

# RANGE UNIT ROOT (RUR) TESTS: ROBUST AGAINST NONLINEARITIES, ERROR DISTRIBUTIONS, STRUCTURAL BREAKS AND OUTLIERS



BY FELIPE APARICIO, ALVARO ESCRIBANO AND ANA E. SIPOLS

*Universidad Carlos III de Madrid and Universidad Rey Juan Carlos*

*First Version received July 2004*

**Abstract.** Since the seminal paper by Dickey and Fuller in 1979, unit-root tests have conditioned the standard approaches to analysing time series with strong serial dependence in mean behaviour, the focus being placed on the detection of eventual unit roots in an autoregressive model fitted to the series. In this paper, we propose a completely different method to test for the type of long-wave patterns observed not only in unit-root time series but also in series following more complex data-generating mechanisms. To this end, our testing device analyses the unit-root persistence exhibited by the data while imposing very few constraints on the generating mechanism. We call our device the *range unit-root* (RUR) test since it is constructed from the running ranges of the series from which we derive its limit distribution. These nonparametric statistics endow the test with a number of desirable properties, the invariance to monotonic transformations of the series and the robustness to the presence of important parameter shifts. Moreover, the RUR test outperforms the power of standard unit-root tests on near-unit-root stationary time series; it is invariant with respect to the innovations distribution and asymptotically immune to noise. An extension of the RUR test, called the forward backward range unit-root (FB-RUR) improves the check in the presence of additive outliers. Finally, we illustrate the performances of both range tests and their discrepancies with the Dickey Fuller unit-root test on exchange rate series.

**Keywords.** Unit-roots tests; structural breaks; nonlinearities; additive outliers; running ranges; exchange rates.

## 1. INTRODUCTION

Many overwhelming low frequency non periodic components in time series are associated with the presence of unit roots in their data generating processes (DGP). Such time series are said to be ‘integrated’. The pioneering work of Nelson and Plosser (1982) led to the belief that many economic time series were best described in this way. This prompted a large amount of research on unit root time series, covering both theoretical and empirical aspects. The unit root paradigm has important practical implications since it entails that shocks have a permanent effect on a variable, or equivalently that the fluctuations they cause are not transitory.

The existence of unit roots in time series is investigated by means of unit root tests. The application of standard unit root tests, such as the Dickey Fuller (DF)

test (Dickey and Fuller, 1979), has been an important step in the construction of a useful parametric model for many economic time series.

Unit root time series models impose, however, severe restrictions on the DGPs of the data. Many real world time series exhibit nonlinearities, outliers and structural breaks either in the mean or in the variance. All these features, which cannot be properly captured with random walk like models, affect the size/power to standard unit root tests.

Many economic and financial time series such as inflation, unemployment rate, and nominal and real interest rates can be trend stationary with a structural break in the unconditional mean which affects the standard inferential procedures and often makes constant coefficient models to perform poorly in practice (see for instance, Perron, 1990; Malliaropoulos, 2000). The literature on testing for unit roots in the presence of both known and unknown break points is large (see Maddala and Kim, 1998 for a review). Perron (1989), Volgelnsang (1990) and Perron and Volgelnsang (1992) reported evidence that structural breaks can make an  $I(0)$  time series behave locally as  $I(1)$  and, as a result, these breaks are able to affect the size/power of standard unit root tests. The appropriate handling of such departures as parameter shifts, trend breaks and nonlinearities calls for the development of robust unit root tests.

In practice, it is difficult and even sometimes impossible to know whether a time series exhibiting unit root like behaviour is really  $I(1)$ , or rather a monotonic nonlinear transformation of an  $I(1)$  series. With standard unit root tests, misspecification of the true time series model may affect the rate of divergence of the test statistic, making it behave inconsistently. The invariance to such nonlinearities, would be therefore, a desirable property of a unit root test.

Granger and Hallman (1991) looked at the autocorrelation function of several nonlinear transformations of the original series and proposed a test invariant to monotonic transformations based on ranks.

Ermini and Granger (1993) worked with the Hermite polynomial expansion of different nonlinear transformations of random walks, possibly with drift, and showed that the autocorrelation function is not always a reliable indicator of the degree of memory of nonlinear time series.

Outlying observations is another source of problem for time series analysis. These may occur for different reasons, ranging from measurement errors to recordings of unusual events such as wars, disasters and dramatic policy changes. Some commonplace outlier inducing events in economic time series are union strikes, hoarding consumer behaviour in response to a policy announcement and computer breakdown effects on unemployment or sales data collection and processing, to name a few. Outliers can also appear as a result of misspecified estimated relationships or omitted variables (see, for instance Peña, 2001).

There is a sort of duality between the effects of additive outliers (AOs) and those of structural breaks on time series. Indeed, while  $I(0)$  time series, subject to level shifts, could be misinterpreted as  $I(1)$ ,  $I(1)$  time series corrupted by AOs might look like  $I(0)$  provided that the outliers are sufficiently frequent and

important in magnitude. In particular, it is known that the presence of AOs leads to a downward bias of the ordinary least square (OLS) parameter estimates in a stationary AR(1) process (Bustos and Yohai, 1986; Martin and Yohai, 1986) and thereby the DF test will have an actual size in excess of the nominal size, thus rejecting the unit root hypothesis too often. The size distortion of the DF test in the presence of AOs was quantified by Franses and Haldrup (1994).

In this paper, we introduce a nonparametric range unit root (RUR) test, which has a remarkable superiority with respect to the standard approaches (see Aparicio *et al.*, 2004a,b). First, it is invariant to monotonic transformations and to the distribution of the model errors. Secondly, it is robust against many structural breaks, parameter shifts and certain additive outliers. Thirdly, it does not depend on the variance of any stationary alternative and thereby outperforms standard tests also in terms of power on near unit root stationary time series. Finally, a modified RUR test (FB RUR) is not affected by the presence of additive noise on the series.

The structure of the paper is as follows. In Section 2 we introduce the RUR test, discuss its small sample behaviour under the null hypothesis of a single unit root and derive the asymptotic null distribution of the test. Section 3 studies its power performances and its consistency against stationary, integrated and trending alternatives. Section 4 analyses the robustness of the test statistic under different departures from standard unit root test assumptions. Section 5 presents a modification of the former RUR test that improves both its small sample power in the presence of level shifts, and its size when additive outliers corrupt the series early. In Section 6 we analyse the size distortion with serial correlation and heteroskedastic error. In Section 7 we apply our testing methodology to a real time series and compare the results with those obtained by means of standard unit root tests. After the concluding remarks of Section 8, an Appendix is devoted to the proofs of the main theoretical results.

## 2. RANGE UNIT ROOT TEST

Many time series not generated by unit root models exhibit mean behaviour similar to those which are. The objective of this section is to investigate alternative procedures for assessing the presence of unit root like features, not necessarily caused by unit roots. We will begin by studying the behaviour of the sequence of running ranges in both stationary and random walk time series.

The range of a data sample is defined in terms of its extremes. Formally, for a given time series  $x_t$ , the statistics  $x_{1,i} = \min\{x_1, \dots, x_i\}$  and  $x_{i,i} = \max\{x_1, \dots, x_i\}$  are called the  $i$ th extremes. When the sample comes from a time series  $x_t$ , a monotonically increasing sequence of ranges can be obtained as  $R_i^{(x)} = x_{i,i} - x_{1,i}$ , for  $i = 1, 2, 3, \dots, n$ , where  $n$  denotes the sample size. The total number of ‘new extremes’ or records in a sample of size  $n$  is given by the quantity  $\sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0)$ , where  $\mathbf{1}(\cdot)$  is the indicator function.

It can be shown that the long run frequency of new records,  $n^{-1} \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0)$ , vanishes faster for stationary time series than for series containing a unit root; these latter series are often referred to as integrated ‘of order 1’, or briefly as  $I(1)$ . In particular, for independent and identically distributed (i.i.d.) sequences of random variables we have (see, for instance Embrechts *et al.*, 1999):

$$\frac{1}{\log n} \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0) \quad O(1)$$

This result still holds for stationary series satisfying the ‘Berman condition’, which requires the covariance sequence of the series  $\{c_i = \text{cov}(x_t, x_{t+i})\}_{i \geq 1}$  to decrease faster than  $(\log i)^{-1}$ , that is  $c_i \log i \rightarrow 0$  as  $i \rightarrow \infty$  (see Lindgren and Rootzén, 1987).<sup>1</sup> As will be shown later in this paper, the frequency of new records for  $I(1)$  time series decreases at a slower rate:

$$\frac{1}{n^{1/2}} \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0) \quad O(1)$$

However, if  $x_t$  is a random walk with drift, then the frequency of new records decreases at an even slow speed, in this case:

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0) \quad O(1)$$

We remark that for the random walk, the sequence of running ranges escalates indefinitely, whereas it does not in the stationary cases (see Aparicio *et al.*, 2005). However, having thick tailed error distributions or mere infinite variance does not imply the divergence of the running ranges. Such a divergence is caused by strong first order serial or stationary frequency dependence.

In what follows, we introduce the RUR test statistic upon which the proposed unit root testing methodology is based. Then we provide some asymptotic results, and analyse its small sample behaviour under the null hypothesis of a single unit root. Finally, we study its small sample power performances against AR(1) stationary alternatives.

### 2.1. The test statistic

In the sequel we consider the statistic  $J_0^{(n)}$  defined below for testing the null hypothesis of a random walk  $x_t = x_{t-1} + \epsilon_t$  where the errors  $\{\epsilon_t\}_{t \geq 1}$  are a sequence of i.i.d. random variables having zero mean and variance  $\sigma_\epsilon^2$ . The corresponding testing device will be referred to as the RUR test.

$$\text{RUR} \equiv J_0^{(n)} = n^{-1/2} \sum_{t=1}^n \mathbf{1}(\Delta R_t^{(x)} > 0). \quad (1)$$

Notice that  $n^{-1/2}J_0^{(n)}$  represents the proportion of these prediction errors in a sample of size  $n$ , while  $n^{1/2}J_0^{(n)}$  is the number of new records of the time series  $x_t$  up to time  $n$ .

Given the non ergodic nature of  $x_t$  under the null hypothesis, the normalized number of records in the sample,  $J_0^{(n)}$ , does not converge to zero but to a non degenerate random variable, as will be shown later. On the contrary, when  $x_t \sim I(0)$ ,  $J_0^{(n)}$  converges in probability to zero. Therefore, we can consider the left tail of the distribution of  $J_0^{(n)}$  to discriminate between  $I(1)$  and  $I(0)$  series. This means that when  $x_t \sim I(0)$ ,  $R_{t-1}^{(x)}$  is a more efficient predictor of  $R_t^{(x)}$  than when  $x_t$  contains a unit root. Consequently, the RUR test statistic  $J_0^{(n)}$  will be expected to take comparatively large values for  $I(1)$  time series and small values for  $I(0)$  time series. We also show that the RUR test is robust to a number of departures from the null hypothesis (no size distortions).

## 2.2. Small sample behaviour under the null

Table I shows estimates of the critical values of  $J_0^{(n)}$  obtained from 10,000 replications of the null model, and for eight different sample sizes and six significance levels ( $\alpha = 0.01, 0.025, 0.05, 0.10, 0.90, 0.95$ ) where the model errors follow  $\epsilon_t \sim \text{n.i.d.}(0,1)$ .

Figure 1 shows the corresponding empirical density of  $J_0^{(n)}$  estimated by kernel smoothing, using the Epanechnikov kernel.

## 2.3. Asymptotic distribution

A basic result regarding the behaviour of the records of a random walk is that the sample size increasing the frequency of these records is equal to zero. Proposition 1 formally establishes this result, which is proved in the Appendix and will be used subsequently.

**PROPOSITION 1.** *Let  $x_t = x_{t-1} + \epsilon_t$  where  $\{\epsilon_t\}_{t \geq 1}$  satisfies the mixing condition of Phillips and Perron (1988) and let  $x_{1,t} = \max\{x_1, \dots, x_t\}$  and  $x_{1,t} = \min\{x_1, \dots, x_t\}$ . Then we have*

TABLE I  
CRITICAL VALUES OF THE UNIT-ROOT TEST

$\alpha$	$n$							
	100	250	500	1000	2000	3000	4000	5000
0.01	0.9	0.9391	1.0119	1.0435	1.1180	1.1137	1.1420	1.1455
0.025	1.0	1.0752	1.1180	1.1700	1.2075	1.2232	1.2301	1.2304
0.05	1.1	1.2017	1.2075	1.2649	1.2746	1.3145	1.3123	1.3152
0.10	1.3	1.3282	1.3864	1.4230	1.4530	1.4534	1.4606	1.4506
0.90	2.8	2.9725	3.04	3.06	3.08	3.1038	3.108	3.11
0.95	3.1	3.2888	3.3541	3.3520	3.4435	3.4324	3.44	3.47

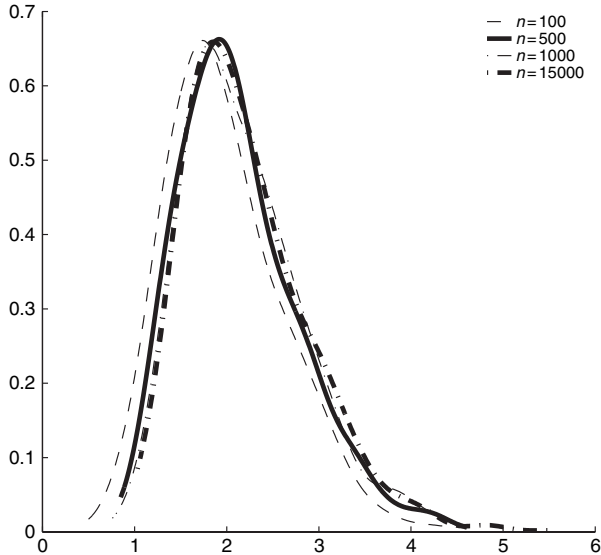


FIGURE 1. Plot of the empirical density of  $J_0^{(n)}$ , for different sample sizes, under the null hypothesis of a random walk with n.i.d.(0,1) errors.

$$\lim_{t \rightarrow \infty} P(x_t = x_{1,t}) = \lim_{t \rightarrow \infty} P(x_t = x_{1,t}) = 0. \quad (2)$$

PROOF. In Appendix A1. □

Appropriate scaling is needed for the sequence of partial sums  $\sum_{t=1}^n \mathbf{1}(\Delta R_t^{(x)} > 0)$  to converge to a non degenerate random variable under the null hypothesis  $H_0$ . Our main result of Theorem 1 establishes that under  $H_0$  the normalized sequence of partial sums

$$J_0^{(n)} = n^{-1/2} \sum_{t=1}^n \mathbf{1}(\Delta R_t^{(x)} > 0)$$

converges weakly to a random variable. Under the alternative hypothesis of stationarity, and under mild conditions on the degree of serial dependence of  $x_t$ , the sequence of partial sums  $\sum_{t=1}^n \mathbf{1}(\Delta R_t^{(x)} > 0)$  diverges at a much lower rate, thus leading to  $J_0^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ . From Revuz and Yor (1991) we have the following definition of local time processes.

DEFINITION 1 (LOCAL TIME OF A BROWNIAN MOTION PROCESS) (LÉVY, 1948). Let  $B(\cdot)$  represent a Brownian motion process in  $\mathfrak{R}$ , and let  $l_B(x, t)$  be defined as

$$l_B(x, t) = \lim_{\delta \downarrow 0} \frac{1}{2\delta} \int_0^t \mathbf{1}[x - \delta \leq B(s) \leq x + \delta] ds \quad (3)$$

where  $l_B(x, t)$  is a continuously increasing process in  $x$  called the local time of  $B$  at  $x$ . It measures the amount of time the Brownian motion spends in the neighbourhood of  $x$ . It can also be interpreted as the 'spatial density' of the occupation time  $\int_0^t \mathbf{1}[x - \delta \leq B(s) \leq x + \delta] ds$ .

THEOREM 1. Let  $x_t = \sum_{i=1}^t \epsilon_i$  where  $\{\epsilon_i\}_{i \geq 1}$  are continuous i.i.d. random variables with bounded probability density, function (pdf), zero mean and finite variance  $\sigma_\epsilon^2$ . Suppose that  $x_0$  also has a bounded pdf and finite variance. Let

$$J_0^{(n)} = J_1^{(n)} + J_2^{(n)}$$

with

$$J_1^{(n)} = n^{-1/2} \sum_{t=1}^n \mathbf{1}(x_t - x_{t,t}) \quad \text{and} \quad J_2^{(n)} = n^{-1/2} \sum_{t=1}^n \mathbf{1}(x_t - x_{1,t}).$$

Then we have

(1)

$$J_1^{(n)} \Rightarrow \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} l_B(0, 1) \quad (4)$$

$$J_2^{(n)} \Rightarrow \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} l_B(0, 1) \quad (5)$$

$$P\{J_i^{(\infty)} \leq h\} = \frac{2}{\sqrt{2\pi} \left(\frac{E\{|\epsilon_1|\}}{\sigma_\epsilon}\right)^2} \int_0^h \exp\left(-\frac{u^2}{2\left(\frac{E\{|\epsilon_1|\}}{\sigma_\epsilon}\right)^2}\right) du, \quad (6)$$

$$h \geq 0, \quad i = 1, 2 \text{ which depends on } a = \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} \quad (7)$$

(2) The limiting distribution of the RUR test statistic is given by,

$$RUR \equiv J_0^{(n)} \Rightarrow \left[ \frac{w^2}{a^2} \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2} \frac{w^2}{a^2}} \right] \\ \sqrt{\frac{2}{\pi}} (\xi + \eta)^2 e^{-\frac{1}{2}(\xi + \eta)^2}$$

independent of  $a = \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon}$  where  $\xi \rightarrow |B(1)|$  and  $\eta \rightarrow l_B(0, 1)$

- (3) If  $x_t$  is a stationary Gaussian series with covariance sequence  $\{c_i \text{cov}(x_t, x_{t+i})\}_i$  satisfying  $c_i \log i \rightarrow 0$  as  $n \rightarrow \infty$  (Berman condition). Then we have

$$J_0^{(n)} \xrightarrow{P} 0, \quad (8)$$

and thus the RUR test is consistent against this stationary alternative.

PROOF. In Appendix A2. □

In Figure 2 we plot the asymptotic distribution and the empirical distribution for a sample size of 1000. It is clear that the critical values from both tails of the two distributions are similar as shown in Table I.

To ensure the consistency of the RUR test statistic against general stationary alternatives, we impose certain restrictions on the serial dependence of the process. The following condition is similar in spirit (although weaker) to the strong mixing condition and allows us to use the results from the asymptotic theory of records for i.i.d. processes.

CONDITION 1.  $D(u_n)$ : Let  $\{x_t\}_{t \geq 1}$  be a stationary sequence of random variables with

$$F_{i_1, \dots, i_n}(u_1, \dots, u_n) = P\{x_{i_1} \leq u_1, \dots, x_{i_n} \leq u_n\}$$

representing its finite dimensional distribution function. Write  $F_{i_1, \dots, i_n}(u)$   $F_{i_1, \dots, i_n}(u, \dots, u)$  for economy of notation and define

$$\alpha_{n,l} = \max\{|F_{i_1, \dots, i_p, j_1, \dots, j_q}(u) - F_{i_1, \dots, i_p}(u)F_{j_1, \dots, j_q}(u)|\}$$

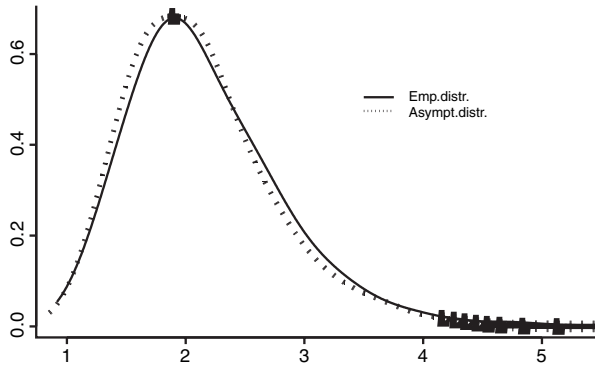


FIGURE 2. Plot of the empirical density of  $J_0^{(n)}$  for  $n = 1000$  under the null hypothesis  $H_0 : x_t = x_{t-1} + \epsilon_t$ , where  $\epsilon_t \sim \text{n.i.d.}(0,1)$  and the asymptotic distribution.



with  $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$ ,  $j_1 - i_p \geq l$ . The sequence  $\{x_t\}_{t \geq 1}$  is said to satisfy condition  $D(u_n)$  if there exists a sequence of numbers  $l_n = o(n)$  such that  $\alpha_{n,l_n} \rightarrow 0$ .

Among the processes that satisfy condition  $D(u_n)$  are the Gaussian processes satisfying the Berman condition. If  $x_t$  is one such process, then the joint distribution of any fixed set of extreme statistics converges to the same limit as if the variables were i.i.d. (Lindgren and Rootzén, 1987). As a consequence, we must expect  $n^{1/2}J_0^{(n)} \sim O(\log n)$ , or equivalently,  $J_0^{(n)} \sim O(n^{-1/2} \log(n)) \rightarrow 0$  as  $n \rightarrow \infty$ , and the consistency against this class of alternatives is proved. Note that the Berman condition is not very demanding, since it is satisfied by any process with exponentially decaying covariance function, among which are all the stationary Gaussian autoregressive moving average (ARMA) processes. The Gaussian condition seems to be too restrictive; however, as we see in Section 3, the size and power of the RUR test statistic do not vary significantly under stationary alternatives and under different error distributions (such as Cauchy's and the Student's  $t$ ).

### 3. SIZE, POWER AND CONSISTENCY OF RUR TEST

In this section we investigate the power performances of the RUR test and its consistency against stationary, trending and integrated alternatives. First of all, it is easy to show that the test is consistent against stationary alternatives [ $H_0 : I(1)$  against  $H_1 : I(0)$ ]. To show it, recall from Section 2 that for such alternatives we can expect the sequence of ranges to behave similarly as if  $x_t$  was an i.i.d. sequence, that is:

$$n^{-1/2} \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0) \sim O(n^{-1/2} \log n)$$

$$J_0^{(n)} \sim O(n^{-1/2} \log n)$$

Therefore, the test is consistent since  $n^{-1/2} \log n \rightarrow 0$  as  $n \rightarrow \infty$ , while  $P(J_0^{(\infty)} = 0 | H_0) = 0$ . A similar behaviour applies to  $I(k)$  time series with  $k > 0$  since the degree of mean reversion is even more pronounced in this case. The following simple device allows us to discriminate between the stationary and the integrated case. Let  $B$  denote the lag operator and let  $\tilde{x}_t^{(0)} \Delta x_t$ . Note that if  $x_t \sim I(0)$  then the time series defined by

$$\tilde{x}_t^{(1)} \Delta \sum_{j=0}^{\infty} B^j x_{t-j}$$

is  $I(1)$ . Similarly, if  $x_t \sim I(k)$  with  $k > 0$  then  $k + 1$  will be the smallest positive integer such that  $\tilde{x}_t^{(k+1)} \sim I(1)$ , or equivalently, such that  $J_0^{(n)}$  does not vanish

asymptotically. By mere inversion of the argument, if  $k$  is the smallest non negative integer such that the null hypothesis is not rejected on  $\tilde{x}_t^{(k)}$  then  $x_t$  will very likely be  $I(k)$ .

The small sample power of the test against stationary AR(1) alternatives is shown in Table II below using the estimated critical values at the 5% significance level, and from 10,000 replications of the alternative model  $x_t = bx_{t-1} + \epsilon_t$ , with  $\epsilon_t \sim \text{n.i.d.}(0,1)$ , the DF test shown in parentheses.

These results show that the DF test outperforms the RUR test in only two cases: (i) when the sample size is comparatively small ( $n = 100$ ), and (ii) when the autoregression parameter  $b$  is far from the nonstationary values of ( $b$ ). For near unit root stationary time series, however, the RUR test outperforms the DF test. Therefore, when compared with the DF test, the RUR test establishes a sharper distinction between the null hypothesis of unit root and the stationary AR(1) alternatives. This can be explained by the invariance of the RUR test statistic  $J_0^{(n)}$  with respect to the finite constant variance  $\sigma_x^2$  of the stationary alternative, which follows from the fact that

$$\mathbf{1}(\Delta R_t^{(x)} > 0) = \mathbf{1}(\sigma_x^{-1} \Delta R_t^{(x)} > 0).$$

On trending alternatives, the RUR test is also consistent if we use the right tail of the distribution of  $J_0^{(n)}$  under  $H_0$ . To see this, we invoke a classical result by Feller (1971) which states that on random walks with nonzero drift, that is when  $\mu_\epsilon = E(\epsilon_t) / \sigma_\epsilon > 0$ , the renewal counting process of records  $N(n) = \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0)$  satisfies:

$$\lim_{n \rightarrow \infty} n^{-1} N(n) = O(1).$$

As a consequence,  $J_0^{(n)} = O(n^{1/2}) \rightarrow \infty$  as  $n \rightarrow \infty$  under such alternatives.

A similar divergent behaviour of the RUR test statistic occurs when  $x_t \sim I(k)$  with integration order  $k > 1$ , or when  $x_t$  is a stationary time series fluctuating around a deterministic trend. To distinguish between these two cases, consider the following time series models:

- (a)  $x_t = x_{t-1} + \epsilon_t$  with  $E(\epsilon_t) = \mu_\epsilon > 0$ .
- (b)  $x_t = y_t + \mu t$  where  $y_t \sim I(0)$ .

TABLE II  
EMPIRICAL SIZE AND POWER OF RUR TEST FROM 10,000 REPLICATIONS FOR DIFFERENT SAMPLE SIZES  $n$  AND FOR DIFFERENT VALUES OF  $b$

$n$	$b$					
	0.5	0.8	0.9	0.95	0.99	1
100	0.8 (1)	0.6 (0.99)	0.5 (0.5)	0.4 (0.18)	0.12 (0.0375)	0.051 (0.04)
250	1 (1)	1 (1)	1 (1)	0.8 (0.7)	0.47 (0.0760)	0.049 (0.05)
500	1 (1)	1 (1)	1 (1)	1 (0.99)	0.72 (0.39)	0.05 (0.05)

Notice that under model (a)  $\Delta x_t \sim I(0)$ , while under model (b)  $\Delta x_t \sim I(-1)$ . So discrimination between models (a) and (b) is reduced to determine the order of integration, as discussed before.

#### 4. DEPARTURES FROM THE STANDARD CONDITIONS UNDER $H_0$

Another important property of the RUR test is its robustness to departures from the standard assumptions. In this paper, we consider three types of departures: (a) stationary alternatives with different error distributions, (b) when a stationary time series undergoes structural breaks; (c) when  $I(1)$  time series are corrupted by additive outliers and  $I(1)$  time series are nonlinear transformed. In the sequel we study the small sample behaviour of the RUR test in the presence of each of the above mentioned departures from the standard unit root test assumptions.

##### 4.1. Robustness of RUR test statistic against alternative error distributions

Consider  $J_0^{(n)}$  in the following form:

$$J_0^{(n)} = n^{-1/2} \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0) \\ + n^{-1/2} \sum_{i=1}^n \mathbf{1}(x_i - x_{i,i} > 0) + n^{-1/2} \sum_{i=1}^n \mathbf{1}(x_i - x_{1,i} > 0)$$

and then realizing that each term in this sum is the normalized number of visits to the origin of the two  $I(1)$  processes with asymptotically i.i.d. innovations:

$$y_t = x_t - x_{t,t} \quad (9)$$

$$y'_t = x_t - x_{1,t}. \quad (10)$$

The  $I(1)$  nature of  $y_t$  and  $y'_t$  allows the application of a result by Burridge and Guerre (1996) to the asymptotic distribution of the normalized number of level crossings of a random walk, which leads straightforwardly to ours. The asymptotic distribution of  $J_i^{(n)}$  depends on the innovation distribution (in particular, of their variance,  $\sigma_\epsilon^2$ ). This dependence comes from the scaling factor  $a = E\{|\epsilon|\}/\sigma_\epsilon$  which varies from one error distribution to another. For example, if the innovations  $\epsilon_t$  are Gaussian then  $a = \sqrt{2/\pi}$ , and thereby even the asymptotic distribution of the normalized number of upper (or lower) records,  $J_1^{(n)}(J_2^{(n)})$ , is unaffected by error variance,  $\sigma_\epsilon^2$ . However, this case is rather exceptional since for all other common distributions the value of  $a$  is sensitive to its shape, or equivalently to the tails. This is shown below for some typical error distributions with the shape parameter denoted by  $v$ .

Probability distribution of model errors $\{\epsilon_t\}_{t \geq 1}$	$a = E\{ \epsilon \}/\sigma_\epsilon$
Student's $t$ with $\nu$ degrees of freedom	$\sqrt{(\nu - 2)/\pi} \Gamma(\frac{1}{2}(\nu - 1))/\Gamma(\frac{\nu}{2})$
Log-normal	$1/\sqrt{\exp(\nu^2)} - 1$
Gamma	$(\Gamma(c + 1))/\sqrt{c}\Gamma(c)$
Weibull	$(\Gamma(\frac{c+1}{c}))/\sqrt{\Gamma(\frac{c+2}{c})}$

Therefore, in general, the asymptotic distribution of the statistics  $J_1^{(n)}$  and  $J_2^{(n)}$  has different support depending on the shape of the model error distribution, which acts as a nuisance parameter. However, the asymptotic distribution of the RUR test statistic  $J_0^{(n)}$  is free of nuisance parameters. This is in contrast with the unit root testing device suggested by BurrIDGE and Guerre (1996), based on the number of crossings, which in fact, depends on these nuisance parameters,  $a = \frac{E\{|\epsilon|\}}{\sigma_\epsilon}$ .

The empirical size and power of the RUR test against stationary AR(1) alternatives is shown in Table III below. We consider the estimated 5% critical values from 10,000 replications of the model  $x_t = bx_{t-1} + \epsilon_t$ , with the following distributions:  $\epsilon_t \sim$  n.i.d. (0,1) when  $\epsilon_t$  has a Student's  $t$  distribution with 5 degrees of freedom, when  $\epsilon_t$  has a mixture of  $N(-4, 9.766)$  and  $U(-1, 9)$  (notice that this distribution has a mean and median equal to zero but it is asymmetric), and finally the same AR(1) model but with  $\epsilon_t$  following a Cauchy distribution. The autoregressive parameter  $b$  was allowed to take different stationary values (power of the test) and a nonstationary value (size of the test) for different sample sizes  $n$ . The DF performances appear in parentheses.

The size of the RUR test is included in the last column of Table III. It is clear that there is no size distortion for different error distributions even if the distributions are asymmetric. As expected, the power of the DF test is higher than that of the RUR test for stationary alternatives that are far from 1 (say 0.5 and

TABLE III  
EMPIRICAL SIZE AND POWER OF RUR TEST FROM 10,000 REPLICATIONS FOR DIFFERENT SAMPLE SIZES  
 $n$  VALUES OF  $b$  AND ERROR DISTRIBUTIONS

$n$	$b$					
	0.5	0.8	0.9	0.95	0.99	1
Cauchy						
100	0.8 (1)	0.6 (0.99)	0.49 (0.4)	0.4 (0.09)	0.12 (0.03)	0.052 (0.051)
250	1 (1)	1 (1)	1 (1)	0.81 (0.6)	0.5 (0.07)	0.05 (0.05)
500	1 (1)	1 (1)	1 (1)	(0.9)	0.7 (0.1)	0.05 (0.05)
Student's $t$						
100	0.8 (1)	0.6 (0.99)	0.5 (0.4)	0.39 (0.09)	0.1 (0.02)	0.05 (0.04)
250	1 (1)	1 (1)	1 (0.99)	0.8 (0.65)	0.45 (0.06)	0.05 (0.05)
500	1 (1)	1 (1)	1 (1)	1 (0.97)	0.7 (0.08)	0.05 (0.05)
Asymmetric						
100	0.8 (0.99)	0.6 (0.98)	0.5 (0.4)	0.4 (0.08)	0.1 (0.02)	0.05 (0.04)
250	1 (1)	1 (1)	1 (1)	0.82 (0.7)	0.5 (0.06)	0.05 (0.05)
500	1 (1)	1 (1)	1 (1)	1 (1)	0.7 (0.1)	0.05 (0.05)

0.8). However, in contrast with the RUR test, the power of the DF decreases dramatically for autoregressive values near the unit root (0.95, 0.99).

#### 4.2. Power of the test against stationary alternatives with level shifts

We consider the case of a single structural break in the series in the middle of the sample. The break is modelled as a dummy variable defined by  $D_t = 0$  for  $t \leq n/2$  and  $D_t = 1$  for  $t > n/2$ . Specifically, we consider the alternative model to be  $x_t = 0.5x_{t-1} + s D_t + \epsilon_t$ . Table IV provides power estimates at the 5% significance level from 10,000 replications for different values of the sample size  $n$  and of the local break size  $s$ . The power performances of the DF test appear in parentheses.

We remark that except for the case of  $s = 4$  and  $n = 500$ , the DF test has no power. The RUR test is more powerful for sample sizes larger than 250 and therefore are less prone to misinterpret structural breaks as are permanent stochastic disturbances. In a scenario allowing for multiple breaks, we should expect a larger decrease in power for both the RUR and the DF tests. In order to assess these power losses, we performed another experiment which included two breaks at different locations in time. The alternative model was now  $x_t = 0.5x_{t-1} + s_1 D_{t,1} + s_2 D_{t,2} + \epsilon_t$  with  $D_{t,i}$  ( $i = 1, 2$ ) representing dummy variables defined by  $D_{t,i} = 0$  for  $t \leq in/4$  and  $D_{t,i} = 1$  for  $in/4 < t \leq in/2$ . Table V shows the power results at the 5% significance level obtained from 10,000 replications of this model, for both the RUR and the DF tests shown in parentheses. Here  $s_{1,2} = (s_1, s_2)'$ . The power estimates are given for different values of the sample size  $n$  (100, 250, 500), and of the break magnitudes  $s_1$  and  $s_2$  ( $s_1 = 2, 4, 8$  and  $s_2 = 4, 8, 12$ ,

TABLE IV  
EMPIRICAL POWER AGAINST THE ALTERNATIVE  $x_t = 0.5x_{t-1} + sD_t + \epsilon_t$ , FOR DIFFERENT VALUES OF THE SAMPLE SIZE  $n$  AND OF THE LOCAL BREAK SIZE  $s$

$n$	$s$		
	4	8	12
100	0.2 (0.00)	0.08 (0.00)	0.07 (0.00)
250	0.7 (0.00)	0.6 (0.00)	0.6 (0.00)
500	1 (0.86)	1 (0.00)	1 (0.00)

TABLE V  
EMPIRICAL POWER OF THE RUR TEST AGAINST THE ALTERNATIVE  $x_t = 0.5x_{t-1} + s_1 D_{t,1} + s_2 D_{t,2} + \epsilon_t$ , FOR DIFFERENT VALUES OF THE SAMPLE SIZE  $n$  AND OF THE LOCAL BREAK SIZE  $s$

$n$	$S_{1,2}$		
	(2, 4)	(4, 8)	(8, 12)
100	0.07 (0.000)	0.005 (0.000)	0.000 (0.000)
250	0.5 (0.000)	0.200 (0.000)	0.05 (0.000)
500	1 (0.453)	0.7 (0.000)	0.6 (0.000)

respectively). Once again, the RUR test outperforms the DF results in all cases, and is powerful for the sample size  $n = 500$ , as long as the break size is not too large.

To explain this robustness of the RUR test theoretically, consider the following AR(1) models, where we allow for the possibility of a single breakthrough innovation dynamics:

$$(a) \quad x_t = ax_{t-1} + \zeta_t, |a| < 1 \quad \text{with } E(\zeta_t) = 0,$$

$$(b) \quad x_t = ax_{t-1} + \epsilon_t, |a| < 1 \quad \text{with } E(\epsilon_t) = s\mathbf{1}(t = t_0).$$

Let  $J_\zeta^{(n)}$  and  $J_\epsilon^{(n)}$  be the RUR test statistics associated with the processes in models (a) and (b), respectively. Now if  $|a| > 0$  and  $t_0 > 0$  we will have

$$R_{t_0} = a + R_{t_0-1},$$

since on  $I(0)$  processes  $P(\Delta R_t > 0) = O(t^{-1}) \sim 0$  for  $t$  large enough, that is  $\Delta R_{t_0} = a$  with probability close to one. As a result,  $J_\epsilon^{(n)} \simeq J_\zeta^{(n)} + n^{-1/2}$  for both  $n$  and  $t_0$  large enough. But then  $J_\epsilon^{(n)} - J_\zeta^{(n)} \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , which means that the consistency of the test is not affected by the presence of a level break, whatever the size of such break.

When several level breaks are involved, say  $m$  breaks, we can write

$$E(\epsilon_t) = \sum_{i=1}^m s_i \mathbf{1}(t = t_i).$$

Now suppose

$$t^* \Delta \min_{1 \leq i \leq m} \{t_i\} > 0 \quad \text{and} \quad 0 < s^* \Delta \min_{1 \leq i \leq m} \{s_i\} < \infty,$$

such that

$$\text{and thereby} \quad P(\Delta R_{t^*} > 0) = O(t^{*-1}) \simeq 0$$

$$J_\epsilon^{(n)} \simeq J_\zeta^{(n)} + n^{-1/2} \sum_{i=1}^m s_i$$

for  $n$  large enough. Therefore  $J_\epsilon^{(n)} - J_\zeta^{(n)} \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , for finite  $m$ . What is more, the number of level breaks,  $m$ , can even grow indefinitely as  $o(n^{1/2})$  without affecting the consistency of the test.

#### 4.3. Size of the RUR test against level shifts, nonlinearities and additive outliers

We want to show that the asymptotic size of the RUR test is unaltered by the presence of as much as  $m = o(n^{1/2})$  level shifts superimposed on a  $I(1)$  time series. For that, we consider the following two AR(1) models:

$$(a) \quad x_t = x_{t-1} + \zeta_t, \quad \text{with } E(\zeta_t) = 0,$$

$$(b) \quad x_t = x_{t-1} + \epsilon_t, \quad \text{with } E(\epsilon_t) = \sum_{i=1}^m s_i \mathbf{1}(t = t_i),$$

with  $t^* > 0$  and  $0 < s^* < \infty$ , such that

$$P(\Delta R_r > 0) \quad O(t^{*-1/2}) \simeq 0 \quad \text{and thereby} \quad J_\epsilon^{(n)} \simeq J_\xi^{(n)} + n^{-1/2} \sum_{i=1}^m s_i.$$

Then as far as  $m = o(n^{1/2})$  we will get  $J_\epsilon^{(n)} - J_\xi^{(n)} \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , and the asymptotic size will be the same as in model (a).

Notice that if, on the contrary,  $m$  is allowed to be  $O(n^{1/2+\gamma})$  with  $\gamma > 0$ ,  $J_\epsilon^{(n)}$  will behave as if  $x_t$  had a trend, that is,  $J_\epsilon^{(n)} \rightarrow \infty$  as  $n \rightarrow \infty$ . Level breaks will then shift the null distribution to the left indefinitely leading to the rejection of the null hypothesis of an  $I(1)$  time series.

Next we analyse the small sample behaviour of the RUR test in the face of several nonlinear transformations of random walks, and show that it is invariant to monotonic transformations even in small samples. Table VI shows the size estimated at the 5% significance level from 10,000 replications of the different models and for  $n = 100, 250$ , and  $500$  and the DF results are in parentheses.

It can be observed that the size of the RUR test tends towards its correct 5% value in all the cases except when the transformation is non monotonic (case 8). In case (7), the transformation makes the series stationary and therefore the table reports the power, not the size. To study more precisely the effect of the logarithmic nonlinearities, in case (6), we forced the variable to take most of its values in the interval  $(0,1)$ . This was done by transforming the series linearly prior to applying the logarithmic transformation. Since, in this interval, the function is not so well approximated by a straight line, one would expect a more noticeable size distortion, at least for the smaller sample size of  $n = 100$ . Overall, however, all the empirical sizes for the purely monotonic transformations seem to converge to the nominal size of 0.05 as the sample size grows. The invariance of  $J_0^{(n)}$  to monotonic nonlinear transformations  $g(\cdot)$  applied to the series  $x_t$  follows immediately from the relations:

TABLE VI

EMPIRICAL SIZE OF THE RUR TEST AGAINST DIFFERENT FORMS OF NONLINEARITY APPLIED TO A RANDOM WALK  $x_t = x_{t-1} + \epsilon_t$

Transformations	100	250	500
<i>Monotonic</i>			
(1) $x_t^2$ with $x_t > 0, \forall$	0.03 (0.397)	0.059 (0.406)	0.048 (0.420)
(2) $x_t^3$	0.038 (0.456)	0.057 (0.532)	0.049 (0.533)
(3) $\exp(x_t)$	0.03 (0.92)	0.05 (1)	0.0469 (1)
(4) $\exp(\frac{x_t}{\sqrt{2}})$	0.054 (0.271)	0.0526 (0.271)	0.05 (0.301)
(5) $\log(x_t + 100)$	0.043 (0.275)	0.064 (0.331)	0.051 (0.354)
(6) $\log(\frac{x_t+2\sqrt{T}}{4\sqrt{T}}, \frac{x_t+2\sqrt{T}}{4\sqrt{T}} \in (0, 1)$	0.072(0.347)	0.054(0.349)	0.051(0.354)
<i>Non-monotonic</i>			
(7) $\sin(x_t)$	0.8828 (1)	0.9986 (1)	1 (1)
(8) $x_t^2$	0.079 (0.397)	0.170 (0.406)	0.178 (0.420)

$$\mathbf{1}(g(x_t) > g(x_{t-1,t-1})) \quad \mathbf{1}(x_t > x_{t-1,t-1})$$

$$\mathbf{1}(g(x_t) < g(x_{1,t-1})) \quad \mathbf{1}(x_t < x_{1,t-1}).$$

Notice that such invariance holds not only under the null hypothesis but also under any alternative. This result is in fact related to the invariance of the number of level crossings in a series (in this case, the first differences of the sequence of running ranges) to monotonic transformations.

The results in Table VII show that the size distortions caused by the presence of AO in the middle of the series and beyond are significantly smaller for the RUR test than for the DF test (shown in parentheses). Our alternative hypothesis was now represented by the model  $y_t = x_t + s\delta_{t,\tau}$ , where  $x_t = x_{t-1} + \epsilon_t$ ,  $\tau$  denotes an integer no larger than the sample size, and  $\delta_{t,\tau}$  is a dummy variable defined by  $\delta_{t,\tau} = 1$  if  $t = \tau$  and zero elsewhere. The sizes were estimated at the 5% significance level, for different values of both  $\tau$  ( $\tau = n/25, n/10, n/5$ ) and the sample size  $n$  (100, 250, 500). It can be seen that when the AO appears near the end of the series, the RUR test has even lower than nominal sizes.

Unfortunately, an early AO will produce a jump in the sequence of ranges which may prevent other jumps from being counted by the RUR test statistic, thus biasing our test towards the rejection of the null hypothesis of unit root. The bias will be larger, the sooner the outlier appears in the series. In order to grasp this problem, we performed another Monte Carlo experiment in which a single AO is introduced near the origin.

The results show that when the AO appears within the first quarter of the sample, the RUR test seems to offer no real improvement over the DF test. To give a flavour of what is going on in this case, suppose we have an AO early in the series at time  $t = t_0$ , and suppose that its magnitude,  $s$ , is such that

TABLE VII  
EMPIRICAL SIZE OF RUR TEST AGAINST THE MODEL  $y_t = x_t + s\delta_{t,\tau}$ , WHERE  $x_t = x_{t-1} + \epsilon_t$  AND DIFFERENT LOCATIONS OF THE AO

$n$	$\tau$		
	$(n/2)$	$(n/2) + 1$	$(n/2) + 2$
100	0.0826 (0.2978)	0.0830 (0.2964)	0.0812 (0.2958)
250	0.0800 (0.1682)	0.0800 (0.1688)	0.0798 (0.1670)
500	0.0644 (0.1130)	0.0640 (0.1102)	0.0642 (0.1096)
	$n$ ( $n/20$ )	$n$ ( $n/10$ )	$n$ ( $n/5$ )
100	0.0212 (0.2964)	0.0244 (0.2990)	0.0352 (0.2980)
250	0.0392 (0.1704)	0.0422 (0.1660)	0.0484 (0.1656)
500	0.0446 (0.1106)	0.0472 (0.1104)	0.0510 (0.1118)
	$n/25$	$n/10$	$n/5$
100	0.3778 (0.2956)	0.3192 (0.2964)	0.2432 (0.3002)
250	0.2746 (0.1672)	0.2230 (0.1668)	0.1700 (0.1676)
500	0.1930 (0.1114)	0.1588 (0.1112)	0.1188 (0.1110)



$\Delta R_{t_0}^{(x)} \geq \max_{1 \leq t \leq n} \Delta R_t^{(x)}$ . Such a large outlier will prevent new records from occurring at  $t > t_0$ , and therefore  $\Delta R_t^{(x)} = 0$  for  $t > t_0$ . It follows that

$$J_0^{(n)} = n^{-1/2} \sum_{t=1}^{t_0} \mathbf{1}(\Delta R_t^{(x)} > 0) \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty,$$

and the test will then be likely to reject the null hypothesis. Notice that the previous result still holds when the AO location is allowed to increase with the sample size as fast as  $O(n^{1-\gamma})$  with  $\gamma > 0$ . Obviously, when more than one early AO appears, the record count will be determined by the largest AO's location, but the real size of the test will grow to one, in the same way, as  $n \rightarrow \infty$ .

The relatively large size distortion of the RUR test in the presence of early AOs can be solved, however, by slightly modifying the test statistic so as to also count the records appearing when the series is observed in reverse order. This is the purpose of Section 5.

## 5. THE FORWARD-BACKWARD RANGE UNIT-ROOT TEST

Unless we know the outlier locations, the amount of size distortion or bias of the RUR test, based on the statistic  $J_0^{(n)}$ , when confounded with time series with AOs will be uncertain. By means of a simple resampling technique, we obtain an extension of the RUR test, called the FB RUR test, based on the statistic here noted as  $J_*^{(n)}$ , which reduces the size distortion when the AO occurs at the beginning of the sample, and which turns out to be smaller than with the DF test. The FB RUR test also improves the power performances of the former RUR test.

This extension consists of running the RUR test first forwards (from the beginning to the end of the sample) and then backwards (from the end to the beginning). The total jump count corresponds therefore to a sample size twice the original one, thus leading to improved size and power performances, in general. The FB RUR test statistic  $J_*^{(n)}$  can be formulated as follows:

$$J_*^{(n)} = \frac{1}{\sqrt{2n}} \sum_{t=1}^n \left\{ \mathbf{1}(\Delta R_t^{(x)} > 0) + \mathbf{1}(\Delta R_t^{(x')} > 0) \right\}, \quad (11)$$

where  $x'_t = x_{n-t+1}$ ,  $t = 1, 2, \dots, T$ , denotes the time reversed series.

The asymptotic null distribution of the FB RUR test statistic  $J_*^{(n)}$  can be obtained from the asymptotic null distribution of  $J_0^{(n)}$ . Indeed, we can write:

$$J_*^{(n)} = \frac{1}{\sqrt{2n}} (J_x^{(n)} + J_{x'}^{(n)})$$

where we have set

$$J_x^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{1}(\Delta R_t^{(x)} > 0),$$

$$J_{x'}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{1}(\Delta R_t^{(x')} > 0).$$

On the other hand, since

$$\begin{aligned} x'_t &= x_n - \sum_{i=t+1}^n \epsilon_i, \\ x'_{t'} &= x_0 + \sum_{i=1}^{t'} \epsilon_i, \end{aligned}$$

we obtain, for  $t' \leq t$ :

$$E\{x'_t x'_{t'}\} = E\{x_0 x_T\} = E(x_0^2).$$

Now suppose  $x_0$  is fixed and the  $\epsilon_i$  are Gaussian  $\text{Cov}(x_0, x_n) = 0$ . These assumptions entail that  $x'_t$  and  $x_t$  are statistically independent as long as  $t' \leq t$ . Thus, let  $k_n = n^{1-\gamma}$  for some  $\gamma$  such that  $0 < \gamma < \min(1, \ln 2/\ln n)$ . We can write:

$$J_*^{(n)} = \frac{1}{\sqrt{2n}} (J_x^{(n-k_n)} + J_{x'}^{(n-k_n)} + J_x^{(n-k_n+1,n)} + J_{x'}^{(n-k_n+1,n)}),$$

where

$$\begin{aligned} J_x^{(n-k_n+1,n)} &= \frac{1}{\sqrt{n}} \sum_{t=n-k_n+1}^n \mathbf{1}(\Delta R_t^{(x)} > 0), \\ J_{x'}^{(n-k_n+1,n)} &= \frac{1}{\sqrt{n}} \sum_{t=n-k_n+1}^n \mathbf{1}(\Delta R_t^{(x')} > 0). \end{aligned}$$

Notice that  $k_n \geq n/2$ , and thereby the random variables  $J_x^{(n-k_n)}$  and  $J_{x'}^{(n-k_n)}$  are independent. Secondly, the term  $J_x^{(n-k_n+1,n)} + J_{x'}^{(n-k_n+1,n)}$  is asymptotically negligible with respect to the term  $J_x^{(n-k_n)} + J_{x'}^{(n-k_n)}$ . Finally, both  $J_x^{(n-k_n)}$  and  $J_{x'}^{(n-k_n)}$  converge weakly to the same limiting variable  $J_0^{(\infty)}$ , by virtue of the duality theorem (Feller, 1971, vol. 2, p. 443).<sup>2</sup> Therefore  $J_*^{(n)} \Rightarrow 1/\sqrt{2n} J_0^{(\infty)}$ , where

$$f_{J_0^{(\infty)}}(u) = (u)^2 \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}(u)^2} \quad \text{where } u = \xi + \eta$$

independent of  $a$  where  $\xi \rightarrow |B(1)|$  and  $\eta \rightarrow l_B(0,1)$ .

The power of the FB RUR test against the alternative of a stationary AR(1) time series with n.i.d. (0,1) model errors is shown in Table VIII. Notice the improvements in power performances, especially for the smaller sample sizes, where now DF (in parentheses) is outperformed, in all the cases except when the value of the autoregression parameter  $b$  is far from the unit root ( $b = 0.5$  and  $0.8$ ).

TABLE VIII  
EMPIRICAL SIZE AND POWER FROM 10,000 REPLICATIONS FOR DIFFERENT SAMPLE SIZES  $n$  AND FOR DIFFERENT VALUES OF  $b$

$n$	$b$					
	0.5	0.8	0.9	0.95	0.99	1
100	1.00 (1.00)	0.80 (0.99)	0.60 (0.5)	0.5(0.18)	0.2 (0.0375)	0.05 (0.05)
250	1.00 (1.00)	1.00 (1.00)	1.00 (1)	0.9(0.7)	0.52 (0.0760)	0.05 (0.05)
500	1.00 (1.00)	1.00 (1.00)	1.00 (1)	1 (0.99)	0.8 (0.39)	0.05 (0.05)

### 5.1. Behaviour of the FB RUR test under departures from the standard assumptions

#### 5.1.1. Robustness against outliers, structural breaks and monotonic nonlinearities

In order to quantify, in finite samples, the size distortion of the FB RUR test in the presence of additive outliers, we used the same experimental framework as for the original RUR test. Table IX shows the Monte Carlo results, depending on whether the single outlier's location is at the beginning, in the middle, or at the end of the sample. For comparison, we let the DF test results appear in parentheses. (See also Franses and Haldrup, 1994.)

On the one hand, notice that even in cases where the outlying observations appear at the beginning of the data sample, the FB RUR test is more robust than the DF test, Dickey and Fuller (1979). Indeed, since in this case the last jump occurs at the largest outlier's location,  $J_0^{(n)}$  tends to be very small (and, asymptotically, zero), whereas  $J_*^{(n)}$  will only be approximately reduced by a factor of  $1/\sqrt{2}$  with respect to the case of no outliers. Notice also that we should not expect any improvement in performances of the FB RUR test over the RUR test when the AOs occur in the middle of the sample. Finally, this competitive edge of the FB RUR test disappears when outliers occur at both the beginning and the end of the sample. However, this situation is more unlikely.

TABLE IX  
EMPIRICAL SIZE AGAINST THE MODEL  $y_t = x_t + s\delta_{t,\tau}$  AND DIFFERENT LOCATIONS OF THE AO

$n$	$\tau$		
	$n/25$	$n/10$	$n/5$
100	0.1206 (0.2956)	0.0880 (0.2964)	0.0550 (0.3002)
250	0.1156 (0.1672)	0.0950 (0.1668)	0.0678 (0.1676)
500	0.0918 (0.1114)	0.0726 (0.1112)	0.0638 (0.1110)
$n/\tau$	$n/2$	$n/2 + 1$	$n/2 + 2$
100	0.04 (0.2978)	0.0630 (0.2964)	0.0512 (0.2958)
250	0.0500 (0.1682)	0.0500 (0.1688)	0.0598 (0.1670)
500	0.0544 (0.1130)	0.0540 (0.1102)	0.042 (0.1096)
$n/\tau$	$n$ $n/20$	$n$ $n/10$	$n$ $n/5$
100	0.0212 (0.2964)	0.0244 (0.2990)	0.0552 (0.2980)
250	0.0692 (0.1704)	0.0522 (0.1660)	0.0584 (0.1656)
500	0.0546 (0.1106)	0.0572 (0.1104)	0.0510 (0.1118)

To have a closer look at this property of the FB RUR test, suppose we have an  $I(1)$  time series corrupted by an isolated AO of size  $a$  at time  $t = t_0$ . That is, let

$$x_t = y_t + a\mathbf{1}(t = t_0) \quad \text{with } y_t = y_{t-1} + \epsilon_t$$

where the i.i.d. random variables  $\{\epsilon_i\}_{i \geq 1}$  are supposed to be zero mean and finite variance as well as a symmetric pdf. The worst case corresponds to when  $a$  is large enough so that

$$\Delta R_{t_0}^{(x)} = a + \Delta R_{t_0}^{(y)}$$

$$\Delta R_t^{(x)} = 0, \quad \forall t > t_0,$$

implying, for the RUR test statistic

$$J_x^{(n)} = J_y^{(t_0)}$$

and thereby  $J_x^{(n)} \xrightarrow{p} 0$ . We obtain the same result when the AO's location  $t_0$  increases with  $T$  as long as  $t_0 = o(n)$ . In this case,

$$J_y^{(t_0)} \sim \left(\frac{t_0}{n}\right)^{1/2} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

The real size of the RUR test in the presence of such an AO will tend to its maximum distortion asymptotically.

Things are quite different as far as the FB RUR test statistic

$$J_*^{(n)} = \frac{1}{\sqrt{2n}} (J_x^{(n)} + J_{x'}^{(n)})$$

is concerned. Indeed, when  $J_*^{(n)}$  is used instead of  $J_x^{(n)}$  we get  $J_{x'}^{(n)} = J_y^{(n-t_0)}$ , and therefore

$$\begin{aligned} J_*^{(n)} &= \frac{1}{\sqrt{2n}} (J_y^{(t_0)} + J_y^{(n-t_0)}) \\ &\Rightarrow \frac{1}{\sqrt{2n}} J_y^{(\infty)} \stackrel{d}{=} \frac{1}{\sqrt{2n}} J_y^{(\infty)}, \end{aligned} \quad (12)$$

by virtue of the aforementioned duality theorem (Feller, 1971). Once again, this result still holds when  $t_0$  is allowed to increase more slowly than  $n$ . As a consequence, an early outlier affects the asymptotic distribution of  $J_*^{(n)}$  only by a factor of  $1/\sqrt{2}$ . Correspondingly, the real size of the FB RUR test in the presence of this type of outliers will be only slightly increased.

When considering the alternative of a stationary AR(1) time series about a single structural break, we obtain a remarkable improvement in power performance over the former RUR test. As expected, the results deteriorate when two breaks are present in the DGP of the time series, and thereby a larger sample size ( $n = 500$ ) is required in order to notice these improvements. (See Arranz and Escribano, 2000.)

Finally, as regards the robustness of the FB RUR test to monotonic nonlinearities, no significant differences are obtained with respect to the former RUR test. It is also straightforward to show that the FB RUR test, based on  $J_*^{(n)}$ , has the same invariance properties and asymptotics as the one based on  $J_0^{(n)}$ .

## 6. FURTHER ANALYSIS OF SIZE DISTORTIONS WITH SERIAL CORRELATION AND HETEROSKEDASTICITY IN THE ERRORS

It is well known that parametric unit root tests run into serious problems if the errors are generated by an MA process with a root close to one (Schwert, 1989; Agiakloglou and Newbold, 1992). The Phillips Perron test has been shown to suffer important size distortions in this case.

We consider the DGP

$$\begin{aligned} x_t &= bx_{t-1} + \epsilon_t, \quad \text{with } b = 1, 0.8 \\ \epsilon_t &= u_t - \beta u_{t-1} \quad \text{with } u_t \sim N.i.d.(0, 1) \end{aligned} \tag{13}$$

Table X presents the rejection frequencies of the augmented Fuller test (ADF) where the lags are selected according to Ng and Perron (2001) using modified information criteria (MAIC) using a local asymptotic framework. The rejection frequencies are based on 5000 replications of model (13), sample size  $n = 100$  and the nominal significance level is 0.05, for  $b = 1$  (i.e. the actual size) and  $b = 0.8$  (i.e. the empirical power).

For stationary cases with  $b = 0.8$ , the power of the three unit root tests considered is similar. For nonstationary cases, the size of the ADF test with the MAIC criteria of Ng and Perron (2001) outperforms the rest of the tests.

Finally, we also consider the size of the RUR and FB RUR in two cases of heteroskedasticity: nonstationary autoregressive AR(1) processes with

TABLE X  
EMPIRICAL SIZE AND EMPIRICAL POWER

$\beta$	ADF <sub>MAIC</sub>	RUR	FB-RUR
<i>b = 1 (size)</i>			
0.5	0.04	0.004	0.006
0	0.05	0.039	0.04
0.2	0.04	0.05	0.05
0.4	0.043	0.1	0.07
0.6	0.044	0.2	0.09
0.8	0.03	0.6	0.1
<i>b = 0.8 (power)</i>			
0.5	0.3	0.3	0.3
0	0.5	0.5	0.53
0.2	0.7	0.7	0.8
0.4	0.9	0.9	0.95
0.6	1	1	1
0.8	1	1	1

heteroskedastic errors, following Politis *et al.* (1997), and a nonstationary autoregressive AR(1) processes with generalized autoregressive conditional heterostedastic GARCH (1,1) errors. The empirical size of the test converges to the nominal size for sample sizes of 1000 observations. The convergence is faster for FB RUR which converges to the nominal size for sizes of 500 observations. The tables with the simulation results are available upon request.

## 7. EMPIRICAL APPLICATION

In this section, we illustrate the performances of our robust unit root testing methodology on real time series. Our example studies the annual US/Finland real exchange rate series from 1900 to 1987, which is contaminated with both additive and innovation outliers.

### 7.1. Analysis of the annual US/Finland real exchange rates: 1900 1987

In this section, the RUR and FB RUR tests were applied to the annual series of US/Finland real exchange rates, whose logarithm is plotted in Figure 3. This series, which contains a total of  $n = 88$  observations (from 1900 to 1987), was constructed using the gross domestic product (GDP) deflator. Previous analyses on this series done by Volgelsang (1990), Franses and Haldrup (1994), Perron and Volgelsang (1992), and Perron and Rodriguez (2000), point to the presence of an AO in 1918 together with innovation outliers (IOs) that produce temporary changes in 1917, 1932, 1949 and 1957.

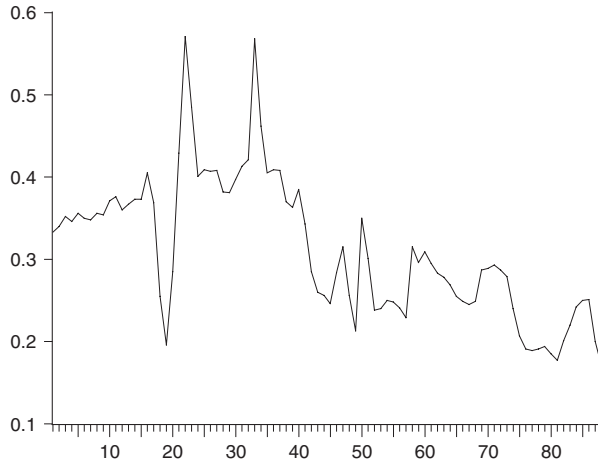


FIGURE 3. Logarithm of the US/Finland real exchange rates deflated annual series from 1900 to 1987.

Using Mackinnon's critical values for the ADF test (Mackinnon, 1994 [31]), the null hypothesis of a unit root is rejected at the 5% significance level (the ADF test statistic took the value  $-3.732041$  while the 5% critical values was  $-3.4614$ ).

Alternatively, with a value of  $J_0 = 1.4924$  obtained for the RUR test statistic, and the corresponding estimated critical value of  $1.1726$  at the 5% significance level and for  $n = 88$ , the null is not rejected. Similarly, for the FB RUR test we obtained a value for its test statistic of  $J_{0,*} = 2.11$ , which is also larger than the corresponding estimated critical value, that is  $1.7337$ .

## 8. CONCLUDING REMARKS

Standard unit root tests suffer from a number of drawbacks when the usual assumptions are no longer justified. Apart from having low power on stationary near unit root time series, they are also seriously affected by other aspects of real data such as parameter shifts, outliers and neglected nonlinearities. Burridge and Guerre (1996) proposed a nonparametric unit root testing device based on the number of crossings. This test was sensitive to the tails of the error distribution. We have presented a nonparametric testing device, called the RUR test, which is robust to important structural breaks either in the mean or in the variance, as well as to the presence of non early additive outliers. The new method is also invariant to monotonic nonlinearities in the DGP and outperforms the DF test in terms of power on stationary near unit root alternatives. Finally, it is asymptotically immune to the presence of additive noise superimposed on an unobserved variable. A drawback of the test is its sensitivity to early additive outliers, which may lead to a size distortion comparable with those of DF tests. However, by simply running the test forwards and backwards it is possible to circumvent this problem and improve other aspects of previous test performances. A few real time series were selected to illustrate our tests and compare their results to those of the DF test. In spite of the small sample size considered, we found discrepancies in all cases between both types of tests, which question the validity of the standard test's outcome.

## APPENDIX

In this section, we provide the proofs for the theoretical results presented in previous sections. For this we need to invoke the following lemmas.

### A0. PRELIMINARY LEMMAS

**LEMMA 1 (HERRNDORF'S INVARIANCE PRINCIPLE).** *Let  $\{\epsilon_t\}_{t=1,\infty}$  be a random sequence satisfying the mixing condition of Phillips and Perron (1988), then defining*

$$x_n(r) = \sigma^{-1} n^{-1/2} \sum_{t=1}^{[nr]} \epsilon_t \Rightarrow B(r)$$

where  $B(\cdot)$  is a Brownian motion process on the interval  $[0,1]$ ,  $\sigma$  represents the long run, and ' $\Rightarrow$ ' denotes convergence in distribution as  $n \rightarrow \infty$ .

PROOF. See Herrndorf (1984). □

LEMMA 2 (CONTINUOUS MAPPING THEOREM). *Let  $T$  be continuous function (except possibly on a set with Lebesgue measure equal to zero) such that:  $T : C[0,1] \mapsto C[0,1]$ , where  $C[0,1]$  denotes the space of cadlag functions on the interval  $[0,1]$ . Let  $x_n(r)$  be defined as in Lemma 1. Then*

$$T(x_n(r)) \Rightarrow T(B(r)).$$

PROOF. See Billingsley (1968). □

LEMMA 3. *Let  $S_t = \sup_{s \in [0,t]} \{B(s)\}$ ,  $I_t = \inf_{s \in [0,t]} \{B(s)\}$  and  $x_n(r)$  defined as in Lemma 1. Under the mixing condition of Phillips and Perron (1988) we have:*

$$x_n(r) = \max_{s \in [0,1]} \{x_n(s)\} \quad T_1(x_n(r)) \Rightarrow B(r) \quad S_t$$

$$x_n(r) = \min_{s \in [0,1]} \{x_n(s)\} \quad T_2(x_n(r)) \Rightarrow B(r) \quad I_t$$

PROOF. The proof follows from the CMT (Lemma 2) and the continuity of the functions  $T_1$  and  $T_2$ . □

LEMMA 4 (LÉVY, 1948). *Let  $\{B(r)\}_{r \in [0,1]}$  represent a Brownian motion process on the interval  $[0,1]$ , and let  $\tilde{B}_1(r) = B(r) - S_t$  and  $\tilde{B}_2(r) = B(r) - I_t$ . The processes  $\{|B(r)|\}_{r \in [0,1]}$ ,  $\{\tilde{B}_1(r)\}_{r \in [0,1]}$  and  $\{\tilde{B}_2(r)\}_{r \in [0,1]}$  have the same probability distribution.*

PROOF. See Karatzas and Shreve (1991). □

LEMMA 5. *The dimensional processes  $(S_t = B(r), S_t)$ ,  $(|B_t|, I_B(0, r))$  and  $(|B(r)|, \frac{1}{2} I_B(0, r))$  have the same law.*

PROOF. See Revuz and Yor (1991, p. 240 and 244). □

LEMMA 6. *The joint law of  $(S_t = B(r), S_t)$  has density*

$$f(a, b) = \sqrt{\frac{2}{\pi t^3}} (a + b) \exp(- (a + b)^2 / 2t)$$

for  $a, b \geq 0$ .



PROOF. See Revuz and Yor (1991, p. 245). □

LEMMA 7. Let  $x_t = x_{t-1} + \epsilon_t$  where  $\{\epsilon_t\}_{t \geq 1}$  are i.i.d. random variables with zero mean and finite variance  $\sigma_\epsilon^2$ , and let

$$J_0^{(n)}(b) = n^{-1/2} \sum_{t=1}^n [\mathbf{1}(x_{t-1} < b, x_t \geq b) + \mathbf{1}(x_{t-1} > b, x_t \leq b)]$$

denote the normalized number of crossings of level  $b$ . If  $x_0$  and  $\epsilon_1$  have bounded pdfs with finite variance then we must have:

$$J_0^{(n)}(b) \Rightarrow \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} |Z|,$$

where  $Z$  is a standard Normal random variable.

PROOF. See Theorem 1 in Burrige and Guerre (1996). □

LEMMA 8 (LÉVY, 1948). Let  $Z$  be a standard normal random variable and let

$$I_B(0, 1) = \lim_{\delta \downarrow 0} \frac{1}{2\delta} \int_0^\delta \mathbf{1}[\delta \leq B(s) \leq \delta] ds,$$

where  $\{B(r)\}_{r \in [0,1]}$  is a Brownian motion process on  $[0,1]$ . Then

$$|Z|^d \stackrel{d}{=} I_B(0, 1).$$

PROOF. See Theorem 2.3 in Revuz and Yor (1991). □

LEMMA 9. Let  $\{x_t\}_{t \geq 1}$  be a stationary Gaussian sequence with covariances  $\{c_i\}_{i \geq 1}$  satisfying the Berman condition:  $c_i \log i \rightarrow 0$  as  $n \rightarrow \infty$ . Then all extreme statistics have the same asymptotic distributions as an i.i.d. Gaussian sequence.

PROOF. Theorem 2.5.2 in Leadbetter and Rootzén (1988). □

LEMMA 10. If  $\{x_t\}_{t \geq 1}$  is a sequence of i.i.d. random variables then for large  $n$

$$E\{n^{1/2} J_0^{(n)}\} = O(\log n)$$

$$\text{Var}\{n^{1/2} J_0^{(n)}\} = O(\log n).$$

PROOF. See, for instance Port (1994). □

LEMMA 11. Let  $\{\xi_i\}_{i \geq 1}$  a sequence of random variables such that  $\lim_{i \rightarrow \infty} E(\xi_i) = \mu$ , and  $\lim_{i \rightarrow \infty} \text{Var}(\xi_i) = 0$ . Then

$$\xi_i \xrightarrow{P} \mu.$$

PROOF. See, for instance Arnold (1990). □

LEMMA 12. Let  $x_i = x_{i-1} + \epsilon_i$  where  $\{\epsilon_i\}_{i \geq 1}$  are continuous i.i.d. random variables with finite variance  $\sigma_\epsilon^2$  and symmetric pdf around a zero mean. If  $t'$  is the random time of occurrence of the maximum of  $\{x_i\}_{1 \leq i \leq t}$  then for any  $u \in [0, 1]$ :

$$P\{t'/t \leq u\} = \frac{2}{\pi} \int_0^u \arcsin \sqrt{v} \, dv$$

PROOF. See Lévy (1948). □

#### A1. PROOF OF PROPOSITION 1

Let  $x_t = x_{t-1} + \epsilon_t$  where  $\epsilon_t$  satisfy mixing condition of Phillips and Perron (1988), and let

$$\begin{aligned} \Psi^{(n)} &= n^{-1} \sum_{t=1}^n \mathbf{1}(R_t^{(x)} > 0) - n^{-1/2} J_0^{(n)} \\ &= n^{-1} \sum_{t=1}^n \mathbf{1}(x_t > x_{1,t}) + n^{-1} \sum_{t=1}^n \mathbf{1}(x_t < x_{1,t}) \\ &= \Psi_1^{(n)} + \Psi_2^{(n)} \end{aligned}$$

Note that  $\Psi^{(n)}$  is the frequency of upper and lower records in the sample  $\{x_1, \dots, x_n\}$ , and that we could also split this frequency into the sum of the frequencies of upper and lower records as:

$$\begin{aligned} \Psi_1^{(n)} &= \sum_{t=1}^n \mathbf{1} \left[ \frac{n^{-1/2} x_t}{\sigma} - \frac{n^{-1/2} x_{1,t}}{\sigma} > 0 \right] \left[ \frac{t}{n} - \frac{t-1}{n} \right] \\ \Psi_2^{(n)} &= \sum_{t=1}^n \mathbf{1} \left[ \frac{n^{-1/2} x_t}{\sigma} - \frac{n^{-1/2} x_{1,t}}{\sigma} < 0 \right] \left[ \frac{t}{n} - \frac{t-1}{n} \right]. \end{aligned}$$

Now defining  $r = t/n$ , where  $t = 1, 2, \dots, n$ , and letting  $n \rightarrow \infty$  we obtain from direct application of Lemmas 1, 2 and 3:

$$\Psi_1^{(n)} \Rightarrow \int_0^1 \mathbf{1}[B(r) - \sup_{s \in [0,1]} \{B(s)\} > 0] dr$$

$$\Psi_2^{(n)} \Rightarrow \int_0^1 \mathbf{1}[B(r) \leq \inf_{s \in [0,1]} \{B(s)\} - 0] dr$$

finally, it Follows from Lemma 4 and from the definition of local time that

$$\begin{aligned} \Psi_i^{(n)} &\Rightarrow \int_0^1 \mathbf{1}[|B(r)| \leq 0] dr, \quad i = 1, 2 \\ &\int_0^1 \mathbf{1}[B(r) \leq 0] dr \\ &0 \end{aligned}$$

$$\lim_{t \rightarrow \infty} P(x_t \leq x_{1,t}) = 0$$

$$\lim_{t \rightarrow \infty} P(x_t \leq x_{1,t}) = 0.$$

## A2. PROOF OF THEOREM 1

1. Consider a time series process  $x_t = \sum_{i=1}^t \epsilon_i$  where  $\{\epsilon_i\}_{i \geq 1}$  are continuous i.i.d. random variables with zero mean and variance  $\sigma_\epsilon^2$ . Let  $y_t = x_t - x_{1,t}$  and  $y'_t = x_t - x_{1,t}$  and split the RUR test statistic as

$$J_0^{(n)} = J_1^{(n)} + J_2^{(n)},$$

with

$$\begin{aligned} J_1^{(n)} &= n^{-1/2} \sum_{t=1}^n \mathbf{1}(y_t \leq 0) \\ &= n^{-1/2} \sum_{t=1}^n \mathbf{1}(y_{t-1} < 0, y_t \leq 0) + n^{-1/2} \sum_{t=1}^n \mathbf{1}(y_{t-1} \geq 0, y_t \leq 0) \\ &= n^{-1/2} \sum_{t=1}^n \mathbf{1}(y_{t-1} < 0, y_t \leq 0), \\ J_2^{(n)} &= n^{-1/2} \sum_{t=1}^n \mathbf{1}(y'_t \leq 0) \\ &= n^{-1/2} \sum_{t=1}^n \mathbf{1}(y'_{t-1} > 0, y'_t \leq 0) + n^{-1/2} \sum_{t=1}^n \mathbf{1}(y'_{t-1} \leq 0, y'_t \leq 0) \\ &= n^{-1/2} \sum_{t=1}^n \mathbf{1}(y'_{t-1} > 0, y'_t \leq 0), \quad \text{since } P(y'_{t-1} \leq 0, y'_t \leq 0) = 0. \end{aligned}$$

where we have used the fact that

$$\sum_{t=1}^n \mathbf{1}(y_{t-1} \geq 0, y_t \leq 0) = o\left(\sum_{t=1}^n \mathbf{1}(y_{t-1} < 0, y_t \leq 0)\right)$$

$$\sum_{i=1}^n \mathbf{1}(y'_{i-1} \leq 0, y'_i = 0) \quad o\left(\sum_{i=1}^n \mathbf{1}(y'_{i-1} > 0, y'_i = 0)\right).$$

Notice that the number of lower records of  $x_t$  in any given time interval is the same as the number of upper records of  $-x_t$  in that same interval. Therefore, the asymptotic distribution of  $J_1^{(n)}$  and  $J_2^{(n)}$  must be identical. To obtain this distribution we will proceed by first showing that the time series processes defined as  $y_t$  and  $y'_t$  are asymptotic random walks. By symmetry, the behaviour of  $y'_t$  must be statistically equal to that of  $y_t$ . It is therefore enough to study the properties of the process  $\{y_t\}_{t \geq 1}$ .

The conditional variance of  $y_t$  given that  $x_{t,t'} \quad x_t(t' \in [1, n] \cap \mathcal{Z})$  is

$$\text{var}(y_t | x_{t'} \quad x_{t,t'}) = \text{var}\left(\sum_{i=t'+1}^t \epsilon_i\right) = (t - t')\sigma_\epsilon^2$$

From Lemma 12, the random variable  $t'/t$  has an arcsine distribution with pdf:

$$f(t'/t) = \frac{2}{\pi\sqrt{1 - (t'/t)^2}}, \quad t'/t \in [0, 1],$$

from which we obtain the following expression for the unconditional variance:

$$\text{var}(y_t) = \frac{2\sigma_\epsilon^2}{\pi} \int_0^1 \frac{t - t'}{\sqrt{1 - (t'/t)^2}} d(t'/t) = t \frac{\sigma_\epsilon^2}{2}.$$

As a consequence,  $y_t$  cannot be an  $I(0)$  time series process. In fact, if we write  $y_t = y_{t-1} + \eta_t$  where  $\eta_t$  is  $I(0)$  and force the equality between this representation and the definition, we get

$$\begin{aligned} \eta_t &= \epsilon_t - \Delta x_{t,t} \\ &= \epsilon_t - (x_{t,t} - x_{t-1,t-1}) \\ &= \epsilon_t, \text{ if } x_t \leq x_{t-1,t-1} \\ &= \epsilon_t - (x_t - x_{t-1,t-1}), \text{ if } x_t \geq x_{t-1,t-1} \end{aligned}$$

Now, from Proposition 1, we know that the long run frequency of records is equal to zero, and thus  $\lim_{t \rightarrow \infty} P(x_t \geq x_{t-1,t-1}) = 0$ . It follows that  $\eta_t = \epsilon_t$  with probability  $p_t = P(x_t < x_{t-1,t-1}) \rightarrow 1$ . In particular:

$$E(\eta_t) = 0$$

$$\text{var}(\eta_t | x_{t'} \quad x_{t,t'}) = \sigma_\epsilon^2 \text{ with probability } p_t \rightarrow 1$$

$$\text{var}(\eta_t | x_{t'} \quad x_{t,t'}) = (t - t' - 1)\sigma_\epsilon^2 \text{ with probability } 1 - p_t \rightarrow 0,$$

from where the unconditional variance of  $\eta_t$  is obtained:

$$\begin{aligned} \text{var}(\eta_t) &= \frac{2\sigma_\epsilon^2}{\pi} \int_0^1 \frac{t-t'}{\sqrt{1-(t'/t)^2}} d(t'/t) \\ &= \sigma_\epsilon^2 \left( t - \frac{1}{2} \right) \quad \text{with probability } 1 - p_t \rightarrow 0. \\ \text{var}(\eta_t) &= \frac{2\sigma_\epsilon^2}{\pi} \int_0^1 \frac{1}{\sqrt{1-(t'/t)^2}} d(t'/t) \\ &= \sigma_\epsilon^2 \quad \text{with probability } p_t \rightarrow 1. \end{aligned}$$

Since, in practice, it can be assumed that the process  $x_t$  was generated at  $t \rightarrow \infty$ , we conclude that  $\eta_t$  is  $I(0)$ .

It can also be shown that for  $t$  small enough the process  $y_t$  has a stochastic unit root. The heuristic reasoning is as follows. Writing  $y_t = a_t y_{t-1} + \epsilon_t$  and assuming  $y_{t-1} / 0$  (event whose long run frequency equals one) we obtain the expected value of the process  $a_t$  given the past of  $y_t$ :

$$E(a_t | y_{t-1}) = 1 + \frac{\eta_t - \epsilon_t}{y_{t-1}}.$$

Thus there is a possibly non observable period of time during which  $\eta_t$  can be less than  $\epsilon_t$ , implying a transitory short memory behaviour for  $y_t$ . Notice however that as  $t \rightarrow \infty$  we get  $E(a_t | y_{t-1}) \rightarrow 1$ , and thus  $y_t$  becomes an  $I(1)$  process.

Given that  $y_t$  is an  $I(1)$ , and noting that for this process a zero ‘crossing’ amounts to a visit to the origin (crossing over the zero level is impossible), it follows from Lemma 9 that

$$J_1^{(n)} \Rightarrow \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} |Z|,$$

where  $Z$  is a standard normal random variable. From Lemma 7 the distribution of  $|Z|$  is the same as the local time at zero of a Brownian motion in  $[0, 1]$ , say  $l_B(0, 1)$ . Therefore, we can write

$$J_1^{(n)} \Rightarrow \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} l_B(0, 1).$$

By the same token we have:

$$J_0^{(n)} \Rightarrow \frac{E\{|\epsilon_1|\}}{\sigma_\epsilon} l_B(0, 1).$$

Since the pdf of the absolute value of a standard normal random variable  $Z$  is given by

$$f_{|Z|}(u) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{u^2}{2}\right), \quad u \geq 0,$$

we can easily obtain for the pdf of  $J_i^{(\infty)}(i = 1, 2)$  the following expression:

$$f_{J_i^{(\infty)}}(u) = \frac{2}{\sqrt{2\pi(E\{|\epsilon_1|\}/\sigma_\epsilon)^2}} \exp\left(-\frac{h^2}{2(E\{|\epsilon_1|\}/\sigma_\epsilon)^2}\right), \quad h \geq 0, \quad i = 1, 2.$$

2. From the Lemmas 5 and 6 we know that joint law of  $(|B(1)|, l(0,1))$  has density

$$f(\xi, \eta) = \sqrt{\frac{2}{\pi}}(\xi + \eta) \exp(-(\xi + \eta)^2/2) \quad \text{for } \xi, \eta \geq 0.$$

$J_0^{(n)} \rightarrow a[\xi + \eta]$  where  $\xi \rightarrow |B(1)|$  and  $\eta \rightarrow l(0,1)$ . Let  $w = a(\xi + \eta)$  then we consider the transformation

$$\begin{aligned} w &= a(\xi + \eta) \\ \tau &= \eta \\ \xi &= \frac{w - a\tau}{a} \Rightarrow |J| = \frac{1}{a} \end{aligned}$$

We know the joint density of  $w$  and  $\eta$

$$\begin{aligned} h(w; \eta) &= f(\xi(w; \tau); \eta(w; \tau))|J| \\ &= \sqrt{\frac{2}{\pi a^2}} \exp\left[-\frac{w^2}{2a^2}\right] \\ & \quad w - \tau \geq 0 \\ & \quad \tau \geq 0 \end{aligned}$$

and therefore the marginal density of  $w = a(\xi + \eta)$ , is what we need, and is given by

$$\begin{aligned} h(w) &= \int_0^w \sqrt{\frac{2}{\pi a^2}} \exp\left[-\frac{w^2}{2a^2}\right] d\eta \\ &= \left[ w^2 \frac{\sqrt{2}}{\sqrt{\pi a^2}} e^{-\frac{1}{2}\frac{w^2}{a^2}} \right] \sqrt{\frac{2}{\pi}} (\xi + \eta)^2 e^{-\frac{1}{2}(\xi + \eta)^2} \end{aligned} \quad (14)$$

3. To prove the consistency of the test against stationary alternatives satisfying the Berman condition, we invoke Lemmas 9 and 10, following which  $E\{J_0^{(n)}\} \sim \text{Var}\{J_0^{(n)}\} \sim O(n^{-1/2} \log n) \rightarrow 0$  as  $n \rightarrow \infty$ . Finally, we apply Lemma 11 to obtain:  $J_0^{(n)} \xrightarrow{P} 0$ .

#### ACKNOWLEDGEMENTS

The last two authors are especially grateful to Felipe Aparicio who recently passed away. We are also grateful to the editor, to an anonymous referee and to Daniel Peña, Juan Romo, Esther Ruiz from Universidad Carlos III de Madrid and to Andreu Sansó from the Universidad de Baleares for their comments and suggestions. Spanish CYCIT under Grant Spanish MCYT BEC2002 00279, BEC2002 03769 the Secretaria de Estado de universidades PR2003 0305. All errors are the entire responsibility of the authors.

1. Any time series with exponentially decaying covariances satisfies the Berman condition.
2. This theorem establishes that for a symmetric random walk  $x_t$ , the joint probability distributions of the random variables  $\{x_0, \dots, x_T\}$  and  $\{x'_0, \dots, x'_T\}$  are identical. Since the distribution of  $J_x^{(T-k)}$  and  $J_{x'}^{(T-k)}$  depend on the joint distributions of  $\{x_0, \dots, x_{T-k}\}$  and  $\{x'_0, \dots, x'_{T-k}\}$ , respectively, both  $J_x^{(T-k)}$  and  $J_{x'}^{(T-k)}$  must have the same distribution when the model errors  $\epsilon_j$  have a symmetric pdf.

Corresponding author: Alvaro Escribano, Departamento de Economía, Universidad Carlos III de Madrid. E mail: alvaroe@eco.uc3m.es

## REFERENCES

- AGIAKLOGLOU, C. and NEWBOLD, P. (1992) Empirical evidence on Dickey-Fuller type tests. *Journal of Time Series Analysis* 13, 471–83.
- APARICIO, F. M., ESCRIBANO, A. and GARCIA, A. (2004a) *A Range Unit Root Test*. Working Paper 04-11, University Carlos III of Madrid.
- APARICIO, F. M., ESCRIBANO, A. and GARCIA, A. (2004b) *Range Unit Root Tests: Improvements and Empirical Analysis*. Mimeo, University Carlos III of Madrid.
- APARICIO, F. M., ESCRIBANO, A. and GARCIA, A. (2005) Synchronicity Between Financial Time Series: An Exploratory Analysis. In *Progress in Financial Markets Research* (ed. C. KYSTOUS). New York Nova Publishers.
- ARNOLD, S. F. (1990) *Mathematical Statistics*. Prentice-Hall, NJ.
- ARRANZ, M. A. and ESCRIBANO, A. (2000) Cointegration testing under structural breaks: a robust extended error correction model. *Oxford Bulletin of Economic and Statistics* 62, 23–52.
- BILLINGSLEY, P. (1968) *Convergence of Probability Measures*. John Wiley, New York.
- BURRIDGE, P. and GUERRE, E. (1996) The limit distribution of level crossings of a random walk. *Econometric Theory* 12, 705–23.
- BUSTOS, O. H. and YOHAI, V. J. (1986) Robust estimates for ARIMA models. *Journal of the American Statistical Association* 81, 155–68.
- DICKEY, A. and FULLER, W. A. (1979) Distribution of the estimators for AR time series with a unit root. *Journal of the American Statistical Association* 74, 427–31.
- EMBRECHTS, P., KLÜPPELBERG, C. and MIKÖSCH, T. (1999) *Modelling Extremal Events (for Insurance and Finance)*. Springer-Verlag, Heidelberg.
- ERMINI, L. and GRANGER, C. W. J. (1993) Generalizations on the algebra of I(1) processes. *Journal of Econometrics* 58, 369–84.
- FELLER, W. (1971) *An Introduction to Probability Theory and its Applications* (vol. 2). John Wiley and Sons, New-York.
- FRANSES, P. H. and HALDRUP, N. (1994) The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economic Statistics*. 12, 471–8.
- GRANGER, C. W. J. and HALLMAN, J. J. (1991) Nonlinear transformations of integrated time series. *Journal of Time Series Analysis* 12, 207–24.
- HERRNDORF, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12, 141–53.
- KARATZAS, I. and SHREVE, S. E. (1991) *Brownian Motion and Stochastic Calculus*. Springer, New-York.
- LEADBETTER, M. R. and ROOTZÉN, H. (1988) Extremal theory for stochastic processes. *The Annals of Probability* 16, 431–78.
- LÉVY, P. (1948) *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, Paris.

- LINDGREN, G. and ROOTZÉN, H. (1987) Extremal values: theory and technical applications. *Scandinavian Journal of Statistics* 14, 241–79.
- MACKINNON, J. G. (1994) Approximate asymptotic distribution functions for unit root and cointegration tests. *Econometrica* 60, 145–57.
- MADDALA, G. S. and KIM, I.-M. (1998) *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge.
- MALLIAROPULOS, D. (2000) A note on nonstationarity, structural breaks and the Fisher effect. *Journal of Banking and Finance* 24, 695–797.
- MARTIN, R. D. and YOHAI, V. J. (1986) Influence functionals of time series. *The Annals of Statistics* 14, 781–818.
- NELSON, C. R. and PLOSSER, C. I. (1982) Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* 10, 139–62.
- NG, S. and PERRON, P. (2001) Lag length selection and the construction of Unit-root tests with good size and power. *Econometrica* 69, 1519–54.
- PEÑA, D. (2001) Outliers, influential observations and missing data. In: *A course in time series* (eds D. PEÑA, G. C. TIAO and R. S. TSAI). New York: John Wiley.
- PERRON, P. (1989) The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–401.
- PERRON, P. (1990) Testing for a unit-root in a time series with a changing mean. *Journal of Business and Economic Statistics* 8, 153–62.
- PERRON, P. and VOLGELSANG, T. J. (1992) Nonstationary time series and level shifts with an application to purchasing power parity. *Journal of Business and Economic Statistics* 10, 301–20.
- PERRON, P. and RODRIGUEZ, G. (2000) *Searching for Additive Outliers in Nonstationary Time Series*. Working Paper 0005E, Department of Economics, University of Ottawa.
- PHILLIPS, P. C. B. and PERRON, P. (1988) Testing for a unit-root in time series regression. *Biometrika* 75, 335–46.
- POLITIS, D.N., ROMANO, J.B. and WOLF, M. (1997) Subsampling for heteroskedastic time series. *Journal of Econometrics* 81, 281–317.
- PORT, S. (1994) *Mathematical Probability for Applications*. John Wiley & Sons, New York.
- REVUZ, D. and YOR, M. (1991) *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin.
- SCHWERT, G.W. (1989) Test for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–59.
- VOLGELSANG, T. J. (1990) Testing for a unit-root in a time series with a changing mean. *Journal of Business and Economic Statistics* 8, 153–62.