Working Paper Series
Department of Economics
University of Verona

The Fragility of the KPSS Stationarity Test

Nunzio Cappuccio, Diego Lubian

WP Number: 67   December 2009

The Fragility of the KPSS Stationarity Test[*]

Nunzio Cappuccio
Department of Economics
University of Padova
via del Santo 33
35123 Padova - Italy
email: nunzio.cappuccio@unipd.it

Diego Lubian
Department of Economics
University of Verona
via dell'Università
37129 Verona - Italy
email: diego.lubian@univr.it

Abstract

Stationarity tests exhibit extreme size distortions if the observable process is stationary yet highly persistent. In this paper we provide a theoretical explanation for the size distortion of the $KPSS$ test for DGPs with a broad range of first order autocorrelation coefficient. Considering a near-integrated, nearly stationary process we show that the asymptotic distribution of the test contains an additional term, which can potentially explain the amount of size distortion documented in previous simulation studies.

Keywords: $KPSS$ stationarity test, size distortion, nearly white noise nearly integrated model.

1   Introduction

Disentangling a stationary process from a unit root one has attracted the attention of many researchers. The testing strategy may have either non-stationarity or stationarity as the null hypothesis and many of the test statistics proposed so far under both approaches are now available in econometric software and routinely applied by empirical researchers. In this paper we focus our attention on the, by now very popular, test of the null hypothesis of stationarity proposed Kwiatkowski et al. (1992), hereafter KPSS. This test statistic builds on the work by Nabeya and Tanaka (1988) who, in a framework with i.i.d. normal errors, obtained the local best invariant (LBI) test to verify the coefficient constancy in a linear regression model (but see also Nyblom, 1986; Nyblom and Mäkeläinen, 1983). To take into account the possible strong autocorrelation of most macroeconomic time series, KPSS proposed an extension of the LBI test, involving a different standardization of the numerator of the test, which can then be used as a test of the null hypothesis of stationarity.

Casting the discussion in formal terms, and assuming for the sake of simplicity that the deterministic component of the series is just the constant term[1], the data generating process (DGP) of the observable variable y_t in KPSS is given by (see also (Stock, 1994)):

$$y_t = \beta + r_t + \epsilon_t \quad t = 1, \dots, T \tag{1a}$$

$$r_t = r_{t-1} + \eta_t \tag{1b}$$

$$\eta_t \sim \text{i.i.d.}(0, \sigma_\eta^2), \quad \epsilon_t \sim I(0), \quad \eta_t \perp\!\!\!\perp \epsilon_s \quad \forall \, t \text{ and } s \tag{1c}$$

The null hypothesis of stationarity constrains the variance of the error η_t to be zero, namely $H_0 : \sigma_\eta^2 = 0$, so that, under the null, the random walk process r_t collapses to a constant term and y_t is level stationary. This DGP can also be written as the ARMA model

$$y_t = \alpha y_{t-1} + w_t, \quad w_t = v_t + \theta v_{t-1}, \quad \alpha = 1 \tag{2}$$

with the null hypothesis given $\theta = -1$ where, now, α is nuisance parameter.

The KPSS test statistic is based on the behaviour of the rescaled sum of the squared partial sums of the residuals from the regression of y_t on the deterministic

---

[1]In general, when a generic deterministic component is present in the DGP, the limit theory we present in the next section can be obtained by simply substituting the Wiener process $W(r)$ with the detrended Wiener process $\underline{W}(r) = W(r) - (\int_0^1 dW(s)X(s)')(\int_0^1 X(s)X(s)'ds)^{-1} \int_0^r X(r)$, where we assume that there exists some standardizing matrix Υ such that $\Upsilon^{-1} x_{[Tr]} \to X(r)$ and $X(r) = (1, r, \dots, r^p)'$. For instance, under the trend stationary null hypothesis we have $x_t = (1, t)'$ and the limit theory should be rephrased in terms of the second-level Brownian bridge given by $V_2(r) = W(r) + (2r - 3r^2)W(1) + (6r^2 - 6r)\int_0^1 W$. Here, we focus on the case known as "local-level" unobserved component model (Harvey, 1995).

component. In model (1a)-(1c), defining the residual as $e_t = y_t - \bar{y}$ and the partial sum $S_t = \sum_{j=1}^{t} e_j$, the $KPSS$ test statistic is given by

$$KPSS = T^{-2} \sum_{t=1}^{T} S_t^2 \Big/ \omega_\epsilon^2 \qquad (3)$$

The crucial point in (3) is the choice of the scale factor ω_ϵ^2, which depends on the behaviour of ϵ_t. If these errors are i.i.d., then it is sufficient to rescale by the usual estimator of the variance, i.e. $\hat{\sigma}_\epsilon^2 = T^{-1} \sum_{t=1}^{T} e_t^2$. In addition, if ϵ_t is Gaussian, as in Nabeya and Tanaka (1988), then the test (3) is LBI. If ϵ_t is dependent, satisfying the (strong mixing) regularity conditions of Phillips and Perron (1988, p. 336), or the linear process conditions of Phillips and Solo (1992, Theorems 3.3, 3.14), KPSS propose to replace ω_ϵ^2 by a consistent estimate, say $s^2(m_T)$, of the long run variance of the residual e_t, where m_T is a bandwidth parameter with $m_T \to \infty$ as $T \to \infty$ so that $m_T/T \to 0$.

As $T \to \infty$, under the null hypothesis the test converges to

$$KPSS \Rightarrow \int_0^1 V(r)^2 \mathrm{d}r \qquad (4)$$

where $V(r)$ is a standard Brownian bridge $V(r) = W(r) - rW(1)$ and $W(r)$ is a Wiener process.

As emphasized by Müller (2005), the rescaling by $s^2(m_T)$ is aimed at achieving the correct size of the test in the presence of autocorrelated series while, at the same time, it has to provide power when the strong autocorrelation arises from an integrated process. Thus, this balance between size and power relies crucially on a "good" estimator of the long run variance of the process and, as testified by several simulation studies, it might be difficult to achieve in practice. Among the many simulation studies, Lee (1996) is a typical example of this problem: his results indicate that estimators of the long run variance producing a test with the correct size also lead to dramatic loss in power.

Concentrating on size distortion, extreme size distortion of stationarity tests has been documented in several previous works. Simulation results in Kwiatkowski et al. (1992) and Caner and Kilian (2001) indicate that over-rejection is quite important whenever the largest autoregressive root is close to unity. For instance, considering an AR(1) process the effective size is 11.4% when the autoregressive parameter is equal to 0.7 but it peaks at 55.4% (70.8%) when it is equal to 0.95 (0.98). Thus, for economically plausible parameter values, the $KPSS$ test exhibits effective size much greater than the nominal size leading to a serious over-rejection in empirical works[2]. Müller (2005) advocates the use of local-to-unity asymptotics to obtain more

---

[2]Caner and Kilian (2001) and Lanne and Saikkonen (2003) present evidence on the relevance of this problem on the stationarity test proposed by Leybourne and McCabe (1994), say LMC.

accurate approximation of the small sample distribution of the stationarity test, as it has been done for unit root tests. He shows that the behavior of stationarity test depends heavily on long run variance estimator and on the rate of growth of the bandwidth parameter but, more importantly, he points out that the remedy to reduce the large size distortion in the presence of strongly autocorrelated series might open the door to test inconsistency.

Previous studies investigated the issue of size distortion focussing on the implications of high persistence, in relation to the role played by the large autoregressive root in the ARMA representation, thus considering a near-integrated processes as in Müller (2005). In this paper, the objective is to improve our understanding of the behaviour of the $KPSS$ test under the null hypothesis of stationarity and to point out its fragility in the presence of small deviations from the maintained hypothesis of $\alpha = 1$ when, at the same time, we induce small deviations of the moving average parameter θ from -1, the parameter value under the null hypothesis. Assuming that ϵ_t is an i.i.d. process (this assumption will be removed in the Appendix), we will consider a specific generating mechanism for these deviations so that the observable process will be a stationary ARMA process with no common factor in any finite sample whereas, in the limit as $T \to \infty$, the process y_t will approach an i.i.d. process since the autoregressive root approaches unity and the moving average root approaches minus unity, thus canceling out each other. Thus, our study concerns the behavior of the test under a sequence of DGPs exhibiting different degrees of autocorrelation in any finite sample but converging, as the sample size grows large, to the setup where the $KPSS$ is known to be the locally best invariant test under normality.

Formally, we follow Nabeya and Perron (1994) and Perron and Ng (1996) by considering a so-called nearly-integrated nearly-white noise process which is a sequence of stationary ARMA(1, 1) processes such that the autoregressive root approaches unity and the MA root tends to -1 as $T \to \infty$. In other words, the ARMA(1,1) has an asymptotic common factor. Nabeya and Perron (1994) and Perron and Ng (1996) have originally used this process but in a different context to investigate the behavior of unit root tests under a sequence of alternative hypotheses converging to a specific alternative hypothesis, namely a white noise process, but this specification of the DGP is especially well suited for our purposes[3]. By combining a possibly large autoregressive root and a moving average root close to -1, the nearly-integrated nearly-white noise DGP is capable of generating a wide range of first-order autocorrelation coefficient but, because of the presence of an asymptotic common factor, it

---

[3]The size distortion of unit root tests has been studied by Pantula (1991) with a similar specification. His process has a unit root in finite samples and converges to a white noise as $T \to \infty$.

collapses to a white noise process as $T \to \infty$ generating the "ideal" settings for the test statistic.

The nearly-white-noise, nearly-integrated sequence of DGPs should be helpful in providing a better approximation to the exact distribution of the test statistics when the time series is stationary but autocorrelated with an MA structure with large negative correlation, as it is the case for several macroeconomic time series. We show that the limiting distribution of the $KPSS$ test depends upon the degree of nearly stationarity and on the vicinity of the autoregressive parameter to the nonstationarity region.

The plan of the paper is as follows. In the following section, we provide a more accurate description of the nearly-white noise, nearly integrated DGPs and we present the main result of the paper on the limiting behavior of the $KPSS$ statistic under this sequence of DGPs. Next, we conduct a small sample analysis via a MonteCarlo experiment to assess the extent of size distortion under this DGP and the importance of bandwidth and kernel choice.

2   The $KPSS$ test under a nearly-integrated nearly-white noise DGP

Following Nabeya and Perron (1994), we consider a setting in which the error term ϵ_t in (1a)-(1c) is i.i.d. with zero mean and finite variance σ_ϵ^2, the general case is considered in the Appendix[4]. Then, defining $\phi = \sigma_\eta^2 / \sigma_\epsilon^2$, we may write $\eta_t = \sigma_\epsilon \sqrt{\phi} h_t$ where h_t is i.i.d. with zero mean and unit variance by construction. Next, we let ϕ be dependent on the sample size in such a way that $\phi \to 0$ as $T \uparrow \infty$, that is $\sqrt{\phi} = \delta/T$. Finally, autocorrelation in the observable process y_t is introduced by modeling r_t as an AR(1) process with a local-to-unity autoregressive root given by $\exp(c/T)$, with $c < 0$, and error term h_t scaled by the $o(1)$ factor $\delta\sigma_\epsilon/T$. Our sequence of DGPs is given then by

$$y_t = \beta + r_t + \epsilon_t \quad t = 1, \ldots, T \tag{5a}$$

$$r_t = \exp(c/T)r_{t-1} + \frac{\delta\sigma_\epsilon}{T}h_t, \tag{5b}$$

$$h_t \sim \text{i.i.d.}(0,1), \quad \epsilon_t \sim \text{i.i.d.}(0,\sigma_\epsilon^2), \quad h_t \perp\!\!\!\perp \epsilon_s \ \forall \, t \text{ and } s \tag{5c}$$

---

[4]In the Appendix we show that the generalization to the linear process $\epsilon_t = a(L)u_t$, where u_t is i.i.d.$(0; \sigma_u^2)$, and $a(L) = \sum_{j=0}^\infty a_j L^j$ is a polynomial in the lag operator L with $a(1) = \sum_{j=0}^\infty a_j \neq 0$ and $\sum_{j=1}^\infty j^2 a_j^2 < \infty$ can be handled easily. Yet, it would overshadow our main point: to consider a highly autocorrelated local to a white noise process series y_t. Allowing for serially correlated errors, one would then obtain a nearly-$I(0)$, nearly-integrated process. Briefly, in the Appendix we point out that under the more general specification, our results will hold by simply substituting δ by $\delta a(1)$ both in the sequence of DGPs and in the limiting distributions.

Simple inspection of (5a)-(5c) reveals that, for any fixed T, the observable process y_t is near-integrated process

$$
\begin{aligned}
y_t &= \beta[1 - \exp(c/T)] + \exp(c/T)y_{t-1} + w_t \\
w_t &= \epsilon_t - \exp(c/T)\epsilon_{t-1} + \frac{\delta\sigma_\epsilon}{T}h_t
\end{aligned}
$$

where the composite error term w_t has a moving average root approaching -1 as $T \to \infty$. It follows that, asymptotically, the autoregressive near unit root cancels out the moving average root and, at the same time, the influence of h_t vanishes so that y_t collapses to a white noise process. This specification allows us to investigate the behavior of the $KPSS$ stationarity test under a sequence of null hypothesis relevant in empirical works where persistent stationary processes are often encountered.

In finite samples, the process w_t has an MA(1) representation with MA parameter given by the negative root of

$$
E(w_t w_{t-1}) + E(w_t^2)\xi + E(w_t w_{t+1})\xi^2 = 0 \tag{6}
$$

with $E(w_t^2) = (1 + \exp(2c/T))\sigma_\epsilon^2 + (\delta/T)^2\sigma_\epsilon^2$ and $E(w_t w_{t-1}) = E(w_t w_{t+1}) = -\sigma_\epsilon^2\exp(c/T)$.

Table 1 provides evidence on the degree of persistence of y_t, as measured by the first-order autocorrelation coefficient. The nearly-integrated, nearly white noise process is capable of generating a wide range of degree of persistence, from almost 0 in correspondence of the common AR and MA factor to 0.9 and beyond. As the first-order autocorrelation decreases, the root of the MA component tends to approach -1 and the process gets close to a white noise process.

The presence of autocorrelation in finite samples would call for a consistent estimator of the long run variance in the computation of the $KPSS$ test in (3), even though the observable process is local to a white noise, which would suggest the use of a standard estimator of the (short run) variance of the residuals e_t. As a matter of fact, this is quite a delicate issue for the $KPSS$ test. As it is well-known, the usage of a data-dependent automatic bandwidth selection according Andrews (1991) or AR-prewhitening combined with automatic bandwidth as in Andrews and Mohanan (1992) is not allowed because both of them lead to inconsistent tests since the estimated bandwidth would grow at the rate $O_p(T)$ which implies that the $KPSS$ test would be $O_p(1)$ under the alternative hypothesis, as discussed in Choi (1994). In fact, test inconsistency is the price paid by Müller (2005), who makes use of these data-dependent bandwidth choices, to get stationarity tests with size closer to the nominal level when the observable process is nearly integrated. However, we believe that test consistency is an important feature of the test statistic to be maintained. Therefore, we consider both a fixed bandwidth as in the original

Table 1: First-order autocorrelation of the nearly-white noise, nearly integrated process (5a)-(5c) for selected values of c and δ ($\sigma_\epsilon^2 = 1$)

			δ						
c	T	$\exp(c/T)$	1	5	10	20	30	40	50
-1	50	0.980	0.010	0.199	0.495	0.787	0.884	0.924	0.943
	100	0.990	0.005	0.111	0.332	0.662	0.812	0.881	0.917
	500	0.998	0.001	0.024	0.091	0.286	0.473	0.615	0.713
-5	50	0.905	0.002	0.047	0.164	0.424	0.602	0.705	0.766
	100	0.951	0.001	0.024	0.090	0.282	0.462	0.596	0.689
	500	0.990	0.000	0.005	0.020	0.074	0.152	0.242	0.332
-10	50	0.819	0.001	0.024	0.089	0.268	0.427	0.540	0.616
	100	0.905	0.000	0.012	0.047	0.164	0.300	0.424	0.525
	500	0.980	0.000	0.002	0.010	0.038	0.082	0.138	0.199
-15	50	0.741	0.001	0.016	0.060	0.194	0.329	0.435	0.510
	100	0.861	0.000	0.008	0.032	0.115	0.222	0.329	0.423
	500	0.970	0.000	0.002	0.007	0.026	0.056	0.096	0.142

Kwiatkowski et al. (1992) or the data-dependent bandwidth choice suggested by Newey and West (1994), which has been shown to deliver test consistency by Hobijn et al. (2004).

The limiting behavior of the test statistic relies on a Functional Central Limit Theorem by Phillips and Solo (1992, Theorem 3.3) for the partial sums $S_{[Tr]} = \sum_{t=1}^{[Tr]} e_t$ built from the residuals $e_t = y_t - \bar{y}$ and the Continuous Mapping Theorem. As $T \to \infty$, we have that

$$\frac{1}{\sqrt{T}} S_{[Tr]} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} e_t \Rightarrow \sigma_\epsilon V(r) + \sigma_\epsilon \delta \int_0^r \underline{K}_c(s) \mathrm{d}s \equiv \sigma_\epsilon V_{c,\delta}(r) \qquad (7)$$

where $V(r) = W_1(r) - rW_1(1)$ is a standard Brownian bridge, $\underline{K}_c(r) = K_c(r) - \int_0^1 K_c(s)\mathrm{d}s$ and $K_c(r) = \int_0^r e^{(r-s)c}\mathrm{d}W_2(s)$ is a diffusion (Ornstein-Uhlenbeck) process. Further, $W_1(r)$ and $W_2(r)$ are two independent Wiener processes such that $T^{-1/2} \sum_{i=1}^{[Tr]} \epsilon_t \Rightarrow \sigma_\epsilon W_1(r)$ and $T^{-1/2} \sum_{i=1}^{[Tr]} h_t \Rightarrow W_2(r)$. Since we may also write

$$K_c(r) = W_2(r) + c \int_0^r e^{(r-s)c} W_2(s)\mathrm{d}s$$

we have the following expression, whose components will enter the asymptotic distribution of the $KPSS$ test under a nearly-white-noise, nearly-integrated process,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} e_t \Rightarrow \sigma_\epsilon V(r) + \sigma_\epsilon \delta \int_0^r \underline{W}_2(s)\mathrm{d}s +$$

$$+ \sigma_\epsilon \delta c \int_0^r \left[\int_0^s e^{(s-u)c} W_2(u)\mathrm{d}u - \int_0^1 \left(\int_0^v e^{(v-u)c} W_2(u)\mathrm{d}u \right) \mathrm{d}v \right] \mathrm{d}s$$

where $\underline{W}_2(r) = W_2(r) - \int_0^1 W_2(s)\mathrm{d}s$. The first term is just the usual Brownian bridge appearing in the limiting distribution of the $KPSS$ test (see Kwiatkowski et al. (1992)) under the null hypothesis. The second term appears in the limiting distribution of the $KPSS$ test statistics under a sequence of local alternatives, that is DGPs (5a)-(5b) with $c = 0$, as in Stock and Watson (1998) and Cappuccio and Lubian (2006) and it is relevant when the interest is in the local asymptotic power of the test statistic. Notice that this bias term is present despite the fact that the nearly-white-noise, nearly-integrated DGP is stationary both in finite samples and in the limit as $T \to \infty$. The third component reflects both the degree of nearly-integration of r_t via the parameter c and the influence of this component, via the scale factor δ, in shaping the time dependence structure of the observable process y_t.

The consistent estimator $s^2(m_T)$ of the long run variance of e_t may be obtained either following the original suggestion of Kwiatkowski et al. (1992) by choosing $m_T = O(T^{1/4})$ or by applying the procedure proposed by Newey and West (1994) outlined in Table 3. Under both choices, the following proposition characterizes the asymptotic behavior of the $KPSS$ test under a sequence of a nearly-white noise, nearly-integrated processes.

PROPOSITION 2.1 *Under DGP (5a)-(5c), the asymptotic distribution of the KPSS test is given by*

$$ KPSS \Rightarrow \int_0^1 V_{c,\delta}^2(r)\mathrm{d}r = \int_0^1 V(r)^2\mathrm{d}r + \delta \int_0^1 \left(V_{\delta,c}(r) \int_0^r \underline{K}_c(s)\mathrm{d}s \right) \mathrm{d}r \qquad (8) $$

This result indicates that the asymptotic behavior the $KPSS$ test in the presence of a nearly-white noise, nearly integrated process is affected both by the local-to-unity parameter c and by the scale factor δ. Evidently, when $\delta = 0$, the order of magnitude of c becomes irrelevant since r_t is just a constant. On the other hand, when the process is persistent the $KPSS$ suffers from size distortion as reported in simulation studies (see Saikkonen and Luukkonen (1993); Caner and Kilian (2001)).

Asymptotic rejection rates based on the right-hand side of (8) are reported in Table 2 as a function of c and δ. When $\delta = 1$ there is no size distortion whereas, in general, for a given c size distortion increases with δ and for given δ it is decreasing in c. Even for small values of the population first-order autocorrelation coefficient of the nearly-white noise nearly integrated DGP implied by $c = -10$ or $c = -15$ (see Table 1), asymptotic size distortion may by substantial for $\delta \geq 20$. This suggests that the second component of the right-hand side of (8) might be helpful to explain the small sample size distortions observed in simulation and empirical research. Thus, the

Table 2: Asymptotic rejection rates based on (8) for selected values of c and δ

	δ						
	1	5	10	20	30	40	50
$c = 0$	6.33	29.79	61.48	86.85	94.67	98.21	99.09
$c = -1$	5.63	21.23	50.25	81.31	92.51	97.04	98.82
$c = -5$	5.01	10.05	26.20	60.79	80.71	90.77	95.39
$c = -10$	4.93	6.88	14.07	39.43	62.69	78.53	88.75
$c = -15$	5.09	6.32	9.98	25.71	45.71	63.32	76.84

asymptotic result in Proposition 2.1 provides a quantitatively useful approximation to the size distortion issue[5].

As for the consistency of the test under the alternative hypothesis, it has been established by Kwiatkowski et al. (1992) that under a fixed bandwidth choice ($m_T/T \to 0$ as $T \to \infty$) the test is $O_p(T/m_T)$. Hobijn et al. (2004) have shown that under the automatic bandwidth selection procedure suggested in Newey and West (1994) the test retains its consistency being $O_p(T^{14/25})$ when using the Bartlett kernel and $O_p(T^{92/125})$ with the Quadratic Spectral kernel. Again, details on kernels and bandwidth selection procedure are provided in Table 3.

3   Finite sample properties

In a simulation study we investigate both the role played in finite samples by the nearly-white noise, nearly integrated sequence of DGPs and the influence of different choices of the kernel function and bandwidth parameter on the effective size of the $KPSS$ test. For simplicity, in the simulation we consider the local-level DGP (5a)-(5c) where the deterministic component is given by the constant term. We consider the sequence $\delta = \{1, 5, 10, 20, 30, 40, 50\}$, values of c ranging from 0 to -16 and 10000 replications.

The long run variance of the residuals e_t is estimated using either the Bartlett or the Quadratic Spectral kernels. As for the choice of the bandwidth parameter, for comparison with previous studies, we adopt the fixed bandwidth as in the original paper by Kwiatkowski et al. (1992) and the automatic bandwidth procedure by Newey and West (1994) whose use has been advocated in Hobijn et al. (2004). For the fixed bandwidth, we follow the latter authors by setting $m_B(4) = [4(T/100)^{1/4}]$ and $m_{QS}(4) = [\frac{8}{3}(T/100)^{2/9}]$.

Rejection rates of the $KPSS$ test reported in Tables 4 and 5 are computed using the 5% critical value of 0.463 as published in Kwiatkowski et al. (1992). In each

---

[5]We thank an anonymous referre for pointing this out.

Table 3: Kernels and Bandwidths

A. Kernels	
Bartlett	$k_B(j) = \begin{cases} (1 - j/(m+1)) & j \leq m \\ 0 & \text{otherwise} \end{cases}$
Quadratic Spectral	$k_{QS}(j) = \dfrac{25}{12\pi^2(j/m)^2} \left[\dfrac{\sin(6\pi(j/m)/5)}{6\pi(j/m)/5} - \cos(6\pi(j/m)/5) \right]$

B. Fixed Bandwidth: $m_B(x), m_{QS}(x)$	
Bartlett	$m_B(x) = [x(T/100)^{1/4}]$
Quadratic Spectral	$m_{QS}(x) = \left[\frac{2}{3}x(T/100)^{2/9}\right]$

C. Automatic Bandwidth Choice (Newey and West (1994))

	Bartlett	Quadratic Spectral
Initial bandwidth parameter	$m_B(x)$	$m_{QS}(x)$
Compute	$\hat{s}^{(0)} = \hat{\gamma}_0 + 2\sum_{i=1}^{m} \hat{\gamma}_i$	
	$\hat{s}^{(1)} = 2\sum_{i=1}^{m} i\hat{\gamma}_i$	
	$\hat{s}^{(2)} = 2\sum_{i=1}^{m} i^2\hat{\gamma}_i$	
Compute	$\hat{\gamma} = 1.1447\left(\left[\dfrac{\hat{s}^{(1)}}{\hat{s}^{(0)}}\right]^2\right)^{1/3}$	$\hat{\gamma} = 1.3221\left(\left[\dfrac{\hat{s}^{(2)}}{\hat{s}^{(0)}}\right]^2\right)^{1/5}$
Select Bandwidth parameter	$n_B(x) = \min\{T, [\hat{\gamma}T^{1/3}]\}$	$n_{QS}(x) = \min\{T, [\hat{\gamma}T^{1/5}]\}$

panel of Table 4 we keep the parameter c constant, with values ranging from $c = 0$ to $c = -15$, and we look at the effect of increasing δ, for different growing sample sizes. For $c = 0$ (panel A), the process (5a)-(5c) is nonstationary irrespective of the value taken by δ and, therefore, rejection rates provide the empirical power function of the test statistic. As expected, power grows with δ and with the sample size. As δ gets large, for fixed T, the informative content in the nonstationary component r_t increases and thus the observable process will behave more and more as a unit root process. On the other hand, as T grows for fixed δ, the amount of information available to the econometrician increases thereby affecting positively the power properties of the test. The Quadratic Spectral kernel delivers higher power than the Bartlett one irrespective of the bandwidth adopted and it seems less sensitive to the bandwidth choice than the Bartlett kernel given that power with fixed bandwidth is close to the power with automatic bandwidth whereas the Bartlett kernel generates greater power when used together with a fixed bandwidth.

Results in Panels B through E refer to cases where $c \neq 0$ leading therefore to consider stationary processes. Thus, the reported rejection rates are in fact the

effective sizes of the test. In each Panel, reading the Table by row we observe that the effective size worsens as δ increases. This result is expected from Table 1 where, for a given c, higher values of δ generate higher population first-order autocorrelation coefficients making the observable process more persistent.

When reading Table 4 by column we notice that increasing the sample size does not make the effective size closer to the nominal one but, on the contrary, it induces even higher size distortion. This is the effect of the second term in the asymptotic distribution of the $KPSS$ test given by (8). These results are consistent with the simulation evidence provided in Caner and Kilian (2001, Table 1), where it is reported that for a sample size as large as $T = 500$ the effective size of the test may be as high as 60% or 50% according to the bandwidth choice.

Furthermore, automatic bandwidth yields lower size distortion under either choice of the kernel function and, noticeably, the Bartlett kernel outperforms the QS kernel.

Table 5 provides a complementary look at the simulation results. Panels A through F contain the rejection rates of the test for (fixed) values of $\exp(c/T)$ approaching unity. In each panel, three increasing sample sizes and increasing values of δ are considered. According to the values taken by the first-order autocorrelation coefficients in Table 1, for a given $\exp(c/T)$ when we move from the north-east to the south-west of each panel we should observe effective sizes similar to the nominal size. This should happen because in the north-east of each panel, the autocorrelation structure of the process r_t displays higher persistence than in the south-west region. As the autoregressive root of the process approaches unity and we move from panel A to panel F, the size of the test worsens for any sample size. However, when the root is very close to unity, say $\exp(c/T) = 0.99$, only for $\delta = 1$ the effective size is close to the nominal one and it deteriorates quickly as δ increases, reaching extremely high values for $\delta = 50$ with little beneficial effects provided by an increased sample size. Our results also indicate that the automatic bandwidth leads to smaller size distortion than the fixed bandwidth and that the Bartlett kernel performs better than the Quadratic Spectral one, with large improvements in the effective size when the autoregressive root is up to 0.8.

Figures 1 report the effective size of the $KPSS$ for the sample sizes $T = 50$ as a function of c as δ increases for different choices of the kernel function and bandwidth parameter. In general, the effective size of the test is greater then nominal size and slightly lower when the automatic bandwidth is used. Considering the highly persistent yet stationary processes obtained when $\delta = 10$ and c is between 0 and -2 for a sample size of $T = 50$, the effective size is between 45% and 25%. Even though it is hardly impossible to discriminate between kernels and bandwidth and to find strong support for a particular kernel or bandwidth choice, our results suggest that

Bartlett kernel used together with automatic bandwidth choice might be able to reduce size distortion an to provide a more accurate approximation to the limiting distribution under i.i.d. errors. Unfortunately, δ must be relatively small and the autoregressive root enough far away form unity for the effective size to be close to the nominal size.

4   Conclusion

In this paper we have set forth an analytic explanation for the size distortion of the $KPSS$ stationarity test. We studied the asymptotic behavior of the $KPSS$ test when the DGP is a nearly white noise, nearly-integrated process. Under this sequence of processes the DGP is always stationary converging the the i.i.d. settings under which the $KPSS$ test statistic is known to be LBI. Our theoretical results rationalize the size distortion found in simulation experiments by, e.g., Kwiatkowski et al. (1992); Leybourne and McCabe (1994); Caner and Kilian (2001); Lanne and Saikkonen (2003). Our simulation results indicate that even though the DGP is a local to white noise, the bias in the effective size may be important not only when the autoregressive root is close to unity. How this size distortion issue can be tackled and possible solved, is a topic for future research.

A Proof of Proposition 2.1

We provide a proof of Proposition 2.1 under general conditions on the sequence of DGPs. In particular, we generalize the nearly-white noise, nearly integrated process by considering a nearly-$I(0)$, nearly integrated process. DGP (5a)-(5c) is then modified as

$$y_t = \beta + r_t + \epsilon_t \quad t = 1, \dots, T \tag{9a}$$

$$r_t = \exp(c/T)r_{t-1} + \frac{\delta a(1)}{T}h_t, \tag{9b}$$

$$h_t \sim \text{i.i.d.}(0,1), \quad \epsilon_t = a(L)u_t, \quad h_t \perp\!\!\!\perp u_s \ \forall\, t \text{ and } s \tag{9c}$$

where u_t is i.i.d. with zero mean and finite variance σ_u^2, and $a(L) = \sum_{j=0}^{\infty} a_j L^j$ is a polynomial in the lag operator L with $a(1) = \sum_{j=0}^{\infty} a_j \neq 0$ and $\sum_{j=1}^{\infty} j^2 a_j^2 < \infty$. It will be useful to define the long run variance of ϵ_t as $\omega_\epsilon^2 = a(1)^2 \sigma_u^2$ and to recall its decomposition $\omega_\epsilon^2 = \sigma_\epsilon^2 + 2\kappa_\epsilon$ where $\sigma_\epsilon^2 = E(\epsilon_0^2)$ and $\kappa_\epsilon = \sum_{k=1}^{\infty} E(\epsilon_0 \epsilon_k)$.

The convergence in (7) is simply modified as

$$\frac{1}{\sqrt{T}}S_{[Tr]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} e_t \Rightarrow \sigma_u a(1)^{1/2}\left(V(r) + \delta \int_0^r \underline{K}_c(s)\mathrm{d}s\right) \tag{10}$$

Routine application of the continuous mapping theorem yields

$$T^{-1}S_{[Tr]}^2 \Rightarrow \sigma_u^2 a(1)\left(V(r) + \delta \int_0^r \underline{K}_c(r)\mathrm{d}s\right)^2$$

and

$$T^{-2}\sum_{t=1}^{T} S_t^2 \Rightarrow \sigma_u^2 a(1)\left[\int_0^1 V(r)^2\mathrm{d}r + \delta \int_0^1 \left(V_{\delta,c}(r)\int_0^r \underline{K}_c(s)\mathrm{d}s\right)\mathrm{d}r\right]$$

where $V_{\delta,c}(r) = 2V(r) + \delta\int_0^r \underline{K}_c(s)\mathrm{d}s$, $V(r) = W(r) - rW(1)$ is a standard Brownian bridge and $\underline{K}_c(s) = K_c(s) - \int_0^1 K_c(v)\mathrm{d}v$ is a demeaned Ornstein-Uhlenbeck process. Next, letting $k(s/m_T)$ be the kernel function as defined in Table 3 we need to show that

$$s^2(m_T) = \frac{1}{T}\sum_{t=1}^{T} e_t^2 + \frac{2}{T}\sum_{s=1}^{m_T} k(s/m_T)\sum_{t=s+1}^{T} e_t e_{t-s} \tag{11}$$

is a consistent estimator of $\sigma_u^2 a(1)^2$.

Letting $H_t = \sum_{j=1}^{t} \exp\{(t-j)c/T\}h_j$ and $\bar{H} = T^{-1}\sum_{t=1}^{T} H_t$, under DGP (9a)-(9c) we have

$$e_t = \frac{\delta a(1)\sigma_u}{T}(H_t - \bar{H}) + (\epsilon_t - \bar{\epsilon})$$

where $\bar{\epsilon}_t = T^{-1}\sum_{t=1}^{T}\epsilon_t$. After substitution in the first term of (11), we obtain

$$
\frac{1}{T}\sum_{t=1}^{T}e_t^2 = \frac{1}{T}\sum_{t=1}^{T}(\epsilon_t - \bar{\epsilon})^2 + \frac{1}{T}\left[\frac{\delta^2 a(1)^2\sigma_u^2}{T^2}\sum_{t=1}^{T}(H_t - \bar{H})^2 + 2\frac{\delta a(1)\sigma_u}{T}\sum_{t=1}^{T}(H_t - \bar{H})(\epsilon_t - \bar{\epsilon})\right]
$$

$$
= \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 + o_p(1)
$$

$$
\xrightarrow{p} \sigma_\epsilon^2
$$

We turn to the analysis of the second term in (11), which, after tedious algebra and apart from the constant, can be written as

$$
\frac{1}{T}\sum_{s=1}^{m_T}k(s/m_T)\sum_{t=s+1}^{T}e_t e_{t-s} = \sum_{s=1}^{m_T}k(s/m_T)\frac{1}{T}\sum_{t=s+1}^{T}(\epsilon_t - \bar{\epsilon})(\epsilon_{t-s} - \bar{\epsilon})+
$$

$$
+\frac{\delta^2 a(1)^2\sigma_u^2}{T}\sum_{s=1}^{m_T}k(s/m_T)\left(\frac{1}{T^2}\sum_{t=s+1}^{T}(H_t - \bar{H})(H_{t-s} - \bar{H})\right)+
$$

$$
+\frac{\delta a(1)\sigma_u}{T}\sum_{s=1}^{m_T}k(s/m_T)\left(\frac{1}{T^2}\sum_{t=s+1}^{T}(H_{t-s} - \bar{H})(\epsilon_t - \bar{\epsilon})\right)+
$$

$$
+\frac{\delta a(1)\sigma_u}{T}\sum_{s=1}^{m_T}k(s/m_T)\left(\frac{1}{T^2}\sum_{t=s+1}^{T}(H_t - \bar{H})(\epsilon_{t-s} - \bar{\epsilon})\right)
$$

which, for convenience, we rewrite as

$$
\frac{1}{T}\sum_{s=1}^{m_T}k(s/m_T)\sum_{t=s+1}^{T}e_t e_{t-s} = \sum_{s=1}^{m_T}k(s/m_T)\frac{1}{T}\sum_{t=s+1}^{T}(\epsilon_t - \bar{\epsilon})(\epsilon_{t-s} - \bar{\epsilon})+
$$

$$
+\frac{m_T}{T}\delta^2 a(1)^2\sigma_u^2\left[\frac{1}{m_T T}\sum_{s=1}^{m_T}k(s/m_T)\left(\frac{1}{T}\sum_{t=s+1}^{T}(H_t - \bar{H})(H_{t-s} - \bar{H})\right)\right]+
$$

$$
+\frac{m_T}{T}\delta a(1)\sigma_u\left[\frac{1}{m_T T}\sum_{s=1}^{m_T}k(s/m_T)\left(\frac{1}{T}\sum_{t=s+1}^{T}(H_{t-s} - \bar{H})(\epsilon_t - \bar{\epsilon})\right)\right]+
$$

$$
+\frac{m_T}{T}\delta a(1)\sigma_u\left[\frac{1}{m_T T}\sum_{s=1}^{m_T}k(s/m_T)\left(\frac{1}{T}\sum_{t=s+1}^{T}(H_t - \bar{H})(\epsilon_{t-s} - \bar{\epsilon})\right)\right]
$$

The first term is the standard expression for the kernel consistent estimator of $\kappa_\epsilon = \sum_{k=1}^{\infty}E(\epsilon_0\epsilon_k)$. By Phillips (1991, formula between (A.10) and (A.11)) the second term in brackets is $O_p(1)$, and the third and fourth terms in brackets are $O_p(1)$ too by (A.13). This follows by substituting in Phillips (1991) the Ornstein-Uhlenbeck process $K_c(r)$ for the Wiener process and by making the usual assumption that the kernel $k(\cdot)$ is a bounded, even function with $\int |k(x)|\mathrm{d}x < \infty$. Since $m_t/T \to 0$ as $T \uparrow \infty$, it follows that

$$
\frac{1}{T}\sum_{s=1}^{m_T}k(s/m_T)\sum_{t=s+1}^{T}e_t e_{t-s} \xrightarrow{p} \kappa_\epsilon
$$

13

Finally, combining all convergence in probability established so far we obtain the desired result, namely $s^2(m_T) \xrightarrow{p} \omega_\epsilon^2 = \sigma_\epsilon^2 + 2\kappa_\epsilon = a(1)^2 \sigma_u^2$.
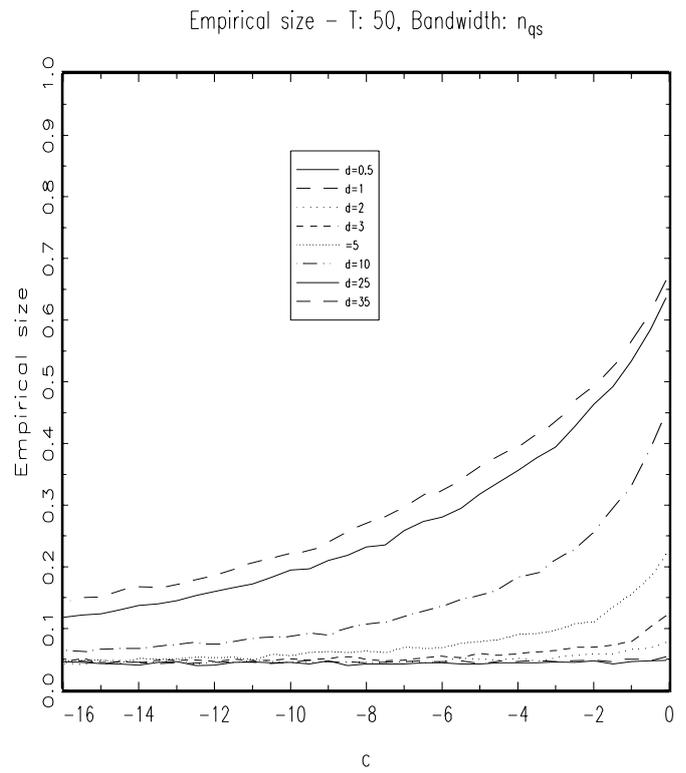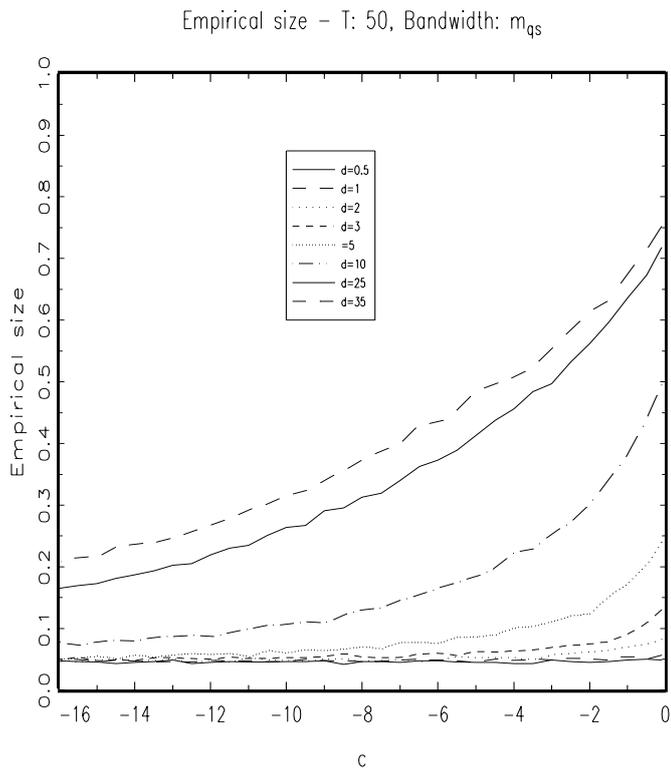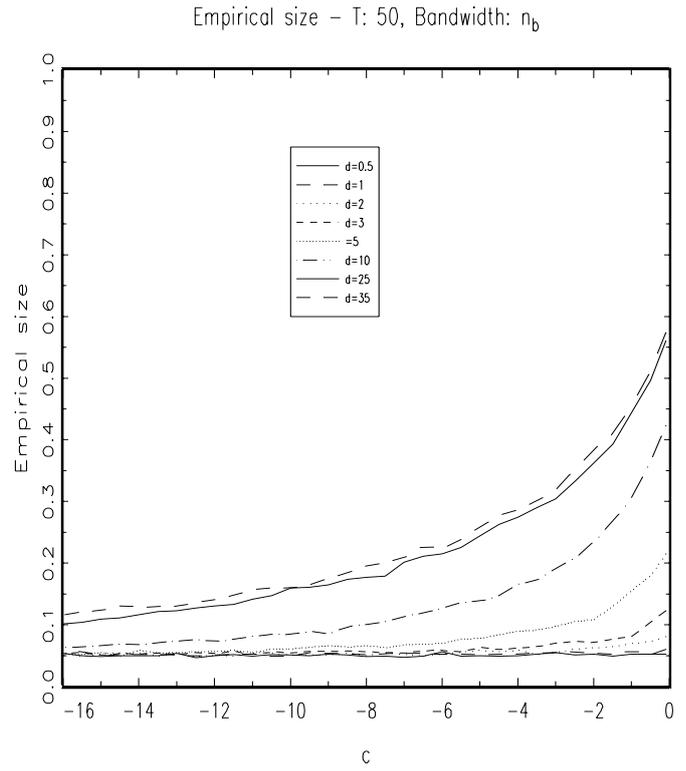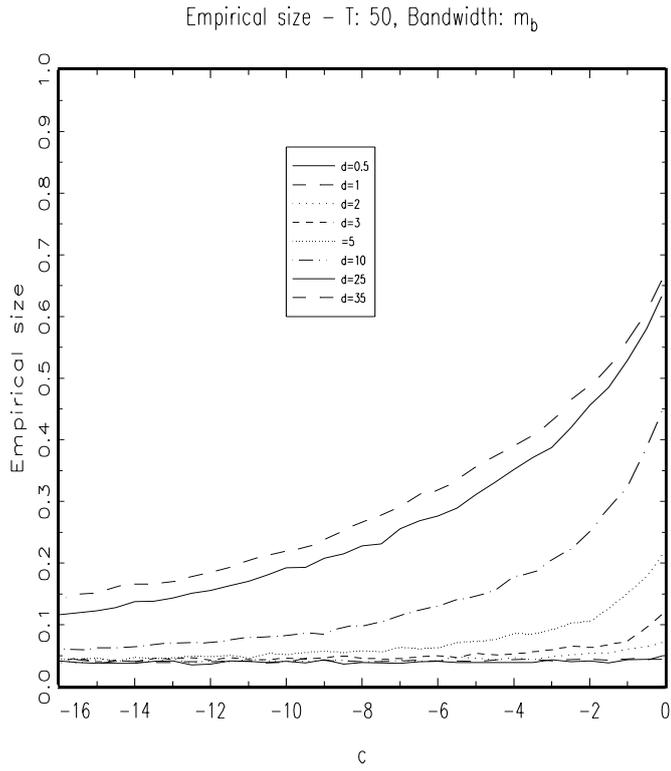
Figure 1: Empirical size of $KPSS$ test, $T = 50$

Table 4: Rejection rates (per cent) of the $KPSS$ test for the process (5a)-(5c).

		Fixed bandwidth, $m(4)$							Automatic bandwidth, $n(4)$						
A. $c=0$					δ		Bartlett Kernel					δ			
T	$e^{c/T}$	1	5	10	20	30	40	50	1	5	10	20	30	40	50
50	1	4.8	24.4	43.2	60.6	64.4	67.6	66.0	5.5	24.4	41.8	54.7	55.7	57.5	56.9
100	1	4.7	25.5	52.1	68.8	76.6	79.2	79.2	5.1	25.6	49.8	60.7	66.7	68.0	66.5
500	1	7.0	29.9	59.3	82.5	91.4	94.8	96.4	7.3	29.5	58.3	80.1	86.9	88.1	90.9
							Quadratic Spectral Kernel								
50	1	5.6	26.4	47.8	70.1	73.3	78.0	77.7	5.8	25.4	43.5	60.6	65.0	67.8	66.6
100	1	5.3	28.7	56.8	78.6	87.6	89.7	90.9	5.1	26.2	51.9	68.4	76.3	78.9	78.5
500	1	7.0	30.0	60.8	83.2	93.0	96.3	97.6	7.1	29.9	60.1	82.9	92.2	95.4	97.1
B. $c=-1$					δ		Bartlett Kernel					δ			
T	$e^{c/T}$	1	5	10	20	30	40	50	1	5	10	20	30	40	50
50	98.0	4.0	11.1	26.3	39.7	44.5	48.7	51.3	4.7	12.3	32.1	48.5	57.8	62.6	65.3
100	99.0	5.2	13.1	29.8	49.9	61.8	60.9	65.6	5.3	15.2	35.0	62.0	76.3	81.4	81.9
500	99.8	5.0	15.2	41.1	71.6	81.9	91.8	94.2	5.1	14.9	41.5	72.9	84.5	93.6	95.9
							Quadratic Spectral Kernel								
50	98.0	5.2	11.4	25.0	34.0	35.0	39.0	40.1	4.1	11.5	27.6	39.9	45.3	49.6	52.0
100	99.0	5.7	13.3	27.7	40.7	48.5	47.3	45.5	5.4	13.9	29.9	49.4	61.5	60.7	64.9
500	99.8	5.0	15.3	39.9	66.8	74.5	81.4	83.4	5.4	15.2	40.9	72.2	83.1	92.5	94.9
C. $c=-5$					δ		Bartlett Kernel					δ			
T	$e^{c/T}$	1	5	10	20	30	40	50	1	5	10	20	30	40	50
50	0.905	3.7	8.4	14.7	28.4	35.3	35.2	34.7	5.5	9.4	14.1	23.6	26.2	24.9	23.9
100	0.951	4.5	8.9	19.1	34.7	43.5	50.4	52.3	4.3	9.3	18.2	28.9	34.0	36.4	36.5
500	0.990	5.2	9.1	23.6	53.5	70.7	84.2	87.0	5.5	8.9	23.5	49.4	61.8	71.7	71.5
							Quadratic Spectral Kernel								
50	0.905	4.8	10.7	17.6	36.5	46.8	45.5	50.1	4.8	9.4	15.0	28.7	36.3	36.2	35.6
100	0.951	5.4	10.4	23.7	46.4	58.2	69.2	71.1	4.9	9.0	19.2	34.4	43.2	50.1	51.4
500	0.990	5.6	9.4	24.2	55.8	74.7	87.9	90.9	5.4	9.3	23.8	54.7	72.1	85.4	88.8
D. $c=-10$					δ		Bartlett Kernel					δ			
T	$e^{c/T}$	1	5	10	20	30	40	50	1	5	10	20	30	40	50
50	0.819	4.1	4.7	6.8	17.4	22.8	21.5	26.8	5.6	4.6	7.1	15.1	17.5	16.2	17.5
100	0.905	5.0	5.5	10.7	19.7	28.9	33.9	39.0	5.6	5.9	10.8	17.3	22.2	24.3	27.3
500	0.980	6.0	6.2	12.2	34.2	49.6	63.7	74.8	6.1	6.5	12.5	32.4	41.7	53.2	57.7
							Quadratic Spectral Kernel								
50	0.819	5.0	5.5	8.4	22.7	30.6	34.8	36.0	4.5	5.6	7.0	17.7	23.2	21.4	26.7
100	0.905	5.9	6.6	12.8	27.8	41.4	52.1	55.7	5.4	5.7	11.0	19.6	28.2	32.3	37.9
500	0.980	5.9	6.6	12.8	36.8	55.1	69.7	80.2	6.2	6.5	12.8	35.1	52.0	65.4	76.8

16

		Fixed bandwidth, $m(4)$							Automatic bandwidth, $n(4)$						
E. $c = -15$					δ		Bartlett Kernel					δ			
T	$e^{c/T}$	1	5	10	20	30	40	50	1	5	10	20	30	40	50
50	0.741	3.8	5.7	7.3	9.8	12.4	14.6	17.6	4.8	5.8	7.3	9.7	9.7	12.4	14.1
100	0.861	4.3	5.3	6.8	14.8	19.7	20.6	25.0	4.7	5.2	7.4	13.3	15.9	16.1	16.5
500	0.970	4.9	5.2	9.4	19.4	37.4	45.7	57.2	5.3	5.3	9.1	18.4	32.5	34.9	41.9
					Quadratic Spectral Kernel										
50	0.741	5.2	6.2	8.8	12.5	19.1	22.0	26.7	4.7	6.1	7.8	10.1	12.8	14.8	17.6
100	0.861	4.8	5.7	8.3	19.2	28.8	33.6	43.0	4.6	5.6	6.8	14.7	18.8	19.7	24.3
500	0.970	5.2	5.6	9.9	21.8	41.5	51.3	64.0	5.0	5.4	9.8	20.8	39.1	48.2	60.4

Table 5: Rejection rates (per cent)of the $KPSS$ test for the process (5a)-(5c).

	Fixed bandwidth							Automatic bandwidth						
A. $e^{c/T} = 0.7$			δ		Bartlett Kernel					δ			
T	1.0	5.0	10.0	20.0	30.0	40.0	50.0	1.0	5.0	10.0	20.0	30.0	40.0	50.0
50	3.8	4.8	6.2	9.4	11.3	13.3	14.3	4.9	5.6	6.8	9.0	9.8	11.2	11.0
100	4.7	4.1	4.7	6.2	8.0	9.1	10.7	4.8	4.4	4.7	6.3	7.7	8.2	9.0
500	4.9	4.8	4.7	4.5	5.0	5.0	5.5	4.9	4.9	4.8	4.5	5.0	4.9	5.3
				Quadratic Spectral Kernel										
50	4.3	5.5	7.4	12.9	16.1	19.9	21.3	4.3	5.1	6.6	9.8	11.5	13.3	14.0
100	5.1	4.7	5.2	7.7	10.5	13.4	16.5	5.0	4.5	4.9	6.4	7.8	8.8	9.9
500	4.9	4.8	4.8	4.6	5.0	5.2	5.6	4.9	4.8	4.7	4.6	5.1	5.0	5.6
B. $e^{c/T} = 0.8$			δ		Bartlett Kernel					δ			
T	1.0	5.0	10.0	20.0	30.0	40.0	50.0	1.0	5.0	10.0	20.0	30.0	40.0	50.0
50	3.9	5.4	7.8	14.6	18.3	20.8	22.4	5.1	6.2	8.0	12.9	14.6	15.3	15.5
100	4.4	4.4	6.1	9.2	12.7	15.5	17.7	4.7	4.5	6.2	8.8	11.1	12.1	12.9
500	4.6	4.7	5.2	4.8	5.4	6.0	6.6	4.6	4.7	5.2	4.7	5.3	6.0	6.6
				Quadratic Spectral Kernel										
50	4.7	6.2	9.9	19.6	26.4	30.3	32.9	4.5	5.8	8.3	14.8	18.6	21.0	22.5
100	4.9	4.8	7.0	12.2	18.7	24.8	28.3	4.6	4.6	6.2	9.1	12.4	14.7	16.7
500	4.7	4.7	5.4	5.0	5.6	6.5	7.2	4.6	4.7	5.3	4.9	5.4	6.2	7.0
C. $e^{c/T} = 0.9$			δ		Bartlett Kernel					δ			
T	1.0	5.0	10.0	20.0	30.0	40.0	50.0	1.0	5.0	10.0	20.0	30.0	40.0	50.0
50	4.3	6.9	14.2	26.8	32.9	36.7	37.4	5.6	7.5	13.5	22.0	24.4	25.9	26.5
100	4.6	5.7	9.8	20.2	28.4	33.6	35.5	4.7	5.8	9.5	17.6	21.7	23.9	23.1
500	4.6	4.8	5.4	6.5	8.0	9.9	13.3	4.7	4.8	5.3	6.3	7.7	9.1	11.7
				Quadratic Spectral Kernel										
50	4.9	8.1	17.8	35.1	43.6	49.2	50.6	4.7	7.3	14.8	27.5	33.5	37.2	38.0
100	5.2	6.6	11.8	27.2	40.4	49.3	53.6	4.8	6.0	9.9	19.9	27.7	32.5	34.2
500	4.6	4.8	5.4	6.8	8.5	11.1	15.4	4.6	4.9	5.3	6.7	8.3	10.3	14.1

	Fixed bandwidth							Automatic bandwidth						
D. $e^{c/T} = 0.95$				δ		Bartlett Kernel					δ			
T	1.0	5.0	10.0	20.0	30.0	40.0	50.0	1.0	5.0	10.0	20.0	30.0	40.0	50.0
50	4.3	10.4	22.4	37.9	43.4	46.3	48.7	5.4	10.6	21.0	31.4	33.1	35.3	36.0
100	4.6	8.1	17.8	35.2	44.0	49.2	53.0	4.6	8.1	16.5	29.4	32.8	34.4	36.7
500	4.9	4.7	6.5	11.4	18.0	26.6	33.8	4.8	4.8	6.3	11.1	16.5	22.1	26.0
						Quadratic Spectral Kernel								
50	4.9	12.3	26.9	47.2	55.2	59.7	62.9	4.8	11.0	23.0	38.7	44.1	47.1	49.4
100	5.2	9.4	21.8	45.8	59.4	67.4	72.9	4.8	8.4	17.8	34.7	43.1	48.3	52.0
500	5.0	4.8	6.7	12.2	19.9	30.0	38.6	5.0	4.8	6.6	11.8	18.9	28.0	35.7
E. $e^{c/T} = 0.97$				δ		Bartlett Kernel					δ			
T	1.0	5.0	10.0	20.0	30.0	40.0	50.0	1.0	5.0	10.0	20.0	30.0	40.0	50.0
50	4.1	12.7	28.2	44.5	51.4	53.4	55.2	5.1	13.2	26.7	37.6	41.0	42.2	43.3
100	4.9	10.7	25.2	45.3	54.7	58.2	61.2	5.1	10.9	23.3	37.7	42.5	42.9	44.2
500	4.5	5.9	8.9	21.0	35.0	47.5	55.8	4.6	6.0	8.8	19.9	30.8	37.5	41.3
						Quadratic Spectral Kernel								
50	4.8	14.8	33.4	53.8	62.2	66.4	68.3	4.4	13.5	28.7	45.0	52.0	54.1	55.8
100	5.4	12.4	29.7	56.7	70.2	76.3	79.9	5.0	10.9	25.2	44.9	53.9	57.2	60.1
500	4.4	6.1	9.4	22.7	38.5	52.8	62.6	4.5	6.0	9.2	21.8	36.4	49.5	58.7
F. $e^{c/T} = 0.99$				δ		Bartlett Kernel					δ			
T	1.0	5.0	10.0	20.0	30.0	40.0	50.0	1.0	5.0	10.0	20.0	30.0	40.0	50.0
50	4.7	18.5	37.7	54.4	59.2	61.8	63.8	5.7	18.7	35.7	47.8	50.0	51.7	52.5
100	5.0	16.3	38.4	58.9	67.1	69.6	71.4	5.0	16.2	36.3	50.8	54.6	54.8	55.7
500	4.9	9.2	24.7	53.4	71.9	80.6	86.4	4.9	9.1	24.0	49.8	63.2	67.9	71.3
						Quadratic Spectral Kernel								
50	5.2	21.6	43.0	63.0	69.0	72.6	75.4	5.1	19.2	38.3	54.9	59.7	62.3	64.4
100	5.4	18.4	43.3	69.8	80.8	84.0	87.2	5.2	16.6	38.4	58.5	66.4	68.8	70.5
500	5.0	9.4	25.5	55.9	75.1	84.7	90.2	5.1	9.4	25.1	54.7	73.3	82.3	88.2

References

Andrews, D.W.K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

Andrews, D.W.K. and C. Mohanan (1992), 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica* **60**, 953–966.

Caner, M. and L. Kilian (2001), 'Size distortions of tests of the null hyopthesis of statioanrity: Evidence and implications for the PPP debate', *Journal of International Money and Finance* **20**, 639–657.

Cappuccio, N. and D. Lubian (2006), Understanding size distortion of kpss test, Technical report, mimeo.

Choi, I. (1994), 'Residual based tests for the null of stationarity with applications to u.s. macroeconomic time series', *Econometric Theory* **10**, 720–746.

Harvey, A.C. (1995), 'Trends and cycles in macroeconomic time series', *Journal of Business and Economic Statistics* **3**, 216–227.

Hobijn, B., P. H. Franses and M. Ooms (2004), 'Generalizations of the kpss-test for stationarity', *Statistica Neerlandica* **58**, 483–502.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (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.

Lanne, M. and P. Saikkonen (2003), 'Reducing size distortions of parameteric stationarity tests', *Journal of Time Series Analysis* **24**, 423–439.

Lee, J. (1996), 'On the power of stationarity tests using optimal bandwidth estimates', *Econoic Letters* **51**, 131–137.

Leybourne, S.J. and B.P.M. McCabe (1994), 'A consistent test for a unit root', *Journal of Business & Economic Statistics* **12**, 157–166.

Müller, U. (2005), 'Size and power of tests of stationarity in highly autocorrelated time series', *Journal of Econometrics* **118**, 195–213.

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

Nabeya, S. and P. Perron (1994), 'Local asymptotic distribution related to the $ar(1)$ model with dependent errors', *Journal of Econometrics* **62**, 229–264.

Newey, W. K. and K. D. West (1994), 'automatic lag selection in covariance matrix estimation', *Review of Economic Studies* **61**, 631–653.

Nyblom, J. (1986), 'Testing for deterministic linear trend in time series', *Journal of the American Statistical Association* **81**, 545–549.

Nyblom, J. and Timo Mäkeläinen (1983), 'Comparisons of tests for the presence of random walk coefficients in a simple linear model', *Journal of the American Statistical Association* **78**, 856–864.

Pantula, S.G. (1991), 'Asymptotic distributions of the unit-root tests when the process is nearly statioanry', *Journal of Business and Economic Statistics* **9**, 325–353.

Perron, P and S. Ng (1996), 'Useful modifications to some unit root tests with dependent errors and their local asymptotic properties', *Review of Economic Studies* **63**, 435–563.

Phillips, P. C. B. and P. Perron (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**, 335–346.

Phillips, P.C.B. (1991), Spectral regression for cointegrated time series, *in* W.Barnett, J.Powell and G.Tauchen, eds, 'Nonparametric and Semiparametric Methods in Econometrics and Statistics', Cambridge University Press, Cambridge, pp. 413–435.

Phillips, P.C.B. and V. Solo (1992), 'Asymptotic theory for linear processes', *The Annals of Statistics* **20**, 971–1001.

Saikkonen, P and R. Luukkonen (1993), 'Testing for a moving average unit root in autoregressive integrated moving average models', *Journal of the American Statistical Association* **88**, 596–601.

Stock, J. H. (1994), Unit root, structural breaks and trends, *in* R. F.Engle and D. L.McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, Amsterdam, pp. 2739–2841.

Stock, J. H. and M. W. Watson (1998), 'Median unbiased estimation of coefficient variance in a time-varying parameter model', *Journal of the American Statistical Association* **93**, 349–358.