Econometric Theory, **24**, 2008, 616–630. Printed in the United States of America.
doi: 10.1017/S0266466608080250

THE IMPOSSIBILITY OF CONSISTENT DISCRIMINATION BETWEEN I(0) AND I(1) PROCESSES

ULRICH K. MÜLLER
Princeton University

An I(0) process is commonly defined as a process that satisfies a functional central limit theorem, i.e., whose scaled partial sums converge weakly to a Wiener process, and an I(1) process as a process whose first differences are I(0). This paper establishes that with this definition, it is impossible to consistently discriminate between I(0) and I(1) processes. At the same time, on a more constructive note, there exist consistent unit root tests and also nontrivial inconsistent stationarity tests with correct asymptotic size.

1. INTRODUCTION

A cornerstone of current time series econometrics is the distinction between processes that are integrated of order zero, I(0), and processes that are integrated of order one, I(1). To classify a given time series accordingly, researchers typically rely on either "unit root tests" with the null hypothesis of the series being I(1) or "stationarity tests" with a null hypothesis of the series being I(0). Both approaches have generated very large literatures, with numerous suggestions for unit root and stationarity test statistics. Under apparently weak assumptions on the underlying processes, both unit root and stationarity tests are typically shown to be consistent and to control asymptotic size. What is more, these assumptions typically also enable consistent discrimination between I(0) and I(1) process by letting the level of the unit root or stationarity tests shrink to zero at an appropriate rate as the sample size increases. Explicitly consistent discrimination procedures have been developed by Stock (1994), Phillips and Ploberger (1996), and Corradi (1999).

By considering a common and arguably relevant definition of I(0) and I(1) processes in a double-array framework, this paper establishes the impossibility of such consistent discrimination. The result is entirely driven by the impossibility of constructing a consistent stationarity test with asymptotic size smaller than one, because there exists a consistent and asymptotically valid unit root test.

The author thanks Benedikt Pötscher, James Davidson, and two anonymous referees for helpful comments on an earlier draft and the NSF for financial support by Grant SES 0518036. Address correspondence to Ulrich Müller, Department of Economics, Princeton University, Princeton, NJ 08544, USA; e-mail: umueller@princeton.edu.

  0266-4666/08 $15.00

The next section introduces the I(0) and I(1) definition employed and emphasizes differences to a related impossibility literature. Section 3 contains the main results, whose implications are discussed in Section 4. Proofs are collected in an Appendix.

2. DEFINITIONS OF I(0) AND I(1) AND THEIR PROPERTIES

In the following discussion, the observed data are thought to be the Tth row $\{y_{T,t}\}_{t=1}^{T}$ of the triangular double-array random process $\{\{y_{T,t}\}_{t=1}^{T}\}_{T=1}^{\infty}$, defined on a probability space $(\Omega, \mathfrak{F}, P)$, on which all following random elements are defined, too. To ease notation, we write $\{y_{T,t}\}$ for $\{\{y_{T,t}\}_{t=1}^{T}\}_{T=1}^{\infty}$ throughout. As in Stock (1994), Davidson (1999, 2002), White (2001, p. 179), and Breitung (2002), among others, the I(0)/I(1) property is defined in terms of the asymptotic properties of the rescaled data and the partial sum process of the data, respectively. Specifically, the process $\{y_{T,t}\}$ is defined to be I(0) and I(1) as follows:

$$\begin{aligned} &\{y_{T,t}\} \text{ is I(0)} \quad \text{if and only if } \exists\, \sigma > 0 : T^{-1/2}\sigma^{-1} \sum_{t=1}^{[\cdot T]} y_{T,t} \Rightarrow W(\cdot), \\ &\{y_{T,t}\} \text{ is I(1)} \quad \text{if and only if } \exists\, \sigma > 0 : T^{-1/2}\sigma^{-1} y_{T,[\cdot T]} \Rightarrow W(\cdot), \end{aligned} \tag{1}$$

where $[\cdot]$ is the largest smaller integer function, $W(\cdot)$ is a standard Wiener process, and $\Rightarrow$ indicates weak convergence (here, in the space of càdlàg functions on the unit interval equipped with the Skorohod metric), and all limits here and later are taken as $T \to \infty$. With this definition, an I(0) process is defined as being any possibly heterogenous process that satisfies a functional central limit theorem (FCLT), and an I(1) process is any process whose increments are I(0). In a more general setup, one might model both I(0) and I(1) processes with an additional deterministic component; for the purpose of this paper, however, it is useful to (unrealistically) assume that this deterministic component is known, such that $y_{T,t}$ in (1) is observed.

As the definition (1) concerns the asymptotic behavior of $\{y_{T,t}\}$ only, the interesting question is whether it is possible to construct discrimination procedures with attractive pointwise asymptotic properties. A (possibly randomized) test φ_T is a sequence of measurable functions of $\mathbb{R}^T \mapsto [0,1]$, where $\varphi_T(\{x_t\}_{t=1}^{T})$ indicates the conditional probability of rejection of the null hypothesis when observing the sample $\{x_t\}_{t=1}^{T}$, such that $E\varphi_T(\{x_t\}_{t=1}^{T})$ is the unconditional rejection probability. We refer to a stationarity test φ_T of asymptotic level α as consistent if $E\varphi_T(\{y_{T,t}\}_{t=1}^{T}) \to 1$ for all processes $\{y_{T,t}\}$ that are I(1) in the sense of definition (1), and we say it controls asymptotic size if $\overline{\lim}_{T\to\infty} E\varphi_T(\{y_{T,t}\}_{t=1}^{T}) \le \alpha$ for all processes $\{y_{T,t}\}$ that are I(0) in the sense of definition (1), with an analogous definition for unit root tests. A consistent discrimination procedure is a unit root or stationarity test of asymptotic level zero that controls asymptotic size, because with such a test both types of classification errors converge to zero. Note that these are the usual pointwise definitions of consistency and size control for unit root and stationarity tests; it is clearly not possible to construct nontrivial tests with uniform size control under definition (1), because one can assign arbitrary properties to $\{y_{T,t}\}_{t=1}^{T}$ for any fixed T without any implication for the eventual convergence in definition (1).

Despite being highly "nonparametric," definition (1) entails the salient features commonly associated with I(1) processes (and, by implication, of scaled partial sums of I(0) processes). Define E_K as a truncated expectation operator with truncation point K; i.e., for any random variable X, $\mathrm{E}_K X = \mathbf{1}[|X| < K]X$. Assume that $\{y_{T,t}\}$ is I(1) in the sense of (1). Then, e.g., (i) unit root test statistics that can be written as continuous functionals (in the sup norm) of $T^{-1/2}y_{T,[\cdot T]}$ have the usual asymptotic distributions; (ii) the ratio of the truncated variances of $y_{T,[sT]}$ and $y_{T,[rT]}$, $\mathrm{E}_K y_{T,[sT]}^2/\mathrm{E}_K y_{T,[rT]}^2$, converges to s/r when $r > 0$ for all truncation points $K > 0$; (iii) with $r > s > 0$, $\mathrm{E}_K[T^{-1}(y_{T,[rT]} - y_{T,[sT]})^2] - \min_{\{\alpha_j\}} \mathrm{E}_K[T^{-1}(y_{T,[rT]} - \sum_{j=1}^{p} \alpha_j y_{T,[\tau_j T]})^2] \to 0$ for all $p \in \mathbb{N}$ and $0 \le \tau_1 < \tau_2 < \cdots < \tau_p = s$, so that $y_{T,[sT]}$ is an asymptotically efficient linear long-range forecast of $y_{T,[rT]}$ in a truncated mean square error sense, with arbitrary truncation point $K > 0$. If in addition to $\{y_{T,t}\}$ being I(1) in the sense of (1), $\{T^{-1}y_{T,t}^2\}$ is assumed to be uniformly integrable, then (ii) and (iii) also hold without any truncation (and the results that follow also hold for this more restricted class of I(1) processes).

What is more, as demonstrated by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2005), the I(0) definition in (1), i.e., the assumption that $\{y_{T,t}\}$ satisfies a FCLT, is sufficient to enable asymptotically justified inference in time series regression models. It is possible, e.g., to do asymptotically justified inference for μ when observing $\{y_{T,t} + \mu\}$ whenever $\{y_{T,t}\}$ is I(0) in the sense of definition (1). It would hence be highly desirable to be able to consistently discriminate (at least pointwise) between I(0) and I(1) processes as defined in (1), the possibility of which is the subject of this paper.

It might be useful to relate this question to the results of Blough (1992), Faust (1996), and Pötscher (2002). These authors are concerned with parametric time series models, and they define a process as I(0) if it is stationary with positive spectral density at frequency zero and as I(1) if the first differences are I(0). The papers explore the possibility of constructing nontrivial unit root and stationarity tests that control size *uniformly* over the parameter space. The conclusion is that such tests do not exist in common parametric models. To illustrate the point, consider the class of processes

$$y_t = \varepsilon_t + \lambda \sum_{l=1}^{t} \eta_l \quad \text{with } (\varepsilon_t, \eta_t)' \text{ i.i.d.} N(0, I_2) \tag{2}$$

with $\lambda \geq 0$. As long as λ is positive, $\{y_t\}_{t=1}^T$ is considered to be I(1), because the spectral density of the first differences of $\{y_t\}_{t=1}^T$ is positive at frequency zero. But for any given sample size T, one can choose λ so small that there is no basis for discrimination from the I(0) case $\lambda = 0$ whatsoever. This is true for any sample size, such that it is impossible to construct a unit root test with power larger than size *uniformly* over $\lambda > 0$, even asymptotically. A similar point can be made about the impossibility of constructing nontrivial stationarity tests in the Gaussian first-order autoregressive model $y_t = \rho y_{t-1} + \varepsilon_t$ with power larger than size uniformly over $\rho \in [0;1)$. The point of this literature is not to criticize specific I(0) versus I(1) discrimination procedures; rather, it shows that discrimination between I(0) and I(1) processes, as defined via the properties of the spectral density at frequency zero, is an "ill-posed" statistical problem for common parametric classes of time series models without further restrictions on the parameter space.

Note, however, that in example (2), the problematic part of the parameter space with λ very small but positive does not generate good approximations of an ensuing I(1) analysis based on the FCLT anyway: The scaled partial sums of the first differences, of course, recover $\{T^{-1/2}y_t\}_{t=1}^T$, which is far from being close to a scaled Wiener process in any sense when λ is small. Similarly, an efficient linear forecast for the sum of an I(0) series and a "sufficiently small" independent random walk is dominated by the I(0) component and is hence typically very different from the last observation of the random walk component, even if it were observed. Following the spectral density definition of I(0) and I(1) processes hence by no means implies that the scaled I(1) and scaled partial sums of I(0) processes have close to random walk behavior, even asymptotically. If the purpose of the discrimination between I(0) and I(1) processes is to employ Wiener process approximations or to infer the usual implications for forecasting, the spectral density definition is not the relevant one.

Also, the impossibility results of Blough (1992), Faust (1996), and Pötscher (2002) have no implication for whether or not it is possible to construct nontrivial (pointwise) asymptotically valid unit root and stationarity tests under definition (1). What example (2) demonstrates is only that for any sample size T there exists $\lambda_T > 0$ such that the difference between the power and level of any unit root test is less than, say, 10 percentage points, which obviously excludes the existence of a 5% level consistent unit root test with uniform size control. But it is not at all clear whether the "near stationary" double-array process, resulting from the data generating process (2) with λ equal to these λ_T for each T, satisfies the I(1) definition in (1), simply because λ_T might converge to zero too quickly.

It is hence an open question whether it is possible to consistently discriminate, at least pointwise, between I(0) and I(1) processes as defined in (1), and given the appeal of definition (1), it is a relevant question for current time series econometrics.

3. CONSISTENT UNIT ROOT AND STATIONARITY TESTS

This section contains the two findings of this paper: under definition (1), there exist consistent unit root tests with positive asymptotic level that control asymptotic size, but any consistent scale-invariant stationarity test necessarily has arbitrarily bad asymptotic size control.

As a consistent (and also scale-invariant) unit root test, consider a test based on the statistic suggested by Breitung (2002):

$$G_T = T^{-2} \frac{\sum_{t=1}^{T} \left[\sum_{l=1}^{t} y_{T,l} \right]^2}{\sum_{t=1}^{T} y_{T,t}^2}$$

that rejects the I(1) null hypothesis for small values. As shown in Breitung (2002), under the null hypothesis of $\{y_{T,t}\}$ being I(1) in the sense of (1), $G_T \Rightarrow \int_0^1 [\int_0^s W(u)\,du]^2\,ds / \int_0^1 W(s)^2\,ds$. For any choice of critical value cv > 0, the asymptotic level is hence determined by the probability $P(\int_0^1 [\int_0^s W(u)\,du]^2\,ds / \int_0^1 W(s)^2\,ds < \text{cv}) > 0$, and unit root tests based on G_T control asymptotic size. Breitung (2002) also demonstrates the consistency of a test based on G_T against stationary and ergodic I(0) alternatives with bounded second moments; the following theorem shows in addition the consistency of a unit root test based on G_T with positive critical value against the general I(0) alternative in the sense of definition (1).

THEOREM 1. *Whenever* $\{y_{T,t}\}$ *is* I(0) *in the sense of (1),* $G_T \xrightarrow{p} 0$.

Definition (1) hence allows the construction of consistent unit root tests with positive asymptotic level that control asymptotic size. At the same time, under definition (1), it is not possible to construct a consistent scale-invariant stationarity test that controls asymptotic size:

A scale invariant test φ_T satisfies $\varphi_T(\{cx_t\}_{t=1}^T) = \varphi_T(\{x_t\}_{t=1}^T)$ for all $\{x_t\}_{t=1}^T$, $T \geq 1$ and nonzero scalars c. The rationale for imposing this invariance is that the scale of time series typically depends on an arbitrary choice of units. More generally, the overall scale is only informative for distinguishing I(0) from I(1) processes as defined in (1) when something is known about the respective σ's. I am not aware of a stationarity test or I(0)/I(1) discrimination procedure suggested in the literature that is not invariant to scale.

The main result of the paper, which implies the impossibility of consistent discrimination between I(0) and I(1) processes, is the following theorem.

THEOREM 2. *Let* $y_{T,t} = \sum_{l=1}^{t} \varepsilon_l$, *where* ε_t *are* i.i.d.$N(0,1)$. *For any scale-invariant stationarity test* φ_T *that satisfies* $\varphi_T(\{y_{T,t}\}_{t=1}^T) \xrightarrow{p} 1$, *there exist a standard Wiener process W and a sequence of* $T \times 1$ *Gaussian random* *vectors* $Z_T = (z_{T,1},\ldots,z_{T,T})'$ *satisfying* $\varphi_T(\{z_{T,t}\}_{t=1}^T) \xrightarrow{p} 1$ *and* $\sup_{0\le s\le 1}|T^{-1/2}\sum_{t=1}^{[sT]} z_{T,t} - W(s)| \xrightarrow{\text{a.s.}} 0$, *so that* $\{z_{T,t}\}$ *is* I(0) *in the sense of (1).*

Heuristically, the argument for Theorem 2 proceeds as follows. Recall that a Wiener process can be written as a randomly weighted sum of certain trigonometric functions, i.e., with W_0 a standard Wiener process and ξ_l i.i.d.$N(0,1)$, $l = 1,2,\ldots$:

$$\frac{\sqrt{2}}{\pi}\sum_{l=1}^{\infty}\frac{\sin((l-\frac{1}{2})\pi s)}{l-\frac{1}{2}}\xi_l = W_0(s)$$

because the left-hand side converges almost surely (a.s.) and uniformly on $s \in [0,1]$ to a Wiener process (see Itô and Nisio (1968); and, e.g., Phillips, 1998, for further discussion). The fact that a Wiener process is smooth, i.e., dominated by low-frequency components, is reflected in the fact that high-frequency components $\sin((l-\frac{1}{2})\pi s)$ with l large receive weight $\xi_l/(l-\frac{1}{2})$ of small variance $(l-\frac{1}{2})^{-2}$. A consistent stationarity test rejects with probability converging to one for a Gaussian random walk, and so $\varphi_T(\{T^{1/2}W_0(t/T)\}_{t=1}^T) \xrightarrow{p} 1$. Now for some $n > 1$, consider the process

$$\sqrt{2}\sum_{l=1}^{n}\sin((l-\tfrac{1}{2})\pi s)\xi_l + \frac{\sqrt{2}}{\pi}\sum_{l=n+1}^{\infty}\frac{\sin((l-\frac{1}{2})\pi s)}{l-\frac{1}{2}}\xi_l = W_n(s). \tag{3}$$

The n lowest frequency components $\sin((l-\frac{1}{2})\pi s)$, $l = 1,\ldots,n$ of W_n now receive stochastic weights of the same variance, just as one would expect for white noise, and the decline in the variances of higher frequency components characteristic of a Wiener process only starts after the n lowest frequency components. What is more, the first summand on the left-hand side of (3) dominates the overall variability of W_n for n large, as the relative variance of any two components $i \le n$ and $j > n$ is at least $\pi^2(n-\frac{1}{2})^2$. So with n large, up to a small remainder, W_n is equal to a continuous time process whose low-frequency behavior is just like white noise. Of course, there is very little high-frequency variability in W_n, and so for T large, the time series $\{W_n(t/T)\}_{t=1}^T$ exhibits very strong positive autocorrelation.

Now W_n is derived from W_0 by a change of the variance of n independent Gaussian variates only. The probability measures of two multivariate Gaussian vectors with positive definite covariance matrices are evidently equivalent, and so a sequence of events whose probability converges to zero under one measure also has probability converging to zero under the other measure. Therefore $\varphi_T(\{T^{1/2}W_0(t/T)\}_{t=1}^T) \xrightarrow{p} 1$ implies $\varphi_T(\{T^{1/2}W_n(t/T)\}_{t=1}^T) \xrightarrow{p} 1$. As $\varphi_T(\{T^{1/2}W_n(t/T)\}_{t=1}^T) \xrightarrow{p} 1$ holds for any fixed n, it is possible to construct an increasing sequence $n_T \to \infty$ such that also $\varphi_T(\{T^{1/2}W_{n_T}(t/T)\}_{t=1}^T) \xrightarrow{p} 1$. For

$n \to \infty$, one might imagine that W_n approximates white noise arbitrarily well; in fact, the limit

$$\sqrt{2}\sum_{l=1}^{\infty} \sin((l-\tfrac{1}{2})\pi s)\xi_l \tag{4}$$

is a continuous analogue of white noise in the generalized stochastic process sense (see, e.g., Hida, 1980, Ch. 3). The process $\{z_{T,t}\}$ of Theorem 2 is simply the scale-normalized version of this sequence, $\{z_{T,t}\}_{t=1}^{T} = \{T^{-1/2}W_{n_T}(t/T)\}_{t=1}^{T}$ for $T \geq 1$, which ensures that partial sums of $\{z_{T,t}\}$ converge to a standard Wiener process. The rate at which $n_T \to \infty$ as $T \to \infty$ depends on the choice of test φ_T, and the slower this rate, the larger the autocorrelations in $\{z_{T,t}\}_{t=1}^{T}$ for T large.

4. DISCUSSION

1. Although Theorem 1 shows the existence of a consistent unit root test with correct asymptotic size under definition (1), this does not imply that popular unit root tests share this desirable feature. Most unit root test statistics rely on a consistent estimator $\hat{\sigma}$ of the scale of the limiting process σ under the I(1) null hypothesis. The results in Müller (2007), however, show that no estimator $\hat{\sigma}$ exists that is consistent for all processes satisfying the I(1) definition (1), such that the resulting tests will typically not control asymptotic size. Also, despite the result in Theorem 1, the application of a consistent unit root test with asymptotic level zero (which, for a unit root test based on G_T, would require a sequence of critical values cv_T that converge to zero at an arbitrarily slow rate) cannot control asymptotic size. Such a sequence of tests would imply the existence of a consistent stationarity test that controls asymptotic size, which is ruled out by Theorem 2.

2. One might argue against the relevance of Theorem 2 by pointing to the double-array character of the argument. In econometric practice one is typically concerned with one data set of fixed sample size. These data can be thought of as being a part of an infinite single-index time series or as a particular row of a double-array process. Theorem 2 implies that for any consistent and scale-invariant stationarity test φ_T and real numbers $\epsilon > 0$ and $\delta > 0$, there exist for a large enough T a $T \times 1$ Gaussian random vector $(z_{T,1}, \ldots, z_{T,T})'$ and a standard Wiener process W such that

$$\mathrm{E}\varphi_T(\{z_{T,t}\}_{t=1}^{T}) > 1-\epsilon \quad \text{and} \quad P\left(\sup_{0\leq s\leq 1}\left|T^{-1/2}\sum_{t=1}^{[sT]} z_{T,t} - W(s)\right| < \delta\right) > 1-\epsilon.$$

For any large enough fixed T, there is hence an infinite time series $\{y_t^*\}_{t=1}^{\infty}$ whose observable part $\{y_t^*\}_{t=1}^{T}$ shares these properties—simply let $y_t^* = z_{T,t}$ for $t = 1, \ldots, T$. So for any large enough T, there exists a standard infinite time series $\{y_t^*\}_{t=1}^{\infty}$ for which the consistent stationarity tests reject with very high probability, yet this time series is close to being I(0) in the sense that the scaled partial sums of the sample $\{y_t^*\}_{t=1}^{T}$ "almost" behave like a Wiener process. In this sense Theorem 2 also implies undesirable properties of consistent stationarity tests when applied to data that are thought of as being part of an infinite single-index series.

3. Size control problems of popular consistent stationarity tests, such as the test suggested by Kwiatkowski, Phillips, Schmidt, and Shin (1992), have been noted before—see, e.g., the simulation study of Caner and Kilian (2001) and the analysis in Müller (2005). Theorem 2 demonstrates that the problems associated with consistent stationarity tests cannot be resolved by considering alternative tests. In applications where a null hypothesis of a general I(0) process as defined in (1) makes sense, the outcome of consistent stationarity tests should be interpreted with caution. Rejections might be due to this inherent fragility, rather than by evidence in the data that the I(0) null hypothesis does not hold.

Consistent stationarity tests should be thought of as testing jointly the I(0) property as defined in (1) and additional restrictions on the behavior of the process, which may very well be a useful enterprise. Given the crucial nature of these additional restrictions, however, it is important to assess whether they are appropriate for a given application.

4. The restriction to scale-invariant tests in Theorem 2 can be dispensed with[1] when σ in the definition (1) of I(0) and I(1) processes is allowed to depend on T, i.e., under the definition

$$
\begin{aligned}
&\{y_{T,t}\} \text{ is I}(0) \quad \text{if and only if } \exists\, \{\sigma_T\}_{T=1}^{\infty} : T^{-1/2}\sigma_T^{-1}\sum_{t=1}^{[\cdot T]} y_{T,t} \Rightarrow W(\cdot), \\
&\{y_{T,t}\} \text{ is I}(1) \quad \text{if and only if } \exists\, \{\sigma_T\}_{T=1}^{\infty} : T^{-1/2}\sigma_T^{-1} y_{T,[\cdot T]} \Rightarrow W(\cdot).
\end{aligned}
\tag{5}
$$

In other words, if I(0) and I(1) processes are defined in terms of weak convergence to a Wiener process *in some sample-size-dependent scale*, then it is impossible to consistently discriminate between the two, because no consistent stationarity test can control asymptotic size. Note that such an alternative I(0)/I(1) definition is equally sensible as far as the discussion at the beginning of Section 2 is concerned.

5. Theorems 1 and 2 display a surprising asymmetry between the properties of consistent unit root and stationarity tests. An attempt to adjust the reasoning of Theorem 2 to yield the analogous impossibility result for consistent unit root tests fails: although it is possible to modify the weights on the low-frequency components of an appropriate version of Gaussian white noise (4) in a similar manner, the remainder term $\sqrt{2}\sum_{l=n+1}^{\infty}\sin((l-\frac{1}{2})\pi s)\xi_l$ does not become small for n large.

Intuitively, under definition (1) and scale invariance, the testable difference between I(1) and I(0) processes is that I(1) processes are dominated by low-frequency components, whereas I(0) processes have similar variability across all frequencies. A consistent unit root test may thus compare the variability across all frequencies (the denominator $T^{-2}\sum_{t=1}^{T} y_{T,t}^2$ of G_T) with the variability in one fixed band of low-frequency components (the numerator $T^{-4}\sum_{t=1}^{T}[\sum_{l=1}^{t} y_{T,l}]^2$ of G_T), because under the alternative, the high-frequency components are known to dominate in relative terms. In contrast, for a stationarity test to achieve consistency, it must assess whether an *increasing* number of low-frequency components are too variable compared to higher frequency components. This necessarily involves an implicit assumption on how quickly the I(0) process exhibits the eventually similar variability across all frequencies, which can always be mistaken.

6. As pointed out in Section 2, definition (1) does not allow the construction of discrimination procedures with attractive properties uniformly over all processes that satisfy the definition. The reason is simply that definition (1) has no implication for any fixed sample size T, as the respective convergences might occur "later." Theorems 1 and 2 have nevertheless an interesting implication for the possibility of obtaining correct inference uniformly over subsets of I(0) and I(1) processes with controlled speed of convergence: Let r be a metric that metrizes weak convergence of probability measures on the metric space $\mathcal{D}$ of the càdlàg functions on the unit interval (equipped with the Skorohod metric). Let π_W be Wiener measure and $\pi_T(x_{[\cdot T]})$ the probability measure on $\mathcal{D}$ induced by the stochastic process $\{x_t\}_{t=1}^{T}$. For $\{\delta_T\}_{T=1}^{\infty}$ a sequence of real numbers, let

$$\mathcal{H}_0(\{\delta_T\}_{T=1}^{\infty}) = \Bigg\{\{y_{T,t}\} : \exists\sigma > 0 \quad \text{so that for all } T \geq 1,\ r\left(\pi_T\left(T^{-1/2}\sigma^{-1}\sum_{t=1}^{[\cdot T]} y_{T,t}\right), \pi_W\right) < \delta_T\Bigg\},$$

$$\mathcal{H}_1(\{\delta_T\}_{T=1}^{\infty}) = \{\{y_{T,t}\} : \exists\sigma > 0 \quad \text{so that for all } T \geq 1,\ r(\pi_T(T^{-1/2}\sigma^{-1} y_{T,[\cdot T]}), \pi_W) < \delta_T\}.$$

As long as $\delta_T \to 0$, the double-array processes $\mathcal{H}_0(\{\delta_T\}_{T=1}^{\infty})$ and the first difference of the processes in $\mathcal{H}_1(\{\delta_T\}_{T=1}^{\infty})$ satisfy a FCLT *uniformly*, with the speed of convergence determined by the sequence $\{\delta_T\}_{T=1}^{\infty}$. Theorem 1 and its proof in the Appendix now show that for *any* choice of $\delta_T \to 0$, a level α unit root test φ_T^G based on G_T is uniformly consistent over $\mathcal{H}_0(\{\delta_T\}_{T=1}^{\infty})$ and controls asymptotic size uniformly over $\mathcal{H}_1(\{\delta_T\}_{T=1}^{\infty})$, i.e.,

$$\lim_{T\to\infty} \inf_{\{y_{T,t}\}\in\mathcal{H}_0(\{\delta_T\}_{T=1}^\infty)} \mathrm{E}\varphi_T^G(\{y_{T,t}\}_{t=1}^T) \to 1,$$

$$\lim_{T\to\infty} \sup_{\{y_{T,t}\}\in\mathcal{H}_1(\{\delta_T\}_{T=1}^\infty)} \mathrm{E}\varphi_T^G(\{y_{T,t}\}_{t=1}^T) \to \alpha.$$

Theorem 2, on the other hand, implies that for any scale-invariant stationarity test that is consistent against a Gaussian random walk there exists a double-array process $\{z_{T,t}\}$ and $\delta_T \to 0$ such that $\{z_{T,t}\} \in \mathcal{H}_0(\{\delta_T\}_{T=1}^\infty)$ and $\lim_{T\to\infty} \mathrm{E}\varphi_T(\{z_{T,t}\}_{t=1}^T) \to 1$. In other words, it is possible to construct uniformly valid and consistent unit root tests that do not require knowledge of the speed of convergence, whereas it is impossible to construct a consistent and asymptotically valid stationarity test without knowledge of the speed of convergence under the null hypothesis. As observed under comment 3 in this section, for any set of I(0) processes $\mathcal{H}_0(\{\delta_T\}_{T=1}^\infty)$ with given speed of convergence $\{\delta_T\}_{T=1}^\infty$, it might be possible to construct consistent and asymptotically valid stationarity tests, but these tests are then necessarily joint tests of the I(0) hypothesis as defined in (1) and the given speed of convergence.

7. A large body of work has generalized the I(0)/I(1) dichotomy to general fractional integration $\mathrm{I}(d)$, where d is not constrained to be an integer. When $\mathrm{I}(d)$ processes are defined in analogy to (1) in terms of weak convergence of the scaled process (or partial sum process) to fractional Brownian motion, Theorem 2 implies the impossibility of consistently estimating $d \in D \subset \mathbb{R}$ as long as $D \supset \{0,1\}$. Inference methods that rely on consistent estimation of $d \in D$ necessarily fail for some process that could have been well approximated by (fractional) Brownian motion.

8. Because the I(0) definition (1) is equivalent to $\{y_{T,t}\}$ satisfying a FCLT, stationarity tests can be thought of as special cases of tests of the FCLT property. Such tests might be motivated by the prominent role that the FCLT condition plays in Kiefer et al. (2000) type inference. But Theorem 2 has a sobering implication: Any test that consistently rejects for a Gaussian random walk (which surely constitutes an extreme deviation of the FCLT null hypothesis) necessarily also consistently rejects for some process that satisfies a FCLT. In particular, there are instances where using a consistent stationarity test as a pretest for subsequent Kiefer et al. type inference leads with high probability to the erroneous conclusion that Kiefer et al. type inference is inadequate, despite the fact that it would have resulted in a highly accurate approximation.

9. As pointed out by one referee, one might conclude from Theorem 2 that definition (1) of I(0) and I(1) processes is not adequate, after all. This view leads to the question of how to appropriately strengthen the definition to ensure the possibility of consistent discrimination between I(0) and I(1) processes.

10. Note, however, that the results of this paper are not entirely destructive in the sense that Theorem 2 only rules out the existence of a *consistent* stationarity test that controls asymptotic size. It is perfectly possible to derive asymp-

totically valid inconsistent but nontrivial stationarity tests. Consider, e.g., the test statistic

$$J_T = \frac{\sum_{t=[T/2]+1}^{T} y_{T,t}}{\sum_{t=1}^{[T/2]} y_{T,t}}.$$

When $\{y_{T,t}\}$ is I(0), J_T has a Cauchy asymptotic distribution, whereas with $\{y_{T,t}\}$ being I(1), $J_T \Rightarrow \int_{1/2}^{1} W(s)\,ds / \int_0^{1/2} W(s)\,ds$, which is not distributed Cauchy.

The development of (necessarily inconsistent) stationarity tests and estimators of the fractional parameter d, in addition to (possibly consistent) unit root tests that remain valid under definition (1) and that are efficient in some sense, seems an interesting topic for future research.

NOTE

1. Define $\{z_{T,t}^n\}_{t=1}^T = \{T^{1/2} W_n(t/T)\}_{t=1}^T$ instead of $\{z_{T,t}^n\}_{t=1}^T = \{T^{-1/2} W_n(t/T)\}_{t=1}^T$ in the proof of Theorem 2, so that no scale-invariance argument is necessary to obtain $\varphi_T(\{z_{T,t}^n\}_{t=1}^T) \xrightarrow{p} 1$ for all n, and note that $T^{-3/2}\sum_{t=1}^{[sT]} z_{T,t}$ recovers the right-hand side of (A.1) in the Appendix; i.e., $\{z_{T,t}\}$ so constructed is I(0) under definition (5) with $\sigma_T = T$.

REFERENCES

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

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

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

Corradi, V. (1999) Deciding between I(0) and I(1) via FLIL-based bounds. *Econometric Theory* 14, 643–663.

Davidson, J. (1999) When Is a Time Series I(0)? Evaluating the Memory Properties of Nonlinear Dynamic Models. Mimeo, Cardiff University.

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

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

Hida, T. (1980) *Brownian Motion*. Springer-Verlag.

Itô, K. & M. Nisio (1968) On the convergence of sums of independent Banach space valued random variables. *Osaka Journal of Mathematics* 5, 35–48.

Kiefer, N.M. & T.J. Vogelsang (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

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

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

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

Müller, U.K. (2007) A theory of robust long-run variance estimation. *Journal of Econometrics* 141, 1331–1352.

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

Phillips, P.C.B. & W. Ploberger (1996) An asymptotic theory of Bayesian inference for time series. *Econometrica* 64, 381–412.

Pollard, D. (2002) *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

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, 1035–1065.

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

White, H. (2001) *Asymptotic Theory for Econometricians*, rev. ed. Academic Press.

APPENDIX: Proofs

Proof of Theorem 1. Because G_T is scale invariant, we may set $\sigma = 1$ without loss of generality. From standard arguments, $T^{-2}\sum_{t=1}^{T}[\sum_{l=1}^{t} y_{T,l}]^2 \Rightarrow \int_0^1 W(s)^2\,ds$. It hence suffices to show existence of a sequence $c_T \to \infty$ such that $P(\sum_{t=1}^{T} y_{T,t}^2 > c_T) \to 1$. Let $\{\hat{\xi}_{T,k}\}_{k=1}^{T-1}$ be the last $T-1$ elements of the type II discrete cosine transform of $\{y_{T,t}\}_{t=1}^{T}$, i.e., $\hat{\xi}_{T,k} = \sqrt{2/T}\sum_{t=1}^{T}\cos(\pi k(t-\frac{1}{2})/T)y_{T,t}$, $k = 1,\ldots,T-1$, so that $\sum_{t=1}^{T} y_{T,t}^2 = \sum_{k=1}^{T-1}\hat{\xi}_{T,k}^2 + T^{-1}(\sum_{t=1}^{T} y_{T,t})^2$. By summation by parts and the trigonometric identity $\cos(2u) - \cos(2v) = -2\sin(u+v)\sin(u-v)$,

$$\hat{\xi}_{T,k} = -\sqrt{2}\sum_{t=0}^{T-1} S_{T,t}[\cos(\pi k(t+\tfrac{1}{2})/T) - \cos(\pi k(t-\tfrac{1}{2})/T)]$$

$$= \sqrt{8}\sin(k\pi/(2T))\sum_{t=0}^{T-1} S_{T,t}\sin(\pi kt/T),$$

where $S_{T,t} = T^{-1/2}(\sum_{l=1}^{t} y_{T,l} - tT^{-1}\sum_{l=1}^{T} y_{T,l})$ (so that $S_{T,0} = S_{T,T} = 0$ for all $T \geq 1$). By definition (1), standard calculations, and the continuous mapping theorem, we find $S_{T,[sT]} \Rightarrow W(s) - sW(1) \equiv B(s)$. Denote by $L^2[0,1]$ the set of square integrable functions on the unit interval. Note that $\{\sqrt{2}\sin(k\pi s)\}_{k=1}^{\infty}$ is an orthonormal basis in $L^2[0,1]$ and that the covariance kernel of the Gaussian process $B(\cdot)$, $\mathrm{E}[B(s)B(u)] = s \wedge u - su$, can be written as $\mathrm{E}[\mathrm{B(s)B(u)}] = 2\sum_{l=1}^{\infty}\sin(\pi sl)\sin(\pi ul)/(\pi^2 l^2)$ (see, e.g., Phillips, 1998). Another application of the continuous mapping theorem thus yields $(\hat{\xi}_{T,1},\ldots,\hat{\xi}_{T,m})' \Rightarrow N(0,I_m)$ for any fixed m, because $\{T\sin(k\pi/(2T))\}_{k=1}^{m} \to \{\frac{1}{2}k\pi\}_{k=1}^{m}$. Let T_m be the smallest T^* such that for all $T \geq T^*$, $|P(\sum_{l=1}^{m}\hat{\xi}_{T,l}^2 > m/2) - P(\chi_m^2 > m/2)| < m^{-1}$, where χ_m^2 is a chi-squared random variable with m degrees of freedom. For any T, let m_T be the largest m^* such that $(\max_{m \leq m^*} T_m) < T$. Note that $m_T \to \infty$, because for any m, T_m is finite. But $P(\sum_{t=1}^{T} y_{T,t}^2 > m_T/2) \geq P(\sum_{k=1}^{m_T}\hat{\xi}_{T,k}^2 > m_T/2) \geq P(2\chi_{m_T}^2/m_T > 1) - m_T^{-1} \to 1$. ■

Proof of Theorem 2. Let ξ_l be i.i.d.$N(0,1)$, $l = 1,2,\ldots$, and denote with $\mathcal{C}$ the set of continuous functions on the unit interval, equipped with the sup norm. Theorem 5.2

of Itô and Nisio (1968) shows that for $\{\zeta_l\}_{l=1}^{\infty}$ an orthonormal basis in $L^2[0,1]$, $\sum_{l=1}^{\infty} \xi_l \int_0^s \zeta_l(u)\,du$ converges to a Wiener process a.s. on $\mathcal{C}$, i.e., uniformly in $s \in [0,1]$.

Specializing this result, define $\phi_l(s) = \sqrt{2}\sin((l-\frac{1}{2})\pi s)/((l-\frac{1}{2})\pi)$ for $l = 1,2,\ldots$. Because $\{\sqrt{2}\cos((l-\frac{1}{2})\pi s)\}_{l=1}^{\infty}$ is an orthonormal basis in $L^2[0,1]$, $\sum_{l=1}^{\infty} \phi_l(s)\xi_l = W_0(s)$, where W_0 is a standard Wiener process, so that $\{y_{T,t}\}_{t=1}^{T}$ has the same distribution as $\{T^{1/2}W_0(t/T)\}_{t=1}^{T}$ for all T. For any fixed $n > 0$, let

$$W_n(s) = \sum_{l=1}^{n} \pi(l-\tfrac{1}{2})\phi_l(s)\xi_l + \sum_{l=n+1}^{\infty} \phi_l(s)\xi_l$$

and define $\{z_{T,t}^n\} = \{T^{-1/2}W_n(t/T)\}_{t=1}^{T}$ for $n \geq 0$. We first show that the measures of W_0 and W_n on $\mathcal{C}$ are equivalent.

For $x \in \mathcal{C}$, let $\psi_l(x) = \pi(l-\frac{1}{2})\int_0^1 \phi_l(u)x(u)\,du$. Consider the continuous functions $h: \mathcal{C} \mapsto \mathcal{C} \times \mathbb{R}^n$ and $g: \mathcal{C} \times \mathbb{R}^n \mapsto \mathcal{C}$ with

$$\begin{aligned} h(x) &= (h_1(x), h_2(x)) \\ &= \left(x - \pi\sum_{l=1}^{n}(l-\tfrac{1}{2})\psi_l(x)\phi_l, (\psi_1(x),\ldots,\psi_n(x))'\right) \end{aligned}$$

and $g(x,(v_1,\ldots,v_n)') = x + \pi\sum_{l=1}^{n}(l-\frac{1}{2})v_l\phi_l$ (where the metric on $\mathcal{C} \times \mathbb{R}^n$ is chosen as the sum of the sup norm in $\mathcal{C}$ and the euclidian norm in $\mathbb{R}^n$). Because $\{\pi(l-\frac{1}{2})\phi_l\}_{l=1}^{\infty} = \{\sqrt{2}\sin((l-\frac{1}{2})\pi s)\}_{l=1}^{\infty}$ are orthonormal functions in $L^2[0,1]$, $h_1(x)$ and $h_2(x)$ are the residual and coefficients of a continuous time regression of x on $\{\pi(l-\frac{1}{2})\phi_l\}_{l=1}^{n}$, respectively. Clearly, $g(h(x)) = x$ for all $x \in \mathcal{C}$. For any measurable set $A \subset \mathcal{C}$ and $j \in \{0,n\}$, we thus have $P(W_j \in A) = P(h(W_j) \in g^{-1}(A))$, where $g^{-1}(A) = (G_1,G_2)$ with $G_1 \subset \mathcal{C}$ and $G_2 \subset \mathbb{R}^n$ is the inverse image of A under g. It therefore suffices to show equivalence of the measures of $h(W_0)$ and $h(W_n)$. By construction of W_0 and W_n, $h_1(W_0) = h_1(W_n) = \sum_{l=n+1}^{\infty}\xi_l\phi_l$ a.s., and $h_2(W_j)$ is a function of $\{\xi_l\}_{l=1}^{n}$. Therefore, $h_1(W_j)$ and $h_2(W_j)$ are independent, $P(h(W_j) \in g^{-1}(A)) = P(h_1(W_j) \in G_1)P(h_2(W_j) \in G_2)$, and $P(h_1(W_0) \in G_1) = P(h_1(W_n) \in G_1)$. But $h_2(W_0) \sim N(0, \pi^{-2}\mathrm{diag}((\frac{1}{2})^{-2},(\frac{3}{2})^{-2},\ldots,(n-\frac{1}{2})^{-2}))$ and $h_2(W_n) \sim N(0,I_n)$, so that $P(h_2(W_0) \in G_2) = 0$ if and only if $P(h_2(W_1) \in G_2) = 0$, and we conclude that also $P(W_0 \in A) = 0$ if and only if $P(W_n \in A) = 0$.

Note that for all T, $\{z_{T,t}^j\}_{t=1}^{T}$ are continuous functions of W_j for $j \in \{0,n\}$, so that for any $\varepsilon > 0$, the event $\varphi_T(\{z_{T,t}^j\}_{t=1}^{T}) < 1 - \varepsilon$ for $T \geq 1$ can be equivalently expressed as $W_j \in A_T$ for some measurable sequence $A_T \subset \mathcal{C}$. As $\{y_{T,t}\}_{t=1}^{T}$ has the same distribution as $\{T^{1/2}W_0(t/T)\}_{t=1}^{T}$ for all T and φ_T is scale invariant, $\varphi_T(\{y_{T,t}\}_{t=1}^{T}) \xrightarrow{p} 1$ implies $\varphi_T(\{z_{T,t}^0\}_{t=1}^{T}) \xrightarrow{p} 1$, so that $P(W_0 \in A_T) \to 0$. From the equivalence of the measures of W_0 and W_n it also follows that $P(W_n \in A_T) \to 0$ (see Pollard, 2002, p. 55). Because ε was arbitrary, we thus have that $\varphi_T(\{z_{T,t}^n\}_{t=1}^{T}) \xrightarrow{p} 1$, too.

There hence exists for any n a finite number T_n such that $P(|\varphi_T(\{z_{T,t}^n\}_{t=1}^{T}) - 1| > n^{-1}) < n^{-1}$ for all $T \geq T_n$. For any T, let $\tilde{n}_T$ be the largest n^* such that $(\max_{n \leq n^*} T_n) < T$. Note that $\tilde{n}_T \to \infty$ as $T \to \infty$, as T_n is finite for any n.

With $n_T = \min([T^{1/5}], \tilde{n}_T)$, let $\{z_{T,t}\}_{t=1}^{T} = \{z_{T,t}^{n_T}\}_{t=1}^{T}$. By construction of $\{z_{T,t}\}_{t=1}^{T}$, $P(|\varphi_T(\{z_{T,t}\}_{t=1}^{T}) - 1| > n_T^{-1}) < n_T^{-1}$, such that $\varphi_T(\{z_{T,t}\}_{t=1}^{T}) \xrightarrow{p} 1$.

Furthermore,

$$T^{-1/2}\sum_{t=1}^{[sT]} z_{T,t} = T^{-1}\sum_{t=1}^{[sT]}\sum_{l=1}^{n_T}\pi(l-\tfrac{1}{2})\phi_l(t/T)\xi_l + T^{-1}\sum_{t=1}^{[sT]}\sum_{l=n_T+1}^{\infty}\phi_l(t/T)\xi_l. \tag{A.1}$$

For the second term on the right-hand side of (A.1), we have

$$\sup_{0\le s\le 1} T^{-1}\left|\sum_{t=1}^{[sT]}\sum_{l=n_T+1}^{\infty}\phi_l(t/T)\xi_l\right| \le \sup_{0\le s\le 1} T^{-1}\sum_{t=1}^{[sT]}\left|\sum_{l=n_T+1}^{\infty}\phi_l(t/T)\xi_l\right| \xrightarrow{\text{a.s.}} 0$$

because $\sup_{0\le s\le 1}|\sum_{l=n_T+1}^{\infty}\phi_l(s)\xi_l| \xrightarrow{\text{a.s.}} 0$ as $n_T\to\infty$ by the result of Itô and Nisio (1968). For the first term on the right-hand side of (A.1), we find

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{[sT]}\sum_{l=1}^{n_T}\pi(l-\tfrac{1}{2})\phi_l(t/T)\xi_l\\
&\quad = T^{-1}\sqrt{2}\sum_{l=1}^{n_T}\xi_l\sum_{t=1}^{[sT]}\sin((l-\tfrac{1}{2})\pi t/T)\\
&\quad = \sqrt{2}\sum_{l=1}^{n_T}\frac{\cos(\pi(l-\frac{1}{2})/(2T)) - \cos(\pi(l-\frac{1}{2})([sT]+\frac{1}{2})/T)}{2T\sin(\pi(l-\frac{1}{2})/(2T))}\,\xi_l
\end{aligned}$$

because for any $0<\omega<\pi$ and $1\le t\le T$, with $\mathbf{i}=\sqrt{-1}$,

$$\begin{aligned}
2\sum_{s=1}^{t}\sin(\omega s) &= \mathbf{i}^{-1}\sum_{s=1}^{t}(e^{\mathbf{i}\omega s}-e^{-\mathbf{i}\omega s})\\
&= \mathbf{i}^{-1}e^{\mathbf{i}\omega}\,\frac{1-e^{\mathbf{i}\omega t}}{1-e^{\mathbf{i}\omega}} - \mathbf{i}^{-1}e^{-\mathbf{i}\omega}\,\frac{1-e^{-\mathbf{i}\omega t}}{1-e^{-\mathbf{i}\omega}}\\
&= -\frac{\mathrm{Im}[(1-e^{\mathbf{i}\omega t})(1-e^{\mathbf{i}\omega})]}{\mathrm{Re}[1-e^{\mathbf{i}\omega}]}\\
&= \frac{\sin(\omega t)+\sin(\omega)-\sin(\omega(t+1))}{1-\cos(\omega)}\\
&= \frac{\cos(\omega/2)-\cos(\omega(t+\frac{1}{2}))}{\sin(\omega/2)},
\end{aligned}$$

where the last equality applies $1-\cos(2u)=2(\sin u)^2$ and $\sin(2u)-\sin(2v)=2\cos(u+v)\sin(u-v)$.

Now for all $T\ge 1$, $\inf_{1\le l\le n_T} 2T\sin(\pi(l-\frac{1}{2})/(2T))\ge \pi/4$, and from Taylor expansions, $\sup_{1\le l\le n_T}|1-\cos(\pi(l-\frac{1}{2})/(2T))|\le \pi n_T/(2T)$, $\sup_{1\le l\le n_T}|2T\sin(\pi(l-\frac{1}{2})/(2T))-\pi(l-\frac{1}{2})|\le \pi^2 n_T^2/(2T)$, and $\sup_{0\le s\le 1, 1\le l\le n_T}|\cos(\pi(l-\frac{1}{2})([sT]+\frac{1}{2})/T)-\cos(\pi(l-\frac{1}{2})s)|\le \pi n_T/(2T)$. Noting that for any $a,\tilde{a}\in\mathbb{R}$ and $b>0$, $\tilde{b}>0$, $|a/b-\tilde{a}/\tilde{b}|\le |a||b-\tilde{b}|/(b\tilde{b})+|a-\tilde{a}|/\tilde{b}$, we obtain

$$\sup_{0\leq s\leq 1}\left|T^{-1}\sum_{t=1}^{[sT]}\sum_{l=1}^{n_T}\pi(l-\tfrac{1}{2})\phi_l(t/T)\xi_l-\sqrt{2}\sum_{l=1}^{n_T}\frac{1-\cos(\pi(l-\frac{1}{2})s)}{\pi(l-\frac{1}{2})}\xi_l\right|$$

$$\leq 12\sqrt{2}\,\frac{n_T^3}{T}\sum_{l=1}^{n_T}|\xi_l|\xrightarrow{\text{a.s.}}0,$$

where the convergence follows from a strong law of large numbers for independent and identically distributed (i.i.d.) random variables and $n_T^4/T \leq T^{-1/5} \to 0$. But $\sqrt{2}\sum_{l=1}^{n_T}(1-\cos(\pi(l-\frac{1}{2})s))\xi_l/(\pi(l-\frac{1}{2}))$ converges to a standard Wiener process a.s. uniformly in $s\in[0,1]$ as $n_T\to\infty$ by applying once more the result of Itô and Nisio (1968), because $\{\sqrt{2}\sin(s\pi(l-\frac{1}{2}))\}_{l=1}^{\infty}$ is an orthonormal basis in $L^2[0,1]$. ∎