Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Econometrics 128 (2005) 195–213

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Size and power of tests of stationarity in highly autocorrelated time series

Ulrich K. Müller

Princeton University, Economics Department, Princeton, NJ 08544, USA

Accepted 1 April 2004
Available online 12 October 2004

Abstract

Tests of stationarity are routinely applied to highly autocorrelated time series. Following Kwiatkowski et al. (J. Econom. 54 (1992) 159), standard stationarity tests employ a rescaling by an estimator of the long-run variance of the (potentially) stationary series. This paper analytically investigates the size and power properties of such tests when the series are strongly autocorrelated in a local-to-unity asymptotic framework. It is shown that the behavior of the tests strongly depends on the long-run variance estimator employed, but is in general highly undesirable. Either the tests fail to control size even for strongly mean reverting series, or they are inconsistent against an integrated process and discriminate only poorly between stationary and integrated processes compared to optimal statistics.

JEL classification: C12; C22

Keywords: Local-to-unity asymptotics; Long-run variance estimation; Mean reversion; Efficient stationarity tests

1. Introduction

Most macroeconomic time series in levels exhibit strong positive autocorrelation. The largest autoregressive root of reasonable models for such series is hence in the

E-mail address: umueller@princeton.edu (U.K. Müller).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.08.012

neighborhood of unity. For various reasons, researchers have been interested in knowing whether this largest autoregressive root is indeed one—in which case the series is called 'integrated'—or smaller than one, leading to a stationary series. The standard tool for this purpose are unit root tests, which test the null hypothesis of a unit root against the alternative hypothesis of stationarity. Efficient unit root tests (cf. Dufour and King, 1991; Elliott et al., 1996) direct their power at distinguishing a series with a unit root from a series with a large, but stationary autoregressive root.

In some instances, it might be desirable to rather perform a test of the null hypothesis of stationarity against the alternative of a unit root, a so-called 'test of stationarity', or 'stationarity test'. For the derivation of such a test it seems appealing to rely on a model in which the null hypothesis of stationarity naturally arises as a restricted form of the model. Research has concentrated on a components model, in which, omitting deterministics, the observable time series is the sum of stationary disturbances and an integrated component. An especially simple version of the model arises when the stationary component consists of independent Gaussian random variables, and derivations of the locally best test have focussed on this case (cf. Nyblom, 1989). In this model the null hypothesis restricts the integrated component to be constant and the series becomes Gaussian white noise, whereas under the local alternative the process is a sum of a small integrated component and Gaussian white noise. Loosely speaking, tests of stationarity in such a model direct their power at detecting the small integrated component in the dominating white noise.

While this simple set-up is attractive for the construction of stationarity tests in the sense that the null hypothesis of stationarity becomes naturally a single hypothesis, it does not lend itself to the analysis of most macro time series in levels: neither the null nor the local alternative model generate data with even nearly the same amount of persistence as the observed series. Despite this background, tests of stationarity are routinely applied to very strongly autocorrelated series. Researchers justify the applicability of tests of stationarity to such series by referring to a correction suggested by Kwiatkowski et al. (1992), abbreviated KPSS in the following. The idea of KPSS is to account for the autocorrelation by dividing the locally best test statistic by an estimator of the so called long-run variance λ of the stationary component. Intuitively, this rescaling has to accomplish a delicate task: On the one hand, it has to compensate the change in the test statistic induced by a strong, but stationary autocorrelation in order to control size under the null hypothesis of stationarity. On the other hand, its presence must not compromise the ability of the test statistic to correctly reject the null hypothesis when the strong sample autocorrelation is in fact the result of an integrated process.

And indeed, the literature contains some evidence that various estimators of the long-run variance, $\hat{\lambda}$, yield unsatisfactory results. In their simulations, KPSS recognized the potential fragility of size control of their stationarity tests with respect to disturbances that follow an order-one autoregressive process. Caner and Kilian (2001) demonstrate by means of a Monte Carlo study that the tests massively overreject in the presence of strong autocorrelation. Lee (1996) investigates different estimators of the long-run variance and finds that some lead to acceptable size control, but at the cost of dramatically reduced power. The Monte Carlo results of Hobijn et al. (1998) corroborate this picture.

This paper develops a deeper understanding of the issues involved by analyzing size and power of tests of stationarity under local-to-unity asymptotics. The idea of analyzing the distribution of tests of stationarity under such non-standard asymptotics was already briefly mentioned in the survey article of Stock (1994, p. 2826), but no special attention is given there to $\hat{\lambda}$. The local-to-unity framework, developed by Chan and Wei (1987) and Phillips (1987), generates relevant asymptotics for series whose dynamics are dominated by a large autoregressive root. They give much more accurate approximations to small sample distributions compared to standard asymptotics when the largest autoregressive root ρ of a series is such that $T(1-\rho)$ is smaller than, say, 30, where T is the sample size. Stock and Watson (1998) estimate values for $T(1-\rho)$ in the region of 3–15 for U.S. annual series of GDP, consumption, investment, government purchases, 10-year Treasury Bond interest rates and 90-day Treasury Bill interest rates with $T = 44$ (OLS estimates of Tables 6 and 7). Analyses of real exchange rate data find half-lives of deviations from Purchasing Power Parity of about 3–5 years (cf. Rogoff (1996)), implying a $T(1-\rho)$ in the region of 14–23 for 100 years of data. An analysis in a local-to-unity framework reveals the behavior of tests of stationarity when applied to such series, which in turn helps the applied econometrician to understand and correctly interpret the test outcomes.

The paper shows that the behavior of tests of stationarity crucially hinges upon the estimator of the long-run variance in local-to-unity asymptotics. There are two key results: First, estimators of the long-run variance that employ a bandwidth that goes to infinity more slowly than the sample size lead to tests of stationarity that reject even highly mean reverting series with probability one for a large enough sample size. Second, for some estimators of λ that employ a bandwidth of the same order as the sample size, the resulting tests of stationarity do reject more often for less mean reverting series, but the exact properties depend crucially on which estimator $\hat{\lambda}$ is used.

It is well understood that there cannot exist a statistic that perfectly discriminates between stationary and integrated processes in the local-to-unity framework—in fact, much of the appeal of this asymptotic device stems precisely from the fact that discrimination remains difficult even as the sample size increases without bound. The failure of tests of stationarity to reliably discriminate between the two hypotheses under local-to-unity asymptotics hence does not come as a surprise, and is per se no compelling argument against their usage.

But surely researchers should aim at using a test with as good a size versus power trade-off as possible. By comparing the performance of KPSS-type tests of stationarity with optimal stationarity tests in a local-to-unity framework, it is shown that KPSS-type tests have much less discriminatory power than efficient tests. In this sense, the properties of current tests of stationarity are much worse than they need to be in highly autocorrelated time series.

The remainder of the paper is organized as follows. The next section introduces the test statistics and the local-to-unity asymptotic framework, and derives the size and power properties of tests of stationarity for various estimators of the long-run variance. Section 3 compares the performance of the tests of stationarity which employ the most promising estimators of the long-run variance with optimal stationarity tests in a local-to-unity set-up. Section 4 concludes. Proofs are collected in an appendix.

2. Tests of stationarity under local-to-unity asymptotics

The Data Generating Process tests of stationarity are build upon is given by

$$y_t = d_t + w_t + \psi_t \quad \text{with} \quad w_t = w_{t-1} + \upsilon_t, \tag{1}$$

where y_t, $t = 1, \ldots, T$, is the observed sample, d_t is a deterministic component and $\{\psi_t\}$ and $\{\upsilon_t\}$ are independent zero mean stationary series. Under the null hypothesis of stationarity, the variance of υ_t is restricted to be zero, such that $\{\psi_t + w_0\}$ is stationary. Under the alternative hypothesis, $\mathrm{E}[\upsilon_t^2]>0$, so that $\{w_t\}$ is an integrated series, and the disturbances $\{w_t + \psi_t\}$ are a sum of an integrated component $\{w_t\}$ and a stationary component $\{\psi_t\}$.

The test statistic of KPSS is constructed as follows: Regress $\{y_t\}$ on deterministic components which consist either of a constant (indicated by a superscript μ throughout the paper) or of a constant and time trend (indicated by a superscript τ) by ordinary least squares. Denote the resulting residuals with $\{y_t^i\}$, where $i = \mu, \tau$, and compute $S_t^i = \sum_{s=1}^t y_s^i$. The test statistic is then given by

$$L^i(\hat{\lambda}) = \frac{T^{-2} \sum_{t=1}^T (S_t^i)^2}{\hat{\lambda}}, \tag{2}$$

where $\hat{\lambda}$ is an estimator of the long-run variance of $\{\psi_t\}$, $\lambda = \sum_{j=-\infty}^{\infty} \mathrm{E}[\psi_t \psi_{t-j}]$,[1] and the null hypothesis of stationarity is rejected for large values of $L^i(\hat{\lambda})$. KPSS show that under some regularity conditions and an appropriate choice of $\hat{\lambda}$ the asymptotic distribution of $L^i(\hat{\lambda})$ under the null hypothesis of stationarity is given by $L^i(\hat{\lambda}) \Rightarrow \int W^i(s)^2\, \mathrm{d}s$, where '$\Rightarrow$' denotes weak convergence as $T \to \infty$, $W(s)$ is a Wiener process, $W^\mu(s) = W(s) - sW(1)$, $W^\tau(s) = W(s) + (2s - 3s^2)W(1) + 6(s^2 - s)\int W(l)\,\mathrm{d}l$ and for notational simplicity, the limits of integration are understood to be zero and one, if not indicated otherwise. 5% critical values of $L^i(\hat{\lambda})$ are known to be 0.463 and 0.146 in the mean and mean and time trend case, respectively (cf. MacNeill (1978) and Nabeya and Tanaka (1988)).

In contrast to the assumptions in KPSS, we analyze the behavior of $L^i(\hat{\lambda})$ when y_t is generated by a Data Generating Process which is standard in the unit root testing literature. Specifically, let

$$y_t = d_t + u_t \quad \text{with} \quad u_t = \rho u_{t-1} + v_t, \tag{3}$$

[1]KPSS define the long-run variance λ (which is σ^2 in their notation) as $\lim_{T\to\infty} T^{-1}\mathrm{E}[(S_T^\tau)^2]$ (p. 164). S_T^τ is identical zero, however. One obtains the asymptotic distributions derived by KPSS when $\hat{\lambda}$ is a consistent estimator of the long-run variance of $\{\psi_t\}$, which is the definition employed in this paper.

where, if $|\rho|<1$, $u_0 = \sum_{s=0}^{\infty} \rho^s v_{-s}$, u_0 is arbitrary for $\rho = 1$ and d_t consists either of a mean or of a mean and time trend. Throughout the paper we assume that if a time trend is present in (3), then the τ-version of $L^i(\hat{\lambda})$ is used; in this sense the deterministics are assumed to be correctly specified.

If $\rho = 1$, then different values of u_0 induce mean shifts of $\{y_t\}$. But the residuals y_t^i are independent of the mean of $\{y_t\}$, so that no additional assumption concerning u_0 is necessary when $\rho = 1$. If $|\rho|<1$, the assumption on the generation of u_0 leads to a stationary series $\{u_t\}$ as long as $\{v_t\}$ is stationary. While somewhat natural, this assumption might considerably affect the asymptotic distributions derived below. See Müller and Elliott (2003) for discussion.

The innovations $\{v_t\}$ that underlie the autoregressive process $\{u_t\}$ have not yet been given any structure. For most of the asymptotic derivations below, we only need to impose the following, rather weak condition.

Condition 1. The zero mean process $\{v_t\}$ is covariance-stationary with finite autocovariances $\gamma(j) = \mathrm{E}[v_t v_{t-j}]$ such that

(a) $\omega^2 = \sum_{j=-\infty}^{\infty} \gamma(j)$ is finite and nonzero,
(b) the scaled partial-sum process $T^{-1/2} \sum_{t=1}^{[sT]} v_t \Rightarrow \omega W(s)$.

In contrast to the reasoning of KPSS, the following derivations employ local-to-unity asymptotics, i.e. ρ in (3) is made a function of the sample size such that $\rho = \rho_T = 1 - \gamma T^{-1}$, where $\gamma \geqslant 0$ is a fixed number. Lemma 2 in Elliott (1999) shows that under Condition 1, the process u_t can then be asymptotically characterized by

$$T^{-1/2}(u_{[Ts]} - u_0) \Rightarrow \omega M(s) \equiv \begin{cases} \omega W(s) & \text{for } \gamma = 0, \\ \omega \zeta (\mathrm{e}^{-\gamma s} - 1)(2\gamma)^{-1/2} + \omega \int_0^s \mathrm{e}^{-\gamma(s-l)}\, \mathrm{d}W(l) & \text{else,} \end{cases} \tag{4}$$

where ζ is a standard normal variable independent of $W(\cdot)$. Note that for $\gamma > 0$, the weak limit of the covariance-stationary series $T^{-1/2}\omega^{-1} u_{[T\cdot]}$, $\tilde{M}(\cdot) = M(\cdot) + \zeta(2\gamma)^{-1/2}$, is a stationary continuous time process.

The relationship between (3) and the Data Generating Process (1) assumed by KPSS is straightforward: For $|\rho|<1$, (3) is a special case of (1) under the null hypothesis of $\mathrm{E}[\upsilon_t^2] = 0$ with $\psi_t = u_t$, and for $\rho = 1$ (3) is a special case of (1) with $\upsilon_t = v_t$ and $\psi_t = 0$. KPSS have derived the properties of $L^i(\hat{\lambda})$ under the null hypothesis of stationarity with standard asymptotics, which corresponds to an asymptotic reasoning with fixed $|\rho|<1$ in (3). As shown below, $L^i(\hat{\lambda})$ has radically different properties in a local-to-unity framework. This raises the question which asymptotic reasoning inference should be based upon.

The ultimate goal of all asymptotic reasoning is to provide useful small sample approximations. It was shown elsewhere (cf., for instance, Nabeya and Tanaka (1990) or Perron and Vodounou (2001)) that local-to-unity asymptotics provide much more accurate small sample approximations when a series contains a large autoregressive root than standard ($|\rho|<1$ fixed) asymptotics. The following results may hence be interpreted as more accurate small sample approximations of the behavior of tests of stationarity when applied to highly persistent processes, which include most macroeconomic series in levels.

The local-to-unity process $M(s)$ with $\gamma > 0$ is an asymptotic representation of a series that reverts to a constant mean. As long as $\gamma > 0$ we thus analyze the behavior of tests of stationarity under the null hypothesis of stationarity. Given that KPSS-type tests were designed to detect a small integrated component, one might argue that the analysis should consider power not only against the 'purely' integrated alternative $\gamma = 0$, but also against an alternative where the data suitably scaled converges weakly to $M(\cdot) + \kappa W_w(\cdot)$, where $W_w(\cdot)$ is a standard Wiener process independent of $M(\cdot)$ and the constant κ describes the relative size of the integrated component. Such an extension is straightforward; the results below, however, reveal such poor properties of KPSS-type stationarity tests under the—in almost all applications at least plausible—null hypothesis of weak mean reversion that the issue is not pursued further.

From (4), straightforward calculations reveal (cf., for instance, Stock, 1994, p. 2772) that the residuals y_t^i satisfy

$$T^{-1/2} y^i_{[Ts]} \Rightarrow \omega M^i(s), \tag{5}$$

where $M^\mu(s) = M(s) - \int M(l)\,\mathrm{d}l$ and $M^\tau(s) = M(s) - (4-6s)\int M(l)\,\mathrm{d}l - 6(2s-1)\int lM(l)\,\mathrm{d}l$. The asymptotic distribution of the (scaled) numerator of $L^i(\hat{\lambda})$ now follows from an application of the continuous mapping theorem (CMT):

$$T^{-4}\sum_{t=1}^{T}(S_t^i)^2 \Rightarrow \omega^2 \int_0^1 \left[\int_0^s M^i(l)\,\mathrm{d}l\right]^2 \mathrm{d}s. \tag{6}$$

Note that under local-to-unity asymptotics, the numerator of $L^i(\hat{\lambda})$ must be divided by an additional T^2 in order to obtain a stable and nondegenerate asymptotic distribution. For this to happen, $\hat{\lambda}$ must hence be of order $\mathrm{O_p}(T^2)$.

Following KPSS, we first consider estimators of λ that are a weighted sum of sample covariances: Define

$$\hat{\lambda}_k(B_T) = \hat{\eta}(0) + 2\sum_{j=1}^{T} k\left(\frac{j}{B_T}\right)\hat{\eta}(j), \tag{7}$$

where $\hat{\eta}(j) = T^{-1}\sum_{t=1}^{T-j} y_t^i y_{t+j}^i$. The continuous function $k : [0,\infty) \to [-1;1]$ serves as the weighting function of the sample autocovariances and is assumed to satisfy $k(0) = 1$, $\int_0^\infty |k(s)|\,\mathrm{d}s < \infty$ and $\lim_{s\to\infty} k(s) = 0$. The bandwidth B_T is, for now, a deterministic function of the sample size. The larger B_T the more weight is attached in (7) to higher-order sample autocovariances. These assumptions on the form of spectral density estimators are very similar to those made in Andrews (1991) and encompass all usual weighting schemes. The popular Bartlett estimator with lag truncation parameter m, for instance, can be represented in this notation with $k(x) = k_{\mathrm{B}}(x) = 1 - |x|$ for $|x| < 1$, $k_{\mathrm{B}}(x) = 0$ for $|x| \geqslant 1$ and $B_T = m + 1$.

In a standard asymptotic framework it can usually be shown that long-run variance estimators of the form (7) are consistent when $B_T = \mathrm{o}(T^{1/2})$ or $B_T = \mathrm{o}(T)$—see Andrews (1991). KPSS, for instance, employ a Bartlett weighting with $B_T = o(T^{1/4})$ in their simulations, and such a choice is also popular in applied work. Finally, while making B_T dependent on the sample, the long-run estimators suggested by Hobijn et al. (1998) satisfy $B_T = o_p(T)$, too.

The following proposition establishes the behavior of $L^i(\hat{\lambda}_k(B_T))$ in a local-to-unity asymptotic framework when $B_T = o_p(T)$.

Proposition 1. *Under Condition* 1 *and for any* $\gamma = T(1-\rho_T) \geqslant 0$, *if* $B_T = o_p(T)$, *then for any critical value* $\text{cv} \in \mathbb{R}$, $P(L^i(\hat{\lambda}_k(B_T)) > \text{cv})$ *converges to one as* $T \to \infty$.

In other words, tests based on $L^i(\hat{\lambda}_k(o_p(T)))$ reject the null hypothesis of stationarity with probability one under local-to-unity asymptotics. To demonstrate the relevance of this result, imagine that the observations y_t stem from a discrete sampling on the [0, 1] interval of the continuous time process $\tilde{M}(\cdot)$ with $\gamma = 70$. This process is highly stationary, the half-life period of a deviation from the mean is less than 1% of the sample size. It might be that a test based on $L^i(\hat{\lambda}_k(B_T))$ with a choice of bandwidth of order $o(T)$ only rarely rejects the null hypothesis of stationarity when the frequency of the observations is, say, $\frac{1}{100}$ (such that $T = 100$ and $y_1 = \tilde{M}(.01)$, $y_2 = \tilde{M}(.02)$, $\ldots$, $y_{100} = \tilde{M}(1)$). But Proposition 1 implies that, as the continuous time process is sampled more and more frequently (which leads to a larger sample size T), there must be a point where the test almost always rejects. As a real-world example, imagine that real exchange rates are mean reverting with a half-life of one year. If 100 years of exchange rate data are employed in a test of stationarity with $B_T = o_p(T)$, then the test is bound to reject the stationarity hypothesis as the sampling frequency increases from yearly data to monthly data to daily data, etc.

Proposition 1 also implies consistency of $L^i(\hat{\lambda}_k(B_T))$ with $B_T = o_p(T)$ in the sense that an integrated process ($\gamma = 0$) will be rejected with probability one. This is the reason that the above-mentioned authors promote bandwidths that are of order $o_p(T)$. But Proposition 1 reveals the steep price which has to be paid for this consistency result: tests of stationarity with $B_T = o_p(T)$ control size arbitrarily badly in the sense that for any amount of mean reversion measured by $\gamma = T(1-\rho_T)$ a high enough sample frequency will lead to rejection with probability one. This necessarily leads to a contradiction with the outcome of any reasonable unit root test, which have nontrivial power for any $\gamma > 0$ and at a 5% level achieve power of about 90% and 70% already for $\gamma = 20$ in the mean and trend cases, respectively (cf. Elliott, 1999).

Taking the asymptotic result of Proposition 1 as an approximation for finite samples, one would expect frequent rejections of $L^i(\hat{\lambda}_k(B_T))$ with $B_T = o(T)$ for highly autocorrelated, but stationary series. And this is precisely what Caner and Kilian (2001) find in a Monte Carlo study with a Bartlett weighting and $B_T = [12(T/100)^{1/4}]$. At a 5% nominal level and in a Gaussian sample with $T = 100$, for instance, the rejection rates are 55.4% for an autoregressive process of order one and root 0.95 in the mean case and 38.0% in the trend case (Caner and Kilian (2001), Table 1).

We now turn to the asymptotic analysis of $L^i(\hat{\lambda})$ for some classes of long-run variance estimators that are of order $O_p(T^2)$. A first such class is given by $\hat{\lambda}_k(hT)$, where $\hat{\lambda}_k(\cdot)$ is defined above and h is positive constant—this corresponds to the estimators recently suggested by Kiefer and Vogelsang (2002). A second $O_p(T^2)$ estimator arises when $\hat{\lambda}$ is estimated by an autoregressive long-run variance estimator. These estimators are popular in time series econometrics and try to capture the correlations in $\{v_t\}$ by an autoregressive parametrization. $\hat{\lambda}_{AR}$ is computed by running the ordinary least squares regression

$$y_t^i = a_1 y_{t-1}^i + a_2 y_{t-2}^i + \cdots + a_p y_{t-p}^i + e_t \tag{8}$$

followed by the computation of $\hat{\lambda}_{AR} = \hat{\sigma}_e^2/(1 - \sum_{i=1}^p \hat{a}_i)^2$, where $\hat{a}_i$ and $\hat{\sigma}_e^2$ are the estimated parameters in (8).

A third class of estimators first 'prewhitens' the data by a low-order autoregression just like (8) and then applies a standard spectral density estimator to the residuals—see Andrews and Monahan (1992) for further discussion. Specifically, we consider a prewhitening scheme where the autoregression is of order one, i.e.

$$y_t^i = \rho_w y_{t-1}^i + e_{w,t} \tag{9}$$

and the spectral density estimator $\hat{\omega}_e^2$ of the residuals $\hat{e}_{w,t}$ is constructed analogously to (7) with a bandwidth $b_T = o(T^{1/2})$. The long-run variance estimator is then given by $\hat{\lambda}_{PW} = (1 - \hat{\rho}_w)^{-2}\hat{\omega}_e^2$.

Finally, we consider spectral density estimators (7) where the bandwidth is endogenously determined by the data, as suggested by Andrews (1991). The computation of the bandwidth requires the estimation of a parametric model, and we follow Andrews (1991) by estimating the AR(1) specification (9). We concentrate the discussion on two kernels, the Bartlett kernel $k_B(\cdot)$ introduced above and the quadratic spectral kernel $k_{QS}(\cdot)$ as defined in Andrews (1991, p. 821). The endogenous bandwidths for these two kernels are given by

$$B_{B,T} = 1.1447\left[\frac{4\hat{\rho}_w^2}{(1-\hat{\rho}_w)^2(1+\hat{\rho}_w)^2}T\right]^{1/3} \quad \text{and}$$
$$B_{QS,T} = 1.3221\left[\frac{4\hat{\rho}_w^2}{(1-\hat{\rho}_w)^4}T\right]^{1/5} \tag{10}$$

and the resulting estimators are denoted $\hat{\lambda}_{A,B}$ and $\hat{\lambda}_{A,QS}$, respectively.[2]

[2]For the AR(p) estimator, the prewithening estimator and the automatic bandwidth selection estimators of the long-run variance the question arises how to treat explosive estimates of the AR processes. The probability of such estimates is below 5% in the mean case and below 1% in the trend case even for an integrated process, at least asymptotically, due to the heavily skewed distribution of the estimator of the largest autoregressive root. In order to keep things as straightforward as possible, we chose to treat negative $(1-\hat{\rho})$ just like their positive counterparts. An alternative solution is to trim the estimates away from zero, as suggested in Andrews (1991) and Andrews and Monahan (1992). These authors propose a trimming at $\hat{\rho} = .97$ for $T = 128$, which corresponds to a trimming of $T(1-\hat{\rho})$ at 3.84. Results not reported here show that such a trimming has very little impact on the asymptotic local rejection rates of $L^\tau(\hat{\lambda})$ for any considered $\hat{\lambda}$, moderately increases asymptotic rejection rates of $L^\mu(\hat{\lambda}_{AR})$ and $L^\mu(\hat{\lambda}_{PW})$ and leads to very few rejections of $L^\mu(\hat{\lambda}_{A,B})$ and $L^\mu(\hat{\lambda}_{A,QS})$ at a 5% level.

Proposition 2. *For any* $\gamma = T(1-\rho_T) \geqslant 0$,

(i) *under Condition* 1

$$L^i(\hat{\lambda}_k(hT)) \Rightarrow \frac{\int_0^1 \left[\int_0^s M^i(l)\,\mathrm{d}l\right]^2 \mathrm{d}s}{2\int_0^1 k(hl)\int_0^{1-s} M^i(l)M^i(s+l)\,\mathrm{d}l\,\mathrm{d}s},$$

(ii) *if* $\{v_t\}$ *is a stable autoregressive process of order* $p-1$ *where the underlying disturbances* $\{\varepsilon_t\}$ *satisfy* $\mathrm{E}[\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\ldots]=0$, $\mathrm{E}[\varepsilon_t^2]=\sigma^2>0$ *and* $\mathrm{E}[\varepsilon_t^4]<\infty$

$$L^i(\hat{\lambda}_{\mathrm{AR}}) \Rightarrow \frac{\int_0^1 \left[\int_0^s M^i(l)\,\mathrm{d}l\right]^2 \mathrm{d}s[M^i(1)^2 - M^i(0)^2 - 1]^2}{4[\int M^i(s)^2\,\mathrm{d}s]^2},$$

(iii) *if in addition to Condition* 1, $\{v_t\}$ *satisfies Assumption A of Andrews* (1991), $b_T \to \infty$ *and* $b_T = \mathrm{o}(T^{1/2})$

$$L^i(\hat{\lambda}_{\mathrm{PW}}) \Rightarrow \frac{\int_0^1 [\int_0^s M^i(l)\,\mathrm{d}l]^2\,\mathrm{d}s[M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2}]^2}{4[\int M^i(s)^2\,\mathrm{d}s]^2},$$

(iv) *under Condition* 1, *for* $j \in \{B, \mathrm{QS}\}$

$$L^i(\hat{\lambda}_{A,j}) \Rightarrow \frac{\int_0^1 \left[\int_0^s M^i(l)\,\mathrm{d}l\right]^2 \mathrm{d}s}{2\int_0^1 k_j(\frac{l}{B_{A,j}})\int_0^{1-s} M^i(l)M^i(s+l)\,\mathrm{d}l\,\mathrm{d}s},$$

$$\textit{where } B_{A,B} = 1.8171\left|\frac{2\int M^i(s)^2\,\mathrm{d}s}{M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2}}\right|^{2/3} \textit{ and}$$

$$B_{A,\mathrm{QS}} = 1.7445\left|\frac{2\int M^i(s)^2\,\mathrm{d}s}{M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2}}\right|^{4/5}.$$

The additional assumption that is invoked in part (iii) of the proposition is technical and requires fourth-order stationarity and certain bounds on the fourth-order cumulants of $\{v_t\}$—see Andrews (1991) for details. The condition on $\{v_t\}$ in part (ii) is such that the proof can rely on the reasoning of Stock (1991).

The asymptotic distributions of $L^i(\hat{\lambda}_{\mathrm{PW}})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,\mathrm{QS}})$ of Proposition 2 depend on the ratio $\gamma(0)/\omega^2$, whereas this is not the case for $L^i(\hat{\lambda}_k(hT))$ and $L^i(\hat{\lambda}_{\mathrm{AR}})$. The local-to-unity asymptotic rejection profiles of the two former versions of $L^i(\hat{\lambda})$ are hence not only a function of γ, but also of the correlation structure of $\{v_t\}$. Note that when $\gamma(0) = \omega^2$, the asymptotic distributions of $L^i(\hat{\lambda}_{\mathrm{PW}})$ and $L^i(\hat{\lambda}_{\mathrm{AR}})$ coincide.

Fig. 1 depicts the asymptotic rejection rates of $L^i(\hat{\lambda}_k(hT))$ with a Bartlett weighting $k = k_{\mathrm{B}}$ and $h = 0.05$, 0.1 and 0.2, $L^i(\hat{\lambda}_{\mathrm{AR}})$, $L^i(\hat{\lambda}_{\mathrm{PW}})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,\mathrm{QS}})$ for the 5% nominal level as a function of γ. For $L^i(\hat{\lambda}_{\mathrm{PW}})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,\mathrm{QS}})$, rejection rates are investigated when $\gamma(0)/\omega^2$ is set to $\frac{1}{2}$, 1 and 2, respectively. The asymptotic behavior of $L^i(\hat{\lambda})$ is very different for the different estimators $\hat{\lambda}$ considered. A test based on $L^i(\hat{\lambda}_{k_{\mathrm{B}}}(hT))$ has arejection profile that is monotonically decreasing in γ, and

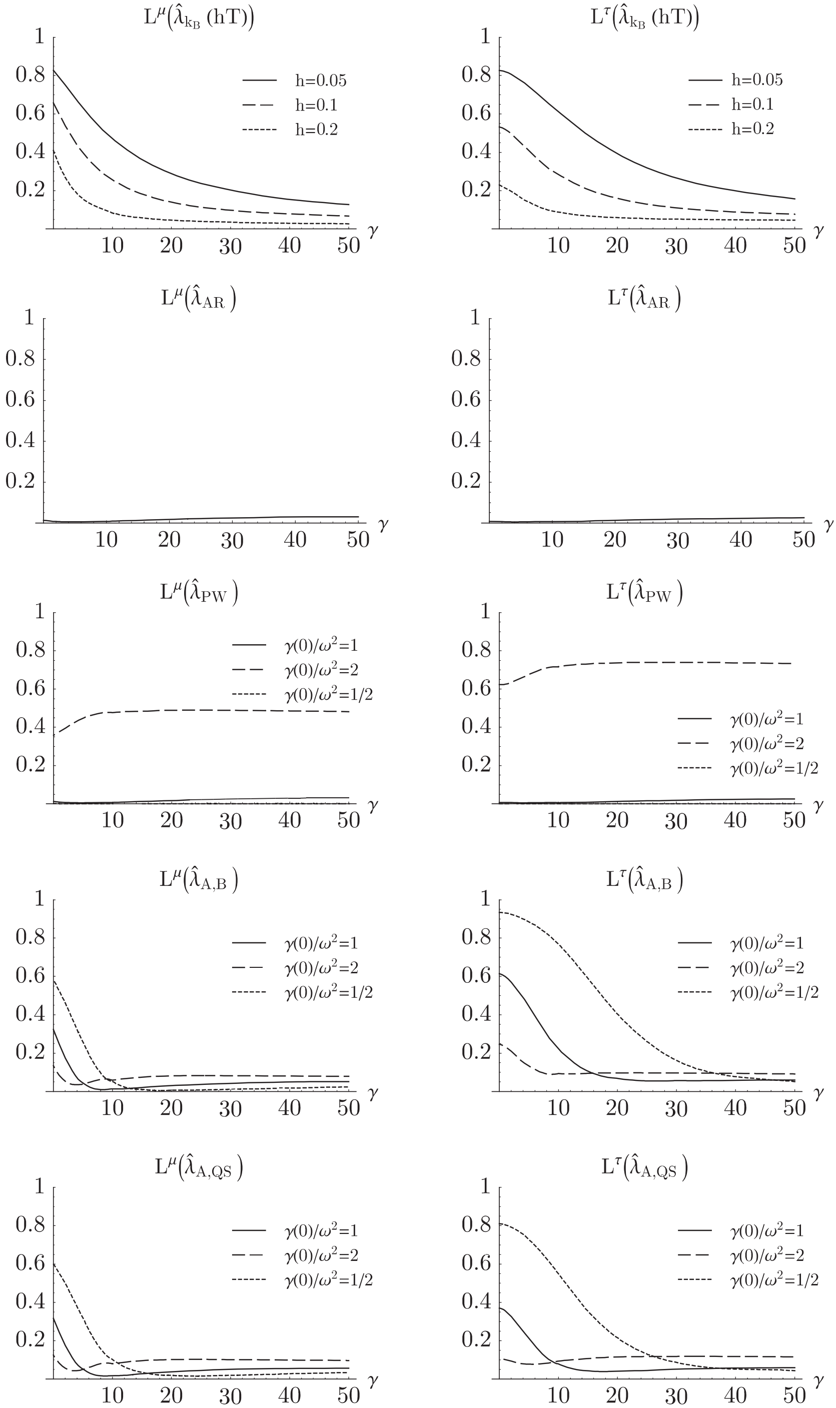

Fig. 1. Asymptotic rejection rates of stationarity tests based on various long-run variance estimators.

larger values of h lead to fewer rejections. As the rate of mean reversion, γ, increases, the rejection rates approach the nominal level. $L^i(\hat{\lambda}_{\text{AR}})$ leads to a rate of rejections that is consistently below the level and increases in γ. This increasing slope is also found for $L^i(\hat{\lambda}_{\text{PW}})$, but there the frequency of rejections depends crucially on the value of $\gamma(0)/\omega^2$: When this ratio is 2, the test rejects much more often than $L^i(\hat{\lambda}_{\text{AR}})$, and for $\gamma(0)/\omega^2 = \frac{1}{2}$ the test practically ceases to reject. A reversed picture can be found for $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,\text{QS}})$. Here, the smaller value of $\gamma(0)/\omega^2$ leads to more rejections.

Part (i) of Proposition 2 implies that it is the ratio of the bandwidth B_T to the sample size T that determines the behavior of tests of stationarity in highly autocorrelated series. For a sample size of $T = 100$ and a Bartlett weighting window, for instance, a choice of $B_T = 10$ will approximately yield a test of stationarity with 5% nominal level with power of 70% and 50% against an integrated process in the mean and trend case, respectively, and size will be roughly 20% when the largest autoregressive root is $\rho = 1 - 10/100 = 0.9$. Caner and Kilian (2001) simulate Gaussian first-order autoregressive processes with $T = 100$ and a root of 0.9. They find for a Bartlett window and $B_T = 13$ that size is 17.6% in the mean case and 18.6% in the trend case.

Part (iii) of Proposition 2 explains Lee's (1996) results from a Monte Carlo study, in which he find that the rejection rates of $L^i(\hat{\lambda}_{\text{PW}})$ against a stationary process are well within the nominal level, but power against a pure random walk is also far below the nominal level. Furthermore, in the light of part (iv) of Proposition 2 and Fig. 1, Engel's (2000) observation that $L^{\mu}(\hat{\lambda}_{A,B})$ has very low power in 100 years of simulated quarterly exchange rate data which consists of a slowly mean reverting component ($\gamma \approx 30$) and a random walk component is no longer surprising.

All of the tests considered in Proposition 2 are inconsistent in the sense that they fail to reject an integrated process ($\gamma = 0$) with probability one. A similar point, although without providing asymptotic rejection rates, has already been made by Choi (1994) and Hobijn et al. (1998) with respect to $L^i(\hat{\lambda}_{\text{PW}})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,\text{QS}})$. Moreover, Hobijn et al. (1998) argue by a different set of arguments that the parametric correction suggested by Leybourne and McCabe (1994) and refined in Leybourne and McCabe (1999) suffers from the same drawback. Small sample simulations by Hobijn et al. (1998) reveal a rejection rate of the Leybourne and McCabe (1994) test of around 30% for processes with a dominating random walk. At least qualitatively, the Leybourne and McCabe tests therefore seem to behave similarly to KPSS-type tests with an $\text{O}_\text{p}(T)$ bandwidth. A more detailed asymptotic analysis, however, is beyond the scope of this paper.

Even for highly mean reverting series with $\gamma = 50$, only the size of $L^i(\hat{\lambda}_{k_\text{B}}(hT))$ with $h = 0.1$, 0.2 and $L^i(\hat{\lambda}_{\text{AR}})$ and, when $\gamma(0)/\omega^2 = 1$, $L^i(\hat{\lambda}_{A,B})$, $L^i(\hat{\lambda}_{A,\text{QS}})$ and $L^i(\hat{\lambda}_{\text{PW}})$, are close to the nominal level of 5%. Nevertheless, when compared to the asymptotic behavior of $L^i(\hat{\lambda}_k(B_T))$ with $B_T = \text{o}_\text{p}(T)$ revealed in Proposition 1, it still seems relatively preferable to use one of these long-run variance estimators for highly autocorrelated time series.

3. Comparison with optimal stationarity tests

One could conclude from the results of the last section that all $L^i(\hat{\lambda})$ fail to reliably discriminate between strongly autocorrelated, but stationary and integrated series and hence should not be used. Such a reasoning does not take into account, however, that this is true of any statistic: The near observational equivalence between models with ρ very close to unity and $\rho = 1$ makes it impossible to obtain the ideal asymptotic rejection profile of no (or rare) rejections for any $\gamma > 0$ and rejection with probability one for $\gamma = 0$.

It is possible, however, to derive optimal stationarity tests for model (3), that efficiently discriminate between a null hypothesis of at least a given amount of mean reversion and the alternative of integration

$$\mathrm{H}_0 : \gamma \geqslant g_0 > 0 \quad \text{against} \quad \mathrm{H}_1 : \gamma = 0. \tag{11}$$

Such an approach requires the specification of a minimal amount of mean reversion g_0 for the null hypothesis, which might or might not be appealing in practice. But whatever its absolute merits, such an optimal stationarity test is ideally suited to assess the relative discriminatory power of KPSS-type stationarity tests.

In order to derive asymptotically optimal stationarity tests for model (3), we note that the optimal unit root test statistics derived by Elliott et al. (1996) and further studied by Elliott (1999) and Müller and Elliott (2003) are point-optimal statistics that, based on the Neyman–Pearson Lemma, optimally discriminate between a fixed level of mean reversion $\gamma = g > 0$ and no mean reversion $\gamma = 0$. The asymptotic optimality property of the statistics requires Gaussian disturbances, but allows for unknown and very general correlations. Usually, these optimal statistics are employed to perform a hypothesis test with integration as the null hypothesis. But the Neyman–Pearson Lemma is not directional in the sense that it is the same statistic (the likelihood ratio or a monotonic transformation thereof) that optimally discriminates between two single hypotheses. A reversal of the null and alternative hypothesis only requires to reject for large (small) values when the original null hypothesis was rejected for small (large) values of the test statistic.

In the local-to-unity asymptotic framework an optimal test of stationarity may hence be based on an optimal unit root test statistic. Such a test maximizes the (positive) difference of rejection rates at $\gamma = 0$ and $\gamma = g$. While not achieving the ideal rejection profile either, this test maximizes power at $\gamma = 0$ for a given size at $\gamma = g$, or, equivalently, minimizes mistaken rejections when $\gamma = g$ for a given power at $\gamma = 0$. The question remains how to choose g; but Elliott (1999) finds that the asymptotic properties of the optimal statistics are rather insensitive to the specific choice of g. In other words, the optimal statistic for a specific g has also close to optimal discriminating power for values of $\gamma \neq g$. We follow Elliott's (1999) recommendation and set $g = 10$ in the mean case and $g = 15$ in the trend case in the following analysis.

Following Müller and Elliott (2001) the asymptotically optimal statistic to discriminate between $\rho = 1$ and $\rho_T = 1 - gT^{-1}$ in model (3) is, in the notation developed here, given by

$$Q^i(g) = q_1^i(\hat{\omega}^{-1}T^{-1/2}y_T^i)^2 + q_2^i(\hat{\omega}^{-1}T^{-1/2}y_1^i)^2 + q_3^i(\hat{\omega}^{-1}T^{-1/2}y_T^i)(\hat{\omega}^{-1}T^{-1/2}y_1^i) + q_4^i\hat{\omega}^{-2}T^{-2}\sum_{t=1}^{T}(y_t^i)^2, \tag{12}$$

where large values are evidence of nonstationarity, $q_1^\mu = q_2^\mu = g(1+g)/(2+g)$, $q_3^\mu = -2g/(2+g)$, $q_4^\mu = g^2$ and, $q_1^\tau = q_2^\tau = g^2(8+5g+g^2)/(24+24g+8g^2+g^3)$, $q_3^\tau = 2g^2(4+g)/(24+24g+8g^2+g^3)$, $q_4^\tau = g^2$ and $\hat{\omega}^2$ is a consistent estimator of the long-run variance ω^2 of $\{v_t\}$ under local-to-unity asymptotics. (The q_j^i differ from those in Theorem 3 of Müller and Elliott (2001) because their y_t^i is defined differently.) An example for a consistent estimator of ω^2 is the spectral density estimator $\hat{\omega}_e^2$ of the residual of regression (9); see the proof of part (iii) of Proposition 2 for details. Also see Stock (1994) for a general discussion. The local-to-unity asymptotic distribution of $Q^i(g)$ follows directly from the CMT

$$Q^i(g) \Rightarrow q_1^i M^i(1)^2 + q_2^i M^i(0)^2 + q_3^i M^i(1)M^i(0) + q_4^i \int M^i(s)^2\,\mathrm{d}s. \tag{13}$$

If one wanted to employ a stationarity tests based on $Q^i(g)$, one would need to choose g_0 and determine the appropriate critical value such that the hypothesis test (11) is correctly sized. For the comparison of the efficient stationarity tests and $L^i(\hat{\lambda})$ in Fig. 2, however, it suffices to find critical values such that the rejection rates of the efficient tests coincide with the rejection rates of $L^i(\hat{\lambda})$ at $\gamma = 0$.[3] In the standard terminology for tests of stationarity we compare 'size control' by looking at rejection rates for $\gamma > 0$. For all considered $\hat{\lambda}$, the rejection profile of $L^i(\hat{\lambda})$ (fine lines) is consistently above the rejection profile of the corresponding $Q^i(g)$ (fat lines) for $\gamma > 0$. This must be true for $\gamma = 10$ in the mean case and $\gamma = 15$ in the trend case by the asymptotic optimality property of $Q^\mu(10)$ and $Q^\tau(15)$, but also holds for all other considered values of γ. The relative inferiority of $L^i(\hat{\lambda})$ is most striking for $L^i(\hat{\lambda}_{k_B}(hT))$, and still considerable for $L^i(\hat{\lambda}_{A,\mathrm{QS}})$ and $L^i(\hat{\lambda}_{A,B})$.

The discriminating power of $L^i(\hat{\lambda})$ in highly autocorrelated series must hence be considered poor compared to what can be achieved. In other words, $L^i(\hat{\lambda})$ contains much less information about the integration or stationarity of the series than is available. In a ranking of the long-run estimators considered here, tests of stationarity constructed with the automatic bandwidth selection procedures suggested by Andrews (1991) do relatively best. Their awkward dependence on the correlation structure of $\{v_t\}$ via $\gamma(0)/\omega^2$ could be avoided by either running the AR(p) regression (8) instead of (9) and by using $\hat{\rho}_\mathrm{p} = \sum_{i=1}^{p}\hat{a}_i$ in place of $\hat{\rho}_w$, or by a correction in the spirit of Phillips and Perron (1988). Nevertheless, if the stationarity of a strongly autocorrelated series is in doubt, it seems not advisable to base

[3]The depicted local-to-unity analysis does not necessarily imply that the efficient stationarity tests control size under standard asymptotics, whereas the critical values of $L^i(\hat{\lambda})$ are chosen for this purpose. But more detailed calculations in the working paper of this contribution reveal that the efficient tests are in fact undersized under standard asymptotics.

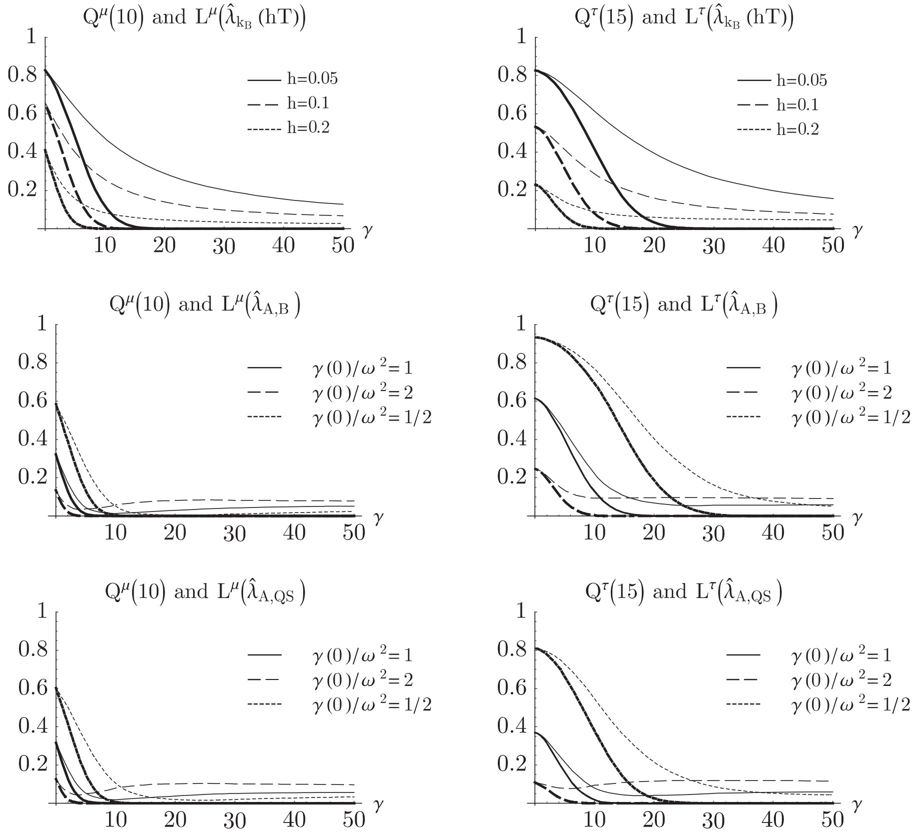

Fig. 2. Asymptotic rejection rates of efficient stationarity tests.

inference on KPSS-type tests. An application of optimal unit root test statistics yields far superior results.

4. Conclusions

This paper has analyzed the size and power properties of KPSS-type tests of stationarity in the presence of high autocorrelation in an asymptotic framework. The analysis reveals a strong dependence of the behavior of such tests on the estimator of the long-run variance, and the tests are shown to possess highly undesirable properties in such circumstances.

The undesirability of the behavior of tests of stationarity in highly autocorrelated time series comes in two forms. On the one hand, for many estimators of the long-run variance that are employed in practice, tests are bound to reject the null hypothesis of stationarity even if the true process is strongly mean-reverting for a high enough sample frequency. In finite samples, this leads at least to a very awkward dependence of the outcome of the tests on the sampling frequency of the observations, where a higher frequency increases the probability of a mistaken rejection. On the other hand, while preventing a degenerate behavior, other estimators of the long-run variance yield tests with an undesirable rejection profile. Not only are these tests inconsistent in the sense that they reject integrated series with probability far below one; their discriminating power between stationary and integrated series is also much inferior compared to optimal tests. While KPSS-type tests of stationarity are a valuable tool in other contexts, these properties cast strong doubts on their usefulness when applied to strongly persistent macroeconomic series.

One alternative is to use tests of stationarity that are based on optimal unit root statistics, but that reject for values of the statistic that indicate stationarity. The appeal of such a solution is limited, however, since the decision of such a test is a one-to-one mapping of the *p*-value of the corresponding optimal unit root test, so no additional information is gained by separately computing such a test of stationarity. This outcome also makes intuitive sense: If a statistic optimally summarizes the mean-reverting property of a time series, then both a hypothesis of mean reversion (stationarity) and a hypothesis of no mean reversion (integration) should be decided by this statistic. Following this reasoning further leads to the (almost) optimal confidence intervals for the mean reverting parameter, derived by Elliott and Stock (2001), as the best description of our knowledge about potential mean reversion.

The findings of the present paper might also have implications for higher-order systems. The multivariate analogue of tests of stationarity are cointegration tests with the null hypothesis of cointegration, and generalizations of KPSS statistic for such cases have been derived by Shin (1994) and Harris and Inder (1994), among others. If the stationary linear combination of the series is only slowly mean reverting, then these methods are likely to suffer from drawbacks similar to those found here for univariate tests of stationarity.

Acknowledgements

The author thanks Graham Elliott, Matthias Lutz and four anonymous referees for very helpful comments.

Appendix A

A.1. Proof of Proposition 1

From $\mathrm{P}(L^i(\hat{\lambda}_k(B_T)) > \mathrm{cv}) = \mathrm{P}(T^{-4}\sum_{t=1}^{T}(S_t^i)^2 > \mathrm{cv}\,T^{-2}\hat{\lambda}_k(B_T))$ and (6) it suffices to show that $T^{-2}\hat{\lambda}_k(B_T) \xrightarrow{p} 0$. Now with $T^{-1}\hat{\eta}(j) \leqslant (\sup_t T^{-1/2}|y_t^i|)^2$ for all j we have for all $T > T^{1/2}B_T^{1/2}$

$$T^{-2}|\hat{\lambda}_k(B_T)| \leqslant 2(\sup_t T^{-1/2}|y_t^i|)^2 T^{-1}\left[\sum_{j=0}^{T^{1/2}B_T^{1/2}}\left|k\left(\frac{j}{B_T}\right)\right| + \sum_{j=T^{1/2}B_T^{1/2}+1}^{T}\left|k\left(\frac{j}{B_T}\right)\right|\right].$$

From Condition 1 and the CMT, $(\sup_t T^{-1/2}|y_t^i|)^2 \Rightarrow \omega^2(\sup_s|M^i(s)|)^2$, so that $(\sup_t T^{-1/2}|y_t^i|)^2 = \mathrm{O_p}(1)$. Furthermore, since $|k(s)| \leqslant 1$ for all s, $T^{-1}\sum_{j=0}^{T^{1/2}B_T^{1/2}}|k(\frac{j}{B_T})| \leqslant T^{-1/2}B_T^{1/2} \xrightarrow{p} 0$, and from $\lim_{s\to\infty}k(s)=0$, $T^{-1}\sum_{j=T^{1/2}B_T^{1/2}+1}^{T}|k(\frac{j}{B_T})| \xrightarrow{p} 0$, such that $T^{-2}|\hat{\lambda}_k(B_T)| \xrightarrow{p} 0$, completing the proof.

A.2. Proof of Proposition 2

(i) The result is proved along the same lines as part (iv) below and is omitted.

(ii) Part (ii) is an implication of the results of Stock (1991). We concentrate on the time trend case, the reasoning for the mean case is analogous. Consider the least-squares regression

$$\begin{aligned} u_t &= c + \delta t + b_1 u_{t-1} + b_2\,\Delta u_{t-1} + \cdots + b_p\,\Delta u_{t-p-1} + \varepsilon_t \\ &= X_t'\beta + \varepsilon_t \end{aligned} \tag{14}$$

with $X_t = (1, t, u_{t-1}, \Delta u_{t-1}, \ldots, \Delta u_{t-p+1})'$ and $\beta = (c, \delta, b_1, \ldots, b_p)'$.

Let $\Lambda = diag(T^{1/2}, T^{3/2}, T, T^{1/2}I_{p-1})$. Then Stock's results imply that $\Lambda(\hat{\beta}-\beta)$ has a nondegenerate asymptotic distribution,

$$T(\hat{b}_1 - 1) \Rightarrow \frac{\sigma}{\omega}\left[\frac{\int M^i(s)\,\mathrm{d}W(s)}{\int M^i(s)^2\,\mathrm{d}s} - \gamma\right]$$

and the standard error of regression (14) converges to σ in probability. We cannot directly apply these results, however, because regression (8) does not contain a mean and a time trend, but rather has y_t^τ in place of u_t.

Now clearly a substitution of u_t by y_t^τ in (14) does not alter the estimated coefficient vector $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_p)'$, and $T^{-3/2}\sum_{t=p+1}^{T} y_{t-j}^\tau(1, T^{-1}t)' \xrightarrow{p} 0$ for $j = 0, 1$ and $T^{-1}\sum_{t=p+1}^{T}\Delta y_{t-j}^\tau(1, T^{-1}t)' \xrightarrow{p} 0$ for $j = 1, \ldots, p-1$, so that in the appropriate scaling, both the regressors and the explained variable are asymptotically orthogonal to the mean and the time trend. It follows that the coefficient vector $\hat{b}^* = (\hat{b}_1^*, \ldots, \hat{b}_p^*)'$ of the short least squares regression

$$y_t^\tau = b_1^* y_{t-1}^\tau + b_2^*\,\Delta y_{t-1}^\tau + \cdots + b_p^*\,\Delta y_{t-p-1}^\tau + \varepsilon_t^* \tag{15}$$

satisfies $diag(T, T^{1/2}I_{p-1})(\hat{b}^* - \hat{b}) \xrightarrow{p} 0$, and the standard error of regression (15) converges to σ in probability, too.

But the regressors in regression (8) in the trend case are a linear transformation of the regressors in (15). By standard linear regression algebra $1 - \sum_{j=1}^{p}\hat{a}_j = 1 - \hat{b}_1^*$, so that

$$T^{-2}\hat{\lambda}_{\mathrm{AR}} \Rightarrow \omega^2 \left| \gamma - \frac{\int M^i(s)\,\mathrm{d}W(s)}{\int M^i(s)^2\,\mathrm{d}s} \right|^{-2} = \omega^2 \left[\frac{2\int M^i(s)^2\,\mathrm{d}s}{M^i(1)^2 - M^i(0)^2 - 1} \right]^2.$$

The result now follows from another application of the CMT.

(iii) We find for the autoregressive estimator in the 'whitening' regression (9)

$$\begin{aligned} T(1-\hat{\rho}_w) &= \frac{T^{-1}(y_T^i)^2 - T^{-1}(y_1^i)^2 - T^{-1}\sum_{t=2}^T (\Delta y_t^i)^2}{2T^{-2}\sum_{t=2}^T (y_{t-1}^i)^2} \\ &\Rightarrow \frac{\omega^2 M^i(1)^2 - \omega^2 M^i(0)^2 - \gamma(0)}{2\omega^2 \int M^i(l)^2\,\mathrm{d}l}, \end{aligned}$$

where the last line follows from $\sup_t|\Delta y_t^i - v_t| = \mathrm{O_p}(T^{-1/2})$, the law of large numbers and the CMT. It hence remains to show that the estimator $\hat{\omega}_e^2$ of the long-run variance of $e_{w,t}$, is consistent for ω^2. We find for the estimated residuals $\hat{e}_{w,t}$

$$\begin{aligned} \hat{e}_{w,t} &= y_t^i - \hat{\rho}_w y_{t-1}^i \\ &= v_t + (y_t^i - u_t + u_0) - \rho_T(y_{t-1}^i - u_{t-1} + u_0) - (1-\rho_T)u_0 + (\rho_T - \hat{\rho}_w)y_{t-1}^i \\ &\equiv v_t + \xi_t^i. \end{aligned}$$

Let $V^\mu(s) = \gamma\int M(l)\,\mathrm{d}l$ and $V^\tau(s) = 4\gamma\int M(l)\,\mathrm{d}l - 6\gamma\int lM(l)\,\mathrm{d}l - (6\int M(l)\,\mathrm{d}l - 12\int lM(l)\,\mathrm{d}l)(\gamma s + 1)$. Then from a direct calculation and the CMT

$$T^{1/2}\omega^{-1}\xi_{[Ts]}^i \Rightarrow \left[\frac{M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2}}{2\int M^i(l)^2\,\mathrm{d}l} - \gamma \right] M^i(s) - (2\gamma)^{1/2}\zeta/2 - V^i(s),$$

so that $\Xi_T \equiv T^{1/2}\sup_t|\xi_t^i| = \mathrm{O_p}(1)$. Under the conditions of part (iii), Proposition 1 of Andrews (1991) implies that $T^{-1}\sum_{t=1}^T v_t^2 + 2\sum_{j=1}^T k(\frac{j}{b_T})T^{-1}\sum_{t=1}^{T-j} v_t v_{t+j} \xrightarrow{\mathrm{p}} \omega^2$. But

$$\begin{aligned} &\left| \sum_{j=0}^T k\left(\frac{j}{b_T}\right) T^{-1} \sum_{t=1}^{T-j} (v_t v_{t+j} - \hat{e}_{w,t}\hat{e}_{w,t+j}) \right| \\ &\quad \leqslant \sum_{j=0}^T \left| k\left(\frac{j}{b_T}\right) \right| T^{-3/2}\Xi_T \sum_{t=1}^{T-j} (T^{-1/2}\Xi_T + |v_{t+j}| + |v_t|) \\ &\quad \leqslant \left[T^{-1/2}\Xi_T^2 + 2\Xi_T T^{-1}\sum_{t=1}^T |v_t| \right] T^{-1/2} \sum_{j=0}^T \left| k\left(\frac{j}{b_T}\right) \right|. \end{aligned}$$

Since $\mathrm{E}[v_t^2] = \gamma(0)$, $\mathrm{E}[|v_t|] \leqslant \gamma(0)^{1/2}$ by Jensen's inequality, so that $T^{-1}\mathrm{E}\left[\sum_{t=1}^T |v_t|\right] \leqslant \gamma(0)^{1/2}$ and $T^{-1}\sum_{t=1}^T |v_t| = \mathrm{O_p}(1)$ by Markov's inequality. Note that $\int_0^\infty |k(s)|\,\mathrm{d}s < \infty$ implies $b_T^{-1}\sum_{j=0}^\infty |k(\frac{j}{b_T})| = \mathrm{O_p}(1)$, so that $T^{-1/2}\sum_{j=0}^T |k(\frac{j}{b_T})| \xrightarrow{\mathrm{p}} 0$ from $b_T = \mathrm{o_p}(T^{1/2})$. The result now follows from the CMT.

(iv) Standard arguments show that the mapping $f \mapsto 2 \int k(\frac{s}{B(f)}) \int_0^{1-s} f(l) f(l+s) \, dl \, ds$ with $B(f) = b_0 \left| \frac{2 \int f(s)^2 \, ds}{f(1)^2 - f(0)^2 - \gamma(0)\omega^{-2}} \right|^{b_1}$ for $b_0, b_1 > 0$ from the set of continuous functions on the unit interval to the reals has a discontinuity set of measure zero with respect to the sup norm when applied to $M(\cdot)$, so that the result follows from the CMT.

References

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D., Monahan, J., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–966.

Caner, M., Kilian, L., 2001. Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. Journal of International Money and Finance 20, 639–657.

Chan, N., Wei, C., 1987. Asymptotic inference for nearly nonstationary AR(1) Processes. The Annals of Statistics 15, 1050–1063.

Choi, I., 1994. Residual based tests for the null of stationarity with applications to US macroeconomic time series. Econometric Theory 10, 720–746.

Dufour, J.-M., King, M., 1991. Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. Journal of Econometrics 47, 115–143.

Elliott, G., 1999. Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. International Economic Review 40, 767–783.

Elliott, G., Stock, J., 2001. Confidence intervals for autoregressive coefficients near one. Journal of Econometrics 103, 155–181.

Elliott, G., Rothenberg, T., Stock, J., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.

Engel, C., 2000. Long-run PPP may not hold after all. Journal of International Economics 51, 243–273.

Harris, D., Inder, B., 1994. A test for the null of cointegration. In: Hargreaves, C. (Ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, Oxford, pp. 133–152.

Hobijn, B., Franses, P. Ooms, M., 1998. Generalizations of the KPSS-test for stationarity. Discussion Paper 9802, Econometric Institute, Erasmus University Rotterdam.

Kiefer, N., Vogelsang, T., 2002. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory 18, 1350–1366.

Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54, 159–178.

Lee, J., 1996. On the power of stationarity tests using optimal bandwidth estimates. Economics Letters 51, 131–137.

Leybourne, S., McCabe, B., 1994. A consistent test for a unit root. Journal of Business and Economic Statistics 12, 157–166.

Leybourne, S., McCabe, B., 1999. Modified stationarity tests with data dependent model selection rules. Journal of Business and Economic Statistics 17, 264–270.

MacNeill, I., 1978. Properties of sequences of partial sums of polynomial regression residuals with applications to test for change of regression at unknown times. Annals of Statistics 6, 422–433.

Müller, U., Elliott, G., 2001. Tests for unit roots and the initial observation. UCSD Working Paper 2001-19.

Müller, U., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269–1286.

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

Nabeya, S., Tanaka, K., 1990. A general approach to the limiting distribution for estimators in time series regression with nonstable autoregressive errors. Econometrica 58, 145–163.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.

Perron, P., Vodounou, C., 2001. Asymptotic approximations in the near-integrated model with a non-zero initial condition. Econometrics Journal 4, 143–169.

Phillips, P., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.

Phillips, P., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Rogoff, K., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34, 647–668.

Shin, Y., 1994. A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.

Stock, J., 1991. Confidence intervals for the largest autoregressive root in US macroeconomic time series. Journal of Monetary Economics 28, 435–459.

Stock, J., 1994. Unit roots, structural breaks and trends. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, New York, pp. 2740–2841.

Stock, J., Watson, M., 1998. Business cycle fluctuations in US macroeconomic time series. NBER Working Paper 6528.