Range Unit Root (RUR) Tests: Robust against Nonlinearities, Error Distributions, Structural Breaks and Outliers

Felipe Aparicio
Departamento de Estadística
Universidad Carlos III de Madrid
Alvaro Escribano
alvaroe@eco.uc3m.es
Departamento de Economía
Universidad Carlos III de Madrid
Corresponding author
Ana E. Sipols
Departamento de Estadística e Investigación Operativa
Universidad Rey Juan Carlos
anaelizabeth.garcia@urjc.es

Abstract

Since the seminal paper by Dickey and Fuller in 1979, unit-root tests have conditioned the standard approaches to analyzing time series with strong serial dependence in the mean behavior, 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 (DF) unit root test on exchange rate series.

Key Words: Unit Roots Tests, Structural Breaks, Nonlinearities, Additive Outliers, Running Ranges and 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 Process (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 hereafter) 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 DGP's 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, fool standard unit root tests.

Many economic and financial time series such as inflation, unemployment rate, 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, and Malliaropulos, 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), Vogelsang (1990) and Perron and Vogelsang (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 fool 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 behavior is really $I(1)$, or rather a monotonically 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 problems for the 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 behavior 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 AO's 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 AO's 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 AO's leads to a downward bias of the 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 AO's was quantified by Franses and Haldrup (1994).

In this paper, we introduce a nonparametric *Range Unit Root* (RUR hereafter) test whose superiority with respect to the standard approaches is remarkable, see Aparicio, Escribano and García (2004a,b) . First, it is invariant to monotonic transformations and to the distribution of the model errors. Second it is robust against many structural breaks, parameter shifts and certain additive outliers. Third, 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, we discuss its small sample behavior under the null hypothesis of a single unit root, we derive the asymptotic null distribution of the test. Section 3 studies its power performances and its consistency against both stationary, integrated and trending alternatives. Section 4 analyses robustness of the test statistic under different departures from the standard unit root tests' 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 corrupts 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 (RUR) test

Many time series not generated by unit-root models exhibit similar mean behavior 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 behavior 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 $x_{1,i} = \min\{x_1, \cdots, x_i\}$ and $x_{i,i} = \max\{x_1, \cdots, x_i\}$ 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 $i = 1, 2, 3, \cdots, n$, 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 $\mathbf{1}(.)$ 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 *i.i.d.* sequences of random variables we have (see for instance Embrechts, Klüppelberg and Mikosch, 1999):

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

This result still holds for stationary series satisfying the "*Berman condition*", which requires the covariance sequence of the series $\{c_i = Cov(x_t x_{t+i})\}_{i \geq 1}$ to decrease faster than $(\log i)^{-1}$, that is $c_i \log i \to 0$ as $i \to \infty$ (see Lindgren and H. Rootzén, 1987)[1]. As will be shown later in this paper, the frequency of new records for $I(1)$ time series decreases at the slower rate:

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

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

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

We remark that for the random walk, the sequence of running ranges escalates indefinitely, whereas it does not in the stationary cases, see Aparicio, Escribano and Garcia (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.

[1] Any time series with exponentially decaying covariances satisfies the "Berman condition".

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 behavior 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 will 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 $\{\epsilon_i\}_{i\geq 1}$ are a sequence of *i.i.d.* random variables having zero mean and variance σ_ϵ^2. The corresponding testing device will be referred to as the *Range Unit Root* (RUR hereafter) test.

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

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 while small for $I(0)$ time series. We will show also that the RUR test is robust to a number of departures from the null hypothesis (no size distortions).

2.2 Small-sample behavior under the null

Table 1 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 follows $\epsilon_t \sim Nid(0, 1)$.

$\alpha \mid n$	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

Table 1. Critical values of the unit root test

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

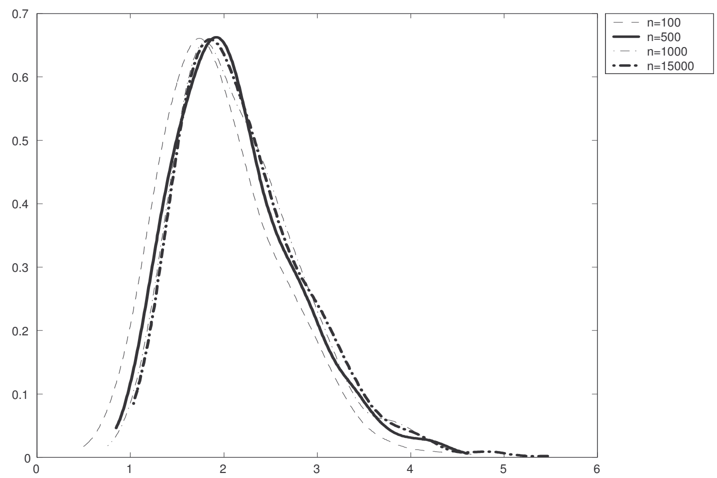

Figure 1. Plot of the empirical density of $J_0^{(n)}$, for different sample sizes, under the null hypothesis of a random walk with $Nid(0,1)$ errors.

2.3 Asymptotic distribution

A basic result regarding the behavior 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* $x_{t,t} = \max\{x_1, ..., x_t\}$ *and* $x_{1,t} = \min\{x_1, ..., x_t\}$. *Then we have*

$$\lim_{t\to\infty} P\left(x_t = x_{t,t}\right) = \lim_{t\to\infty} P(x_t = x_{1,t}) = 0. \tag{2}$$

Proof. In Appendix A1. ■

The 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 3 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 stationary, 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 $J_0^{(n)} \to 0$ as $n \to \infty$.

From Revuz and Yor (1991) we have the following definition of local time processes:

Definition 2 ***(Local Time of a Brownian Motion Process)*** *(Lévy, 1948) Let $B(.)$ represent a Brownian motion process in $\Re$, and let $l_B(x,t)$ be defined as*

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

$l_B(x,t)$ is a continuous increasing process in x called the local time of B at x. It measures the amount of time the Brownian motion spends in the neighborhood of x. It can also be interpreted as the "spatial density" of the occupation time $\int_0^t 1\left[x - \delta \leq B(s) \leq x + \delta\right] ds$.

Theorem 3 *Let $x_t = \sum_{i=1}^{t} \epsilon_i$ where $\{\epsilon_i\}_{i\geq 1}$ are continuous i.i.d. random variables with bounded pdf, zero mean and finite variance σ_ϵ^2. Suppose that x_0 has also a bounded pdf and finite variance. And let $J_0^{(n)} = J_1^{(n)} + J_2^{(n)}$ with $J_1^{(n)} = n^{-1/2} \sum_{t=1}^{n} \mathbf{1}(x_t = x_{t,t})$ and $J_2^{(n)} = n^{-1/2} \sum_{t=1}^{n} \mathbf{1}(x_t = x_{1,t})$. Then we have*

1.

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

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

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

$$h \geq 0, \quad i = 1, 2 \textit{ which depends on } a = \frac{E\left\{|\epsilon_1|\right\}}{\sigma_\epsilon} \tag{7}$$

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

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

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

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

$$J_0^{(n)} \xrightarrow{p} 0, \tag{8}$$

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

Proof. In Appendix A2. ■

In Figure 2 we plot the asymptotic distribution and the empirical distribution for a sample size of 1,000. It is clear that the critical values from both tails of the two distribution are similar as was shown in Table 1.

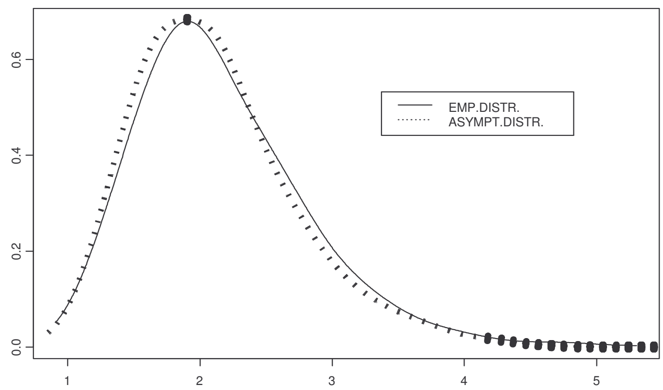

Figure 2. Plot of the empirical density of $J_0^{(n)}$ for $n = 1000$ under the null hypothesis $H_0:\ x_t = x_{t-1} + \epsilon_t$, where $\epsilon_t \sim Nid(0,1)$ and the Asymptotic distribution.

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 4 $D(u_n)$*: Let* $\{x_t\}_{t\geq 1}$ *be a stationary sequence of random variables with*

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

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

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

with $1 \leq i_1 < \ldots < i_p < j_1 < \ldots < j_q \leq n$, $j_1 - i_p \geq l$. *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* $\alpha_{n,l_n} \to 0$.

Among the processes that satisfy condition $D(u_n)$ are the Gaussian processes satisfying the so-called "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, $J_0^{(n)} \sim O(n^{-1/2}\log(n)) \to 0$ as $n \to \infty$, 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 ARMA processes. The Gaussian condition seems to be too restrictive; however, as we will see in the next section, 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 t-Student).

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:

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}\mathbf{1}(\Delta R_i^{(x)} &> 0) = O(n^{-1/2}\log n) \\ J_0^{(n)} &= O(n^{-1/2}\log n) \end{aligned}$$

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

$k > 0$ then $k+1$ will be the smallest positive integer such that $\widetilde{x}_t^{(k+1)} \sim I(1)$, or equivalently, such that $J_0^{(n)}$ does not vanish asymptotically. By mere inversion of the argument, if k is the smallest nonnegative integer such that the null hypothesis is not rejected on $\widetilde{x}_t^{(k)}$ 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 2 below using the estimated critical values at the 5% significance level, and from 10,000 replications of the alternative model $x_t = b\, x_{t-1} + \epsilon_t$, with $\epsilon_t \sim Nid(0,1)$, the DF test shown in parenthesis.

$n \mid b$	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)

Table 2. Empirical size and power of RUR test from 10,000 replications for different sample sizes n and for different values of b. (DF in parenthesis)

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 as compared to 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 σ_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 $\mu_\epsilon = E(\epsilon_t) \neq 0$, the renewal counting process of records $N(n) = \sum_{i=1}^n \mathbf{1}(\Delta R_i^{(x)} > 0)$ satisfies:

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

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

A similar divergent behavior 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:

$$a) \quad x_t = x_{t-1} + \epsilon_t \quad \text{with } E(\epsilon_t) = \mu_\epsilon \neq 0.$$

$$b) \quad x_t = y_t + \mu t \quad \text{where } y_t \sim I(0).$$

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

4 Departures from the standard conditions under $\mathbf{H}_0$

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 behavior of the RUR test in the presence of each of the above-mentioned departures from the standard unit-root tests assumptions.

4.1 Robustness of RUR test statistic against alternative error distributions

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

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

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:

$$y_t = x_t - x_{t,t} \tag{9}$$

$$y_t' = x_t - x_{1,t}. \tag{10}$$

The $I(1)$ nature of y_t and y_t' allows the application of a result by Burridge and Guerre (1996) for the asymptotic distribution of the normalized number of level crossings of a random walk, and which leads straightforwardly to ours. The asymptotic distribution of $J_i^{(n)}$ depends on the innovations' distribution (in particular of their variance, σ_ϵ^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 ϵ_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 errors' variance, σ_ϵ^2. However, this case is rather exceptional since for all other common distribution the value of a is sensitive to its shape, or equivalently to the tails. This is shown below for some typical error distributions with shape parameter denoted by ν.

Probability Distribution of Model Errors $\{\epsilon_t\}_{t\geq 1}$	$a = \frac{E\{\lvert\epsilon\rvert\}}{\sigma_\epsilon}$
Student$-t$ with ν degrees of freedom	$\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma\left(\frac{1}{2}(\nu-1)\right)}{\Gamma(\frac{\nu}{2})}$
Log-Normal	$\frac{1}{\sqrt{\exp(\nu^2)-1}}$
Gamma	$\frac{\Gamma(c+1)}{\sqrt{c}\Gamma(c)}$
Weibull	$\frac{\Gamma(\frac{c+1}{c})}{\sqrt{\Gamma(\frac{c+2}{c})}}$

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, $a = \frac{E\{\lvert\epsilon\rvert\}}{\sigma_\epsilon}$.

The empirical size and power of the RUR test against stationary $AR(1)$ alternatives is shown in Table 3 below. We consider the estimated 5% critical values from $10,000$ replications of the model $x_t = b\, x_{t-1} + \epsilon_t$, with the following distributions: $\epsilon_t \sim Nid(0,1)$ when ϵ_t has a Student-t distribution with 5 degrees of freedom, when ϵ_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 ϵ_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 parenthesis.

$n \mid b$	0.5	0.8	0.9	0.95	0.99	1
			Cauhy			
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)
			t-student			
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)

Table 3. Empirical size and power of RUR test from $10,000$ replications for different sample sizes n, values of b and error distributions

The size of the RUR test is included in the last column of Table 3. 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 the power of the RUR test for stationary alternatives that are far from 1 (say 0.5 and 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 will consider the case of a single structural break in the series in the middle of the sample. The break is modeled 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 is $x_t = 0.5\ x_{t-1} + s\ D_t + \epsilon_t$. Table 4 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 parenthesis.

$n \mid s$	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 4. Empirical power against the alternative $x_t = 0.5\ x_{t-1} + s\ D_t + \epsilon_t$, for different values of the sample size n and of the local break size s

We remark that except for the case of $s = 4$ and $n = 500$, the Dickey-Fuller (DF) test has no power. The RUR test is more powerful for sample sizes larger 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 $x_t = 0.5\ x_{t-1} + s_1\ D_{t,1} + s_2 D_{t,2} + \epsilon_t$ 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 5 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 brackets. 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 $s_2 = 4, 8, 12$,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.

n |$\mathbf{s}_{1,2}$	(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)

Table 5. Empirical power of the RUR test against the alternative $x_t = 0.5\, x_{t-1} + s_1\, D_{t,1} + s_2 D_{t,2} + \epsilon_t$, for different values of the sample size n and of the local break size s

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

$$\begin{aligned} &a) \quad x_t = a x_{t-1} + \xi_t, \quad |a| < 1 \ \text{ with } E(\xi_t) = 0, \\ &b) \quad x_t = a x_{t-1} + \epsilon_t, \quad |a| < 1 \ \text{ with } E(\epsilon_t) = s\mathbf{1}(t = t_0). \end{aligned}$$

Let $J_\xi^{(n)}, J_\epsilon^{(n)}$ 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 $P(\Delta R_t > 0) = O(t^{-1}) \simeq 0$ for t large enough, that is $\Delta R_{t_0} = a$ with probability close to one. As a result, $J_\epsilon^{(n)} \simeq J_\xi^{(n)} + n^{-1/2}$ for both n and t_0 large enough. But then $J_\epsilon^{(n)} - J_\xi^{(n)} \xrightarrow{p} 0$ as $n \to \infty$, 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 $t^* \triangleq \min_{1 \le i \le m}\{t_i\} > 0$ and $0 < s^* \triangleq \min_{1 \le i \le m}\{s_i\} < \infty$, such that $P(\Delta R_{t^*} > 0) = O(t^{*-1}) \simeq 0$ and thereby $J_\epsilon^{(n)} \simeq J_\xi^{(n)} + n^{-1/2} \sum_{i=1}^m s_i$ for large enough n. Therefore $J_\epsilon^{(n)} - J_\xi^{(n)} \xrightarrow{p} 0$ as $n \to \infty$, 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:

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

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

with $t^* > 0$ and $0 < s^* < \infty$, such that $P(\Delta R_{t^*} > 0) = O(t^{*-1/2}) \simeq 0$ and thereby $J_\epsilon^{(n)} \simeq J_\xi^{(n)} + n^{-1/2} \sum_{i=1}^{m} s_i$. Then as far as $m = o(n^{1/2})$ we will get $J_\epsilon^{(n)} - J_\xi^{(n)} \xrightarrow{p} 0$ as $n \to \infty$, 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, $J_\epsilon^{(n)} \to \infty$ as $n \to \infty$. Level breaks will then shift indefinitely the null distribution to the left leading to the rejection of the null hypothesis of an $I(1)$ time series.

Next we analyze the small sample behavior 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 6 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 parenthesis.

Transformations	100	250	500
Monotonic			
1) x_t^2, with $x_t > 0, \forall t$	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) $exp(\frac{x_t}{75})$	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)
Non monotonic			
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)

Table 6. Empirical size of the RUR test against different forms of nonlinearity applied to a random walk $x_t = x_{t-1} + \epsilon_t$

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 linearly the series 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 $g(.)$ applied to the series x_t follows immediately from the relations:

$$\mathbf{1}\left(g(x_t) > g(x_{t-1,t-1})\right) = \mathbf{1}(x_t > x_{t-1,t-1})$$
$$\mathbf{1}\left(g(x_t) < g(x_{1,t-1})\right) = \mathbf{1}(x_t < x_{1,t-1}).$$

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 7 show that the size distortions caused by the presence of additive outliers (AO) in the middle of the series and beyond are significantly smaller for the RUR test than for the DF test (shown in parenthesis). Our alternative hypothesis was now represented by the model $y_t = x_t + s\delta_{t,\tau}$ where $x_t = x_{t-1} + \epsilon_t$, τ 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 τ ($\tau = n/25, n/10, n/5$) 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.

$n \mid \boldsymbol{\tau}$	$(n/2)$	$(n/2)+1$	$(n/2)+2$
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)
$n \mid \boldsymbol{\tau}$	$n-(n/20)$	$n-(n/10)$	$n-(n/5)$
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)
$n \mid \boldsymbol{\tau}$	$n/25$	$n/10$	$n/5$
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)

Table 7. 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

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 flavor 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 $\Delta R_{t_0}^{(x)} \geq \max_{1 \leq t \leq n} \Delta R_t^{(x)}$. 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

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

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 $n \to \infty$.

The relatively large size distortion of the RUR test in the presence of early AO's 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 the next section.

5 The forward-backward range unit root (FB-RUR) 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 to time series with AO's will be uncertain. By means of a simple resampling technique, we obtain an extension of the RUR test, called the *Forward-Backward RUR (FB-RUR) Test*, based on the statistic here noted as $J_*^{(n)}$, which reduces the size distortion when the additive outlier (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:

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

where $x'_t = x_{n-t+1}$, $t = 1, 2, ..., T$, 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

$$x'_t = x_n - \sum_{i=t+1}^{n} \epsilon_i,$$
$$x_{t'} = x_0 + \sum_{i=1}^{t'} \epsilon_i,$$

we obtain for $t' \leq t$:

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

Now suppose x_0 is fixed and the ϵ_i are Gaussian $Cov(x_0, x_n) = 0$. 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 γ 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

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

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)[2]. Therefore $J_*^{(n)} \Rightarrow \frac{1}{\sqrt{2n}} J_0^{(\infty)}$,

[2] This theorem establishes that for a symmetric random walk x_t the joint probability distributions of the random variables $\{x_0, ..., x_T\}$ and $\left\{x'_0, ..., x'_T\right\}$ are identical. Since the distribution of $J_x^{(T-k)}$ and $J_{x'}^{(T-k)}$ depend on the joint distributions of $\{x_0, ..., x_{T-k}\}$ and $\left\{x'_0, ..., x'_{T-k}\right\}$, respectively, both $J_x^{(T-k)}$ and $J_{x'}^{(T-k)}$ must have the same distribution when the model errors ϵ_i have a symmetric *pdf*.

where

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

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

The power of the FB-RUR test against the alternative of a stationary $AR(1)$ time series with $Nid(0,1)$ model errors is shown in Table 8. Notice the improvements in power performances, especially for the smaller sample sizes, where now DF (in parenthesis) 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).

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)

Table 8. Empirical size and power from 10000 replications for different sample sizes n and for different values of b

5.1 Behavior 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 9 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 brackets. See also Franses and Haldrup (1994).

$n\ \vert\tau$	$n/25$	$n/10$	$n/5$
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)
$n\ \vert\tau$	$n/2$	$n/2+1$	$n/2+2$
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)
$n\vert\tau$	$n-n/20$	$n-n/10$	$n-n/5$
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)

Table 9. Empirical size against the model $y_t = x_t + s\delta_{t,\tau}$ and different locations of the AO

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 AO's 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.

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

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

where the *iid* 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 $J_y^{(t_0)} \sim (\dfrac{t_0}{n})^{1/2} \to 0$, as $n \to \infty$. 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

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

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 only affects the asymptotic distribution of $J_*^{(n)}$ by a factor of $\frac{1}{\sqrt{2}}$. 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 asymptotic as the one based on $J_0^{(n)}$.

6 Further analysis of size distortions with serial correlation and heteroskedastic 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

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

Table 13 presents the rejection frequencies of the ADF where the lags are selected according to Ng and Perron (2001) using MAIC criteria. The rejection frequencies are based on $5,000$ 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).

β		ADF$_{MAIC}$	RUR	FB-RUR
	$b = 1$ (size)			
-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
	$b = 0.8$ (power)	ADF$_{MAIC}$	RUR	FB-RUR
-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

Table 13. Empirical size and 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) outperforme the rest of the tests.

Finally we will also consider the size of the RUR and FB-RUR in two cases of heteroskedasticity: nonstationary autoregressive $AR(1)$ processes with heteroskedastic errors, following Politis et al. (1997), and a nonstationary autoregressive $AR(1)$ processes with $GARCH(1,1)$ 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 anual US/Finland real exchange rates 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 Vogelsang (1990), Franses and Haldrup (1994), Perron and Vogelsang (1992), and Perron and Rodriguez (2000), point to the presence of an AO in 1918 together with IO's that produce temporary changes in 1917, 1932, 1949 and 1957.

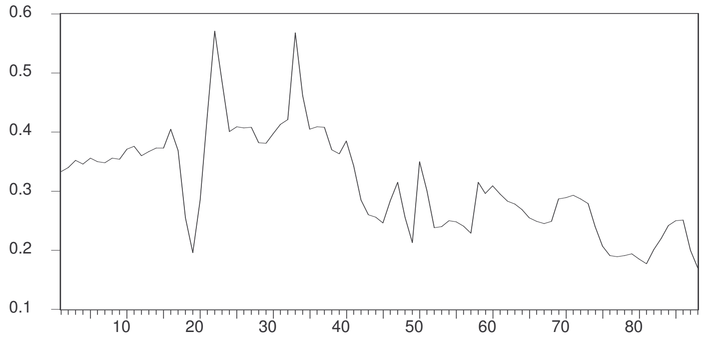

Figure 3. Logarithm of the US/Finland real exchange rates deflated annual series from 1900 to 1987.

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. In 1996 Burridge and Guerre 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 an nonparametric testing device, called the *Range Unit Root* (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 comes from its sensitivity to early additive outliers, which may lead to a size distortion comparable to DF's. 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.

9 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 5 ***(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 $B(.)$ *is a Brownian motion process on the interval [0,1],* σ *represents the long-run, and "*$\Rightarrow$*" denotes convergence in distribution as* $n \to \infty$*.*

Proof. See Herrndorf (1984) ■

Lemma 6 ***(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* $x_n(r)$ *defined as in Lemma 5. Then*

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

Proof. See Billingsley (1968). ■

Lemma 7 *Let $S_t = \sup_{s\in[0,1]}\{B(s)\}$, $I_t = \inf_{s\in[0,1]}\{B(s)\}$ and $x_n(r)$ defined as in Lemma 5. Under the mixing-condition of Phillips and Perron (1988) we have:*

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

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

Proof. The proof follows from the CMT (Lemma 6) and the continuity of the functions T_1 and T_2. ■

Lemma 8 *(Lévy, 1948). Let $\{B(r)\}_{r\in[0,1]}$ represent a Brownian motion process on the interval [0,1], and let $\widetilde{B}_1(r) = B(r) - S_t$ and $\widetilde{B}_2(r) = B(r) - I_t$. The processes $\{|B(r)|\}_{r\in[0,1]}$, $\left\{\widetilde{B}_1(r)\right\}_{r\in[0,1]}$ and $\left\{\widetilde{B}_2(r)\right\}_{r\in[0,1]}$ have the same probability distribution.*

Proof. See Karatzas and Shreve (1988). ■

Lemma 9 *The dimensional processes $(S_t - B(r), S_t)$, $(|B_t|, l_B(0,r))$ and $\left(|B(r)|, \frac{1}{2}l_{|B|}(0,r)\right)$ have the same law.*

Proof. See Revuz and Yor (1998) page 240 and 244. ■

Lemma 10 *The joint law of $(S_t - B(r), S_t)$ has density*

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

for $a, b \geq 0$.

Proof. See Revuz and Yor (1998) page 245. ■

Lemma 11 *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 σ_ϵ^2, and let*

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

denote the normalized number of crossings of level b. If x_0 and ϵ_1 have bounded $pdf's$ 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 Burridge and Guerre (1996). ■

Lemma 12 *(Lévy, 1948) Let Z be a standard Normal random variable and let*

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

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

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

Proof. See Theorem 2.3 in Revuz and Yor (1991). ■

Lemma 13 *Let $\{x_t\}_{t\geq 1}$ be a stationary Gaussian sequence with covariances $\{c_i\}_{i\geq 1}$ satisfying the "Berman condition": $c_i \log i \to 0$ as $n \to \infty$. 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 14 *If $\{x_t\}_{t\geq 1}$ is a sequence of i.i.d. random variables then for large n*

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

Proof. See for instance Port (1994). ■

Lemma 15 *Let $\{\xi_i\}_{i\geq 1}$ a sequence of random variables such that $lim_{i\to\infty} E(\xi_i) = \mu$, and $lim_{i\to\infty} Var(\xi_i) = 0$. Then*

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

Proof. See for instance Arnold (1990). ■

Lemma 16 *Let $x_i = x_{i-1} + \epsilon_i$ where $\{\epsilon_i\}_{i \geq 1}$ are continuous i.i.d. random variables with finite variance σ_ϵ^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 $u \in [0,1]$:*

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

Proof. See levy (1948). ■

A1. Proof of Proposition 1

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

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

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

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

Now defining $r = t/n$, where $t = 1, 2, ..., n$, and letting $n \to \infty$ we obtain from direct application of lemmas 5, 6 and 7:

$$\begin{aligned}
\Psi_1^{(n)} &\Rightarrow \int_0^1 \mathbf{1} \left[B(r) - \sup_{s \in [0,1]} \{B(s)\} = 0 \right] dr \\
\Psi_2^{(n)} &\Rightarrow \int_0^1 \mathbf{1} \left[B(r) - \inf_{s \in [0,1]} \{B(s)\} = 0 \right] dr
\end{aligned}$$

finally, it follows from lemma 8 and from the definition of local time that

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

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

A2. Proof of Theorem 3

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 σ_ϵ^2. Let $y_t = x_t - x_{t,t}$ 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

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

where we have used the fact that

$$\begin{aligned}
\sum_{t=1}^{n} \mathbf{1}(y_{t-1} \geq 0, y_t = 0) &= o\left(\sum_{t=1}^{n} \mathbf{1}(y_{t-1} < 0, y_t = 0)\right)\\
\sum_{t=1}^{n} \mathbf{1}(y'_{t-1} \leq 0, y'_t = 0) &= o\left(\sum_{t=1}^{n} \mathbf{1}(y'_{t-1} > 0, y'_t = 0)\right).
\end{aligned}$$

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 asymptotically random walks. By symmetry, the behavior 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 $x_{t,t} = x_{t'}$ $(t' \in [1, n] \cap Z)$ is

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

From lemma 16, 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:

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

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 η_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 $\lim_{t\to\infty} P(x_t \geq x_{t-1,t-1}) = 0$. It follows that $\eta_t = \epsilon_t$ with probability $p_t = P(x_t < x_{t-1,t-1}) \to 1$. In particular:

$$\begin{aligned}
E(\eta_t) &= 0 \\
var(\eta_t|x_{t'} = x_{t,t}) &= \sigma_\epsilon^2 \text{ with probability } p_t \to 1 \\
var(\eta_t|x_{t'} = x_{t,t}) &= (t - t' - 1)\sigma_\epsilon^2 \text{ with probability } 1 - p_t \to 0,
\end{aligned}$$

from where the unconditional variance of η_t is obtained:

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

Since in practice it can be assumed that the process x_t was generated at $t = -\infty$, we conclude that η_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 $y_{t-1} \neq 0$ (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 η_t can be less than ϵ_t, implying a transitory short-memory behavior for y_t. Notice however that as $t \to \infty$ we get $E(a_t|y_{t-1}) \to 1$, 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 13 that

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

where Z is a standard Normal random variable. From lemma 11 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

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

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

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

2. From the lemmas 9 and 10 we know that joint law of $(|B(1)|, l(0,1))$ has density

$$f(\xi, \eta) = \sqrt{\frac{2}{\pi}} (\xi + \eta) \exp\left(-(\xi + \eta)^2/2\right)$$

for $\xi, \eta \geq 0$.

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

$$\begin{aligned} w &= a(\xi+\eta) \\ \tau &= \eta \end{aligned}$$

$$\begin{aligned} \xi &= \tfrac{w-a\tau}{a} \\ \eta &= \tau \end{aligned} \quad \Longrightarrow |J| = \frac{1}{a}$$

We know the joint density of w and η

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

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

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

3. To prove the consistency of the test against stationary alternatives satisfying the Berman condition we invoke lemmas 13 and 14, following which $E\left\{J_0^{(n)}\right\} \sim Var\{J_0^{(n)}\} \sim O(n^{-1/2}\log n) \rightarrow 0$ as $n \rightarrow \infty$. finally, we apply lemma 15 to obtain: $J_0^{(n)} \xrightarrow{p} 0$

A3. Proof of Proposition 6

Letting $x_t = w_t + s_t$, the proof is a straight consequence of the fact that

$$n^{-1/2}s_t \xrightarrow{p} 0, \quad \text{as } n \rightarrow \infty$$

Now since

$$\begin{aligned} R_t^{(x)} &= x_{t,t} - x_{1,t} \\ &\leq w_{t,t} + s_{t,t} - w_{1,t} - s_{1,t}, \end{aligned}$$

we obtain

$$\begin{aligned}\mathbf{1}(R_t^{(x)} > 0) &= \mathbf{1}(\sigma_\epsilon^{-1} n^{-1/2} R_t^{(x)} > 0) \\ &= \mathbf{1}(\sigma_\epsilon^{-1} n^{-1/2} R_t^{(w)} + \sigma_\epsilon^{-1} n^{-1/2} R_t^{(s)} > 0).\end{aligned}$$

Thus

$$\begin{aligned}J_0^{(n)} &= n^{-1/2} \sum_{t=1}^{n} \mathbf{1}(\sigma_\epsilon^{-1} n^{-1/2} R_t^{(w)} > -\sigma_\epsilon^{-1} n^{-1/2} R_t^{(s)}) \\ &\sim n^{-1/2} \sum_{t=1}^{n} \mathbf{1}(\sigma_\epsilon^{-1} n^{-1/2} R_t^{(w)} > 0) \quad \text{for large enough } n\end{aligned}$$

since for $0 < t \leq n \quad R_t^{(s)} \leq R_n^{(s)} = o(n^{-1/2})$.

References

[1] Aparicio F.M., Escribano A. and Garcia A.(2004a), "A Range Unit Root Test", Working paper 04-11, University Carlos III of Madrid.

[2] Aparicio F.M., Escribano A. and Garcia A.(2004b), "Range Unit Root Tests: Improvements and Empirical Analysis", mimeo, University Carlos III of Madrid.

[3] Aparicio F.M., A. Escribano and A. Garcia (2005), "Synchronicity Between Financial Time Series: An Exploratory Analysis", Kluwer Academic Press (forthcoming).

[4] Arranz M.A. and A. Escribano (2000), "Cointegration Testing under Structural Breaks: A Robust Extended Error Correction Model", Oxford Bulletin of Economic and Statistics, vol. 62, 23-52.

[5] S. F. Arnold (1990), Mathematical Statistics, Prentice-Hall, New-Jersey, 1990.

[6] P. Billingsley (1968), "Convergence of Probability Measures", John Wiley, New York.

[7] P. Burridge and E. Guerre (1996), "The Limit Distribution of Level Crossings of a Random Walk", Econometric Theory, 12, 705-723.

[8] O.H. Bustos and V.J. Yohai (1986), "Robust Estimates for ARIMA Models", Journal of the American Statistical Association, 81, 155-168.

[9] A. Dickey and W.A. Fuller (1979), "Distribution of the Estimators for AR Time Series with a Unit Root", Journal of the American Statistical Association, 74, 427-431.

[10] P. Embrechts, C. Klüppelberg and T. Mikösch (1999), Modelling Extremal Events (for Insurance and finance), Springer-Verlag, Heidelberg.

[11] L. Ermini and C.W.J. Granger (1993), "Generalizations on the Algebra of I(1) Processes", Journal of Econometrics, 58, 369-384.

[12] P.H. Franses and N. Haldrup (1994), "The Effects of Additive Outliers on Tests for Unit Roots and Cointegration", Journal of Business and Economic Statistics, 12, 4, 471-478.

[13] W. Feller (1971), "An Introduction to Probability Theory and its Applications" (vol. 2), John Wiley and Sons, New-York.

[14] C.W.J. Granger and Halmman J.J. (1991), "Nonlinear Transformations of Integrated Time Series", Journal of Time Series Analysis, 12, 3, 207-224.

[15] N. Herrndorf (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", Annals of Probability, 12, 141-153.

[16] I. Karatzas and S.E. Shreve (1991), Brownian Motion and Stochastic Calculus, Springer, New-York.

[17] M.R. Leadbetter and H. Rootzén (1988), "Extremal Theory for Stochastic Processes", The Annals of Probability, 16, 2, 431-478.

[18] P. Lévy (1948), Processus Stochastiques et Mouvement Brownien, Gauthier-Villars, Paris.

[19] G. Lindgren and H. Rootzén (1987), "Extremal Values: Theory and Technical Applications", Scandinavian Journal of Statistics, 14, 241-279.

[20] J.G. Mackinnon (1994), "Approximate Asymptotic Distribution Functions for Unit Root and Cointegration tests", Econometrica, 60, 1, 145-157.

[21] G.S. Maddala and In-Moo Kim (1998), "Unit Roots, Cointegration and Structural Change", Cambridge University Press, Cambridge.

[22] D. Maliiaropulos (2000), "A Note on Nonstationarity, Structural Breaks and the fisher Effect", Journal of Banking and finance, 24, 695-797.

[23] R.D. Martin and V.J. Yohai (1986), "Influence Functionals of Time Series", The Annals of Statistics, 14, 781-818.

[24] C.R. Nelson and C.I. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", Journal of Monetary Economics, 10, 139-162.

[25] D. Peña (2001), "Outliers, Influential Observations and Missing Data" in A Course in Time Series, eds. D. Peña, G.C. Tiao and R.S. Tsay, New York: John Wiley.

[26] P. Perron (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", Econometrica, 57, 1361-1401.

[27] P. Perron (1990), "Testing for a Unit Root in a Time Series with a Changing Mean", Journal of Business and Economic Statistics, 8, 70, 153-162.

[28] P. Perron and T.J. Volgelsang (1992), "Nonstationary Time Series and Level Shifts with an Application to Purchasing Power Parity", Journal of Business and Economic Statistics, 10, 3, 301-320.

[29] P. Perron and G. Rodriguez (2000), "Searching for Additive Outliers in Nonstationary Time Series". Working Paper 0005E, Department of Economics, University of Otawa.

[30] P.C.B. Phillips and P. Perron (1988), "Testing for a Unit Root in Time Series Regression", Biometrika, 75, 335-346.

[31] S. Port (1994), "Mathematical Probability for Applications", John Wiley & Sons, New-York.

[32] D. Revuz and M. Yor (1991), Continuous Martingales and Brownian Motion, Springer-Verlag, Berlin.

[33] NG. Serena and P. Perron (2001), "Lag length selection and the construction of Unit Root Tests with good Size and Power", Econometrica, 69, 1519-1554.

[34] T.J. Vogelsang (1990), "Testing for a Unit Root in a Time Series with a Changing Mean", Journal of Business and Economic Statistics, 8, 70, 153-162.

Aknowledgments

The last two authors are especially grateful to Felipe Aparicio who recently passed away. We are also grateful to the editor, to an anonmous 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 responsability of the authors.