Universität St.Gallen

Size and Power of Tests for Stationarity in Highly Autocorrelated Time Series

Ulrich K. Müller

November 2002 Discussion paper no. 2002-26

Department of Economics University of St. Gallen

Editor: Prof. Jörg Baumberger
University of St. Gallen
Department of Economics
Bodanstr. 1
CH-9000 St. Gallen
Phone ++41 71 224 22 41
Fax ++41 71 224 28 85
Email joerg.baumberger@unisg.ch
Publisher: Forschungsgemeinschaft für Nationalökonomie
an der Universität St. Gallen
Dufourstrasse 48
CH-9000 St. Gallen
Phone ++41 71 224 23 00
Fax ++41 71 224 26 46
Electronic Publication: www.fgn.unisg.ch/public/public.htm

Size and Power of Tests for Stationarity in Highly Autocorrelated Time Series[1]

Ulrich K. Müller

Author's address: Dr. Ulrich K. Müller
SIAW
Dufourstr. 48
9000 St. Gallen
Tel. ++41 71 224 23 41
Fax ++41 71 224 22 98
Email ulrich.mueller@unisg.ch

[1] The author thanks Graham Elliott and Matthias Lutz for very helpful comments on an earlier version of this paper. All errors are mine.

Abstract

Tests for stationarity are routinely applied to highly persistent time series. Following Kwiatkowski, Phillips, Schmidt and Shin (1992), standard stationarity employs 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 for 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.

Keywords

Tests for stationarity, local-to-unity asymptotics, long-run variance estimation, mean reversion

JEL Classification

C12; C22

1. Introduction

Discrimination between stationary and integrated series is a cornerstone of current time series analysis. Popular procedures for this purpose come in two distinct forms: On the one hand, there is a huge literature on 'tests for a unit root' with the null hypothesis of integration and the alternative hypothesis of stationarity. On the other hand, attempts have been made to derive tests that start from the null hypothesis of stationarity and that reject for integrated series — so-called 'tests for stationarity'. The only difference between unit root tests and tests for stationarity is the role of the null and alternative hypothesis. At first sight, it might hence seem surprising that the suggested test statistics differ between the two classes of tests: If, say, a certain unit root test statistic discriminates well between integrated and stationary series, then surely it must also discriminate well between stationary and integrated series.

The two types of tests rely on different test statistics because they are motivated by very different forms of 'stationary' and 'integrated' series. Tests for stationarity were originally derived to test the null hypothesis of a constant mean in a sample of independent Gaussian random variables. In fact, Nyblom (1989) has proved the local optimality of the test statistics against changes in the mean that form a martingale. Under the null hypothesis the mean is constant, and the series is Gaussian white noise. Under the local alternative the process is a sum of a small integrated component and Gaussian white noise. Loosely speaking, these statistics direct their power at testing whether a strongly mean reverting series reverts to either a constant mean (null hypothesis) or a slowly varying mean (alternative hypothesis).

In contrast, tests for unit roots assume a 'pure' integrated series under the null hypothesis, and consider highly autocorrelated, yet stationary series under the alternative. Efficient unit root tests (cf. Dufour and King (1991) and Elliott, Rothenberg and Stock (1996)) direct their power at detecting whether there is mean reversion in the series, not at checking whether the series reverts to a constant mean. Tests for stationarity and tests for unit roots are hence naturally suited for very different circumstances: While the former require strong mean reversion and try to detect a slowly varying component, the latter question the mean reversion itself.

Despite this theoretical background, tests for stationarity are routinely applied to what are at best slowly mean reverting series. Researchers justify the applicability of tests for stationarity to such series by referring to a correction suggested by Kwiatkowski, Phillips, Schmidt and Shin (1992), abbreviated KPSS in the following. The idea of KPSS is to account for the strong autocorrelation by dividing the 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 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. 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, Franses and Ooms (1998) corroborate this picture.

This paper develops a deeper understanding of the issues involved by analyzing size and power of tests for stationarity under local-to-unity asymptotics. The idea of analyzing the distribution of tests for 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), provides the asymptotic representation of processes that arise in unit root testing, i.e. integrated and slowly mean reverting series. From the above discussion, one would not expect tests for stationarity to do particularly well for such series, even when they are rescaled by $\hat{\lambda}$. But local-to-unity asymptotics 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 a constant between zero and 30, where T is the sample size. Stock and Watson (1998) estimate values for $T(1-\rho)$ in the region of 3 to 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 three to five years (cf. Rogoff (1996)), implying a $T(1-\rho)$ in the region of 14 to 23 for 100 years of data. An analysis in a local-to-unity framework reveals the behavior of tests for 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 for stationarity crucially hinges upon the estimator of the long-run variance in local-to-unity asymptotics. There are three 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 for stationarity that reject even highly mean reverting series with probability one for a large enough sample size. Second, if the true long-run estimator was known and used in the test statistic, then the rejection rate of the tests is below the level and *increases* in the rate of mean reversion. A desired rejection profile of more rejections for less mean reversion must therefore stem from inaccurate estimators of λ. Third, for some estimators of λ that employ a bandwidth of the same order as the sample size, the resulting tests for 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 for stationarity to reliably discriminate between the two alternatives under local-to-unity asymptotics hence does not come as a surprise, and is *per se* no compelling argument against their usage. The question is, however, how inefficient tests for stationarity are in a local-to-unity framework compared to optimal statistics. If there are tests that reject an integrated process as often as the test for stationarity, but that have much lower rates of rejections for strongly autocorrelated but stationary series, then the appropriateness of tests for stationarity for highly autocorrelated series is put in doubt.

There is no need to develop a new theory of efficient tests for stationarity to carry out this comparison. Efficient unit root test statistics, pioneered by Elliott et al. (1996) in the local-to-unity framework and further studied by Elliott (1999) and Müller and Elliott (2001), are derived as point-optimal tests that optimally discriminate between a given rate of mean reversion and no mean reversion. Usually, of course, these test statistics are used in hypothesis tests with the null hypothesis of integration. But the optimality of the discrimination of these statistics also holds when the hypotheses are reversed, making them ideally suited to assess the relative merits of the rejection profile of tests for stationarity. It turns out that the optimal unit root test statistics have much more discriminating power than the tests for stationarity. In this sense, the properties of most tests for 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 for stationarity for various estimators of the long-run variance. Section 3 compares the performance of the tests for stationarity which employ the most promising estimators of the long-run variance with tests based on optimal unit root test statistics. Section 4 concludes. Proofs are collected in an appendix.

2. Tests for Stationarity under Local to Unity Asymptotics

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

$$\begin{aligned} y_t &= d_t + w_t + \psi_t \qquad (1) \\ w_t &= w_{t-1} + v_t \end{aligned}$$

where y_t, $t = 1, \cdots, T$, is the observed sample, d_t is a deterministic component and $\{\psi_t\}$ and $\{v_t\}$ are independent stationary series. Under the null hypothesis of stationarity, the variance of v_t is zero, such that the disturbances $\psi_t + w_0$ are stationary. Under the alternative hypothesis, $E[v_t^2] > 0$, 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 (ψ_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

$$S_t^i = \sum_{s=1}^{t} y_s^i. \tag{2}$$

The test statistic is then given by

$$L^i(\hat{\lambda}) = \frac{T^{-2} \sum_{t=1}^{T} (S_t^i)^2}{\hat{\lambda}} \tag{3}$$

where $\hat{\lambda}$ is an estimator of the long-run variance of ψ_t, $\lambda = \sum_{j=-\infty}^{\infty} E[\psi_t \psi_{t-j}]$,[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 of ψ_t is given by

$$L^i(\hat{\lambda}) \Rightarrow \int W^i(s)^2 ds \tag{4}$$

where '$\Rightarrow$' denotes weak convergence as $T \to \infty$, $W(s)$ is a Wiener process, $W^\mu(s) = W(s) - sW(1)$, $W^\tau(s) = W(s) - (2s - 3s^2)W(1) + 6(s^2 - s)\int W(l)dl$ and for notational simplicity, the limits of integration are understood to be zero and one, if not indicated otherwise. The asymptotic critical values of $L^i(\hat{\lambda})$ can be calculated from these expressions and are given by KPSS for the 5% level by 0.463 and 0.146 in the mean and mean and time trend case, respectively.

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

$$\begin{aligned} y_t &= d_t + u_t \\ u_t &= \rho u_{t-1} + \nu_t \end{aligned} \tag{5}$$

where, if $|\rho| < 1$, $u_0 = \sum_{s=0}^{\infty} \rho^s \nu_{-s}$, 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 (5), then the τ-version of $L^i(\hat{\lambda})$ is used; in this sense the deterministics are assumed to be correctly specified. Note that $L^i(\hat{\lambda})$ is then independent of the particular value of d_t in the Data Generating Process.

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 if $\rho = 1$. If $|\rho| < 1$, the assumption of the generation of u_0 leads to a stationary series $\{u_t\}$ as long

[1] KPSS define the long-run variance λ (which is σ^2 in their notation) as $\lim_{T\to\infty} T^{-1}E[S_T^2]$ (p. 164). S_T 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 ψ_t, which is the definition employed in this paper.

as $\{\nu_t\}$ is stationary. While somewhat natural, this assumption might considerably affect the asymptotic distributions derived below. See Müller and Elliott (2001) for a detailed discussion.

The innovations $\{\nu_t\}$ 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 $\{\nu_t\}$ is covariance-stationary with finite autocovariances $\gamma(k) = E[\nu_t \nu_{t-k}]$ such that*
(a) $\omega^2 = \sum_{k=-\infty}^{\infty} \gamma(k)$ is finite and nonzero
(b) the scaled partial-sum process $T^{-1/2} \sum_{t=1}^{[sT]} \nu_t \Rightarrow \omega W(s)$.

In contrast to the reasoning of KPSS, the following derivations employ local-to-unity asymptotics, i.e. ρ in (5) is made a function of the sample size such that $\rho = \rho_T = 1 - \gamma T^{-1}$, where $\gamma \geq 0$ is a fixed number. Lemma 2 in Elliott (1999) shows that under Condition 1, the process u_t can then be asymptotically characterized by

$$\begin{aligned} T^{-1/2}(u_{[Ts]} - u_0) &\Rightarrow \begin{cases} \omega W(s) \text{ for } \gamma = 0 \\ \omega \zeta (e^{-\gamma s} - 1)(2\gamma)^{-1/2} + \omega \int_0^s e^{-\gamma(s-l)} dW(l) \text{ else} \end{cases} \\ &= \omega M(s) \end{aligned} \tag{6}$$

where ζ is a standard normal variable independent of $W(\cdot)$.

The relationship between (5) and the Data Generating Process (1) assumed by KPSS is straightforward: For $|\rho| < 1$, (5) is a special case of (1) under the null hypothesis of $E[v_t^2] = 0$ with $\psi_t = u_t$, and for $\rho = 1$ (5) is a special case of (1) with $v_t = \nu_t$ and $w_t = 0$. 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 a fixed $|\rho| < 1$ in (5). As we shall see, $L^i(\hat{\lambda})$ 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 the largest autoregressive root of the sample is such that $T(1 - \rho)$ is smaller than, say, 30 than standard ($|\rho| < 1$fixed) asymptotics. The following results may hence be interpreted as more useful small sample descriptions of the behavior of tests for stationarity when applied to highly autocorrelated series.

The local-to-unity process $M(s)$ with $\gamma > 0$ is an asymptotic representation of a series that slowly reverts to a constant mean. As long as $\gamma > 0$ we thus analyze the behavior of tests for stationarity under the null hypothesis of stationarity. It is possible to extend the analysis to asymptotics representing highly autocorrelated series reverting to a varying mean by including an additional unit root process $\{w_t\}$ in (5), where $\{w_t\}$ is defined just as in (1). Many results derived in the following then continue to hold with $M(\cdot)$ replaced by $M(\cdot) + \kappa W_w(\cdot)$, where $W_w(\cdot)$

is a Wiener process independent of $W(\cdot)$ and κ is the ratio of the long-run variances of $\{\upsilon_t\}$ and $\{\nu_t\}$. But prior to considering asymptotic power of tests for stationarity in such a manner, it is arguably more important to analyze their size control in such a framework. We hence stick to the simpler formulation (5), where the alternative of nonstationarity only arises when $\rho = 1$ (which corresponds to $\gamma = 0$).

From (6), straightforward calculations reveal (cf., for instance, Stock (1994), p. 2772) that the residuals y_t^i satisfy

$$T^{-1/2} y^i_{[Ts]} \Rightarrow \omega M^i(s) \tag{7}$$

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

$$T^{-4}\sum_{t=1}^{T}(S_t^i)^2 \Rightarrow \omega^2 \int \left[\int_0^s M^i(l)dl\right]^2 ds \tag{8}$$

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 $O_p(T^2)$.

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

$$\hat{\eta}(j) = T^{-1}\sum_{t=1}^{T-j} y_t^i y_{t+j}^i \tag{9}$$

and define

$$\hat{\lambda}_k(B_T) = \hat{\eta}(0) + 2\sum_{j=1}^{T} k(\frac{j}{B_T})\hat{\eta}(j). \tag{10}$$

The even and continuous function $k : [0, \infty) \to [-1; 1]$ serves as the weighting function of the sample autocovariances and is assumed to satisfy $k(0) = 1$, $\int_0^\infty |k(s)|ds < \infty$ 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 (10) 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 $k(x) = k_B(x) = 1 - |x|$ for $|x| < 1$, $k_B(x) = 0$ for $|x| \geq 1$ and $B_T = m + 1$.

In a standard asymptotic framework it can usually be shown that long-run variance estimators of the form (10) are consistent when $B_T = o(T^{1/2})$ or $B_T = o(T)$ — 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 $\gamma = T(1-\rho_T) \geq 0$, 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 realization of a continuous time process $M(s)$ 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 $o(T)$ does not reject the null hypothesis of stationarity when the frequency of the observations is, say, 1/100 (such that $T = 100$ and $y_1 = M(.01)$, $y_2 = M(.02)$, $\cdots$, $y_{100} = M(1)$). 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 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 for 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 for 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.

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 $\rho = .95$ in the mean case and 38.0% in the trend case (Caner and Kilian (2001), Table 1).

As a next step in the analysis, we consider the behavior of $L^i(\hat{\lambda})$ when $\hat{\lambda}$ is replaced by the true value λ. This is a purely theoretical exercise, since λ is finite only if $|\rho| < 1$, so that the knowledge of λ allows to infer whether $\rho = 1$ or $|\rho| < 1$. But it is still interesting to disentangle the effect of estimation inaccuracy of $\hat{\lambda}$ from the overall behavior of the test statistic.

Proposition 2. *Under Condition 1, for any $\gamma = T(1-\rho_T) > 0$, $\lambda = \gamma^{-2}\omega^2 T^2$ and*

$$L^i(\lambda) \Rightarrow \gamma^2 \int \left[\int_0^s M^i(l)dl \right]^2 ds.$$

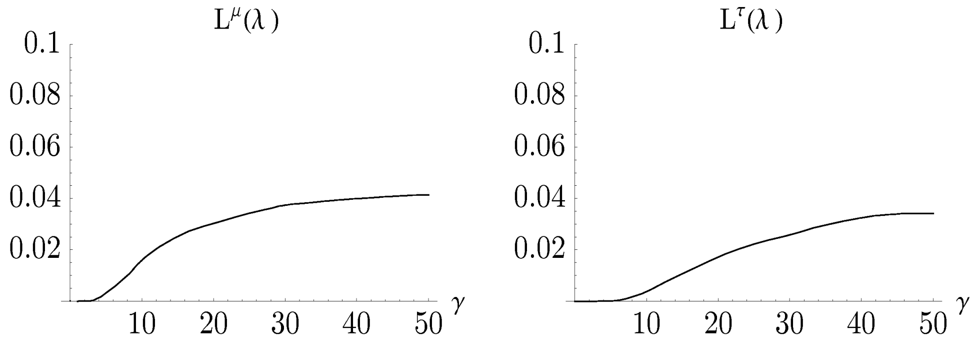

FIGURE 1. Asymptotic rejection rates for true long-run variance

Figure 1 depicts the asymptotic rejection rates of $L^i(\lambda)$ as a function of γ. Strikingly, while keeping below the nominal level of 5%, the rejection rates *increase* in γ. This runs, of course, counter to the expected behavior of a test for stationarity to reject more often for less mean reverting series.

The positively sloped rejection profiles in Figure 1 are the result of two countervailing influences. On the one hand, $\int \left[\int_0^s M^i(l)dl\right]^2 ds$ decreases in γ on average, since less mean reverting $M(s)$ lead to larger deviations of $M^i(s)$ from zero. On the other hand, γ^2 obviously increases in γ, and this second effect dominates the first. If tests based on $L^i(\hat{\lambda})$ are to have a rejection profile that decreases in γ, then it must be the result of estimation error in $\hat{\lambda}$. Intuitively, the ratio of $\hat{\lambda}/\lambda$ must be increasing in γ in order to produce larger values of $L^i(\hat{\lambda})$ for smaller γ.

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. 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 $\{\nu_t\}$ by an autoregressive parametrization. $\hat{\lambda}_{AR}$ is computed by running the ordinary least squares regression

$$y_t^i = a_1 y_{t-1}^i + a_2 y_{t-2}^i + \cdots + a_p y_{t-p}^i + e_t \tag{11}$$

followed by the computation of

$$\hat{\lambda}_{AR} = \frac{\hat{\sigma}_e^2}{(1 - \sum_{i=1}^p \hat{a}_i)^2} \tag{12}$$

where $\hat{a}_i$ and $\hat{\sigma}_e^2$ are the estimated parameters in (11).

A third class of estimators first 'prewhitens' the data by a low order autoregression just like (11) 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.

$$y_t^i = \rho_w y_{t-1}^i + e_{w,t} \tag{13}$$

and the spectral density estimator $\hat{\omega}_e^2$ of the residuals $\hat{e}_{w,t}$ is constructed in analogy to (10) with a bandwidth $b_T = o(T^{1/2})$. The long-run variance estimator is then given by

$$\hat{\lambda}_{PW} = (1 - \hat{\rho}_w)^{-2} \hat{\omega}_e^2. \tag{14}$$

Finally, we consider spectral density estimators (10) 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 (13). 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

$$B_{B,T} = 1.1447 \left[\frac{4\hat{\rho}_w^2}{(1-\hat{\rho}_w)^2(1+\hat{\rho}_w)^2} T \right]^{1/3} \tag{15}$$

$$B_{QS,T} = 1.3221 \left[\frac{4\hat{\rho}_w^2}{(1-\hat{\rho}_w)^4} T \right]^{1/5} \tag{16}$$

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

Proposition 3. *For any $\gamma = T(1 - \rho_T) \geq 0$,*
(i) under Condition 1

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

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

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

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

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

[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 as long ν_t is uncorrelated), because of 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. Alternative solutions are to trim the estimates as suggested in Andrews (1991) and Andrews and Monahan (1992) away from zero. 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 but available from the author on request 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 local power for $L^\mu(\hat{\lambda}_{AR})$ and $L^\mu(\hat{\lambda}_{PW})$ and leads to very few rejections of $L^\mu(\hat{\lambda}_{A,B})$ and $L^\mu(\hat{\lambda}_{A,QS})$.

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

$$L^i(\hat{\lambda}_{A,j}) \Rightarrow \frac{\int \left[\int_0^s M^i(l)dl\right]^2 ds}{2\int k_j(\frac{l}{B_j})\int_0^{1-s} M^i(l)M^i(s+l)dlds}$$

where $B_{A,QS} = 1.7445 \left| \frac{2\int M^i(s)^2 ds}{M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2}} \right|^{4/5}$ *and*
$B_{A,B} = 1.1447 \left| \frac{2\int M^i(s)^2 ds}{M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2}} \right|^{2/3}$.

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 $\{\nu_t\}$ — see Andrews (1991) for details. The condition on $\{\nu_t\}$ in part (ii) is such that the proof can rely on the reasoning of Stock (1991).

The asymptotic distributions of $L^i(\hat{\lambda}_{PW})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,QS})$ of Proposition 3 depend on the ratio $\gamma(0)/\omega^2$, whereas this is not the case for $L^i(\hat{\lambda}_k(hT))$ and $L^i(\hat{\lambda}_{AR})$. 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 $\{\nu_t\}$. Furthermore, the asymptotic distribution of $L^i(\hat{\lambda}_{PW})$ and $L^i(\hat{\lambda}_{AR})$ are the same when the variance $\gamma(0)$ of ν_t and its long-run variance ω^2 coincide.

Figure 2 depicts the asymptotic rejection rates of $L^i(\hat{\lambda}_k(hT))$ with a Bartlett weighting $k = k_B$ and $h = 0.05$, 0.1 and 0.2, $L^i(\hat{\lambda}_{AR})$, $L^i(\hat{\lambda}_{PW})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,QS})$ for the 5% nominal level as a function of γ. For $L^i(\hat{\lambda}_{PW})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,QS})$, rejection rate are investigated when $\gamma(0)/\omega^2$ is set to 1/2, 1 and 2, respectively. Overall, the rejection rates are similar but slightly larger for the trend case than for the mean case for all considered $\hat{\lambda}$. These similarities aside, the asymptotic behavior of $L^i(\hat{\lambda})$ is very different for the different estimators $\hat{\lambda}$ considered. A test based on $L^i(\hat{\lambda}_{k_B}(hT))$ has a rejection profile that is monotonically decreasing in γ, and larger values of h lead to fewer rejections. As the rate of mean reversion, γ, increases, the rejection rates approach the nominal level. The behavior of $L^i(\hat{\lambda}_{AR})$ is very much comparable to the (infeasible) test based on $L^i(\lambda)$ as the rate of rejections remains consistently below the level and is increasing in γ. This increasing slope is also found for $L^i(\hat{\lambda}_{PW})$, 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 $L^i(\hat{\lambda}_{AR})$, and for $\gamma(0)/\omega^2 = 1/2$ the test practically ceases to reject. A somewhat reversed picture can be found for $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,QS})$. Here, the smaller value of $\gamma(0)/\omega^2$ leads to more rejections.

Part (i) of Proposition 3 implies that it is the ratio of the bandwidth B_T to the sample size T that determines the behavior of tests for 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 for 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

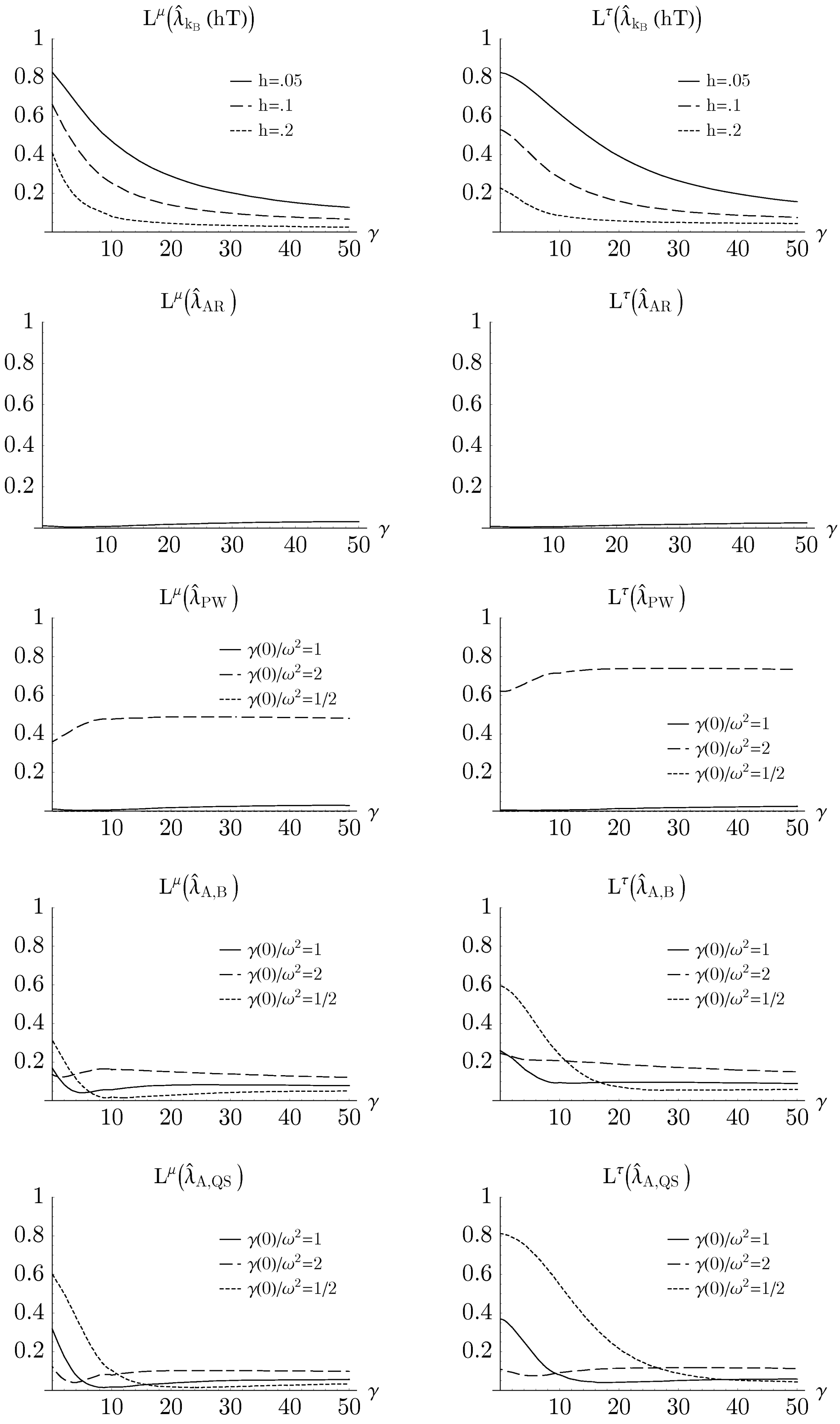

FIGURE 2. Asymptotic rejection rates for various long-run variance estimators

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 3 explains Lee's (1996) results from a Monte Carlo study, in which he find that the rejection rates of $L^i(\hat{\lambda}_{PW})$ 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 3 and Figure 2, 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 3 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 Hobijn et al. (1998) with respect to $L^i(\hat{\lambda}_{PW})$, $L^i(\hat{\lambda}_{A,B})$ and $L^i(\hat{\lambda}_{A,QS})$. Moreover, these authors argue by a different set of arguments that the parametric correction suggested by Leybourne and McCabe (1994), which is not easily analyzed in the framework considered here, suffers from the same drawback. At the same time, even for highly mean reverting series with $\gamma = 50$, only the size of $L^i(\hat{\lambda}_{k_B}(hT)$ with $h = 0.1$, 0.2 and $L^i(\hat{\lambda}_{AR})$ and, when $\gamma(0)/\omega^2 = 1$, $L^i(\hat{\lambda}_{A,QS})$ and $L^i(\hat{\lambda}_{PW})$, are close to the nominal level of 5%. Nevertheless, when compared to the asymptotic behavior of $L^i(\hat{\lambda}_k(B_T))$ with $B_T = o_p(T)$ 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 Unit Root Tests Statistics

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 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 few) rejections for any $\gamma > 0$ and rejection with probability one for $\gamma = 0$. This impossibility has led some authors to criticize the whole idea of trying to distinguish between the two models (cf. Blough (1992), Cochrane (1991)). Rather than taking this all-or-nothing view, this section tries to assess the relative merits of the asymptotic rejection profiles of $L^i(\hat{\lambda})$ compared to the optimal discrimination that can be achieved.

For this purpose, 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 (2001) 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. The statistics of Elliott et al. (1996) and Elliott (1999) differ in how they treat the initial disturbance u_0. Elliott (1999) assumes a distribution of u_0 such that $\{u_t\}$ becomes stationary if $|\rho| < 1$, which is also the assumption made here. Müller and Elliott (2001) give a full account of the importance of such an assumption for unit root testing. Usually, these optimal statistics are employed to perform a hypothesis test with integration as the null hypothesis. But the Neyman-Pearson Lemma is undirectional 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 a test for stationarity based on an optimal unit root test statistic hence maximizes the (positive) difference of rejection rates at $\gamma = 0$ and $\gamma = g$. While not achieving the ideal rejection profile either, such a test maximizes power at $\gamma = 0$ for a given size at $\gamma = g$, or, equivalently, minimizes mistaken rejections at $\gamma = g$ for a given power at $\gamma = 0$. An obvious question is how to choose g; but Elliott (1999) and Müller and Elliott (2001) find 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 good 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 (5) is, in the notation developed here, given by

$$Q^i(g) = q_1^i \left(\hat{\omega}^{-1} T^{-1/2} y_T^i\right)^2 + q_2^i \left(\hat{\omega}^{-1} T^{-1/2} y_1^i\right)^2 + q_3^i \left(\hat{\omega}^{-1} T^{-1/2} y_T^i\right)\left(\hat{\omega}^{-1} T^{-1/2} y_1^i\right) + q_4^i \hat{\omega}^{-2} T^{-2} \sum_{t=1}^{T} (y_t^i)^2 \quad (17)$$

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 ν_t under local-to-unity asymptotics. (The q_i^j 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 (13); see the proof of part (iii) of Proposition 3 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

$$Q^i(g) \Rightarrow q_1^i M^i(1)^2 + q_2^i M^i(0)^2 + q_3^i M^i(1) M^i(0) + q_4^i \int M^i(s)^2 ds. \quad (18)$$

If the statistics $Q^i(g)$ are to be used as tests for stationarity, a critical value has to be chosen such that the rejection rates under the null of stationarity do not exceed the nominal level. The following proposition establishes that any positive critical value possesses this feature under usual asymptotics with $|\rho| < 1$ fixed, provided $\hat{\omega}^2$ is bounded away from zero. Also see Theorem 3 of Elliott et al. (1996) for a similar argument.

Proposition 4. *If Condition 1 holds, $|\rho| < 1$ is fixed and $P(\hat{\omega}^2 > \delta) = 1$ for some $\delta > 0$, then $Q^i(g) \xrightarrow{p} 0$ as $T \to \infty$.*

One possible choice for a $\hat{\omega}^2$ which has the property invoked in Proposition 4 in a very general context is the standard autoregressive spectral density estimator; see Lemma 1 of Stock (2000).

In order to compare the asymptotic rejection rates of the more promising versions of $L^i(\hat{\lambda})$ with a test for stationarity based on $Q^i(g)$, the critical values in Figure 3 are chosen such that the rejection rates of $Q^i(g)$ coincide with the rejection rates of $L^i(\hat{\lambda})$ at $\gamma = 0$. All these critical values are positive, so that Proposition 4 implies that the tests based on $Q^i(g)$ are asymptotically undersized under standard asymptotics. 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 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 $L^i(\hat{\lambda}_{A,QS})$ 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 mean reversion of a series than is available. In a ranking of the long-run estimators considered here, tests for stationarity constructed with the automatic bandwidth selection procedures suggested by Andrews (1991) do relatively best. Their awkward dependence on the correlation structure of $\{\nu_t\}$ via $\gamma(0)/\omega^2$ could be avoided by either running the AR(p) regression (11) instead of (13) and by using $\hat{\rho}_p = \sum_{i=1}^p \hat{a}_i$ in place of $\hat{\rho}_w$, or by a correction in the spirit of Phillips and Perron (1988). Nevertheless, if the mean reversion of a series is in doubt, it seems not advisable to base inference on $L^i(\hat{\lambda})$. 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 for 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 for 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

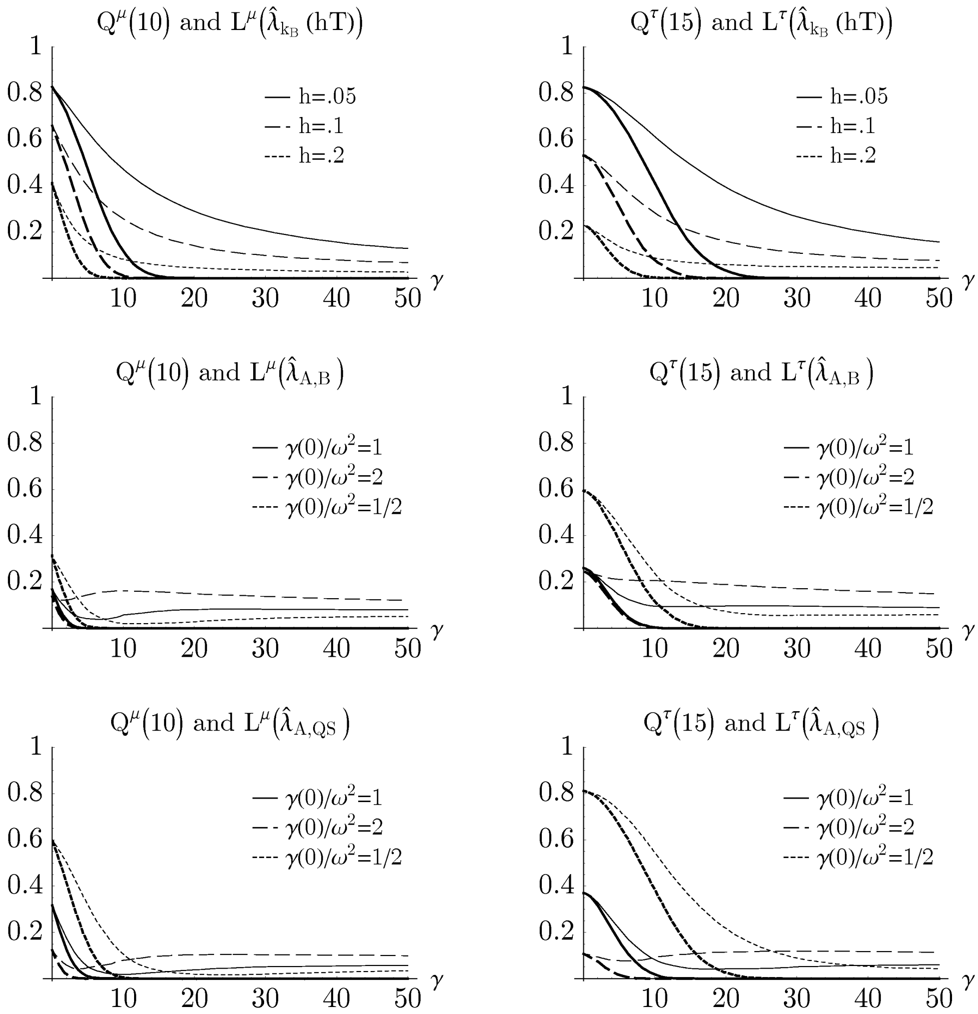

FIGURE 3. Asymptotic rejection rates of optimal tests for stationarity

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. These properties cast strong doubts on the usefulness of tests for stationarity for (at most) weakly mean reverting macroeconomic series.

One alternative is to use tests for 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 for 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 for 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 for stationarity.

APPENDIX

In the following proofs, all limits are taken as $T \to \infty$, if not indicated otherwise.

Proof of Proposition 1:
From $P(L^i(\hat{\lambda}_k(B_T)) > \text{cv}) = P(T^{-4}\sum_{t=1}^{T}(S_t^i)^2 > \text{cv}\, T^{-2}\hat{\lambda}_k(B_T))$ and (8) it clearly suffices to show that $T^{-2}\hat{\lambda}_k(B_T) \xrightarrow{p} 0$.

Now with $T^{-1}\hat{\eta}(k) \leq (\sup_t T^{-1/2}|y_t^i|)^2$ for all k we have

$$T^{-2}|\hat{\lambda}_k(B_T)| \leq 2(\sup_t T^{-1/2}|y_t^i|)^2 T^{-1}\sum_{j=0}^{T}|k(\frac{j}{B_T})|.$$

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 $(\sup_t T^{-1/2}|y_t^i|)^2 = O_p(1)$.

Furthermore, since $\lim_{s\to\infty} k(s) = 0$, for any $\varepsilon > 0$ there exists a N such that $|k(s)| < \varepsilon$ for all $s \geq N$. By assumption $B_T = o_p(T)$, so that the probability of the event $T^{-1}B_T \leq \varepsilon N^{-1}$ can be made arbitrarily close to one by choosing T large enough. Now in this event

$$\begin{aligned} T^{-1}\sum_{j=0}^{T}|k(\frac{j}{B_T})| &= T^{-1}\sum_{j=0}^{NB_T}|k(\frac{j}{B_T})| + T^{-1}\sum_{j=NB_T+1}^{T}|k(\frac{j}{B_T})| \\ &\leq 2\varepsilon \end{aligned}$$

But ε was chosen arbitrarily, which implies to $T^{-1}\sum_{j=0}^{T}|k(\frac{j}{B_T})| \xrightarrow{p} 0$ and hence $T^{-2}\hat{\lambda}_k(B_T) \xrightarrow{p} 0$, completing the proof.

Proof of Proposition 2:
Since ν_t is covariance-stationary, it has a spectral density function $f_\nu(\cdot)$, and $\omega^2 = 2\pi f_\nu(0)$. It follows that $u_t = (1-\rho_T L)^{-1}\nu_t$ is covariance-stationary, too, and has spectral density $f_u(\theta) = (1-\rho_T e^{-i\theta})^{-1}(1-\rho_T e^{i\theta})^{-1}f_\nu(\theta)$. Hence

$$\lambda = 2\pi f_u(0) = \frac{\omega^2}{(1-\rho_T)^2}$$

With $\rho_T = 1-\gamma T^{-1}$ we obtain $(1-\rho_T)^2 = \gamma^2 T^{-2}$, so that $\lambda = \gamma^{-2}\omega^2 T^2$. The weak convergence of $L^i(\lambda)$ now follows immediately from (8).

Proof of Proposition 3:

(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). The assumption made here concerning u_0 differs from Stock's, but that does not change anything substantial in the argument.

We concentrate on the time trend case, the reasoning for the mean case is analogous. Consider the least squares regression

$$\begin{aligned} u_t &= c + \delta t + b_1 u_{t-1} + b_2 \Delta u_{t-1} + \cdots + b_p \Delta u_{t-p-1} + \varepsilon_t \\ &= X_t' \beta + \varepsilon_t \end{aligned} \tag{19}$$

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

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 non-degenerate distribution,

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

and the standard error of regression (19) converges to σ in probability. We cannot directly apply these results, however, because regression (11) does not contain a mean and a time trend, but rather has y_t^τ in place of u_t.

Now a substitution of u_t by y_t^τ in (19) does not alter the estimated coefficient vector $\hat{b} = (\hat{b}_1, \cdots, \hat{b}_p)'$ (the Dickey-Fuller regression is invariant to such changes). Furthermore, we have $T^{-3/2} \sum_{t=p+1}^{T} y_{t-j}^\tau (1, T^{-1} t)' \xrightarrow{p} 0$ 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 $j = 1, \cdots, p-1$, 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 $\hat{b}^* = (\hat{b}_1^*, \cdots, \hat{b}_p^*)'$ of the short least squares regression

$$y_t^\tau = b_1^* y_{t-1}^\tau + b_2^* \Delta y_{t-1}^\tau + \cdots + b_p^* \Delta y_{t-p-1}^\tau + \varepsilon_t^* \tag{20}$$

satisfies $diag(T, T^{1/2} I_{p-1})(\hat{b}^* - \hat{b}) \xrightarrow{p} 0$, and the standard error of regression (20) converges to σ in probability, too.

But the regressors in regression (11) in the trend case are a linear transformation of the regressors in (20). By standard linear regression algebra $1 - \sum_{j=1}^{p} \hat{a}_j = 1 - \hat{b}_1^*$, so that

$$\begin{aligned} T^{-2} \hat{\lambda}_{AR} &\Rightarrow \omega^2 \left[\frac{\int M^i(s) dW(s)}{\int M^i(s)^2 ds} - \gamma \right]^{-2} \\ &= \omega^2 \left[\frac{2 \int M^i(s)^2 ds}{M^i(1)^2 - M^i(0)^2 - 1} \right]^2 \end{aligned}$$

The result now follows from another application of the CMT.

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

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

where the last line follows from $\Delta y_t^i = \Delta y_t + O_p(T^{-1/2}) = \nu_t + O_p(T^{-1/2})$ 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 $\hat{e}_{w,t}$

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

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

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

so that $\Xi \equiv T^{1/2}\sup_t |\xi_t| = O_p(1)$. Under the conditions of part (iii), Proposition 1 of Andrews (1991) implies that

$$\omega^2 - T^{-1}\sum_{t=1}^T \nu_t^2 - 2\sum_{j=1}^{b_T} k(\frac{j}{b_T})T^{-1}\sum_{t=1}^{T-j}\nu_t\nu_{t+j} \xrightarrow{p} 0.$$

But

$$\begin{aligned}
&\left|\sum_{j=0}^{b_T} k(\frac{j}{b_T})T^{-1}\sum_{t=1}^{T-j}(\nu_t\nu_{t+j} - \hat{e}_{w,t}\hat{e}_{w,t+j})\right| \\
&\leq \sum_{j=0}^{b_T}|k(\frac{j}{b_T})|T^{-3/2}\Xi\sum_{t=1}^{T-j}(T^{-1/2}\Xi + |\nu_{t+j}| + |\nu_t|) \\
&\leq \Xi^2 b_T T^{-1} + 2\Xi b_T T^{-3/2}\sum_{t=1}^T |\nu_t|.
\end{aligned}$$

Since $E[\nu_t^2] = \gamma(0)$, $E[|\nu_t|] < \gamma(0)^{1/2}$ by Jensen's inequality, so that $T^{-1}E\left[\sum_{t=1}^T |\nu_t|\right] < \gamma(0)^{1/2}$ and $T^{-1}\sum_{t=1}^T |\nu_t| = O_p(1)$ by Markov's inequality. With $b_T = o(T^{1/2})$ the estimator $\hat{\omega}_e^2$ of the long-run variance of $e_{w,t}$ hence converges to ω^2, and the result follows from the CMT.

(iv) The result follows from the CMT when we can show that $X(f) = 2\int k(\frac{s}{B(f)})\int_0^{1-s} f(l)f(l+s)dlds$ 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, $b_1 > 0$ is continuous in f. Since $M(\cdot)$ has almost surely continuous sample paths, it suffices to show continuity of X in the space $\mathcal{C}$ of all continuous functions on $[0,1]$ in the sup-norm. Let $Q_1(\cdot)$ and $Q_2(\cdot)$ be two elements of $\mathcal{C}$ with the property $\sup_s |Q_1(s) - Q_2(s)| < \varepsilon$. Then

$$\begin{aligned}
X(Q_1) &= \int\left[k(\frac{s}{B(Q_2)}) - \left(k(\frac{s}{B(Q_2)}) - k(\frac{s}{B(Q_1)})\right)\right] \\
&\quad \times \int_0^{1-s}\left[Q_2(l)Q_2(l+s) - (Q_2(l)Q_2(l+s) - Q_1(l)Q_1(l+s))\right]dlds \\
&= X(Q_2) - A_1 - A_2 + A_3
\end{aligned}$$

where

$$
\begin{aligned}
A_1 &= \int k(\frac{s}{B(Q_2)}) \int_0^{1-s} \left(Q_2(l)Q_2(l+s) - Q_1(l)Q_1(l+s)\right) dlds \\
A_2 &= \int \left(k(\frac{s}{B(Q_2)}) - k(\frac{s}{B(Q_1)}) \right) \int_0^{1-s} Q_2(l)Q_2(l+s) dlds \\
A_3 &= \int \left(k(\frac{s}{B(Q_2)}) - k(\frac{s}{B(Q_1)}) \right) \int_0^{1-s} \left(Q_2(l)Q_2(l+s) - Q_1(l)Q_1(l+s)\right) dlds
\end{aligned}
$$

We have to show that A_1, A_2 and A_3 converge to zero as $\varepsilon \to 0$. Now

$$
\begin{aligned}
|A_1| &\leq \int |k(\frac{s}{B(Q_2)})| \int_0^{1-s} \left(2\varepsilon \sup_r |Q_2(r)| + \varepsilon^2 \right) dlds \\
&\leq \varepsilon \left(2 \sup_r |Q_2(r)| + \varepsilon \right) \int_0^1 |k(\frac{s}{B(Q_2)})| ds
\end{aligned}
$$

$$
|A_2| \leq \sup_r Q_2(r)^2 \sup_s \left| k(\frac{s}{B(Q_2)}) - k(\frac{s}{B(Q_1)}) \right|
$$

and

$$
|A_3| \leq \varepsilon \left(2 \sup_r |Q_2(r)| + \varepsilon \right) \sup_s \left| k(\frac{s}{B(Q_2)}) - k(\frac{s}{B(Q_1)}) \right|.
$$

But $B(f)$ is continuous in f as long as $f(1)^2 - f(0)^2 + \gamma(0)\omega^{-2} \neq 0$, and $k(\cdot)$ is a continuous function, so that if $Q_j(1)^2 - Q_j(0)^2 + \gamma(0)\omega^{-2} \neq 0$ for $j = 1, 2$, $\sup_s \left| k(\frac{s}{B(Q_2)}) - k(\frac{s}{B(Q_1)}) \right|$ converges to zero as $\varepsilon \to 0$. As $\int |k(\frac{s}{B(Q_2)})| ds \leq 1$ and, with probability one, $\sup_r |M^i(r)| < \infty$ and $M^i(1)^2 - M^i(0)^2 - \gamma(0)\omega^{-2} \neq 0$, the CMT is applicable and yields the desired result.

Proof of Proposition 4:

From the covariance-stationarity of ν_t and the assumption on u_0, u_t is covariance-stationary, too, and $E[u_t^2] = \sigma_u^2 < \infty$. From the definition of y_t^i

$$
\begin{aligned}
y_t^\tau &= u_t + (c_1 + tT^{-1}c_2)T^{-1} \sum_{s=1}^T u_s + (c_3 + c_4 tT^{-1})T^{-2} \sum_{s=1}^T s u_s \\
y_t^\mu &= u_t - T^{-1} \sum_{s=1}^T u_s
\end{aligned}
$$

where $c_i = O(1)$ for $i = 1, 2, 3, 4$. With $E[u_t^2] = \sigma_u^2$ Jensen's inequality implies $E[|u_t|] \leq \sigma_u$, so that $T^{-1}E\left[\sum_{t=1}^T |u_t|\right] \leq \sigma_u$ and $T^{-1}\left|\sum_{t=1}^T u_t\right| \leq T^{-1}\sum_{t=1}^T |u_t| = O_p(1)$ by Markov's inequality. Similarly, $T^{-2}|\sum_{s=1}^T s u_s| \leq T^{-1}\sum_{s=1}^T |u_s| = O_p(1)$. Furthermore, $T^{-1}E\left[\sum_{t=1}^T u_t^2\right] = \sigma_u^2$, so that Markov's inequality implies $T^{-1}\sum_{t=1}^T u_t^2 = O_p(1)$. After some straightforward algebra these results imply $T^{-1/2}y_T^i \xrightarrow{p} 0$, $T^{-1/2}y_1^i \xrightarrow{p} 0$ and $T^{-2}\sum_{t=1}^T (y_t^i)^2 \xrightarrow{p} 0$, which concludes the proof.

References

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

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

Blough, S., 1992, The relationship between power and level for generic unit root tests in finite samples, Journal of Applied Econometrics 7, 295–308.

Caner, M. and L. Kilian, 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. and C. Wei, 1987, Asymptotic inference for nearly nonstationary AR(1) processes, The Annals of Statistics 15, 1050–1063.

Cochrane, J., 1991, A critique of the application of unit root tests, Journal of Economic Dynamics and Control 15, 275–284.

Dufour, J.-M. and M. King, 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., Rothenberg, T. and J. Stock, 1996, Efficient tests for an autoregressive unit root, Econometrica 64, 813–836.

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

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

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

Kwiatkowski, D., Phillips, P., Schmidt, P. and Y. Shin, 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. and B. McCabe, 1994, A consistent test for a unit root, Journal of Business and Economic Statistics 12, 157–166.

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

Nabeya, S. and K. Tanaka, 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. and C. Vodounou, 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. and P. Perron, 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 u.s. macroeconomic time series, Journal of Monetary Economics 28, 435–459.

Stock, J., 1994, Unit roots, structural breaks and trends, in: R. Engle and D. McFadden, eds., Handbook of Econometrics, Vol. 4 (North Holland, New York) 2740–2841.

Stock, J., 2000, A class of tests for integration and cointegration, in: R. Engle and H. White, eds., Cointegration, Causality, and Forecasting — A Festschrift in Honour of Clive W.J. Granger (Oxford University Press, Oxford) 135–167.

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