Economics Department Discussion Papers Series

ISSN 1473 – 3307

When is a Time Series I(0)?

James Davidson

Paper number 08/11

URL: http://business-school.exeter.ac.uk/economics/papers/

When is a Time Series I(0)?*

James Davidson†
University of Exeter

Forthcoming in *The Methodology and Practice of Econometrics*, a festschrift for David F. Hendry, Eds. Jennifer Castle and Neil Shepherd, Oxford University Press.

Abstract

This paper surveys the extensive recent literature on the problems of deciding what is meant by an I(0) process, and then deciding how to test for the property. A formidable difficulty exists in the construction of consistent and asymptotically correctly sized tests for the I(0) hypothesis, and this may appear to place a question mark over the validity of a large area of econometric theory and practice. To overcome these difficulties in practical applications, the paper proposes that a slightly different question needs to be posed, relating to the adequacy of approximation to asymptotic inference criteria in finite samples. A simulation-based test, aimed at discriminating between data sets on this basis, is examined in a Monte Carlo experiment.

1 Introduction

Since the inception of integrated time series modelling in econometrics, the question of what constitutes a 'non-integrated' process has remained troublingly elusive. The inferential techniques developed for cointegration and related analyses require for their validity that the differences of the data series possess certain critical properties. These properties are nearly the same as those required for 'classical' asymptotics or, in other words, the application of the central limit theorem to approximate the distribution of regression coefficients and similar quantities. The project of doing time series econometrics could hardly be viable, one would suppose, unless these properties could be both clearly delineated, and subject to verification.

Before the advent of cointegration these problems were often resolved willy-nilly, by an assumption of correct specification in the context of a fairly heroic conditioning exercise, whereby the explanatory variables in a model were held to be 'fixed in repeated samples'. The only stochastic components left to model (the disturbances) could then be treated as independently and identically distributed, and their treatment was elementary. However implausible these classical assumptions may always have been, they are manifestly inadequate to deal with cointegration models, because here it not possible to hold the data conditionally fixed. It is the *observed series themselves*, not constructed disturbances, whose distributions must satisfy the critical regularity conditions.

*This paper shares a title with the first version of a working paper that subsequently appeared as Davidson (2002). It further explores some themes that the earlier working paper broached rather briefly. I am glad of this excuse to revive a nice title, although there is in practice minimal overlap between the content of this paper and its predecessor.

†Address for correspondence: james.davidson@exeter.ac.uk.

2 Defining I(0)

Early contributions to the cointegration literature tended to be fairly casual in their treatment of I(0), perhaps because this component of the theory was viewed as inherited from the pre-existing modelling methodology. The following definitions are culled from some widely cited articles and monographs.

1. "Definition: A series with no deterministic component which has a stationary, invertible ARMA representation after differencing d times is said to be integrated of order d..." (Engle and Granger 1987, p. 252.)
2. "It follows that [...] a short-memory series is I(0), as it needs differencing zero times" (Engle and Granger 1991, p. 3)
3. "... if the series must be differenced exactly k times to achieve stationarity then the series is I(k), so that a stationary series is I(0)." (Banerjee, Dolado, Galbraith and Hendry 1993, p. 7.)
4. "A finite (non-zero) variance stochastic process which does not accumulate past errors is said to be integrated of order zero..." (Hendry 1995, p. 43)
5. "A stochastic process Y_t which satisfies $Y_t - E(Y_t) = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$ is called I(0) if [$\sum_{i=0}^{\infty} C_i z^i$ converges for $|z| < 1$ and] $\sum_{i=0}^{\infty} C_i \neq 0$." (Johansen 1995, p. 34-35, the condition $\varepsilon_t \sim iid(0, \sigma^2)$ being understood.)

Of these (chronologically ordered) quotations, 2, 3 and 4 can be thought of as informal and descriptive, while 1 and 5 are intended as more rigorous. Even so, it's interesting to note that they are by no means equivalent. The concepts of stationarity, short memory and finite variance are each singled out as 'defining' descriptive characteristics, but it is not yet clear how these might be connected with one another. On the other hand, the more formal definitions restrict attention to a limited class of linear models, in which the three characteristics of stationarity, short memory and (under Gaussianity) finite variance are united in a single parametric restriction. Note that in a more general framework it easy to dispense with one while retaining another. The inclusion of deterministic components (e.g. 'trend stationarity') is only one of the many ways these models might be generalized.

Another approach to definition is the pragmatic one of simply specifying conditions under which the asymptotic theory is valid; see for example Stock (1994), Davidson (2002) and Müller (2008). These conditions are of course what motivate the technical and informal definitions just given, but in many ways it simplifies the analysis to state the desired properties directly, rather than conditions sufficient for them. Thus,

Definition 1 *A time series $\{x_t\}_{t=1}^{\infty}$ is I(0) if the partial sum process X_T defined on the unit interval by*

$$X_T(\xi) = \omega_T^{-1} \sum_{t=1}^{[T\xi]} (x_t - Ex_t), \ 0 < \xi \leq 1 \tag{2.1}$$

where $\omega_T^2 = \mathrm{Var}(\sum_{t=1}^{T} x_t)$, converges weakly to standard Brownian motion B as $T \to \infty$.

This definition first makes it clear that I(0) is an attribute of an infinite stochastic sequence. In other words, it is not a well-defined concept for observed time series except in the context of limit arguments as $T \to \infty$. Next, note that it implies the property $\omega_T^2 \sim T\omega^2$ for $0 < \omega^2 < \infty$, because otherwise the limit process cannot have the Brownian property $E(B(s) - B(r))^2) = s - r$

for $0 \leq r < s \leq 1$. For full applicability, it might need to be supplemented by the condition that a consistent estimator of ω^2 exists, which typically will be one of the class of kernel estimators; see Newey and West (1987) and Andrews (1991) inter alia. However, the best known sufficient conditions for these twin convergences, in distribution and probability, are in fact quite similar; see de Jong and Davidson (2000). It is quite possible that the best conditions actually coincide. Moreover, Kiefer, Vogelsang and Bunzel (2000) have shown that valid inference is possible without consistent variance estimation, although as pointed out below, their results don't have application for testing the I(0) hypothesis, in particular.

What is clear is that a very wide class of processes satisfy these conditions, of which the cases cited by Engle-Granger (1987) and Johansen (1995), respectively, form only a small subset.

3 Conditions for I(0)

Davidson (2002 and 2006, Section 5.5) provides a convenient summary of the technical conditions that ensure the property given in Definition 1 holds. A set of conditions is given for linear models that are effectively necessary for I(0), in the sense that convergence to a non-Brownian limit process (fractional Brownian motion) can be demonstrated in cases where they are violated.

Summability of the autocovariances (though not necessarily absolute summability) is the fundamental necessary condition for I(0), because on this condition depends the property $E(\omega_T^2) \sim T\omega^2$. Consider the class of covariance stationary moving average processes defined by

$$x_t = \sum_{j=0}^{\infty} a_j u_{t-j}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty, \quad u_t \sim \text{i.i.d.}(0, \sigma^2). \tag{3.1}$$

Since the mth order autocovariance is $\gamma_m = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+m}$, note that

$$\omega^2 = \sum_{m=-\infty}^{\infty} \gamma_m = \sigma^2 \left(\sum_{j=0}^{\infty} a_j \right)^2$$

so that summability of the autocovariances is equivalent to summability of the moving average coefficients. However, the conditions in (3.1) can be substantially relaxed by allowing dependence in the process $\{u_t\}$ itself, which can in its turn be weakly dependent with summable autocovariances. This can be illustrated by the obvious, though typically redundant, case where

$$u_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2).$$

Then we simply obtain

$$\omega^2 = \sigma^2 \left(\sum_{j=0}^{\infty} a_j \right)^2 \left(\sum_{j=0}^{\infty} b_j \right)^2$$

and this "Russian doll" layering of the dependence structure could be iterated any finite number of times.

More pertinent are the cases where u_t exhibits some form of nonlinear dependence. In these cases, restrictions on the autocovariances may need to be supplemented by more general restrictions on dependence. The simplest is to let u_t be a stationary ergodic martingale difference. A variety of mixing conditions are also popular in the literature, although these have the drawback of non-transparency. Being restrictions on the entire joint distribution of the process at long

range, they are difficult to test, either in an efficient manner, or at all. 'Geometric ergodicity' is a property of Markov chains which can be established for certain nonlinear difference equations (see, e.g. Tong 1990). The condition of 'near-epoch dependence' links the distribution of an observed process to that of the near epoch of a specified underlying forcing process, which can for example be mixing. However, in a variety of nonlinear models driven by independent shocks, it is comparatively easy to specify testable (in principle) parametric restrictions which are sufficient for near-epoch dependence of specified 'size' (rate of memory decay) and in turn sufficient for I(0) in the sense of Definition 1. The cases of ARCH and GARCH models, bilinear models and SETAR models, among others, are analysed in Davidson (2002).

The obvious difficulty with Definition 1 is that it specifies an asymptotic property that cannot be verified in any finite sample. Summability of the autocovariances can never be resolved, one way or the other, from sample information. It is not unreasonable to ask whether sample autocorrelations 'look' summable, in the sense that they decline at such a rate as the lag increases that some implicit smoothness constraint must be violated, were they to behave differently at long range. However, a number of authors have examined difficult cases that place our ability to make this discrimination in doubt, even in large samples.

Leeb and Pötscher (2001) consider processes u_t that are covariance stationary, and for which there exists no covariance stationary process v_t such that $u_t = \Delta v_t$ – in other words, are not over-differenced. They exhibit cases having these properties, yet lacking a spectral density (i.e., the spectral distribution function is non-differentiable) which also lack the characteristic property (necessary for Brownian asymptotics) that the partial sum variance increases proportionately to sample size. Accordingly, such processes cannot be regarded as I(0). Their results emphasize the fact that attributes such as 'stationary' or 'short memory', cannot substitute for Definition 1.

Müller (2008), on the other hand, considers processes generated by expansions of the form

$$Y(s) = \frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} g_k \sin(\pi s(k - \tfrac{1}{2}))\xi_k, \quad s \in [0, 1] \tag{3.2}$$

where $\xi_k \sim$ i.i.d.N$(0, 1)$. Setting $g_k = 1/(k - \frac{1}{2})$ defines a Brownian motion (see Phillips 1998) and sampling it at T points $s = 1/T, \ldots, 1$, yields a discrete integrated series. On the other hand, setting $g_k = 1$ yields, in the corresponding manner, a sample of Gaussian white noise. The interesting cases are found by setting $g_k = 1$ for $k = 1, ..., n$, for some $n < \infty$, and $g_k = 1/(k - \frac{1}{2})$ for $k > n$. For quite modest values of n, one can obtain a series that appears stationary, yet is also highly autocorrelated at long range. By letting n increase with T in just the right way, one can manufacture a series which is I(0) on Definition 1, yet the probability of rejection in any of a wide class of tests for (in effect) summable covariances converges to 1. This example is again artificial, but it illustrates the pitfalls that await those who seek to test the conditions embodied in the definition. As we show in more detail in the next section, there are always cases for which no sample is large enough to discriminate effectively.

4 Testing I(0)

Testing the hypothesis embodied in Definition 1 has been called an "ill-posed" inference problem, and a number of recent research contributions have highlighted different aspects of the difficulty.

Consider three possible approaches to the testing problem. 1) perform a test in the context of a specified parametric or semiparametric model; 2) test a specific restriction on the sample distribution, such as the value of the spectrum at zero; 3) construct a nonparametric statistic whose null distribution depends directly on the conditions of Definition 1. In practice these approaches will to a large degree overlap, but it is instructive to consider the difficulties implicit

in each. A fourth approach is to devise a consistent criterion for choosing between the specific alternatives of I(0) and I(1); see Stock (1994) and Corradi (1999). However, these latter methods have a rather specialized application, since they are predicated on the assumption that these two cases exhaust the possibilities. Given the existence of fractionally integrated processes in particular, this assumption appears unduly restrictive for our purposes.

4.1 Parametric Hypotheses

Start with the parametric framework. In an autoregressive or ARMA model, the null hypothesis takes the form "the largest autoregressive root lies strictly inside the unit circle".[1] The size control problems are immediately obvious, for the null hypothesis is defined by a non-compact set in the parameter space, say Ω_0, whose closure contains the leading case of the alternative (the unit root). If a test is consistent, then as sample size increases,

$$\text{size} = \sup_{\omega \in \Omega_0} P_\omega(\text{test rejects}) \to 1.$$

One can certainly test the hypothesis that the largest autoregressive root lies in a specified stable region which does not have 1 as a boundary point. This approach has the virtue that a failure to reject the restricted hypothesis implies a failure to reject the I(0) hypothesis at at most the same significance level. However, it does not tell us how to interpret a rejection and hence it cannot be considered as a test of I(0) in the strict sense.

Another approach which has proved popular is to embed the I(0) case in the class of I(d) models, where d represents the fractional integration (long memory) parameter. Note that $d \neq 0$ is incompatible with Definition 1, since the limit of the normalized partial sum process is a fractional Brownian motion. The LM-type tests of Robinson (1991), Agiakloglou and Newbold (1994), Tanaka (1999), and Breitung and Hassler (2002) are all of this form. These tests are constructed, in effect, as functions of the sample autocovariances. One might also construct a confidence interval for the parameter d itself, using either a parametric or a semiparametric procedure – see Robinson (1994), Geweke and Porter-Hudak (1983), Moulines and Soulier (2000) inter alia. Being based on the periodogram, these estimators can again be thought of as functions of the sample autocovariances. The problem with all these tests is that autoregressive components, if present, assume the role of nuisance parameters. Local dependence is known to induce small sample bias in these estimators, so that conventional significance tests for d have to be treated with caution.[2] For correct asymptotic size, these tests require that autoregressive components be controlled for by some method of pre-whitening. A valid test of $d = 0$ requires that any such autoregressive roots are in the stable region. However, a unit root is, of course, observationally equivalent to the case $d = 1$. The previous problem of size control now re-emerges in a new form. If the prewhitening is done consistently, these tests must have power equal to size against the alternative of a unit root.

4.2 "Ill-posed" Estimation Problems

A number of authors including Blough (1992), Dufour (1997), Faust (1996, 1999), Pötscher (2002), and Müller (2005, 2008) have investigated a class of estimation problems in which testing of integration order (whether I(0)or I(1)) features prominently. As Dufour points out, there are two distinct cases that give rise to similar difficulties in practice. One is a failure of identification

[1] There is also the parameterization which places stable roots outside the unit circle, but it is convenient for expository purposes to adopt the parameterization in which root and lag coefficient coincide in the AR(1) case.

[2] Davidson and Sibbertsen (2007) suggest a pre-test for bias.

at points of the parameter space; in other words, the existence of observationally equivalent points. The second case is where the object of interest is a function of the underlying parameters, and the parameter space contains points of discontinuity of this function.

Of the various analyses offered in these papers, Faust (1996, 1999) demonstrates the second case neatly, as follows. Consider the class of processes in (3.1). For the purposes of the argument let the shocks be Gaussian, and since $a_0 = 1$ is not imposed there is no loss of generality is assuming $\varepsilon_t \sim NI(0,1)$. Define $A = \{a_0, a_1, a_2, \ldots\}$ to be a point in the space of square-summable sequences $\mathcal{A} \subset \mathbb{R}^\infty$. Let the distance $\|\cdot\|$ be defined on $\mathcal{A}$ such that

$$\|A_1 - A_2\| = \sqrt{\sum_{j=0}^{\infty}(a_{1j} - a_{2j})^2}.$$

If $\{A_1, A_2, \ldots\}$ defines a sequence in $\mathcal{A}$ such that $\|A_k - A\| \to 0$, and the corresponding stochastic sequences are $\{X_{kt}\}$ such that

$$X_{kt} = \sum_{j=0}^{\infty} a_{kj}\varepsilon_{t-j}$$

then the distributions of the $\{X_{kt}\}$, say $\{P_{A_k}, k \geq 1\}$, converge weakly to P_A, the distribution of $\{X_t\}$. To demonstrate this, it is sufficient in view of the Gaussianity to show that the autocovariances of the processes converge. Given A, let $A^m = \{0, \ldots, 0, a_m, a_{m+1}, \ldots\} \in \mathcal{A}$, and note that $\|A_k^m - A^m\| \to 0$ if $\|A_k - A\| \to 0$. Also note that if $\gamma_{km} = E(X_{kt}X_{k,t-m})$ then for each $m \geq 0$,

$$\begin{aligned}
|\gamma_{km} - \gamma_m| &= \left|\sum_{j=0}^{\infty} a_{kj}a_{k,j+m} - \sum_{j=0}^{\infty} a_j a_{j+m}\right| \\
&= \left|\sum_{j=0}^{\infty} a_{kj}(a_{k,j+m} - a_{j+m}) + \sum_{j=0}^{\infty} a_{j+m}(a_{kj} - a_j)\right| \\
&\leq \|A_k^m - A^m\|\,\|A_k\| + \|A_k - A\|\,\|A^m\| \\
&\to 0 \text{ as } k \to \infty,
\end{aligned}$$

using the triangle and Schwarz inequalities. In other words, if $\|A_k - A\|$ is small then the difference between the distributions of $\{X_{kt}\}$ and $\{X_t\}$ is correspondingly small. Now consider the sequence $A_k = \{a_1, a_2, \ldots a_k, 0, 0, \ldots\}$, such that $A_k \to A \in \mathcal{A}$ but suppose $\sum_{j=0}^{\infty} a_j = \infty$. The sums $\sum_{m=0}^{\infty} \gamma_{km}$ are accordingly diverging as $k \to \infty$. $\{X_{kt}\}$ is an I(0) sequence for each k, but the limit is not I(0) in spite of lying arbitrarily close in distribution to I(0) sequences.

The implications for tests of the I(0) hypothesis should be clear. Supposing we seek to construct a confidence interval of level α for the spectral density at 0, say $f(0) = \pi^{-1}(\frac{1}{2}\gamma_0 + \sum_{m=1}^{\infty}\gamma_m)$. Let $(\Omega, \mathcal{F}, \mu)$ represent the probability space generating the process innovations, and also let $\mathcal{B}$ represent the Borel sets of the real line. An α-level confidence interval depending on a sample $\{X_1, \ldots, X_T\}$ is a measurable mapping $C_T(\alpha)$:$\mathcal{A} \times \Omega \longmapsto \mathcal{B}$ such that

$$\inf_{\mathcal{A}} P_A\left(f_A(0) \in C_T(\alpha)\right) \geq 1 - \alpha.$$

In words, a valid $C_T(\alpha)$ needs to contain $f_A(0)$ with probability at least $1 - \alpha$, no matter how the data are generated. It is evident that for any $\alpha > 0$, $C_T(\alpha)$ is unbounded. More alarmingly, this is also the case if attention is confined just to the subset $\mathcal{A}_0 = \{A \in \mathcal{A} : f_A(0) < \infty\}$, since this set is not compact, as demonstrated. Note that $\mathcal{A} \subset \overline{\mathcal{A}_0}$ (the closure of $\mathcal{A}_0$). *Every* non-summable element of $\mathcal{A}$ can be constructed as the limit of a sequence of summable elements, and $\overline{\mathcal{A}} = \overline{\mathcal{A}_0}$. The closure of the set of square-summable sequences contains the non-square-summable sequences.

This property of confidence intervals holds for any finite T. A standard kernel estimator of $f_A(0)$ should tend in distribution to the normal, with variance shrinking at the rate K_T/T where K_T is the bandwidth. However, the implied approximate confidence interval is an arbitrarily poor approximation to the true confidence interval. There exist data generation processes arbitrarily close to A for which the kernel estimate is diverging at the rate K_T, and has no well defined limiting distribution.

A closely related analysis considers the distribution of the difference processes $x_t = \Delta X_t$, having the representation

$$x_t = \sum_{j=0}^{\infty} a_j^* \varepsilon_{t-j}$$

where $a_0^* = a_0$ and $a_j^* = a_j - a_{j-1}$ for $j \geq 1$. Denote the generic sequence constructed in this way from an element A of $\mathcal{A}$ by $A^* \in \mathcal{A}$. If $A \in \mathcal{A}_0$ then $A^* \in \mathcal{A}_0^*$, where $\mathcal{A}_0^*$ is the subset of $\mathcal{A}$ having the property $\sum_{j=0}^{\infty} a_j = 0$. If attention is restricted to exponential lag decay processes, having the property $\sum_{j=m}^{\infty} a_j = O(a_m)$, we may further say that $\{X_t\}$ is I(0) if and only if the difference process belongs to $\mathcal{A}_0^*$. Evidently, sequences of elements of $\mathcal{A} - \mathcal{A}_0^*$ can be constructed whose limits lie in $\mathcal{A}_0^*$. In other words, there exist sequences of non-I(0) processes whose weak limits are I(0).

Pötscher (2002) points out that the existence of such points implies that consistent estimation is not a *uniform* property with respect to the parameter space. In other words, letting $\hat{\theta}_T$ denote an estimator of $f(0)$ the quantity $\sup_{A \in \mathcal{A}} E_A |\hat{\theta}_T - f_A(0)|^2$ is infinite, for every $T \geq 1$. A more subtle implication of the Faust-Pötscher analysis is that $\mathcal{A} - \mathcal{A}_0$ is dense in $\mathcal{A}$. *Every* model A with $f_A(0) < \infty$ is arbitrarily close to a case A' with $f_{A'}(0) = \infty$. Now, it might be thought that this result depends on the parameter space being explicitly infinite dimensional. Parametric representations of linear processes, such as the ARMA(p, q), are defined by subspaces of $\mathcal{A}$, (the images of mappings from $\Theta \subset \mathbb{R}^{p+q+1}$ to $\mathcal{A}$) which, it might be hoped, exclude most problematic regions. However, Pötscher shows that even the ARMA$(1, 1)$ class contains problematic points such that the uniform consistency criterion fails. Hence it also fails for every superset thereof.

4.3 The ARMA(1,1) Process

Consider the element of $\mathcal{A}$ defined by

$$(1 - \phi L) X_t = \sigma (1 - \psi L) \varepsilon_t$$

so that $a_0 = \sigma$ and $a_j = \sigma(\phi - \psi)\phi^{j-1}$ for $j \geq 1$. Consider initially just the AR(1), by fixing $\psi = 0$, and note that the sequence A_k defined by setting $\phi = \phi_k$ for $\phi_k = 1 - 1/k$ lies in $\mathcal{A}_0$, with limit $A \in \overline{\mathcal{A}} - \mathcal{A}_0$. In this case $A \notin \mathcal{A}$, and there is also a failure of the weak convergence of the distributions. The discontinuity in the space of probability measures at the stationarity boundary is a familiar feature of this class. However, as noted previously, the null hypothesis of I(0) is represented by the open set $\Omega_0 = \{\phi : |\phi| < 1\}$, such that the leading case of the alternative $\phi = 1$ lies in its closure. It follows that if a test of I(0) is defined by a statistic s_T and a critical region W_T, such that the hypothesis of I(0) is rejected if $s_T \in W_T$, then for any $T \geq 1$ the power of the test against the alternative $\phi = 1$ can never exceed the size defined as $\sup_{A \in \Omega_0} P_A(s_T \in W_T)$.

A special feature of the ARMA(1,1) class, closely related to the present problem although distinct from it, is the existence of the set of unidentified structures with $\phi = \psi$. Having the same likelihood corresponding to the case $\phi = \psi = 0$, all these structures represent i.i.d. data, although the case $\phi = \psi = 1$ is arbitrarily close in model space to I(1) cases with $\phi = 1$, $\psi < 1$. Pötscher

(2002) considers the following example. Construct a sequence of coefficient pairs, $\{\phi_k, \psi_k\}$ such that the sequence of spectral densities is

$$f_k(\omega) = \frac{\sigma^2}{2\pi} \frac{1 + \psi_k^2 - 2\psi_k \cos\omega}{1 + \phi_k^2 - 2\phi_k \cos\omega}.$$

Choose $M \geq 0$, and set $0 < \phi_k < 1$ and $\psi_k = 1 - M(1 - \phi_k)$, also requiring $\phi_k > (M-1)/M$ in the cases with $M > 1$ so that $\psi_k > 0$. Otherwise, $\{\phi_k\}$ can be an arbitrary sequence converging to 1. Note that $\psi_k \uparrow 1$ as $\phi_k \uparrow 1$, and also that along these sequences, $f_k(0) = \frac{1}{2}\pi^{-1}M^2\sigma^2$ for every k. Except at the limit point, the sequences of models have $\phi_k \neq \psi_k$ and hence they are technically identified, but depending on the path chosen they can have effectively any nonnegative spectral density at 0, in spite of being arbitrarily close to one another as the limit is approached.

As in the examples of the previous section, a confidence interval for $f(0)$ must be either unbounded, or have level zero. For a more familiar insight into this issue, consider the one parameter IMA(1, 1) class of models, defined by the MA parameter ψ. This has nonsummable lag coefficients for every $\psi \in (-1, 1)$, yet the case $\psi = 1$, lying in closure of this set, defines the i.i.d. case. Be careful to note that the fact this point is unidentified in the ARMA(1,1) class is irrelevant, for it is perfectly well identified in the IMA class. This problem is related strictly to the discontinuity of $f(0)$ as a function of ψ.

4.4 Nonparametric Tests

The most popular procedures for checking I(0) involve computing statistics that address the question of summability of the autocovariances directly. Among tests in this class are the modified R/S test (Lo, 1991), the KPSS test (Kwiatkowski et al., 1992), the LM test of Lobato and Robinson (1998), the V/S test of Giraitis et al. (2003), the 'remote autocorrelations' test of Harris et. al. (2008), and the increment ratio test of Surgailis et al. (2008). Except for the last, these tests all depend on an estimator of the long run variance of the process, which is assumed finite under the null hypothesis. In fact, it is true to say that the properties of the tests are completely defined by the properties of these variance estimators. It is necessary to specify the null by specifying a finite lag, beyond which the sum of the autocovariances is either exactly zero or arbitrarily close to zero. Different choices of truncation point effectively define different null hypotheses, all of which are strictly contained in the 'I(0) hypothesis' proper.

The force of this point is nicely illustrated by the fact that the KPSS statistic, if constructed using the Bartlett kernel with bandwidth set equal to sample size, has a degenerate distribution with value $\frac{1}{2}$ (see Kiefer and Vogelsang 2002). In other words the KPSS test can be viewed as comparing two variance estimators, respectively imposing and not imposing a truncation point smaller than sample size. The problem is there are $T - 1$ such comparisons that can be made in a sample of size T, and no formal constraints on the proper choice. Since the null hypothesis imposes no finite truncation point, as such, the test is bound to be oversized for any finite truncation; equivalently, there is always a valid truncation point which sets power equal to size.[3]

5 Fingerprinting I(0)

The literature surveyed in this paper may appear to place a question mark over large areas of econometric practice. If there are serious problems in discriminating between I(0) models and alternatives, what is the future for methods of analysis which depend critically on making this

[3] Interestingly, $\frac{1}{2}$ actually exceeds the 5% critical value of the limiting KPSS null distribution, so there always exists a truncation to guarantee rejection under both null and alternative at the nominal 5% level.

assessment reliably at the outset? Indeed, some authors have evidenced a certain satisfaction at pouring cold water on the efforts of time series analysts in this area.

Before going too far in this pessimistic direction, however, we do well to remind ourselves of the actual question usually being posed. In almost every application, this is "Will asymptotic distribution results based on the assumption of I(0) provide more accurate approximate inferences than alternatives, in my sample?" Call this Question 1. It is clearly a different question from the following, which we will call Question 2: "Will the distributions obtained by extending my sample indefinitely match the asymptotic distributions implied by the I(0) hypothesis?" It is Question 2 that has proved to be difficult to answer in the conventional manner. However, this is of little concern if there is no actual prospect of extending the sample indefinitely, and if there were then the difficulties would resolve themselves by the same token. As to Question 1 it is, arguably, reasonable to be guided by the popular adage: "If it walks like a duck, and quacks like a duck, then (let's assume) it's a duck."

The problem is to find an independent yardstick by which to judge, in a simulation experiment for example, whether the answer to Question 1 is affirmative. Linking back to Definition 1, this is essentially the question of whether the partial sums of the process approximate to Brownian motion in a sufficiently large sample. A natural approach to answering this question is to formulate a real-valued statistic whose limiting distribution corresponds to a unique functional of Brownian motion. Unfortunately, most statistics known to converge to pivotal Brownian functionals (for example, the Dickey-Fuller statistic and variants) are dependent on unknown scale factors, and embody estimates of the long-run variances. As previously noted, invoking these would tend to make the problem circular.

There is one nice exception, however. Consider the statistic $T^{-1}\hat{\varrho}_T$ where

$$\hat{\varrho}_T = \frac{\sum_{t=1}^T U_t^2}{T\sum_{t=1}^T u_t^2}$$

where $U_t = u_1 + \cdots + u_t$, and either $u_t = x_t - \bar{x}$ with $\bar{x}$ denoting the sample mean, or $u_t = x_t - \hat{\delta}' z_t$ where z_t is a vector of deterministic regressors, such as intercept and time trend. For simplicity we consider only the former case, but the extension is very easily handled. Note that $\hat{\varrho}_T$ is similar to the KPSS statistic, except that the variance estimate is not autocorrelation-corrected. This statistic is proposed by Breitung (2002) as a nonparametric test of I(1). Suppose that $v_t \sim \mathrm{I}(0)$ with mean 0 and long-run variance $\sigma^2 < \infty$, and $x_t = \sum_{s=1}^t v_s$. Then (by definition)

$$T^{-1/2} x_{[T\cdot]} \xrightarrow{d} \sigma W(\cdot)$$

where W is standard Brownian motion, and accordingly, by the continuous mapping theorem, $T^{-1}\hat{\varrho}_T \xrightarrow{d} \Xi_0$ where

$$\Xi_0 = \frac{\int_0^1 \left(\int_0^\tau W(s)ds - \tau \int_0^1 W(s)ds\right)^2 d\tau}{\int_0^1 W(\tau)^2 d\tau - \left(\int_0^1 W(\tau)d\tau\right)^2}. \tag{5.1}$$

Breitung points out that under the alternative hypothesis $u_t \sim \mathrm{I}(0)$, $T^{-1}\hat{\varrho}_T = O_p(T^{-1})$, and hence, using the lower tail as a rejection region yields a consistent test of I(1) against the alternative of I(0).

The test does not provide a consistent test against the alternative of $\mathrm{I}(1+d)$ for $d > 0$ (and hence by implication a test of I(0) applied to the partial sums) because the distribution of $T^{-1}\hat{\varrho}_T$ has bounded support. In fact, it never exceeds $1/\pi^2$ regardless of the distribution of $\{x_t\}$ (see Davidson, Magnus and Wiegerinck 2007). However, consider the case where v_t is $I(d)$ for $d > 0$.

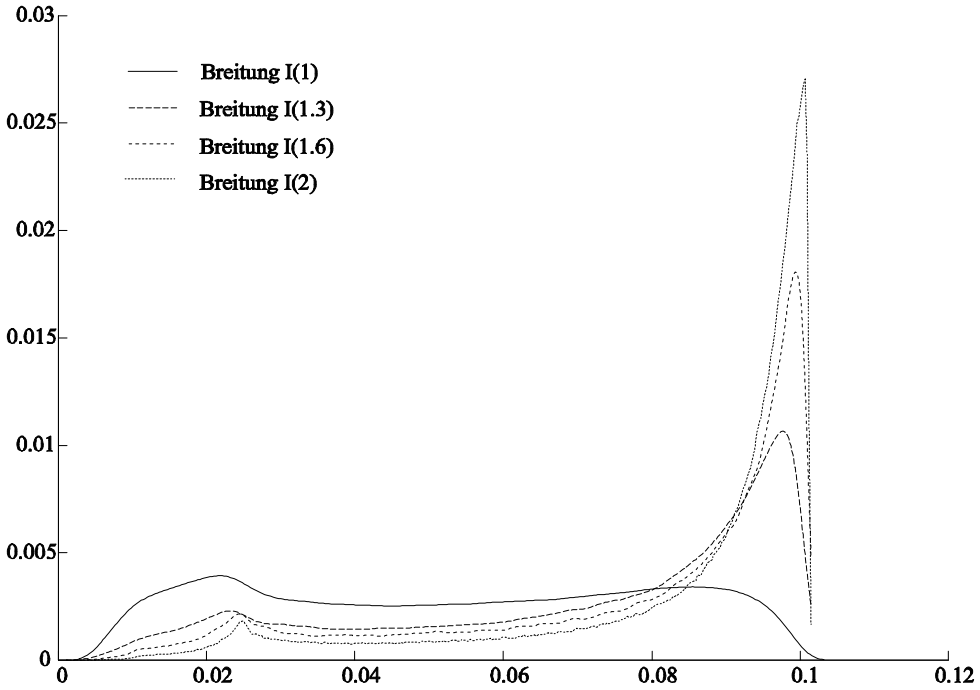

Figure 1: Breitung (2002) statistic with cumulated I(d) increments . The case I(1) is Breitung's null distribution. (Kernel density plots from 1 million replications.)

If $d < \frac{1}{2}$, then under mild assumptions on the increments (see for example Davidson and de Jong 2000) we have the result

$$T^{-d-1/2} \sum_{s=1}^{[T\tau]} v_s \xrightarrow{d} \sigma W_d(\tau)$$

where σ is the long-run variance of the fractional differences $(1-L)^d v_t$, and W_d is fractional Brownian motion as defined by Mandelbrot and Van Ness (1968) for $-\frac{1}{2} < d < \frac{1}{2}$. The Breitung statistic then has the limit

$$\Xi_d = \frac{\int_0^1 \left(\int_0^\tau W_d(\zeta) d\zeta - \tau \int_0^1 W_d(s) ds \right)^2 d\tau}{\int_0^1 W_d(\tau)^2 d\tau - \left(\int_0^1 W_d(\tau) d\tau \right)^2}. \tag{5.2}$$

On the other hand, if $\frac{1}{2} < d < \frac{3}{2}$ then

$$T^{-d-1/2} \sum_{s=1}^{[T\tau]} v_s \xrightarrow{d} \int_0^\tau W_{d-1}(\zeta) d\zeta, \qquad 0 \le \tau \le 1$$

and

$$\Xi_d = \frac{\int_0^1 \left(\int_0^\tau \int_0^\alpha W_{d-1}(\zeta) d\zeta d\alpha - \tau \int_0^1 \int_0^\alpha W_{d-1}(\zeta) d\zeta d\alpha \right)^2 d\tau}{\int_0^1 \left(\int_0^\tau W_{d-1}(\zeta) d\zeta \right)^2 d\tau - \left(\int_0^1 \int_0^\tau W_{d-1}(\zeta) d\zeta d\tau \right)^2}. \tag{5.3}$$

Be careful to note how the extra normalization factors T^{-2d} cancel in the ratio, as does σ, so that these distributions remain $O_p(1)$ and free of nuisance parameters other than d. These

	T	50				100			200	
	ϕ	0.3	0.5	0.7	0.9	0.5	0.7	0.9	0.7	0.9
KPSS:	Bw = 4	0.062	0.087	0.145	0.270	0.095	0.172	0.403	0.189	0.509
	Bw = 12	0.043	0.043	0.051	0.080	0.055	0.077	0.174	0.087	0.236
	Andrews	0.044	0.029	0.011	0.028	0.041	0.026	0.003	0.043	0.010
	N-W	0.060	0.077	0.112	0.212	0.073	0.114	0.264	0.115	0.316
K-S	for $T^{-1}\hat{\varrho}_T$	0.727	**1.486**	**2.786**	**6.790**	0.737	**1.434**	**4.404**	0.802	**2.401**

Table 1: KPSS Rejections in Gaussian AR(1) models with parameter ϕ, in 100,000 replications. The last row shows the Kolmogorov-Smirnov statistic for comparison of partial sums with the Breitung distribution.

distributions have been tabulated by simulation for four values of d, using 1000 NID(0,1) drawings to represent the v_s (see Figure 1). While any I(0) process v_s must yield (5.1) in the limit, it is clear that the passage to the limit may be substantially different, depending on the strength of dependence. Thus, the distribution of $T^{-1}\hat{\varrho}_T$ where v_t is an autoregressive process, with a root close to unity, is likely to resemble Ξ_1 more closely than Ξ_0 in samples of moderate size.

The idea to be explored here is to use the null distribution of Breitung's statistic to fingerprint (the partial sums of) an I(0) process. If the latter distribution cannot be distinguished from the former, in a sample of given size, it is a reasonable conjecture that the dependence in the process is innocuous from the point of view of applying asymptotic inference. Of course, this is by no means the only statistic that might be used for this purpose, but it does have two notable advantages, independence of scale parameters and bounded support. The latter is a particularly convenient feature for implementing a comparison of distributions.

In Table 1, data have been simulated from five I(0) processes, the Gaussian AR(1) with coefficients $\phi = 0$, 0.3, 0.5, 0.7 and 0.9, and three sample sizes, $T = 50$, 100 and 200. In all these cases the correct answer to Question 2 is affirmative. The KPSS test has been computed for these series with HAC variance estimator computed using the Bartlett kernel and four choices of bandwidth, two fixed, and two selected by data-based 'plug-in' methods as proposed by, respectively, Andrews (1991) and Newey and West (1994) (denoted N-W in the table).[4] To provide critical values, 1.5 million Gaussian i.i.d. samples were used to construct tabulations for each choice of T, so ensuring that all features of the data and test procedure, except the dependence, are correctly modelled. Viewed as attempts to answer Question 2, all of these procedures appear to represent an unsatisfactory compromise. Only the Andrews method is never over-sized, but its power against a unit root alternative appears in doubt.

The last row of the table shows the Kolmogorov-Smirnov tests of the Breitung distributions generated from the Monte Carlo replications for each case, using the tabulations from the i.i.d. data to provide the benchmark distributions. Those cases exceeding the asymptotic 5% critical value, of 1.35, are shown in boldface in the table.[5] Suppose we take rejection on this test as a negative answer to Question 1. On this criterion, only the case $\phi = 0.3$ is included in the null hypothesis in a sample of size 50. In a sample of size 100, $\phi = 0.5$ enters the acceptance region, and in a sample of 200, so does $\phi = 0.7$. The point to be emphasized here is that the KPSS

[4] The plug-in formulae have the form bandwidth $= 1.447(\alpha T)^{1/3}$ where $\alpha = \alpha_A$ and $\alpha = \alpha_{NW}$, respectively, and $\alpha_A = 4\hat{\rho}^2/(1-\hat{\rho})^2(1+\hat{\rho})^2$ where $\hat{\rho}$ is the first-order autocorrelation coefficient, and $\alpha_{NW} = \left[2\sum_{j=1}^{[n_T]} j\hat{\gamma}_j / \left(\hat{\gamma}_0 + 2\sum_{j=1}^{[n_T]} \hat{\gamma}_j\right)\right]^2$ where $\hat{\gamma}_j$ is the jth order sample autocovariance. Here, [.] is the floor function, and $n_T = 3(T/100)^{2/9}$ so that $[n_{50}] = 2$, and $[n_{100}] = [n_{200}] = 3$. Newey and West advocate a pre-whitening step using an autoregression before applying their kernel estimator, but this step has been omitted here.

[5] For clarity the table shows only the most extreme cases of the null hypothesis, as indicated by the K-S statistic.

tests are even less satisfactory as a means for answering Question 1 than for answering Question 2. Except for the Andrews method, which has no power, the rejection rates for a given ϕ all *increase* with sample size, whereas on the criterion of Question 1, as indicated by the last row, we should like them to decrease. It is, manifestly, the evidence contained in the last row of Table 1 that we should most like to possess, when evaluating Question 1. The next section attempts to operationalize this insight.

6 A Bootstrap Test of I(0)

A test of I(0) in the sense of Question 1, based directly on the comparison of fingerprinting distributions, might be implemented by the following steps.

1. Formulate and fit a model of the data generation process.
2. Use this estimate to simulate the series many times and tabulate the Breitung statistic $T^{-1}\hat{\varrho}_T$ for the partial sums.
3. Use the Kolmogorov-Smirnov test to compare the distribution of this statistic with the benchmark case based on independent increments.

Given an implementation of Step 1, which we discuss in detail below, Step 2 might be performed using a Gaussian random number generator, or by bootstrap draws from the Step 1 residuals. In the latter case it is very important to generate the benchmark distribution from the same sample as the test distribution, to avoid a spurious difference. The drawings are recoloured by the estimated filter to create the test distribution, and used unfiltered to create the benchmark. Note that differences in the variances of the two draws are unimportant, since scale effects cancel in the construction of the Breitung statistic. For Step 3, the benchmark distribution should preferably be estimated in parallel with matching sample size, and compared by the two-sided Kolmogorov-Smirnov test. This is to ensure that it is exclusively the dependence that influences the test outcome, not the accuracy of the asymptotic approximation.

Estimation of the DGP is clearly the trickiest step, in effect the counterpart of the bandwidth selection problem in conventional tests, although the constraints it imposes are different and generally more favourable. Note that nonparametric methods for bootstrapping under dependence, such as the block bootstrap or Fourier bootstrap, are not attractive in this context because of the problem of matching the distributions under the null hypothesis. Given a suitable estimator of the autocovariance function, it would be feasible to simulate using the Choleski method or the circulant embedding algorithm (Davies and Harte, 1987). However, this estimation problem is precisely the source of the difficulties described in Section 4.2. Therefore, parametric modelling as in Step 1 appears the most promising approach.

For power against unit and near-unit root autoregressive alternatives, an autoregressive model naturally suggests itself. However, this is a less attractive option from the point of view of detecting fractional alternatives, since unrestricted estimation of a hyperbolic AR(∞) lag structure poses obvious efficiency problems. Therefore it seems important that the autocorrelation model contain a fractional integration component. One possibility is to fit an ARFIMA model to the data, although there are well-known identification and numerical problems involved in simultaneously fitting an autoregressive root and fractional d parameter. Multi-modal and poorly conditioned likelihoods are commonly encountered in these models. For the purposes of a Monte Carlo study, where a routine of model checking and evaluation at each replication is not feasible, three options have been compared. The first is a sieve autoregression, using the Akaike criterion to select the AR order from the set $0, \ldots, [0.6T^{1/3}]$. The second is to fit an ARFI(1,d)

two-parameter model by nonlinear least squares. The third alternative considered is to fit a truncated fractional model, of the form

$$x_t = - \sum_{j=1}^{\min(\tau,t-1)} b_j x_{t-j} + e_t$$

where $b_j = (j-d-1)b_{j-1}/j$, with $b_0 = 1$, and the fitted parameters are d and τ. Think of this as a restricted version of the sieve autoregression, parsimoniously approximating either a low-order autoregressive alternative with τ small, or a fractional alternative with τ large.

Table 2 shows the results of replicating these three test procedures, using 500 bootstrap draws to generate the test distributions at Step 2. For four sample sizes, T = 50, 100, 200 and 500, the rows of the table show the results for these three estimation methods augmented by the 'True' model, where the known data generation process has been used to create the bootstrap replications. This test is of course infeasible in practice, but it provides a yardstick against which to gauge the effectiveness of the alternative feasible methods.

The table entries show the proportion of rejections in the Kolmogorov-Smirnov test comparing the distribution of Breitung's statistic constructed from the re-coloured data with that of the statistic constructed from the same number of i.i.d. bootstrap drawings. Each statistic was first tabulated under the null hypothesis from 10,000 replications using i.i.d. normal drawings, so as to provide correct critical values for each sample size. Taking the critical values for the 5%-level 'True' test as the yardstick (so that these table entries are 0.05 by construction, note) the first column of the table shows the estimated sizes of the nominal 5% tests. The remaining columns show estimates of the true powers (using the null tabulations to provide critical values) against seven alternatives, based on 5000 replications of each case. The cases are four AR(1) processes with parameter ϕ and three ARFIMA(0,d,0) processes, with i.i.d. Gaussian shocks and zero start-up values in each case.

Some important points of interpretation need to be borne in mind, in studying this table. In the limiting case as $T \to \infty$, we should expect to find power = size for each of the four cases of the I(0) hypothesis, and power = 1 for each of the three cases of the I(d) alternative. In finite samples, however, rejection in the I(0) cases is not an incorrect outcome. The issue is whether the autocorrelation is strong enough to put asymptotic inference criteria into question. The infeasible 'True' cases represent the ideal outcomes from this point of view, against which the feasible tests can be judged. If this test were to be adopted as a pre-test before a conventional inference procedure, we can even see it as a means of discriminating between data sets which (by chance) tend to satisfy our validity criteria, from those which violate it. Failure to reject can be conjectured to indicate that subsequent tests with these data may not be too badly sized.

In the event, the truncated fractional model appears to have the best all-round performance. The sieve AR method performs generally closest to the infeasible test in the I(0) cases, but has poor power properties against the fractional alternatives. The ARFI method suffers the worst from spurious rejection and so diverges furthest from the 'True' benchmark under I(0), while the truncated fractional method appears to offer the best compromise in both cases. Of course, this is chiefly due to the fact that it gives a good approximation to both the AR(1) and FI alternatives tested. To determine how it performs in a more general setting calls for more experiments. In practical implementations (as opposed to a Monte Carlo experiment) the test should be performed following the specification and estimation of a time series model by the investigator, and so tailored more accurately to the data set in question.

		Size	Power						
		(Nominal	AR(1): ϕ				FI: d		
T	Test	5% Test)	0.3	0.5	0.7	0.9	0.3	0.5	0.7
50	True	0.050	0.206	0.682	0.997	1	1	1	1
	Sieve AR	0.068	0.164	0.492	0.828	0.975	0.293	0.672	0.924
	ARFI(1,d)	0.174	0.064	0.092	0.305	0.797	0.383	0.622	0.861
	Trunc. FI	0.129	0.135	0.366	0.730	0.964	0.458	0.733	0.926
100	True	0.050	0.085	0.207	0.673	1	1	1	1
	Sieve AR	0.054	0.086	0.204	0.557	0.962	0.105	0.479	0.846
	ARFI(1,d)	0.153	0.073	0.092	0.193	0.810	0.554	0.758	0.730
	Trunc. FI	0.095	0.112	0.168	0.473	0.953	0.548	0.761	0.877
200	True	0.050	0.069	0.098	0.261	0.986	1	1	1
	Sieve AR	0.050	0.071	0.096	0.249	0.867	0.197	0.516	0.857
	ARFI(1,d)	0.165	0.080	0.096	0.155	0.602	0.842	0.914	0.855
	Trunc. FI	0.094	0.077	0.063	0.139	0.818	0.750	0.862	0.884
500	True	0.050	0.050	0.056	0.075	0.381	1	1	1
	Sieve AR	0.050	0.050	0.056	0.074	0.344	0.070	0.208	0.755
	ARFI(1,d)	0.143	0.083	0.109	0.159	0.369	0.960	0.975	0.928
	Trunc. FI	0.081	0.044	0.035	0.045	0.283	0.890	0.967	0.985

Table 2: Bootstrap I(0) test: Rejection rates for the Kolmogorov-Smirnov test of Breitung's statistic in 5000 replications

7 Concluding Remarks

The hypothesis that a time series is I(0) has been justly described as an "ill-posed" problem for statistical investigation. A number of studies have shown that this question, as conventionally posed, is unsuited to standard methods of inference. This paper suggests that there are more suitable hypotheses to test, relating directly to the implications of the distribution of the data for asymptotic (i.e., approximate) inference. A convenient asymptotically pivotal statistic is used as a yardstick, to assess how far data features such as local dependence affect the distribution, in a sample of given size. The null hypothesis under test is not 'I(0)' in the strict sense, but the arguably more useful hypothesis that the assumption of I(0) is innocuous from the point of view of the asymptotic approximation of test distributions.

It's important to emphasize that this test is strictly of the properties of a *model* (or DGP), not a direct test on an observed series, as such. The link between the model and the data has to be supplied by the explicit modelling exercise, which is accordingly the key component of the procedure. The reported Monte Carlo results, which show simple models fitted mechanically to series with a known simple structure, need to be interpreted with care in this light. Whereas reproducing the observed autcorrelation structure of the data is a key requirement, don't overlook the fact that (for example) an uncorrelated IGARCH process is a case of the alternative. Power against such cases depends on a suitable choice of model. In view of the cited result of Müller (2008), there are bound to be cases which defy the ability of popular time series models to capture the dependence structure, although being non-causal it is questionable whether processes of the type (3.2) can feature in observed economic time series. It will be useful to compare the performance of the test in alternative DGPs, especially with nonlinear dynamics, and also to calibrate the performance of conventional tests, such as the Dickey Fuller, in conjunction with bootstrap 'pre-testing'. Among other important questions is whether the Breitung statistic is the best candidate for comparison, or whether a range of benchmarks might be implemented. Such

exercises must however be left for future work.

8 References

Agiakloglou, Ch. and P. Newbold (1994) Lagrange multiplier tests for fractional difference. *Journal of Time Series Analysis* 15, 253–262.

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation *Econometrica* 59 (3), 817-858.

Andrews, D. W. K. and J. C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60 (4).953-966.

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Blough, S. R. (1992): The Relationship hetween Power and Level for Generic Unit Root Tests in Finite Samples, *Journal of Applied Econometrics* 7, 295-3013,

Breitung, J. (2002), Nonparametric tests for unit roots and cointegration, *Journal of Econometrics* 108, 343–363.

Breitung, J. and U. Hassler (2002), Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics* 110, 167–185

Breitung, J., and A. M. R. Taylor (2003), Corrigendum to "Nonparametric tests for unit roots and cointegration, *Journal of Econometrics* 117 401–404.

Busetti, F. and A. M. R. Taylor (2004) Tests of stationarity against a change in persistence, *Journal of Econometrics* 123, 33–66.

Cavaliere, G. and A. M. R. Taylor (2005) Stationarity tests under time-varying second moments *Econometric Theory* 21, 1112–1129

Corradi, V., (1999), Deciding between I(0) and I(1) via Flil-based bounds, *Econometric Theory* 15 (5) 643-663.

Davidson, J. (2002) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243-269.

Davidson, J. (2006) Asymptotic Methods and Functional Central Limit Theorems. Chapter 5 of *Palgrave Handbook of Econometrics: Vol. 1 Econometric Theory* , eds. T. C. Mills and K. Patterson, Palgrave Macmillan.

Davidson, J. and R. M. de Jong (2000) Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68(2) 407-424

Davidson, J. and P. Sibbertsen (2007) Tests of bias in log-periodogram regression. Working paper.

Davidson, J., J. Magnus and J. Wiegerinck (2007) A general bound for the limiting distribution of Breitung's statistic. Working paper

Davies, R. B. and D. S. Harte (1987) Tests for Hurst effect. *Biometrika* 74, 95-102

de Long, R. M. and J. Davidson (2000) The functional central limit theorem and weak convergence to stochastic integrals I: weakly dependent processes *Econometric Theory* 16, 621-42.

Den Haan, W. J. and A. Levin (1997) A practitioner's guide to robust covariance matrix estimation, Chapter 12 Chapter 12 (pp291-341) of *Handbook of Statistics 15: Robust Inference*, eds. G. S. Maddala and C. R. Rao, Elsevier.

Dufour, J.-M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models, *Econometrica* 65 (6), 1365-1387.

Engle, R. F. and Granger, C. W. J. (1987) Co-integration and error correction: representation, estimation and testing, *Econometrica* 35, pp.251-276.

Engle, R. F. and Granger, C. W. J. (1991) *Long-run Economic Relationships: Readings in Cointegration*, Oxford University Press.

Faust, J. (1996) Near observational equivalence and theoretical size problems with unit root tests, *Econometric Theory* 12 (4) 724-731.

Faust, J. (1999) Conventional confidence intervals for points on spectrum have confidence level zero, *Econometrica* 67 (3) 629-637.

Geweke, J. and S. Porter-Hudak (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221-237

Giraitis, L., P. Kokoszka, R. Leipus and G. Teyssiere (2003) Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* 112, 265-294.

Harris, D., B. M. McCabe and S. Leybourne (2008) Testing for long memory, *Econometric Theory* 24,1 (forthcoming)

Hendry, D. F. (1995) *Dynamic Econometrics.* Oxford: Oxford University Press.

Hurvich, C. M., R. Deo and J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of a long memory time series, *Journal of Time Series Analysis* 19, 19–46

Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press

Kiefer, N. M., T. J. Vogelsang and H. Bunzel (2000)Simple robust testing of regression hypotheses. *Econometrica*, 68(3) 695-714.

Kiefer, N. M. and T. J. Vogelsang (2002) Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation, *Econometrica* 70 (5), 2093-2095.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159-178.

Leeb, H. and B. M. Pötscher (2001) The variance of an integrated process need not diverge to infinity, and related results on partial sums of stationary processes, *Econometric Theory* 17(4) 671-685.

Lo, Andrew W. (1991) Long-term memory in stock market prices, *Econometrica* 59, 5, 1279-1313

Lobato, I. N. and P. M.. Robinson (1998) A Nonparametric Test for I(0), *Review of Economic Studies* 65 (3), 475-495.

Moulines, E. and P. Soulier (1999) Broad ban log-periodogram estimation of time series with long-range dependence. *Annals of Statistics* 27, 1415–1439

Müller, U. K. (2005) Size and power of tests of stationarity in highly autocorrelated time series, *Journal of Econometrics* 128, 195–213.

Müller, U. K. (2008) The impossibility of consistent discrimination between I(0) and I(1) processes. *Econometric Theory* 24, 616 – 630.

Newey, W. K., and K. D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-8

Newey, W. K. and K. D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–54.

Phillips, P. C. B. (1998), New tools for understanding spurious regression. *Econometrica* 66, 1299–1325.

Pötscher, B. M. (2002), Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots and estimation of long memory parameters, *Econometrica* 70 (3) 1035–1065.

Robinson, P.M. (1991), Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regressions. *Journal of Econometrics* 47, 67–84.

Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420–1437.

Stock, J. H. (1994) Deciding between I(1) and I(0), *Journal of Econometrics* 63, 105-131.

Surgailis, D., G. Teyssiere and M. Vaiciulis (2008) The increment ratio statistic. *Journal of Multivariate Analysis* 99, 510–541.

Tanaka, K. (1999) The nonstationary fractional unit root, *Econometric Theory* 15, 549–582

Tong, H (1990) *Non-linear Time Series* Oxford: Oxford University Press