the median of the distribution of the errors and  $B_1$  and  $B_2$  are analogous to the previous ones.

### 3. LAD-based unit root tests

We have seen that the distribution of the statistic  $n(\hat{\beta}_{\text{LAD}} - 1)$  is rather complicated and depends on the distribution of the errors through  $f(0)$  and  $\Sigma$ . Even if we multiply by a consistent estimator of the density, the distribution still depends on the covariance matrix and the tabulation of the critical values would be specific for each case. Herce (1996) suggests two different tests based on this statistic. First, he proposes a linear combination of the LAD estimator with the least-squares estimator, namely

$$L_\beta = \frac{2\hat{f}(0)\hat{\sigma}_1^2}{\hat{\Delta}^{1/2}}n(\hat{\beta}_{\text{LAD}} - 1) - \frac{\hat{\sigma}_{12}}{\hat{\Delta}^{1/2}}n(\hat{\beta}_{\text{LS}} - 1) \xrightarrow{w} \left[ \int_0^1 W_1^2 \right]^{-1/2} \eta, \quad (9)$$

where  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_{12}$  are estimators of the variance of the errors and the covariance of the errors and their sign,  $\hat{f}(0)$  is a consistent estimator of the density evaluated at zero, and  $\eta$  is a standard normal random variable independent of  $W_1(1)$ . The second test statistic he proposes is

$$L_t = \left[ \hat{\sigma}_1^{-2} n^{-2} \sum_{t=1}^n X_{t-1}^2 \right]^{1/2} L_\beta \xrightarrow{w} \eta.$$

Both statistics are asymptotically distribution free and can be tabulated. Despite of this, they still present a great drawback: because of the form of the expression  $L_\beta$ , both tests have no power if the distribution of the errors is normal.

In this work we propose a different approach for testing. The goal is to avoid the estimation of the covariance matrix and specially the estimation of  $f(0)$ , using the bootstrap approximation. To do it we have proved in Section 2 that the asymptotic distribution of  $n(\hat{\beta}_{\text{LAD}}^* - 1)$  when the errors are independent and identically distributed is the same as that of  $n(\hat{\beta}_{\text{LAD}} - 1)$ . The bootstrap methodology provides a very powerful tool in this situation: once we have proved that the resampled version of the statistic has the same asymptotic distribution as the estimator, a large number of resamples can be used to obtain empirical critical values for the asymptotic distribution. This approach has a double advantage over the proposal of Herce (1996). The first one is that, as we have already mentioned, the estimation of the distribution parameters is avoided; the second one is that the new test has power where the other one failed.

To check the behaviour of the proposed test in finite samples, an empirical study was designed. We compare in it the behaviour of several tests proposed in the literature; three of them are based on the least-squares estimator, the Dickey and Fuller test (1979), and two bootstrap tests: the first one was proposed by Basawa et al. (1991b) and the second by Ferretti and Romo (1996). The other three are based on LAD estimators; one is the test based on  $L_\beta$  and the other two are our new proposals. The first one uses the results in Section 2 and is based on the statistic  $n(\hat{\beta}_{\text{LAD}}^* - 1)$  calculated as was proposed there. The second one is based on a different type of resampling that only centres the residuals in the mean (the behaviour of this test in the simulation indicates that probably the procedure is correct in this case also). In both cases the tests are going to be based on empirical bootstrap critical values.

Li and Maddala (1996) review bootstrap methods for time-series models presenting the different ways of resampling from residuals for testing. They suggest the possibility of resampling from the residuals obtained directly from the model under the null hypothesis. The bootstrap tests we propose in this paper have been conducted also in this fashion, the results are not included because their behaviour in the cases studied were remarkably worse than the ones based on the residuals obtained with LAD estimation.

The simulation has been conducted as follows: 2000 series were created of sizes 50, 100 and 250 for different values of  $\beta$ : 0.8, 0.9, 0.95, 0.99, 1.01, 1.05, 1.1 and 4000 series of the same sizes for  $\beta = 1$ . For all the bootstrap tests, 4000 resamples were taken. The critical values for Herce's test were calculated using 40 000 series,  $f(0)$  was estimated using a gaussian kernel with a bandwidth of  $h = n^{-1/5}$ . Five models were considered for the distribution of the errors:

1.  $N(0, 1)$ .
2. Double exponential.
3. Mixture of  $N(0, 1)$  and  $N(0, 25)$  with weights 0.6 and 0.4, respectively.
4. Student's  $t$  with 3 degrees of freedom.
5. Mixture of  $N(-4, 9.766)$  and  $U(-1, 9)$  (this distribution has mean and median equal to zero but is asymmetric).

Distributions 2, 3 and 4 are heavy tailed distributions and both Herce's and our tests should outperform Dickey-Fuller's. Distribution 5 is considered to study the behaviour of our test when the distribution is not symmetric. The results are summarized in Tables 1-5.

In the tables we have used the following notation:  $D_n$  is the Dickey-Fuller statistic,  $B_n^*$  is the test proposed by Basawa et al.,  $D_n^*$  is the bootstrap version proposed by Ferretti and Romo (1996).  $L_\beta$  and  $n(\hat{\beta}_{LAD}^* - 1)$  have already been defined in (9) and (3), respectively; finally, we have used  $M_n^*$  to note the test based on the residuals centred in the mean.

In the case of the standard normal distribution the behaviour of our tests is comparable with those of Dickey-Fuller, Basawa et al. and Ferretti-Romo and significantly better than Herce's. In models 2, 3 and 4 (symmetric distributions with heavy tails but finite variance) it can be observed that Herce's works better when the size of the sample is very small and when  $\beta$  is between 0.99 and 0.95; the important fact is that in all the remaining cases our tests outperform the rest. It is also remarkable how the empirical size of the tests is pretty close to the nominal size in almost all the cases.

In model 5 (asymmetric distribution) Herce's test presents the worst behaviour; this is the only case where the empirical level is clearly above the expected level. Despite of the asymmetry, our tests still work properly and the results, though a little worse, are comparable with those of the tests based on least-squares estimation. (Notice that in this case the distribution is not heavy tailed.)

We present also two figures that show graphically the power of the tests considered here compared to ours. Fig. 1 shows that even with a small sample size our method is comparable to those that use least squares when the innovations are normally



Table 1  
 Empirical power with model 1,  $\alpha = 5\%$

$n$	Test	$\beta$							
		0.80	0.90	0.95	0.99	1.00	1.01	1.05	1.1
50	$D_n$	61.30	20.60	9.20	5.35	7.15	12.45	69.50	96.55
	$B_n^*$	56.10	18.55	7.60	3.40	5.40	10.90	70.25	95.85
	$D_n^*$	66.30	23.35	10.90	6.20	7.65	12.75	69.85	96.60
	$L_\beta$	31.60	18.10	11.40	6.65	6.10	4.60	1.70	0.60
	$n(\hat{\beta}_{LAD}^* - 1)$	51.50	21.20	10.05	7.55	7.30	8.30	43.25	96.40
	$M_n^*$	48.75	17.90	8.35	5.45	5.17	5.45	31.25	95.45
100	$D_n$	99.20	55.25	18.35	3.75	5.95	25.35	96.95	100.00
	$B_n^*$	98.80	56.15	17.55	3.65	4.95	23.05	96.75	99.85
	$D_n^*$	99.30	58.50	20.05	4.15	5.85	25.20	97.00	100.00
	$L_\beta$	47.05	30.20	18.20	6.70	4.65	3.10	0.20	0.95
	$n(\hat{\beta}_{LAD}^* - 1)$	91.15	47.75	17.30	6.55	6.40	8.80	97.30	100.00
	$M_n^*$	92.40	47.75	19.25	7.05	4.35	7.10	96.55	100.00
250	$D_n$	100.00	100.00	76.15	7.60	5.15	70.80	100.00	100.00
	$B_n^*$	100.00	99.80	73.10	7.50	4.97	71.75	100.00	100.00
	$D_n^*$	100.00	100.00	76.75	7.65	5.15	70.55	100.00	100.00
	$L_\beta$	67.00	53.50	36.25	12.00	5.35	0.95	0.25	100.00
	$n(\hat{\beta}_{LAD}^* - 1)$	100.00	97.40	63.85	8.40	5.45	42.05	100.00	100.00
	$M_n^*$	100.00	97.70	61.75	7.20	4.40	33.15	100.00	100.00

Table 2  
 Empirical power with model 2,  $\alpha = 5\%$

$n$	Test	$\beta$							
		0.80	0.90	0.95	0.99	1.00	1.01	1.05	1.1
50	$D_n$	61.85	19.75	8.50	4.40	4.95	10.50	70.75	96.45
	$B_n^*$	54.80	16.70	6.95	4.05	4.82	11.15	72.85	96.10
	$D_n^*$	66.90	22.45	10.00	5.10	5.42	10.90	70.75	96.50
	$L_\beta$	53.15	28.85	12.60	5.15	4.17	3.95	1.95	2.15
	$n(\hat{\beta}_{LAD}^* - 1)$	79.35	28.00	9.95	4.95	4.97	7.15	71.50	97.65
	$M_n^*$	76.10	22.40	7.85	3.65	3.92	5.00	50.15	97.45
100	$D_n$	98.85	58.90	20.30	4.50	4.70	24.35	97.05	100.00
	$B_n^*$	98.85	55.70	18.25	3.70	4.95	22.95	97.00	99.95
	$D_n^*$	99.05	60.45	21.60	5.00	4.77	24.25	97.05	100.00
	$L_\beta$	80.65	57.80	28.10	5.85	3.47	3.45	1.75	42.80
	$n(\hat{\beta}_{LAD}^* - 1)$	99.65	83.15	33.70	5.40	4.87	13.10	98.65	100.00
	$M_n^*$	99.65	79.70	26.55	4.75	3.57	8.50	98.20	99.95
250	$D_n$	100.00	99.80	75.20	7.55	5.35	72.05	100.00	100.00
	$B_n^*$	100.00	99.90	73.50	6.70	4.77	74.00	100.00	100.00
	$D_n^*$	100.00	99.95	75.80	7.60	5.42	72.00	100.00	100.00
	$L_\beta$	98.70	91.35	69.50	10.00	3.75	2.35	83.10	100.00
	$n(\hat{\beta}_{LAD}^* - 1)$	100.00	100.00	96.10	9.40	5.12	78.60	100.00	100.00
	$M_n^*$	100.00	100.00	95.35	9.55	4.35	61.10	100.00	100.00

Table 3  
Empirical power with model 3,  $\alpha = 5\%$

$n$	Test	$\beta$							
		0.80	0.90	0.95	0.99	1.00	1.01	1.05	1.1
50	$D_n$	62.95	18.70	8.20	4.10	5.60	10.30	69.60	96.05
	$B_n^*$	54.95	15.20	7.10	3.75	4.85	10.85	70.90	96.90
	$D_n^*$	66.95	21.35	9.15	4.55	5.92	10.25	69.75	96.15
	$L_\beta$	93.75	72.65	39.95	8.75	6.57	5.20	11.40	72.00
	$n(\hat{\beta}_{LAD}^* - 1)$	96.95	54.40	14.40	5.20	5.82	8.70	83.75	98.75
	$M_n^*$	95.10	45.10	11.45	3.60	3.92	4.95	59.40	98.05
100	$D_n$	99.10	57.55	17.85	3.50	4.57	22.80	97.10	100.00
	$B_n^*$	98.50	53.60	17.30	3.60	4.85	22.50	97.85	100.00
	$D_n^*$	99.25	59.55	19.20	3.50	4.65	22.80	97.15	100.00
	$L_\beta$	100.00	95.60	74.40	12.75	5.40	3.90	82.85	99.65
	$n(\hat{\beta}_{LAD}^* - 1)$	100.00	98.55	58.90	5.45	5.10	20.80	99.15	100.00
	$M_n^*$	100.00	97.50	48.15	4.40	3.05	8.95	98.85	100.00
250	$D_n$	100.00	99.95	75.90	7.70	5.85	71.40	100.00	100.00
	$B_n^*$	100.00	99.80	73.85	6.70	4.52	70.60	100.00	100.00
	$D_n^*$	100.00	99.95	76.20	7.70	5.90	71.40	100.00	100.00
	$L_\beta$	100.00	99.90	98.60	34.30	5.92	9.45	100.00	100.00
	$n(\hat{\beta}_{LAD}^* - 1)$	100.00	100.00	99.85	18.30	6.32	86.80	100.00	100.00
	$M_n^*$	100.00	100.00	99.90	14.25	4.52	67.10	100.00	100.00

Table 4  
Empirical power with model 4,  $\alpha = 5\%$

$n$	Test	$\beta$							
		0.80	0.90	0.95	0.99	1.00	1.01	1.05	1.1
50	$D_n$	60.30	19.00	7.35	3.15	4.65	10.50	71.65	97.05
	$B_n^*$	56.55	16.00	6.70	3.25	5.02	10.00	71.70	96.35
	$D_n^*$	65.30	21.90	8.95	3.55	5.15	11.15	71.75	97.10
	$L_\beta$	59.00	36.80	19.85	6.05	5.30	3.90	3.40	11.35
	$n(\hat{\beta}_{LAD}^* - 1)$	76.25	29.80	10.45	4.85	5.10	6.65	64.45	98.15
	$M_n^*$	76.10	28.95	10.05	3.95	3.97	5.25	41.60	97.45
100	$D_n$	99.20	60.10	18.60	4.75	5.07	21.85	96.70	100.00
	$B_n^*$	98.55	53.70	15.35	3.45	4.10	22.55	97.60	100.00
	$D_n^*$	99.35	62.30	19.80	5.10	5.30	22.25	96.70	100.00
	$L_\beta$	79.90	60.85	36.90	8.05	5.50	4.10	10.50	51.20
	$n(\hat{\beta}_{LAD}^* - 1)$	99.50	78.80	31.15	5.60	5.17	10.70	98.50	100.00
	$M_n^*$	99.55	77.80	29.40	4.80	4.30	6.70	97.70	99.95
250	$D_n$	100.00	99.80	75.65	6.90	4.62	72.85	100.00	100.00
	$B_n^*$	100.00	99.80	74.65	6.35	5.15	70.95	100.00	100.00
	$D_n^*$	100.00	99.80	75.75	6.95	4.75	73.05	100.00	100.00
	$L_\beta$	97.60	90.70	73.15	16.00	4.55	3.85	81.25	100.00
	$n(\hat{\beta}_{LAD}^* - 1)$	100.00	99.95	91.95	10.95	5.10	72.80	100.00	100.00
	$M_n^*$	100.00	100.00	92.05	8.50	4.60	50.50	100.00	100.00

Table 5  
Empirical power with model 5,  $\alpha = 5\%$

$n$	Test	$\beta$							
		0.80	0.90	0.95	0.99	1.00	1.01	1.05	1.1
50	$D_n$	60.55	20.30	8.20	4.50	5.45	10.85	70.65	96.45
	$B_n^*$	58.00	19.10	8.30	3.85	6.02	10.45	70.45	96.95
	$D_n^*$	65.65	23.75	10.20	5.10	6.22	10.90	70.55	96.45
	$L_\beta$	54.55	35.25	23.50	15.70	12.95	11.30	6.05	8.60
	$n(\hat{\beta}_{LAD}^* - 1)$	39.80	16.35	9.50	6.30	7.07	6.60	32.75	95.90
	$M_n^*$	41.05	16.25	8.85	5.30	4.82	5.50	23.30	95.20
100	$D_n$	99.30	58.50	18.30	4.90	5.97	23.20	97.25	100.00
	$B_n^*$	98.75	55.70	17.45	4.40	5.77	23.05	96.10	100.00
	$D_n^*$	99.45	61.05	19.65	5.40	6.30	23.10	97.15	100.00
	$L_\beta$	67.35	48.65	31.00	13.85	11.45	7.15	3.40	24.10
	$n(\hat{\beta}_{LAD}^* - 1)$	85.15	39.60	14.85	6.90	6.67	7.40	96.20	100.00
	$M_n^*$	85.70	36.45	13.80	5.55	4.72	5.35	95.80	99.90
250	$D_n$	100.00	99.95	74.70	7.25	5.45	72.85	100.00	100.00
	$B_n^*$	100.00	99.90	75.15	8.40	4.97	73.20	100.00	100.00
	$D_n^*$	100.00	99.95	75.25	7.45	5.52	72.70	100.00	100.00
	$L_\beta$	78.55	66.20	48.05	16.05	7.20	2.15	21.50	100.00
	$n(\hat{\beta}_{LAD}^* - 1)$	100.00	94.75	54.85	6.15	5.27	41.80	100.00	100.00
	$M_n^*$	100.00	94.05	53.05	6.55	3.95	29.95	100.00	100.00

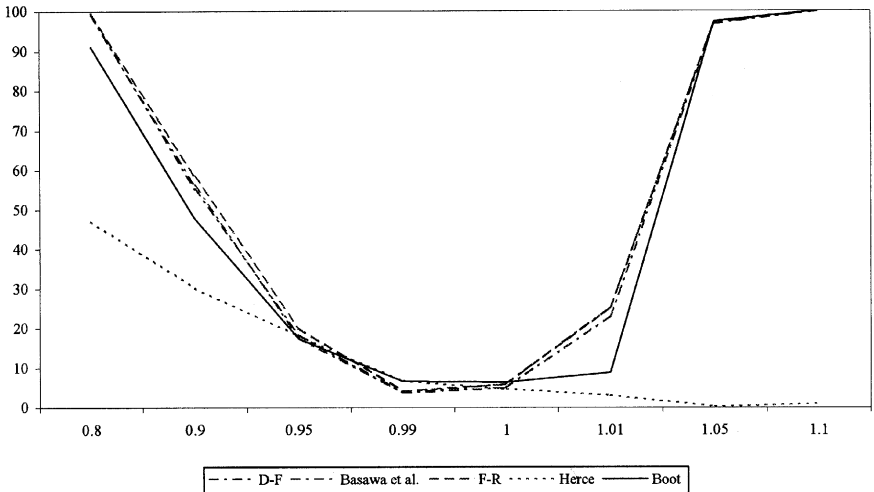


Fig. 1. Empirical power with model 1,  $n = 100$ .

distributed. It can be appreciated also that Herce's test has considerably small power in this case.

In Fig. 2 our test presents better behaviour than the others when the distribution is heavy tailed, in this case we have used the double exponential with  $n = 100$ .

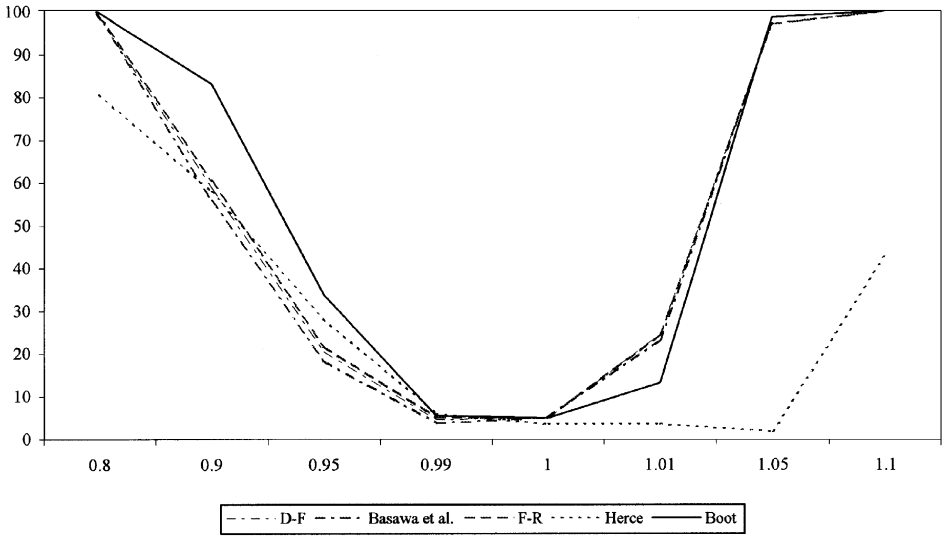


Fig. 2. Empirical power with model 2,  $n = 100$ .

#### 4. Conclusions

In this paper we propose a new test for detecting unit roots in a AR(1) process with independent innovations. This new test exploits the power of the resampling procedures in the sense that it is automatic, natural and easy to compute. Its main advantage is that it avoids the problem of estimation of the variance matrix and the density at zero and at the same time the problem of the artificial linear combination with the least-squares estimator needed to obtain a distribution free statistic. On the other hand, the test we propose presents a good behaviour when the distribution is heavy tailed, it has the advantages of a robust test and it works properly when the distribution is normal. It also improves Herce's proposal in the sense that it has power when  $\beta > 1$ . Another interesting question that is being currently studied is what happens when the variance is infinite and we use bootstrap to obtain the critical values for the test.

#### References

- Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991a. Bootstrapping unstable first-order autoregressive processes. *Ann. Statist.* 19, 1098–1101.
- Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991b. Bootstrap test of significance and sequential bootstrap estimation for unstable first-order autoregressive processes. *Comm. Statist. — Theory Methods* 20, 1015–1026.
- Basset, G.W., Koenker, R.W., 1978. The asymptotic distribution of least absolute error estimator. *J. Am. Statist. Assoc.* 73, 618–622.
- Bloomfield, P., Steiger, W.L., 1983. *Least Absolute Deviations*. Birkhäuser, Boston.
- Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* 74, 427–431.

- Dudley, R.M., 1989. *Real Analysis and Probability*, Mathematic Series. Chapman & Hall, New York.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Ferretti, N., Romo, J., 1996. Unit root bootstrap tests for AR(1) models. *Biometrika* 83, 849–860.
- González Manteiga, W., Prada Sánchez, J.M., Romo, J., 1994. The bootstrap — a review. *Comp. Statist.* 9, 165–205.
- Herce, M.A., 1996. Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econom. Theory* 12, 129–153.
- Karatzas, I., Shreve, S.E., 1991. *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics. Springer, New York.
- Knight, K., 1989. Limit theory for autoregressive parameter estimates in an infinite variance random walk. *Can. J. Statist.* 17, 261–278.
- Knight, K., 1991. Limit theory for  $M$ -estimates in an integrated infinite variance process. *Econom. Theory* 7, 200–212.
- Koul, H.L., 1991. A weak convergence result useful in robust autoregression. *J. Statist. Plann. Inference* 29, 291–308.
- Li, H., Maddala, G.S., 1996. Bootstrapping time series models. *Econom. Rev.* 15, 115–158.
- Phillips, P.C.B., 1991. A shortcut to LAD estimator asymptotics. *Econom. Theory* 7, 450–463.
- Pollard, D., 1984. *Convergence of Stochastic Processes*. Springer, New York.
- Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* 7, 186–199.
- White, H., 1984. *Asymptotic Theory for Econometricians*. Academic Press, London.