Econometrics Journal (2007), volume **10**, pp. 82–112.
doi: 10.1111/j.1368-423X.2007.00200.x

How useful are tests for unit-root in distinguishing unit-root processes from stationary but non-linear processes?

CHI-YOUNG CHOI[†] AND YOUNG-KYU MOH[‡]

[†]*Department of Economics, University of Texas at Arlington, Arlington, TX 76019, USA*
E-mail: cychoi@uta.edu

[‡]*Department of Economics, Tulane University, New Orleans, LA 70118, USA*
E-mail: ymoh@tulane.edu

Received: February 2005

Summary Standard unit-root tests are known to be biased towards the non-rejection of a unit-root when they are applied to time series with non-linear dynamics. Unfortunately, not much is known about the source of the power loss mainly because the analysis on nonstationarity and nonlinearity to this date has been fragmentary. By means of a Monte Carlo study, the current paper investigates the finite sample performance of five popular unit-root tests against a wide class of non-linear dynamic models. In contrast to the common perception, our simulation results suggest that what determines the power of unit-root tests is not the specific type of nonlinearity in the alternative model, but how far the alternative model is away from the unit-root process. The presence of nonlinearity seems immaterial to the performance of unit-root tests if the non-linear process is far away from the unit-root process, which is in line with the fact established in linear framework. Among the five tests under study, the ADF test outperforms when the sample size is relatively small while the inf-*t* due to Park and Shintani (2005) is more powerful for relatively large sample size regardless of the form of true models. We then illustrate the empirical relevance of our analysis by reexamining the issue of mean reversion in real interest rates, often referred to the Fisher hypothesis.

Key words: *Unit-root tests*, *Non-linear dynamic models*, *Monte Carlo simulation*, *the Fisher hypothesis*.

1. INTRODUCTION

Stationary but non-linear processes have been often blamed for the poor performance of conventional unit-root tests. Ever since pointed out by Perron (1989, 1990) that the standard Dickey–Fuller (DF) tests underreject the unit-root null hypothesis when they are applied to a stationary time series with structural shifts, the properties of conventional linear unit-root tests have been assessed by numerous subsequent studies in the presence of diverse non-linear models [e.g. Pippenger and Goering (1993) and Taylor (2001) for threshold models; Kim, Leybourne and Newbold (2002) and Leybourne, Mills and Newbold (1998) for structural break models; Hall, Psaradakis and Sola (1997) and Nelson, Piger and Zibot (2001) for Markov-switching models; among many others]. In general, their evaluations are not encouraging for the tests which reveal poor discriminatory power against non-linear alternatives that are beyond the conventional linear ARMA framework. Given that earlier unit-root tests were constructed under the maintained assumption of linear and symmetric adjustment, a departure from linearity through curvature or kinkedness could lead the tests to misinterpret it as permanent stochastic disturbances. In response, more recent studies have developed testing procedures that are designed to accommodate specific non-linear dynamics against unit-root process and they often reversed the empirical conclusions on nonstationarity established by the standard DF-type tests. For all the contributions and advancement in the literature, the analysis on nonstationarity and nonlinearity thus far has been fragmentary in the sense that extant studies have predominantly focused on the comparison with the standard DF-type tests only under a specific class of parametric non-linear models. The current literature consequently leaves important issues intact such as the source of power loss and the relative performance of tests under a wide variety of non-linear models.

The primary objective of this study is to address these issues via simulation experiments under more comprehensive setup. Specifically, we evaluate the finite sample properties of popular tests for unit-root against a broad class of non-linear dynamic models. Beware that our focus here is not to address whether macroeconomic time series is best characterized by a linear or a non-linear model but to investigate what mistakes we can make in drawing inference on stationarity of series when we rely on usual testing procedures for unit-root. To this end, we consider five tests for unit-root, the conventional ADF test together with four recent tests that are designed for diverse non-linear processes under the alternative hypothesis: the M-TAR test proposed by Enders and Granger (1998), the sign test given by So and Shin (2001), the test due to Kapetanios, Shin and Snell (2003, hereafter KSS), and the inf-t test developed by Park and Shintani (2005). It is important to note that the five tests have the common null hypothesis of unit-root but different alternative hypotheses. The standard ADF test primarily concentrates on linear stationarity as the alternative, while the M-TAR test has the Threshold Autoregressive (TAR) model, the KSS test has the Smooth Transition Autoregressive (STAR) model, the sign test has general linear and non-linear stationary AR models, and the inf-t test has general transitional AR models under the alternatives. As such the selection of the tests were governed by the treatment for non-linear dynamic models under the alternative.[1]

Extant studies on unit-root tests tend to highlight the usefulness of their tests by showing the power improvement over the conventional DF-type tests under specific non-linear models maintained in the alternative hypothesis. However, the outperformance could be a natural consequence from the design of the tests and there is no guarantee that it will be witnessed in other non-linear models as well. Moreover, since the true underlying model is usually unknown in practice, it is imperative to evaluate the performance of tests under a wide class of non-linear models. Given that the class of non-linear models includes virtually an infinite number of models and specifications that are not linear, we focus on a subset of autoregressive non-linear models which are popularly adopted in the literature of macroeconomics and international finance. Table 1 reports the data generating processes considered in the current study that encompasses a

[1] Other popular tests that are not considered here fall into a subset of these tests in terms of parametric specification of non-linear dynamic models under the alternative. For example, the testing procedures proposed by Bec, Guay and Guerre (2002), Caner and Hansen (2001), Seo (2003) and Kapetanios and Shin (2003) test unit-root against the alternative of multi-regime TAR model (two-regime for Caner and Hansen and Seo while three-regime for Bec *et al.* and Kapetanios and Shin).

Table 1. Summary of DGPs.

DGP no.	Data Generating Process	Model
1	$y_t = \rho y_{t-1} + \varepsilon_t$	AR(1)
2	$y_t = \rho y_{t-1} + \phi y_{t-1}^2 + e_t,\ e_t \sim i.i.d.(0, \sigma_1^2)$	Generalized AR(1)
3	$y_t = \rho y_{t-1} + \phi y_{t-1}\, e_{t-1} + e_t,\ e_t \sim N(0, \sigma_1^2)$	Bilinear (BL)
4	$y_t = (\rho\ \lvert y_{t-1}\rvert)/(\lvert y_{t-1}\rvert + c) + \varepsilon_t$	Non-Linear AR
5	$y_t = x_t^2 + \epsilon_t,\ x_t = \rho x_{t-1} + \varepsilon_t,\ \epsilon_t \sim N(0, 1)$	Squared relation (SR)
6	$y_t = exp(x_t) + \epsilon_t,\ x_t = \rho x_{t-1} + \varepsilon_t,\ \epsilon_t \sim N(0, 1)$	Exponential relation (ER)
7	$y_t = \alpha + [1 + e^{-\gamma(y_{t-1}-x_t)}]^{-1} + [1 + e^{-\gamma(y_{t-1}+x_t)}]^{-1} + v_t,$	Binary neural
	$x_t = \rho x_{t-1} + e_t,\ \ v_t \sim N(0, \sigma_1^2), e_t \sim N(0, \sigma_2^2)$	Network (BNN)
8	$y_t = \rho_1\, y_{t-1} + \rho_2\, y_{t-1}\, I(y_{t-1} \geq c) + \varepsilon_t$	SETAR(1)
9	$y_t = y_{t-1} + \varepsilon_t$, if $\lvert y_{t-1}\rvert \leq k$	EQ-TAR
	$y_t = \rho y_{t-1} + \varepsilon_t$, if $\lvert y_{t-1}\rvert > k$	
10	$y_t = k(1-\rho) + \rho y_{t-1} + \varepsilon_t$, if $y_{t-1} > k$	Band-TAR
	$y_t = y_{t-1} + \varepsilon_t$, if $\lvert y_{t-1}\rvert \leq k$	
	$y_t = -\,k(1-\rho) + \rho y_{t-1} + \varepsilon_t$, if $y_{t-1} < -k$	
11	$y_t = \alpha + \rho_1\, y_{t-1} + \theta \cdot (\beta + \rho_2\, y_{t-1}) + \varepsilon_t,$	ESTAR
	where $\theta = 1 - e^{-\gamma(y_{t-1}-c)^2}$	
12	$y_t = \alpha + \rho_1\, y_{t-1} + \theta \cdot (\beta + \rho_2\, y_{t-1}) + \varepsilon_t,$	LSTAR
	where $\theta = [1 + e^{-\gamma(y_{t-1}-c)}]^{-1}$	
13	$y_t = \rho_t\, y_{t-1} + e_t,\ e_t \sim i.i.d.(0, 0.4)$	Markov-switching (MS) in
	$\rho_t = \rho_1\, S_t + \rho_2(1 - S_t)$	AR coefficients
14	$y_t = \alpha_1 + \rho y_{t-1} + \varepsilon_t$, if $t \leq \lambda T$ where $0 < \lambda < 1$	Structural change (SC)
	$y_t = \alpha_2 + \rho y_{t-1} + \varepsilon_t$, if $t > \lambda T$	in level
15	$y_t = \alpha_1 + \rho y_{t-1} + \varepsilon_t$, if $t \leq \lambda_1\, T$ where $0 < \lambda_i < 1$	Multiple SCs
	$y_t = \alpha_2 + \rho y_{t-1} + \varepsilon_t$, if $\lambda_1\, T < t \leq \lambda_2\, T$	
	$y_t = \alpha_3 + \rho y_{t-1} + \varepsilon_t$, if $\lambda_2\, T < t \leq T$	
16	$y_t = \alpha + \rho y_{t-1} + \sigma_1\, \varepsilon_t$, if $t \leq \lambda T$ where $0 < \lambda < 1$	SC in innovation
	$y_t = \alpha + \rho y_{t-1} + \sigma_2\, \varepsilon_t$, if $t > \lambda T$	
17	$y_t = \alpha + y_{t-1} + \varepsilon_t$	Unit-root process
18	$y_t = y_{t-1} + \sigma_t \varepsilon_t,\ \sigma_t = \sigma_1 S_t + \sigma_2(1 - S_t)$	Regime switching
		with unit-root

Note: $I(s)$ denotes an *indicator* function which takes on the value of 1 if the argument is true and 0 otherwise. Parameter values in simulations are set to $k = 3$, $\phi = -0.1$, $\gamma = 100$, $\alpha = \alpha_1 = 0$, $\alpha_2 = -0.5$, $\alpha_3 = 1.5$, $\sigma_1 = 0.01$, $\sigma_2 = 0.05$, $\varepsilon_t \sim N(0, 1)$, $P_{11} = \text{Prob}(S_t = 1 \mid S_{t-1} = 1) = 0, 95$, $P_{22} = \text{Prob}(S_t = 2 \mid S_{t-1} = 2) = 0, 9$ where S_t is a discrete, unobserved state variable that takes on the value of 1 or 2 in the regime switching models of DGPs 13 and 18.

family of well-known non-linear dynamic models such as bilinear (BL), generalized AR (GAR), non-linear AR (NAR), TAR, STAR, Markov-switching and structural break models. It should be noted that these non-linear time series models are based on mixtures of local autoregressive models in which the persistence of process is influenced by the autoregressive parameter (hereafter *'associated'* AR parameter). In our simulation we consider two different values for the associated autoregressive parameter ($\rho = \rho_1 = 0.5$ and 0.9) which are set to be identical across models in order to examine the impact of the associated AR parameters on test performances. Another notable feature of our simulation study is to probe the potential impact of nonlinearity on the performance of tests. This is carried out by looking at the response of test performance to the different values of parameters which are believed to be relevant for non-linear configurations such as convexity, curvature, or kinkedness. Take DGP 8 (SETAR model) and DGP 13 (Markov-switching model), for example, the parameters are pertaining to the kinkedness in SETAR or the transition probabilities in the case of the Markov-switching model. By so doing, we attempt to throw additional light on the source of power loss of unit-root tests in the presence of non-linear dynamic processes.

Our simulation results reveal several interesting points. First, the performance of unit-root tests is more affected by the magnitude of associated AR parameters than by nonlinearity *per se*. All the tests display decent discriminatory power irrespective of model specifications if the associated AR parameter is mild indicating that the non-linear model is rather far away from the unit-root, whereas the power drops sharply when the associated AR parameter increases. Nonlinearity seems playing a part in the power loss only when the associated AR parameter is close to unity. Second, the power loss is a finite sample problem as in the case of linear models. The power loss is most serious when the sample size is relatively small while it improves with the sample size in the vast majority of non-linear models. Third, among the tests under comparison, the ADF test and the inf-*t* test stand out. In many non-linear models considered, the standard ADF test outperforms the other tests under comparison particularly when the sample size is relatively small. The inf-*t* test is most powerful for large sample sizes as its power reaches unity regardless of the types of model. Fourth, in contrast to our prior beliefs, all the unit-root tests have certain discriminatory power against the models beyond the ones stipulated in the alternative hypothesis. For example, the ADF test has a satisfactory power property for various non-linear models, and the M-TAR test frequently rejects the false null of unit-root even when the true underlying model follows other non-linear process than TAR-type models. In this context, it will be misleading if not dangerous to interpret rejection of the unit-root null as compelling evidence of the non-linear model under the alternatives because the rejection can be driven by a myriad of other models.

On the empirical plane of this paper, we illustrate the practical relevance of our analysis by investigating the mean reversion of real interest rates, or the Fisher hypothesis, which has been a popular subject of research in macroeconomics and finance. Despite extensive research, the empirical evidence on the hypothesis remains inconclusive largely due to econometric challenges involved in identifying stationarity of time series. By implementing the five unit-root tests to the real interest rates of 12 OECD countries, we find supportive evidence of non-linear but stationary behavior of the postwar real interest rates. Specifically, our analysis points toward possible structural changes in the variables, which has been challenging to detect by the standard unit-root tests.

The remainder of this paper is structured as follows. In the following section, we review the five univariate tests employed in the current study. In Section 3, we present the simulation results on the finite sample performance of the tests. Section 4 discusses the potential impact of nonlinearity on the test performance. Section 5 demonstrates the empirical relevance of our analysis and Section 6 concludes.

2. UNIVARIATE TESTS

This section briefly outlines the five univariate tests employed in our analysis. For details on these test procedures, the reader is referred to their original work. Throughout the paper, we focus on the regression with no time trend which is more compatible with many empirical topics of interest such as the PPP hypothesis or the growth convergence.

2.1. The ADF test

The ADF test is formulated by the following regression equation:

$$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{k} \phi_j \Delta y_{t-j} + \epsilon_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where ϵ_t is a white noise error term. Under the null hypothesis of unit-root ($H_0 : \rho = 0$) against the alternative of stationarity ($H_A : \rho < 0$), the test statistic has a non-normal and non-standard limiting distribution which is tabulated by Fuller (1976). As stressed in many studies, the choice of lag length for k in equation (1) has a large effect on the test performance such that too few lags adversely affect the size of tests while too many lags can reduce power. In the current paper, we follow Ng and Perron (1995) to adopt the sequential t-test method with the maximum lag length set as integer[$8(T/100)^{1/4}$].[2]

2.2. The M-TAR test

Enders and Granger (1998) generalize the DF test to consider the null hypothesis of a unit-root against the alternative of the momentum-threshold autoregressive (M-TAR) model which allows a variable to display more momentum in one direction than the other that can be parametrized as

$$\Delta y_t = I_t \rho_1 [y_{t-d} - \tau] + (1 - I_t \rho_2 [y_{t-d} - \tau]) + \epsilon_t \tag{2}$$

where I_t is the indicator function such that

$$I_t = \begin{cases} 1 & \text{if } \Delta y_{t-d} > \tau \\ 0 & \text{if } \Delta y_{t-d} \leq \tau \end{cases} \tag{3}$$

and τ denotes the value of the threshold and d represents the delay parameter. They set $d = 1$ so that the threshold depends on the previous period's change in y. The value of threshold τ is unknown and hence needs to be estimated along with ρ_1 and ρ_2. Using the estimated value of τ, the F-statistic (Φ_μ) is constructed under the null hypothesis of unit-root ($\mathrm{H}_A : \rho_1 = \rho_2 = 0$). Since the distribution of the test statistics is nonstandard due to the presence of nuisance parameters, we use the critical values reported in Enders (2001).

[2] We obtain similar results from using Ng and Perron's (2001) modified information criterion (MIC) and with another maximum lag length rule of integer[$12(T/100)^{1/4}$].

2.3. *The sign test*

So and Shin (2001) develop a nonparametric sign test for the unit-root null against the alternative of general linear and non-linear stationary AR process. Let $\{y_t\}$ be a monotone transformation of a possibly non-linear AR process $\{x_t\}$,

$$x_t = \rho(x_{t-1}, \ldots, x_{t-k}) + u_t, \quad t = 1, 2, \ldots, T,$$

where $\rho\ (x_{t-1}, \ldots, x_{t-k})$ is an unknown regression function of interest and $\{u_t\}$ is a sequence of conditional zero median errors. The test statistic is then constructed as

$$S_T(1) = \sum_{t=1}^{T} \operatorname{sign}(y_t - y_{t-1}) \operatorname{sign}(y_{t-1} - \hat{m}_{t-1}), \tag{4}$$

where $\hat{m}_{t-1}$ denotes the median of $\{y_i\}_{i=0}^{t-1}$. The unit-root null hypothesis ($H_0 : x_t = x_{t-1} + u_t$) is rejected if $S_T(1) \leq 2B_T(\alpha) - T$ where $B_T(\alpha)$ denotes the lower α_{th} quantile of the binomial distribution $B(T, 1/2)$. Using Monte Carlo experiments, So and Shin show that the sign test has stable size and locally better power than the parametric DF test as well as other nonparametric tests.

2.4. *The KSS test*

Kapetanios *et al.* (2003) propose a testing procedure under the unit-root null against the alternative of a STAR model. In an ESTAR model

$$\Delta y_t = \gamma y_{t-1}\left[1 - \exp\left(-\theta y_{t-1}^2\right)\right] + \varepsilon_t,$$

testing unit-root against non-linear stationarity is equivalent to testing $H_0 : \theta = 0$ against $H_A : \theta > 0$. Since θ is not directly identifiable under the null, KSS use the following auxiliary regression derived from a first-order Taylor expansion,

$$\Delta y_t = \delta y_{t-1}^3 + \text{error}. \tag{5}$$

Under the null ($\delta = 0$), a t-type test statistic has a limiting distribution of functionals of the standard Brownian motion. Asymptotic critical values of test statistics are tabulated by KSS.

2.5. *The Inf-t test*

Park and Shintani (2005) develop a unit-root test against general transitional AR models that embrace a wide range of AR models with threshold, discrete and smooth transition dynamics. Consider an autoregressive model,

$$\Delta y_t = \lambda(\theta) y_{t-1} \pi(y_{t-d}, \theta) + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \varepsilon_t, \tag{6}$$

where y_{t-d} is the transition variable with delay lag $d \geq 1$, θ is an m-dimensional nuisance parameter which can be identified only under the alternative hypothesis of stationarity and π denotes a real-valued transition function on $\mathbb{R} \times \mathbb{R}^m$. Since the regression model in (6) can represent a broad class of non-linear partial adjustment AR models with relevant choice of the transition functions, the test is claimed to be useful in identifying unit-root in diverse transitional AR models.

Testing the unit-root null hypothesis is equivalent to testing $H_0 : \lambda = 0$ against $H_1 : \lambda < 0$, which involves estimating λ in (6) using least squares for each possible value of the transition parameter $\theta \in \Theta_n$ to obtain the following t-ratio

$$T_n(\theta) = \frac{\hat{\lambda}_n(\theta)}{s(\hat{\lambda}_n(\theta))} \tag{7}$$

where $s(\hat{\lambda}_n(\theta))$ is the standard error of the estimate $\hat{\lambda}_n(\theta)$. The inf-t test is then defined as

$$T_n = \inf_{\theta \in \Theta_n} T_n(\theta),$$

which is the infinum of t-ratios in (7) taken over all possible values of $\theta \in \Theta_n$, where Θ_n is a random sequence of parameter spaces given for each n as functions of the sample $(y_1, \ldots, y_n)$. The limit distribution of inf-t statistic is free from any nuisance parameters and depends only on the transition function and the limit parameter space. Throughout the paper, we stick to the transition function for ESTAR model to account for the unknown underlying model in practice.

3. FINITE SAMPLE PERFORMANCE

Table 1 summarizes 18 models of linear and non-linear dynamics that we adopt for our simulation experiments. They encompass simple extensions of the conventional AR model (DGPs 1 through 7), the endogenous and exogenous regime switching models (DGPs 8 through 13) as well as the models with structural shifts (DGPs 14 through 16). Note that DGPs 1 through 16 are stationary models to investigate the power performance of unit-root tests, whereas DGPs 17 and 18 are unit-root processes for the size properties. Note that equilibrium errors in these models are stationary but with different short-term dynamics. In fact, DGPs 14–16 violate the usual definition of stationarity as the distribution is not same through time, but they are classified as stationary here because they would be rejected with probability one by a standard unit root test when the sample size is large enough as shown by Perron (1990, p. 156). These models include many interesting non-linear dynamic models which are popularly employed by a multitude of studies [e.g. Hamilton (1989), Garcia and Perron (1996), Michael, Nobay and Peel (1997), Rothman (1998), and Sarno, Taylor and Chowdhury (2004), amongst many others] to characterize the dynamic behaviour of macroeconomic variables such as real output, industrial production, unemployment rates, interest rate and real exchange rates. Figure 1 plots the simulated sample series from each DGP for $T =$ 100.

We consider the sample sizes of $T \in \{50, 100, 200, 500\}$ that are likely to be encountered in empirical analysis. Each simulation run is carried out with 5,000 replications. At each replication $T + 500$ random numbers are generated, of which the first 500 observations are discarded to minimize the influence of initial condition.[3] Pseudo-random numbers are generated using the GAUSS (version 6.0) RNDNS and RNDUS procedures. One noteworthy feature in our simulation design is that the values for the associated autoregressive parameter (ρ and ρ_1) are set to be identical across DGPs 1 through 16. We consider two scenarios for ρ by setting it to 0.5 and 0.9, in order

[3] Müller and Elliott (2003) note that the performance of some unit-root tests depends on the initial conditions, but we find that the tests considered here are not sensitive to the initial conditions. For DGPs 14–16, additional initial 500 observations are generated for the first regime only.

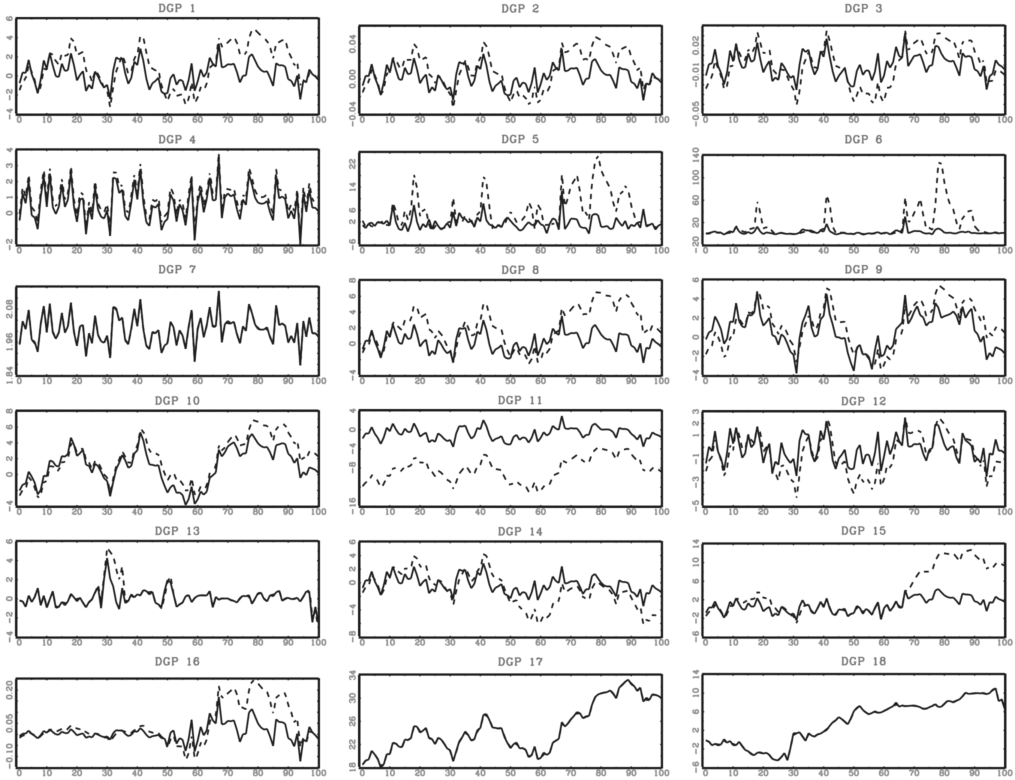

Figure 1. Simulated DGPs (T = 100) for $\rho = 0.5$ (solid line) and $\rho = 0.9$ (dashed line).

to explore the impact of the magnitude of associated AR parameter on the power performance of unit-root tests. Unless specified otherwise, the other parameters are set to $k = 3$, $\phi = -0.1$, $\gamma = 100$, $\alpha = \alpha_1 = 0$, $\alpha_2 = -0.5$, $\alpha_3 = 1.5$, $\sigma_1 = 0.01$, $\sigma_2 = 0.05$, $P_{11} = 0.95$, $P_{22} = 0.9$ throughout the experiments.[4]

Table 2 presents the rejection rates of the unit-root tests which represent the number of times out of 5,000 simulation that the unit-root null hypothesis is rejected at the 10% nominal size.[5] Recall that they are related to the power performance of tests for DGPs 1 through 16, while to the size performance for DGPs 17 and 18. The results bear several important features of note. First, the power performance is highly sensitive to the value of associated AR parameter (ρ) particularly when the sample size is relatively small. When $\rho = 0.5$, all the testing procedures exhibit decent

[4] P_{ij} denotes the probability that process switches from regime *i* to regime *j* such that $P_{11} = \text{Prob}(S_t = 1 \mid S_{t-1} = 1)$, $P_{22} = \text{Prob}(S_t = 2 \mid S_{t-1} = 2)$ where S_t is a discrete, unobserved state variable that takes on the value of 1 or 2 in the regime switching models of DGPs 13 and 18.

[5] While the table only reports the results for i.i.d. error term, a broadly similar story is told when the error terms are serially correlated but with a slight power deterioration except for a large negative MA root in which the tests are known to suffer from severe size distortions.

Table 2. Rejection rates (at 10% significance level).

DGP/	ADF test				MTAR test				Sign test				KSS test				inf − t test			
/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
									$\rho = \rho_1 = 0.5; \rho_2 = 0.05$											
1	0.94	0.98	**1.00**	**1.00**	**0.95**	**1.00**	**1.00**	**1.00**	0.79	0.99	**1.00**	**1.00**	0.50	0.79	0.86	0.93	0.90	**1.00**	**1.00**	**1.00**
2	0.94	0.98	**1.00**	**1.00**	**0.95**	**1.00**	**1.00**	**1.00**	0.79	0.99	**1.00**	**1.00**	0.51	0.79	0.86	0.93	0.90	**1.00**	**1.00**	**1.00**
3	**0.94**	0.97	**1.00**	**1.00**	**0.94**	**1.00**	**1.00**	**1.00**	0.77	0.99	**1.00**	**1.00**	0.51	0.77	0.85	0.91	0.90	**1.00**	**1.00**	**1.00**
4	0.96	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.99	**1.00**	**1.00**	**1.00**	0.58	0.77	0.79	0.84	**1.00**	**1.00**	**1.00**	**1.00**
5	0.95	0.98	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.98	**1.00**	**1.00**	**1.00**	0.48	0.66	0.73	0.75	0.99	**1.00**	**1.00**	**1.00**
6	0.92	0.96	**1.00**	**1.00**	**0.99**	**1.00**	**1.00**	**1.00**	0.93	**1.00**	**1.00**	**1.00**	0.59	0.77	0.88	0.94	0.97	**1.00**	**1.00**	**1.00**
7	0.96	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.99	**1.00**	**1.00**	**1.00**	0.58	0.77	0.79	0.84	**1.00**	**1.00**	**1.00**	**1.00**
8	**0.93**	0.98	**1.00**	**1.00**	**0.93**	**1.00**	**1.00**	**1.00**	0.77	0.98	**1.00**	**1.00**	0.49	0.79	0.86	0.94	0.87	**1.00**	**1.00**	**1.00**
9	0.49	0.84	0.97	**1.00**	0.24	0.78	**1.00**	**1.00**	0.20	0.39	0.54	0.85	**0.63**	0.94	**1.00**	**1.00**	0.54	**0.99**	**1.00**	**1.00**
10	**0.28**	0.38	0.81	**1.00**	0.11	0.18	0.56	**1.00**	0.19	0.35	0.49	0.81	0.23	0.49	0.87	**1.00**	0.22	**0.52**	**0.98**	**1.00**
11	**0.92**	0.97	**1.00**	**1.00**	0.88	**1.00**	**1.00**	**1.00**	0.71	0.98	**1.00**	**1.00**	0.50	0.80	0.89	0.96	0.84	**1.00**	**1.00**	**1.00**
12	0.95	0.98	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.94	**1.00**	**1.00**	**1.00**	0.42	0.70	0.75	0.76	0.97	**1.00**	**1.00**	**1.00**
13	0.93	0.97	**1.00**	**1.00**	**0.97**	**1.00**	**1.00**	**1.00**	0.88	**1.00**	**1.00**	**1.00**	0.48	0.71	0.79	0.86	0.91	**1.00**	**1.00**	**1.00**
14	**0.84**	0.88	0.94	**1.00**	0.82	**1.00**	**1.00**	**1.00**	0.71	0.97	**1.00**	**1.00**	0.41	0.69	0.72	0.78	0.73	**1.00**	**1.00**	**1.00**
15	0.24	0.21	0.16	0.14	0.32	**0.97**	**1.00**	**1.00**	**0.61**	**0.97**	**1.00**	**1.00**	0.06	0.10	0.07	0.04	0.14	0.76	**1.00**	**1.00**
16	0.81	0.91	0.99	**1.00**	**0.90**	**1.00**	**1.00**	**1.00**	0.80	0.99	**1.00**	**1.00**	0.43	0.60	0.66	0.80	0.86	**1.00**	**1.00**	**1.00**
17	0.13	0.10	0.09	0.11	0.08	0.11	0.10	0.11	0.03	0.02	0.02	0.03	0.10	0.09	0.10	0.09	0.09	0.09	0.09	0.10
18	0.15	0.13	0.12	0.10	0.06	0.05	0.04	0.03	0.07	0.12	0.11	0.10	0.14	0.14	0.12	0.11	0.14	0.12	0.11	0.11

Table 2. Continued

DGP/ /T	ADF test 50	ADF 100	ADF 200	ADF 500	MTAR test 50	MTAR 100	MTAR 200	MTAR 500	Sign test 50	Sign 100	Sign 200	Sign 500	KSS test 50	KSS 100	KSS 200	KSS 500	inf − t test 50	inf − t 100	inf − t 200	inf − t 500
$\rho = \rho_1 = 0.9; \rho_2 = 0.05$																				
1	**0.29**	**0.52**	**0.88**	**1.00**	0.08	0.21	0.73	**1.00**	0.19	0.47	0.72	0.99	0.19	0.37	0.62	0.94	0.19	0.42	0.85	**1.00**
2	**0.29**	**0.52**	**0.88**	**1.00**	0.08	0.21	0.73	**1.00**	0.19	0.46	0.72	0.99	0.19	0.37	0.61	0.93	0.19	0.41	0.85	**1.00**
3	**0.29**	**0.56**	0.86	**1.00**	0.10	0.20	0.65	**1.00**	0.17	0.39	0.64	0.96	0.22	0.40	0.53	0.75	0.25	0.54	**0.93**	**1.00**
4	0.95	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.99	**1.00**	**1.00**	**1.00**	0.59	0.77	0.81	0.83	**1.00**	**1.00**	**1.00**	**1.00**
5	**0.69**	0.86	0.96	**1.00**	0.50	0.81	**0.99**	**1.00**	0.59	0.88	**0.99**	**1.00**	0.44	0.60	0.76	0.88	0.66	**0.89**	**0.99**	**1.00**
6	0.76	0.82	0.94	0.99	0.76	0.96	**1.00**	**1.00**	0.58	0.84	0.96	**1.00**	0.67	0.88	0.96	0.99	**0.87**	0.97	**1.00**	**1.00**
7	0.95	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.99	**1.00**	**1.00**	**1.00**	0.59	0.77	0.81	0.83	**1.00**	**1.00**	**1.00**	**1.00**
8	**0.22**	**0.36**	**0.67**	**0.99**	0.06	0.10	0.39	0.99	0.16	**0.36**	0.53	0.93	0.15	0.28	0.47	0.86	0.14	0.28	0.63	**0.99**
9	**0.22**	0.33	0.71	0.99	0.06	0.10	0.37	**1.00**	0.17	0.33	0.48	0.85	0.16	**0.34**	0.66	0.98	0.16	0.32	**0.78**	**1.00**
10	**0.19**	0.21	0.32	0.90	0.04	0.05	0.11	0.66	0.12	**0.24**	0.32	0.63	0.12	0.18	0.32	0.88	0.13	0.17	**0.35**	**0.97**
11	**0.19**	0.25	**0.51**	**0.97**	0.05	0.07	0.19	0.88	0.14	**0.29**	0.43	0.81	0.13	0.21	0.36	0.78	0.13	0.20	0.41	0.95
12	**0.50**	**0.84**	0.97	**1.00**	0.25	0.66	**1.00**	**1.00**	0.35	0.69	0.93	**1.00**	0.26	0.46	0.64	0.86	0.37	0.75	**1.00**	**1.00**
13	**0.46**	**0.77**	0.96	**1.00**	0.24	0.57	**0.97**	**1.00**	0.41	0.75	0.96	**1.00**	0.30	0.49	0.71	0.90	0.37	0.70	**0.97**	**1.00**
14	**0.08**	0.08	0.21	0.76	0.02	0.03	0.05	0.80	**0.08**	**0.19**	**0.35**	0.77	**0.08**	0.12	0.30	0.80	0.06	0.09	0.29	**0.98**
15	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	**0.05**	**0.24**	**0.73**	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.49
16	0.30	0.47	0.73	0.99	0.09	0.25	0.68	**1.00**	0.23	0.48	0.72	0.98	0.28	0.41	0.57	0.83	**0.30**	**0.54**	**0.86**	**1.00**

Note: Entries represent the fraction of times when the null hypothesis is rejected out of 5,000 replications. Numbers in bold face indicate dominance.

finite sample performance in terms of power and size in the vast majority of non-linear models. The power is above 0.9 even for the sample size of $T = 50$ and it gets consistently closer to unity when the sample size increases to $T \geq 100$. However, the picture changes dramatically when the associated AR parameter (ρ) rises to 0.9 as the power drops sharply in small samples. It is important to note that the power loss is a finite sample phenomenon because the power performance improves with the sample size and approaches unity in most DGPs for the sufficiently large sample size, say $T = 500$. Interestingly, this is in line with the stylized fact established in linear models that unit-root tests have poor discriminatory power in finite samples against highly persistent linear processes.

Second, unlike our original expectations, the unit-root tests have certain discriminatory power for the models beyond the ones assumed in the alternative hypothesis. The ADF test has a satisfactory power performance against unit-root process even if the true underlying models are not linear. This is particularly true when ρ is modest or when ρ is high but T is large. For example, the probability that the ADF test rejects the false null of unit-root for the stationary EQ-TAR model with $\rho = 0.5$ is as high as 84% when the sample size is just $T = 100$. For the same model, the power gets close to unity for $T = 500$ even when ρ is as high as 0.9. In this context it is reasonable to posit that the general perception on the poor performance of the DF-type tests in the presence of non-linear models is driven by high ρ and small T. Although striking, this result is consistent with the findings by some previous studies. Balke and Fomby (1997), for instance, note that standard unit-root tests designed for linear case are valid asymptotically for non-linear models to the extent that error terms satisfy certain mixing conditions. Enders and Granger (1998) also report the robustness of the DF-type test against TAR models satisfying geometric ergodicity. A similar story is told from the other tests based on certain non-linear models under the alternative. The M-TAR and inf-*t* tests exhibit decent power of distinguishing unit-root process from non-linear stationary models other than TAR or transitional AR models stipulated in the alternatives. Consequently one should exercise considerable care when interpreting rejection of the unit-root null by these tests as a conclusive evidence of the specific models under the alternative because the rejection can be caused by a plethora of other models.

Third, among the five tests under comparison the ADF test and the inf-*t* test stand out. The ADF test dominates in the majority of DGPs when the sample size is relatively small, while the inf-*t* test is more powerful when the sample size is large. The good performance of the inf-*t* test is interesting in view of the fact that we set the transition function as ESTAR model throughout the experiments. The M-TAR test displays acceptable power performance for moderate ρ but is outperformed by the ADF-test for large ρ as echoed in Enders and Granger (1998). Moreover, the M-TAR test appears to suffer from undersize problem as it rejects the unit-root null far less frequently than it ought to. The sign test shows comparable but not dominant performance, while the KSS turns out to be outperformed by the other tests except for several cases in EQ-TAR model.

Fourth, all the tests commonly exhibit significantly low power against DGP 9 (EQ-TAR model), DGP 10 (Band-TAR model) and DGP 15 (Multiple structural break model) for relatively small sizes regardless of the magnitude of ρ. While the power appears to improve with the sample size (T) for DGPs 9 and 10, the poor discriminatory power sustains for DGP 15 even in relatively large sample size when ρ is high although the power would approach unity asymptotically as shown by Perron (1990). It is interesting to note that the sign test has certain discriminatory power in DGP 15 for moderate sample size. Therefore, inference drawn from the sign test is more reliable when the time series under study is suspected to have undergone several structural changes in the form of infrequent changes in the mean. We will revisit this issue in our empirical application in the upcoming section.

In sum, our simulation results suggest that what determines the power of unit-root tests is not the specific type of non-linearity in the alternative model, but how far the alternative model is away from the unit-root process. In contrast to the general perception that the standard unit-root tests are subject to have poor discriminatory power when they are applied to time series with non-linear dynamics, the discriminatory power of unit-root tests is reasonable in the vast majority of non-linear models under consideration when the associated AR parameter is mild, whereas it deteriorates sharply as the associated AR parameter is close to unity. However, this does not necessarily nullify the impact of nonlinearity on the performance of unit-root tests. For example, the power of all tests is consistently lower in some non-linear models such as DGPs 9, 10, 14 and 15 compared to DGP 1, their linear equivalent. Since the models take the same profile in terms of associated AR parameters, one should expect similar power performance across them if the AR parameters are solely responsible for the performance of unit-root tests. The lower power in those models thereby may indicate some potential role of nonlinearity, which prompts the question of in what situation nonlinearity matters to the performance of unit-root tests. This question is explored in the following section.

4. THE IMPACT OF NONLINEARITY

In dynamic models, a departure from linearity possibly through convexity, curvature, or kinkedness could alter the deviation from long run equilibrium and the extent of dynamics. In the case of structural change model, for instance, structural shifts are known to generate an upward bias on persistence measured by standard AR models that assume a stable mean. As noted by Perron (1989), the power problem of unit-root tests gets exacerbated because the shift in mean induces a bias of the autoregressive coefficient towards unity and thus makes it appear to be unit-root process.[6] As such, nonlinearity could be consequential to the performance of unit-root tests so long as it manifests itself by affecting the finite sample distribution of test statistics through a change in curvature or so. Despite its potential importance, however, not much is known about the potential connection between nonlinearity and the performance of unit-root tests and much less is known about the extent of the impact. In this section, we explore the potential impact of nonlinearity on the performance of unit-root tests focusing on the several non-linear models where the unit-root tests exhibited lower power than in the linear equivalent. Given that the power loss is a finite sample phenomenon, we follow much of the literature to provide simulation evidence rather than theoretical foundation.

4.1. Parameters relevant for nonlinearity

Though carefully designed, our experiment in the previous section is open to criticism as design-specific because it is conducted by fixing the values of other parameters than AR parameters. As often reported in the literature, however, performance of unit-root tests in non-linear dynamic

[6]Levin and Piger (2004) also report a persistence parameter for the United States GDP deflator of 0.92 over the period 1984:Q1–2003:Q4 without accounting for possible shifts, but persistence drops to 0.36 once a structural break is allowed for. Since these estimates of persistence are obtained from dynamic linear models which crucially depends on the assumption of specific long run equilibrium level, exaggeration in the degree of persistence is resulted from the violation of the assumption in the case of structural break models.

models also hinges upon other parameters that were held constant in our earlier experiments. Take DGPs 9 and 10 for example, Balke and Fomby (1997) report that not only the autoregressive parameters but also the constants in the threshold autoregression contribute to the power of standard tests. They also find that unit-root tests are more powerful against the Band-TAR model than the EQ-TAR model in spite of the identical autoregressive parameters in both models, which is in line with our results that the unit-root null is rejected less frequently in DGP 10 than in DGP 9. As such, since the evaluation of unit-root tests can vary with different values of parameters other than AR parameter, it would be informative to assess the performance of tests in wider parameter spaces.

In general, it is not straightforward to pin down the parameters dictating nonlinearity partly because they are often intermingled with other parameters like AR parameters. In bilinear model (DGP 3), for instance, ϕ is allegedly related to nonlinearity in the sense that the process degenerates to linear model if $\phi = 0$, but it is also a part of AR parameter. By contrast, in some models it is relatively easier to identify. For example, in EQ- and Band-TAR models (DGPs 9 and 10) where the AR parameter in the inner regime is unity while those in the outer regimes are less than unity, nonlinearity can be manifested by the difference in slopes between two regimes, or $(1 - \rho)$, which is dubbed by Sarno *et al.* (2004) as the threshold effect.[7] Likewise in SETAR model (DGP 8) where the behaviour of process is governed by the joint behaviour of the linear AR models on the two boundary threshold regimes, the difference between ρ_1 and ρ_2 reflects the kinkedness of the boundary and hence the departure from linearity. Here, the degree of nonlinearity can be measured by the *ratio* of the two AR parameters, $\zeta = \frac{\rho_1}{(\rho_1+\rho_2)}$, so that higher ratio implies less kinkedness in the sense that the ratio becomes unity when it is linear ($\rho_2 = 0$). The ratio can be also applicable to other transition models like STAR models (DGPs 11 and 12) and Markov-switching model (DGP 13). In STAR models, however, there is another source of nonlinearity. It is the curvature of transition function (θ) which is jointly decided by the smoothness parameter (γ) that determines the speed with which the transition from one regime and to the other takes place, as well as the threshold size (c). Given the value of γ, the threshold parameter (c) decides the distance between regimes and hence positively involve the degree of nonlinearity. But the role of γ is not that straightforward as it depends on the type of STAR model. ESTAR model collapses to a linear model either when $\gamma \to 0$ or $\gamma \to \infty$, while it is the case for LSTAR only when $\gamma \to 0$.[8] In Markov-switching model, nonlinearity also stems from the transition matrix of the Markov chain which is governed by the transition probabilities, (P_{11}, P_{22}) where P_{ij} denotes the probability that the process switches from regime i to regime j.[9]

To investigate the impact of these parameters on the performance of unit-root tests, we conduct another simulation experiment. If nonlinearity really matters, the performance of unit-root tests should be responsive to the values of relevant parameters. Table 3 reports the power performance of unit-root tests in four transition models where the degree of nonlinearity is measured by $\zeta = \frac{\rho_1}{(\rho_1+\rho_2)}$. We consider several sets of parameter values, $(\rho_1, \rho_2) = \{(0.3, 0.3), (0.4, 0.2), (0.4, 0.4), (0.6, 0.2), (0.5, 0.4), (0.6, 0.3), (0.8, 0.1)\}$ such that the corresponding ratio (ζ) is equal to

[7]In these TAR models, a tradeoff exists between the degree of nonlinearity and the magnitude of AR parameter in the outer regimes so that an increase in persistence in the outer regimes lowers the degree of nonlinearity.

[8]When $\gamma \to \infty$, LSTAR model nests a TAR model where the transition occurs abruptly rather than smoothly.

[9]As summarized by Yang (2000, p. 31–32), the process collapses to a single AR(1) process if (P_{11}, P_{22})=(0,1) or (1,0), while it becomes two AR(1) regimes that never interact with probability 1 when (P_{11}, P_{22}) = (1,1). If (P_{11}, P_{22}) = (0,0), the behaviour of the Markov chain is periodic and switches regimes every period with probability 1.

Table 3. Rejection rates and the degree of nonlinearity (DGPs 8,11,12,13).

(ρ_1, ρ_2)	DGP/	ADF test				Sign test				MTAR test				KSS test				inf $-$ t test			
	/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
(0.3,0.3)	8	0.93	0.98	1.00	1.00	0.79	0.99	1.00	1.00	0.96	1.00	1.00	1.00	0.53	0.79	0.86	0.93	0.91	1.00	1.00	1.00
$\zeta = 1/2$	11	0.90	0.97	1.00	1.00	0.66	0.96	1.00	1.00	0.77	1.00	1.00	1.00	0.46	0.78	0.89	0.97	0.76	1.00	1.00	1.00
	12	0.95	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.42	0.69	0.72	0.72	0.98	1.00	1.00	1.00
	13	0.94	0.98	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.53	0.73	0.78	0.87	0.98	1.00	1.00	1.00
(0.4,0.2)	8	0.93	0.97	1.00	1.00	0.77	0.99	1.00	1.00	0.94	1.00	1.00	1.00	0.51	0.78	0.86	0.94	0.88	1.00	1.00	1.00
$\zeta = 2/3$	11	0.90	0.97	1.00	1.00	0.66	0.96	1.00	1.00	0.77	1.00	1.00	1.00	0.46	0.78	0.89	0.97	0.76	1.00	1.00	1.00
	12	0.95	0.98	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.42	0.70	0.73	0.74	0.97	1.00	1.00	1.00
	13	0.93	0.98	1.00	1.00	0.90	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.51	0.73	0.78	0.87	0.96	1.00	1.00	1.00
(0.4,0.4)	8	0.82	0.95	0.99	1.00	0.55	0.90	0.99	1.00	0.61	0.99	1.00	1.00	0.43	0.73	0.87	0.95	0.65	0.99	1.00	1.00
$\zeta = 1/2$	11	0.55	0.90	0.98	1.00	0.36	0.74	0.95	1.00	0.21	0.72	1.00	1.00	0.28	0.57	0.83	0.97	0.35	0.81	1.00	1.00
	12	0.94	0.98	1.00	1.00	0.92	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.38	0.64	0.72	0.74	0.92	1.00	1.00	1.00
	13	0.93	0.97	1.00	1.00	0.88	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.51	0.74	0.80	0.89	0.95	1.00	1.00	1.00

Table 3. Continued.

(ρ_1, ρ_2)	DGP/	ADF test				Sign test				MTAR test				KSS test				inf − t test			
	/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
(0.6,0.2)	8	0.74	0.95	0.99	1.00	0.49	0.87	0.99	1.00	0.44	0.97	1.00	1.00	0.37	0.68	0.86	0.95	0.55	0.96	1.00	1.00
$\zeta = 3/4$	11	0.55	0.90	0.98	1.00	0.36	0.74	0.95	1.00	0.21	0.72	1.00	1.00	0.29	0.57	0.83	0.97	0.36	0.81	1.00	1.00
	12	0.93	0.97	1.00	1.00	0.84	0.99	1.00	1.00	0.92	1.00	1.00	1.00	0.36	0.63	0.74	0.78	0.84	1.00	1.00	1.00
	13	0.90	0.96	1.00	1.00	0.79	0.99	1.00	1.00	0.86	1.00	1.00	1.00	0.44	0.71	0.81	0.90	0.83	1.00	1.00	1.00
(0.5,0.4)	8	0.54	0.86	0.97	1.00	0.35	0.69	0.91	1.00	0.25	0.72	1.00	1.00	0.33	0.60	0.79	0.93	0.40	0.84	1.00	1.00
$\zeta = 5/9$	11	0.29	0.53	0.90	1.00	0.21	0.46	0.71	0.98	0.08	0.20	0.71	1.00	0.19	0.33	0.62	0.93	0.17	0.39	0.84	1.00
	12	0.92	0.97	1.00	1.00	0.84	0.99	1.00	1.00	0.90	1.00	1.00	1.00	0.33	0.58	0.69	0.74	0.82	1.00	1.00	1.00
	13	0.92	0.97	1.00	1.00	0.82	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.48	0.73	0.80	0.90	0.90	1.00	1.00	1.00
(0.6,0.3)	8	0.50	0.84	0.97	1.00	0.33	0.67	0.90	1.00	0.22	0.66	1.00	1.00	0.30	0.57	0.78	0.93	0.35	0.79	1.00	1.00
$\zeta = 6/9$	11	0.29	0.53	0.90	1.00	0.21	0.46	0.71	0.98	0.08	0.20	0.71	1.00	0.19	0.33	0.62	0.93	0.17	0.39	0.84	1.00
	12	0.90	0.97	1.00	1.00	0.80	0.99	1.00	1.00	0.84	1.00	1.00	1.00	0.31	0.58	0.71	0.75	0.76	0.99	1.00	1.00
	13	0.90	0.96	1.00	1.00	0.77	0.99	1.00	1.00	0.85	1.00	1.00	1.00	0.44	0.71	0.81	0.90	0.83	1.00	1.00	1.00
(0.8,0.1)	8	0.38	0.72	0.95	1.00	0.27	0.57	0.84	0.99	0.12	0.40	0.95	1.00	0.23	0.45	0.73	0.95	0.25	0.59	0.98	1.00
$\zeta = 8/9$	11	0.29	0.53	0.90	1.00	0.21	0.46	0.71	0.98	0.08	0.20	0.71	1.00	0.19	0.33	0.62	0.93	0.17	0.39	0.84	1.00
	12	0.78	0.95	0.99	1.00	0.57	0.93	1.00	1.00	0.52	0.97	1.00	1.00	0.29	0.54	0.75	0.84	0.57	0.96	1.00	1.00
	13	0.69	0.92	0.99	1.00	0.59	0.91	0.99	1.00	0.45	0.91	1.00	1.00	0.35	0.58	0.78	0.91	0.54	0.91	1.00	1.00

Note: Entries represent the fraction of times when the null hypothesis is rejected out of 5,000 replications at the 10% significance level. See the notes in Table 2 for the values of parameters set in simulations.

DGP8 : $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-1} I(y_{t-1} \geq c) + \varepsilon_t$.

DGP11 : $y_t = \alpha + \rho_1 y_{t-1} + \theta \cdot (\beta + \rho_2 y_{t-1}) + \varepsilon_t$ where $\theta = 1 - e^{-\gamma (y_{t-1} - c)^2}$.

DGP12 : $y_t = \alpha + \rho_1 y_{t-1} + \theta \cdot (\beta + \rho_2 y_{t-1}) + \varepsilon_t$ where $\theta = [1 + e^{-\gamma (y_{t-1} - c)}]^{-1}$.

DGP13 : $y_t = \rho_t y_{t-1} + e_t, \ e_t \sim i.i.d.(0, 0.4)$ where $\rho_t = \rho_1 S_t + \rho_2 (1 - S_t)$.

{1/2, 2/3, 1/2, 3/4, 5/9, 2/3, 8/9}. In all models considered, the power of unit-root tests is highly sensitive to the associated AR parameters in particular ρ_1, but not much to the nonlinearity ratio (ζ). For the last three sets of parameters which take the same values for ($\rho_1 + \rho_2$) but different values for ρ_1, the power of tests decreases with ρ_1 while the departure from linearity in fact gets smaller. This reinforces our prior finding that power performance hinges upon AR parameter rather than kinkedness of slopes. We further investigate the performance of unit-root tests in STAR models (DGPs 11 and 12) by looking at several choices of c and γ that impinge on the curvature of transition function and hence nonlinearity. As presented in Table 4, the power changes dramatically with ρ_1 and ρ_2 but little with c and γ when ρ_1 and ρ_2 are fixed. Interestingly, c and γ exert bigger impact on the power when AR parameters are larger, implying that the impact of nonlinearity depends on the magnitude of AR parameter. A similar story is told in Table 5 which gives the results from Markov-switching models. The power performance hardly responds to the different probabilities in the transition matrix, but substantially to the associated AR parameters.

Table 6 summarizes the results for EQ- and Band-TAR models in which the roots outside of the band are chosen from the set of {0.1, 0.3, 0.5, 0.7, 0.9} so that the consequent threshold effects are {0.9, 0.7, 0.5, 0.3, 0.1.} As can be seen from Table 6, the power of tests decreases with the magnitude of outer roots but increases with the difference in slopes between the two regimes that represents the departure from linearity. Our results therefore confirm the finding by earlier studies [e.g. Balke and Fomby (1996) and Sarno *et al.* (2004)] that the power is affected more by the variations in the outer root (ρ) than by the threshold effect ($1 - \rho$). In the Band-TAR model (DGP 10), however, we find that the power of tests is inversely affected by the value of threshold parameter (k), as previously reported by Pippenger and Goering (1993), Balke and Fomby (1997) and Park and Shintani (2005). This is well expected because larger threshold parameter implies broader zone for unit-root process, and hence unit-root behaviour is observed more often. It is worth noting that the inf-t test dominates in terms of power performance regardless of the threshold band except when ρ is large in small samples where it is marginally outperformed by the ADF test.

In the models with structural break as presented in Table 7, inference on unit roots is highly affected not just by AR parameters but by the size of mean shifts and the timing of break point (e.g. Lee 2000). The ADF test tends to substantively underreject the unit-root null even when the mean shift is relatively small, as reported by numerous authors. By contrast, the sign, M-TAR and inf-t tests display reasonable power performance for mild AR parameter especially when the sample size is large. However, their power performance deteriorates rapidly as the associated AR parameter increases irrespective of sample size. A remarkable exception is the sign test which retains decent power in large sample size even when the AR parameter is large. This suggests that inference drawn from the sign test is more reliable when the time series under study is suspected to have undergone several structural changes in the form of infrequent changes in the mean.

Overall, our results in this section suggest that nonlinearity may play only a supplementary role in the power performance of unit-root tests while it is primarily influenced by the magnitude of associated AR parameters. Major exceptions include Band-TAR models and structural break models in which low power is observed even when the associated AR parameters are rather mild.

4.2. Nonlinearity and stationarity conditions

A process is said to be nonstationarity if it violates certain conditions for stationarity. In linear autoregressive model, for example, the stationarity condition concerns that the roots of the

Table 4. Rejection rates in STAR models (DGPs 11 and 12).

c	DGP/	ADF test				Sign test				MTAR test				KSS test				inf $-$ t test			
	/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
$\rho_1 = 0.3$; $\rho_2 = 0.3$; $\gamma = 100$																					
c = −5	11	0.90	0.97	0.99	1.00	0.68	0.97	1.00	1.00	0.36	0.57	0.77	1.00	0.45	0.76	0.83	0.95	0.77	1.00	1.00	1.00
	12	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.12	0.14	0.12	0.46	1.00	1.00	1.00	1.00
c = 0	11	0.90	1.00	1.00	1.00	0.68	0.97	1.00	1.00	0.36	1.00	1.00	1.00	0.45	0.76	0.87	0.95	0.77	1.00	1.00	1.00
	12	0.93	1.00	1.00	1.00	0.79	0.97	1.00	1.00	0.99	1.00	1.00	1.00	0.76	0.14	0.86	0.93	0.91	1.00	1.00	1.00
c = 5	11	0.90	0.97	1.00	1.00	0.68	0.97	1.00	1.00	0.34	0.55	0.83	1.00	0.45	0.76	0.86	0.95	0.77	1.00	1.00	1.00
	12	0.90	0.97	1.00	1.00	0.68	0.97	1.00	1.00	0.34	0.55	0.83	1.00	0.45	0.76	0.86	0.95	0.77	1.00	1.00	1.00
$\rho_1 = 0.8$; $\rho_2 = 0.1$; $\gamma = 100$																					
c = −5	11	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00
	12	0.96	0.99	1.00	1.00	0.61	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.93	0.96	0.99	1.00	1.00	1.00	1.00
c = 0	11	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00
	12	0.38	0.72	0.95	1.00	0.27	0.57	0.84	0.99	0.13	0.41	0.96	1.00	0.23	0.45	0.73	0.95	0.25	0.59	0.98	1.00
c = 5	11	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00
	12	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00

Table 4. Continued

c	DGP/	ADF test				Sign test				MTAR test				KSS test				inf − t test			
	/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
										$\rho_1 = 0.3; \rho_2 = 0.3; c = 0$											
$\gamma = 1$	11	0.90	0.97	1.00	1.00	0.68	0.97	1.00	1.00	0.36	0.58	0.86	1.00	0.45	0.76	0.86	0.95	0.77	1.00	1.00	1.00
	12	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.69	0.71	0.71	1.00	1.00	1.00	1.00
$\gamma = 100$	11	0.90	0.97	1.00	1.00	0.68	0.97	1.00	1.00	0.36	0.58	0.86	1.00	0.45	0.76	0.86	0.95	0.77	1.00	1.00	1.00
	12	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.27	0.44	0.48	0.52	1.00	1.00	1.00	1.00
$\gamma = 10000$	11	0.90	0.97	1.00	1.00	0.68	0.97	1.00	1.00	0.36	0.58	0.86	1.00	0.45	0.76	0.86	0.95	0.77	1.00	1.00	1.00
	12	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.27	0.45	0.47	0.52	1.00	1.00	1.00	1.00
										$\rho_1 = 0.8; \rho_2 = 0.1; c = 0$											
$\gamma = 1$	11	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00
	12	0.93	0.97	1.00	1.00	0.78	0.99	1.00	1.00	0.95	1.00	1.00	1.00	0.53	0.77	0.84	0.91	0.89	1.00	1.00	1.00
$\gamma = 100$	11	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00
	12	0.95	0.97	1.00	1.00	0.64	0.93	1.00	1.00	0.99	1.00	1.00	1.00	0.78	0.90	0.93	0.98	0.99	1.00	1.00	1.00
$\gamma = 10000$	11	0.29	0.54	0.91	1.00	0.21	0.46	0.72	0.98	0.08	0.22	0.73	1.00	0.19	0.34	0.63	0.94	0.18	0.40	0.85	1.00
	12	0.95	0.98	1.00	1.00	0.63	0.93	1.00	1.00	0.99	1.00	1.00	1.00	0.79	0.90	0.93	0.98	0.99	1.00	1.00	1.00

Note: Entries represent the fraction of times when the null hypothesis is rejected out of 5,000 replications at the 10% significance level.

$$\text{DGP11}: \quad y_t = \alpha + \rho_1 y_{t-1} + \theta \cdot (\beta + \rho_2 y_{t-1}) + \varepsilon_t \quad \text{where } \theta = 1 - e^{-\gamma(y_{t-1}-c)^2}.$$

$$\text{DGP12}: \quad y_t = \alpha + \rho_1 y_{t-1} + \theta \cdot (\beta + \rho_2 y_{t-1}) + \varepsilon_t \quad \text{where } \theta = [1 + e^{-\gamma(y_{t-1}-c)}]^{-1}$$

Table 5. Rejection rates in Markov-switching model (DGP 13).

p_{11}	p_{22}	ADF test				Sign test				MTAR test				KSS test				inf − t test			
	/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
										$\rho_1 = 0.3$; $\rho_2 = 0.3$											
0.6	0.5	0.95	0.98	1.00	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.76	0.80	0.86	0.98	1.00	1.00	1.00
	0.6	0.95	0.98	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.49	0.75	0.80	0.85	0.98	1.00	1.00	1.00
	0.7	0.95	0.98	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.49	0.73	0.79	0.85	0.98	1.00	1.00	1.00
	0.8	0.95	0.98	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.47	0.72	0.79	0.83	0.98	1.00	1.00	1.00
	0.9	0.95	0.98	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.46	0.68	0.76	0.82	0.98	1.00	1.00	1.00
0.95	0.5	0.94	0.98	1.00	1.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.54	0.79	0.82	0.89	0.98	1.00	1.00	1.00
	0.6	0.95	0.98	1.00	1.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.55	0.79	0.82	0.89	0.98	1.00	1.00	1.00
	0.7	0.95	0.98	1.00	1.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.55	0.78	0.81	0.89	0.98	1.00	1.00	1.00
	0.8	0.95	0.98	1.00	1.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.55	0.77	0.81	0.88	0.98	1.00	1.00	1.00
	0.9	0.94	0.98	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.53	0.73	0.78	0.87	0.98	1.00	1.00	1.00
										$\rho_1 = 0.5$; $\rho_2 = 0.4$											
0.6	0.5	0.94	0.97	1.00	1.00	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.49	0.78	0.85	0.93	0.90	1.00	1.00	1.00
	0.6	0.93	0.97	1.00	1.00	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.49	0.78	0.85	0.92	0.89	1.00	1.00	1.00
	0.7	0.94	0.97	1.00	1.00	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.78	0.84	0.92	0.90	1.00	1.00	1.00
	0.8	0.93	0.97	1.00	1.00	0.80	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.77	0.83	0.92	0.89	1.00	1.00	1.00
	0.9	0.92	0.97	1.00	1.00	0.80	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.48	0.73	0.81	0.91	0.89	1.00	1.00	1.00
0.95	0.5	0.94	0.98	1.00	1.00	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.43	0.75	0.82	0.89	0.90	1.00	1.00	1.00
	0.6	0.94	0.97	1.00	1.00	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.43	0.73	0.81	0.89	0.90	1.00	1.00	1.00
	0.7	0.93	0.97	1.00	1.00	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.42	0.71	0.81	0.88	0.90	1.00	1.00	1.00
	0.8	0.93	0.98	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.42	0.70	0.80	0.87	0.90	1.00	1.00	1.00
	0.9	0.94	0.98	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.42	0.65	0.77	0.85	0.90	1.00	1.00	1.00

Table 5. Continued

p_{11}	p_{22}	ADF test				Sign test				MTAR test				KSS test				inf − *t* test			
	/T	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
										$\rho_1 = 0.8; \rho_2 = 0.1$											
0.6	0.5	0.93	0.97	1.00	1.00	0.82	0.99	1.00	1.00	0.92	1.00	1.00	1.00	0.59	0.80	0.86	0.92	0.91	1.00	1.00	1.00
	0.6	0.93	0.97	1.00	1.00	0.83	0.99	1.00	1.00	0.93	1.00	1.00	1.00	0.57	0.79	0.85	0.91	0.91	1.00	1.00	1.00
	0.7	0.94	0.97	1.00	1.00	0.86	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.58	0.76	0.84	0.89	0.92	1.00	1.00	1.00
	0.8	0.94	0.98	1.00	1.00	0.91	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.54	0.74	0.82	0.86	0.93	1.00	1.00	1.00
	0.9	0.94	0.98	1.00	1.00	0.94	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.48	0.66	0.76	0.78	0.94	1.00	1.00	1.00
0.95	0.5	0.65	0.93	0.99	1.00	0.45	0.83	0.98	1.00	0.33	0.88	1.00	1.00	0.36	0.63	0.84	0.95	0.47	0.91	1.00	1.00
	0.6	0.66	0.92	0.99	1.00	0.46	0.84	0.98	1.00	0.34	0.88	1.00	1.00	0.36	0.63	0.84	0.95	0.47	0.91	1.00	1.00
	0.7	0.66	0.92	0.99	1.00	0.48	0.85	0.98	1.00	0.35	0.88	1.00	1.00	0.36	0.63	0.84	0.95	0.48	0.91	1.00	1.00
	0.8	0.67	0.92	0.99	1.00	0.51	0.88	0.99	1.00	0.37	0.89	1.00	1.00	0.36	0.62	0.83	0.94	0.50	0.91	1.00	1.00
	0.9	0.69	0.92	0.99	1.00	0.59	0.91	0.99	1.00	0.44	0.90	1.00	1.00	0.35	0.58	0.78	0.91	0.54	0.91	1.00	1.00

Note: Entries represent the fraction of times when the null hypothesis is rejected out of 5,000 replications at the 10% significance level. DGP is

$$y_t = \rho_t y_{t-1} + e_t, \quad e_t \sim i.i.d.(0, 0.4) \text{ where } \rho_t = \rho_1 S_t + \rho_2(1 - S_t).$$

Table 6. Rejection rates in TAR models (DGPs 9 and 10).

ρ	DGP	ADF test				Sign test				MTAR test				KSS test				inf − t test			
		50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
										$k = 1$											
0.1	9	0.95	0.98	1.00	1.00	0.80	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.82	0.93	0.98	1.00	1.00	1.00	1.00	1.00
	10	0.89	0.96	1.00	1.00	0.48	0.84	0.97	1.00	0.79	1.00	1.00	1.00	0.73	0.96	1.00	1.00	0.86	1.00	1.00	1.00
0.3	9	0.93	0.98	1.00	1.00	0.73	0.97	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.92	0.97	1.00	0.98	1.00	1.00	1.00
	10	0.83	0.95	1.00	1.00	0.44	0.79	0.96	1.00	0.57	1.00	1.00	1.00	0.62	0.92	0.99	1.00	0.73	1.00	1.00	1.00
0.5	9	0.92	0.97	1.00	1.00	0.63	0.93	1.00	1.00	0.90	1.00	1.00	1.00	0.62	0.88	0.95	1.00	0.88	1.00	1.00	1.00
	10	0.68	0.93	0.99	1.00	0.37	0.74	0.93	1.00	0.33	0.96	1.00	1.00	0.48	0.85	0.97	1.00	0.53	0.98	1.00	1.00
0.7	9	0.75	0.95	0.99	1.00	0.46	0.84	0.98	1.00	0.43	0.98	1.00	1.00	0.40	0.76	0.91	0.99	0.56	0.98	1.00	1.00
	10	0.43	0.85	0.97	1.00	0.29	0.60	0.85	1.00	0.15	0.58	1.00	1.00	0.31	0.67	0.93	1.00	0.30	0.80	1.00	1.00
0.9	9	0.27	0.51	0.90	1.00	0.20	0.44	0.70	0.97	0.07	0.19	0.70	1.00	0.18	0.34	0.64	0.94	0.16	0.38	0.86	1.00
	10	0.22	0.34	0.73	1.00	0.16	0.34	0.53	0.88	0.06	0.10	0.37	1.00	0.15	0.26	0.56	0.95	0.13	0.26	0.66	1.00
										$k = 3$											
0.1	9	0.81	0.94	0.99	1.00	0.16	0.31	0.42	0.71	0.57	1.00	1.00	1.00	0.93	1.00	1.00	1.00	0.90	1.00	1.00	1.00
	10	0.31	0.53	0.87	1.00	0.19	0.36	0.50	0.83	0.14	0.25	0.89	1.00	0.35	0.68	0.96	1.00	0.28	0.79	1.00	1.00
0.3	9	0.66	0.90	0.98	1.00	0.17	0.34	0.46	0.77	0.36	0.96	1.00	1.00	0.83	0.98	1.00	1.00	0.76	1.00	1.00	1.00
	10	0.29	0.45	0.85	1.00	0.19	0.36	0.50	0.82	0.13	0.21	0.77	1.00	0.29	0.60	0.93	1.00	0.24	0.68	1.00	1.00
0.5	9	0.47	0.84	0.96	1.00	0.20	0.38	0.53	0.84	0.23	0.77	1.00	1.00	0.64	0.93	1.00	1.00	0.53	0.99	1.00	1.00
	10	0.27	0.38	0.81	1.00	0.18	0.34	0.48	0.81	0.11	0.17	0.57	1.00	0.23	0.48	0.87	1.00	0.20	0.53	0.99	1.00
0.7	9	0.35	0.72	0.93	1.00	0.21	0.40	0.59	0.91	0.16	0.39	0.99	1.00	0.38	0.77	0.98	1.00	0.31	0.89	1.00	1.00
	10	0.24	0.32	0.67	0.99	0.17	0.31	0.45	0.77	0.08	0.13	0.32	1.00	0.18	0.33	0.73	0.99	0.16	0.34	0.87	1.00
0.9	9	0.24	0.34	0.74	1.00	0.17	0.33	0.50	0.83	0.07	0.12	0.38	1.00	0.16	0.30	0.65	0.98	0.15	0.31	0.79	1.00
	10	0.19	0.21	0.34	0.90	0.12	0.23	0.33	0.60	0.05	0.07	0.12	0.66	0.12	0.18	0.34	0.87	0.11	0.17	0.34	0.97

Table 6. Continued.

ρ	DGP	ADF test				Sign test				MTAR test				KSS test				inf − t test			
		50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
										$k = 5$											
0.1	9	0.38	0.77	0.95	1.00	0.11	0.17	0.19	0.28	0.19	0.52	0.99	1.00	0.76	0.95	1.00	1.00	0.69	0.97	1.00	1.00
	10	0.25	0.28	0.41	0.92	0.14	0.25	0.30	0.52	0.11	0.15	0.22	0.97	0.23	0.34	0.67	0.98	0.20	0.38	0.83	1.00
0.3	9	0.32	0.59	0.90	1.00	0.11	0.18	0.20	0.30	0.17	0.33	0.95	1.00	0.67	0.91	0.99	1.00	0.57	0.95	1.00	1.00
	10	0.24	0.27	0.38	0.91	0.14	0.25	0.31	0.53	0.09	0.14	0.20	0.91	0.21	0.31	0.60	0.98	0.18	0.33	0.73	1.00
0.5	9	0.30	0.42	0.79	1.00	0.11	0.19	0.23	0.37	0.16	0.25	0.70	1.00	0.52	0.78	0.97	1.00	0.41	0.86	1.00	1.00
	10	0.23	0.26	0.35	0.88	0.14	0.25	0.31	0.53	0.08	0.12	0.18	0.79	0.18	0.27	0.51	0.96	0.16	0.28	0.60	1.00
0.7	9	0.27	0.32	0.64	0.98	0.15	0.23	0.30	0.51	0.13	0.19	0.38	1.00	0.32	0.54	0.86	1.00	0.26	0.59	0.99	1.00
	10	0.21	0.24	0.31	0.83	0.13	0.24	0.31	0.52	0.07	0.10	0.15	0.55	0.14	0.22	0.40	0.91	0.14	0.23	0.42	1.00
0.9	9	0.23	0.26	0.38	0.92	0.15	0.26	0.35	0.60	0.07	0.11	0.17	0.88	0.15	0.23	0.48	0.96	0.14	0.25	0.57	1.00
	10	0.17	0.19	0.25	0.53	0.11	0.19	0.25	0.42	0.04	0.05	0.09	0.20	0.12	0.15	0.24	0.64	0.11	0.15	0.23	0.75
										$k = 7$											
0.1	9	0.25	0.43	0.83	1.00	0.09	0.14	0.14	0.18	0.12	0.22	0.63	1.00	0.59	0.85	0.98	1.00	0.54	0.86	0.99	1.00
	10	0.22	0.24	0.29	0.60	0.13	0.21	0.24	0.37	0.09	0.12	0.17	0.31	0.20	0.27	0.39	0.85	0.17	0.29	0.42	0.99
0.3	9	0.26	0.34	0.67	0.98	0.09	0.15	0.15	0.19	0.13	0.21	0.42	1.00	0.55	0.77	0.96	1.00	0.48	0.80	0.99	1.00
	10	0.21	0.23	0.28	0.56	0.12	0.21	0.24	0.37	0.08	0.11	0.16	0.27	0.18	0.25	0.36	0.82	0.16	0.26	0.38	0.99
0.5	9	0.26	0.30	0.45	0.93	0.09	0.15	0.16	0.21	0.13	0.19	0.28	0.99	0.44	0.63	0.88	1.00	0.36	0.64	0.96	1.00
	10	0.20	0.23	0.27	0.51	0.13	0.21	0.25	0.38	0.07	0.10	0.14	0.24	0.15	0.22	0.32	0.76	0.14	0.23	0.34	0.97
0.7	9	0.24	0.27	0.34	0.82	0.11	0.18	0.19	0.28	0.11	0.16	0.22	0.76	0.31	0.41	0.68	0.98	0.25	0.42	0.79	1.00
	10	0.19	0.22	0.26	0.43	0.12	0.21	0.25	0.38	0.06	0.08	0.13	0.19	0.13	0.19	0.28	0.67	0.12	0.20	0.30	0.87
0.9	9	0.21	0.23	0.29	0.64	0.13	0.22	0.26	0.42	0.07	0.10	0.15	0.32	0.15	0.22	0.34	0.83	0.14	0.23	0.36	0.99
	10	0.16	0.17	0.22	0.30	0.10	0.17	0.21	0.32	0.04	0.05	0.08	0.13	0.11	0.14	0.20	0.39	0.10	0.14	0.20	0.42

Note: Entries represent the fraction of times when the null hypothesis is rejected out of 5,000 replications at the 10% significance level.

$$\text{DGP9}: \quad y_t = y_{t-1} + \varepsilon_t, \quad \text{if } |y_{t-1}| \leq k; \quad y_t = \rho y_{t-1} + \varepsilon_t, \text{ if } |y_{t-1}| > k,$$

$$\text{DGP10}: \quad y_t = k(1-\rho) + \rho y_{t-1} + \varepsilon_t, \text{ if } y_{t-1} > k; \quad y_t = y_{t-1} + \varepsilon_t, \quad \text{if } |y_{t-1}| \leq k; y_t = -k(1-\rho) + \rho y_{t-1} + \varepsilon_t, \quad \text{if } y_{t-1} < -k.$$

Table 7. Rejection rates in structural break model (DGP 15).

$[\alpha_2, \alpha_3]$	(λ_1, λ_2)	ADF test				Sign test				MTAR test				KSS test				inf − t test			
		50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
										$\rho = 0.5$											
(−0.5, 1.5)	(1/3, 1/3)	0.34	0.36	0.44	0.50	0.32	0.75	0.96	1.00	0.14	0.83	1.00	1.00	0.19	0.35	0.34	0.44	0.18	0.77	1.00	1.00
	(1/4, 3/4)	0.29	0.21	0.17	0.14	0.75	0.99	1.00	1.00	0.58	1.00	1.00	1.00	0.04	0.02	0.01	0.00	0.15	0.80	1.00	1.00
	(2/5, 3/5)	0.21	0.23	0.18	0.16	0.45	0.89	1.00	1.00	0.13	0.86	1.00	1.00	0.15	0.21	0.20	0.22	0.13	0.73	1.00	1.00
(−1.0, 2.0)	(1/3, 1/3)	0.10	0.26	0.17	0.14	0.21	0.63	0.91	1.00	0.02	0.31	1.00	1.00	0.15	0.28	0.30	0.42	0.08	0.61	1.00	1.00
	(1/4, 3/4)	0.08	0.05	0.04	0.02	0.71	0.99	1.00	1.00	0.27	0.98	1.00	1.00	0.01	0.00	0.00	0.00	0.07	0.68	1.00	1.00
	(2/5, 3/5)	0.03	0.09	0.05	0.02	0.38	0.85	0.99	1.00	0.01	0.32	1.00	1.00	0.09	0.16	0.18	0.21	0.05	0.59	1.00	1.00
(−1.5, 3.0)	(1/3, 1/3)	0.01	0.01	0.02	0.00	0.15	0.54	0.87	1.00	0.00	0.00	0.48	1.00	0.10	0.19	0.26	0.43	0.02	0.45	1.00	1.00
	(1/4, 3/4)	0.00	0.00	0.00	0.00	0.67	0.99	1.00	1.00	0.03	0.67	1.00	1.00	0.00	0.00	0.00	0.00	0.01	0.47	1.00	1.00
	(2/5, 3/5)	0.00	0.00	0.00	0.00	0.33	0.83	0.99	1.00	0.00	0.00	0.36	1.00	0.03	0.10	0.18	0.23	0.01	0.39	1.00	1.00
										$\rho = 0.9$											
(−0.5, 1.5)	(1/3, 1/3)	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.34	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.18	0.02	0.00	0.00	0.46
	(1/4, 3/4)	0.00	0.00	0.00	0.00	0.12	0.56	0.93	1.00	0.04	0.03	0.05	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.59
	(2/5, 3/5)	0.00	0.00	0.00	0.00	0.01	0.10	0.39	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.44
(−1.0, 2.0)	(1/3, 1/3)	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.02	0.00	0.00	0.05
	(1/4, 3/4)	0.00	0.00	0.00	0.00	0.07	0.45	0.91	1.00	0.06	0.01	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
	(2/5, 3/5)	0.00	0.00	0.00	0.00	0.00	0.05	0.31	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
(−1.5, 3.0)	(1/3, 1/3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	(1/4, 3/4)	0.00	0.00	0.00	0.00	0.03	0.30	0.85	1.00	0.22	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(2/5, 3/5)	0.00	0.00	0.00	0.00	0.00	0.02	0.19	0.91	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: Entries represent the fraction of times when the null hypothesis is rejected out of 5,000 replications at the 10% significance level. $\alpha_1 = 0$ in

$$\begin{aligned} y_t &= \alpha_1 + \rho y_{t-1} + \varepsilon_t, \quad \text{if } t \le \lambda_1 T. \\ &= \alpha_2 + \rho y_{t-1} + \varepsilon_t, \quad \text{if } \lambda_1 T < t \le \lambda_2 T. \\ &= \alpha_3 + \rho y_{t-1} + \varepsilon_t, \quad \text{if } \lambda_2 T < t \le T. \end{aligned}$$

autoregressive polynomial lie outside the unit circle in the complex plane. By analogy, a non-linear process is subject to lead to the rejection of unit-root null provided that it satisfies stationarity conditions.

In many non-linear models, the conditions that guarantee stationarity are not available with the exceptions of several parametric non-linear models such as bilinear model, TAR model and Markov-switching model.[10] In Bilinear model and Markov-switching model, stationarity conditions hinge on the associated AR parameters as well as the parameters pertaining to nonlinearity. As a result, nonlinearity can exert certain influence on the performance of unit-root tests when it leads to the violation of the conditions. Take a bilinear model (DGP 3) for example, nonlinearity is approximated via the second-order Taylor expansion, and the necessary condition for stationarity is shown by Granger and Andersen (1978) as $\rho^2 + \phi^2\sigma_e^2 < 1$ where the process nests to linear if $\phi = 0$. Also in Markov-switching model (DGP 13) where dynamics change in accordance with a non-observed Markov chain, stationarity conditions depend on the characteristic roots as well as the transition matrix of the Markov chain. As illustrated by Yang (2000, p. 31–32), the condition for boundedness and stationarity in the case of Markov-switching in AR coefficients is given as $\max_{i=1,2}\{P_{i1}|\rho_1|^2 + P_{i2}|\rho_2|^2\}$ where P_{ij} denotes the transition probability from regime i to j. In these cases, the parameters associated with nonlinearity can affect the performance of unit-root tests even asymptotically through stationarity conditions.

In some other non-linear models, the stationarity condition is primarily governed by the associated AR parameters. For instance, the sufficient condition for global stationarity in SETAR model is known as the roots of the autoregression in the outer regime should be less than unity in absolute value [e.g. Tong (1990), Granger and Teräsvirta (1993), Sarno *et al.* (2004), amongst many others]. Similarly in TAR model with identical symmetric threshold autoregression, a sufficient condition for stationarity is that the roots of the autoregression in the outer regimes are less than unity in absolute value (e.g. Tjøstheim 1990).[11] Since nonlinearity does play little role in the stationarity condition in these models, the properties of unit-root tests will be hardly affected by nonlinearity no matter whether we take into account the specific non-linear structure in the construction of the tests in the previous sections.

5. EMPIRICAL APPLICATION

The time series property of real interest rates has been the subject of a substantial amount of research in macroeconomics and finance owing to its important implication in the theoretical models such as the consumption-based asset pricing and the Fisher hypothesis. Although theoretical models in general predict the mean reversion of real interest rates, empirical evidence

[10]Refer to Chan *et al*. (1985, theorem 2.1 in p. 270), Liu and Susko (1992), Tjøstheim (1990), for TAR model, and Yang (2000) for general conditions of the ergodic behaviour of bilinear model, TAR model, and Markov–switching model. See also Carrasco (2002, p. 243–246) for the stationarity conditions of the structural change AR (SCA) model, Markov–switching AR model and TAR model. Kristensen (2005) documents the conditions of geometrically ergodic (β-mixing) for a general non-linear Markov model. As noted by Altissimo and Violante (2001, p. 465), they commonly rely on Markov chain theory to verify the existence of a central set in the space of the time series towards which the stochastic trajectories drift almost certainly.

[11]See also Chan and Tong (1985) for a set of sufficient conditions on the general threshold autoregression. The conditions with regard to other parameters such as intercepts or threshold parameter are not available. Investigating this would be an interesting avenue of research but is beyond the scope of this paper.

has been rather mixed. A stream of research (e.g. Mishikin 1992; Evans and Lewis 1995; Crowder and Hoffman 1996, amongst others) looked at the dynamic behavior of the real interest rate within a cointegrated framework taking the nonstationarity of inflation and nominal interest as a maintained hypothesis. Despite the wide acceptance in empirical research, it remains an open question whether both inflation and nominal interest rate are integrated variables. In fact, some studies based on panel data techniques [e.g. Wu and Zhang 1996; Wu and Chen 2001) have found evidence against nonstationarity for nominal interest rates, which violates the prerequisite for applying a cointegration technique. They attribute the earlier findings of unit-root for nominal interest to an artefact of the low finite sample power of the standard univariate tests. Another criticism against the cointegration approach is related to the poor performance of the conventional unit-root test techniques when potential nonlinearities in the dynamics are overlooked. Indeed, more recent studies (e.g. Garcia and Perron 1996; Bekdache 1999; Lai 2004) attempt to compromise the empirical findings with the theoretical prediction on real interest rates by demonstrating that the stochastic behaviour of real interest rate is better described by stationary but non-linear characterizations. Given that the wedge between the empirical evidence and the theoretical models rests in large part on the inference around nonlinearity and nonstationarity; the issue is highly relevant to the main theme of this paper.

Table 8 reports the empirical results when the five unit-root test techniques discussed in the previous sections are implemented to the real interest rates of 12 industrial countries: Australia, Belgium, Canada, Denmark, France, Italy, the Netherlands, Norway, Switzerland, New Zealand, the United Kingdom and the United States. The selection of countries was governed by the requirement of having longest continuous data series. The data used are quarterly long-run government bond yield (IFS line code 61..ZF) and CPI (IFS line code 64) based inflation rate during 1957:1–2003:3 retrieved from the International Monetary Fund's *International Financial Statistics (IFS)*. Here, we follow the common practice of using *ex post* real interest rates defined by the difference between the nominal interest rate and *realized* inflation which is regarded as proxies for the expected inflation rate.[12]

As can be seen from Table 8, the ADF test provides little evidence of mean reversion as it fails to reject the unit-root null in most countries, which mirrors similar results found in the preceding studies based on the conventional unit-root tests. However, this conclusion is reversed by the tests designed to have power against the alternative of specific non-linear models. In the majority of countries, the M-TAR, sign and inf-*t* tests provide strong evidence of mean reversion by soundly rejecting the unit-root null even at the 5% significance level. Taken together, it might be fair to interpret this seemingly contradictory result as convincing evidence of non-linear mean-reverting behaviour of real interest rates except for Italy and the United States where all the tests reach an agreement on the unit-root. The next step might be to ask in which non-linear process the real interest rates follow. Unfortunately, the extant techniques are of little help in providing guidance as to which non-linear form is most appropriate. However, comparison of the empirical results

[12]In the literature, it is widely agreed that distinction between *ex post* and *ex ante* real interest rate is immaterial to the mean reversion of the real interest rate because the difference between the two is the inflation forecasting error which is assumed to be stationary under rational expectations. As well documented in Sun and Phillips (2004), however, the use of *ex post* series could substantially underestimate the true degree of persistence in the *ex ante* variables if the former is viewed as noisy observations of the latter. Nevertheless, we stick to the *ex post* real interest rates in our study partly because data for expected inflation are unavailable in many countries for the long time span under study and largely for the purpose of comparison with earlier studies based on *ex post* series. Furthermore, the issue of adjusting for inflation seems not been settled yet.

Table 8. Empirical results for quarterly real interests.

	Unit-root tests					Datings of
Country	ADF	MTAR	Sign	KSS	inf-t	structural breaks
AUS	−1.97	8.59**	−20.00**	−2.75*	−3.58**	70:2 [69:1–70:3] 90:3 [89:4–92:1]
BEL	−7.37**	81.98**	−29.00**	−6.43**	−9.19**	None
CAN	−1.45	6.25**	−23.00**	−1.33	−2.94*	66:4 [66:3–67:1] 91:1 [90:3–94:2]
DEN	−1.44	11.98**	−18.00**	−1.47	−2.99*	69:2 [67:1–70:1] 83:3 [83:1–87:4] 90:3 [89:2–91:2]
FRA	−1.63	6.43**	−19.00**	−1.91	−4.44**	91:3 [91:2–91:4]
ITA	−1.59	3.28	−16.00	−2.79*	−3.11	None
NET	−2.05	32.71**	−30.00**	−1.83	−6.04**	68:3 [66:4–71:2] 84:3 [83:2–86:2]
NOR	−1.32	14.88**	−31.00**	−1.08	−3.54**	69:3 [68:3–70:3] 79:3 [72:1–81:1] 91:1 [90:2–92:2]
NZL	−2.66*	15.65**	−26.00**	−3.00**	−5.15**	69:3 [65:2–69:4] 90:3 [90:1–92:2]
SWI	−2.73*	6.55**	−25.00**	−4.73**	−5.33**	64:4 [57:1–67:2] 93:1 [92:3–97:3]
U.K.	−2.03	11.28**	−29.00**	−2.77*	−3.91**	69:3 [67:1–70:1] 82:1 [81:3–87:1] 92:1 [91:1–93:1]
U.S.	−1.55	3.75	−4.00	−2.62	−3.27	67:1 [66:4–67:1] 90:3 [89:3–95:1]

Note: * and ** denote the cases where the null hypothesis can be rejected at the 10%, 5%, respectively, significance level. Entries in the dating represent the occurrence of break points in year and quarter estimated by the sequential procedure estimation method of Bai and Perron (1998, 2003). In brackets are the 95% confidence intervals for the end dates.

in Table 8 with our simulation results in Table 2 hints a promising candidate: multiple structural break model (DGP 15) in which the M-TAR, sign and inf-t tests exhibit decent discriminatory against unit-root while the ADF test does not. As shown in Table 7, in the sample size of $T = 200$ which is comparable to our empirical application, the rejection rate of the ADF test is very low even for large sample size, whereas the rejection rates of the three tests are quite high when the associated AR parameter is relatively mild. Our argument is reinforced by more recent studies on real interest rate (e.g. Bai and Perron 1998, 2003; Caporale and Grier 2000; Garcia and Perron 1996; Rapach and Wohar 2005) that report evidence of structural breaks in the mean which has been elusive for the standard tests to capture. For example, by applying the Bai and Perron's

(1998, 2003) structural break testing methodology to 13 industrialized countries during 1957–1998, Rapach and Wohar (2005) find structural breaks in the mean for *ex post* tax-adjusted real interest rates. Once we find the existence of breaks, we need to know their locations. Exploiting the findings in Rapach and Wohar, we apply the Bai and Perron's methodology to our data sets.[13] Table 8 also presents the dates for the structural breaks in the real interest rates and their 95% confidence intervals for the 12 countries. As in Rapach and Wohar, we witness that the breaks for the real interest rates occur relatively close to one another in many countries. However, unlike Rapach and Wohar, we fail to find any convincing evidence of commonality in the number of structural breaks and occurring dates. In the number of breaks, for example, Belgium and Italy exhibit no break, while three breaks are found in the real interest rates for Denmark, Norway and the United Kingdom.

6. CONCLUDING REMARKS

Standard unit-root tests are known to be biased towards the non-rejection of a unit root when they are applied to time series with non-linear dynamics. Although it is widely perceived that nonlinearity may constitute a significant source of loss of power in standard unit-root tests, not much is known about the source of power loss mainly because the analysis on nonstationarity and nonlinearity to this date has been fragmentary. By means of a Monte Carlo study, the current paper attempts to provide an advance over the existing literature by investigating the finite sample performance of unit-root tests against a wide class of stationary but non-linear processes. For this purpose, we apply five popular tests under the same unit-root null but with different stationary models in the alternatives to a broad family of non-linear dynamic processes. In contrast to the common perception, our simulation results suggest that what determines the power of unit-root tests is not the specific type of nonlinearity in the alternative model, but how far the alternative model is away from the unit-root process. The presence of nonlinearity does not necessarily result in power loss unless the associated AR parameter is close to unity. Moreover, the unit-root tests under study have decent discriminatory power against the models beyond the ones assumed in the alternative hypothesis. The ADF test has a satisfactory power property for various non-linear models and the M-TAR and inf-t tests have reasonable power in identifying unit-root process against other non-linear models than TAR or transitional AR models. Consequently, it is misleading if not dangerous to interpret the rejection of the unit-root null as a compelling evidence of the specific models under the alternative hypothesis. Among the five tests under study, the ADF test outperforms when the sample size is relatively small while the inf-t due to Park and Shintani is more powerful for large sample size irrespective of the form of models. We also illustrate the empirical relevance of our simulation study by reevaluating the mean reversion issue of real interest rates, often referred to as Fisher hypothesis. Our analysis suggests that the inconclusive evidence on the mean reversion was largely driven by the infrequent structural breaks in the mean

[13] Other than the data span, our data set is identical to the one used by Rapach and Wohar except that Ireland is excluded in our data set due to its discontinuity after 1998. As discussed in the text, we also take a different approach to deriving nominal interest and inflation rates. See Appendix for a brief description of the Bai and Perron's methodology. Readers can refer their original work for further details regarding the method. We thank Jushan Bai and Pierre Perron for the use of their publicly available Gauss code available at http://people.bu.edu/perron/.

of real interest rates which the standard testing techniques do not effectively distinguish from nonstationarity.

In the literature, it is often claimed that a unit-root test needs to be applied before establishing whether the model is linear or not. However, our results indicate that this approach will be of reduced merit when the associated AR parameters take large values which makes challenging for the extant technical devices to detect stationarity of processes with high AR parameter regardless of linearity.

ACKNOWLEDGEMENTS

We are very grateful to co-editor Pierre Perron and three anonymous referees for numerous corrections and valuable comments that helped to improve the paper. We also thank Pok-sang Lam for helpful comments and Jushan Bai and Pierre Perron for the use of their publicly available Gauss code. All computations were performed in GAUSS (version 6.0). All remaining errors are the authors'.

REFERENCES

Altissimo, F. and G. L. Violante (2001). The non-linear dynamics of output and unemployment in the U.S. *Journal of Applied Econometrics 16*, 461–86.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22.

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

Bec, F., A. Guay and E. Guerre (2002). Adaptive consistent unit root tests based on autoregressive threshold model. *mimeo*.

Bekdache, B. (1999). The time-varying behaviour of real interest rates: A re-evaluation of the recent evidence. *Journal of Applied Econometrics 14*, 171–90.

Caner, M. and B. E. Hansen (2001). Threshold autoregression with a unit root. *Econometrica 69*, 1555–96.

Caporale, T. and K. B. Grier (2000). Political regime change and the real interest rates. *Journal of Money, Credit and Banking 32*, 320–34.

Carrasco, M. (2002). Misspecified structural change, threshold and Markov-switching models. *Journal of Econometrics 109*, 239–73.

Chan, K. S., J. Petruccelli, H. Tong and S. Woolford (1985). A multiple-threshold AR(1) model. *Journal of Applied Probability 22*, 267–79.

Chan, K. S. and H. Tong (1985). On the use of the deterministic Lyapunov functions for the ergodicity of stochastic difference equations. *Advances in Applied Probability 17*, 666–78.

Crowder, W. J. and D. L. Hoffman (1996). The long-run relationship between nominal interest rates and inflation: The Fisher equation revisited. *Journal of Money, Credit and Banking 28*, 102–18.

Enders, W. (2001). Improved critical values for the Enders-Granger test. *Applied Economic Letters 8*, 257–61.

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

Evans, M. D. and K. K. Lewis (1995). Do expected shifts in inflation affect estimates of the long-run Fisher relation? *Journal of Finance*. *50*, 225–53.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. (New York, Wiley).

Garcia, R. and P. Perron (1996). An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics 78*, 111–25.

Granger, C. W. J. and A. P. Andersen (1978). *An Introduction to Bilinear Time Series Models*. (Gottingen: Vandenhoeck and Ruprecht).

Granger, C. W. J. and T. Teräsvirta (1993). *Modelling Non-linear Economic Relationships*. (Oxford, UK, Oxford University Press).

Hall, S., Z. Psaradakis and M. Sola (1997). Cointegration and changes in regime: The Japanese consumption function. *Journal of Applied Econometrics 12*, 151–68.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–84.

Kapetanios, G. and Y. Shin (2003). Unit root tests in three-regime SETAR models, mimeo, University of Edinburg.

Kapetanios, G., Y. Shin and A. Snell (2003). Testing for a unit root in the non-linear STAR framework. *Journal of Econometrics 112*, 359–79.

Kim, T.-H., S. Leybourne and P. Newbold (2002). Unit root tests with a break in innovation variance. *Journal of Econometrics 109*, 365–87.

Kristensen, D. (2005). On stationarity and ergodicity of the bilinear model with applications to GARCH Models, mimeo, University of Wisconsin.

Lai, K. S. (2004). On structural shifts and stationarity of the ex ante real interest rate. *International Review of Economics and Finance 13*, 217–28.

Lee, J. (2000). On the end-point issue in unit root tests in the presence of a structural break. *Economics Letters 68*, 7–11.

Leybourne, S., T. C. Mills and P. Newbold (1998). Spurious rejections by Dickey-Fuller tests in the presence of a break under the null. *Journal of Econometrics 87*, 191–203.

Levin, A. T. and J. M. Piger (2004). Is inflation persistence intrinsic in industrial economies? mimeo, The Federal Reserve Board.

Liu, J. and E. Susko (1992). On strict stationarity and ergodicity of a non-linear ARMA model. *Journal of Applied Probability 29*, 363–73.

Michael, P., A. R. Nobay and D. A. Peel (1997). Transactions costs and non-linear adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy 105*, 862–79.

Mishikin, F. (1992). Is the Fisher effect real? A reexamination of the relationship between inflation and interest rates. *Journal of Monetary Economics 30*, 195–15.

Müller, U. K. and G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica 71*, 1269–86.

Nelson, C. R., J. Piger and E. Zibot (2001). Markov regime switching and unit-root tests. *Journal of Business and Economic Statistics 19*, 404–15.

Ng, S. and P. Perron (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association 90*, 268–81.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica 69*, 1519–54.

Park, J. and M. Shintani (2005). Testing for a unit root against transitional autoregressive models, mimeo, Vanderbilt University.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica 57*, 1361–1401

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics 8*, 153–62.

Pippenger, M. and G. Goering (1993). A note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics 55*, 473–81.

Rapach, D. E. and M. E. Wohar (2003). Regime changes in international real interest rates: Are they a monetary phenomenon? *Journal of Money, Credit and Banking*, forthcomin.

Rothman, P. (1998). Forecasting asymmetric unemployment rates. *Review of Economics and Statistics 80*, 164–68.

Sarno, L., M. P. Taylor, and I. Chowdhury (2004). Non-Linear dynamics in deviations from the law of one price: A broad-based empirical study. *Journal of International Money and Finance 23*, 1–25.

Seo, M. (2003). Unit root test in a threshold autoregression: Asymptotic theory and residual-based block bootstrap, mimeo, University of Wisconsin.

So, B. S. and D. W. Shin (2001). An invariant sign test for random walks based on recursive median adjustment. *Journal of Econometrics 102*, 197–29.

Sun, Y. and P. C. B. Phillips (2004). Understanding the Fisher equation. *Journal of Applied Econometrics*, forthcoming.

Taylor, A. M. (2001). Potential pitfalls for the PPP puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica 69*, 473–98.

Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability 22*, 587–11.

Tong, H. (1990). Non-linear time series: A dynamical system approach. (Oxford, UK, Oxford University Press).

Tsay, W.-L. (2000). Long memory story of the real interest rate. *Economics Letters 67*, 325–30.

Wu, J.-L. and S.-L. Chen (2001). Mean reversion of interest rates in the eurocurrency market. *Oxford Bulletin of Economics and Statistics 63*, 459–73.

Wu, Y. and H. Zhang (1996). Mean Reversion in interest rates: New evidence from a panel of OECD countries. *Journal of Money, Credit and Banking 28*, 604–21.

Yang, M. (2000). Some properties of vector autoregressive processes with Markov-switching coefficients. *Econometric Theory 16*, 23–43.

APPENDIX: THE BAI–PERRON METHODOLOGY

In estimating unknown multiple structural breaks in dynamic linear regression models, Bai and Perron's (1998, 2003) method largely consists of two stages. The first stage pertains to estimating the number of unknown structural breaks. To this end, BP suggests several testing procedures: double maximum tests and a test for l versus $l + 1$ breaks. The former tests are constructed under the null hypothesis of no structural break against the alternative of an unknown number of breaks given some upper bound, while the latter, labelled $\sup F_T(l + 1|l)$, involves testing the null of l breaks against the alternative of $l + 1$ breaks. BP recommends to apply double maximum tests first to see whether at least one break exists. If the tests suggest the presence of at least one break, then the number of breaks is decided based on a sequential examination of the $\sup F_T(l + 1|l)$ statistics. According to BP, this approach leads to the best results and hence recommended for empirical applications.

Once the number of break is identified, the second stage of BP method is related to estimating breakpoints as well as coefficients of interest using the least squares principle. To illustrate, consider a linear regression model with $m - 1$ breaks (and thus m regimes which are identified in the first stage),

$$y_t = \delta^{(j)} + \varepsilon_t, \quad t = T_{j-1} + 1, T_{j-1} + 2 \ldots, T_j,$$

for $j = 1, \ldots, m$ where $\delta^{(j)}$ is the mean level of the series in the j_{th} regime. To estimate breakpoints, we consider every possible m-partition of T, $(T_1, \ldots, T_m)$. For each m-partition, the regression coefficients ($\delta^{(j)}$s) are estimated by minimizing

$$\sum_{j=1}^{m} \sum_{t=T_{j-1}+1}^{T_j} \left[y_t - \delta^{(j)}\right]^2.$$

Let $S_T(T_1, \ldots, T_m)$ be the sum of squared residuals such that $S_T(T_1, \ldots, T_m) = \sum_{j=1}^{m} \sum_{t=T_{j-1}+1}^{T_j} [y_t - \hat{\delta}^{(j)}]^2$. Then the breakpoints $(\hat{T}_1, \ldots, \hat{T}_m)$ are estimated by choosing the m-partition that has minimal $S_T(T_1, \ldots, T_m)$ such that

$$(\hat{T}_1, \ldots, \hat{T}_m) = argmin_{T_1,\ldots,T_m} S_T(T_1, \ldots, T_m).$$

With these breakpoint estimates, the associated regression parameters are estimated subsequently. BP develop an efficient algorithm for the minimization problem based on the principle of dynamic programming.