THEMA Working Paper n°2022-14
CY Cergy Paris Université, France

Power of unit root tests against nonlinear and noncausal alternatives

Frédérique Bec, Alain Guay, Heino Bohn Nielsen, Sarra Saïdi

August 2022

Power of unit root tests against nonlinear and noncausal alternatives*

Frédérique Bec[†], Alain Guay[‡], Heino Bohn Nielsen[§] and Sarra Saïdi[¶]

Abstract

The increasing sophistication of economic and financial time series modelling creates a need for a test of the time dependence structure of the series which does not require a proper specification of the alternative. Indeed, the latter is unknown beforehand. Yet, the stationarity has to be established before proceeding to the estimation and testing of causal/noncausal or linear/nonlinear models as their econometric theory has been developed under the maintained assumption of stationarity. In this paper, we propose a new unit root test statistics which is both asymptotically consistent against all stationary alternatives and still keeps good power properties in finite sample. A large simulation study is performed to assess the power of our test compared to existing unit root tests built specifically for various kinds of stationary alternatives, when the true DGP is either causal or noncausal, linear or nonlinear stationary. Based on various sample sizes and degrees of persistence, it turns out that our new test performs very well in terms of power in finite sample, no matter the alternative under consideration.

Keywords: Unit root test, Threshold autoregressive model, Noncausal model.

JEL classification: C12, C22, C32.

* F. Bec and S. Saïdi acknowledge financial support from the Labex MMEDII (ANR11-LBX-0023-01). H. B. Nielsen acknowledges financial support from the Institute for Advanced Studies at CY Cergy Paris University and the Danish Council for Independent Research, DSF Grant 015-00028B.

[†]Thema, CY Cergy Paris University, and CREST, France, frederique.bec@cyu.fr.

[‡]Université du Québec à Montréal, guay.alain@uqam.ca.

[§]Department of Economics, University of Copenhagen, heino.bohn.nielsen@econ.ku.dk.

[¶]Thema, CY Cergy Paris University, France, sarra.saidi@cyu.fr.

1 Introduction

Nonlinear models play an important role in modern applied econometrics. Important examples include threshold autoregressions (TAR), mixed causal noncausal autoregression (MAR), autoregressive conditional root models (ACR), inter alia. The theoretical underpinning of these nonlinear models typically requires stationarity (absence of unit roots, in particular). More precisely, the knowledge of the potential presence of a unit root allows the proper choice of probability measure for further inference. For each model, unit root tests have been suggested to properly make the unit-root decision given the choice of the nonlinear framework. Important unit root tests include e.g. Dickey and Fuller [1979], Elliott et al. [1996], Kapetanios et al. [2003], see also Choi [2015] for a recent survey.

In practice, however, unit-root testing is a delicate issue in a nonlinear context: Firstly, nonlinear models can easily be mistaken for linear ones or mistakenly taken for alternative nonlinear specifications, see Breidt and Davis [1992] and Gouriéroux and Zakoian [2017]. Secondly, since the seminal paper of Dickey and Fuller [1979], unit root tests have become increasingly sophisticated, but most of them are built against a specific alternative. This means that, for a given time series at hand, the econometrician might have a hard time determining the potential stationarity of the series, without knowing the underlying data generating process (DGP).

A few non-parametric unit-root tests have been suggested, e.g. Bierens [1997a] or Breitung [2002]. These tests possess asymptotic power against any stationary alternative, but may suffer from low or moderate power in finite samples. In the parametric test category, a few tests have been shown to be consistent against any stationary alternative, e.g. Dickey and Fuller [1979] and Bec et al. [2008a]. Because of their asymptotic power against a broad range of stationary alternatives, we may think of these as *de facto* non-parametric tests and it would be extremely useful to see how they compare for a wide selection of linear, nonlinear and noncausal DGPs. Our hope is that a single unit-root test keeps satisfying power against any stationary alternative in finite samples, in which case the econometrician may decide on the stationarity of the time series without knowing the true DGP. After the unit-root decision, the econometrician may investigate the precise nature of the nonlinearity, knowing that the most important regularity condition for valid inference in the nonlinear models is satisfied.

In this paper, beyond evaluating a large variety of existing unit root tests, we propose a new one which is both asymptotically consistent against all stationary alternatives and still keeps good power properties in finite sample. A large simulation study is then used to assess the power of these tests, when the true DGP is either causal or noncausal, linear or nonlinear stationary. Based on various sample sizes and degrees of persistence, our main findings can be summed up as follows. As conjectured, our new test performs very well in terms of power in finite sample, no matter the alternative under consideration. There are only two cases where another test could be preferable. The first one is when the OLS estimate of the root from a linear autoregression is larger than 0.95. In this case, the test of Elliott et al. [1996] routinely outperforms the other ones. The second case is when there is a good motive to believe that the DGP is a noncausal one. In this particular case, tests built for noncausal alternatives, such as the ones by Saikkonen and Sandberg [2016] or Bec et al. [2020], could be preferred.

The article is organized as follows. Section 2 describes the various unit root tests that are used in the simulation exercises, including *(i)* noncausalparametric tests, *(ii)* tests which consider linear autoregressive alternatives, *(iii)* tests designed against threshold-like nonlinear alternatives and *(iv)* tests built for noncausal alternatives. It also introduces our new test. Section 3 presents the results of the Monte-Carlo experiments for each class of alternatives considered, namely linear causal autoregression (AR), nonlinear causal (Self Exciting Threshold AutoRegression (SETAR), Exponential Smooth Transition AutoRegression (ESTAR) and the ACR model developed by Bec, Rahbek and Shephard [2008b] (ACR)) and noncausal (MAR). Section 4 concludes.

2 Unit root tests under consideration

As emphasized in the Introduction, four distinct categories of unit root tests are considered below: *(i)* non parametric tests, *(ii)* tests with linear autoregressive alternatives, *(iii)* tests designed against threshold nonlinear alternatives and *(iv)* tests built for noncausal alternatives. Our new test belongs to the third category and hence will be developed there. This section motivates the choice of tests retained within each category for the subsequent Monte-Carlo study.

2.1 Non parametric tests

Even though they usually result in a loss of efficiency (the ability to detect a false hypothesis) and hence are generally found to be less powerful, non parametric tests[1] are relevant for our purpose because they do not require predefined specification of the DGP. Moreover, nonparametric tests may have power superiority for heavy-tailed distributions (see e.g. Zimmerman and Zumbo [1993]). In our study, Breitung [2002]'s unit root test will be considered. This simple variance ratio nonparametric unit root test statistic is asymptotically free of nuisance parameters and robust against misspecification and structural breaks in the short-run components. Moreover, it has been shown by its author to be more powerful than e.g. Bierens [1997a]'s tests.

Breitung [2002]'s test statistic is built like a Lagrange Multiplier statistic and more precisely as a variance ratio of the partial sums of the original series and the original series itself, say y_t:

$$LM = \frac{T^{-2} \sum_{t=1}^{T} Y_t^2}{T^{-1} \sum_{t=1}^{T} y_t^2} \tag{1}$$

where $Y_t = y_1 + ... + y_t$ denotes the partial sum process. Obviously, no estimation is needed: the series itself is used to compute the variance ratio. The limiting distribution of the test statistic under the null does not depend on nuisance parameters because the variance parameter σ^2 vanishes from the variance ratio (see Breitung [2002]).

2.2 Parametric tests devised against linear stationary alternatives

The second class of unit root tests considered in our analysis relies on linear autoregressive alternatives. The standard augmented Dickey-Fuller (ADF) test is retained as

[1] See e.g. Park and Choi [1988], Park [1990], Bierens [1997a], Bierens [1997b].

a benchmark. Recently, from a simulation study performed for various sample sizes and persistence degrees of a linear autoregressive DGP, Arltová and Fedorová [2016] conclude that the ADF test power behaves well in general compared to various unit root tests of the same class, namely the ones developed by Phillips and Perron [1988], Kwiatkowski et al. [1992], Elliott et al. [1996], Ng and Perron [1995] and Ng and Perron [2001]. Nevertheless, Arltová and Fedorová [2016]'s simulation study also reveals that for local alternatives, i.e. for roots typically greater than 0.95 but lesser than one, Elliott et al. [1996]'s test (ERS) reaches the largest power. For this reason, the latter will also be included in our simulation study. It consists in a General Least Squares version of the ADF test, sometimes referred to as ADF-GLS, where the deterministic component is estimated in a first step. In a second step, the ADF t-type test is performed on the demeaned/detrended series.

2.3 Parametric tests devised against threshold nonlinear stationary alternatives

The third class of tests considered consists in tests specifically designed against threshold nonlinear alternatives. Since the late nineties, a large number of this kind of tests have been developed such as the ones by Enders and Granger [1998], Lo and Zivot [2001], Bec et al. [2004], Kapetanios and Shin [2006] or Bec et al. [2008a] for stationary Threshold AutoRegressions (TAR) alternatives, or Kapetanios et al. [2003] and Bec et al. [2010] amongst others for stationary Smooth Transition AutoRegressions (STAR) alternatives.

2.3.1 SETAR stationary alternative

BGG Wald and *t* tests: In this paper, the unit root test of Bec et al. [2008a] is retained as it is built against Self-Exciting TAR (SETAR) stationary alternatives and has the desirable property of being consistent against any stationary alternative, noncausal ones included. It relies on a SETAR model with 3 regimes given by:

$$y_t = a(L)\Delta y_{t-1} + \begin{cases} \mu + \rho_1 y_{t-1} & \text{if } y_{t-1} \leq -\lambda \\ \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda \\ -\mu + \rho_1 y_{t-1} & \text{if } y_{t-1} \geq \lambda \end{cases} + u_t \tag{2}$$

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. The test of the joint hypothesis $H_0 : \rho_1 = \rho_2 = 1$ is based on a Wald statistic denoted $W_T(\lambda)$. In the original paper, the parameters are estimated so as to maximize the Wald statistic over a threshold set $\mathit{\Lambda}_T = [\underline{\lambda}_T, \overline{\lambda}_T]$: $SupW(\mathit{\Lambda}_T) = sup_{\lambda \in \mathit{\Lambda}_T} W_T(\lambda)$. We focus on the adaptive bounded threshold set[2] statistics, denoted here W_b^s, which provides the best power performance in Bec et al. [2008a]. The following average and exponential average versions of this test are also considered:

$$W_b^a = \frac{1}{\sharp\Lambda_T} \sum_{i=1}^{\sharp\Lambda_T} W_T(\lambda_i), \quad W_b^e = \frac{1}{\sharp\Lambda_T} \sum_{i=1}^{\sharp\Lambda_T} \exp\left(\frac{W_T(\lambda_i)}{2}\right),$$

[2] Adaptive means that the set is wider under the stationary alternative than under the null. Bounded refers to the property that the set size does not go to infinity as the sample size does.

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.

In order to enhance the power of this test, we follow Balke and Fomby [1997], Michael et al. [1997], Kapetanios et al. [2003] or Kapetanios and Shin [2006] and simplify further the auxiliary SETAR model by imposing that $\rho_2 = 1$. This way, the unit root test is not a joint hypotheses test anymore as it consists in $H_0 : \rho_1 = 0$. The test statistics are then build like Student's t infimum, average or exponential average statistics, denoted by superscript i, a and e respectively. Three versions of these statistics, depending on the choice of Λ_T's boundaries, are considered. The first one (denoted by t_b^i, t_b^a and t_b^e) is the t version of the Wald statistics described above. It is built from the same adaptive bounded thresholds set as described in Bec et al. [2008a] and its asymptotic properties derive straightforwardly from the ones of W_b therein. The two other versions, namely the one of Kapetanios and Shin [2006] and a new one we propose to gain power, are described below.

KS *t* test: The statistics built from the same non-adaptive unbounded threshold set as in Kapetanios and Shin [2006], denoted t_{ks}, is considered for comparison's sake, even though it is not consistent against any stationary alternative. This approach is described in Section 4, page 264, of Kapetanios and Shin [2006]: The set of thresholds Λ_T consists in equally spaced points between the 10^{th} percentile (lower bound) and the 90^{th} percentile (upper bound).

A new version of the *t* test: We develop here a new version of the t test, with the aim to gain power. Most theorems and proofs are gathered in the Appendix. As the one of Kapetanios and Shin [2006], it retains a grid set that is unbounded and non adaptive. But unlike the latter, we show that ours is consistent against any stationary alternative. 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. For this reason, it will be denoted t_{all}. It is worth noting that 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 $\underline{\lambda}_T$ and the $(T-1-k)^{th}$ value as the upper bound $\overline{\lambda}_T$, 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.

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

Theorem 1 *Consider the TAR specification (2) with $\rho_2 = 0$. 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, $t^i(\Lambda_T)$ diverges in probability, with*

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

Importantly, the inequality (4) indicates that the $t_T^i(\Lambda_T)$ 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. Then, Theorem A2 in the Appendix shows that $t_T^i(\Lambda_T)$ has a pivotal null distribution. Beside the infimum of the t statistic, denoted by t_{all}^i from now on[3], its average and exponential average defined below will also be considered:

$$t_{all}^a(\Lambda_T) = \frac{1}{\sharp\Lambda_T}\sum_{i=1}^{\sharp\Lambda_T} t_T(\lambda_i), \quad t_{all}^e(\Lambda_T) = \frac{1}{\sharp\Lambda_T}\sum_{i=1}^{\sharp\Lambda_T} \exp\left(\frac{t_T(\lambda_i)}{2}\right),$$

where $\sharp\Lambda_T$ and λ_i have been defined earlier. Theorem A3 gives their asymptotic distribution. The empirical critical values of these t_{all} statistics are reported in Table A1 — see Appendix — for T =100, 250, 500, 1000 and 10000. They are used to compute the empirical size corresponding to a theoretical size of 5%, reported in Table 1 below. As can be seen therein, the empirical size of our tests is quite accurate, especially the ones of t_{all}^a and t_{all}^e. It is worth noticing that t_{all}^i test seems slightly more liberal than the two other ones.

Table 1: Empirical size of t_{all} statistics

T	t_{all}^i	t_{all}^a	t_{all}^e
100	0.063	0.055	0.055
250	0.056	0.049	0.052
500	0.058	0.053	0.053
1000	0.054	0.052	0.05
10000	0.056	0.052	0.051

NOTE: *Size computed from 10,000 replications, for a theoretical size of 5%.*

2.3.2 ESTAR stationary alternative

Within this class of tests specifically designed against threshold nonlinear alternatives, the unit root test of Kapetanios et al. [2003] designed against ESTAR stationary alternatives is also considered as it is remarkably easy to implement. Using a first-order Taylor series approximation of the ESTAR model, these authors get the following simple auxiliary model:

$$\Delta y_t = a(L)\Delta y_{t-1} + \delta y_{t-1}^3 + u_t, \tag{5}$$

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. If needed, the series has to be demeaned or detrended beforehand and the critical values adjusted accordingly.

[3]The subscript 'all' refers to the fact that the thresholds set includes all possible values of $|y_{t-1}|$.

2.4 Parametric tests devised against noncausal stationary alternatives

The fourth class of tests considered here is built against noncausal stationary alternatives. First introduced in the statistics literature by Weiss [1975] and further explored by e.g. Breidt and Davis [1992], Rosenblatt [1993], Cambanis and Fakhre-Zakeri [1994] or Rosenblatt [2000][4], mixed causal-noncausal autoregressions have been recently considered as possibly relevant and parsimonious representations for macroeconomic and financial time series[5]. As emphasized in Lanne and Saikkonen [2011], by allowing predictable future errors, noncausal autoregressions can be used to empirically approach rational expectations models with nonfundamental solutions. The latter can also be seen as evidence that the econometrician uses a smaller information set than economic agents do. Noncausal autoregressions popularity also stems from their ability to capture epochs of bubble build-up and burst (see for instance Gouriéroux and Zakoian [2017], Fries and Zakoian [2019] or Fries [2021]).

Saikkonen and Sandberg [2016] have developed a unit root test based on Mixed causal-noncausal AutoRegression (MAR) stationary alternatives. The amended version of this test which we proposed in a previous work (Bec et al. [2020]) is retained here. Indeed, this modified version aims to circumvent likelihood bimodality issues which might affect Saikkonen and Sandberg [2016]'s unit root test power.

In the perspective of the unit root test, let us write the MAR(1,1) model as:

$$(1 - \rho B)(1 - \varphi B^{-1})y_t = u_t, \tag{6}$$

where B is the backward shift operator $By_t = y_{t-1}$, and u_t is a sequence of zero-mean non-Gaussian[6] independent, identically distributed random variables, with $E(u_t^2) < \infty$. This representation can be defined as an AR(1) process with errors following a purely noncausal MAR(0,1) process. Saikkonen and Sandberg [2016] propose to follow Lanne and Saikkonen [2011] to estimate the MAR model but depart from their framework by allowing for a unit root in the backward polynomial $\rho = 1$, which governs the causal component of the MAR. For tractability reasons, they maintain the assumption that the forward polynomial has his roots outside the unit circle.

Saikkonen and Sandberg [2016] propose a t-like test statistics for the null $\rho = 0$ against $\rho < 0$. However, Bec et al. [2020] show that a bimodality issue occurs in the Student's t case and leads to interchanged roots. Indeed, the backward root can be estimated as the forward root and *vice versa*. Since the unit root test is based on the estimation of the backward root, this can pose a problem. As a consequence, the estimation strategy proposed by Bec et al. [2020] relies on a grid search procedure in order to characterize the entire likelihood surface and list all local maxima. Then, if there are multiple maxima, the maximum with the backward root higher than the

[4]See also among others Findley [1986], Lawrance [1991], Breidt et al. [1991], Breidt et al. [1992] or Cambanis and Fakhre-Zakeri [1996].

[5]See e.g. Lanne and Saikkonen [2011], Lanne et al. [2012], Lanne and Saikkonen [2013], Hencic and Gouriéroux [2015], Gouriéroux and Zakoian [2015], Gouriéroux and Jasiak [2016], Hecq et al. [2020] or Fries [2021].

[6]With Gaussian distributed u_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].

forward root ($\hat{\rho} > \hat{\varphi}$) is selected even if it exists a maximum with a higher likelihood but with $\hat{\rho} < \hat{\varphi} \approx 1$. This choice stems from the fact that a unit root in the forward component would lack reasonable economic interpretation: It would mean that agents look into the infinite future without discounting. Beside, the test statistic is similar to Saikkonen and Sandberg [2016]'s. It is also a 't-ratio' type test statistic defined as :

$$t_{BNS} = \frac{(\hat{\rho}_T - 1)}{s.e.(\hat{\rho}_T)}$$

where $s.e.(\hat{\rho}_T)$ is the element of the inverse standardized Hessian matrix corresponding to $\hat{\rho}_T$. However, the limiting distribution is shown to be non-standard, and depends on a nuisance parameter. For Student's $t-$distributed errors, the limiting distribution depends on $\mathcal{J} = \frac{\nu_0(\nu_0+1)}{(\nu_0-2)(\nu_0+3)}$, where $\nu_0 > 2$ is the true degrees of freedom parameter.

3 A Monte-Carlo simulation approach

As mentioned in the Introduction, nonlinear causal models can be confounded with noncausal linear specifications. Since the true DGP is ignored beforehand in practice, it is useful to evaluate the power of the various unit root tests described above from simulated causal/noncausal and linear/nonlinear autoregressive DGPs. In this Section, the power analysis general framework is first described before turning to the overall power comparison.

3.1 Power Analysis framework

The simulation study, on which the comparison of the power of these various unit root tests relies, involves 10,000 replications of each DGP for three sample sizes — $T \in \{100, 250, 500\}$ — and many degrees of persistence[7]. The latter will include values which are typically found in quarterly or monthly macroeconomic and financial time series, namely close to but lesser than one: values of ρ ranking from 0.5 to 0.99 are considered. In order to compare properly the power of the different tests, we compute size-corrected empirical rejection rates (ERR hereafter). To this end, empirical critical values are computed under the null for all DGPs, based on 10,000 replications.

Let us first present the models considered under the alternative. Then, we will justify our choice to focus on 7 of the 17 test statistics discussed in Section 2.

3.1.1 Models considered under the alternative

Three classes of stationary DGPs will be used in our simulation study: linear causal (AR), nonlinear causal (SETAR, ESTAR, Autoregressive Conditional Root model (ACR)) and noncausal (MAR). This choice is motivated by the fact that the AR model is the stationary alternative of ADF and ERS unit root tests, the SETAR is the one of W_b, t_b, t_{ks} and t_{all} tests, the ESTAR is the one of t_{NL} test while the MAR model is the stationary alternative of t_{BNS} test. The ACR nonlinear model — first introduced by Gouriéroux and Robert [2006] and Bec, Rahbek and Shephard [2008b] — can generate

[7] In fact, 200, 350 and 600 realizations are simulated and the first 100 ones are removed from these samples.

SETAR-like dynamics with the difference that the threshold does not need to be fixed. It is retained even though none of the parametric tests is built against this stationary alternative. It makes this case interesting *per se*.

In order to make the various DGPs comparable for the power analysis, two autoregressive lags in levels are included in the AR, SETAR, ESTAR and ACR models while there is one lead and one lag in the MAR model. For all DGPs, $\rho = 1$ under the unit root null, while $\rho < 1$ under the alternative. Unless otherwise mentioned, $u_t \sim \mathcal{N}(0,1)$ in the following DGPs:

AR(2) Gaussian or Student: $y_t = \rho y_{t-1} + 0.3\Delta y_{t-1} + u_t$, with $u_t \sim \mathcal{N}(0,1)$ (Gaussian case) or $u_t \sim t(5)$ (Student case).

SETAR(2): $y_t = 0.3\Delta y_{t-1} + \rho y_{t-1}\mathbb{I}_{|y_{t-1}|>\lambda} + u_t$, where $\mathbb{I}_{|y_{t-1}|>\lambda}$ denotes the indicator function which takes value one if $|y_{t-1}| > \lambda$ and zero otherwise.

ESTAR(2): $y_t = y_{t-1} + \rho y_{t-1}(1 - exp(-0.2y_{t-1}^2)) + 0.3\Delta y_{t-1} + u_t$.

MAR(1,1): $(1 - \rho B)(1 - 0.5B^{-1})y_t = u_t$, whit $u_t \sim t(3)$.

ACR(2): $y_t = (1+\rho)^{s_t} y_{t-1} + 0.3\Delta y_{t-1} + u_t$, with
$P(s_t = 1|y_{t-1}, u_t) = [1 + \exp(-(-20 + 13|y_{t-1}|^{1/2}))]^{-1}$.

3.1.2 Pre-selection of unit root tests

As mentioned above, 17 unit root test statistics have been discussed in Section 2, and the simulation study has been performed for them all. Among these 17 test statistics, 12 have been developed against the SETAR alternative: the three versions (supremum, average and exponential) of Bec et al. [2008a]'s W_b, the three versions (infimum, average and exponential) of its single hypothesis analogue t_b, of Kapetanios and Shin [2006]'s t_{ks} and of our new test statistics t_{all}. A quick glance at Figure A1 in Appendix, which shows the empirical rejection rates of t_{all}^i, t_{all}^a and t_{all}^e for all DGPs with T=250, reveals *(i)* a clear domination of t_{all}^a and t_{all}^e over t_{all}^i and *(ii)* very similar power performances for t_{all}^a and t_{all}^e, with a slight advantage for t_{all}^e in some rare cases, as in the MAR case here. As can be seen from Figures A2 and A3 reported in the Appendix, the same conclusions can be drawn for t_b and t_{ks} test statistics. Then, comparing t_{all}^e , t_b^e and t_{ks}^e — see Figure A4 in Appendix which zooms on local alternatives— it appears that they all behave very much alike, with a very slight advantage for our new test for all DGPs and especially the ACR one.[8] This illustrates the gain in power as well as the consistency property of this test. Regarding Bec et al. [2008a]'s W_b's statistics ERRs, the average and exponential average are very similar and outperform W_b^s again — see Figure A5. For these reasons, within this class of test statistics built against SETAR alternatives, our attention will be focused on W_b^s and t_{all}^e in the subsequent analysis. To sum up, the following test statistics will be considered below: LM, ADF, ERS, W_b^s, t_{all}^e, t_{NL} and t_{BNS}.

[8] All these conclusions hold also for $T = 100$ et $T = 500$, but are not reported to save space.

3.2 Comparison of selected unit root tests

Figures 1 to 3 below present the size-corrected ERRs obtained for $T = 100, 250$ and 500 respectively.

Expectedly, the ERR of all tests increases with the sample size. For example, in the Gaussian AR(2) case, while for T=100 they rank from 10% (t_{BNS}) to 29% (ERS) when $(1-\rho) = 0.05$ and from 21% (t_{BNS}) to 65% (ERS) when $(1-\rho) = 0.1$, for T=500 they rank from 73% (LM) to 100% (t_{BNS}) when $(1-\rho) = 0.05$ and from 92% (LM) to 100% (all other tests but ERS) when $(1-\rho) = 0.1$.

Then, let us start with the results obtained for $T = 100$ as reported in Figure 1. For values of the root greater than 0.9 (i.e. $(1-\rho) < 0.1$), Elliott et al. [1996]'s ERS test outperforms the other ones for all DGPs but the MAR. It is also worth noticing that the ADF test keeps a relatively good power excepted when the true DGP is SETAR or MAR. For the former, t_{NL} has the highest power while for the latter, t_{BNS} does remarkably well compared to other tests. Remember that t_{BNS} is specifically designed for this noncausal alternative. Overall, for this sample size, our new test t^e_{all} seems to be a good option as its power is always the second or third best one, meaning that it never drops drastically. Beside, the performance of the various tests depends crucially on the true DGP. For instance, t_{BNS} lacks power for the AR with Gaussian errors, the SETAR and the ESTAR. The non parametric LM test of Breitung [2002] performs poorly in general, with the exception of the AR with Gaussian errors — in which case all tests but t_{BNS} do better. To conclude with this sample size, for the values of the root typically found in monthly or quarterly economic data (i.e. $(1-\rho) < 0.1$), the ERS test seems to be a safe choice, unless there are good reasons to suspect a MAR-like DGP, in which case it can be worth using t_{BNS} test statistic. Overall, the second best seems to be t^e_{all} as it keeps power against all kinds of stationary alternatives.[9]

As the results obtained with $T = 250$ and $T = 500$ lead to the same qualitative conclusions, let us focus on the former ones, plotted in Figure 2. First, the ERRs of LM confirm the lack of power of the non parametric test approach, as it performs very poorly in general compared to the other ones, especially when $(1-\rho) \geq 0.1$. Then, the ERS outperforms other tests again for very local alternatives — namely when $(1-\rho) < 0.05$ — and with the exception of the MAR DGP where t_{BNS} is the best again. Then, t^e_{all} and ADF reach rather similar and good ERRs overall, when the sample size is either 250 or 500. When their ERRs differ, it is always in favor of t^e_{all}, the most noticeable example being the SETAR DGP with $T = 250$.

For this reason, and because it is easy to implement, we recommend the use of t^e_{all} in finite sample, unless

- a simple linear AR OLS estimate of the root gives a value greater than 0.95 (in which case ERS is recommended),
- there is a good reason to suspect a noncausal underlying dynamics (in which case t_{BNS} is recommended).

[9]Remark that even though this consistency property is shared by W^s_b, it performs poorly with the MAR DGP when $T = 100$.

4 Conclusion

In practice, the kind of DGP of the data in hands is most likely unknown beforehand. For instance, such series as stock or commodity prices could be generated by nonlinear or noncausal processes. Yet, the econometric theory of the latter has been developed under the maintained assumption of stationarity. Unfortunately, using standard unit root tests as e.g. the ADF, those series are commonly found to be non stationary. The goal of this paper was to propose a simple test, asymptotically consistent against any stationary alternatives and keeping good power performance in small samples.

Our findings reveal that our new t^e_{all} test performs very well in general, no matter the sample size. Nevertheless, it is outperformed by Elliott et al. [1996]'s test for local alternatives when the root is greater than 0.95 and by Bec et al. [2020]'s t_{BNS} test for noncausal alternatives. It is also worth noticing that the standard ADF test turns out to be very reliable in large samples.

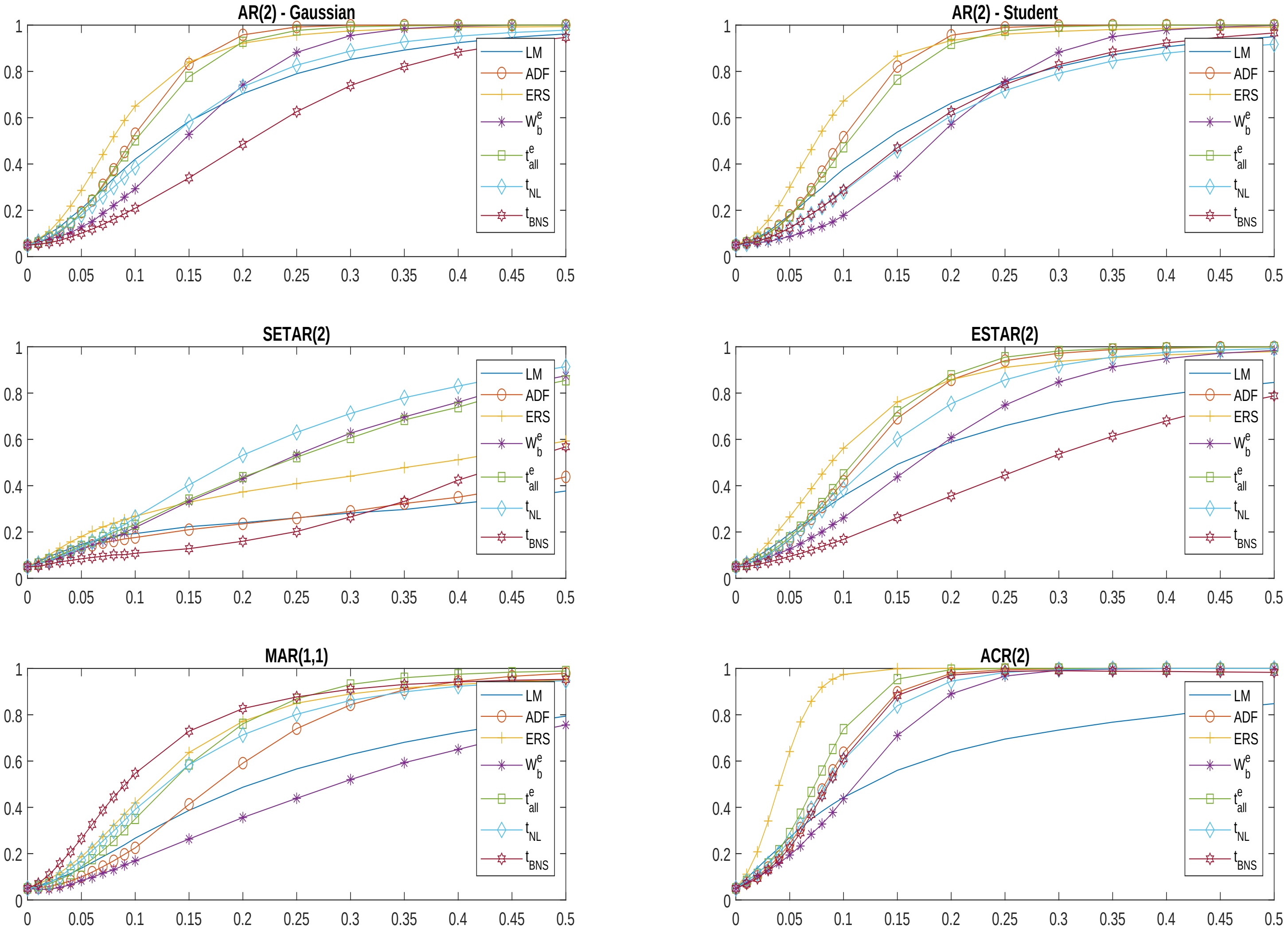

Figure 1: Power of selected unit root tests as a function of $(1 - \rho)$ for $T = 100$

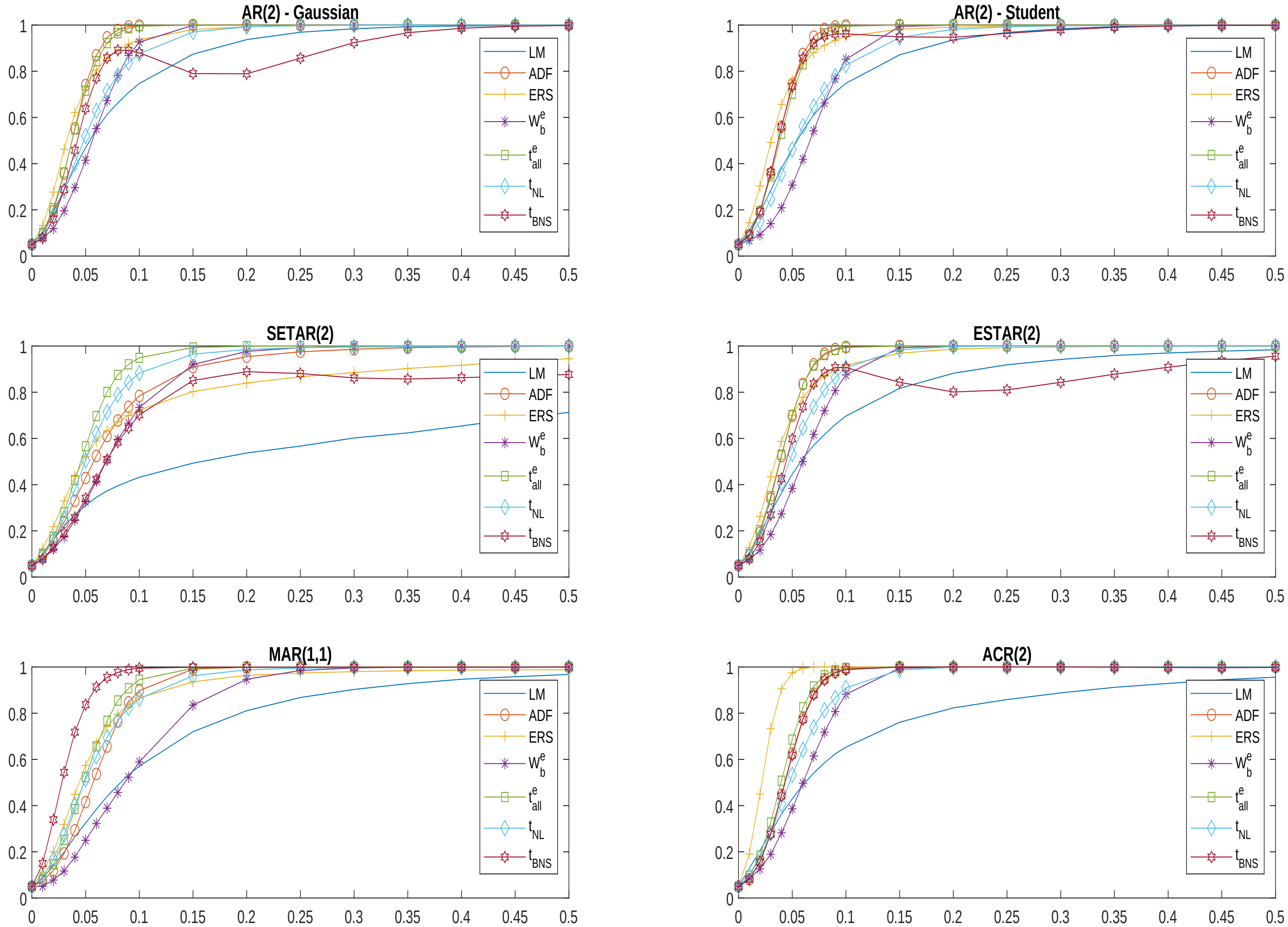

Figure 2: Power of selected unit root tests as a function of $(1-\rho)$ for $T = 250$

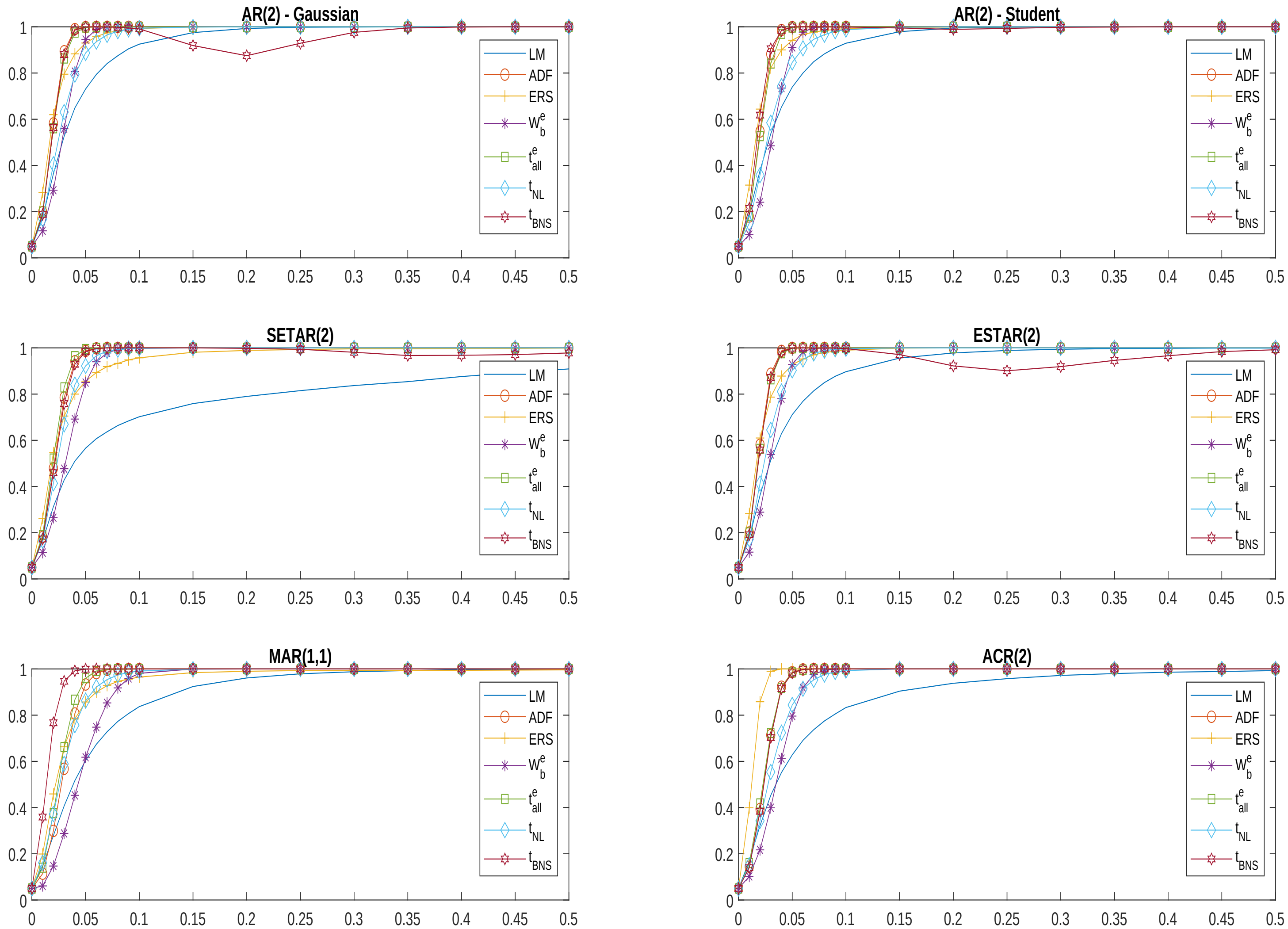

Figure 3: Power of selected unit root tests as a function of $(1-\rho)$ for $T = 500$

References

Arltová, M. and D. Fedorová, Selection of Unit Root Test on the Basis of Length of the Time Series and Value of AR(1) Parameter, *Statistika - Statistics and Economics Journal*, 2016, *96*, 47–64.

Balke, N. and T. 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 Autoregressive Conditional Root (ACR) Model, *Oxford Bulletin of Economics and Statistics*, 2008, *70* (5), 583–618.

_ , H. Bohn Nielsen, and S. Saïdi, Mixed Causal–Noncausal Autoregressions: Bimodality Issues in Estimation and Unit Root Testing, *Oxford Bulletin of Economics and Statistics*, 2020, *82*, 1413–1428.

_ , 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.

_ , M. Ben Salem, and M. Carrasco, Detecting Mean Reversion in Real Exchange Rates from a Multiple Regime STAR Model, *Annals of Economics and Statistics*, 2010, *99/100*, 395–427.

Bierens, H., Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate, *Journal of Econometrics*, 1997a, *81*, 29–64.

_ , Nonparametric cointegration analysis, *Journal of Econometrics*, April 1997b, *77*, 379–404.

Breidt, F. and R. Davis, Time-reversibility, identifiability and independence of innovations for stationary time series, *Journal of Time Series Analysis*, 1992, *13*, 377–390.

_ , _ , and W. Dunsmuir, On Backcasting in Linear Time Series Models, in D. et al. Brillinger, editor, *New Directions in Time Series Analysis. Part I*, New-York: Springer-Verlag, 1992, pp. 25–40.

_ , _ , K.-S. Lh, and M. Rosenblatt, Maximum likelihood estimation for noncausal autoregressive processes, *Journal of Multivariate Analysis*, 1991, *36*, 175–198.

Breitung, Jorg, Nonparametric tests for unit roots and cointegration, *Journal of Econometrics*, 2002, *108*, 343–363.

Cambanis, S. and I. Fakhre-Zakeri, On prediction of heavy-tailed autoregressive sequences: forward versus reversed time, *Teoriya Veroyatnostei i ee Primeneniya*, 1994, *39*, 294–312.

— and — , Forward and Reversed Time Prediction of Autoregressive Sequences, *Journal of Applied Probability*, 1996, *33*, 1053–1060.

Choi, I., *Almost All About Unit Roots: Foundations, Developments, and Applications*, Cambridge, UK: Cambridge University Press, 2015.

Dickey, D. and W. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 1979, *74*, 427–431.

Elliott, G., T. Rothenberg, and J. Stock, Efficient Tests for an Autoregressive Unit Root, *Econometrica*, 1996, *64*, 813–836.

Enders, W. and C.W.J. Granger, Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economic Statistics*, 1998, *16*, 304–11.

Findley, D., On bootstrap estimates of forecast mean square errors for autoregressive processes, in D. Allen, editor, *Computer Science and Statistics: The Interface*, Amsterdam: North-Holland, 1986, pp. 11–17.

Fries, S., Conditional Moments of Noncausal Alpha-Stable Processes and the Prediction of Bubble Crash Odds, *Journal of Business & Economic Statistics*, 2021.

— and J.-M. Zakoian, Mixed Causal-Noncausal AR Processes And The Modelling Of Explosive Bubbles, *Econometric Theory*, 2019, *35*, 1234–1270.

Gouriéroux, C. and C. Robert, Stochastic Unit Root Models, *Econometric Theory*, 2006, *22*, 1052–1090.

— and J. Jasiak, Filtering, Prediction and Simulation Methods for Noncausal Processes, *Journal of Time Series Analysis*, 2016, *37*, 405–430.

— and J.-M. Zakoian, On Uniqueness of Moving Average Representation of Heavy Tailed Stationary Processes, *Journal of Time Series Analysis*, 2015, *36*, 876–887.

— and — , Local explosion modelling by noncausal process, *J. R. Stat. Soc. B*, 2017, *79*, 737–756.

Hecq, A., J. V. Issler, and S. Telg, Mixed causal–noncausal autoregressions with exogenous regressors, *Journal of Applied Econometrics*, 2020, *35*, 328–343.

Hencic, A. and C. Gouriéroux, Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates, in Studies in Computational Intelligence: Econometrics of Risk 2015, pp. 17–40.

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 STAR framework, *Journal of Econometrics*, 2003, *112*, 359–379.

Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics*, 1992, *54*, 159–178.

Lanne, M., A. Luoma, and J. Luoto, Bayesian Model Selection And Forecasting In Noncausal Autoregressive Models, *Journal of Applied Econometrics*, 2012, *27*, 812–830.

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

_ and _ , Noncausal Vector Autoregression, *Econometric Theory*, 2013, *29*, 447–481.

Lawrance, A., Directionality and reversibility in time series, *nternational Statistical Review*, 1991, *59*, 67–79.

Lo, M. C. and E. Zivot, Threshold Cointegration and Nonlinear Adjustment to the Law of One Price, *Macroeconomic Dynamics*, 2001, *5*, 533–576.

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*, 862–79.

Ng, S. and P. Perron, Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag, *Journal of the American Statistical Association*, 1995, *90*, 268–281.

_ and _ , Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, *Econometrica*, 2001, *69*, 1519–1554.

Park, J., Testing for unit roots and cointegration by variable addition, *Fomby, T., Rhodes, F. (Eds.), Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots. Jai Press, Greenwich*, 1990.

_ and I.i Choi, A New Approach to Testing for a Unit Root, 1988.

Phillips, P. and P. Perron, Testing for a Unit Root in Time Series Regression, *Biometrika*, 1988, *75*, 335–346.

Rosenblatt, M., A note on prediction and an autoregressive sequence, in S. et al. Cambanis, editor, *Stochastic Processes, A Festschrift in Honour of Gopinath Kallia*, New-York: Springer-Verlag, 1993, pp. 291–295.

_ , *Gaussian and Nno-Gaussian Linear Time Series and Random Fields*, New-York: Springer-Verlag, 2000.

Saikkonen, P. and R. Sandberg, Testing for a Unit Root in Noncausal Autoregressive Models, *Journal of Time Series Analysis*, 2016, *37*, 99–125.

Weiss, G., Time-reversibility of linear stochastic processes, *Journal of Applied Probability*, 1975, *12*, 821–836.

Zimmerman, D. and B. Zumbo, The relative power of parametric and nonparametric statistical methods, in G. Keren and C. Lewis, editors, *A handbook for data analysis in the behavioral sciences: Methodological issues*, Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., 1993, pp. 481–517.

Appendix

Proof of Theorem 1. 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 (2) is equivalent to the autoregressive linear model when the threshold is $\lambda_T = |y|_{(0)}$.[10] In this case, the lower and the central regimes vanish and the SETAR shrinks to the auxiliary model of the ADF test: $\Delta y_t = a(L)\Delta y_{t-1} + \mu + \rho y_{t-1} + u_t$. ■

Before turning to the next theorem, let us first define for the outer regime:

$$\xi_O(\lambda) = \frac{\int_0^1 W(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))dW(v) - \frac{\int_0^1 W(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))dv}{\int_0^1 \mathbb{I}_{I_{out}(\lambda)}(W(v))dv}\int_0^1\left(\mathbb{I}_{I_\ell(\lambda)} - \mathbb{I}_{I_u(\lambda)}\right)(W(v))dW(v)}{\left[\int_0^1 W^2(v)\mathbb{I}_{I_{out}(\lambda)}\left(W(v)\right)dv - \frac{\left(\int_0^1 W(v)\mathbb{I}_{I_{out}(\lambda)}(W(v))dv\right)^2}{\int_0^1 \mathbb{I}_{I_{out}(\lambda)}(W(v))dv}\right]^{1/2}} \tag{7}$$

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 A2 *Consider the TAR specification (2) with $\rho_2 = 0$. 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 $\inf_{\lambda\in\Lambda}\left(\xi_O(\lambda/\sigma)\right)$, which has a pivotal distribution.*

Proof of Theorem A2. Theorem A2 follows directly from Theorem 2 of Bec et al. [2008a] but without the inner regime. In particular, it can be seen 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. ■

Theorem A3 *Consider the TAR specification (2) with $\rho_2 = 0$. 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,*

$$t_T^{avg}(\Lambda_T) \Rightarrow \int_0^1 \xi_O(\lambda/\sigma)d\lambda \quad \text{and} \quad t_T^{exp}(\Lambda_T) \Rightarrow \int_0^1 \exp\left(\frac{\xi_O(\lambda/\sigma)}{2}\right)d\lambda \tag{8}$$

where $\Rightarrow$ means convergence in distribution.

Proof of Theorem A3. This result is directly obtained through the application of the continuous mapping theorem (Pollard, 1984). ■

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

Table A1: Empirical critical values of t_{all} statistics

	T=100			T=250			T=500			T=1000			T=10000		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
t^i_{all}	-4.271	-3.715	-3.433	-4.312	-3.788	-3.526	-4.377	-3.858	-3.607	-4.434	-3.921	-3.679	-4.584	-4.112	-3.876
t^a_{all}	-3.278	-2.750	-2.476	-3.229	-2.726	-2.472	-3.226	-2.730	-2.481	-3.242	-2.738	-2.490	-3.238	-2.745	-2.498
t^e_{all}	0.208	0.268	0.305	0.212	0.269	0.304	0.212	0.268	0.302	0.210	0.266	0.299	0.209	0.265	0.298

NOTE: *Empirical critical values computed from 100,000 replications.*

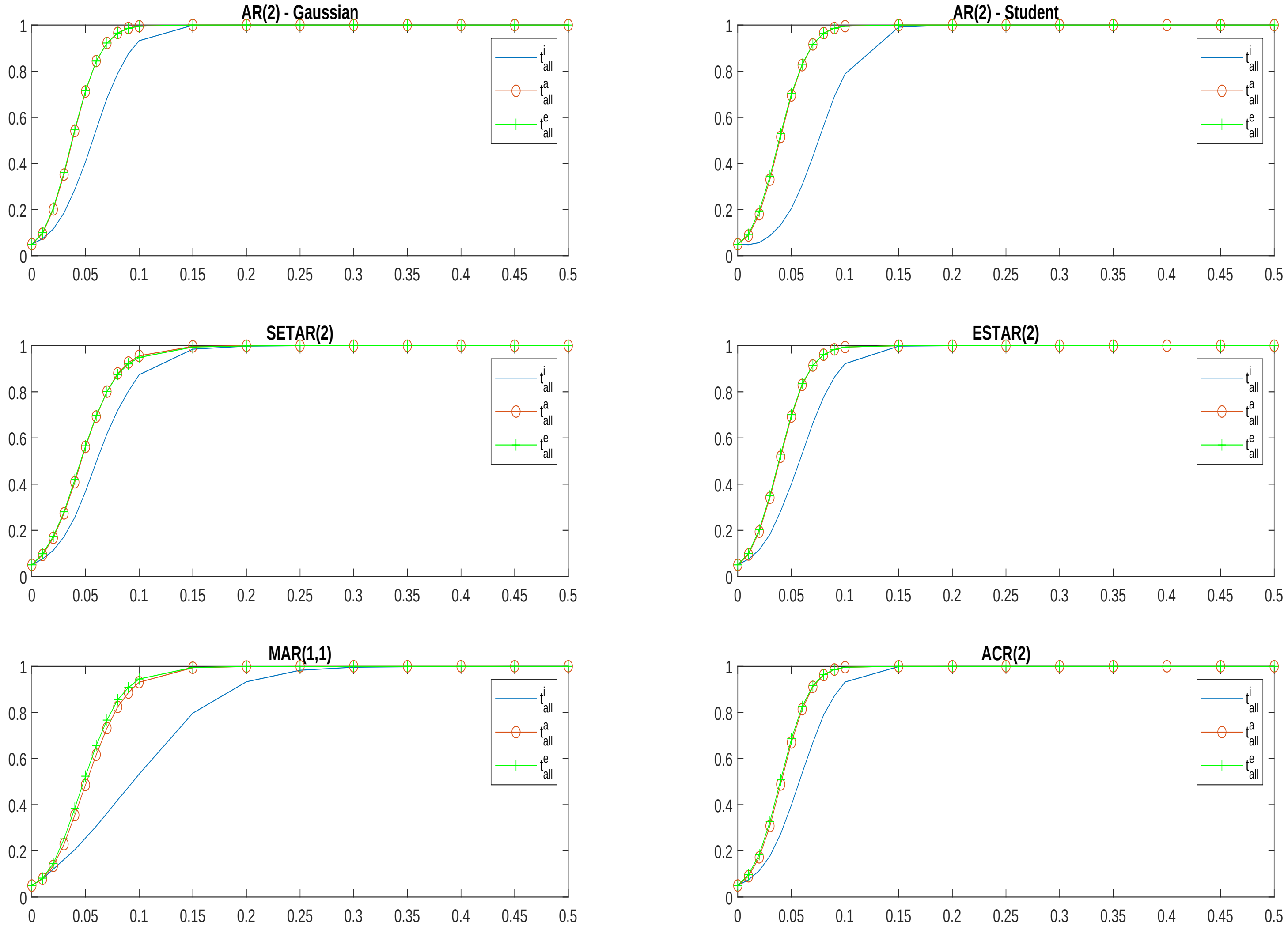

Figure A1: Power of t_{all} unit root tests as a function of $(1-\rho)$ for $T = 250$

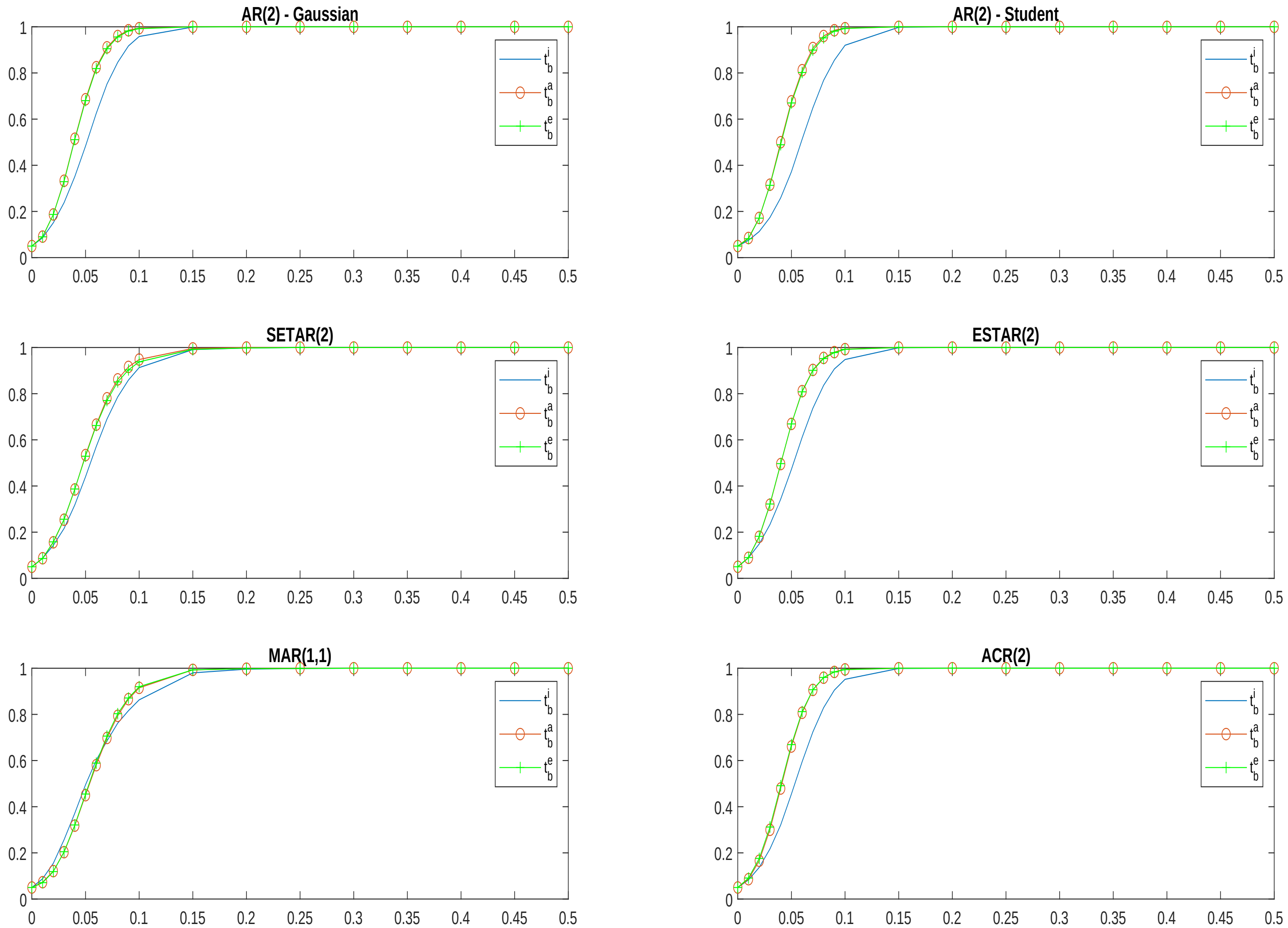

Figure A2: Power of t_b unit root tests as a function of $(1-\rho)$ for $T = 250$

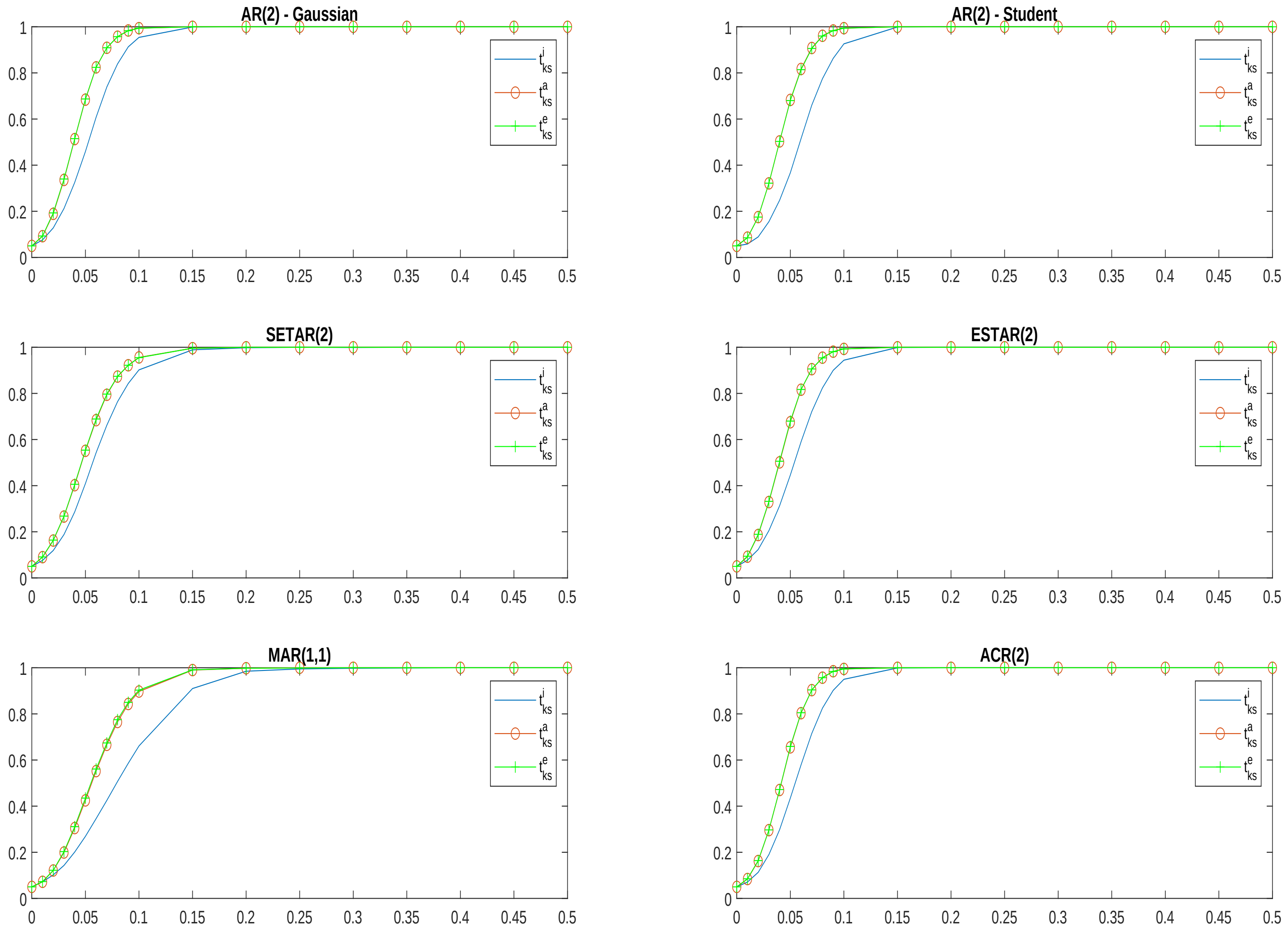

Figure A3: Power of t_{ks} unit root tests as a function of $(1-\rho)$ for $T = 250$

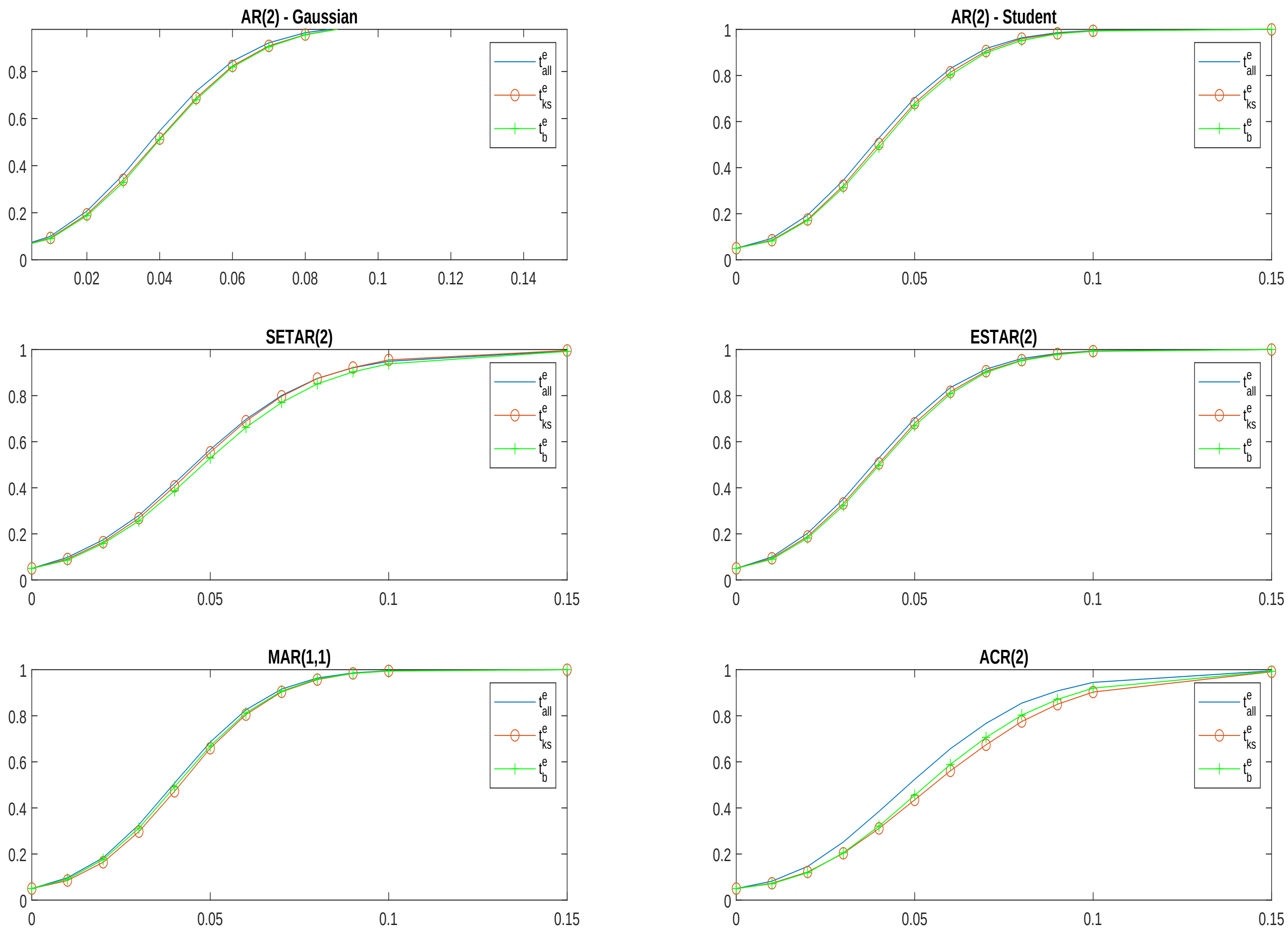

Figure A4: Power of t^e unit root tests as a function of $(1-\rho)$ for $T = 250$

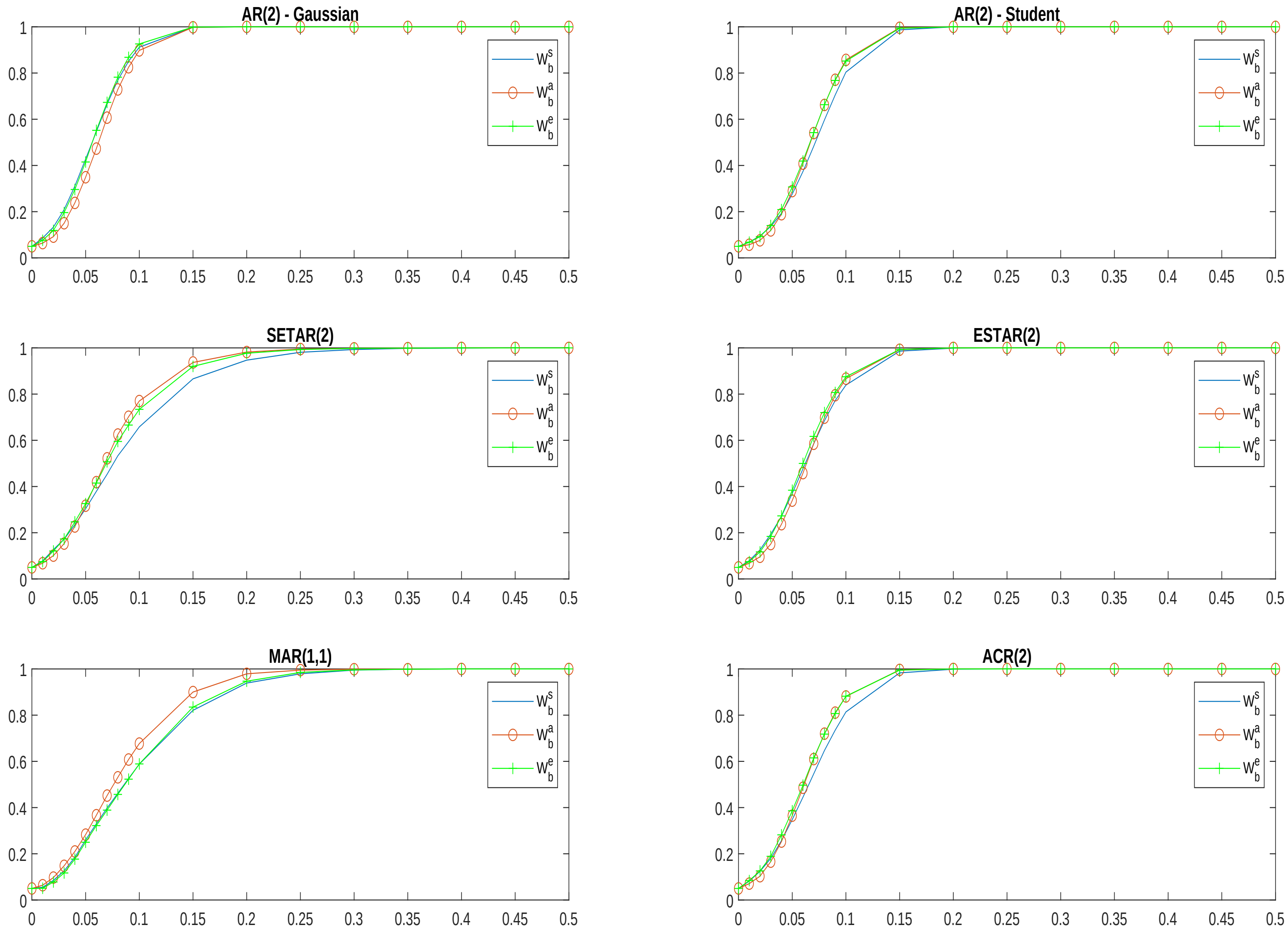

Figure A5: Power of W_b unit root tests as a function of $(1-\rho)$ for $T = 250$