12

Simple Wald Tests of the Fractional Integration Parameter: An Overview of New Results

*Juan J. Dolado, Jesus Gonzalo, and Laura Mayoral**

12.1 Introduction

A well-known feature of tests of $I(1)$ vs. $I(0)$—or $I(0)$ vs. $I(1)$—processes is that they reject their respective null hypotheses very occasionally when the true DGP for a time series $\{y_t\}_1^T$ is a fractionally integrated, $I(d)$, process. This is often the case for the Dickey–Fuller (DF)-type tests if $0.5 < d < 1$ and for the KPSS-type tests if $0 < d < 0.5$. Given that the microfoundations of $I(d)$ processes make them quite plausible in practice, this issue can have serious consequences when analysing the long-run properties of the variables of interest.[1] To mention only a few: (i) shocks could be identified as permanent when in fact they die out eventually, and (ii) two series could be considered as spuriously cointegrated when they are independent at all leads and lags (see, eg, Gonzalo and Lee, 1998). These mistakes are more likely to occur in the presence of deterministic components like, eg, in the case of trending economic variables. Additionally, if the true DGP is an $I(0)$ process subject to structural breaks in its deterministic components, then it could be misinterpreted as a long-memory process, or vice versa.

* We are grateful to Carlos Velasco for many insightful comments. Financial support from the Spanish Ministry of Education through grants SEJ2006-00369 and SEJ2007-63098, and also from the Barcelona Economics Program of CREA is gratefully acknowledged. The usual disclaimer applies.

[1] For explanations of the origin of $I(d)$ processes based on aggregration of individual stationary series with heterogeneous persistence, see Robinson (1978) and Granger (1980), and for models which mimic some of the key properties of $I(d)$ processes based on the existence of shocks that die out at a certain probabilistic rate, see Parke (1999) and Diebold and Inoue (2001). For persuasive macroeconomic applications of these processes, see Michelacci and Zaffaroni (2000) and Lo and Haubrich (2001).

In view of these caveats, the goal of this chapter is four-fold. First, we illustrate the advantages, in terms of power under fixed alternatives, of recently proposed Wald tests of $I(d_0)$ vs. $I(d)$, $d \neq d_0$, with $d_0 = 1$ or $d_0 = 0$, relative to well-known LM and semiparametric tests; for simplicity, we do this in a setup when the time-series has i.i.d. error terms and is free of deterministic components. Second, we extend the previous procedures to allow for these components, possibly subject to structural breaks. Third, we derive new LM and Wald test statistics to test the null that a process is $I(d)$ with constant long-memory parameter, d, against the alternative of a break in d. Finally, the Wald tests are extended to account for autocorrelated disturbances in the DGP.

Specifically, we focus on a modification of the Fractional Dickey–Fuller (FDF) test by Dolado, Gonzalo, and Mayoral (2002; DGM hereafter) recently introduced by Lobato and Velasco (2007; LV hereafter) to achieve an improvement in efficiency over the former. Although this test—henceforth denoted as the EFDF (efficient FDF) test—was originally devised to extend the traditional DF test of $I(1)$ against $I(0)$ to the broader framework of $I(1)$ against $I(d)$ processes, with $d \in [0,1)$, we show that it can be easily generalized to cover the case of $I(0)$ vs. $I(d)$, with $d \in (0,1]$. This testing approach relies upon a simple regression model where both the regressand and the regressor are filtered to become $I(0)$ under the null and the alternative hypotheses, respectively.[2] The test is based on the t-ratio, t_φ, of the estimated slope, φ, of the regressor. Hence, when testing $I(1)$ vs. $I(d)$, Δy_t becomes the dependent variable. As regards the regressor, whereas DGM choose $\Delta^d y_{t-1}$, LV show that $z_{t-1}(d) = (1-d)^{-1}(\Delta^{d-1} - 1)\Delta y_t$ improves the efficiency of the test.[3] These tests belong to the Wald family because their underlying regression models are estimated under the alternative hypothesis. Thus, non-rejection of $H_0 : \varphi = 0$ against $H_1 : \varphi < 0$, implies that the process is $I(1)$ and, conversely, $I(d)$ when the null is rejected. As shown below, the EFDF test for testing $I(0)$ vs. $I(d)$ is based on an analogous t-ratio, t_ψ, this time in a regression of y_t on the regressor $s_{t-1}(d) = d^{-1}(1 - \Delta^d)y_t$.

To compute either version of the EFDF test, an input value for d is required. One could either consider a (known) simple alternative, $H_A : d = d_A < 1$

[2] In the DF setup, these filters are $\Delta = (1 - L)$ and $\Delta^0 L = L$, so that the regressand and regressor are Δy_t and y_{t-1}, respectively.

[3] As explained in DGM (Appendix A; 2002), both regressors can be constructed by filtering the series $\{y_t\}_{t=1}^T$ with the truncated version at the origin (with pre-sample shocks set to 0) of the binomial expansion of $(1 - L)^d$ in the lag operator L. Thus, $\Delta_+^d y_t = \Sigma_{i=0}^{t-1} \pi_i(d)\, y_{t-i}$, where $\pi_i(d)$ is the i-th coefficient in that expansion (for more details, see the end of this section). This 'deadstart' fractional process has been popularized, among others, by Robinson and Marinucci (2001), giving rise to Type-II fractional Brownian motion. Since the limit distributions of the EFDF tests discussed throughout this chapter are always Gaussian, none of the results depend on this choice. To simplify the notation, we will omit the truncation subscript in the sequel and refer to this filter simply as Δ^d.

(or $d_A > 0$) or, more realistically, a composite one, $H_1 : d < 1$ (or $d > 0$). We focus here on the latter case where LV (2007) have proved that the use of a T^{κ}-consistent estimate (with $\kappa > 0$) of the true d suffices to obtain a $N(0,1)$ limiting distribution of the resulting test.

Under a sequence of local alternatives approaching $H_0 : d = 1$ from below at a rate of $T^{-1/2}$, LV (2007, Theorem 1) prove that, under Gaussianity, the EFDF test of $I(1)$ vs. $I(d)$ is asymptotically equivalent to the uniformly most powerful invariant (UMPI) test, ie, the LM test introduced by Robinson (1991, 1994) and later adapted by Tanaka (1999) to the time domain. We show that this result also holds for the $I(0)$ vs. $I(d)$ case. Our first contribution here is to analyse the properties of Wald and LM tests in the case of fixed alternatives using the concept of Bahadur's *asymptotic relative efficiency* (ARE; see Gourieroux and Monfort 1995). Although both tests are consistent and diverge at the same rate under fixed alternatives, we find that the EFDF test fares better using Bahadur's ARE criterion in both setups. This is not surprising, given the well-known result about the better power properties of Wald tests in a wide range of models (see Engle 1984). Moreover, when compared to other tests of $I(1)$ or $I(0)$ vs. $I(d)$ which rely on direct inference about semiparametric estimators of d, the EFDF test also exhibits in general better power properties, under a correct specification of the stationary short-run dynamics of the error term in the auxiliary regression. This is due to the fact that the semiparametric estimation procedures often imply larger confidence intervals of the memory parameter, in exchange with less restrictive assumptions on the error term. By contrast, the combination of a wide range of semiparametric estimators for the input value of d with the auxiliary parametric regressions involved in the EFDF test, yields a parametric rate for the Wald tests.[4] Thus, in a sense, Wald tests combine the favourable features of both approaches in improving power while at the same time they reduce the danger of misspecifying short-run dynamics.

Following the development of unit-root tests in the past, we investigate how to implement Wald tests when some deterministic components are considered in the DGP, a case which is not treated in LV (2007). We first focus on the role of a *polynomial trend of known order* since many (macro) economic time-series exhibit this type of trending behaviour. Our main result is that, in contrast with the results for most tests for $I(1)$ against $I(0)$ or vice versa, the EFDF test remains efficient in the presence of deterministic components and maintains the same asymptotic distribution, insofar as they are correctly removed. This result mimics the one found for LM tests when these components are present; cf. Gil-Alaña and Robinson (1997). Next, we examine the cases where there are

[4] LV (2006, 2007) have shown that a Gaussian semiparametric estimator, such as the one proposed by Velasco (1999), suffices to achieve consistency and asymptotic normality in the analysed Wald tests (see sections 12.2 and 12.3 below) extending the results by DGM (2002) about parametric estimators.

structural breaks in the deterministic components, where we devise tests for $I(d)$ cum constant-parameter deterministic terms vs. $I(0)$ cum breaks in these components, or in the long-memory parameter, d, as well as other alternative time-varying schemes for d. Lastly, we show that the previous asymptotic results obtained for DGPs with i.i.d. disturbances remain valid when the error term is allowed to be parametrically autocorrelated, as in the (augmented) ADF setup. In particular, we propose a linear single-step estimation procedure to account for (parametric) AR disturbances which simplifies the two-step procedure proposed by LV (2007).

The rest of the chapter is structured as follows. Section 12.2 briefly overviews the properties of the EFDF tests when the process is either a driftless random walk or i.i.d. under the null, and derives new results about their power relative to the power of the LM test under fixed alternatives. Section 12.3 extends the previous results to processes containing trending deterministic components with constant parameters. Section 12.4 discusses tests to distinguish between $I(0)$ series whose deterministic terms may be subject to structural breaks and $I(d)$ processes with constant parameters. Section 12.5 deals with how to test for breaks in the long-memory parameter, d, as well as some other alternative time varying structures. Section 12.6 explains how to modify the previous tests when the error terms are autocorrelated. Lastly, section 12.7 concludes.

Proofs of the main results are in an Appendix, available as supplementary material to this chapter (see http://dolado-research.blogspot.com/).

In the sequel, the definition of an $I(d)$ process adopted here is the one proposed by Akonom and Gourieroux (1987) where a fractional process is initialized at the origin. This corresponds to Type-II fractional Brownian motion (see the previous discussion in footnote 3) and is similar to the definitions of an $I(d)$ process underlying the LM test proposed by Robinson (1994) and Tanaka (1999). Moreover, the following conventional notation is adopted throughout the chapter: $\Gamma(.)$ denotes the Gamma function, and $\{\pi_i(d)\}$, with $\pi_i(d) = \frac{\Gamma(i-d)}{\Gamma(-d)\Gamma(i+1)}$, represents the sequence of coefficients associated to the binomial expansion of $(1-L)^d$ in powers of L. The indicator function is denoted by $1_{\{.\}}$. Finally, $\xrightarrow{p}$ means convergence in probability, and $\xrightarrow{w}$ denotes weak convergence in $D[0,1]$ endowed with the Skorohod J_1 topology.

12.2 The EFDF Test

12.2.1 *I(1) vs. I(d)*

Following Robinson (1994), we consider an additive model for the process $\{y_t\}_1^T$ which is generated as the sum of a deterministic component, $\mu(t)$, and

a stochastic component, u_t, that is

$$y_t = \mu(t) + u_t, \tag{12.1}$$

where $u_t = \Delta^{-d}\varepsilon_t 1_{\{t>0\}}$ is a purely stochastic $I(d)$ process, with $d \in [0,1]$, and ε_t is a zero-mean i.i.d. random variable.

When $\mu(t) \equiv 0$,[5] DGM (2002) developed a Wald-type (FDF) test for testing the null hypothesis $H_0 : d = 1$ against the composite alternative $H_1 : d \in [0,1)$, based on the t-ratio on ϕ to test $H_0 : \phi = 0$ in the OLS regression model

$$\Delta y_t = \phi \Delta^{d^*} y_{t-1} + v_t, \tag{12.2}$$

where $d^* \geq 0$ is an input value needed to perform the test. If d^* is chosen such that $d^* = \hat{d}_T$, where $\hat{d}_T$ is a T^κ-consistent estimator of d, with $\kappa > 0$, DGM (2002) and LV (2006) have shown that the asymptotic distribution of the resulting t-statistic, t_ϕ, is $N(0,1)$.

Recently, LV (2007) have proposed the EFDF test based on a modification of (12.2) that is more efficient while keeping the good finite-sample properties of Wald tests. Specifically, their proposal is to use the t-statistic, t_φ, associated to $H_0 : \varphi = 0$ in the OLS regression

$$\Delta y_t = \varphi z_{t-1}(d^*) + \varepsilon_t, \tag{12.3}$$

where $z_{t-1}(d^*)$ is defined as[6]

$$z_{t-1}(d^*) = \frac{\left(\Delta^{d^*-1} - 1\right)}{(1-d^*)} \Delta y_t,$$

such that $\varphi = (d^* - 1)$ and $d^* > 0.5$ is an input value needed to implement the test. Note that, if d^* is the true integration order of the process, d, then $\varphi = 0$ under $H_0 : d = 1$ and the model becomes a random walk, ie, $\Delta y_t = \varepsilon_t$. By contrast, under $H_1 : d \in [0,1)$, it holds that $\varphi < 0$, and the model becomes a pure fractional process, ie, $\Delta^d y_t = \varepsilon_t$.

The insight for the higher efficiency of the EFDF test is as follows. Let $d^* = d$. Then under H_1, the regression model in (12.2) can be written as $\Delta y_t = \Delta^{1-d}\varepsilon_t = \varepsilon_t + (d-1)\varepsilon_{t-1} + 0.5d(d-1)\varepsilon_{t-2} + \ldots = \phi\Delta^d y_{t-1} + \varepsilon_t + 0.5d(d-1)\varepsilon_{t-2} + \ldots$ with $\phi = d - 1$. Thus, the error term $v_t = \varepsilon_t + 0.5d(d-1)\varepsilon_{t-2} + \ldots$ in (12.2) is serially correlated. Although OLS provides a consistent estimator of ϕ, since v_t is orthogonal to the regressor $\Delta^d y_{t-1} = \varepsilon_{t-1}$, it is not the most efficient one. By contrast, the regression model used in the EFDF test does not suffer from this problem since, by construction, it yields an i.i.d. error term. In order to distinguish this test from the one proposed in the next subsection for

[5] Alternatively, $\mu(t)$ could be considered to be known. In this case, the same arguments go through after subtracting it from y_t to obtain a purely stochastic process.

[6] A similar model was first proposed by Granger (1986) in the more general context of testing for cointegration with multivariate series, a modification of which has been recently considered by Johansen (2005).

$H_0 : d = 0$, we denote it in the sequel as the EFDF(1) test. Finally, note that application of L'Hôpital rule to $z_{t-1}(d^*)$ in the limit case as $d^* \to 1$ leads to a regressor equal to $-ln(1-L)\Delta y_t = \Sigma_{j=1}^{\infty} j^{-1}\Delta y_{t-j}$, which is the one used in Robinson's LM test (see section 12.2.3).

Theorem 1 in LV (2007), which we reproduce below (as Theorem 12.1) for completeness, establishes the asymptotic properties of t_φ.

Theorem 12.1. *Under the assumption that the DGP is given by* $y_t = \Delta^{-d}\varepsilon_t 1_{\{t>0\}}$, *where* ε_t *is i.i.d. with finite fourth moment, the asymptotic properties of the t-statistic,* t_φ, *for testing* $\varphi = 0$ *in (12.3), where the input of* $z_{t-1}(\hat{d}_T)$ *is a* T^κ*-consistent estimator of* d^*, *for some* $d^* > 0.5$ *with* $\kappa > 0$, *are given by:*

a) Under the null hypothesis ($d = 1$),

$$t_\varphi(\hat{d}_T) \xrightarrow{w} N(0,1).$$

b) Under local alternatives, ($d = 1 - \gamma/\sqrt{T}$),

$$t_\varphi(\hat{d}_T) \xrightarrow{w} N(-\gamma h(d^*), 1),$$

where $h(d^*) = \Sigma_{j=1}^{\infty} j^{-1}\pi_j(d^* - 1)/\sqrt{\Sigma_{j=1}^{\infty}\pi_j(d^* - 1)^2}$, $d^* > 0.5$, $d^* \neq 1$.

c) Under fixed alternatives $d \in [0,1)$, *the test based on* $t_\varphi(\hat{d}_T)$ *is consistent.*

LV (2007) have shown that the function $h(d^*)$ achieves a global maximum at 1 where $h(1) = \sqrt{\pi^2/6}$, and that $h(1)$ equals the noncentrality parameter of the locally optimal LM test (see subsection 12.2.2 below).[7] Thus, insofar as a T^κ-consistent estimator of d, with $\kappa > 0$, is used as an input of $z_{t-1}(d^*)$, the EFDF test is locally asymptotically equivalent to Robinson's LM test.

In practice, the obtained estimate of d could be smaller than 0.5. In these cases, the input value can be chosen according to the following rule: $\tilde{d}_{1T} = \max\{\hat{d}_T, 0.5 + \epsilon\}$, with $\epsilon > 0$, for which the test can be easily proved to diverge under H_1.

A power-rate consistent estimate of d can be easily obtained by applying some available semiparametric estimators. Among them, the estimators proposed by Abadir *et al*. (2005), Shimotsu (2006a), and Velasco (1999) provide appropriate choices since they also cover the case where deterministic components exist, as we do below.

12.2.2 *I(0) vs. I(d)*

Although the EFDF(1) test was originally derived for testing $I(1)$ vs. $I(d)$ processes, it can be easily extended to cover the case of $I(0)$ vs. $I(d)$, with

[7] DGM (2002, Theorem 3) in turn obtained that the corresponding distribution under local alternatives of the FDF test in (12.2) is $N(-\gamma, 1)$. Hence, the asymptotic efficiency of the FDF test relative to the EFDF(1) test is 0.78 ($\simeq \sqrt{6}/\pi$).

$d \in (0,1]$. This new test is labelled as the EFDF(0) test in the sequel. As before, the maintained hypothesis is taken to be (12.1), but now the null is $H_0 : d = 0$, and the composite alternative $H_1 : 0 < d \leq 1$.[8] We first focus on the simple case where $\mu(t) \equiv 0$. Adding and subtracting y_t to both sides of (12.1) and solving for y_t, yields

$$y_t = \psi s_{t-1}(d) + \varepsilon_t, \tag{12.4}$$

where

$$s_{t-1}(d) = \frac{1 - \Delta^d}{d} y_t,$$

such that $\psi = d$. Like in (12.3), $s_{t-1}(d)$ does not contain the current value of y_t since $(1 - \Delta^d) = (dL + \frac{1}{2}d(d-1)L^2 - \ldots)$. Under H_0, $\psi = 0$, while, under H_1, $0 < \psi \leq 1$. When $\psi = 0$, the model is $y_t = \varepsilon_t$ whereas it becomes $\Delta^d y_t = \varepsilon_t$ for $\psi = d \in (0,1]$. As in the $I(1)$ vs. $I(d)$ case, equation (12.4) motivates a test of $H_0 : \psi = 0$ based on the t-statistic of $\hat{\psi}$, t_ψ, computed in a regression of y_t on $s_{t-1}(d^*)$, where d^* is an input value needed to make the test feasible. Thus, the null is tested by means an upper-side test based on t_ψ. As with the EFDF(1) test, the limit case as $d^* \to 0$ implies that $s_{t-1}(d) \to -ln(1-L)y_t = \Sigma_{j=1}^{\infty} j^{-1} y_{t-j}$, which again corresponds to the regressor used in the LM test.

In this case, the following theory holds

Theorem 12.2. *Under the assumption that the DGP is given by $y_t = \Delta^{-d}\varepsilon_t 1_{\{t>0\}}$, where ε_t is i.i.d. with finite fourth moment, the asymptotic properties of the t-statistic, t_ψ, for testing $\psi = 0$ in (12.4) where the input of the regressor $s_{t-1}(\hat{d}_T)$ is a T^κ-consistent estimator of d^*, for some $d^* < 0.5$ with $\kappa > 0$, are given by:*

a) Under the null hypothesis ($d = 0$),

$$t_\psi(\hat{d}_T) \xrightarrow{w} N(0,1).$$

b) Under local alternatives, ($d = \gamma/\sqrt{T}$),

$$t_\psi(\hat{d}_T) \xrightarrow{w} N(\gamma g(d^*), 1),$$

where $g(d^*) = \Sigma_{j=1}^{\infty} j^{-1}\pi_j(d^*)/\sqrt{\Sigma_{j=1}^{\infty}\pi_j(d^*)^2}$, $d^* < 0.5$, $d^* \neq 0$.

c) Under fixed alternatives ($d \in (0,1)$), the test based on $t_\psi(\hat{d}_T)$ is consistent.

It is easy to show that the function $g(.)$ achieves an absolute maximum at 0, in which case $g(0)$ equals the noncentrality parameter of the locally optimal Robinson's LM test. Therefore, if the input of $s_{t-1}(.)$, $\hat{d}_T$, is a T^κ-consistent estimator of d with $\kappa > 0$, the test based on $t_\psi(\hat{d}_T)$ is locally optimal. In

[8] Note that if we were to take a null of $I(d_0)$, $d_0 \in (0,1]$, and an alternative of $I(0)$, the EFDF regression model would be $\Delta^{d_0} y_t = \rho[d_0^{-1}(\Delta^{-d_0} - 1)]\Delta^{d_0} y_t + \varepsilon_t$, with $\rho = -d_0$. In this case, under H_0, $\rho \neq 0$, whereas, under H_1, $\rho = 0$.

practice, to perform regression (12.4) the input value $\widetilde{d}_{0T} = \min\{\hat{d}_T, 0.5 - \epsilon\}$, with $\epsilon > 0$, can be employed so that it is always strictly smaller than 0.5.

12.2.3 *Power Comparisons Under Fixed Alternatives*

As discussed before, the closer competitor to the EFDF test is the LM test proposed by Robinson (1991, 1994) in the frequency domain, subsequently extended by Tanaka (1999) to the time domain. In this section we discuss the power properties of the two competing tests under the case of fixed alternatives in Bahadur's ARE sense.[9]

We start with the LM test, henceforth denoted as LM_T, which considers $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ for the DGP $\Delta^{d_0+\theta} y_t = \varepsilon_t$. In line with the hypotheses considered in this chapter, we focus on the particular cases where $d_0 = 1$ and $-1 < \theta \leq 0$, and $d_0 = 0$ and $0 < \theta \leq 1$. Assuming that $\varepsilon_t \sim n.i.d.\ (0,\sigma^2)$, the score-LM test is computed as

$$LM_T = \sqrt{\frac{6}{\pi^2}}\, T^{1/2} \sum_{j=1}^{T-1} j^{-1} \hat{\rho}_j \xrightarrow{w} N(0,1)\,, \tag{12.5}$$

where $\hat{\rho}_j = \Sigma_{t=j+1}^{T} \Delta^{d_0} y_t \, \Delta^{d_0} y_{t-j} / \Sigma_{t=1}^{T} (\Delta^{d_0} y_{t-j})^2$ (see Robinson, 1991 and Tanaka, 1999). Breitung and Hassler (2002) have shown that an alternative way to compute the LM test is as the t-ratio (t_λ) in the regression

$$\Delta^{d_0} y_t = \lambda x_{t-1}^* + e_t, \tag{12.6}$$

where $x_{t-1}^* = \Sigma_{j=1}^{t-1} \; j^{-1} \Delta^{d_0} y_{t-j}$.

Under a sequence of local alternatives of the type $\theta = 1 - T^{-1/2}\gamma$ with $\gamma > 0$ for $H_0 : d_0 = 1$, the LM_T (or t_λ) test is the UMPI test. However, as discussed earlier, the EFDF(1) is asymptotically equivalent to the UMPI test whenever an appropriate estimator of d, $\hat{d}_T$, is used since the limit case as $\hat{d}_T \to 1$ in the filter $(\Delta^{\hat{d}_T - 1} - 1)/(1 - \hat{d}_T)$ yields the linear filter used in the LM test. Similar arguments hold for the EFDF(0) test, where $H_0 : d_0 = 0$, and $\theta = T^{-1/2}\,\gamma$.

In the rest of this section, we analyse the case with fixed alternatives where, to our knowledge, results are new. In particular, we first derive the noncentrality parameters of two above-mentioned tests under an $I(d)$ alternative where the DGP is assumed to be $\Delta^d y_t = \varepsilon_t$. The permissible ranges of d in this analysis are $d \in [0,1)$ for the EFDF(1) test, and $d \in (0,0.5)$ for the EFDF(0) test.[10] In the

[9] The available results in the literature establish the consistency of the Wald and LM tests and derive their (identical) speed of divergence under fixed alternatives. However, they do not derive the noncentrality parameters as we do below which can be useful to characterize power differences for a *given* sample size.

[10] The intuition for why the two cases differ is that, under a fixed $I(d)$ alternative, the EFDF(1) test proceeds to first-difference the series, so that $\Delta y_t \sim I(d-1)$, and then, all the variables in regression (12.2) are stationary under the alternative hypothesis of $d < 1$. The EFDF(0) treats the series in levels so that $y_t \sim I(d)$ and then, for values of $d > 0.5$, regression

case of the EFDF(1) test, $H_0 : d = 1$ and, hence, $\Delta y_t = \Delta^{-b}\varepsilon_t$ where $b = d - 1 < 0$. Then, the following result holds.

Theorem 12.3. *If $\Delta^d y_t = \varepsilon_t$ with $d \in [0,1)$, the t-statistic, t_φ, associated to the EFDF(1) test satisfies*

$$T^{-1/2} t_\varphi \xrightarrow{p} -\left(\frac{\Gamma(3-2d)}{\Gamma^2(2-d)} - 1\right)^{1/2} := c_{1,EFDF}(d),$$

while, under the same DGP, the LM test defined in (12.5) *satisfies*

$$T^{-1/2} LM_T \xrightarrow{p} -\sqrt{\frac{6}{\pi^2}} \frac{\Gamma(2-d)}{(1-d)\Gamma(d-2)} \sum_{j=1}^{\infty} \frac{\Gamma(j+d-1)}{j\Gamma(j+2-d)} := c_{1,LM}(d), \qquad (12.7a)$$

where $c_{1,EFDF}(d)$ and $c_{1,LM}(d)$ denote the non-centrality parameters under the fixed alternative $H_1 : d \in [0,1)$ of the EFDF(1) and LM tests, respectively.

Secondly, for the EFDF(0) test, $H_0 : d = 0$, whereby now $y_t = \Delta^{-b}\varepsilon_t$ with $b = d$. Then

Theorem 12.4. *If $\Delta^d y_t = \varepsilon_t$ with $d \in (0,0.5)$, the t-statistic, t_ψ, associated to the EFDF(0) test satisfies,*

$$T^{-1/2} t_\psi \xrightarrow{p} \left(\frac{\Gamma(1-2d)}{\Gamma^2(1-d)} - 1\right)^{1/2} := c_{0,EFDF}(d),$$

while, under the same DGP, the LM test defined in (12.5) satisfies

$$T^{-1/2} LM_T \xrightarrow{p} \sqrt{\frac{6}{\pi^2}} \frac{\Gamma(1+d)}{\Gamma(-d)} \sum_{j=1}^{\infty} \frac{\Gamma(j-d)}{j\Gamma(j+d+1)} := c_{0,LM}(d), \qquad (12.7b)$$

where $c_{0,EFDF}(d)$, and $c_{0,LM}(d)$ denote the noncentrality parameters under the fixed alternative $H_1 : d \in (0,0.5)$ of the EFDF(0) and LM tests, respectively.

Figures 12.1 and 12.2 display the two noncentrality parameters of the LM and EFDF derived in Theorems 12.3. and 12.4.. Their squares correspond to the approximate slopes in Bahadur's ARE so that the test with the greater slope is asymptotically more powerful. As expected, they behave similarly for values of d very close to the corresponding null hypotheses. However, despite being devised as the UIMP test for local alternatives, the LM test performs worse than the EFDF tests, for a given sample size, when the alternative is not local: $c_{1,EFDF}(d)$ ($c_{0,EFDF}(d)$) is much more negative (positive) than $c_{1,LM}(d)$ ($c_{0,LM}(d)$) when d departs from its respective nulls. Hence, the ARE ratio $c^2_{i,EFDF}(d)/c^2_{i,LM}(d)$ is larger than unity, favouring the Wald test. Extensive

(12.3) includes both stationary and nonstationary variables. As a result, the LLN can be applied on the EFDF(1) test for all $d < 1$ but only for values of $d < 0.5$ in the EFDF(0) case. If $d > 0.5$, the noncentrality parameter will converge to a random variable.

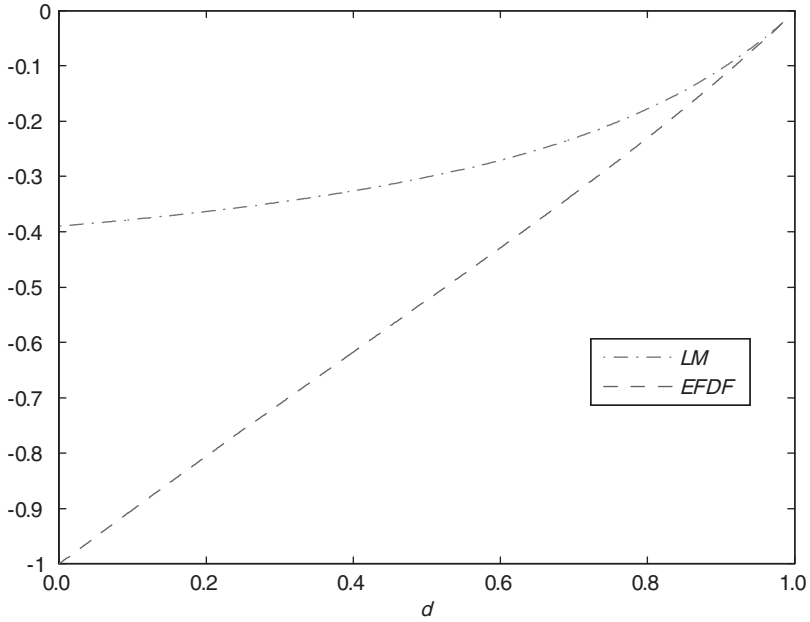

FIG. 12.1. Noncentrality parameters of EFDF(1) and LM tests.

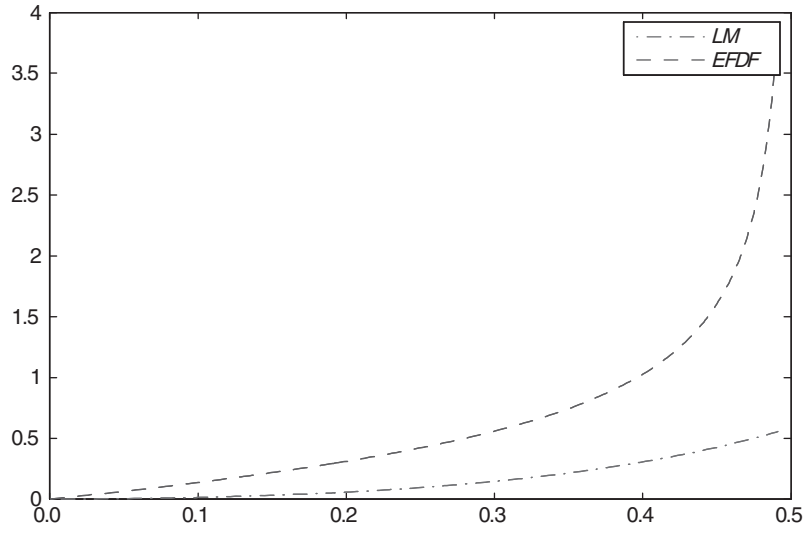

FIG. 12.2. Noncentrality parameters of EFDF(0) and LM tests.

Monte Carlo evidence supporting this better power performance can be found in LV (2007) and DGM (2008). The intuition for the worse power of the LM test is that there is no value for λ in (12.6) that makes e_t both i.i.d. and independent of the regressor for fixed alternatives, implying that x^*_{t-1} does not maximize the correlation with $\Delta^{d_0} y_t$.

As regards the power of semiparametric estimators, whose confidence intervals could be directly used for inference purposes, both the Fully Extended Local Whittle (FELW) (see Abadir *et al.*, 2005) and the Exact Local Whittle estimators (ELW) (see Shimotsu and Phillips, 2005) verify the asymptotic property: $\sqrt{m}(\hat{d}_T - d) \xrightarrow{w} N\left(0, \frac{1}{4}\right)$ for $m = o(T^{\frac{4}{5}})$. For example, test statistics for a unit root are based on $\tau_d = 2\sqrt{m}(\hat{d}_T - 1) \xrightarrow{w} N(0,1)$. Therefore, their rate of divergence under $H_1 : d < 1$ is the nonparametric rate $O_p(\sqrt{m})$ which is smaller than the $O_p(\sqrt{T})$ parametric rate achieved by the Wald test. Of course, this loss of power is just the counterpart of the higher robustness against misspecification achieved by semiparametric tests.

12.3 Deterministic Components without Breaks

In the case where $\mu(t) \neq 0$, DGM (2008) have derived the properties of the EFDF(1) test when the time-series is generated by (12.1) and $\mu(t)$ verifies the following condition.

Condition ET (Evolving trend): $\mu(t)$ is a polynomial in t of known order.

Under Condition *ET*, the DGP is allowed to contain trending regressors in the form of polynomials (of known order) of t. Hence, when the coefficients of $\mu(t)$ are unknown, the test described above is unfeasible. Nevertheless, it is still possible to obtain a feasible test with the same asymptotic properties as in Theorem 12.1. if a consistent estimate of $\mu(t)$ is removed from the original process. Indeed, under H_0, the relevant coefficients of $\mu(t)$ can be consistently estimated by OLS in a regression of Δy_t on $\Delta\mu(t)$. For instance, consider the case where the DGP contains a linear time trend, that is

$$y_t = \alpha + \beta t + \Delta^{-d}\varepsilon_t 1_{\{t>0\}}, \tag{12.8}$$

which, under $H_0 : d = 1$, leads to the popular case of a random walk with drift. Taking first differences, it follows that $\Delta y_t = \beta + \Delta^{1-d}\varepsilon_t 1_{\{t>0\}}$. Then, the OLS estimate of β, $\hat{\beta}$, (ie, the sample mean of Δy_t) is consistent under both H_0 and H_1. In effect, under H_0, $\hat{\beta}$ is a $T^{1/2}$-consistent estimator of β whereas, under H_1, it is $T^{3/2-d}$-consistent with $3/2 - d > 0.5$ (see Hosking, 1996, Theorem 8). Hence, if one uses the regression model

$$\widetilde{\Delta y_t} = \varphi \widetilde{z_{t-1}}\left(\hat{d}_T\right) + e_t, \tag{12.9}$$

where the input of $\widetilde{z_{t-1}}\left(\hat{d}_T\right)$ is a T^{κ}-consistent estimator of d with $\kappa > 0$, $\widetilde{\Delta y_t} = \Delta y_t - \Delta\hat{\mu}(t)$, $\widetilde{z_{t-1}}\left(\hat{d}_T\right) = \frac{\left(\Delta^{\hat{d}_T-1}-1\right)}{(1-\hat{d}_T)}(\Delta y_t - \Delta\hat{\mu}(t))$, and the coefficients of $\Delta\hat{\mu}(t)$ are estimated by an OLS regression of Δy_t on $\Delta\mu(t)$, then the asymptotic properties of the EFDF(1) test in (12.9) are identical to those stated in Theorem 12.1..

A similar result holds for the EFDF(0) test but this time using a T^{κ}-consistent estimator of d, for $d^* < 0.5$, with $\kappa > 0$. In this alternative setup of $I(0)$ vs. $I(d)$, the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ in the regression in levels of y_t on $\mu(t)$ are $T^{1/2-d}$ and $T^{3/2-d}$-consistent estimators of α and β, respectively. Consequently, the estimator of the trend slope, $\hat{\beta}$, is always consistent for $d \in (0,1]$ whereas the estimator of the intercept, $\hat{\alpha}$, is only consistent for $d \in (0,0.5)$, implying that the residuals from the OLS detrending procedure in levels are only valid for $d^* < 0.5$. Under fixed alternatives, since the true value of d could well exceed 0.5, one possibility in order to obtain consistent detrended series is to use Shimotsu's (2006a) detrending approach for $I(d)$ processes. This author notices that if one chooses the initial value of the series, y_1, as an estimator of α, then it holds that the deviations $y_1 - \alpha(= \Delta^{-d}\varepsilon_1 1_{t>0})$ are $O_p(1)$, implying that its variance is dominated by the exploding variance of y_t when $d \in (0.5,1]$. Thus, he recommends to use the above-mentioned FELW estimation procedure to the detrended series in levels $\widetilde{y}_t = \dot{y}_t - \dot{\alpha}(d)$, where $\dot{y}_t = y_t - \hat{\alpha} - \hat{\beta}t$ are the OLS residuals and $\dot{\alpha}(d) = \omega(d)T^{-1}\Sigma\dot{y}_t + [1 - \omega(d)]\dot{y}_1$. Notice that $\dot{\alpha}(d)$ is a weighted average of the two alternative estimators of α earlier discussed with $\omega(d)$ being a smooth (twice continuously differentiable) weight function such that $\omega(d) = 1$ for $d \in (0,0.5)$.[11] Through this alternative detrending procedure, the difference between $\Delta^d\widetilde{y}_t$ and ε_t becomes negligible for any value of $d \in (0,1]$. Therefore, if one considers the regression model

$$\widetilde{\Delta y_t} = \psi\widetilde{s_{t-1}}\left(\hat{d}_T\right) + e_t, \tag{12.10}$$

where the input of $\widetilde{s_{t-1}}(\hat{d}_T)$ is a T^{κ}-consistent estimator of d^*, for some $d^* < 0.5$ with $\kappa > 0$, having used as residuals $(\widetilde{y}_t)$ the ones obtained from an OLS regression of y_t on $\mu(t)$, the asymptotic properties of the EFDF(0) test for testing $\psi = 0$ in (12.10) are identical to those stated in Theorem 12.2.. Likewise, under fixed alternatives, a similar result holds for cases where $d \in (0.5,1)$, this time using Shimotsu's (2006a) residuals and $\widetilde{d}_{0T}$, as an alternative estimator of the corresponding input value of the regressor $\widetilde{s_{t-1}}(.)$.

12.4 Deterministic Components with Breaks

Next we extend the EFDF tests to cover the case where the deterministic component, $\mu(t)$, of the time-series y_t in (12.1) is possibly subject to structural

[11] An example of $\omega(d)$ for $d \in (0.5,1)$ is $(1/2)[1 - cos\ \pi d]$.

breaks, denoted hereafter as $\mu_B(t)$. One possibility is to consider breaks both under the null and the alternative hypotheses discussed in sections 12.2.1 and 12.2.2. In this case, similar two-stage procedures to those described in section 12.3 could be applied.[12] However, it is well known in the statistical literature that some features of long-range dependence (LRD) can be generated by either the process being $I(d)$ with smooth deterministic components or by an $I(0)$ process subject to breaks; see, eg, Bhattacharya *et al*. (1983), Mikosch and Starica (2004), and Berkes *et al*. (2006). Indeed, these studies show that conventional statistics designed to detect long range dependence behave similarly under weak dependence with change-points.[13]

For this reason we focus in the sequel on the pure distinction between these two *alternative* models that can account for the observed strong persistence of y_t: (i) u_t is an $I(d)$ process, with $d \in (0,1)$ and $\mu(t)$ is smooth, and (ii) u_t is a short-memory $I(0)$ process and $\mu(t)$ is subject to breaks. The EFDF approach, where one of the hypotheses encompasses the other, cannot directly accommodate these two types of models. This is so since the $I(d)$ hypothesis clearly nests the $I(0)$ one, but then the $\mu(t)$ component cannot nest $\mu_B(t)$ component at the same time. We therefore follow a comprehensive model approach, whereby non-nested models are tested within an artificially constructed general model that includes them as special cases. This approach was advocated by Atkinson (1970) and later taken up under a different guise by Davidson and MacKinnon (1981) in developing their J-test. In effect, let us think of two alternative models, denoted as $M1$ and $M2$, respectively, defined as follows

$$M1 : y_t = \mu_B^i(t) + \varepsilon_t, \tag{12.11}$$

and

$$M2 : y_t = \mu(t) + \Delta^{-d}\varepsilon_t 1_{t>0}, \text{ with } d \in (0,1), \tag{12.12}$$

where $\mu_B^i(t)$ is a linear deterministic trend function that may contain breaks at known or unknown dates (in principle, just a single break at date T_B would be considered) while $\mu(t)$ does not contain breaks. In line with Perron (1989),

[12] For tests of I(1) vs. I(0) with breaks under both the null and the alternative, see Banerjee and Urga (2005), and Kim and Perron (2006). Extensions of these tests to a fractional setup can be found in DGM (2005), Mayoral (2006), and Shimotsu (2006b).

[13] More recently, a similar issue has re-emerged in the econometric literature dealing with financial data. For example, Ding and Granger (1996), and Mikosch and Starica (2004) claim that the stochastic components of both the absolute and the squared returns of financial prices (bonds, exchange rates, options, etc.) are $I(0)$ and explain the evidence about LRD found in the data as spuriously induced by structural breaks in the parameters of the deterministic components over different subsamples due to significant events, such as the Great Depression of 1929 or the oil-price shocks in the 1970s. On the contrary, Lobato and Savin (1998) conclude that the LRD evidence found in the squared returns is genuine and, thus, not a spurious feature of structural breaks.

three definitions of $\mu_B^i(t)$, $i \in \{A, B, C\}$ will be considered,

$$Case\ A: \ \mu_B^A(t) = \mu_0 + (\mu_1 - \mu_0)DU_t(\omega_B), \tag{12.13}$$

$$Case\ B: \ \mu_B^B(t) = \mu_0 + \beta_0 t + (\beta_1 - \beta_0)DT_t^*(\omega_B), \tag{12.14}$$

$$Case\ C: \ \mu_B^C(t) = \mu_0 + \beta_0 t + (\mu_1 - \mu_0)DU_t(\omega_B) + (\beta_1 - \beta_0)DT_t(\omega_B). \tag{12.15}$$

Case A corresponds to the *crash* hypothesis, case B to the *changing growth* hypothesis and case C to a combination of both. The dummy variables are defined as follows: $DU_t(\omega_B) = 1_{(T_B+1\leq t\leq T)}$, $DT_t^*(\omega_B) = (t - T_B)1_{(T_B+1\leq t\leq T)}$ and $DT_t(\omega_B) = t1_{(T_B+1\leq t\leq T)}$ where $\omega_B = T_B/T$ is a fixed value belonging to the subset of the interval (0,1) that describes the relative location of the break in the sample.

Then, noticing that $M2$ can be rewritten as

$$M2: y_t = y_t - \Delta^d[y_t - \mu(t)] + \varepsilon_t, \tag{12.16}$$

one could follow Davidson and MacKinnon (1981) in considering the following linear combinations of $M1$ and $M2$

$$y_t = (1 - \zeta)\mu_B^i(t) + \zeta\{y_t - \Delta^d[y_t - \mu(t)]\} + \varepsilon_t, \tag{12.17}$$

or

$$y_t = (1 - \zeta)\{y_t - \Delta^d[y_t - \mu(t)]\} + \zeta\mu_B^i(t) + \varepsilon_t, \tag{12.18}$$

so that two J-tests can be applied, depending on whether $M1$ or $M2$ is considered to be the null hypothesis. In the case where $M1$ is taken to be H_0 and $M2$ to be H_1, the unknown parameters in $\{y_t - \Delta^d[y_t - \mu(t)]\}$ are not identified under H_0 since $\zeta = 0$. A solution of this problem is to replace the term $\{y_t - \Delta^d[y_t - \mu(t)]\}$ in (12.17) by $\{y_t - \Delta^{\hat{d}_T}[y_t - \hat{\mu}(t)]\}$, where $\hat{d}_T$ and $\hat{\mu}(t)$ are consistent under H_1, eg, using Shimotsu's (2006a) estimation procedure described in section 12.3. Hence, the following regression can be estimated

$$y_t = \mu_B^{*i}(t) + \zeta\widetilde{\nu_{t-1}} + \varepsilon_t, \tag{12.19}$$

where $\widetilde{\nu_{t-1}} = \{y_t - \Delta^{\hat{d}_T}[y_t - \hat{\mu}(t)]\}$ and $\mu_B^{*i}(t) = (1 - \zeta)\mu_B^i(t)$. Under H_0, it follows that $\zeta = 0$, and this hypothesis can be tested using a t-test on the coefficient of $\widetilde{\nu_{t-1}}$, t_ζ. We will denote this test as the EFDF(B) test.

Conversely, if one chooses $M2$ to be H_0 and $M1$ to be H_1, the corresponding regression model becomes

$$\Delta^d y_t = \Delta^d \mu^*(t) - \zeta\widetilde{\nu_{t-1}}(d) + \varepsilon_t, \tag{12.20}$$

where now $\widetilde{\nu_{t-1}}(d) = \{y_t - \Delta^d y_t - \widehat{\mu_B^i}(t)\}$, $\mu^*(t) = (1 - \zeta)\mu(t)$ where d is taken to be knowns under the null, and $\widehat{\mu_B^i}(t)$ is estimated in a preliminary regression under the alternative of $I(0)$ cum breaks. Again, under H_0, we have that $\zeta = 0$, and a t-test, t_ζ, could be used to test for this hypothesis.

For simplicity, we have operated above as if the break dates were known in regressions (12.17) and (12.18). The more realistic case of unknown breaks when y_t is $I(0)$, are under current investigation following Bai's (1997) or Bai and Perron's (1998) procedures.

Finally, notice that, because non-nested hypothesis tests are designed as specification tests, rather than as procedures for choosing among competing models, it is not at all surprising that sometimes they do not lead us to choose one model over the other. If we would simply wish to choose the *best* model between M1 and M2, one could use some information criteria that help to discriminate between them. This approach is also in our current research agenda.

12.5 Breaks in the Long-Memory Parameter

Granger and Ding (1996) were the first to analyse the consequences of having a variable memory parameter d.[14] They consider two possible scenarios: (i) d_t is a stochastic process, eg, an $AR(1)$ process with mean $\overline{d}$, and (ii) d_t switches between two regimes, eg, $y_t = \lambda_t x_{1t} + (1 - \lambda_t)x_{2t}$, with $x_{1t} \sim I(d_1)$, $x_{2t} \sim I(d_2)$ and λ_t following a 0–1 Markov switching process. Since this chapter is focused on testing, we consider a different setup. The memory parameter d can take two values, d_1 in a first given proportion of the sample and d_2 in the remaining proportion.

Both stationary and nonstationary fractional roots are considered. Although it is not difficult to generalize the analysis to allow for breaks in the deterministic components as well as short-term correlation in the disturbance terms, for simplicity we will focus in the sequel only on the case where the error terms are i.i.d. and no deterministic terms are present. More specifically, we assume that y_t is generated as

$$(1 - L)^{d_0 + \theta D_t(\omega_B)} y_t = \varepsilon_t 1_{t>0}, \tag{12.21}$$

so that y_t is a zero-mean integrated process (with an integer or fractional integration order), that can be either stationary or nonstationary. The order of integration of y_t is allowed to change along the sample at time T_B, with the dummy variable $D_t(\omega_B)$ taking a value equal to 1 if $\omega_B T < t$ and zero otherwise. Then, the process y_t is $I(d + \theta)$ until T_B and $I(d)$ after T_B, where θ can be either larger or smaller than zero.

Under H_0, no change in persistence occurs and therefore $H_0 : \theta = 0$. By contrast, under H_1, a change in persistence occurs at time T_B, that is $H_1 : \theta < 0$ or

[14] Detecting a change in the persistence of a process is usually tackled in the context of AR processes within the $I(0)/I(1)$ framework (see, eg, Busetti and Taylor, 2004), later extended by Hassler and Scheithauer (2007) to $I(0)/I(d)$, $d > 0$. Nevertheless, as argued above, this framework can be too narrow in many empirical applications.

$H_1' : \theta > 0$, where the first (second) case corresponds to an increase (decrease) in persistence after T_B.

Since, to our knowledge, the LM tests have not been used so far to test this type of hypothesis, we start by deriving such a test in the present setup. Under Gaussianity, recall that Tanaka's (1999) time-domain version of the LM statistic for testing $H_0 : d = d_0$ vs. $H_1 : d \neq d_0$ uses the the log-likelihood

$$L\left(\theta,\sigma^2,\omega_B\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{(1-L)^{d_0+\theta D_t(\omega_B)}\, y_t\right\}^2 \tag{12.22}$$

Thus, an LM test for $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$ rejects H_0 for large values of

$$LM_T = \left.\frac{\partial L\left(\theta,\sigma^2,\omega_B\right)}{\partial\theta}\right|_{H_0:\theta=0,\sigma^2=\hat{\sigma}^2} = -\frac{1}{\hat{\sigma}^2}\sum_{t=1}^{T}\left(\left\{\log(1-L)\times D_t(\omega_B)\,\Delta^{d_0}y_t\right\}\Delta^{d_0}y_t,\right.$$

where the estimated variance is $\hat{\sigma}^2 = T^{-1}\Sigma(\Delta^{d_0}y_t)^2$.

Since $log(1-L) = -\left(L + L^2/2 + L^3/3 + \ldots\right)$, and $D_t(\omega_B) = 0$ for $t > T_B$, then

$$LM_T = \frac{1}{\hat{\sigma}^2}\sum_{t=2}^{T_B}\left(\sum_{k=1}^{t-1}\frac{1}{k}\hat{\varepsilon}_{t-k}\right)\hat{\varepsilon}_t = T\sum_{t=2}^{T_B}\frac{\left(\sum_{k=1}^{t-1}\frac{1}{k}\hat{\varepsilon}_{t-k}\hat{\varepsilon}_t\right)}{\sum_{t=1}^{T}\hat{\varepsilon_t}^2} = T\sum_{k=1}^{T_B-1}\frac{1}{k}\hat{\rho}_k^*\left(\hat{\varepsilon}_t\right), \tag{12.23}$$

where $\hat{\rho}_k^*\left(\hat{\varepsilon}_t\right) = \sum_{t=k+1}^{T_B}\hat{\varepsilon}_{t-k}\hat{\varepsilon}_t/\sum_{t=1}^{T}\hat{\varepsilon}_t^2$. Notice that in finite samples $\hat{\rho}_k^*\left(\hat{\varepsilon}_t\right)$ is not identical to the k-th autocorrelation of residuals since in order to compute the numerator, only observations previous to the break are considered whereas all observations are employed to compute the denominator. This difference vanishes asymptotically. The following theorem describes the asymptotic properties of the test under local alternatives when the break date is known.

Theorem 12.5. *Under the hypothesis of $\theta = \delta/\sqrt{T}$, for a known value of T_B and a fixed δ it holds that, as $T \to \infty$,*

$$T^{-1/2}S_T\left(\omega_B\right) = \sqrt{T}\sum_{k=1}^{T_B-1}\frac{1}{k}\hat{\rho}_k^*\left(\hat{\varepsilon}_t\right) \xrightarrow{w} N\left(\frac{\pi^2}{6}\delta, \frac{\pi^2}{6}\omega_B\right). \tag{12.24}$$

Note that, since $\omega_B < 1$, the variance of this distribution is smaller than the variance in the case where no break occurs. This reflects the fact that only a fraction of the data is employed but the data is divided by $\sqrt{T}$. Along the lines of Tanaka (1999), it can also be shown that the test statistic proposed in (12.23) is locally optimal.

An EFDF test, denoted as EFDF(Bd), can also be constructed for this case. Following the derivations in section 12.2, one could consider the following maintained hypothesis

$$\Delta^{d_0}y_t = [1 - \Delta^{\theta D_t(\omega_B)}]\Delta^{d_0}y_t + \varepsilon_t, \tag{12.25}$$

which can be expressed in EFDF format as

$$\Delta^{d_0} y_t = \vartheta \left[\frac{1 - \Delta^{\theta}}{\theta} \right] \Delta^{d_0} y_t D_t(\omega_B) + \varepsilon_t, \tag{12.26}$$

where $\vartheta = \theta$. Thus, conditional upon the choice of ω_B, the EFDF(Bd) test would test $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ by means of a two-sided test based on the t-ratio, t_ϑ, which is estimated with observations $1, \ldots, T\omega_B$, and whose asymptotic distribution, under the null, would be $N(0,1)$ and, under local alternatives, satisfies Theorem 12.5.. To construct the regressor in (12.26), the first step is to apply the 'deadstart' filter $\theta^{-1}[\Delta^{d_0} - \Delta^{d_0+\theta}]$ to $\{y_t\}_{t=1}^{T}$; next, the resulting filtered series is truncated to the first subsample by means of the dummy variable, $D_t(\omega_B)$. If θ is taken to be unknown, one could use a T^{κ}-consistent estimator of d from the first subsample and subtract it from d_0 using any of the estimation procedures discussed above.

This way of testing breaks in the long-memory parameter opens the possibility of testing a wide set of other alternative explanations for time varying long-memory behaviour. For instance, inspired by Granger and Ding (1996), the changes in d could be triggered by a strictly stationary and ergodic variable W_t that characterizes different regimes of the economy. More concretely, we are interested on testing $H_0 : y_t \sim I(d)$ versus $H_A : y_t \sim I(d)$ when $W_{t-1} \leq r$ and $I(d + \theta)$ when $W_{t-1} > r$. Substituting the structural break dummy $D_t(\omega_B)$ by the threshold dummy $I(W_{t-1} > r)$ in (12.26) and running the regression

$$\Delta^{d_0} y_t = \vartheta \left[\frac{1 - \Delta^{\theta}}{\theta} \right] \Delta^{d_0} y_t I(W_{t-1} > r) + \varepsilon_t, \tag{12.27}$$

where $\vartheta = \theta$, the corresponding EFDF test for threshold long memory, denoted by EFDF(Td), is a simple two-sided test based on the t-ratio, t_ϑ, whose asymptotic distribution, under the null, would be $N(0,1)$ assuming r is known (eg, $r = 0$). Further issues stemming from an unknown r are beyond the scope of this chapter and are subject to current investigation by the authors.

12.6 Allowing for Serial Correlation

Lastly, we generalize the DGPs considered in section 12.2 to the case where u_t follows a stationary linear $AR(p)$ process, namely, $\Phi_p(L)u_t = \epsilon_t 1_{t>0}$ with $\Phi_p(L) = 1 - \phi_1 L - \ldots \phi_p L^p$ and $\Phi_p(z) \neq 0$ for $|z| \leq 1$. This motivates the following nonlinear regression model

$$\Delta^{d_0} y_t = \varphi[\Phi_p(L) x_{t-1}(d)] + \sum_{j=1}^{p} \phi_j \Delta^{d_0} y_{t-j} + \varepsilon_t, \tag{12.28}$$

where $x(.) = z(.)$ or $s(.)$, for the EFDF(1) and EFDF(0) test, respectively. The new model is similar to (12.3) and (12.4), except for the inclusion of the lags

of $\Delta^{d_0} y_t$ and for the filter $\Phi_p(L)$ in the regressor $x_{t-1}(d)$. Estimation of this model is cumbersome due to the nonlinearity in the parameters φ and $\Phi = (\phi_1, \ldots, \phi_p)$. Compared with the i.i.d. case, LV (2007) claim that a practical problem arises because the vector Φ is unknown and therefore the regressor $[\Phi_p(L)x_{t-1}(d)]$ is unfeasible. For this reason, they recommend applying the following two-step procedure that allows one to obtain efficient tests also with autocorrelated errors.

Assuming, for simplicity, that $\mu(t) \equiv 0$ (or known),[15] in the first step, the coefficients of $\Phi_p(L)$ are estimated (under H_1) by OLS in the equation $\Delta^{\hat{d}_T} y_t = \sum_{t=1}^{p} \phi_j \Delta^{\hat{d}_T} y_{t-j} + a_t$, where $\hat{d}_T$ satisfies the conditions stated in Theorems 12.1. and 12.2.. The estimator of $\Phi_p(L)$ is consistent with a convergence rate which depends on the rate κ. The second step consists of estimating by OLS the equation $\Delta^{d_0} y_t = \varphi[\hat{\Phi}_p(L)x_{t-1}(\hat{d}_T)] + \sum_{j=1}^{p} \phi_j \Delta^{d_0} y_{t-j} + v_t$, where $\hat{\Phi}_p(L)$ is the estimator from the first step, and $\hat{d}_T$ denotes the same estimated input used in that step as well. As LV (2007, Theorem 2) have shown, for the $I(1)$ vs. $I(d)$ case, the t_φ statistic in this augmented regression is still both normally distributed and locally optimal, but a similar argument applies to the $I(0)$ vs. $I(d)$ case. The tests will be denoted as AEFDF(i), $i = 1,0$, (augmented EFDF) tests in the sequel.

However, in DGM (2008) we claim that a feasible single-step procedure in the case of the AEFDF(1) test can also be applied with the same properties. In effect, under H_1, the process would be $\Phi_p(L)\Delta^d y_t = \varepsilon_t$, so that adding and subtracting the process under H_0, $\Phi_p(L)\Delta y_t$, it becomes

$$\Delta y_t = \varphi[\Phi_p(L)z_{t-1}(d)] + [1 - \Phi_p(L)]\Delta y_t + \varepsilon_t. \tag{12.29}$$

The one-step method we propose is based on the following decomposition of the lag polynomial $\Phi_p(L)$

$$\Phi_p(L) = \Phi_p(1) + \frac{1}{\Delta^{d-1} - 1}\Phi_p^*(L), \tag{12.30}$$

where the polynomial $\Phi_p^*(L)$ is defined by equating (12.30) to the standard polynomial decomposition

$$\Phi_p(L) = \Phi_p(1) + \Delta\widetilde{\Phi}_p(L). \tag{12.31}$$

Hence

$$\Phi_p^*(L) = (\Delta^d - \Delta)\widetilde{\Phi}_p(L) = \Delta^d\widetilde{\Phi}_p(L) - [\Phi_p(L) - \Phi_p(1)]. \tag{12.32}$$

Substitution of (12.32) into (12.29), using (12.30) and noticing that $\varphi = d - 1$, $\Phi_p(1) + \widetilde{\Phi}_p(0) = 1$ and $z_{t-1}(d) = \frac{\Delta^{d-1}-1}{1-d}\Delta y_t$, yields after some simple

[15] For the case where the coefficients of $\mu(t)$ are considered to be unknown, a similar procedure as that described in section 12.2.1 can be implemented and efficient tests will still be obtained.

algebra

$$\Delta y_t = \varphi[\Phi_p(1)]\ z_{t-1}(d) - \widetilde{\Phi}_p(L)[\Delta^d - 1]\ \Delta y_t - [\widetilde{\Phi}_p(L) - \widetilde{\Phi}_p(0)]\ \Delta y_t + \varepsilon_t. \quad (12.33)$$

where notice that the second and third regressors are predetermined since $(\Delta^d - 1)$ and $[\widetilde{\Phi}_p(L) - \widetilde{\Phi}_p(0)]$ do not include contemporaneous values of Δy_t. Hence, a one-step procedure can be implemented in a regression of Δy_t on $z_{t-1}(d)$, contemporaneous and lagged values of $[\Delta^d - 1]\ \Delta y_t$ and lags of Δy_t, by means of a t-test on the coefficient of $z_{t-1}(d)$. For example, in the case of an AR(1) disturbance, ie, $\Phi_1(L) = 1 - \phi L$, we have that $\Phi_1(1) = 1 - \phi$ and $\widetilde{\Phi}_1(L) = \widetilde{\Phi}_1(0) = \phi$, so that (12.33) becomes

$$\Delta y_t = \varphi(1 - \phi)z_{t-1}(d) - \phi[\Delta^d - 1]\Delta y_t + \varepsilon_t. \quad (12.34)$$

A similar one-step testing procedure can be used for the AEFDF(0) test. In effect, adding and subtracting the process under H_0 to the process under H_1, yields

$$y_t = \psi[\Phi_p(L)s_{t-1}(d)] + [1 - \Phi_p(L)]y_t + \varepsilon_t. \quad (12.35)$$

Then, using the decompositions

$$\Phi_p(L) = \Phi_p(0) + \frac{1}{\Delta^d - 1}\Phi_p^*(L), \quad (12.36)$$

$$\Phi_p(L) = \Phi_p(0) + L\Phi_{p-1}(L), \quad (12.37)$$

and operating, yields

$$y_t = \psi s_{t-1}(d) - \frac{\Phi_{p-1}(L)}{\Phi_p(0)}\Delta^d y_{t-1} + \frac{1}{\Phi_p(0)}\varepsilon_t, \quad (12.38)$$

which for the illustrative case of an AR(1) disturbance, ie, $\Phi_1(L) = 1 - \phi L$, becomes

$$y_t = \psi s_{t-1}(d) + \phi\Delta^d y_{t-1} + \varepsilon_t. \quad (12.39)$$

Following LV (2007), one can show that the asymptotic properties of the two single-step AEFDF(i=1,2) tests above are identical to those in Theorems 12.1. and 12.2., except that, under local alternatives ($d = 1 - \gamma/\sqrt{T}$ for AEFDF(1) and $d = \gamma/\sqrt{T}$ for AEFDF(0), with $\gamma > 0$), we have that $t_\varphi(d) \xrightarrow{w} N(-\gamma\omega,1)$ and $t_\psi(d) \xrightarrow{w} N(\gamma\omega,1)$ where

$$\omega^2 = \frac{\pi^2}{6} - \varkappa'\Psi^{-1}\varkappa, \quad (12.40)$$

such that $\varkappa = (\varkappa_1, \ldots, \varkappa_p)'$ with $\varkappa_k = \sum_{j=k}^{\infty} j^{-1}c_{j-k}$, $k = 1, \ldots, p$, c_j's are the coefficients of L^j in the expansion of $1/\Phi(L)$, and $\Psi = [\Psi_{k,j}]$, $\Psi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|}$, $k, j = 1, \ldots, p$, denotes the Fisher information matrix for $\Phi(L)$

under Gaussianity. Note that ω^2 is identical to the drift of the limiting distribution of the LM test under local alternatives (see Tanaka, 1999). The use of semiparametric estimators for d is very convenient here, since one can be agnostic about a parametric specification of the autocorrelation in the error terms when estimating the input value of d. Although it has not been proved yet, we conjecture that the single-step procedure can be generalized to deal with ARMA processes, rather than AR ones, by increasing the number of regressors in (12.33) or (12.38) at a certain rate, along the lines of DGM (2002, Theorem 7).

12.7 Concluding Remarks

Long-memory processes have become a very attractive research topic in econometrics during the last few years, due both to their flexibility and realistic microfoundations. Indeed, they received a lot of attention from the theoretical viewpoint but, in our opinion, so far this has not been sufficiently reflected in empirical work. There must be several reasons for this disconnection. We believe that one of them is that empirical researchers have found difficulties in implementing many of those theoretical results. Thus, our main goal in this chapter has been to frame the long-memory testing procedures in a setup somewhat equivalent to the nowadays familiar unit roots testing approach (à la Dickey–Fuller): t-statistics in simple time-domain regressions, with known conventional asymptotic distributions and easy to implement using standard econometrics softwares. Although our illustrations have focused on univariate processes, extensions to fully-fledged multivariate models should not be hard to derive. For example, a first try at applying the Wald test principles to the reduced-rank analysis in a system of $I(1)$ processes with fractional cointegrating relationships of order $(1 - b)$, $b \in [0,0.5)$, can be found in Avarucci and Velasco (2007).

References

Akonom, J. and Gourieroux, C. (1987). A functional limit theorem for fractional processes. CEPREMAP, Mimeo.

Abadir, K., Distaso, W., and Giraitis, L. (2005). Semiparametric estimation for trending I(d) and related processes. Imperial College, Mimeo.

Atkinson, A. (1970). A method for discriminating between models (with discussion). *Journal of the Royal Statistical Society*, B, B32, 323–353.

Avarucci, M. and Velasco, C. (2007). A Wald test for the cointegration rank in nonstationary fractional systems. Universidad Carlos III, Mimeo.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, **79**, 551–563.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.

Banerjee, A. and Urga, G. (2005). Modelling structural breaks, long memory and stock market volatility: An overview. *Journal of Econometrics*, **129**, 1–34.

Berkes, I., Hovarth, L., Kokoszka, P., and Shao, Q-M. (2006). On discriminating between long-range dependence and changes in mean. *Annals of Statistics*, **34**, 1140–1165.

Bhattacharya, R. N, Gupta, V. K., and Waymire, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, **20**, 649–662.

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

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

Davidson, R. and Mackinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypothesis. *Econometrica*, **49**, 781–793.

Diebold, F. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, **105**, 131–159.

Ding, Z. and Granger, C. W. J. (1996). Varieties of long memory models. *Journal of Econometrics*, **73**, 61–77.

Dolado, J., Gonzalo, J., and Mayoral, L. (2002). A fractional Dickey–Fuller test for unit roots. *Econometrica*, **70**, 1963–2006.

——————(2005). Structural breaks vs. long memory: What is what? Universidad Carlos III, Mimeo.

——————(2008). Wald tests of I(1) against I(d) alternatives: Some new properties and an extension to processes with trending components. *Studies in Nonlinear Dynamics and Econometrics*, **12**(4), 1–32.

Engle, R. F. (1984). Wald, likelihood ratio and Lagrange multiplier tests in econometrics. In Griliches, Z. and Intrilligator, R. (eds.), *Handbook of Econometrics*, vol II, pp. 75–826. Amsterdam: North Holland.

Gil-Alaña, L. A. and Robinson, P. (1997). Testing unit roots and other statistical hypotheses in macroeconomic time series. *Journal of Econometrics*, **80**, 241–268.

Gonzalo, J. and Lee, T. (1998). Pitfalls in testing for long-run relationships. *Journal of Econometrics*, **86**, 129–154.

Gourieroux, C. and Monfort, A. (1995). *Statistics and Econometric Models*, Volume II. Cambridge: Cambridge University Press.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14**, 227–238.

——(1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, **48**, 213–228.

——and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, **73**, 61–77.

Hassler, U. and Scheithauer, J. (2007). Testing against a change from short to long memory. Mimeo.

Hosking, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, **73**(1), 261–284.

Johansen, S. (2005). A representation theory for a class of vector autoregressive models for fractional processes. University of Copenhagen, Mimeo.

Kim, D. and Perron, P. (2006). Unit roots tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. Boston University, Mimeo.

Lo, A. W. and Haubrich, J. G. (2001). The sources and nature of long-term dependence in the business cycle. *Economic Review*, **37**, 15–30.

Lobato, I. and Savin, G. (1998). Real and spurious long memory properties of stock market data. *Journal of Business and Economic Statistics*, **16**, 261–268.

—— and Velasco, C. (2006). Optimal fractional Dickey–Fuller tests for unit roots. *Econometrics Journal*, **9**, 492–510.

—— —— (2007). Efficient Wald tests for fractional unit roots. *Econometrica*, **75**, 575–589.

Mayoral, L. (2006). Testing for fractional integration versus short memory with trends and structural breaks. Universitat Pompeu Fabra, Mimeo.

Michelacci, C. and Zaffaroni, P. (2000). Fractional beta convergence. *Journal of Monetary Economics*, **45**, 129–153.

Mikosch, T. and Starica, C. (2004). Nonstationarities in financial time series, long range dependence and the IGARCH model. *Review of Economics and Statistics*, **86**, 378–390.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–1401.

Parke, W. R. (1999). What is fractional integration? *Review of Economics and Statistics*, **81**, 632–638.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, **5**, 163–168.

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

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

—— and Marinucci, D. (2001). Narrow-band analysis of nonstationary processes. *Annals of Statistics*, **29**, 947–976.

Shimotsu, K. (2006a). Exact local Whittle estimation of fractional integration with unknown mean and time trend. Queen's Economics Dept. Working Paper 1061.

—— (2006b), Simple (but effective) tests of long memory versus structural breaks, Queen's Economics Dept. Working Paper No. 1101.

—— and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, **33**, 1890–1933.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory*, **15**, 249–264.

Velasco, C. (1999). Non-stationary log-periodogram regression. *Journal of Econometrics*, **91**, 325–371.

13

When is a Time-Series I(0)?

*James Davidson**

13.1 Introduction

Since the inception of integrated time-series modelling in econometrics, the question of what constitutes a 'nonintegrated' 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 is 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 chapter 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 chapter and its predecessor.

13.2 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, pp. 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 is easy to dispense with one while retaining another. The inclusion of deterministic components (eg '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 13.1. *A time series* $\{x_t\}_{t=1}^{\infty}$ *is I(0) if the partial sum process* X_T *defined on the unit interval by*

$$X_T(\xi) = \omega_T^{-1} \sum_{t=1}^{[T\xi]} (x_t - E\,x_t), 0 < \xi \leq 1 \tag{13.1}$$

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 (1994) 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 (2002) 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 and Granger (1987) and Johansen (1995), respectively, form only a small subset.

13.3 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 13.1 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

$$x_t = \sum_{j=0}^{\infty} a_j u_{t-j}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty, \quad u_t \sim \text{i.i.d.}(0,\sigma^2). \tag{13.2}$$

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 (13.2) 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, eg 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 13.1. The cases of ARCH and GARCH models, bilinear models and SETAR models, among others, are analysed in Davidson (2002).

The obvious difficulty with Definition 13.1 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 (ie, 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 13.1.

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

$$Y(s) = \frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} g_k \sin(\pi s(k - \tfrac{1}{2}))\xi_k, \quad s \in [0,1] \tag{13.3}$$

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 $k = 1, \ldots, n$, 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 13.1, 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.

13.4 Testing I(0)

Testing the hypothesis embodied in Definition 13.1 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 13.1. 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.

13.4.1 *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 13.1, 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

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

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.

13.4.2 *'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 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 (13.2). 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 in 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 $\left\|A_k^m - A^m\right\| \to 0$ if

[2] Davidson and Sibbertsen (2009) suggest a pre-test for bias.

$\|A_k - A\| \to 0$. Also note that if $\gamma_{km} = E(X_{kt}X_{k,t-m})$ then for each $m \geq 0$,

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

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.

13.4.3 *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 non-negative 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.

13.4.4 *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]

13.5 Fingerprinting I(0)

The literature surveyed in this chapter 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 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 13.1, 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

[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.

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 $v_t \sim$ 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

$$\Xi_0 = \frac{\int_0^1 \left(\int_0^\tau W(s)ds - \tau\int_0^1 W(s)ds\right)^2 d\tau}{\int_0^1 W(\tau)^2 d\tau - \left(\int_0^1 W(\tau)d\tau\right)^2}. \tag{13.4}$$

Breitung points out that under the alternative hypothesis $u_t \sim$ 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 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, 2008). However, consider the case where v_t is $I(d)$ for $d > 0$. 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

$$\Xi_d = \frac{\int_0^1 \left(\int_0^\tau W_d(\zeta)d\zeta - \tau \int_0^1 W_d(s)ds\right)^2 d\tau}{\int_0^1 W_d(\tau)^2 d\tau - \left(\int_0^1 W_d(\tau)d\tau\right)^2}. \tag{13.5}$$

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

$$\Xi_d = \frac{\int_0^1 \left(\int_0^\tau \int_0^\alpha W_{d-1}(\zeta)d\zeta d\alpha - \tau \int_0^1 \int_0^\alpha W_{d-1}(\zeta)d\zeta d\alpha\right)^2 d\tau}{\int_0^1 \left(\int_0^\tau W_{d-1}(\zeta)d\zeta\right)^2 d\tau - \left(\int_0^1 \int_0^\tau W_{d-1}(\zeta)d\zeta d\tau\right)^2}. \tag{13.6}$$

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 distributions have been tabulated by simulation for four values of d, using 1000 NID(0,1) drawings to represent the v_s (see Figure 13.1). While any I(0) process v_s must yield (13.4) in the limit, it is

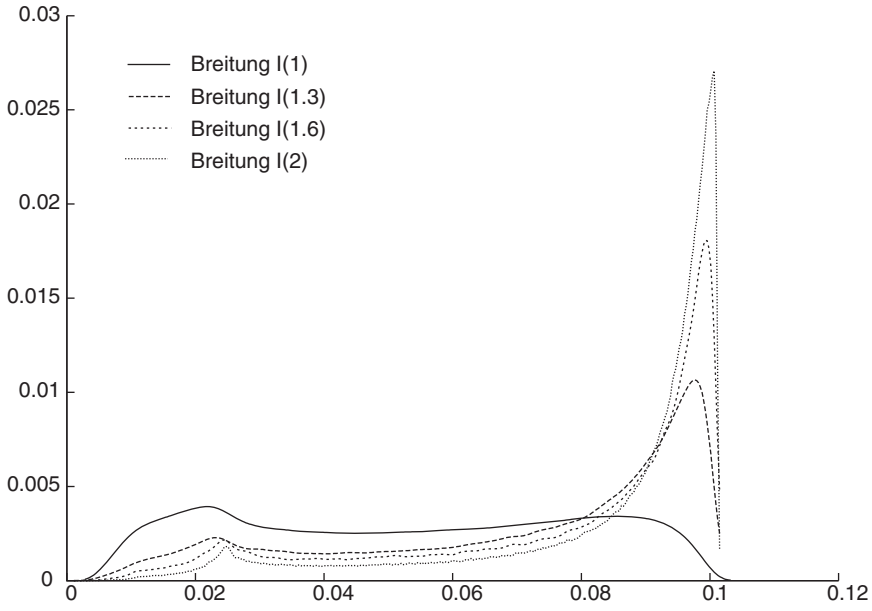

FIG. **13.1.** Breitung (2002) statistic with cumulated I(d) increments. The case I(1) is Breitung's null distribution. (Kernel density plots from 1 million replications.)

Table 13.1. 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

	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**

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 13.1, 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

[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 $\alpha_{NW} = [2\sum_{j=1}^{[n_T]} j\hat{\gamma}_j/(\hat{\gamma}_0 + 2\sum_{j=1}^{[n_T]}\hat{\gamma}_j)]^2$ 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.

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 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 13.1 that we should most like to possess, when evaluating Question 1. The next section attempts to operationalize this insight.

13.6 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

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

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 13.4.2. 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 13.2 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

Table 13.2. Bootstrap I(0) test: Rejection rates for the Kolmogorov–Smirnov test of Breitung's statistic in 5000 replications

T	Test	Size (Nominal 5% Test)	Power						
			AR(1): ϕ				FI: d		
			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

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.

13.7 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 chapter suggests that there are more suitable hypotheses to test, relating directly to the implications of the distribution of the data for asymptotic (ie, 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 (13.3) 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.

References

Agiakloglou, Ch. and Newbold, P. (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.

——and Monahan, J. C. (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 between power and level for generic unit root tests infinite samples. *Journal of Applied Econometrics*, **7**, 295–308.

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

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

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

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

Cavaliere, G. and Taylor, A. M. R. (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.

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

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

——and Sibbertsen, P. (2009). Tests of bias in log-periodogram regression. Forthcoming in *Economics Letters*.

Davidson, J., Magnus, J., and Wiegerinck, J. (2008). A general bound for the limiting distribution of Breitung's statistic. Econometric Theory, **24**(5), 1443–1455.

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

de Jong, R. M. and Davidson, J. (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 Levin, A. (1997). A practitioner's guide to robust covariance matrix estimation. Ch. 12 of Maddala, G. S. and Rao, C. R. (eds.), *Handbook of Statistics 15: Robust Inference*, pp. 291–341. 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**, 251–276.

————(1991). *Long-run Economic Relationships: Readings in Cointegration*. Oxford: Oxford University Press.

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

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

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

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

Harris, D., McCabe, B. M., and Leybourne, S. (2008). Testing for long memory. *Econometric Theory*, **24**(1), 143–175.

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

Hurvich, C. M., Deo, R., and Brodsky, J. (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., Vogelsang, T. J. and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, **68**(3), 695–714.

————(2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, **70**(5), 2093–2095.

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

Leeb, H. and Pötscher, B. M. (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, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, **59**(5), 1279–1313.

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

Mandelbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, 442–437.

Moulines, E. and Soulier, P. (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.

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

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

————(1974). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, **61**, 631–654.

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.

——(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., Teyssiere, G., and Vaiciulis, M. (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.