A simple unit root test consistent against any stationary alternative

Frédérique Bec, Alain Guay

A simple unit root test consistent against any stationary alternative*

Frédérique Bec† and Alain Guay‡

Abstract

This paper proposes $t-$like unit root tests which are consistent against any stationary alternatives, nonlinear or noncausal ones included. It departs from existing tests in that it uses an unbounded grid set including all possible values taken by the series. In our setup, thanks to the very simple nonlinear stationary alternative specification and the particular choice of the thresholds set, the proposed unit root test contains the standard ADF test as a special case. This, in turn, yields a sufficient condition for consistency against any ergodic stationary alternative. From a Monte-Carlo study, it turns out that the power of our unbounded non adaptive tests, in their average and exponential versions, outperforms existing bounded tests, either adaptive or not. This is illustrated by an application to interest rate spread data.

Keywords: Unit root test, Threshold autoregressive model, Interest rate spread.

JEL classification: C12, C22, C32, E43.

1 Introduction

In the existing literature about unit root testing against non-linear threshold alternatives, what differs across the proposed test statistics — beside the non-linear alternative model specification — is basically the choice of the grid set of thresholds. To fix ideas, mainly two categories may be distinguished: the bounded and unbounded sets of thresholds. Unlike the former, the boundaries increase with the sample size for the latter. Within these two categories, two more classes can be separated: the adaptive and not adaptive sets of thresholds. In this paper, we depart from this existing work by considering an unbounded not adaptive set of thresholds which has the particularity to choose its lower bound so as to make the test statistics match exactly the ADF one. Thanks to this characteristic, it is consistent against any stationary process, either linear or not, causal or non-causal. This desirable feature is also found by Bec, Guay and Guerre [2008a]

*The authors would like to thank Heino Bohn Nielsen for fruitful discussions at the early stage of this work.
†THEMA, CY Cergy Paris Université and CREST, Paris.
‡Université du Québec à Montréal and Chaire en macroéconomie et prévisions ESG-UQAM.

in a more sophisticated framework. Here, this is made possible by fixing all the parameters of the inner regime of a three-regime self-exciting threshold autoregression (SETAR) so that it is characterized by a unit root process.

Why is this consistency property desirable? The test statistics developed here keeps good power performance against models which are more difficult to estimate than our auxiliary simplified SETAR model. Indeed, models such as e.g. Smooth transition Threshold AutoRegression, Autoregressive Conditional Root, Markov switching or non-causal models involve time-consuming and tedious maximum likelihood techniques which are not necessary to estimate our auxiliary model. Yet, the econometric theory underlying these models has been developed under the maintained hypothesis of stationarity, but they are typically used to fit such variables as e.g. the inflation rate, the real exchange rate or the crude oil price level which might reasonably be suspected to be (close to) non stationary. As a consequence, there is a strong need for a powerful unit root test before using these models.

Why is our test much more simple and powerful? First, it retains the idea developed in Kapetanios and Shin [2006] — and before them by Balke and Fomby [1997], Michael, Nobay and Peel [1997], Kapetanios, Shin and Snell [2003] — which imposes a unit root in the inner regime of a three-regime SETAR model. As stressed by Kapetanios and Shin [2006], by imposing a unit root in the inner regime, a test should gain power compared to the joint hypothesis Wald test proposed in e.g. Bec, Ben Salem and Carrasco [2004] or Bec et al. [2008a]. Second, it is simplified even more by considering symmetrical thresholds in the lower and upper regimes. By imposing symmetrical thresholds and a lower bound of the set of thresholds which makes the auxiliary model shrink to a standard autoregression, the test statistic obviously amounts to the ADF one for this lower bound. Thirdly, by allowing a mirroring intercept in the outer regimes under the stationary alternative, it does not require to de-mean the series in a first step, as it can accommodate non centered series. This should also contribute to increase our test's power. Finally, this skeleton SETAR auxiliary model does not call for time-consuming ML or EM estimation techniques since it can be estimated by piecewise OLS method.

The paper is organized as follows. We first introduce the new unit root test statistics and derive its asymptotic behaviour. Then, Section 2 presents simulation experiments results to compare its power to other classes of this kind of test. Section 3 illustrates the gain from this new approach using the same interest rate spread data as the ones used in Bec et al. [2008a]. Section 4 concludes.

2 The consistent unbounded, not adaptive unit root test statistics

As in Bec et al. [2008a], we are interested in testing the random walk null hypothesis:

$H_0 : \Delta y_t = a(L)\Delta y_{t-1} + \varepsilon_t,$

where $y_0 = \ldots = y_{-p-1} = 0$ and $\{\varepsilon_t\}$ is a sequence of i.i.d. centered random variables with variance σ^2 and $1 - a(L)$ is of known order $p \geq 0$ with roots outside the unit circle. Assume that $T + 1$ observations $y_0, \ldots, y_T$ are available to test H_0 against

$H_1 : \{y_t\}$ is a non constant stationary ergodic process with a finite non vanishing variance.

The auxiliary test regression we propose to use for our unit root test is restricted to a symmetric mirroring 3-regime dynamic TAR specification:

$$\Delta y_t = u_t + a(L)\Delta y_{t-1} + \begin{cases} -\mu + \rho y_{t-1} & \text{if } y_{t-1} \leq -\lambda, \\ \mu + \rho y_{t-1} & \text{if } y_{t-1} \geq \lambda \end{cases} \tag{1}$$

where $\{u_t\}$ is a sequence of i.i.d. centered random variables with variance σ_u^2 and $a(L)$ is a lag polynomial of order p. Note that in the middle, or inner regime, a random walk (without drift) behavior is imposed, so that no specific parameter has to be estimated here. Beyond this restriction, this auxiliary model is further simplified by assuming that *i*) the threshold defining the lower regime $(-\lambda)$ is just the opposite of the one defining the upper regime (λ) and *ii*) the intercept is symmetrical across regimes.[1] We focus on testing the null $H_0 : \rho = 0$ against the stationary alternative $H_1 : |\rho| < 0$, using the following Student-t type Infinimum statistics:

$$t_T^{inf}(\Lambda_T) = \inf_{\lambda \in \Lambda_T} t_T(\lambda), \tag{2}$$

with

$$t_T(\lambda) = \hat{\rho}_T(\lambda)/s.e.(\hat{\rho}_T(\lambda))$$

where $\hat{\rho}_T$ is the piecewise ordinary linear least squares estimate of ρ and $s.e.(\hat{\rho}_T)$ its standard error. A Wald-type test statistic may be obtained via a similar route, as in Bec et al. [2008a] for instance. The advantage of the t-test over the Wald-test is that the t-test deals with one-sided stationary alternatives explicitly, and thus is expected to be more powerful. Moreover, beside the auxiliary model under consideration, what distinguishes existing statistics of this kind is

[1] Note that this model does not generate zero-mean processes, so that unlike Kapetanios and Shin [2006], this framework does not require to de-mean the data before proceeding to the test.

basically the choice of the set of thresholds Λ_T, as emphasized in the Introduction. Bec et al. [2008a] develop both bounded and unbounded SupWald tests statistics for which the boundaries of Λ_T are adaptive, in the sense that Λ_T is larger under the alternative than under the null hypothesis. Bec et al. [2004]'s SupWald test statistic relies on an unbounded and not adaptive set of thresholds: it consists in all values of y_t lying between its 15^{th} and the 85^{th} quantiles, following Andrews [1993] and the common practice since this seminal paper. Kapetanios and Shin [2006] auxiliary test regression is similar to the one given in Equation (1) but it relaxes the symmetry assumption: the autoregressive parameter is allowed to differ across the upper and lower regimes.[2] So, their unit root null consists in setting both upper and lower autoregressive roots to zero. The corresponding Sup (or average or exponential) Wald test statistics they develop belong, by assumption, to the bounded, non adaptive class of thresholds sets.

The test statistics proposed here belongs to the same class as the one of Bec et al. [2004] in that it retains a grid set that is unbounded and non adaptive. However, it departs from the quantiles approach by considering a grid set including all the values taken by the series, with the exception of the few points needed to estimate the outer regime. Hence, the set of thresholds considered here is larger than the ones typically used in the quantile approach. More concretely, denoting $|y|_{(t)}$, $t = 0, \ldots, T-1$, the ordered $|y_{t-1}|$, it amounts to consider the first value $|y|_{(0)}$ as the lower bound and the $(T-1-k)^{th}$ value as the upper bound, where T is the sample size and k is the number of parameters to estimate in the outer regime, so that:

$$\Lambda_T = [|y|_{(0)}, |y|_{(T-1-k)}]. \tag{3}$$

Hence, the set of thresholds does not adapt its size to the null or the alternative under consideration. Then, it is of course unbounded as its span widens with the sample size.

The Theorem 1 below shows that choosing $|y|_{(0)}$ as the lower bound of Λ_T is sufficient to get consistency against any ergodic alternatives.

Theorem 1 *Consider the TAR specification (1). Assume that Λ_T is such that, for any $\{y_t\}$ in H_1, $\underline{\lambda}_T = |y|_{(0)}$ implies that $t_T(\underline{\lambda}_T) = ADF_T$. Then, under H_1, $\mathrm{t}^{\inf}(\Lambda_T)$ diverges in probability, with*

$$t_T^{inf}(\Lambda_T) \leq t_T(\underline{\lambda}_T) = ADF_T(1 + o_p(1)) \ . \tag{4}$$

This directly results from the fact that the statistic ADF_T diverges with exact order $\sqrt{T}$ in probability for any $\{y_t\}$ in H_1 and the TAR specification (1) is equivalent to the autoregressive

[2] Moreover, as they de-mean the series before proceeding to test, $\mu = 0$ in their setup.

linear model when the threshold is $\lambda_T = |y|_{(0)}$.[3] In this case, the lower and the central regimes vanish and the SETAR shrinks to the auxiliary model of the ADF test:

$$\Delta y_t = a(L)\Delta y_{t-1} + \mu + \rho y_{t-1} + u_t. \tag{5}$$

Importantly, it follows from Theorem 1 that this definition of the grid set of thresholds is a sufficient condition for consistency against *any* ergodic stationary alternative. Then, the inequality (4) indicates that the $t_T^{inf}(\Lambda_T)$ test can be more powerful than the ADF test provided its critical values are close enough to the critical values of the ADF_T statistic (see Table 2).

The next theorem shows that $t_T^{inf}(\Lambda_T)$ has a pivotal null distribution. Let us first define for the outer regime

$$\xi_O(\lambda) = \frac{\int_0^1 W(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))dW(v) - \frac{\int_0^1 W(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))dv}{\int_0^1 \mathbb{I}_{I_{out}(\lambda)}(W(v))dv}\int_0^1\left(\mathbb{I}_{I_\ell(\lambda)} - \mathbb{I}_{I_u(\lambda)}\right)(W(v))dW(v)}{\left[\int_0^1 W^2(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))\,dv - \frac{\left(\int_0^1 W(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))dv\right)^2}{\int_0^1 \mathbb{I}_{I_{out}(\lambda)}(W(v))dv}\right]^{1/2}} \tag{6}$$

where $\mathbb{I}_I$ denotes the indicator function which takes value one if $y_t \in I$ and zero otherwise. $I_{out}(\lambda) = I_\ell(\lambda) \cup I_u(\lambda)$, with $I_\ell(\lambda) = (-\infty, -\lambda]$ and $I_u(\lambda) = [\lambda, +\infty)$.

Theorem 2 *Consider the TAR specification (1). Let Λ_T be as in (3) and assume that Assumption E(s) given in Section 7 of Bec et al. [2008a] for $s > 4$ holds. Then, under H_0, $t_T^{inf}(\Lambda_T)$ converges in distribution to* $\inf_{\lambda\in\Lambda}(\xi_O(\lambda/\sigma))$*, which has a pivotal distribution.*

Theorem 2 follows directly from Theorem 2 of Bec et al. [2008a] but without the inner regime. We can see in particular that for $\underline{\lambda}_T = |y|_{(0)}$,

$$\xi_O(\underline{\lambda}_T) = \frac{\int_0^1 W(v)dW(v) - W(1)\int_0^1 W(v)dv}{\left[\int_0^1 W^2(v)dv - (\int_0^1 W(v)dv)^2\right]^{1/2}}$$

which is the limit distribution of the ADF_T statistic.

As in Kapetanios and Shin [2006], beside the infimum of the t statistic, its average and exponential average defined below will also be considered:

$$t_T^{avg}(\Lambda_T) = \frac{1}{\sharp\Lambda_T}\sum_{i=1}^{\sharp\Lambda_T} t(\lambda_i), \quad t_T^{exp}(\Lambda_T) = \frac{1}{\sharp\Lambda_T}\sum_{i=1}^{\sharp\Lambda_T}\exp\left(\frac{t_T(\lambda_i)}{2}\right),$$

where $\sharp\Lambda_T$ is the number of points included in the grid set Λ_T and λ_i is the ith point of the threshold parameters in Λ_T.

[3] See the proof of Theorem 4 in Bec et al. [2008a].

Theorem 3 *Consider the TAR specification (1). Let Λ_T be as in (3) and assume that Assumption E(s) given in Section 7 of Bec et al. [2008a] for $s > 4$ holds. Then, under H_0,*

$$t_T^{avg}(\Lambda_T) \Rightarrow \int_0^1 \xi_O(\lambda/\sigma)d\lambda \quad \textit{and} \quad t_T^{exp}(\Lambda_T) \Rightarrow \int_0^1 \exp\left(\frac{\xi_O(\lambda/\sigma)}{2}\right)d\lambda \tag{7}$$

where $\Rightarrow$ means convergence in distribution.

This result is directly obtained through the application of the continuous mapping theorem (Pollard, 1984). Lines 3 to 5 of Table 2 in the Appendix give their empirical critical values for different sample sizes.

3 Simulation experiments

In this section, the empirical size corrected power of various unit root tests is evaluated from simulation experiments. First, we present the different tests under scrutiny and then, the various alternatives used to compare their performances.

3.1 The competing unit root tests

Two very simple tests are considered as benchmarks : the ADF and the Kapetanios et al. [2003] t^{NL} test statistics. The latter has been developed to test the unit root null against a stationary Exponential Smooth Transition AR (ESTAR) model — first introduced by Chan and Tong [1986]. As these authors use a first-order Taylor series approximation of the ESTAR, they get the simple auxiliary model below:

$$\Delta y_t = a(L)\Delta y_{t-1} + \delta y_{t-1}^3 + u_t. \tag{8}$$

and the test statistic of the null $\delta = 0$ against $\delta < 0$ is simply given by $t_{T,NL} = \hat{\delta}_T/s.e.(\hat{\delta}_T)$, where $\hat{\delta}_T$ and its standard error are estimated by OLS. Then, our $t_T^{inf}(\Lambda_T)$, $t_T^{avg}(\Lambda_T)$ and $t_T^{exp}(\Lambda_T)$ tests are of course considered. For practical reasons, they are denoted t_{all}^{inf}, t_{all}^{avg} and t_{all}^{exp} in the tables below. The subscript "all" refers to the fact that all possible values of the threshold are considered.

We find it useful to compare these new test statistics to their bounded analogues. To this end, we use the same auxiliary model as the one given by Equation (1), but use the bounded adaptive set of thresholds, Λ_T^B defined by Equation (2.9) of Bec et al. [2008a]:

$$\Lambda_T^B = [\underline{\lambda}_T, \overline{\lambda}_T], \quad \text{with } \underline{\lambda}_T = |y|_{(2)} + \frac{\widehat{\sigma}_{\varepsilon T}}{\ell|DF_T|} \quad \text{and} \quad \overline{\lambda}_T = \underline{\lambda}_T + \ell\widehat{\sigma}_{\varepsilon T}|DF_T|, \tag{9}$$

with $\ell = 6$. Note that the lower bound starts after $|y|_{(2)}$: this is needed in their approach to estimate the inner regime's parameters.[4] The statistics obtained likewise are denoted t_b^{inf}, t_b^{avg} and t_b^{exp}.

Next, so as to compare our proposed tests to another one belonging to the unbounded, non adaptive class, we retain the approach described in Section 4, page 264, of Kapetanios and Shin [2006] in order to build the set of threshold. The latter consists in equally spaced points between the 10^{th} percentile (lower bound) and the 90^{th} percentile (upper bound). Unlike here, these authors do not choose symmetric threshold in general. Hence, they have to estimate the two thresholds which define the three regimes of their model. In their Monte-Carlo study, for each of these two thresholds, they retain eight points between each boundary and the sample mean, and they de-mean the series before proceeding to the unit root test. Since the thresholds are symmetric in our auxiliary SETAR model, we adapt their approach by using a grid set which also starts from the 10^{th} percentile and ends to the 90^{th} percentile, but includes equally spaced quantiles in between instead, so that: $\Lambda_T^U = \{Q(0.10), Q(0.15), Q(0.2), \ldots, Q(0.9)\}$ where $Q(\cdot)$ denote the quantile function of $|y_t|$.[5] As in Kapetanios and Shin [2006], the simulated series are de-meaned first. The corresponding t-statistics are denoted t_{ks}^{inf}, t_{ks}^{avg} and t_{ks}^{exp}.

Finally, in order to evaluate the impact of having developed a one-sided unit root test instead of the two-sided SupWald statistics used when the inner root is not constrained to be unity, the bounded Wald test proposed in Bec et al. [2008a] is also considered, both in its Sup, average and exponential versions — W_b^{sup}, W_b^{avg} and W_b^{exp} respectively.

All the tests are adjusted to match the number of lags introduced in the simulated DGP. All experiments are based on 10,000 Monte Carlo replications. The first 100 realizations are discarded so as to minimize the impact of initial conditions. The powers are empirical size corrected and evaluated at the 5% nominal level. Table 2 in the Appendix gives the critical values based on 40,000 simulations of different sample sizes.

3.2 Power analysis

The power of the tests listed in the first column of Table 2 is compared for various stationary alternatives. The first one is the stationary three-regime SETAR model for which all the tests have been built but the ADF and the t_{NL}. Next, other non-linear stationary alternatives are considered: the Exponential Smooth Transition AutoRegression (ESTAR), from which the t_{NL}

[4]Results of the Bec et al. [2008a]'s unbounded set analogue are not reported, as the latter is found to be slightly less powerful than the bounded one in most cases, as already noticed by these authors.

[5]Note that this approach is similar to the one developed by Bec et al. [2004], but it uses less points in the set of thresholds.

statistic has been built, the Autoregressive Conditional Root (ACR) model developed by Bec, Rahbek and Shephard [2008b] which is a dynamic mixture autoregression which does not require a fixed threshold, and the Mixed causal-noncausal AutoRegression (MAR). Note that none of the tests under study is built for the last two alternatives. Nevertheless, the bounded statistics derived from Bec et al. [2008a] as well as the new test proposed here should keep power against it as they have been shown to be consistent against all stationary alternatives. All the tables reporting the results of this power analysis are gathered in the Appendix.

SETAR alternatives: The model defined by Equation (1) is studied first, using the same set of parameters values as Bec et al. [2008a], as can be seen from Table 3. In particular, $\mu = 1.3 \times |\rho| \times \lambda$ and ε_t is an i.i.d. $\mathcal{N}(0,1)$. These values are inspired from Bec et al. [2004] analysis of real exchange rate data. The first four parameter sets correspond to DGPs which locate less than 5% of the realizations in the stationary upper and lower regimes. This explains the bad performance of the ADF test, as most of the sample lies in the inner random walk band. t_{NL} does slightly better than the ADF when the root is close to unity, here when $\rho = -0.1$, which amounts to a root of (1-ρ)=0.9. Then, comparing the other tests in these four cases, it appears that t_{ks} is always outperformed by t_{all}, t_b and W_b. So, the two adaptive tests are performing rather well. As expected, t_b is more powerful than W_b in the average and exponential versions. Finally, the best performing test in these cases is the t_{all}^{exp}. This is particularly remarkable in the first and third lines of these results, where only 1.7 to 2.9% of the observations lie in the stationary outer regime: it rejects the unit root null in 77.2% and 93.4% respectively. The closest test in terms of power is the t_b^{exp}. The last two sets of parameters generate series with more than 20% of realizations in the stationary regime. It is worth noticing that in these cases, even with a root close to unity (9^{th} and 10^{th} lines), all tests have a rather high power, the ADF included, with a rejection rate of 99.9% for T=200 and 100% for T=300. The last two lines of Table 3 show that when enough observations belong to the stationary regimes and the root is far from unity, then all the tests correctly reject the null in 100% of the cases.

ESTAR alternatives: The ESTAR DGP considered here is given by:

$$y_t = y_{t-1} + \gamma y_{t-1}(1 - \exp(\theta y_{t-1}^2)) + a\Delta y_{t-1} + \varepsilon_t$$

with $\varepsilon_t \sim \mathcal{N}(0,1)$. The rejection rates reported in Table 4, reveal more surprising results with ESTAR alternatives than with SETAR ones. Indeed, even though the test statistic t_{NL} has been built for this specific alternative, it is always dominated by other tests. In particular, when

θ is as small as 0.05 or 0.005, which corresponds to highly persistent series, our t_{all} test in its average and exponential versions reach much larger rejection rates. For instance, line 2 of Table 4 where T=300 and $\theta = 0.005$, t_{NL} rejects the null in 18.8% of the replications while t_{all}^{avg} and t_{all}^{exp} rejection rates are more than 64%. Note that the latter outperform t_{NL} in all the cases considered. In the cases where t_{NL} behaves rather well, as for instance in the last four lines of the table, even the ADF test performs better than it. Then, when $\theta = 0.2$, a case far from the null, the t_{ks}, t_{all} and ADF statistics produce comparable outcomes, close to 100% especially for T=300. Finally, as for the SETAR alternatives, the one-sided version of Bec et al. [2008a] bounded test statisitcs do better than the original Wald statistics, confirming the relevance of the unit root constraint in the central regime.

ACR alternatives: Table 5 reports rejection rates of the various tests for ACR stationary alternatives. The ACR model, proposed by Bec et al. [2008b], is given by:

$$y_t = (1+\rho)^{s_t} y_{t-1} + a\Delta y_{t-1} + \varepsilon_t$$

with $P(s_t = 1|y_{t-1}, \varepsilon_t) = [1 + \exp(-(\alpha + \beta|y_{t-1}|^{1/2}t))]^{-1}$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma)$. This model can generate SETAR-like dynamics with the difference that the threshold does not need to be fixed. Like them, we fix $\sigma = 0.009$ in our simulations. Again, t_{all}^{avg} and t_{all}^{exp} keep a remarkably good performance when a small percentage of the observations belong to the stationary regimes (first and last four lines of the table), compared to the other tests considered. Only the t_b statistics behave not too badly in these cases. Nevertheless, their performances deteriorate much more than the ones of t_{all}^{avg} and t_{all}^{exp} when the outer root gets closer to unity ($\rho = -0.1$).

Noncausal alternatives: The power of our proposed test is also evaluated for the Mixed causal-noncausal AutoRegressions (MAR) which have been shown to be a useful representation for bubble-like dynamics as well as for macroeconomic variables such as inflation or interest rates — see e.g. Lanne and Saikkonen [2011], Lof [2013], Lof and Nyberg [2017],Gouriéroux and Zakoian [2017] or Fries and Zakoian [2019]. Hence, it seems to be able to capture some kinds of nonlinearities. This MAR model is defined in Lanne and Saikkonen [2011] as

$$\phi(B)\varphi(B^{-1})y_t = \epsilon_t, \tag{10}$$

where B is the backward shift operator ($B^k y_t = y_{t-k}$ for $k = 0, \pm 1, ...$) and $\phi(B) = 1 - \phi_1 B - ... - \phi_r B^r$, $\varphi(B^{-1}) = 1 - \varphi_1 B^{-1} - ... - \varphi_s B^{-s}$. Finally, ϵ_t is a sequence of zero-mean non-Gaussian[6] independent, identically distributed random variables, and $E(\epsilon_t^2) < \infty$ unless otherwise mentioned. For the Monte-Carlo study, $r = s = 1$ and the forward, noncausal root, φ, is fixed to 0.5. A Student's t density function is assumed, $f(\epsilon_t \mid \sigma, \nu)$ where σ and ν are the scale and degrees of freedom parameters respectively. Following Lanne and Saikkonen [2011], the scale parameter σ is fixed to 0.1. Two values of the degrees of freedom are considered: $\nu = 3$ to generate fat-tailed distributed ϵ_t's and $\nu = 10$ to generate mildly fat-tailed distributed perturbations which look closer to Gaussian distribution than the case with $\nu = 3$. The causal root values considered are $\phi \in \{0.7, 0.9, 0.95\}$. The results are gathered in Table 6 in the Appendix. It can be seen that no matter the value of ν, the ADF test's empirical rejection rates are 100% or so when $\phi = 0.7$: The backward autoregressive root is far enough from unity for this simple test to reject the null in all drawings or so. Note that the t_{all} and t_{ks} statistics perform as well as the ADF in these cases. These three kinds of tests also perform well for $\phi = 0.9$, especially when $T = 300$. More interesting are the cases where $\phi = 0.95$. Here, t_{all}^{avg} and t_{all}^{exp}'s rejection rates are still around 74% for $T = 200$ and 88% for $T = 300$ when $\nu = 3$, and around 59% for $T = 200$ and 74% for $T = 300$ when $\nu = 10$. By contrast, both ADF, t_{NL}, t_b, W_b and, to a lower extent, t_{ks}, powers drop dramatically.

To sum up this power analysis, it turns out that our unbounded non adaptive tests, in their average and exponential versions, outperform existing bounded tests, either adaptive or not. The only exception to this result occurs when a large proportion of the realizations of the DGP belongs to the stationary regime of a SETAR or ESTAR model. However, the ADF test behaves very well in these cases so that no more sophisticated test is needed.

4 Empirical illustration

Here, we examine the same interest rate spread data as the ones used in Bec et al. [2008a].[7] Let S_F, S_G, S_{NZ} and S_{US} denote the 10-year vs 3-month bons spreads for France, Germany, New-Zealand and the U.S. respectively. As can be seen from Table 4, p.109, in Bec et al. [2008a], the standard ADF (or KPSS in the case of New-Zealand) tests suggest the presence of a unit root in these series. So, this is the kind of cases where using more powerful tests could make a difference. Indeed, from Table 5 of Bec et al. [2008a], it turns out that $\mathcal{W}_B^{Sup}$ rejects the unit root at the 1%-level for S_{NZ} and S_{US}, at the 3.5%-level for S_G and at the 15%-level only for

[6]With Gaussian distributed ϵ_t's, the model could be written equivalently as a backward or a forward autoregression. In this case, these two representations are observationally equivalent asymptotically, as discussed in e.g. Cambanis and Fakhre-Zakeri [1996].

[7]See the data description therein.

S_F. We re-examine these results using exactly the same data and sample, and hence with the same number of autoregressive lags in Eq. (1), namely $p = 1$ for France and Germany and $p = 4$ in New-Zealand and the US. As can be seen from Table 1, only the t^{avg}_{all} and t^{exp}_{all} as well as all

Table 1: Unit-root tests for interest rate spread data

	S_F	S_G	S_{NZ}	S_{US}
♯ obs.	228	228	205	259
ADF	-2.67	-1.89	**-3.12**	-2.73
t_{NL}	-2.82	**-3.56**	**-87.7**	**-145.48**
t^{inf}_{all}	-2.67	-3.08	**-3.99**	**-4.05**
t^{avg}_{all}	**-1.55**	**-1.50**	**-2.72**	**-2.96**
t^{exp}_{all}	**0.48**	**0.52**	**0.29**	**0.24**
t^{inf}_{b}	**-2.64**	**-3.10**	**-4.05**	**-4.09**
t^{avg}_{b}	**-1.54**	**-1.49**	**-2.74**	**-2.99**
t^{exp}_{b}	**0.49**	**0.52**	**0.29**	**0.24**
t^{inf}_{ks}	-2.67	-1.79	**-3.21**	-2.86
t^{avg}_{ks}	-2.05	-1.09	**-2.70**	**-2.41**
t^{exp}_{ks}	**0.37**	0.59	**0.27**	**0.30**
W^{sup}_{b}	10.96	**15.42**	**52.16**	**30.07**
W^{avg}_{b}	5.43	**9.47**	**12.48**	**17.37**
W^{exp}_{b}	28.95	**327.85**	**1.09·10⁹**	**1.10·10⁵**

Notes: Numbers in bold denote rejection of the null at the 5% level according to the corresponding critical values reported in Table 2.

versions of t_b test statistics reject the unit root null for the four series at the 5%-level. These results illustrate two points stressed earlier in the paper, especially in the Monte-Carlo study. First, the average and exponential versions of our unbounded and non adaptive t_{all} tests reject the null more often than the other tests. Second, the bounded Sup test developed by Bec et al. [2008a] rejects the null less often than its one-sided version proposed here, namely t^{inf}_b.

5 Conclusion

This paper proposes a new category of $t-$like unit root tests which are consistent against any stationary alternatives, nonlinear ones included. It departs from existing tests in that it uses an unbounded, not adaptive set of thresholds. In our setup, thanks to the very simple nonlinear stationary alternative specification and the particular choice of the thresholds set, the proposed unit root test contains the standard ADF test as a special case. This, in turn, yields a sufficient condition for consistency against any ergodic stationary alternative.

Our proposed tests power is then evaluated from a Monte-Carlo study. As a result, our unbounded non adaptive tests, in their average and exponential versions, outperform existing bounded tests, either adaptive or not. This suggests that using t_{all}^{avg} and t_{all}^{exp} test statistics on top of the simple ADF could prove very useful, especially for those series which may be suspected to behave in a nonlinear or noncausal way.

References

Andrews, D.W.K., Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica*, 1993, *61* (4), 821–856.

Balke, N.S. and T.B. Fomby, Threshold Cointegration, *International Economic Review*, 1997, *38*, 627–45.

Bec, F., A. Guay, and E. Guerre, Adaptive consistent unit-root tests based on autoregressive threshold model, *Journal of Econometrics*, 2008, *142*, 94–133.

——, A. Rahbek, and N. Shephard, The ACR Model: A Multivariate Dynamic Mixture Autoregression, *Oxford Bulletin of Economics and Statistics*, 2008, *70*, 583–618.

——, M. Ben Salem, and M. Carrasco, Test for Unit-root Versus Threshold Specification with an Application to the PPP, *Journal of Business and Economic Statistics*, 2004, *22*, 382–395.

Cambanis, S. and I. Fakhre-Zakeri, Forward and Reversed Time Prediction of Autoregressive Sequences, *Journal of Applied Probability*, 1996, *33*, 1053–1060.

Chan, K.S. and H. Tong, On estimating thresholds in autoregressive models, *Journal of Time Series Analysis*, 1986, *7*, 178–190.

Fries, S. and J.-M Zakoian, Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles, *Econometric Theory*, 2019, *35*, 1234–1270.

Gouriéroux, C. and J.-M. Zakoian, Local explosion modelling by noncausal process, *Journal of the Royal Statistical Society, Series B*, 2017, *79*, 737–756.

Kapetanios, G. and Y. Shin, Unit Root Tests in Three-Regime SETAR Models, *Econometrics Journal*, 2006, *9*, 252–278.

———, ———, and A. Snell, Testing for a Unit Root in the Nonlinear Framework, *Journal of Econometrics*, 2003, *112*, 359–379.

Lanne, M. and P. Saikkonen, Noncausal Autoregressions for Economic Time Series, *Journal of Time Series Econometrics*, 2011, *3*, 1–32.

Lof, M., Noncausality and asset pricing, *Studies in Nonlinear Dynamics and Econometrics*, 2013, *17*, 211–220.

——— and H. Nyberg, Noncausality and the commodity currency hypothesis, *Energy Economics*, 2017, *65*, 424–433.

Michael, P., A. Nobay, and D. Peel, Transactions costs and nonlinear adjustment in real exchange rates : an empirical investigation, *Journal of Political Economy*, 1997, *105* (4), 862–79.

6 Appendix

Table 2: Critical values (40,000 replications)

T	100	200	250	300	400	500	1000	10000
ADF	-2.89	-2.88	-2.88	-2.87	-2.87	-2.87	-2.87	-2.86
t_{NL}	-2.89	-2.90	-2.94	-2.91	-2.93	-2.91	-2.93	-2.93
t_{all}^{inf}	-2.98	-2.97	-2.97	-2.96	-2.97	-2.97	-2.96	-2.93
t_{all}^{avg}	-0.91	-0.85	-0.81	-0.80	-0.77	-0.74	-0.65	-0.38
t_{all}^{exp}	0.67	0.69	0.70	0.70	0.71	0.72	0.75	0.85
t_{b}^{inf}	-2.47	-2.54	-2.55	-2.58	-2.61	-2.62	-2.65	-2.67
t_{b}^{avg}	-0.93	-0.97	-0.99	-1.00	-1.06	-1.08	-1.23	-1.79
t_{b}^{exp}	0.65	0.64	0.63	0.63	0.61	0.60	0.56	0.42
t_{ks}^{inf}	-2.90	-2.87	-2.87	-2.85	-2.85	-2.84	-2.82	-2.52
t_{ks}^{avg}	-2.32	-2.09	-2.02	-1.95	-1.85	-1.77	-1.53	-0.80
t_{ks}^{exp}	0.33	0.37	0.39	0.40	0.42	0.44	0.50	0.71
W_{b}^{sup}	14.32	14.36	14.34	14.43	14.45	14.35	14.47	14.58
W_{b}^{avg}	6.22	6.15	6.23	6.24	6.35	6.45	6.70	7.47
W_{b}^{exp}	84.32	71.52	68.73	66.15	67.20	65.94	67.09	76.72

Table 3: Empirical rejection rates against SETAR alternatives (10,000 replications)

(λ, a, ρ_1)	T	%	ADF	t_{NL}	t_{all}^{inf}	t_{all}^{avg}	t_{all}^{exp}	t_b^{inf}	t_b^{avg}	t_b^{exp}	t_{ks}^{inf}	t_{ks}^{avg}	t_{ks}^{exp}	W_b^{sup}	W_b^{avg}	W_b^{exp}
(10,0,-0.1)	200	2.9	17.3	29.3	26.4	72.2	77.2	39.4	58.7	59.6	17.8	21.9	22.8	62.5	48.2	59.2
	300	2.9	20.1	38.6	41.6	91.7	93.0	53.1	78.5	79.0	20.6	25.1	26.2	83.4	66.8	80.2
(10,0,-0.3)	200	1.7	21.7	71.2	55.4	92.7	93.4	71.0	89.3	89.4	22.3	30.0	31.7	90.0	81.4	89.2
	300	1.5	28.5	86.6	76.7	98.4	98.5	85.6	96.6	96.5	29.4	41.6	45.0	97.1	92.9	96.6
(10,0.3,-0.1)	200	4.0	17.5	42.4	43.2	93.5	95.1	60.2	86.6	87.0	18.0	17.1	17.5	85.4	72.2	83.2
	300	3.9	23.4	68.7	64.9	99.1	99.4	79.2	97.7	97.9	23.6	17.0	16.5	97.0	91.4	96.1
(10,0.3,-0.3)	200	2.3	36.8	93.0	87.1	99.2	99.3	93.5	98.6	98.5	38.3	43.7	46.4	98.4	97.0	98.3
	300	2.1	74.9	99.1	98.1	99.9	100	98.9	99.9	99.9	75.3	71.2	73.2	99.8	99.6	99.8
(2,0.3,-0.1)	200	41.7	99.9	91.8	99.9	97.7	95.1	95.6	94.9	88.6	99.9	99.9	99.9	96.6	88.0	96.3
	300	41.6	100	98.2	100	99.7	98.5	99.5	99.0	95.6	100	100	100	100	99.8	100
(2,0.3,-0.3)	200	24.0	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	300	23.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Notes: The DGP is Equation (1) with $\mu = 1.3 \times |\rho| \times \lambda$ and $\varepsilon_t \sim \mathcal{N}(0,1)$. The column labeled % reports the percentage of data in the stationary regimes.

Table 4: Empirical rejection rate against ESTAR alternatives (10,000 replications)

(γ, a, θ)	T	ADF	t_{NL}	t_{all}^{inf}	t_{all}^{avg}	t_{all}^{exp}	t_{b}^{inf}	t_{b}^{avg}	t_{b}^{exp}	t_{ks}^{inf}	t_{ks}^{avg}	t_{ks}^{exp}	W_{b}^{sup}	W_{b}^{avg}	W_{b}^{exp}
(-0.1,0,0.005)	200	11.6	11.7	17.4	37.8	39.5	26.2	23.5	20.3	12.1	13.9	14.2	11.5	5.9	8.8
	300	15.5	18.4	27.3	64.3	64.0	35.0	33.7	27.0	16.1	19.3	19.7	17.7	8.0	14.4
(-0.1,0,0.05)	200	36.6	43.5	54.9	86.8	87.1	59.6	75.2	71.0	38.5	48.1	49.9	31.1	17.8	25.9
	300	78.8	74.7	87.4	96.6	96.2	76.3	88.5	84.8	80.5	89.6	91.2	55.4	38.5	52.0
(-0.1,0,0.2)	200	72.8	61.5	77.9	87.3	86.6	66.1	77.9	73.3	74.6	83.0	83.9	43.6	26.6	38.8
	300	98.3	85.7	98.3	96.1	94.9	82.5	89.1	84.7	98.6	99.6	99.7	78.8	56.6	75.2
(-0.1,0.3,0.005)	200	17.4	22.6	29.8	57.2	58.7	39.7	36.9	30.3	18.1	14.1	13.7	19.5	10.5	15.8
	300	31.6	43.2	50.9	85.4	85.0	54.1	57.4	47.8	32.6	25.4	24.9	33.9	19.5	29.9
(-0.1,0.3,0.05))	200	77.0	75.4	85.1	94.5	94.3	77.9	88.1	84.5	78.3	76.5	77.0	54.6	40.4	51.5
	300	99.0	94.3	99.5	99.4	99.3	91.6	96.9	94.8	99.1	99.1	99.2	86.6	76.5	86.2
(-0.1,0.3,0.2)	200	95.2	80.9	95.8	93.8	92.7	80.0	87.2	82.9	95.6	96.2	96.3	71.1	52.1	67.3
	300	100	95.1	100	98.8	98.0	92.8	95.7	92.6	100	100	100	97.0	87.8	96.4

Notes: The DGP is $y_t = y_{t-1} + \gamma y_{t-1}(1 - \exp(\theta y_{t-1}^2)) + a\Delta y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \mathcal{N}(0,1)$.

Table 5: Empirical rejection rates against ACR alternatives (10,000 replications)

(α, β, a, ρ)	T	%	ADF	t_{NL}	t_{all}^{inf}	t_{all}^{avg}	t_{all}^{exp}	t_{b}^{inf}	t_{b}^{avg}	t_{b}^{exp}	t_{ks}^{inf}	t_{ks}^{avg}	t_{ks}^{exp}	W_{b}^{sup}	W_{b}^{avg}	W_{b}^{exp}
(-10,30,0.3,-0.3)	200	4.5	28.1	72.3	61.0	97.6	98.1	74.9	94.1	93.7	29.0	27.4	27.8	72.9	64.5	71.2
	300	4.4	61.2	92.5	89.5	99.8	99.8	90.3	98.9	98.8	62.4	60.3	62.9	91.8	88.1	92.0
(-10,30,0.3,-0.1)	200	8.0	14.8	21.9	26.8	58.2	60.4	38.7	39.6	33.9	15.4	13.0	12.6	24.8	13.2	22.4
	300	7.9	22.3	41.3	46.0	89.0	89.7	55.4	63.7	55.5	22.9	18.5	18.3	42.4	31.5	39.9
(-20,120,0.3,-0.3)	200	19.5	100	99.9	100	100	100	100	100	100	100	100	100	100	100	100
	300	19.4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
(-20,120,0.3,-0.1)	200	35.1	89.3	81.9	94.5	97.1	96.3	88.2	93.3	88.3	90.3	90.8	91.5	69.9	57.3	67.9
	300	35.0	99.8	96.2	99.9	99.8	99.3	97.0	98.4	95.7	99.9	99.9	100	95.9	89.1	96.0
(-10,30,0,-0.3)	200	3.7	14.5	33.3	29.3	83.4	86.0	45.6	69.9	69.2	15.4	21.3	22.8	41.8	29.2	37.6
	300	3.5	22.1	56.3	50.8	96.8	97.2	64.1	88.9	88.0	23.3	32.5	35.3	62.6	48.7	59.8
(-10,30,0,-0.1)	200	6.3	10.7	13.0	14.9	28.1	31.3	23.1	18.1	16.5	11.4	15.0	15.9	13.3	9.3	11.7
	300	6.2	14.5	18.3	23.5	61.0	61.1	32.9	28.8	24.5	15.2	19.7	20.8	21.3	13.3.1	18.5

Notes: The DGP is $y_t = (1+\rho)^{s_t} y_{t-1} + a\Delta y_{t-1} + \varepsilon_t$, with $P(s_t = 1|y_{t-1}, \varepsilon_t) = [1 + \exp(-(\alpha + \beta|y_{t-1}|^{1/2} t))]^{-1}$ and $\varepsilon_t \sim \mathcal{N}(0, 0.009)$. The column labeled % reports the percentage of data in the stationary regimes.

Table 6: Empirical rejection rates against MAR alternatives (10,000 replications)

(ν,ρ)	T	ADF	t_{NL}	t_{all}^{inf}	t_{all}^{avg}	t_{all}^{exp}	t_{b}^{inf}	t_{b}^{avg}	t_{b}^{exp}	t_{ks}^{inf}	t_{ks}^{avg}	t_{ks}^{exp}	W_{b}^{sup}	W_{b}^{avg}	W_{b}^{exp}
(3,0.7)	200	99.9	99.1	99.9	99.8	99.8	96.1	98.2	97.3	99.9	100	100	87.6	82.8	85.4
	300	100	99.9	100	100	100	99.5	99.8	99.7	100	100	100	99.0	97.7	98.6
(3,0.9)	200	71.7	76.2	73.7	93.5	93.8	66.3	70.8	65.2	74.9	90.0	92.5	55.2	36.2	48.0
	300	96.9	92.1	96.7	98.5	98.6	82.8	81.9	77.6	97.7	99.7	99.9	78.1	57.1	71.0
(3,0.95)	200	26.2	38.7	30.2	73.3	74.2	38.4	33.7	27.7	28.4	39.6	43.3	27.3	13.5	22.1
	300	53.1	59.8	54.0	87.9	89.0	51.5	39.9	33.7	56.1	71.3	75.4	44.2	21.7	34.9
(10,0.7)	200	100	96.0	100	99.3	99.4	89.6	90.0	86.6	100	100	100	69.1	43.0	59.6
	300	100	99.7	100	100	100	97.6	97.3	94.3	100	100	100	94.9	78.3	91.9
(10,0.9)	200	73.5	57.0	73.0	81.8	82.9	47.2	36.5	32.0	76.4	88.6	90.0	36.6	8.8	23.7
	300	97.2	79.5	96.6	92.5	92.5	60.3	44.3	34.6	97.9	99.7	99.8	64.4	19.2	48.8
(10,0.95)	200	28.8	27.1	31.0	58.3	59.7	29.8	15.5	12.6	31.3	40.0	40.8	20.4	3.6	11.6
	300	55.6	44.3	55.6	74.3	74.6	38.6	15.3	9.6	58.5	69.8	71.8	35.6	6.0	21.6

Notes: The DGP is $(1-\phi B)(1-0.5B^{-1})y_t = 0.3+\varepsilon_t$, with $\varepsilon_t \sim t_{(\nu,\sigma)}$, where $\nu \in \{3,10\}$ is the degree of freedom and $\sigma = 0.1$ is a scale parameter.