Ann Inst Stat Math (2013) 65:125–147
DOI 10.1007/s10463-012-0363-z

Nonparametric pseudo-Lagrange multiplier stationarity testing

Manuel Landajo · María José Presno

Received: 27 June 2011 / Revised: 19 December 2011 / Published online: 17 April 2012

Abstract The framework of stationarity testing is extended to allow a generic smooth trend function estimated nonparametrically. The asymptotic behavior of the pseudo-Lagrange multiplier test is analyzed in this setting. The proposed implementation delivers a consistent test whose limiting null distribution is standard normal. Theoretical analyses are complemented with simulation studies and some empirical applications.

Keywords Time series · Stationarity testing · Limiting distribution · Nonparametric regression · Nonparametric hypothesis testing

1 Introduction and scope of the paper

A considerable amount of research has focused on developing both unit root and stationarity tests, capable of distinguishing between integrated and stationary-around-a-trend/level stochastic processes. A number of statistical implications are associated with this distinction, which becomes crucial in applied time series forecasting, where it is well known that difference stationary and trend stationary processes often imply very different forecasts (e.g., Diebold and Kilian 2000). Other application fields include analysis of economic/financial time series, health economics (studies on health expenditure and gross domestic product, e.g., Jewell et al. 2003), hydrology (where testing for stationarity is an important topic, e.g., Wang 2006; Van Gelder et al. 2007), or

M. Landajo (✉) · M. J. Presno
Unit of Statistics and Econometrics, Department of Applied Economics,
University of Oviedo, Avenida del Cristo s/n, 33011 Oviedo, Spain
e-mail: landajo@uniovi.es

M. J. Presno
e-mail: mpresno@uniovi.es

climate change studies (air pollutant emissions, Dore and Johnston 2000; temperatures, Gay-García et al. 2009). Furthermore, both unit root and stationarity tests are also regularly applied for model selection.

A potential drawback of standard unit root and stationarity tests stems from their lack of robustness to misspecification of the trend function. For instance, under misspecification of the trend component, standard stationarity tests typically diverge in probability as sample size goes to infinity, so a spurious unit root is detected with probability approaching one. This recognition has led some authors to devise *flexible* tests, that do not depend critically on trend specification. The earliest proposal is due to Bierens (1997), who developed a unit root test that considers as alternative a random process that is stationary around a trend which belongs to a very general class of functions of time. Bierens's approach relies on approximating the trend function by Chebyshev polynomials, and the test becomes more flexible as more complex Chebyshev polynomials are used. The limiting null distribution of the test statistic, which depends on the complexity of the chosen approximant, is computed by Monte Carlo simulations.

In the field of stationarity testing, Becker et al. (2006) proposed two *flexible* tests. The first test relies on approximating the unknown trend function by a model including an intercept, a linear trend and two trigonometric components, namely a sine and cosine function, whose frequency is chosen to maximize goodness of fit. The limiting null distribution of the test, which is nonstandard, is derived in the same paper. A potential limitation stems from the fact that most smooth functions have Fourier expansions with an infinite number of frequencies, so the approximation capabilities of the chosen specification are limited. The same paper also proposes the *cumulative frequency* test, which provides further flexibility as it relies on trigonometric polynomials. The null distribution of the test, which as in Bierens's approach, depends on the order of the trigonometric polynomial, is computed in the same paper by Monte Carlo simulations. The cumulative frequency test relies on a *fixed-order* trigonometric polynomial, whose representation capabilities are inherently limited to trend functions with only a finite number of non-null terms in their Fourier expansion, so in this implementation the test may be regarded as flexible, but not properly as nonparametric. This drawback is well known in nonparametric statistics. A classical remedy is provided by the method of sieves (Grenander 1981): by using an increasing sequence of parametric models, whose complexity grows with sample size at appropriately limited rates, the method delivers consistent nonparametric estimation and hypothesis testing in very general settings (e.g., Hong and White 1995).

The sieve principle can also be exploited in stationarity testing. This is the main goal of the paper. The proposed test relies on nonparametric least squares estimation of the trend component, which is carried out through trigonometric series regression. Our approach has the following features: (a) we focus on smooth trends which can be approximated with arbitrary accuracy, in mean-squared sense, by linear combinations of the elements of a cosine basis (in principle, any squared integrable function on [0, 1] has this property). (b) The stochastic part of the null model is generated by a linear filter process (the performance of the test under several nonlinear time series models is also studied in simulations). (c) The behavior of the test when the long-run variance of the process is estimated upon the residuals of the nonparametric regression is studied.

Finally, (d) the proposed test asymptotically has correct size, and is consistent under unit root alternatives, with its limiting null distribution being standard normal, which enables a fairly simple implementation in practice.

As compared with previous stationarity tests, this proposal provides a number of interesting features: (i) the proposed test is fully nonparametric. It relies on a sieve mechanism which ensures both consistent estimation of the trend function and asymptotically correct behavior of the stationarity test. (ii) The limiting null distribution of the (suitably rescaled) test is standard, unlike those of most unit root and stationarity tests, whose distributions are nonstandard as well as model-dependent, in the sense that they range with each trend specification. The expedient of rescaling the test statistic avoids the burden of computing (usually, by Monte Carlo simulations) a different set of critical values for each choice of model complexity. (iii) The analytical results in this paper are valid for the case when the data are driven by linear processes, so our analyses are not limited to the i.i.d. context. (iv) The complicated issue of estimating the long-run variance of the process in the nonparametric environment is addressed analytically. A proposal (namely, a class of kernel estimators for the long-run variance of the process) is provided in the paper, and its theoretical behavior is analyzed, including derivation of appropriate rates for bandwidth increase in the nonparametric setting. The estimator for the long-run variance of the process is computed upon the residuals of nonparametric regression.

The rest of the paper is structured as follows: in Sect. 2 the nonparametric test is introduced, and its limiting behavior is analyzed. In Sect. 3 the finite sample properties of the test are investigated and, in Sect. 4, some empirical applications are presented. The paper closes with a summary of conclusions. All the mathematical derivations are collected in the Appendix.

2 A nonparametric stationarity test

2.1 The model

The following error-components model may be used as a general framework:

$$\begin{aligned} y_{t,T} &= \mu_t + \theta^*(t/T) + \varepsilon_t, \\ \mu_t &= \mu_{t-1} + u_t; \quad t = 1, \ldots, T; \quad T = 1, 2, \ldots \end{aligned} \tag{1}$$

with $\theta^* : [0, 1] \to \mathbb{R}$ being a smooth function of time (i.e., a trend). We consider approximants to θ^* of the form $\theta_m(x) = \sum_{j=0}^{m} \beta_{j,m}\varphi_j(x)$, with $\beta_m = (\beta_{0,m}, \ldots, \beta_{m,m})' \in \mathbb{R}^{m+1}$, $\varphi_0(x) = 1$, $\varphi_j(x) = \sqrt{2}\cos(j\pi x)$, $j \geq 1$, $x \in [0, 1]$. (The basis $\{\varphi_j, j = 0, 1, \ldots\}$ is complete and orthonormal in $L_2[0, 1]$.) We let $m = m_T$ grow to infinity with sample size T at an appropriate rate and assume that θ^* is the limit of $\{\theta_{m_T}\}$ under the metric $d_T(\theta_{m_T}, \theta^*) \equiv \sqrt{T^{-1} \sum_{t=1}^{T} [\theta_{m_T}(t/T) - \theta^*(t/T)]^2}$. This holds for any function in $L_2[0, 1]$, although further smoothness conditions on θ^* will be imposed below.

In addition, $\{\varepsilon_t\}$ and $\{u_t\}$ are independent of each other zero-mean error processes with characteristics to be detailed below and respective (finite) variances $E(\varepsilon_t^2) = \sigma_\varepsilon^2 > 0$ and $E(u_t^2) = \sigma_u^2 \geq 0$; $\{\mu_t\}$ starts with μ_0, which is assumed to be zero.

Lagrange multiplier (LM) stationarity testing relies on the following setting:

$$\mathrm{H}_0 : q \equiv \frac{\sigma_u^2}{\sigma_\varepsilon^2} = 0, \text{ versus } \mathrm{H}_1 : q > 0 \tag{2}$$

In usual stationarity testing, a parametric model for the trend function, such as $\theta^*(x) = \beta_0 + \beta_1 x$, is specified in advance, and the LM statistic to test (2) has the well-known expression:

$$S_T = \widehat{\sigma}^{-2} T^{-2} \widehat{\boldsymbol{\varepsilon}}' \boldsymbol{C}_T' \boldsymbol{C}_T \widehat{\boldsymbol{\varepsilon}} = \widehat{\sigma}^{-2} T^{-2} \sum_{t=1}^{T} E_t^2, \tag{3}$$

where $\widehat{\boldsymbol{\varepsilon}}$ is the $T \times 1$ vector of OLS residuals (we suppress double indexing for notational simplicity) and $\boldsymbol{C}_T$ is a $T \times T$ lower triangular matrix of ones. $E_t = \sum_{i=1}^{t} \widehat{\varepsilon}_i$ denotes the forward partial sum of the residuals, and $\widehat{\sigma}^2$ is a suitable estimator for the long-run variance of $\{\varepsilon_t\}$, to be denoted as σ^2 and assumed non-null.

We will analyze the behavior of the above stationarity test when the trend function θ^* is estimated nonparametrically and the test is carried out upon the residuals of this regression. We consider the estimator $\widehat{\theta}_{m_T}(x) = \sum_{j=0}^{m_T} \widehat{\beta}_j \varphi_j(x)$, with $\widehat{\boldsymbol{\beta}}_{m_T} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_{m_T})'$, which (given m_T) is computed by OLS regression, i.e., $\widehat{\boldsymbol{\beta}}_{m_T} = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{y}$, with $\boldsymbol{\Phi} = [\varphi_{t,j}]$, $\varphi_{t,j} = \varphi_j(t/T)$, $t = 1, \ldots, T$; $j = 0, \ldots, m_T$; we may write $\boldsymbol{\Phi} = [\varphi_1, \ldots, \varphi_m]$, with $\varphi_j = [\varphi_j(1/T), \ldots, \varphi_j(T/T)]'$.

The pseudo-LM test statistic has the usual form:

$$\widehat{S}_T = \widehat{\sigma}^{-2} T^{-2} \mathbf{e}' \boldsymbol{C}_T' \boldsymbol{C}_T \mathbf{e} \tag{4}$$

with $\mathbf{e} = (e_1, \ldots, e_T)'$ being the vector of OLS residuals from the above nonparametric regression. (As the sum of OLS residuals is null in this setting, it holds $\mathbf{e}' C_T' C_T \mathbf{e} = \mathbf{e}' C_T C_T' \mathbf{e}$, so the nonparametric test may also be computed upon the basis of backward partial residual sums.)

We shall follow the usual conventions: the symbol "$\xrightarrow{L}$" indicates convergence in distribution, "$\xrightarrow{p}$" denotes convergence in probability, and symbols O_p and o_p are used with their usual probability order meanings, as $T \to \infty$ with respect to the probability measure P.

2.2 Assumptions

We consider model (1) under the following assumptions:

Assumption 1 (i) The underlying probability space $(\Omega, \mathbb{F}, P)$ is complete, and the unobservable error process $\{\varepsilon_t\}$ is generated as $\varepsilon_t = \sum_{j=0}^{\infty} \alpha_j v_{t-j}$, with

$\sum_{j=0}^{\infty} |\alpha_j| < \infty$ and $\alpha \equiv \sum_{j=0}^{\infty} \alpha_j \neq 0$. (ii) The process $\{v_t \mid t = 1, 2, \ldots\}$ is independent identically distributed (i.i.d.), with $E(v_t) = 0$, $Var(v_t) = \sigma_v^2 > 0$ and $E|v_t|^r < \infty$ for some $r > 2$. (iii) The process $\{u_t\}$ is independent of $\{v_t\}$, has $E(u_t) = 0$, $Var(u_t) = \sigma_u^2 \geq 0$, $E|u_t|^{2+\delta} < \infty$ for some $\delta > 0$, and $\sum_{t=1}^{T} u_t = O_p(T^{1/2})$.

Assumption 2 As $T \to \infty$, (i) $m_T^{3/2} T\, d_T(\theta_{m_T}, \theta^*) \to 0$, with (ii) $m_T \to \infty$ and $m_T^{9/2} T^{-1} \to 0$.

Assumption 3 $\widehat{\sigma}^2 \geq 0$ and, as $T \to \infty$, (i) $m_T^{1/2}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{p} 0$ under H_0, and (ii) under H_1, $\widehat{\sigma}^2 = O_p(T^{\zeta})$, $0 \leq \zeta < 2$. □

2.3 Results

First we derive the limiting behavior of the (standardized) test when the long-run variance is known.

Proposition 1 *Under Assumptions 1 to 2, let* $Z_T = s_{m_T}^{-1}(\sigma^{-2} S_T - \mu_{m_T})$, *where* $S_T = T^{-2}\mathbf{e}'\mathbf{C}_T'\mathbf{C}_T\mathbf{e}$, $\mathbf{e} = (\mathbf{I}_T - \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}')\mathbf{y}$, $\mathbf{\Phi} = [\varphi_{t,j}]$, $\varphi_{t,j} = \varphi_j(t/T)$, $t = 1, \ldots, T$, $j = 0, \ldots, m_T$, $\mu_{m_T} = \sum_{j=m_T+1}^{\infty}(j\pi)^{-2}$ *and* $s_{m_T}^2 = 2\sum_{j=m_T+1}^{\infty}(j\pi)^{-4}$. *Then as* $T \to \infty$*:*

(a) *under* H_0, $Z_T \xrightarrow{L} N(0, 1)$,
(b) *under* H_1, $P(Z_T > \kappa_T) \to 1$ *for any nonstochastic sequence* $\{\kappa_T\}$ *with* $\kappa_T m_T^{-3/2} T^{-2} \to 0$. □

An analogous result follows when σ^2 is estimated from data.

Proposition 2 *Under Assumptions 1 to 3, let* $\widehat{Z}_T = s_{m_T}^{-1}(\widehat{\sigma}^{-2} S_T - \mu_{m_T})$. *Then as* $T \to \infty$*:*

(a) *under* H_0, $\widehat{Z}_T \xrightarrow{L} N(0, 1)$,
(b) *under* H_1, $P(\widehat{Z}_T > \kappa_T) \to 1$ *if* $\kappa_T m_T^{-3/2} T^{-(2-\zeta)} \to 0$. □

Assumption 3 requires a suitable estimator for σ^2. We follow Pötscher and Prucha (1991), and results are stated for nonparametric estimators with kernel W belonging to the class $\mathbb{W}_\rho$, of functions $W : \mathbb{R} \to [-1, 1]$ satisfying $W(0) = 1$, $W(-x) = W(x)$ for all x, $W(x) = 0$ for $|x| > 1$ and $\lim_{x\to 0} |W(x) - 1|/x^{\rho} < \infty$ for some $\rho > 0$.

The following result states that Proposition 2 holds if σ^2 is replaced by a nonparametric estimator with kernel W belonging to the class $\mathbb{W}_\rho$. (The truncation estimator is embedded into the scheme of Proposition 2 by using $|\widehat{\sigma}^2|$ instead of $\widehat{\sigma}^2$.)

Proposition 3 *Under Assumptions 1 to 3, let* $\widehat{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} w_{i,T}\widehat{\sigma}_i \geq 0$, *with* $\widehat{\sigma}_i = T^{-1}\sum_{t=1+|i|}^{T} e_t^{(d)} e_{t-|i|}^{(d)}$, $\mathbf{e}^{(d)} = (e_1^{(d)}, \ldots, e_T^{(d)})' = (\mathbf{I}_T - \mathbf{\Phi}_d(\mathbf{\Phi}_d'\mathbf{\Phi}_d)^{-1}\mathbf{\Phi}_d')\mathbf{y}$, $\mathbf{\Phi}_d = [\varphi_{t,j}]$, $\varphi_{t,j} = \varphi_j(t/T)$, $t = 1, \ldots, T$, $j = 0, \ldots, m_T^{(d)}$, *and* $w_{i,T} = W(i/(1 + \ell_T))$, *with kernel* $W \in \mathbb{W}_\rho$. *If the following conditions hold:* (i) *either* (i.1) $E|v_t|^4 < \infty$ *or* (i.2) $E|v_t|^r < \infty$, *with* $2 < r < 4$ *and* $\alpha_j = O(j^{-(1+q_v+\epsilon)})$, *where* $\epsilon > 0$ *and* $-q_v \leq \min\{-2(r-1)/(r-2), -(\rho+1)\}$, (ii) $(m_T^{(d)})^{9/2}T^{-1} \to 0$, (iii) $m_T \ell_T^2 m_T^{(d)} T^{-1} \to 0$, (iv) $m_T\ \ell_T^3 T^{-1} \to 0,\ m_T \ell_T^{-2\rho} \to 0$, (v) $m_T^d \to \infty,\ \ell_T \to \infty$, *then as* $T \to \infty$*:*

(a) *under* H_0, $m_T^{1/2}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{p} 0$,

(b) *under* H_1, $\widehat{\sigma}^2 = O_p(\ell_T T)$ *and* $P(\widehat{Z}_T > \kappa_T) \to 1$ *if* $\kappa_T \ell_t m_T^{-3/2} T^{-1} \to 0$. □

In Proposition 3 above we allow possibly different model complexities (respectively, m_T and $m_T^{(d)}$) for the numerator and denominator of the test statistic. To our knowledge, the possibility of this separate treatment of both components in the test statistic has not been exploited in the literature, although our simulations indicate that this strategy may be useful in certain cases, where a more complex model may be advisable in the numerator of the statistic—e.g., this may induce undersmoothing, so reducing the bias of the nonparametric trend estimator, and allowing better control of the test's size—than in the denominator, where stronger complexity control may be useful to better estimate the long-run variance of the process.

3 Monte Carlo study

In this section, we first provide computer simulation results for the performance of the test under i.i.d. errors, and afterwards the research is extended to time series.

3.1 Simulations in i.i.d. environment

We analyze the following trend specifications:

(A) $\theta^*(x) = 0$,
(B) $\theta^*(x) = 1 + 2x + 3x^2$,
(C) $\theta^*(x) = 1 + 2x + 3[1 + \exp\{-50(x - 0.3)\}]^{-1} - 4[1 + \exp\{-40(x - 0.6)\}]^{-1}$,
(D) $\theta^*(x) = 1 + 2x + 2[1 + \exp\{-\gamma(x - 0.3)\}]^{-1}$, with $\gamma = 20, 50, 100$,
(E) $\theta^*(x) = 1 + 2x + 2[1 + \exp\{-\gamma(x - 0.3)(x - 0.6)\}]^{-1}$, for $\gamma = 20, 50, 100$,
(F) $\theta^*(x) = 1 + 2x + 2[1 - \exp\{-\gamma(x - 0.3)^2\}]$, where $\gamma = 20, 50, 100$,
(G) $\theta^*(x) = 1 + 2x + 2x\mathbf{1}(x > 0.3)$,
(H) $\theta^*(x) = 1 + 2x - x\mathbf{1}(x > 0.3)$,
(I) $\theta^*(x) = 1 + 2x - 3(x - 0.3)\mathbf{1}(x > 0.3) + 4(x - 0.6)\mathbf{1}(x > 0.6) - 5(x - 0.8)\mathbf{1}(x > 0.8)$,

with $x \in [0, 1]$ and $\mathbf{1}(\bullet)$ denoting the indicator function.

Specification (A) corresponds to a no-trend (or *level plus noise*) model under the null hypothesis, and a random walk with noise under the alternative, while (B) allows us to analyze an explosive deterministic trend. Specification (C) represents an artificial neural network trend, or equivalently, a linear trend affected by two smooth transitions of large magnitude which are modeled by logistic sigmoids. Sigmoid functions are very flexible and allow the analysis of series with gradual changes. So, several classes of sigmoid curves are considered in specifications (D)–(F). Finally, specifications (G)–(I) correspond to linear trends with breaks (simulation results for the linear trend model with no breaks were very close to the more general cases considered in this section, so we omitted them for brevity).

In simulations we considered sample sizes $T = 100, 300, 500, 1{,}000, 1{,}500, 2{,}000$, and signal-to-noise ratio values $q = 0, 0.01, 0.1$. In the i.i.d. analysis, simulations were based on 5,000 replications, with the processes u_t and ε_t being $N(0, 1)$. We applied the deterministic rule $m_T = m_T^d = [5T^{1/5}]$. This rate of increase is somewhat slower than others which appear in related (mainly cross-sectional) nonparametric regression contexts, although the combined requirements of Assumptions 2 and 3 advised us against being too liberal in this respect. The variance of the process was estimated by using the "unbiased" estimator $\widehat{\sigma}^2 = (T - m_T^d - 1)^{-1} e^{(d)\prime} e^{(d)}$, as our simulations indicated that, in small samples, this estimator outperformed the (asymptotically equivalent) plug-in estimator $\widetilde{\sigma}^2 = T^{-1} e^{(d)\prime} e^{(d)}$ when included in the nonparametric stationarity test. (It can be readily checked that the asymptotic results in Proposition 3 remains valid if these "unbiased" estimators for the autocovariances are used; we omit derivations for brevity). Table 1 displays the rejection rates at 5 % significance level, with the critical value provided by the $N(0, 1)$ limiting distribution. Results indicate that size is close to the nominal significance level (excepting case D with $\gamma = 100$, where a slight oversize is observed) and power figures are very similar for all the trend specifications considered. So, the test seems to perform suitably under a wide spectrum of trend specifications, thus being free from the overrejection problems caused by misspecification of the trend function.

Next, we compared the performance of the nonparametric test with that of the two flexible stationarity tests (hereafter BEL1 and BEL2) proposed by Becker et al. (2006). As commented above, BEL1 approximates the unknown trend function (their paper focuses mainly on smooth breaks of unknown form and number) by a single frequency component from its classical Fourier expansion. The recommendation is to select that frequency (with a maximum of 5) which minimizes the sum of squared residuals. The BEL2 (or cumulative frequency) test also relies on classical Fourier series: the trend function is estimated by least squares regression on a basis of sines and cosines, and in order to avoid power loss Becker et al. (2006) recommend that at most the first two frequencies be included. We checked the performance of the BEL tests—including a linear trend component, τ_τ test—for trend specifications (A)–(I) and sample sizes: $T = 100, 500, 1{,}000, 2{,}000$. Results are reported in Table 2 below.

The BEL2 test displays better control of the test size than BEL1, but less power. These results agree with Becker et al. (2006) conclusions. However, even the BEL2 strategy suffers size distortions in some of the studied cases. The magnitude of oversize becomes very large as sample size increases (indeed, size approaches 1). As outlined in Table 1 above the nonparametric test does not suffer these size distortions. On the other hand, the BEL tests exhibit higher power than the nonparametric test for $T \leq 1{,}000$, though for larger sample sizes the power of both tests is similar. This could be expected in advance, as the nonparametric test includes more frequency terms in order to control test size, which somewhat reduces power.

Table 1 Size and power of the nonparametric stationarity test

$q\backslash T$	Trend A						Trend B						Trend C					
	100	300	500	1000	1500	2000	100	300	500	1000	1500	2000	100	300	500	1000	1500	2000
0	0.036	0.045	0.053	0.057	0.057	0.058	0.042	0.046	0.056	0.060	0.056	0.055	0.041	0.068	0.066	0.069	0.063	0.063
0.01	0.036	0.139	0.343	0.882	0.992	1	0.047	0.145	0.341	0.876	0.993	1	0.047	0.159	0.360	0.886	0.993	1
0.1	0.133	0.874	0.997	1	1	1	0.116	0.871	0.997	1	1	1	0.105	0.872	0.996	1	1	1
	Trend D ($\gamma = 20$)						Trend D ($\gamma = 50$)						Trend D ($\gamma = 100$)					
0	0.042	0.050	0.058	0.062	0.049	0.051	0.032	0.054	0.058	0.045	0.058	0.061	0.049	0.075	0.096	0.107	0.132	0.101
0.01	0.060	0.140	0.311	0.878	0.991	1	0.040	0.142	0.331	0.858	0.992	1	0.073	0.168	0.396	0.874	0.995	1
0.1	0.142	0.838	0.991	1	1	1	0.089	0.831	0.994	1	1	1	0.115	0.841	0.995	1	1	1
	Trend E ($\gamma = 20$)						Trend E ($\gamma = 50$)						Trend E ($\gamma = 100$)					
0	0.049	0.050	0.057	0.050	0.057	0.045	0.033	0.049	0.057	0.059	0.054	0.051	0.034	0.054	0.053	0.059	0.058	0.059
0.01	0.049	0.126	0.321	0.869	0.994	1	0.041	0.131	0.326	0.891	0.993	1	0.030	0.139	0.329	0.878	0.992	0.998
0.1	0.137	0.836	0.992	1	1	1	0.134	0.832	0.991	1	1	1	0.075	0.835	0.994	1	1	1
	Trend F ($\gamma = 20$)						Trend F ($\gamma = 50$)						Trend F ($\gamma = 100$)					
0	0.044	0.050	0.059	0.054	0.057	0.059	0.041	0.046	0.056	0.054	0.049	0.053	0.029	0.049	0.055	0.054	0.057	0.048
0.01	0.037	0.136	0.336	0.867	0.992	1	0.034	0.135	0.323	0.882	0.993	1	0.048	0.130	0.341	0.899	0.992	1
0.1	0.109	0.831	0.992	1	1	1	0.105	0.821	0.991	1	1	1	0.083	0.841	0.992	1	1	1
	Trend G						Trend H						Trend I					
0	0.024	0.057	0.068	0.090	0.105	0.096	0.053	0.052	0.059	0.060	0.054	0.068	0.027	0.049	0.063	0.058	0.060	0.054
0.01	0.038	0.150	0.357	0.871	0.994	1	0.030	0.140	0.324	0.871	0.993	1	0.050	0.143	0.331	0.878	0.993	1
0.1	0.091	0.834	0.994	1	1	1	0.112	0.843	0.990	1	1	1	0.135	0.822	0.993	1	1	1

i.i.d. $N(0, 1)$ errors; 5 % significance

Table 2 Size and power of the Becker et al. (2006) test

$q \backslash T$	BEL1 test												BEL2 test											
	Trend A				Trend B				Trend C				Trend A				Trend B				Trend C			
	100	500	1000	2000	100	500	1000	2000	100	500	1000	2000	100	500	1000	2000	100	500	1000	2000	100	500	1000	2000
0	0.011	0.008	0.007	0.010	0.074	0.196	0.353	0.639	0.999	1	1	1	0.050	0.050	0.060	0.052	0.053	0.078	0.106	0.164	0.257	0.976	1	1
0.01	0.065	0.889	0.995	1	0.123	0.917	0.996	1	0.999	1	1	1	0.098	0.882	0.999	1	0.090	0.896	0.999	1	0.315	0.985	0.999	1
0.1	0.554	0.999	1	1	0.584	1	1	1	0.959	1	1	1	0.522	1	1	1	0.512	1	1	1	0.666	1	1	1
	Trend D ($\gamma = 20$)				Trend D ($\gamma = 50$)				Trend D ($\gamma = 100$)				Trend D ($\gamma = 20$)				Trend D ($\gamma = 50$)				Trend D ($\gamma = 100$)			
0	0.162	0.634	0.928	0.998	0.426	0.996	1	1	0.501	0.999	1	1	0.060	0.117	0.180	1	0.180	0.687	0.954	1	0.228	0.892	0.998	1
0.01	0.236	0.940	0.995	1	0.469	0.955	1	1	0.527	0.978	1	1	0.112	0.901	1	1	0.202	0.946	1	1	0.294	0.968	1	1
0.1	0.605	0.999	1	1	0.672	0.999	1	1	0.699	0.999	1	1	0.521	1	1	1	0.592	1	1	1	0.620	1	1	1
	Trend E ($\gamma = 20$)				Trend E ($\gamma = 50$)				Trend E ($\gamma = 100$)				Trend E ($\gamma = 20$)				Trend E ($\gamma = 50$)				Trend E ($\gamma = 100$)			
0	0.055	0.049	0.059	0.059	0.259	0.915	0.997	1	0.729	1	1	1	0.053	0.063	0.044	0.082	0.055	0.076	0.118	0.223	0.077	0.174	0.408	0.761
0.01	0.143	0.931	0.994	1	0.362	0.956	0.999	1	0.731	0.975	1	1	0.093	0.884	0.999	1	0.095	0.892	0.998	1	0.112	0.918	1	1
0.1	0.599	1	1	1	0.684	0.999	1	1	0.749	0.999	1	1	0.511	1	1	1	0.517	1	1	1	0.535	1	1	1
	Trend F ($\gamma = 20$)				Trend F ($\gamma = 50$)				Trend F ($\gamma = 100$)				Trend F ($\gamma = 20$)				Trend F ($\gamma = 50$)				Trend F ($\gamma = 100$)			
0	0.239	0.919	1	1	0.596	1	1	1	0.523	1	1	1	0.048	0.058	0.062	0.069	0.151	0.600	0.906	0.996	0.429	0.994	1	1
0.01	0.334	0.950	0.999	1	0.600	0.973	0.997	1	0.571	0.979	0.999	1	0.103	0.890	0.999	1	0.210	0.925	0.999	1	0.493	0.978	1	1
0.1	0.648	0.999	1	1	0.708	0.999	1	1	0.735	1	1	1	0.519	1	1	1	0.566	1	1	1	0.699	1	1	1
	Trend G				Trend H				Trend I				Trend G				Trend H				Trend I			
0	0.018	0.118	0.323	0.754	0.010	0.023	0.050	0.117	0.015	0.056	0.071	0.069	0.067	0.151	0.296	0.576	0.066	0.062	0.087	0.164	0.040	0.085	0.129	0.207
0.01	0.104	0.916	0.998	1	0.071	0.894	0.996	1	0.092	0.897	0.998	1	0.112	0.909	0.999	1	0.100	0.888	0.999	1	0.104	0.899	0.999	1
0.1	0.559	1	1	1	0.559	1	1	1	0.593	1	1	1	0.525	1	1	1	0.519	1	1	1	0.524	1	1	1

BEL1 denotes the test where the frequency k is estimated by minimizing the SSR; $k^{max} = 5$
BEL2 denotes the test where the first two cumulative frequencies are included
i.i.d. $N(0, 1)$ errors; 5 % significance

3.2 Simulation analysis in time series

Then we investigated the finite sample properties of the test under more general error processes. In particular, ε_t was generated according to the following DGPs:

1. AR(1) model: $\varepsilon_t = \rho\varepsilon_{t-1} + \upsilon_t$, with $\rho = 0.5, 0.2, 0, -0.2$.

2. MA(1) model: $\varepsilon_t = \upsilon_t + \rho\upsilon_{t-1}$, for $\rho = 0.5, 0.2, -0.2$.
   We also investigated four nonlinear processes analyzed by Hong and Lee (2003) and Escanciano (2006), namely:

3. AR(1) model with heteroskedasticity (ARHET): $\varepsilon_t = \rho_1\varepsilon_{t-1} + h_t\upsilon_t$; $h_t^2 = 0.1 + 0.1\varepsilon_{t-1}^2 + \rho_2\varepsilon_{t-2}^2$.
4. AR(1) model plus a bilinear term (AR-BIL): $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-1}\upsilon_t + \upsilon_t$.
5. Bilinear model (BIL): $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2}\upsilon_{t-1} + \upsilon_t$.
6. Nonlinear moving average model (NLMA): $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\upsilon_{t-1}\upsilon_{t-2} + \upsilon_t$.

For these last models we considered $\rho_1 = 0.5$ and $\rho_2 = 0.1$. Again, the basis processes $\{u_t\}$ and $\{\upsilon_t\}$ were i.i.d. $N(0, 1)$.

For time series only trend specifications (A)–(C) were analyzed and 2,000 replications were carried out. Also, the slightly more conservative rule $m_T = [4T^{1/5}]$, $m_T^{(d)} = [0.85 \times 4T^{1/5}]$ was applied. The choice $m_T^d \leq m_T$ was not mandatory in light of the theoretical results above, although in our simulations it enabled easier control of size.

In this case it is necessary to treat autocorrelation. A number of papers have proposed several methods for long-run variance estimation, and analyzed the finite sample behavior of the stationarity test under these proposals (e.g., Kwiatkowski et al. 1992; Kurozumi 2002; Sul et al. 2005). In our simulations these methods did not provide satisfactory results. This is not surprising given Propositions 1 to 3 above, as the probability orders of most magnitudes differ from their analogues in parametric stationarity testing, this mainly being a consequence of the slower convergence rates of the nonparametric estimators for the trend function, as compared with their parametric counterparts. This implies that standard corrections for autocorrelation are generally invalid in the nonparametric setting.

The poor performance of standard autocorrelation treatments in our nonparametric setting led us to devise a new strategy. We combined (in a somewhat *ad hoc* fashion) the technical apparatus of Proposition 3 above with methods adapted from previous approaches. Our proposal is oriented to ensure a suitable performance of the nonparametric test in most common applications, although it should only be seen as a sensible starting point, and further refinements should be pursued in future research. As estimator for σ^2 we used the following truncation estimator (i.e., a kernel estimator with rectangular kernel): $\widehat{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} (T - |i| - m_T^d - 1)^{-1} \sum_{t=1+|i|}^{T} e_t^{(d)} e_{t-|i|}^{(d)}$. In order to select the bandwidth parameter ℓ_T, a data-driven rule was applied. We considered values in the interval $\ell_T^{(-)} \leq \ell_T \leq \ell_T^{(+)}$, with $\ell_T^{(-)}$ and $\ell_T^{(+)}$ being deterministic limits fixed in advance. By imposing $\ell_T^{(-)} \to \infty$ and (e.g.) $\ell_T^{(+)} = [cT^{1/5}]$ ($c > 0$), the rate $\ell_T = O_p(T^{1/5})$ is achieved and consistency of the test is ensured in many common applications. The following scheme, adequate for AR(1) or MA(1) error processes, is then applied to obtain ℓ_T (as Kurozumi (2002) rule, the scheme uses a tuning parameter, k; in our simulations we set $k = 0.5$; higher values are recommended in case of stronger autocorrelation):

1. Set $\ell_T^{(-)}$ and $\ell_T^{(+)}$. (Here, $\ell_T^{(-)} = 0$, $\ell_T^{(+)} = [2T^{1/5}k]$ were fixed.) Set K_T, the maximum lag order permitted in AR fitting (here, $K_T = [2T^{1/5}k]$ was used).
2. Fit AR(p) models to the residual vector $\mathbf{e}^{(d)}$, with $p = 0, \ldots, K_T$, and select the order p^* that minimizes Schwarz's information criterion (SIC). Go to step 3.

3a. If $p^* = 0$, set $\ell_T = \ell_T^{(-)}$.

3b. If $p^* = 1$, set $\ell_T = \min([20|b_T|k], \ell_T^{(+)})$, with b_T being the regression coefficient of the fitted AR(1) model.

3c. If $p^* > 1$, compute sample autocorrelations (r_i) of $\mathbf{e}^{(d)}$, with $i = 1, \ldots, \ell_T^{(+)}$, and select i^* such that $|r_{i^*}| = \max_{i=1,\ldots,\ell_T^{(+)}}\{r_{i^*}\}$. Set $\ell_T = \min(\max(i^*, p^*), \ell_T^{(+)})$.

The above procedure performed well under a wide range of circumstances (several specifications of the trend, stochastic characteristics of the process, autocorrelation levels). Further refinements would be available in specific cases where more detailed knowledge of the nature of the error process is available *a priori*. Results are reported in Tables 3 and 4 below.

As compared with the i.i.d. case, some impairment of the finite sample performance of the test is observed, with slight distortions in size and loss of power. As expected, the behavior of the test improves as sample size increases (particularly for series with $T \geq 500$). Regarding the error processes considered, it is observed that the test works better in series with nonnegative autocorrelations.

4 Empirical applications

To illustrate the application of the test, we investigated two time series: the daily series of the Japanese yen/US dollar exchange rate and the FTSE Eurotop 100 index.

4.1 The daily Japanese yen/US dollar exchange rate series

The data range from July 7, 2002 to July 7, 2007 (1,827 observations) and were analyzed, together with other exchange rate series, by Brooks (2008) in the context of vector autoregressive estimation. These financial series—or their first differences—appear to exhibit nonlinear patterns (e.g., Mills and Markellos 2008) of the kind analyzed in the above section.

First, we outline the details of stationarity testing for this exchange rate series. The deterministic rule $m_T = [4T^{1/5}]$ was applied, according to results from Sect. 3. As sample size is $T = 1{,}827$, this gives $m_T = 17$. Figure 1 below displays the series and its fitted trend, under this model complexity. The numerator of the test statistic is straightforwardly computed upon the residuals of the OLS regression $y_t = \widehat{\beta}_0 + \sum_{j=1}^{17} \widehat{\beta}_j \sqrt{2} \cos(j\pi t/T) + e_t$. This gives $S_T = 0.377$ for the raw "KPSS" statistic.

For estimation of σ^2 we used the rule outlined in the previous section. In order to have better control over the size of the test, the residuals used to compute $\widehat{\sigma}^2$ were

Table 3 Size and power of the nonparametric stationarity test

	ρ	$q\backslash T$	Trend A						Trend B						Trend C					
			100	300	500	1000	1500	2000	100	300	500	1000	1500	2000	100	300	500	1000	1500	2000
AR(1)	0.5	0	0.256	0.056	0.06	0.079	0.076	0.104	0.213	0.059	0.063	0.074	0.090	0.075	0.185	0.047	0.055	0.088	0.084	0.081
		0.01	0.231	0.072	0.09	0.262	0.487	0.735	0.216	0.069	0.09	0.256	0.486	0.726	0.223	0.057	0.102	0.259	0.487	0.729
		0.1	0.186	0.101	0.247	0.692	0.943	0.999	0.196	0.105	0.256	0.693	0.935	0.996	0.223	0.100	0.260	0.709	0.949	0.996
	0.2	0	0.223	0.156	0.105	0.084	0.092	0.097	0.230	0.175	0.090	0.088	0.098	0.108	0.208	0.155	0.112	0.135	0.123	0.103
		0.01	0.236	0.154	0.244	0.705	0.914	0.983	0.238	0.166	0.258	0.717	0.908	0.985	0.184	0.137	0.249	0.738	0.905	0.983
		0.1	0.202	0.184	0.544	0.961	0.980	0.997	0.214	0.190	0.527	0.971	0.979	0.999	0.143	0.160	0.505	0.973	0.971	0.994
	0	0	0.151	0.075	0.063	0.058	0.061	0.071	0.157	0.068	0.067	0.059	0.063	0.068	0.138	0.087	0.105	0.147	0.094	0.069
		0.01	0.135	0.200	0.485	0.927	0.960	0.984	0.145	0.211	0.495	0.930	0.962	0.993	0.139	0.228	0.493	0.928	0.952	0.979
		0.1	0.185	0.471	0.700	0.974	0.980	0.994	0.213	0.433	0.713	0.975	0.977	0.997	0.188	0.444	0.683	0.968	0.975	0.998
	−0.2	0	0.586	0.285	0.165	0.114	0.117	0.101	0.620	0.292	0.187	0.129	0.135	0.120	0.324	0.234	0.204	0.270	0.185	0.119
		0.01	0.577	0.364	0.607	0.929	0.958	0.987	0.617	0.386	0.620	0.940	0.952	0.986	0.302	0.327	0.595	0.936	0.953	0.987
		0.1	0.241	0.436	0.680	0.978	0.972	0.998	0.258	0.449	0.686	0.970	0.977	0.998	0.188	0.420	0.676	0.978	0.976	0.996
MA(1)	0.5	0	0.211	0.053	0.058	0.055	0.069	0.092	0.240	0.047	0.05	0.058	0.076	0.078	0.180	0.051	0.054	0.094	0.092	0.078
		0.01	0.228	0.074	0.144	0.503	0.788	0.932	0.221	0.078	0.157	0.513	0.783	0.942	0.175	0.074	0.159	0.507	0.781	0.944
		0.1	0.134	0.189	0.493	0.959	0.980	0.997	0.124	0.176	0.507	0.949	0.985	0.998	0.105	0.157	0.496	0.955	0.987	0.996
	0.2	0	0.200	0.126	0.072	0.068	0.058	0.060	0.224	0.122	0.083	0.069	0.06	0.082	0.186	0.133	0.091	0.118	0.078	0.076
		0.01	0.218	0.147	0.250	0.736	0.921	0.985	0.184	0.142	0.242	0.742	0.933	0.982	0.190	0.136	0.258	0.735	0.926	0.985
		0.1	0.201	0.206	0.518	0.97	0.973	0.999	0.244	0.187	0.495	0.959	0.979	0.994	0.153	0.185	0.494	0.964	0.972	0.997
	−0.2	0	0.631	0.343	0.218	0.120	0.107	0.104	0.607	0.350	0.229	0.123	0.135	0.104	0.342	0.242	0.191	0.293	0.179	0.110
		0.01	0.618	0.402	0.632	0.955	0.954	0.991	0.659	0.390	0.629	0.959	0.960	0.993	0.313	0.349	0.592	0.950	0.955	0.99
		0.1	0.270	0.493	0.699	0.978	0.985	0.997	0.238	0.458	0.682	0.977	0.975	0.998	0.170	0.428	0.687	0.979	0.977	0.997

Linear time Series; 5 % significance

Table 4 Size and power of the nonparametric stationarity test

	ρ_1	$q\backslash T$	Trend A						Trend B						Trend C					
			100	300	500	1000	1500	2000	100	300	500	1000	1500	2000	100	300	500	1000	1500	2000
ARHET	0.5	0	0.185	0.056	0.067	0.075	0.080	0.098	0.198	0.075	0.080	0.093	0.102	0.109	0.083	0.03	0.046	0.241	0.140	0.067
		0.01	0.196	0.190	0.490	0.837	0.946	0.993	0.201	0.204	0.544	0.890	0.951	0.991	0.039	0.160	0.519	0.892	0.959	0.991
		0.1	0.043	0.289	0.569	0.908	0.990	0.999	0.036	0.310	0.654	0.945	0.988	0.999	0.027	0.282	0.637	0.935	0.983	0.998
ARBIL	0.5	0	0.209	0.065	0.06	0.069	0.080	0.082	0.215	0.070	0.057	0.069	0.089	0.070	0.197	0.055	0.065	0.097	0.077	0.084
		0.01	0.226	0.075	0.092	0.269	0.463	0.718	0.253	0.074	0.078	0.256	0.483	0.719	0.192	0.066	0.084	0.275	0.457	0.718
		0.1	0.214	0.095	0.253	0.669	0.939	0.990	0.215	0.118	0.262	0.681	0.946	0.992	0.178	0.093	0.267	0.703	0.918	0.994
BIL	0.5	0	0.214	0.067	0.057	0.071	0.080	0.09	0.213	0.058	0.056	0.079	0.081	0.088	0.198	0.054	0.055	0.099	0.088	0.085
		0.01	0.206	0.065	0.121	0.371	0.550	0.737	0.207	0.059	0.117	0.340	0.535	0.739	0.192	0.052	0.115	0.358	0.559	0.732
		0.1	0.186	0.249	0.627	0.916	0.969	0.996	0.191	0.249	0.603	0.906	0.973	0.994	0.153	0.244	0.612	0.922	0.969	0.998
NLMA	0.5	0	0.213	0.057	0.051	0.067	0.06	0.084	0.211	0.053	0.054	0.071	0.082	0.081	0.185	0.047	0.054	0.091	0.085	0.087
		0.01	0.210	0.077	0.113	0.356	0.569	0.747	0.211	0.067	0.112	0.346	0.597	0.769	0.159	0.065	0.108	0.384	0.588	0.783
		0.1	0.215	0.239	0.618	0.913	0.973	0.997	0.177	0.247	0.618	0.931	0.979	0.994	0.153	0.234	0.626	0.913	0.977	0.996

Nonlinear time Series; 5 % significance level

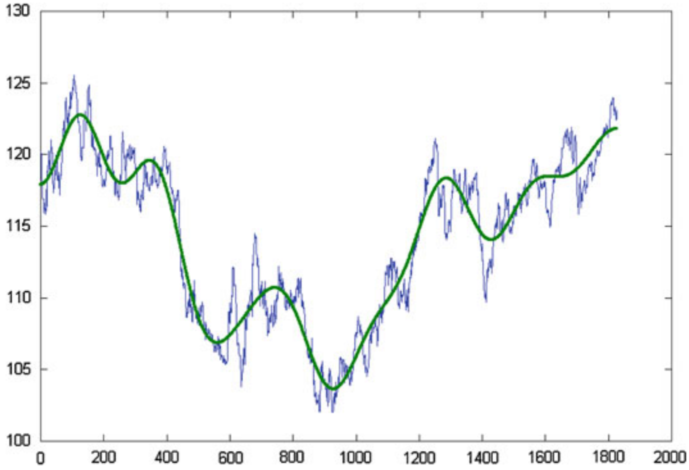

Fig. 1 Daily Japanese yen/US dollar exchange rate series (*broken line*) versus nonparametric fitted trend (*continuous line*)

Table 5 Standardization parameters for the nonparametric stationarity test

m_T	1	2	3	4	5	6	7	8	9	10
μ_{m_T}	0.06535	0.04002	0.02876	0.02242	0.01837	0.01556	0.01349	0.01191	0.01066	0.00964
s_{m_T}	0.04111	0.02017	0.01239	0.00856	0.00636	0.00496	0.00401	0.00333	0.00282	0.00243
m_T	11	12	13	14	15	16	17	18	19	20
μ_{m_T}	0.00881	0.00810	0.00750	0.00698	0.00653	0.00614	0.00579	0.00548	0.00519	0.00494
s_{m_T}	0.00212	0.00187	0.00167	0.00150	0.00135	0.00123	0.00113	0.00104	0.00096	0.00089
m_T	21	22	23	24	25	26	27	28	29	30
μ_{m_T}	0.00471	0.00450	0.00431	0.00413	0.00397	0.00382	0.00368	0.00355	0.00343	0.00332
s_{m_T}	0.00083	0.00077	0.00073	0.00068	0.00064	0.00061	0.00057	0.00054	0.00052	0.00049
m_T	31	32	33	34	35	36	37	38	39	40
μ_{m_T}	0.00322	0.00312	0.00302	0.00294	0.00285	0.00278	0.00270	0.00263	0.00256	0.00250
s_{m_T}	0.00047	0.00045	0.00043	0.00041	0.00039	0.00038	0.00036	0.00035	0.00033	0.00032

obtained from a slightly simpler cosine regression with $m_T^{(d)} = [0.85 \times 4T^{1/5}]$, and the (rectangular kernel) estimator was applied, i.e., $\widehat{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} (T - |i| - m_T^d - 1)^{-1} \sum_{t=1+|i|}^{T} e_t^{(d)} e_{t-|i|}^{(d)}$, with $\ell_T = 3$ obtained by applying the rule outlined in Sect. 3 above. This gives $\widehat{\sigma}^2 = 17.9819$.

In order to facilitate the application of the test, Table 5 below displays the standardization parameters μ_{m_T} and s_{m_T} for a representative range of values of m_T. In our case, as $m_T = 17$, the table gives $\mu_{m_T} = 0.00579$ and $s_{m_T} = 0.00113$. Therefore, the observed value of the test statistic is $\widehat{Z}_T = s_{m_T}^{-1}(\widehat{\sigma}^{-2} S_T - \mu_{m_T}) = 13.446$ which, by checking the $N(0, 1)$ distribution, indicates that the null of stationarity around a deterministic trend is rejected at the critical level $p = 0.000$.

In a second stage the analysis was extended to the first difference of the series, for which the fitted trend was computed again under $m_T = 17$. In this case $\widehat{Z}_T = 0.122$

Table 6 Sensitivity analysis

Japanese yen/US dollar exchange rate						Japanese yen/US dollar exchange rate (differences)					
	$\ell_T = 1$	$\ell_T = 2$	$\ell_T = 3$	$\ell_T = 4$	$\ell_T = 5$		–	–	$\ell_T = 0$	$\ell_T = 1$	$\ell_T = 2$
$m_T = 16$	32.840	18.290	12.077	8.652	6.482	$m_T = 16$	–	–	0.125	−0.0004	−0.355
$m_T = 17$	36.056	20.212	**13.446**	9.716	7.353	$m_T = 17$	–	–	**0.122**	−0.007	−0.372
$m_T = 18$	38.279	21.524	14.369	10.425	7.926	$m_T = 18$	–	–	0.433	0.293	−0.103
FTSE Eurotop 100 index (logs)						FTSE Eurotop 100 (returns)					
	$\ell_T = 3$	$\ell_T = 4$	$\ell_T = 5$	$\ell_T = 6$	$\ell_T = 7$		$\ell_T = 3$	$\ell_T = 4$	$\ell_T = 5$	$\ell_T = 6$	$\ell_T = 7$
$m_T = 16$	12.934	9.408	7.199	5.672	4.558	$m_T = 16$	−0.350	−1.440	−0.377	−0.185	0.943
$m_T = 17$	10.430	7.368	**5.449**	4.123	3.155	$m_T = 17$	−0.228	−1.382	**−0.257**	−0.054	1.140
$m_T = 18$	11.534	8.226	6.154	4.721	3.676	$m_T = 18$	0.030	−1.218	−0.001	0.219	1.512

Values of the test statistic under moderate variations of m_T and ℓ_T
In bold type the value of the statistic for m_T and ℓ_T obtained according to the data-driven rules
Critical values of the N(0,1) distribution: 1.282, 1.645 and 2.326, at 10, 5 and 1 % significance levels, respectively

(critical level $p = 0.451$), with $\ell_T = 0$ selected by the data-driven device. So, the null of stationarity cannot be rejected.

In empirical applications it is advisable to carry out a sensitivity analysis in order to assess the robustness of the test's results under moderate variations of m_T and ℓ_T. Table 6 displays this analysis for the daily Japanese yen/US dollar exchange rate series, which indicates that conclusions remain unaffected.

These results, together with the output from mainstream unit root tests, coincide to suggest that the daily Japanese yen/US dollar exchange rate series has a single unit root. This conclusion is also in accordance with predictions from financial theory.

4.2 The daily series of the FTSE Eurotop 100 index

Then we analyzed the closing prices of the daily series of the FTSE Eurotop 100 index spanning the period from 17 December 2002 to 30 October 2009 (1,739 observations) (available at http://www.fin-rus.com/analysis/export_eng_/default.asp). Figure 2 plots the logarithm of the series and the fitted trend, which was computed for $m_T = 17$. For estimation of the long-run variance the data-driven device selected $\ell_T = 5$, and the observed value of the test statistic was $\widehat{Z}_T = 5.449$ (critical level $p = 0.000$). On the contrary, the extension of the analysis to the first difference of the series (i.e., the return series) indicated that the null of stationarity around a deterministic trend cannot be rejected for returns: the value of the test statistic is $\widehat{Z}_T = -0.257$—critical level $p = 0.601$—with $\ell_T = 5$ selected by the above procedure to estimate σ^2. The robustness of these conclusions under moderate changes of m_T and ℓ_T can be checked in Table 6 above.

These results are in accordance with both financial theories and a large amount of empirical research, all of them indicating that logarithms of asset prices contain a unit root, while asset return series do not.

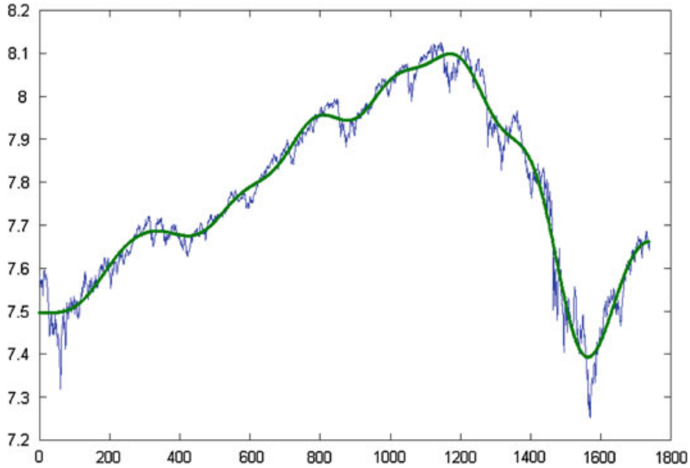

Fig. 2 Daily FTSE Eurotop 100 index series, in logarithms (*broken line*) versus nonparametric fitted trend (*continuous line*)

5 Concluding remarks and further research

We have proposed a nonparametric stationarity test which allows stationarity testing to be carried out without relying on a priori specification of the trend component, which tends to be problematic in practice. The test is consistent under unit root alternatives and its limiting null distribution is standard normal. Simulation analyses indicate that the test performs suitably in a wide range of circumstances (trend shapes, stochastic dependence structures).

The nonparametric stationarity test is a conservative solution, that provides a safeguard against misspecification of the trend function and achieves correct test size, at the cost of relatively large samples. Nonparametric estimation/testing is appropriate when the researcher lacks a reliable parametric model for the trend function, and as an exploratory tool. Its use is particularly indicated in large samples. If the researcher has genuine a priori knowledge that the trend obeys a specific parametric model, the standard stationarity test—under correct specification of the trend function (e.g., Landajo and Presno 2010, for general parametric specifications)—will generally outperform any nonparametric test, exhibiting higher power in small samples.

The issue of long-run variance estimation has also been addressed. The theoretical results allow nonparametric stationarity testing under nonparametric (kernel) estimation of the long-run variance, with a deterministic rule for bandwidth selection. In practice, data-driven bandwidth selection often tends to outperform deterministic rules in estimating long-run variances. A data-driven procedure to treat autocorrelation—with deterministic brackets that ensure the appropriate stochastic order for the estimator—has been outlined. Simulations indicate that this procedure performs suitably in common applications of the nonparametric stationarity test.

The above results suggest a number of interesting research avenues. First, more extensive analyses on autocorrelation treatment in nonparametric stationarity testing are clearly indicated. The theoretical validity of a number of procedures proposed in the literature (e.g., Sun et al. 2008; Hashimzade and Vogelsang 2008; Kurozumi and Tanaka 2010), still has to be established in the nonparametric case. Simulation studies are required in order to assess the empirical performance of the various methods of treating for autocorrelation in this new setting.

Finally, the proposed nonparametric approach relies on trigonometric series estimation of the trend function. This allowed a relatively simple mathematical analysis. Computer simulations suggest that analogue results may be obtained for other classes of series estimators, particularly for algebraic polynomials. A confirmation of this conjecture would be desirable (results in a classical paper by MacNeill 1978, are relevant for this extension), although the technical burden tends to increase dramatically when the cosine basis is replaced by other classes of polynomials.

Appendix Mathematical proofs

Notational issues and previous remarks

1. In the proof of limiting normality the following kernel is used: for any (positive integer) m, let

$$K_m(u,v) = \min(u,v) - uv - \sum_{j=1}^{m} 2(j\pi)^{-2}\sin(j\pi u)\sin(j\pi v);\ u,v\in[0,1] \tag{5}$$

This kernel has the (uniformly convergent) Mercer expansion

$$K_m(u,v) = \sum_{j=m+1}^{\infty} 2(j\pi)^{-2}\sin(j\pi u)\sin(j\pi v), \tag{6}$$

with (reciprocal) eigenvalues $\eta_j = (m+j)^{-2}\pi^{-2}$, $j=1,2,\ldots$, $\mu_m = \int_0^1 K_m(u,u)\mathrm{d}u = \sum_{j=m+1}^{\infty}(j\pi)^{-2}$ and $s_m^2 = 2\int_0^1\int_0^1 K_m^2(u,v)\,\mathrm{d}u\,\mathrm{d}v = 2\sum_{j=m+1}^{\infty}(j\pi)^{-4}$ (for further details, see Tanaka 1996, Chapter 5, p. 153). It is readily checked that $s_m^2 \geq 2\pi^{-4}\int_{m+1}^{\infty}x^{-4}\mathrm{d}x = 2/3\pi^{-4}(m+1)^{-3}$, so $s_{m_T}^{-1} \leq \sqrt{3/2}\pi^2(m+1)^{3/2}$.

2. Let $r_{m_T}(x) = \theta^*(x) - \theta_{m_T}(x)$, $x\in[0,1]$. Given T, the OLS residuals have the decomposition $e_t = \varepsilon_t + r_{m_T}(t/T) + \theta_{m_T}(t/T) - \widehat{\theta}_{m_T}(t/T)$. In matrix form, $\mathbf{e} = (e_1,\ldots,e_T)' = \mathbf{\Pi}_{m_T}\boldsymbol{\varepsilon} + \mathbf{\Pi}_{m_T}\mathbf{r}_{m_T} + \mathbf{\Pi}_{m_T}\mu$, with $\mathbf{\Pi}_{m_T} = (\mathbf{I}_T - \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}')$, $\mathbf{\Phi} = [\varphi_{t,j}]$, $t=1,\ldots,T$, $j=0,\ldots,m_T$; $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_T)'$, $\mu = (\mu_1,\ldots,\mu_T)'$ and $\mathbf{r}_{m_T} = (r_{m_T}(1/T),\ldots,r_{m_T}(T/T))'$.

The following previous lemmas are required.

Lemma 1 *Let* $\widetilde{\mathbf{B}}_m = [\widetilde{b}_{s,t}^{(m)}] = T^{-1}\mathbf{C}_T \mathbf{\Pi}_m \mathbf{C}_T'$, *with* m *fixed. Let* $\widetilde{S}_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} \widetilde{b}_{s,t}^{(m)} u_s u_t$ *and* $S_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K_m(s/T, t/T)u_s u_t$, *where* $\mathbf{u} = (u_1, \ldots, u_T)'$ *is a sequence of i.i.d. random variables with* $E(u_i) = 0$ *and* $Var(u_i) = \sigma_u^2 < \infty$. *Then, for some* $c < \infty$ *not depending on* m *or* T, *as* $T \to \infty$:

(a) $\sup_{s,t=1,\ldots,T} |\widetilde{b}_{s,t} - K_m(s/T, t/T)| \leq cm^3 T^{-1}$, (b) $E|\widetilde{S}_T - S_T| \leq cm^3 T^{-1}$. □

Proof As to part (a), (if $T \geq m+1$) we have $\widetilde{b}_{s,t}^{(m)} = \min(s/T, t/T) - \widetilde{\mathbf{g}}_s' \widetilde{\mathbf{H}}_{m+1}^{-1} \widetilde{\mathbf{g}}_t$, with $\widetilde{\mathbf{H}}_{m+1} = [\widetilde{h}_{j,k}] = T^{-1}\mathbf{\Phi}'\mathbf{\Phi}$, $j, k = 0, \ldots, m$, and $\widetilde{\mathbf{g}}_t = [\widetilde{g}_{o,t}, \ldots, \widetilde{g}_{m,t}]'$, with $\widetilde{g}_{j,t} = T^{-1}\sum_{i=1}^{t} \varphi_j(i/T)$. We also have $K_m(s/T, t/T) = \min(s/T, t/T) - \mathbf{g}_s{}'\mathbf{H}_{m+1}^{-1}\mathbf{g}_t$, where $\mathbf{H}_{m+1} = [h_{j,k}]$, with $h_{j,k} = \int_0^1 \varphi_j(u)\varphi_k(u)\mathrm{d}u$, $j, k = 0, \ldots, m$, and $\mathbf{g}_t = [g_{o,t}, \ldots, g_{m,t}]'$, with $g_{j,t} = \int_0^{t/T} \varphi_j(u)\mathrm{d}u$. Orthonormality of the basis ensures $\mathbf{H}_{m+1} = \mathbf{I}_{m+1}$, so $K_m(s/T, t/T) = \min(s/T, t/T) - \mathbf{g}_s'\mathbf{g}_t$ and $\widetilde{b}_{s,t}^{(m)} - K_m(s/T, t/T) = \mathbf{g}_s'\mathbf{g}_t - \widetilde{\mathbf{g}}_s'\widetilde{\mathbf{H}}_{m+1}^{-1}\widetilde{\mathbf{g}}_t = A_1 + A_2$, with $A_1 = \mathbf{g}_s'\mathbf{g}_t - \widetilde{\mathbf{g}}_s'\widetilde{\mathbf{g}}_t$ and $A_2 = \widetilde{\mathbf{g}}_s'\widetilde{\mathbf{g}}_t - \widetilde{\mathbf{g}}_s'\widetilde{\mathbf{H}}_{m+1}^{-1}\widetilde{\mathbf{g}}_t$. Discretization arguments and standard inequalities for eigenvalues ensure $|A_1| \leq c(1+m)^2 T^{-1}$ and $A_2 = O(m^3 T^{-1})$. Part (b) directly follows from part (a) and Lemma 3 in Nabeya and Tanaka (1988). □

Lemma 2 *Let* $S_{1T} = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{m_T}(s/T, t/T)v_s v_t$, *where the components of* $\mathbf{v} = (v_1, \ldots, v_T)'$ *are i.i.d. random variables with* $E(v_i) = 0$, $Var(v_i) = \sigma_v^2 > 0$ *and* $E|v_i|^{2+\delta} < \infty$, $\delta > 0$. *Let* $Z_{1T} = s_{m_T}^{-1}(\sigma_v^{-2}S_{1T} - \mu_{m_T})$, *with* $\mu_{m_T} = \int_0^1 K_{m_T}(u, u)\mathrm{d}u = \sum_{j=m_T+1}^{\infty}(j\pi)^{-2}$, $s_{m_T}^2 = 2\int_0^1\int_0^1 K_{m_T}^2(u, v)\,\mathrm{d}u\,\mathrm{d}v = 2\sum_{j=m_T+1}^{\infty}(j\pi)^{-4}$. *If* $m_T \to \infty$ *and* $m_T^3 T^{-1} \to 0$, *then* $Z_{1T} \xrightarrow{L} N(0, 1)$ *as* $T \to \infty$. □

Proof Without loss of generality we assume $\sigma_v^2 = 1$. It suffices to check that a central limit theorem for quadratic forms (with nonvanishing diagonal) in i.i.d. random variables holds. We apply Theorem 2.1.(iii) in Bhansali et al. (2007). We have $S_{1T} = \mathbf{u}'\mathbf{D}\mathbf{u} = \sum_{s=1}^{T}\sum_{t=1}^{T} d_{s,t}v_s v_t$, with $\mathbf{D} = [d_{s,t}]$ and $d_{s,t} = T^{-1}K_{m_T}(s/T, t/T)$. Let $\|\mathbf{D}\|_{2,T} = \sqrt{\sum_{s=1}^{T}\sum_{t=1}^{T} d_{s,t}^2}$ and $\|\mathbf{D}\|_{sp,T} = \widetilde{\eta}_1$, with $\widetilde{\eta}_1$ being the largest eigenvalue of $\mathbf{D}$. We start by deriving limiting normality for $\widetilde{Z}_{1T} = (\sqrt{2}\|\mathbf{D}\|_{2,T})^{-1}(S_{1T} - E(S_{1T}))$. This follows under the conditions (1) $\|\mathbf{D}\|_{sp,T}/\|\mathbf{D}\|_{2,T} \to 0$ as $T \to \infty$, and (2) $\sum_{t=1}^{T} d_{t,t}^2 = o(\|\mathbf{D}\|_{2,T}^2)$. These requirements can be readily checked upon the basis of the (Euclidean, spectral) norms of kernel $K_{m_T}(\cdot, \cdot)$, namely,

$$\|K_{m_T}\|_2 = \sqrt{\int_0^1\int_0^1 K_{m_T}^2(u, v)\,\mathrm{d}u\,\mathrm{d}v} = \sqrt{\sum_{j=m_T+1}^{\infty}(j\pi)^{-4}} \tag{7}$$

—this implies $\|K_{m_T}\|_2^{-1} = O(m_T^{3/2})$—, and $\|K_{m_T}\|_{sp} = \eta_1 = (m_T+1)^{-2}\pi^{-2} = O(m_T^{-2})$. Discretization arguments and Aronszajn Theorem give $\|\mathbf{D}\|_{sp,T} = \widetilde{\eta}_1 \leq$

$\eta_1 + |\widetilde{\eta}_1 - \eta_1| = O\big(m_T^{-2}\big)$ and $\|\mathbf{D}\|_{2,T} = O(m_T^{-3/2})$. Hence, we obtain $\|\mathbf{D}\|_{sp,T}/\|\mathbf{D}\|_{2,T} = O(m_T^{-1/2})$. This is directly checked as $\sum_{t=1}^{T} d_{t,t}^2 = O(m_T^{-2}T^{-1})$ and $\|\mathbf{D}\|_{2,T}^2 = O\big(m_T^{-3}\big)$. So, $\widetilde{Z}_{1T} \xrightarrow{L} N(0,1)$, and as $Z_{1T} = \widetilde{Z}_{1T} + O_p(m_T^{5/2}T^{-1})$ the conclusion follows. □

Lemma 3 *Let $S_{1T} = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{m_T}(s/T, t/T)\varepsilon_s\varepsilon_t$, with $\varepsilon_t = \sum_{i=0}^{\infty}\alpha_i v_{t-i}$, under Assumption 1 and $\sigma^2 = \alpha^2\sigma_v^2$. Let $Z_T = s_{m_T}^{-1}(\sigma^{-2}S_{1T} - \mu_{m_T})$. If $m_T \to \infty$ and $m_T^3 T^{-1} \to 0$ then $Z_T \xrightarrow{L} N(0,1)$ as $T \to \infty$.* □

Proof Without loss of generality we assume $\sigma_v^2 = 1$, so $\sigma^2 = \sum_{i=0}^{\infty}\sum_{k=0}^{\infty}\alpha_i\alpha_k > 0$, and proceed as in Tanaka (1990, Theorem 1, Appendix). First, we have

$$\begin{aligned} Z_T &= s_{m_T}^{-1}\sigma^{-2}\left(T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{m_T}(s/T, t/T)\varepsilon_s\varepsilon_t - \mu_{m_T}\sigma^2\right) \\ &= \sigma^{-2}\sum_{i=0}^{\infty}\sum_{k=0}^{\infty}\alpha_i\alpha_k w_{T,i,k} \end{aligned} \tag{8}$$

with $w_{T,i,k} = s_{m_T}^{-1}[T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{m_T}(s/T, t/T)v_{s-i}v_{t-k} - \mu_{m_T}]$. The following decomposition is applicable: $Z_T = Z_{T,M} + V_{T,M}$, where $Z_{T,M} = \sigma^{-2}\sum_{i=0}^{M}\sum_{k=0}^{M}\alpha_i\alpha_k w_{T,i,k}$ and $V_{T,M}$ is the remainder term. It is readily checked that $E|V_{T,M}| \le c_M$ for all T large, with $c_M \equiv 3\sigma^{-2}(1+\delta)\sum_{i=0}^{\infty}|\alpha_i|\cdot\sum_{k=M+1}^{\infty}|\alpha_k| \to 0$ as $M \to \infty$.

As to $Z_{T,M}$, first fix M and let $T \to \infty$. We have $Z_{T,M} = \widetilde{Z}_{T,M} + R_{T,M}$, with $\widetilde{Z}_{T,M} = w_{T,0,0}(\sigma^{-2}\sum_{i=0}^{M}\sum_{k=0}^{M}\alpha_i\alpha_k)$ and $R_{T,M} = \sigma^{-2}\sum_{i=0}^{M}\sum_{k=0}^{M}\alpha_i\alpha_k(w_{T,i,k} - w_{T,0,0})$. As $w_{T,0,0} \xrightarrow{L} N(0,1)$ by Lemma 2, we have (for any fixed M, as $T \to \infty$) that $\widetilde{Z}_{T,M} \xrightarrow{L} \widetilde{Z}_M$, which is Gaussian with mean zero and variance $(\sigma^{-2}\sum_{i=0}^{M}\sum_{k=0}^{M}\alpha_i\alpha_k)^2$.

As to the remainder $R_{T,M}$, for fixed M as $T \to \infty$, it holds $E|w_{T,i,k} - w_{T,0,0}| \le c'M^2 m_T^{5/2}T^{-1}$, with c' depending neither on $0 \le i, k \le M$, nor on M or T. Hence, $\limsup_{T\to\infty} E|w_{T,i,k} - w_{T,0,0}| = 0$ and the same applies, for any fixed M, to $R_{T,M}$.

Hence, we have, for any M, $\widetilde{Z}_{T,M} \xrightarrow{L} \widetilde{Z}_M$ as $T \to \infty$. It also holds $\widetilde{Z}_M \xrightarrow{L} N(0,1)$ as $M \to \infty$, by the continuous mapping Theorem. Since $Z_T = Z_{T,M} + V_{T,M} = \widetilde{Z}_{T,M} + R_{T,M} + V_{T,M}$ and $E|Z_T - \widetilde{Z}_{T,M}| \le E|R_{T,M}| + E|V_{T,M}|$, Theorem 4.2 in Billingsley (1968) and Tchebyshev's inequality give $Z_T \xrightarrow{L} N(0,1)$. □

Proof of Proposition 1 As to part (**a**), since $\mathbf{e} = \mathbf{\Pi}_{m_T}\varepsilon + \mathbf{\Pi}_{m_T}\mathbf{r}$, the decomposition $S_T = S_{1T} + A_1 + A_2$, with $S_{1T} = T^{-2}\varepsilon'\mathbf{\Pi}_{m_T}'\mathbf{C}_T'\mathbf{C}_T\mathbf{\Pi}_{m_T}\varepsilon$, $A_1 = T^{-2}\mathbf{r}'\mathbf{\Pi}_{m_T}'\mathbf{C}_T'\mathbf{C}_T\mathbf{\Pi}_{m_T}\varepsilon + T^{-2}\varepsilon'\mathbf{\Pi}_{m_T}'\mathbf{C}_T'\mathbf{C}_T\mathbf{\Pi}_{m_T}\mathbf{r}$, and $A_2 = T^{-2}\mathbf{r}'\mathbf{\Pi}_{m_T}'\mathbf{C}_T'\mathbf{C}_T\mathbf{\Pi}_{m_T}\mathbf{r}$, is directly obtained. Hence, $Z_T = \widetilde{Z}_{1T} + s_{m_T}^{-1}\sigma^{-2}A_1 + s_{m_T}^{-1}\sigma^{-2}A_2$, with $\widetilde{Z}_{1T} = s_{m_T}^{-1}(\sigma^{-2}S_{1T} - \mu_{m_T})$.

Limiting normality of $\widetilde{Z}_{1T}$ is readily checked. It suffices to derive limiting normality for $\widetilde{Z}_{2T} = s_{m_T}^{-1}(\sigma^{-2}\widetilde{S}_{2T} - \mu_{m_T})$, with $\widetilde{S}_{2T} = T^{-1}\varepsilon'\mathbf{B}_T\varepsilon$ and $\mathbf{B}_T = [b_{s,t}] =$

$T^{-1}\mathbf{C}_T\Pi_{m_T}\mathbf{C}'_T$. First we approximate $\widetilde{S}_{2T}$ by $S_{2T} = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{m_T}(s/T, t/T)\varepsilon_s\varepsilon_t$. Let $R_T = \widetilde{S}_{2T} - S_{2T} = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}[b_{s,t} - K_{m_T}(s/T, t/T)]\varepsilon_s\varepsilon_t$ and $\delta_T \equiv \sup_{s,t=1,\ldots,T}|\widetilde{b}_{s,t} - K_m(s/T, t/T)|$. By Lemma 1, $\delta_T = O(m_T^3 T^{-1})$. A probability inequality in Tanaka (1990, Appendix, Theorem 1) ensures, for any $x > 0$, that $P(|R_T| > x) \le (c/x)\delta_T(\sum_{i=0}^{\infty}|\alpha_i|)^2$ for some constant $c > 0$ not depending on T. Hence, $R_T = O_p(\delta_T) = O_p(m_T^3 T^{-1})$. Therefore, $\widetilde{Z}_{2T} = Z_{2T} + s_{m_T}^{-1}\sigma^{-2}R_T$, with $Z_{2T} = s_{m_T}^{-1}(\sigma^{-2}S_{2T} - \mu_{m_T})$. As $s_{m_T}^{-1} = O(m_T^{3/2})$, we obtain $s_{m_T}^{-1}\sigma^{-2}R_T = O_p(m_T^{9/2}T^{-1})$, which is asymptotically negligible as $m_T^{9/2}T^{-1}\longrightarrow 0$. Lemma 3 gives $Z_{2T} \xrightarrow{L} N(0, 1)$. Therefore, $\widetilde{Z}_{2T} \xrightarrow{L} N(0, 1)$, which amounts to $\widetilde{Z}_{1T} \xrightarrow{L} N(0, 1)$.

It is readily checked that the bias terms have lower probability orders than $\widetilde{Z}_T$. In particular, the basic projection inequality of least squares regression ensures $s_{m_T}^{-1}\sigma^{-2}A_1 = O_p(Tm_T^{3/2}d_T(\theta_{m_T}, \theta^*))$ and $s_{m_T}^{-1}\sigma^{-2}A_2 = O(Tm_T^{3/2}d_T^2(\theta_{m_T}, \theta^*))$. Both quantities are asymptotically negligible under H_0 given Assumption 2. As to (**b**), under H_1 we have $\mathbf{e} = \mathbf{\Pi}_{m_T}\varepsilon + \mathbf{\Pi}_{m_T}\mathbf{r} + \mathbf{\Pi}_{m_T}\mu$, that combined with standard inequalities gives $S_T = O_p(T^2)$. Hence, $Z_T = s_{m_T}^{-1}\sigma^{-2}S_T - s_{m_T}^{-1}\mu_{m_T} = O_p(m_T^{3/2}T^2) - O_p(m_T^{1/2}) = O_p(m_T^{3/2}T^2)$, so $P(Z_T > \kappa_T)\longrightarrow 1$ if $\kappa_T = o(m_T^{3/2}T^2)$. □

Proof of Proposition 2 As to (**a**), we have $\widetilde{Z}_T - Z_T = s_{m_T}^{-1}(\sigma^{-2}S_T - \mu_{m_T})\,\widehat{\sigma}^{-2}(\sigma^2 - \widehat{\sigma}^2) + \widehat{\sigma}^{-2}(\sigma^2 - \widehat{\sigma}^2)s_{m_T}^{-1}\mu_{m_T}$, and as $s_{m_T}^{-1}(\sigma^{-2}S_T - \mu_{m_T}) = O_p(1)$ by Proposition 1 and $\widehat{\sigma}^2 - \sigma^2 = o_p(1)$, it holds $\widetilde{Z}_T - Z_T = o_p(1)$ as Assumption 3 imposed $m_T^{1/2}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{p} 0$ under H_0. As to (**b**), as $S_T = O_p(T^2)$ we have $\widetilde{Z}_T = s_{m_T}^{-1}\widehat{\sigma}^{-2}S_T - s_{m_T}^{-1}\mu_{m_T} = \widehat{\sigma}^{-2}O_p(m_T^{3/2}T^2) - O(m_T^{1/2}) = O_p(m_T^{3/2}T^{2-\zeta})$ and the conclusion follows. □

Lemma 4 *Let* $\{\varphi_j, j = 0, 1, \ldots\}$ *be an orthonormal set in* $L_2[0, 1]$ *and let* $\widetilde{\mathbf{H}}_{m_T+1}^{-1} = (T^{-1}\mathbf{\Phi}'\mathbf{\Phi})^{-1}$, *with* $\mathbf{\Phi} = [\varphi_{t,j}]$, $\varphi_{t,j} = \varphi_j(t/T)$, $t = 1, \ldots, T$; $j = 0, \ldots, m_T$. *Let* $V_T = \mathbf{v}'\mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{v}$, *with* $\mathbf{v} = (v_1, \ldots, v_T)'$, *being a finite sample of the i.i.d. process in Assumption* 1. *Under the conditions:* (i) $\sup_{j\ge 0}\|\varphi_j\|_\infty \le \Delta < \infty$, *and each* φ_j *satisfies the Lipschitz condition* $|\varphi_j(x) - \varphi_j(x')| \le cj|x - x'|$; $x, x' \in [0, 1]$, *with* $c < \infty$ *not depending on* j, (ii) *for any fixed* m, $\|\widetilde{\mathbf{H}}_{m+1} - \mathbf{I}_{m+1}\|_{2,m+1} \le cm^2T^{-1}$ *as* $T \to \infty$, *with c not depending on m or T, and* (iii) $m = m_T \to \infty$ *and* $m_T^{9/2}T^{-1} \to 0$, *as* $T \to \infty$. *Then* $Q_T = (2(1 + m_T))^{-1/2}(\sigma_v^{-2}V_T - (1 + m_T)) \xrightarrow{L} N(0, 1)$ *as* $T \to \infty$. □

Proof This case is analogous to Hong and White (1995, Appendix, Theorem A.1). Without loss of generality we assume $\sigma_v^2 = 1$. Let $V_T = T^{-1}\mathbf{v}'\mathbf{B}_T\mathbf{v}$, with $\mathbf{B}_T = [b_{s,t}] = \mathbf{\Phi}(T^{-1}\mathbf{\Phi}'\mathbf{\Phi})^{-1}\Phi' = \mathbf{\Phi}\widetilde{\mathbf{H}}_{m_T+1}^{-1}\mathbf{\Phi}'$. For any T, we shall use the kernel $K'_{m_T}(u, v) = \sum_{j=0}^{m_T}\varphi_j(u)\varphi_j(v)$; $u, v \in [0, 1]$, which is degenerate, with reciprocal eigenvalues $\eta'_1 = \cdots = \eta'_{m_T+1} = 1$, which implies $\|K'_{m_T}\|_{sp} = 1$ and $\|K'_{m_T}\|_2 = m_T + 1$.

The analysis proceeds as in Lemma 1. First, we approximate V_T by $\widetilde{V}_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} K'_{m_T}(s/T, t/T)v_s v_t$ and obtain, by Lemma 3 in Nabeya and Tanaka

(1988), $V_T = \widetilde{V}_T + O_p(m_T^5 T^{-1})$. The rest of the proof is analogous to that of Lemma 2. $\square$

Lemma 5 *Under the assumptions of Lemma 4, let* $V_T = \varepsilon' \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\varepsilon$, *with* $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$ *being a finite sample from the linear filter process in Assumption 1. Then* $Q_T = (2(1+m_T))^{-1/2}(\sigma^{-2}V_T - (1+m_T)) \xrightarrow{L} N(0,1)$ *as* $T \to \infty$. $\square$

Proof It is analogous to the proof of Lemmas 2 and 4. Without loss of generality we assume $\sigma_v^2 = 1$, so $\sigma^2 = \sum_{i=0}^{\infty}\sum_{k=0}^{\infty}\alpha_i\alpha_k > 0$. Let $V_T = T^{-1}\varepsilon'\mathbf{B}_T\varepsilon$, with $\mathbf{B}_T = [b_{s,t}] = \mathbf{\Phi}(T^{-1}\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}' = \mathbf{\Phi}\widetilde{\mathbf{H}}_{m_T+1}^{-1}\mathbf{\Phi}'$. First, we approximate V_T by $\widetilde{V}_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}K'_{m_T}(s/T, t/T)\varepsilon_s\varepsilon_t$. Lemma 4 and the same probability inequality used in Lemma 2 above (see Tanaka 1990, Theorem 1, Appendix) give $V_T - \widetilde{V}_T = O_p(m_T^5 T^{-1})$. The rest of the proof is analogous to that of Lemma 3. $\square$

Proof of Proposition 3 We have the decomposition $\mathbf{e}^{(d)} = \widetilde{\varepsilon}^{(d)} + \widetilde{\mathbf{r}}^{(d)} + \widetilde{\mu}^{(d)}$, with $\widetilde{\varepsilon}^{(d)} = (\widetilde{\varepsilon}_t^{(d)}) = \mathbf{\Pi}_{m_T^{(d)}}\varepsilon$, $\widetilde{\mathbf{r}}^{(d)} = \widetilde{\mathbf{r}}_{m_T^{(d)}} = (\widetilde{r}_t^{(d)}) = \mathbf{\Pi}_{m_T^{(d)}}\mathbf{r}_{m_T^{(d)}}$, $\mathbf{r}_{m_T^{(d)}} = (r_{m_T^{(d)}}(1/T), \ldots, r_{m_T^{(d)}}(T/T))'$, and $\widetilde{\mu}^{(d)} = (\widetilde{\mu}_t^{(d)}) = \mathbf{\Pi}_{m_T^{(d)}}\mu$. Hence, $\widehat{\sigma}_i = T^{-1}\sum_{t=1+|i|}^{T} e_t^{(d)} e_{t-|i|}^{(d)} = \sum_{k=1}^{6} A_{k,i}$, with $A_{1,i} = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\varepsilon}_t^{(d)}\widetilde{\varepsilon}_{t-|i|}^{(d)}$, $A_{2,i} = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{r}_t^{(d)}\widetilde{r}_{t-|i|}^{(d)}$, $A_{3,i} = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\varepsilon}_t^{(d)}\widetilde{r}_{t-|i|}^{(d)} + T^{-1}\sum_{t=1+|i|}^{T}\widetilde{r}_t^{(d)}\widetilde{\varepsilon}_{t-|i|}^{(d)}$, $A_4 = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\mu}_t^{(d)}\widetilde{\mu}_{t-|i|}^{(d)}$, $A_5 = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\varepsilon}_t^{(d)}\widetilde{\mu}_{t-|i|}^{(d)} + T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\mu}_t^{(d)}\widetilde{\varepsilon}_{t-|i|}^{(d)}$, and $A_6 = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{r}_t^{(d)}\widetilde{\mu}_{t-|i|}^{(d)} + T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\mu}_t^{(d)}\widetilde{r}_{t-|i|}^{(d)}$. Cauchy–Schwarz inequality and the projection inequality of least squares regression give: $|A_{2,i}| \le d_T^2(\theta_{m_T^{(d)}}, \theta^*)$, $A_{3,i} = O_p(d_T(\theta_{m_T^{(d)}}, \theta^*))$, $|A_{4,i}| \le T^{-1}\sum_{t=1}^{T}\mu_t^2 = O_p(T)$, $|A_{5,i}| \le 2\sqrt{T^{-1}\sum_{t=1}^{T}\varepsilon_t^2}\sqrt{T^{-1}\sum_{t=1}^{T}\mu_t^2} = O_p(T^{1/2})$ and $A_{6,i} = O_p(d_T(\theta_{m_T^{(d)}}, \theta^*)T^{1/2})$. We may write $\widehat{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} w_{i,T}\widehat{\sigma}_i = \sum_{i=-\ell_T}^{\ell_T} w_{i,T}A_{1,i} + R_T$, with $R_T = \sum_{i=-\ell_T}^{\ell_T} w_{i,T}\sum_{k=2}^{6}A_{k,i}$. As $|w_{i,T}| \le 1$, we obtain:

(1) under H_0: $|R_T| \le (2\ell_T+1)\{d_T^2(\theta_{m_T^{(d)}}, \theta^*) + 2\sqrt{T^{-1}\sum_{t=1}^{T}\varepsilon_t^2}\, d_T(\theta_{m_T^{(d)}}, \theta^*)\} = O_p(\ell_T d_T(\theta_{m_T^{(d)}}, \theta^*))$,

(2) under H_1: $|R_T| \le \sum_{i=-\ell_T}^{\ell_T}|w_{i,T}|\sum_{k=2}^{6}|A_{k,i}| = O_p(\ell_T T^{-1}\sum_{t=1}^{T}\mu_t^2) = O_p(\ell_T T)$.

Now we analyze $\sum_{i=-\ell_T}^{\ell_T} w_{i,T}A_{1,i}$. As $A_{1,i} = T^{-1}\sum_{t=1+|i|}^{T}\widetilde{\varepsilon}_t^{(d)}\widetilde{\varepsilon}_{t-|i|}^{(d)}$ and $\widetilde{\varepsilon}_t^{(d)} = \varepsilon_t^{(d)} - h_t$, with $\mathbf{h} = (h_1, \ldots, h_T)' = \mathbf{\Phi}_d(\mathbf{\Phi}_d'\mathbf{\Phi}_d)^{-1}\mathbf{\Phi}_d'\varepsilon$, we obtain the decomposition $\sum_{i=-\ell_T}^{\ell_T} w_{i,T}A_{1,i} = \widetilde{\sigma}^2 + \sum_{i=-\ell_T}^{\ell_T} w_{i,T}(B_{1,i} + B_{2,i})$ with $\widetilde{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} w_{i,T}T^{-1}\sum_{t=1+|i|}^{T}\varepsilon_t\varepsilon_{t-|i|}$, $B_{1,i} = -T^{-1}\sum_{t=1+|i|}^{T}\varepsilon_t h_{t-|i|} - T^{-1}\sum_{t=1+|i|}^{T} h_t\varepsilon_{t-|i|}$, and $B_{2,i} = T^{-1}\sum_{t=1+|i|}^{T} h_t h_{t-|i|}$.

It is readily obtained $\sum_{i=-\ell_T}^{\ell_T} w_{i,T}(B_{1,i} + B_{2,i}) = O_p(\ell_T\sqrt{T^{-1}\sum_{t=1}^{T}h_t^2})$. As $T^{-1}\sum_{t=1}^{T}h_t^2 = T^{-1}\varepsilon'\mathbf{\Phi}_d(\mathbf{\Phi}_d'\mathbf{\Phi}_d)^{-1}\mathbf{\Phi}_d'\varepsilon = T^{-1}V_T$, with V_T as in Lemma 5 above, we have $V_T = \sigma^2(Q_T\sqrt{2(1+m_T^{(d)})} + (1+m_T^{(d)})) = O_p(m_T^{(d)})$, and Lemma 5 gives $T^{-1}\sum_{t=1}^{T}h_t^2 = T^{-1}V_T = O_p(m_T^{(d)}T^{-1})$.

Then we apply the decomposition $\widehat{\sigma}^2 = (\widehat{\sigma}^2 - \widetilde{\sigma}^2) + (\widetilde{\sigma}^2 - \overline{\sigma}^2) + (\overline{\sigma}^2 - \sigma^2) + \sigma^2$, with $\overline{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} w_{i,T} T^{-1} \sum_{t=1+|i|}^{T} E(\varepsilon_t \varepsilon_{t-|i|})$. As to the first difference, the above results give, under H_0, $\widehat{\sigma}^2 - \widetilde{\sigma}^2 = \sum_{i=-\ell_T}^{\ell_T} w_{i,T}(B_{1,i} + B_{2,i}) + R_T = O_p(\ell_T (m_T^{(d)})^{1/2} T^{-1/2}) + O_p(\ell_T d_T(\theta_{m_T^{(d)}}, \theta^*)) = O_p(\ell_T (m_T^{(d)})^{1/2} T^{-1/2})$ because of Assumption 2 and $\ell_T = o(T)$.

Corollary 6.3 in Pötscher and Prucha (1991) gives $\widetilde{\sigma}^2 - \overline{\sigma}^2 = O_p(\ell_T^{3/2} T^{-1/2})$ and $\overline{\sigma}^2 - \sigma^2 = O_p(\ell_T^{-\rho})$. Hence, under H_0, $\widehat{\sigma}^2 - \sigma^2 = O_p(\ell_T (m_T^{(d)})^{1/2} T^{-1/2}) + O_p(\ell_T^{3/2} T^{-1/2}) + O_p(\ell_T^{-\rho})$. Under H_0 and the assumptions of this proposition all these terms vanish in probability, even when multiplied by $m_T^{1/2}$. These arguments are valid under assumption (*i.2*), but remain true under (*i.1*), i.e., when $E|\varepsilon|^4$ and $\sum_{j=0}^{\infty} |\alpha_j| < \infty$, as a consequence of Corollary 8.3.1 in Anderson (1971) and Corollary 6.3 in Pötscher and Prucha (1991).

Finally, under H_1 the dominant term in $\widehat{\sigma}^2$ is R_T, so $\widehat{\sigma}^2 = O_p(\ell_T T)$, as well as nonnegative by construction, and the rate of divergence is derived as in Proposition 2. □

Acknowledgments The authors wish to thank the Editor, the Associate Editor and one Anonymous Referee for their valuable comments and suggestions.

References

Anderson, T. W. (1971). *The statistical analysis of time series*. New York: Wiley.

Becker, R., Enders, W., Lee, J. (2006). A stationarity test in the presence of an unknown number of smooth breaks. *Journal of Time Series Analysis, 27*, 381–409.

Bhansali, R. J., Giraitis, L., Kokoszka, P. S. (2007). Convergence of quadratic forms with nonvanishing diagonal. *Statistics & Probability Letters, 77*, 726–734.

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

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Brooks, C. (2008). *Introductory econometrics for finance*. Cambridge: Cambridge University Press.

Diebold, F. X., Kilian, L. (2000). Unit-root tests are useful for selecting forecasting models. *Journal of Business and Economics Statistics, 18*, 265–273.

Dore, M. H. I., Johnston, M. (2000). The carbon cycle and the value of forests as a carbon sink: A boreal case study. In: *Sustainable forest management and global climate change: Selected case studies from the Americas*. Cheltenham, UK: Edward Elgar Publishing.

Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association, 101*, 531–541.

Gay-García, C., Estrada, F., Sánchez, A. (2009). Global and hemispheric temperatures revised. *Climatic Change, 94*, 333–349.

Grenander, U. (1981). *Abstract inference*. New York: Wiley.

Hashimzade, N., Vogelsang, T. J. (2008). Fixed-b asymptotic approximation of the sampling behaviour of nonparametric spectral density estimators. *Journal of Time Series Analysis, 29*, 142–162.

Hong, Y., Lee, Y. J. (2003). Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *Review of Economic Studies, 72*, 499–541.

Hong, Y., White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica, 63*, 1133–1159.

Jewell, T., Lee, J., Tieslau, M., Stracizich, M. C. (2003). Stationarity of health expenditures and GDP: Evidence from panel unit root tests with heterogeneous structural breaks. *Journal of Health Economics, 22*, 313–323.

Kurozumi, E. (2002). Testing for stationarity with a break. *Journal of Econometrics, 108*, 63–99.

Kurozumi, E., Tanaka, S. (2010). Reducing the size distortion of the KPSS test. *Journal of Time Series Analysis, 31*, 415–426.

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

Landajo, M., Presno, M. J. (2010). Stationarity testing under nonlinear models. Some asymptotic results. *Journal of Time Series Analysis, 31*, 392–405.

MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Annals of Statistics, 6*(2), 422–433.

Mills, T. C., Markellos, R. N. (2008). *The econometric modelling of financial time series*. Cambridge: Cambridge University Press.

Nabeya, S., Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *The Annals of Statistics, 16*, 218–235.

Pötscher, B. M., Prucha, I. R. (1991). Basic structure of the asymptotic theory in dynamic nonlinear econometric models. Part II: Asymptotic normality. *Econometric Reviews, 10*, 253–325.

Sul, D., Phillips, P. C. B., Choi, C. (2005). Prewhitening bias in HAC estimation. *Oxford Bulletin of Economics and Statistics, 67*, 517–546.

Sun, Y. X., Phillips, P. C. B., Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica, 76*, 175–194.

Tanaka, K. (1990). The Fredholm approach to asymptotic inference on nonstationary and noninvertible time series models. *Econometric Theory, 6*, 411–432.

Tanaka, K. (1996). *Time series analysis: Nonstationary and noninvertible distribution theory*. New York: Wiley.

Van Gelder, P. H. A. J. M., Wang, W., Vrijling, J. K. (2007). Statistical estimation methods for extreme hydrological events. In: *Extreme hydrological events: New concepts for security*. Netherlands: Springer.

Wang, W. (2006). *Stochasticity, nonlinearity and forecasting of streamflow processes*. Amsterdam: IOS Press.