ASYMPTOTIC THEORY FOR LOCAL TIME DENSITY ESTIMATION AND NONPARAMETRIC COINTEGRATING REGRESSION

BY

QIYING WANG and PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1270

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2009

http://cowles.econ.yale.edu/

Econometric Theory, **25**, 2009, 710–738. Printed in the United States of America.
doi:10.1017/S0266466608090269

ASYMPTOTIC THEORY FOR LOCAL TIME DENSITY ESTIMATION AND NONPARAMETRIC COINTEGRATING REGRESSION

QIYING WANG
University of Sydney

PETER C.B. PHILLIPS
Yale University
University of Auckland
University of York
and
Singapore Management University

Asymptotic theory is developed for local time density estimation for a general class of functionals of integrated and fractionally integrated time series. The main result provides a convenient basis for developing a limit theory for nonparametric cointegrating regression and nonstationary autoregression. The treatment directly involves local time estimation and the density function of the processes under consideration, providing an alternative approach to the Markov chain and Fourier integral methods that have been used in other recent work on these problems.

1. INTRODUCTION

Since the introduction of unit root and cointegration analysis, linear models have dominated empirical work in the application of these methods. This emphasis on linearity is convenient for practical implementation and accords well with the linear framework of partial summation in which the integrated process and cointegration concepts have been developed. Nonetheless, it is restrictive, especially in view of the attention given elsewhere in modern time series to nonlinear and nonparametric estimation, and the fact that theory models, particularly in economics, often suggest nonlinear responses without being specific regarding functional form. In such situations, nonparametric function estimation offers an alternative that is appealing in applied work.

The authors thank the co-editor and two referees for helpful comments on the original version. Wang acknowledges partial research support from the Australian Research Council. Phillips acknowledges partial research support from a Kelly Fellowship and the NSF under grant, SES 04-142254 and SES 06-47086. Wang can be contacted at qiying@maths.usyd.edu.au. Address correspondence to Peter C.B. Phillips, Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268, USA; e-mail: peter.phillips@yale.edu.

  0266-4666/09 $15.00

For stationary time series data, the theory of nonparametric function estimation and inference is well developed, and the methods are widely used in practice. Density function estimation and nonparametric regression involving stochastically nonstationary time series are not so well developed. In linear parametric autoregression, nonstationarity is known to increase the signal from the regressors in the nonstationary direction, which in turn leads to a corresponding increase in the rate of convergence in estimation in that direction. However, in nonparametric estimation, where the focus of attention is on local behavior, nonstationarity typically reduces both the magnitude of the signal and rates of convergence in comparison with stationary time series. These reductions are explained by the fact that time series such as a random walk have wandering characteristics that reduce the amount of time spent by the process in the locality of any single point. Such time series are also recurrent so that they continue to revisit points in the sample space, making consistent estimation possible, but at a reduced rate of convergence that reflects the amount of time spent by the process in the vicinity of each point. These considerations make the study of nonstationary nonparametric regression considerably different from the stationary case. They also mean that the local time of the process plays an important role in the limit theory.

Early contributions to the study of local time estimation with discrete time series include Akonom (1993) and Borodin and Ibragimov (1995). Phillips and Park (1998) studied nonparametric autoregression in the context of a random walk. Karlsen and Tjøstheim (2001) and Guerre (2004) studied nonparametric estimation for certain nonstationary processes in the framework of recurrent Markov chains. Most recently, Karlsen, Muklebust, and Tjøsthein (2007) developed an asymptotic theory for nonparametric estimation of a time series regression equation involving stochastically nonstationary time series. Karlsen et al. (2007) address the function estimation problem for a possibly nonlinear cointegrating relation, providing an asymptotic theory of estimation and inference for nonparametric forms of cointegration.

The present paper has a similar goal to Karlsen et al. (2007) but offers an alternative approach to the asymptotic theory that we hope has some advantages. Whereas Karlsen et al. (2007) use the framework of null recurrent Markov chains, we use a local time density argument that makes the approach more closely related to conventional nonparametric arguments that rely heavily on kernel density estimation and regression.

The starting point in our development is to show the weak convergence of a general class of functionals to the local time density of a certain limiting stochastic process. The functional class is specifically designed to include the type of kernel averages that appear in standard kernel density estimation, thereby making the results applicable to nonparametric density estimation and regression with nonstationary time series. For instance, if x_t is an integrated process and $K_h(\cdot) = h^{-1}K(\cdot/h)$ is a kernel function depending on some bandwidth h, then asymptotic theory of nonparametric regression involving x_t typically requires that we study sums of the form $\sum_{t=1}^{n} K_h(x_t - x)$ and $\sum_{t=1}^{n} K_h^2(x_t - x)$, as shown in the regression application discussed in Section 3. Such asymptotics necessarily involve two sequences ($n \to \infty$, $h \to 0$), a feature that substantially complicates technical arguments and leads to limit results involving the local features of the limiting stochastic process that are germane to the neighborhood $x_t \sim x$.

To begin, consider a triangular array $x_{k,n}, 1 \le k \le n, n \ge 1$ constructed from some underlying time series (e.g., by standardizing an integrated process x_t by $\sqrt{n}$) and assume that there is a continuous limiting Gaussian process $G(t), 0 \le t \le 1$, such that

$$x_{[nt],n} \Rightarrow G(t) \quad \text{on} \quad D[0,1], \tag{1.1}$$

where $[a]$ denotes the integer part of a and $\Rightarrow$ denotes weak convergence on the Skorohod space $D[0,1]$. The functional of interest S_n of $x_{k,n}$ is defined by the sample average

$$S_n = \frac{c_n}{n} \sum_{k=1}^{n} g(c_n x_{k,n}), \tag{1.2}$$

where c_n is a certain sequence of positive constants and g is a real function on R. As intimated previously, such functionals commonly arise in nonlinear regression with integrated time series (Park and Phillips, 1999, 2001) and nonparametric estimation in relation to nonlinear cointegration models (Phillips and Park, 1998; Karlsen and Tjøstheim, 2001; Karlsen et al., 2007). In such cases, g may be a kernel function K or a squared kernel function K^2, and, when the times series is an integrated process, c_n may take the explicit form $\sqrt{n}/h$ involving both the sample size n and the bandwidth h. We then have a sample average S_n that depends on two primary sequences n and c_n, and the asymptotic development requires that both n and c_n tend to infinity. Given that c_n depends on a bandwidth h that tends to zero, the relative rates of divergence of n and c_n are important in the asymptotics. As we will show, the limit behavior of S_n in the situation that $c_n \to \infty$ and $n/c_n \to \infty$ is particularly interesting and important for practical applications involving nonparametric kernel estimation and regression with nonstationary data.

The present paper derives by direct calculation the limit distribution of S_n when $c_n \to \infty$ and $n/c_n \to \infty$, showing that under very general conditions on the function g and the process $x_{k,n}$

$$S_n \to_D \int_{-\infty}^{\infty} g(x)\, dx\, L(1,0), \tag{1.3}$$

where $L(t,s)$ is the local time of the process $G(t)$ at the spatial point s, as defined in Section 2. When the function g is a kernel density, the integral in (1.3) is unity, and the limit is then the local time of G at the origin. Accordingly, the result reveals that the limit of a nonparametric kernel density of a nonstationary time series is simply the local time of the Gaussian process $G(t)$ to which the standardized nonstationary time series converges weakly. When the array $x_{k,n}$ is suitably recentered at some spatial point away from the origin, the local time in the limit (1.3) is correspondingly recentered at that spatial point.

These results relate to those of Jeganathan (2004), who investigated the asymptotic form of similar functionals when $x_{k,n}$ is the partial sum of a linear process. For the particular situation where $c_n x_{k,n}$ is a partial sum of independent and identically distributed (i.i.d.) random variables, some other related results can be found in the work of Borodin and Ibragimov (1995), Akonom (1993), and Phillips and Park (1998).

As in Jeganathan (2004), the approach in this paper involves approximating the difference

$$\frac{c_n}{n}\sum_{k=1}^{n} g\left(c_n x_{k,n}\right) - \frac{c_n}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty} g\left[c_n\left(x_{k,n}+z\epsilon\right)\right]\phi(z)\,dz,$$

for some ϵ and where $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\left(x^2/2\right)\right\}$. However, unlike Jeganathan (2004), who used a traditional Fourier transformation like that of Borodin and Ibragimov (1995) for dealing with this kind of problem, our treatment directly involves the density function of $x_{k,n}$. In this respect our work is related to the approach used in Pötscher (2004) and Berkes and Horváth (2006). The application of this idea gives the results wide applicability to important practical cases where $x_{k,n}$ is an integrated time series and the limit process is Gaussian, including cases of fractional Brownian motion. It also makes for rather simple and neat derivations.

We mention that the limit distribution of S_n in the situation that $c_n = 1$ is very different from that when $c_n \to \infty$ and $n/c_n \to \infty$. When $c_n = 1$, in a series of papers of increasing generality on the conditions for $x_{k,n}$, $g(x)$, and $G(t)$, Park and Phillips (1999), de Jong (2004), Pötscher (2004), De Jong and Wang (2005), and Berkes and Horváth (2006) proved that

$$\frac{1}{n}\sum_{k=1}^{n} g(x_{k,n}) \to_D \int_0^1 g(G(t))\,dt. \tag{1.4}$$

The limit distribution of S_n in this case is an integral of $G(t)$, and the result may be interpreted as an application of weak convergence in conjunction with a version of the continuous mapping theorem. When $c_n \to \infty$, not only is the limit result different, but the rate of convergence is affected, and the result no longer has a form associated with a continuous map.

Some heuristic arguments help to reveal the nature of these differences. Note first that by virtue of the occupation times formula (see eqn. (2.1)) we may write

$$\int_0^1 g(G(t))\,dt = \int_{-\infty}^{\infty} g(s) L_G(1,s)\,ds, \tag{1.5}$$

where $L_G(1,s)$ is the local time at s of the limit process G over the time interval [0, 1], as considered in Section 2. Next, rewrite the average S_n so that it is indexed by twin sequences c_m and n defining $S_{m,n} = \frac{c_m}{n} \sum_{k=1}^{n} g(c_m x_{k,n})$ and noting that $S_{m,n} = S_n$ when $m = n$. If we hold c_m fixed as $n \to \infty$, then from (1.4) and (1.5) we have

$$S_{m,n} \to_D c_m \int_0^1 g(c_m G(t))\,dt = c_m \int_{-\infty}^{\infty} g(c_m s) L_G(1,s)\,ds$$

$$= \int_{-\infty}^{\infty} g(r) L_G\left(1, \frac{r}{c_m}\right) dr := S_{m,\infty}.$$

Clearly, $S_{m,\infty} \to_D \int_{-\infty}^{\infty} g(r)\,dr\,L_G(1,0)$ as $m \to \infty$, so that (1.3) may be regarded as a limiting version of (1.4). The goal is to turn this sequential argument as $n \to \infty$, followed by $m \to \infty$, into a joint limit argument so that c_n may play an active role in the asymptotics, which is needed when c_n involves the bandwidth parameter in density estimation and kernel regression.

The paper is organized as follows. The next section presents our main results. Theorem 2.1 provides a general framework for the limit theory, and its applications to integrated time series and Gaussian limit processes including fractional Brownian motion are given in the following corollaries. Section 3 investigates further applications of Theorem 2.1, which include nonlinear nonparametric cointegrating regressions and the nonparametric estimation of a unit root autoregression. These applications provide a basis for practical nonparametric work with nonstationary series, including both unit root and nonstationary long memory processes. Section 4 concludes by discussing these results and some possible extensions. Section 5 gives proofs of the main results and corollaries. Throughout the paper we use conventional notation, so that $\to_D$ stands for convergence in distribution and $\to_P$ for convergence in probability. The terms $A, A_1, \ldots$ denote constants that may be different at each appearance.

2. FIRST RESULTS

We start by recalling the definition of local time. The process $\{L_\zeta(t,s), t \geq 0, s \in R\}$ is said to be the local time of a measurable process $\{\zeta(t), t \geq 0\}$ if, for any locally integrable function $T(x)$,

$$\int_0^t T[\zeta(s)]\,ds = \int_{-\infty}^{\infty} T(s) L_\zeta(t,s)\,ds, \quad \text{all } t \in R, \tag{2.1}$$

with probability one. Equation (2.1) is known as the occupation times formula. Roughly speaking, $L_\zeta(t,s)$ is a spatial density that records the relative sojourn time of the process $\zeta(t)$ at the spatial point s over the time interval $[0,t]$. For further discussion, alternative definitions, and the various properties of local time, we refer to Geman and Horowitz (1980) and Revuz and Yor (1999) and to Phillips (2001, 2005) for recent economic applications.

We also define a fractional Brownian motion with $0 < \beta < 1$ on $D[0,1]$ as follows:

$$W_\beta(t) = \frac{1}{A(\beta)} \int_{-\infty}^{0} \left[(t-s)^{\beta-1/2} - (-s)^{\beta-1/2}\right] dW(s) + \int_0^t (t-s)^{\beta-1/2}\, dW(s),$$

where $W(s)$ is a standard Brownian motion and

$$A(\beta) = \left(\frac{1}{2\beta} + \int_0^\infty \left[(1+s)^{\beta-1/2} - s^{\beta-1/2}\right]^2 ds\right)^{1/2}.$$

Note that $W_{1/2}(t)$ is a standard Brownian motion and $W_\beta(t)$ has a continuous local time $L_{W_\beta}(t,s)$ with regard to (t,s) in $[0,\infty) \times R$. See, for example, Geman and Horowitz (1980, Thm. 22.1).

As in Section 1, let $x_{k,n}$, $0 \le k \le n$, $n \ge 1$ (define $x_{0,n} \equiv 0$) be a random triangular array and $g(x)$ be a real measurable function on R. In most practical situations, as in Corollaries 2.1 and 2.2 later in this section, $x_{k,n}$ is equal to x_k/d_n, where x_k is a partial sum and $0 < d_n \to \infty$ in such a way that x_n/d_n has a limit distribution. We make the following assumptions.

Assumption 2.1. $|g(x)|$ and $g^2(x)$ are Lebesgue integrable functions on R with $\tau \equiv \int g(x)\,dx \ne 0$.

Assumption 2.2. There exists a stochastic process $G(t)$ having a continuous local time $L_G(t,s)$ such that $x_{[nt],n} \Rightarrow G(t)$, on $D[0,1]$, where weak convergence is understood with respect to the Skorohod topology on the space $D[0,1]$.

Assumption 2.2*. On a suitable probability space, there exists a stochastic process $G(t)$ having a continuous local time $L_G(t,s)$ such that $\sup_{0\le t\le 1} |x_{[nt],n} - G(t)| = o_P(1)$.

In Assumption 2.3 we shall make use of the notation $\Omega_n(\eta) = \{(l,k) : \eta n \le k \le (1-\eta)n,\ k+\eta n \le l \le n\}$, where $0 < \eta < 1$.

Assumption 2.3. For all $0 \le k < l \le n$, $n \ge 1$, there exist a sequence of constants $d_{l,k,n}$ and a sequence of σ-fields $\mathcal{F}_{k,n}$ (define $\mathcal{F}_{0,n} = \sigma\{\phi, \Omega\}$, the trivial σ-field) such that

(a) for some $m_0 > 0$ and $C > 0$, $\inf_{(l,k)\in\Omega_n(\eta)} d_{l,k,n} \ge \eta^{m_0}/C$ as $n \to \infty$,

$$\lim_{\eta\to 0}\lim_{n\to\infty} \frac{1}{n} \sum_{l=(1-\eta)n}^{n} (d_{l,0,n})^{-1} = 0, \tag{2.2}$$

$$\lim_{\eta\to 0}\lim_{n\to\infty} \frac{1}{n} \max_{0\le k\le(1-\eta)n} \sum_{l=k+1}^{k+\eta n} (d_{l,k,n})^{-1} = 0, \tag{2.3}$$

$$\limsup_{n\to\infty} \frac{1}{n} \max_{0\le k\le n-1} \sum_{l=k+1}^{n} (d_{l,k,n})^{-1} < \infty; \tag{2.4}$$

(b) $x_{k,n}$ are adapted to $\mathcal{F}_{k,n}$ and, conditional on $\mathcal{F}_{k,n}$, $(x_{l,n}-x_{k,n})/d_{l,k,n}$ has a density $h_{l,k,n}(x)$ which is uniformly bounded by a constant K and

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{(l,k)\in\Omega_n[\delta^{1/(2m_0)}]}\sup_{|u|\le\delta}\left|h_{l,k,n}(u)-h_{l,k,n}(0)\right|=0. \tag{2.5}$$

We remark that Assumptions 2.1 and 2.2 are quite weak and likely very close to necessary conditions for this kind of problem. Assumption 2.1 excludes the so-called zero energy case $\int g(x)\,dx=0$, where the limit theory is different and a different convergence rate applies. Assumption 2.2* is a stronger version of Assumption 2.2. In certain situations Assumptions 2.2 and 2.2* are equivalent (e.g., in the situation that $x_{k,n}=\sum_{j=1}^{k}\epsilon_j/\sqrt{n}$, where ϵ_j are i.i.d. random variables with $E\epsilon_1=0$ and $E\epsilon_1^2=1$). If Assumption 2.2 holds and $G(t)$ is a continuous Gaussian process, it follows from the so-called Skorohod–Dudley–Wichura representation theorem (e.g., Shorack and Wellner, 1986, Rmk. 2, p. 49) that $x_{k,n}$ may be replaced by a distributionally equivalent process $x^*_{k,n}$ for which $x^*_{k,n}$ satisfies Assumption 2.2*. This is sufficient for many applications if we are only interested in weak convergence.

As for Assumption 2.3, we may choose $\mathcal{F}_{k,n}=\sigma(x_{1,n},\ldots,x_{k,n})$, the natural σ-fields, and the $d_{l,k,n}$ form a numerical sequence such that, conditional on $\mathcal{F}_{k,n}$, $(x_{l,n}-x_{k,n})/d_{l,k,n}$ has a limit distribution as $l-k\to\infty$. Because $x_{k,n}$ satisfies the functional law (1.1), the appropriate form of the sequence $d_{l,k,n}$ will be apparent in applications. Thus, if $x_{k,n}=\sum_{j=1}^{k}\epsilon_j/\sqrt{n}$, where ϵ_j are i.i.d. random variables with $E\epsilon_1=0$ and $E\epsilon_1^2=1$, we may choose $\mathcal{F}_{k,n}=\sigma(\epsilon_1,\ldots,\epsilon_k)$ and $d_{l,k,n}=\sqrt{l-k}/\sqrt{n}$. More examples are given in Corollaries 2.1 and 2.2 later in this section.

We now state our first main result.

THEOREM 2.1. *Suppose Assumptions 2.1–2.3 hold. Then, for any* $c_n\to\infty$, $c_n/n\to 0$, *and* $r\in[0,1]$,

$$\frac{c_n}{n}\sum_{k=1}^{[nr]}g(c_n x_{k,n})\to_D \tau\, L_G(r,0). \tag{2.6}$$

If Assumption 2.2 is replaced by Assumption 2.2, then, for any* $c_n\to\infty$ *and* $c_n/n\to 0$,

$$\sup_{0\le r\le 1}\left|\frac{c_n}{n}\sum_{k=1}^{[nr]}g(c_n x_{k,n})-\tau\, L_G(r,0)\right|\to_P 0, \tag{2.7}$$

under the same probability space defined as in Assumption 2.2.*

Remarks 2.1. Many examples occur in applications where limit results at spatial points other than the origin are relevant. Phillips (2001) gave examples of hazard rate analyses for inflation series, and Hu and Phillips (2004) analyzed federal funds rate market intervention policy on interest rates. To suit such applications, versions of results (2.6) and (2.7) still hold if $x_{i,n}$ is replaced by $y_{i,n} = x_{i,n} + x\, c_n'$ where $c_n' \to 0$ or $c_n' = 1$ and, respectively, $L_G(r, 0)$ is replaced by

$$L_G^*(r) = \begin{cases} L_G(r,0), & \text{if } c_n' \to 0, \\ L_G(r,-x), & \text{if } c_n' = 1. \end{cases}$$

Indeed, if $x_{i,n}$ satisfies Assumption 2.2 (similarly for Assumption 2.2*), then for any given $x \in R$

$$y_{[nt],n} \Rightarrow \begin{cases} G(t), & \text{if } c_n' \to 0, \\ G(t)+x, & \text{if } c_n' = 1. \end{cases}$$

If $x_{i,n}$ satisfies Assumption 2.3 then $y_{i,n}$ also satisfies Assumption 2.3. The claim follows directly from Theorem 2.1 and the fact that $G(t) + x$ has local time $L_G(t, s - x)$.

In what follows we consider applications of Theorem 2.1 to Gaussian processes and general linear processes. Further applications will be investigated in Section 3, where we consider the nonparametric estimation of a nonlinear cointegration regression model.

COROLLARY 2.1. *Suppose Assumption 2.1 holds. Let* $\{\xi_j, j \geq 1\}$ *be a stationary sequence of Gaussian random variables with* $\mathrm{E}\xi_1 = 0$ *and covariances* $\gamma(j-i) = \mathrm{E}\xi_i\xi_j$ *satisfying the following condition, for some* $0 < \alpha < 2$ *and* $\lambda < 1$,

$$d_n^2 \equiv \sum_{1 \le i,j \le n} \gamma(j-i) \sim n^\alpha h(n) \quad \textit{and} \quad |\tilde{\gamma}_{l,k}| \le \lambda\, d_k\, d_{l-k}, \tag{2.8}$$

as $\min\{k, l-k\} \to \infty$, *where* $h(n)$ *is a slowly varying function at* ∞ *and*

$$\tilde{\gamma}_{l,k} = \sum_{i=1}^{k} \sum_{j=k+1}^{l} \gamma(j-i).$$

Let $S_k = \sum_{j=1}^k \xi_j$, $1 \le k \le n$. *Then, for* $r \in [0,1]$ *and any* $c_n > 0$ *satisfying* $c_n n^{\alpha/2} \sqrt{h(n)} \to \infty$ *and* $c_n \sqrt{h(n)}/n^{1-\alpha/2} \to 0$,

$$\frac{c_n \sqrt{h(n)}}{n^{1-\alpha/2}} \sum_{k=1}^{[nr]} g(c_n S_k) \to_D \tau\, L_{W_{\alpha/2}}(r, 0). \tag{2.9}$$

Remarks 2.2. Note that $d_n^2 = \mathrm{E}S_n^2$ and $\tilde{\gamma}_{l,k} = \mathrm{cov}(S_k, S_l - S_k)$. Condition (2.8) is quite weak. For instance, if one of the following conditions is satisfied, then (2.8) holds:

(a) $\gamma(j) = \mathrm{E}(\xi_1\xi_{1+j}) \ge 0$ for all $j \ge 0$ and $\sum_{j=0}^\infty \gamma(j) < \infty$;

(b) $\gamma(k) = \mathrm{E}(\xi_1\xi_{1+k}) \sim C\, k^{-\mu}$ with some $0 < \mu < 1$ and $C > 0$;

(c) $\gamma(k) = \mathrm{E}(\xi_1\xi_{1+k}) \sim -C\,k^{-\mu}$ with some $1 < \mu < 2$, $C > 0$, and $\gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k) = 0$.

Indeed, in situation (a), it is readily seen that $d_n^2 \sim C\,n$ with some constant $C > 0$ and as $\min\{k, l-k\} \to \infty$,

$$\tilde{\gamma}_{l,k} = \sum_{i=0}^{k-1}\sum_{j=1}^{l-k}\gamma(j+i) = o(1)\min\{k, l-k\} \le \frac{1}{2}\,d_k^{1/2}\,d_{l-k}^{1/2}.$$

In both situations (b) and (c), it follows from Taqqu (1975, Lem. 5.1) (also see Berkes and Horváth, 2006, Exp. 2.3) that $d_n^2 = \mathrm{E}S_n^2 \sim K\,n^{2-\mu}$, where K is a constant depending only on μ and C. This yields the first part of (2.8). On the other hand, it can be easily seen that, as $\min\{k, l-k\} \to \infty$,

$$|\tilde{\gamma}_{l,k}| = \frac{1}{2}\left|\mathrm{E}S_l^2 - \mathrm{E}(S_l - S_k)^2 - \mathrm{E}S_k^2\right|$$
$$\sim \frac{1}{2}K\left|l^{\alpha} - (l-k)^{\alpha} - k^{\alpha}\right| \le \frac{1}{2}(1+\varsigma)\max\{1, 2-\mu\}\,d_k\,d_{l-k},$$

for arbitrary $\varsigma > 0$, where we have used the fact that

$$|(x+y)^{\alpha} - x^{\alpha} - y^{\alpha}| \le \max\{1,\alpha\}x^{\alpha/2}y^{\alpha/2}, \qquad x, y \ge 0, \qquad 0 < \alpha < 2.$$

Recall that $0 < \mu < 2$. By letting $\varsigma = \varsigma_0$ be sufficiently small, the second part of (2.8) follows with $\lambda = \frac{1}{2}(1+\varsigma_0)\max\{1, 2-\mu\} < 1$.

COROLLARY 2.2. *Let Assumption 2.1 hold. Let* $\{\xi_j, j \ge 1\}$ *be a sequence of linear processes defined by*

$$\xi_j = \sum_{k=0}^{\infty}\psi_k\,\epsilon_{j-k},$$

where $\{\epsilon_j, -\infty < j < \infty\}$ *is a sequence of i.i.d. random variables with* $\mathrm{E}\epsilon_0 = 0$, $\mathrm{E}\epsilon_0^2 = 1$, *and characteristic function* $\varphi(t)$ *of* ϵ_0 *satisfying* $\int_{-\infty}^{\infty}|\varphi(t)|\,dt < \infty$. *Let* $S_k = \sum_{j=1}^{k}\xi_j$, $1 \le k \le n$ *and* $d_n^2 = \mathrm{E}S_n^2$.

(i) If $\psi_k \sim k^{-\mu}\,h(k)$, *where* $1/2 < \mu < 1$ *and* $h(k)$ *is a function slowly varying at* ∞, *then* $d_n^2 \sim c_\mu\,n^{3-2\mu}\,h^2(n)$ *with* $c_\mu = \frac{1}{(1-\mu)(3-2\mu)}\int_0^{\infty}x^{-\mu}(x+1)^{-\mu}dx$ *and, for* $r \in [0,1]$ *and any* $c_n > 0$ *satisfying* $c_n n^{3/2-\mu}h(n) \to \infty$ *and* $c_n\,h(n)/n^{\mu-1/2} \to 0$,

$$\frac{c_n\,h(n)}{n^{\mu-1/2}}\sum_{k=1}^{[nr]}g(c_n\,S_k) \to_D \left(\sqrt{c_\mu}\right)^{-1}\tau\,L_{W_{3/2-\mu}}(r,0). \tag{2.10}$$

(ii) If $\sum_{k=0}^{\infty}|\psi_k| < \infty$ *and* $\psi \equiv \sum_{k=0}^{\infty}\psi_k \ne 0$, *then* $d_n^2 \sim \psi^2\,n$ *and, for* $r \in [0,1]$ *and any* $c_n > 0$ *satisfying* $c_n\sqrt{n} \to \infty$ *and* $c_n/\sqrt{n} \to 0$,

$$\frac{c_n}{\sqrt{n}}\sum_{k=1}^{[nr]}g(c_n\,S_k) \to_D \psi^{-1}\,\tau\,L_W(r,0). \tag{2.11}$$

Remarks 2.3. Corollary 2.2(a) provides a result similar to Theorem 3 of Jeganathan (2004), who considered the more general situation where ϵ_0 is in the domain of attraction of the stable law. It is possible to restate our corollary in the same setting. However, this is not essential for our purpose in the present paper, and we therefore omit the details. Corollary 2.2(b) essentially improves and extends similar results obtained in Akonom (1993), Park and Phillips (1999), and others.

Remarks 2.4. Consider a fractionally integrated process $\{Z_t\}$ initialized at $Z_0 = 0$ and defined by

$$(1-B)^{d+1} Z_t = \epsilon_t, \tag{2.12}$$

where $0 \le d < 1/2$, B is a backshift operator, and $\{\epsilon_j, -\infty < j < \infty\}$ is a sequence of i.i.d. random variables with $\mathrm{E}\epsilon_0 = 0$, $\mathrm{E}\epsilon_0^2 = 1$, and characteristic function $\varphi(t)$ of ϵ_0 satisfying $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$. The fractional difference operator $(1-B)^\gamma$ is defined by its Maclaurin series (by its binomial expansion, if γ is an integer):

$$(1-B)^\gamma = \sum_{j=0}^{\infty} \frac{\Gamma(-\gamma+j)}{\Gamma(-\gamma)\Gamma(j+1)} B^j \quad \text{where } \Gamma(z) = \begin{cases} \int_0^\infty s^{z-1} e^{-s} ds & \text{if } z > 0 \\ \infty & \text{if } z = 0. \end{cases}$$

If $z < 0$, $\Gamma(z)$ is defined by the recursion formula $z\Gamma(z) = \Gamma(z+1)$.

Write (2.12) as $\Delta Z_t = (1-B)^{-d}\epsilon_t$ and then Z_n has the partial sum form $Z_n = \sum_{t=1}^{n} Z_t^*, n \ge 1$, where $Z_t^* = \sum_{k=0}^{\infty} a(k)\, \epsilon_{t-k}$ with

$$a(k) = \frac{\Gamma(k+d)}{\Gamma(k+1)\,\Gamma(d)} \sim \frac{1}{\Gamma(d)} k^{d-1},$$

as $k \to \infty$. It follows easily from Corollary 2.2 that, for any $r \in [0,1]$, $0 \le d < 1/2$ and c_n satisfying $c_n\, n^{d-1/2} \to 0$ and $c_n\, n^{d+1/2} \to \infty$, we have

$$\frac{c_n}{n^{1/2-d}} \sum_{k=1}^{[nr]} g(c_n Z_n) \to_D C_0^{-1} \int_{-\infty}^{\infty} g(x)\, dx\, L_{W_{1/2+d}}(r,0),$$

where

$$C_0^2 = c_{1-d} / \Gamma^2(d) = \Gamma(1-2d) / \big\{(1+2d)\Gamma(1+d)\Gamma(1-d)\big\}$$

and c_{1-d} is defined by c_μ of Corollary 2.2 with $\mu = 1-d$.

3. NONPARAMETRIC COINTEGRATING REGRESSION

Consider a nonlinear cointegrating regression model:

$$y_t = f(x_t) + u_t, \qquad t = 1, 2, \ldots, n, \tag{3.1}$$

where u_t is a stationary error process and x_t is a nonstationary regressor. Let $K(x)$ be a nonnegative real function and set $K_h(s) = h^{-1}K(s/h)$ where $h \equiv h_n \to 0$. The conventional kernel estimate of $f(x)$ in model (3.1) is given by

$$\hat{f}(x) = \frac{\sum_{t=1}^{n} y_t K_h(x_t - x)}{\sum_{t=1}^{n} K_h(x_t - x)}. \tag{3.2}$$

The limit behavior of $\hat{f}(x)$ has recently been investigated in Karlsen et al. (2007) in the situation where x_t is a recurrent Markov chain. (For related work on nonlinear, nonstationary regressions, see also Phillips and Park, 1998; Karlsen and Tjøstheim 2001; Guerre, 2004; Bandi, 2004). The main theorem in Karlsen, et al. (2007, Thm. 3.1) relies on the asymptotic theory developed in Karlsen and Tjøstheim (2001) involving the conditions on the invariant measure associated with a recurrent Markov chain. These conditions are not always easy to check in practice and do not include some cases of econometric interest such as fractional processes.

This section provides an alternative approach to nonparametric cointegration by making direct use of Theorem 2.1 in developing the asymptotics. In particular, instead of the recurrent Markov chain in Karlsen et al. (2007), we work with partial sum representations of the type $x_t = \sum_{j=1}^{t} \xi_j$ where ξ_j is a Gaussian process or a general linear process defined in Corollary 2.1 or 2.2 and use Theorem 2.1 to obtain the limit behavior for kernel functions of this process. This specification corresponds to the conventional formulation of unit root and cointegration models, and the limit theory has links to traditional nonparametric asymptotics for stationary models even though rates of convergence are different. This approach also allows us to work with cases where the regressor x_t is a nonstationary long memory time series.

The estimation error in the kernel estimator (3.2) has the usual decomposition

$$\hat{f}(x) - f(x) = \frac{\sum_{t=1}^{n} u_t K_h(x_t - x)}{\sum_{t=1}^{n} K_h(x_t - x)} + \frac{\sum_{t=1}^{n} \left[f(x_t) - f(x)\right] K_h(x_t - x)}{\sum_{t=1}^{n} K_h(x_t - x)}. \tag{3.3}$$

The second term of (3.3) affects bias, and, at least when this is of smaller order, it is the first term that determines the asymptotic distribution. Observe that in the special case where the sequence u_t is *iid* $N(0, \sigma^2)$ and independent of x_t, then, given $\{x_t\}_{t=1}^{n}$, the conditional distribution of $\sum_{t=1}^{n} u_t K_h(x_t - x) / \sum_{t=1}^{n} K_h(x_t - x)$ is the same as

$$\sigma \left(\frac{\sum_{t=1}^{n} K_h^2(x_t - x)}{\left(\sum_{t=1}^{n} K_h(x_t - x)\right)^2} \right)^{1/2} Z, \tag{3.4}$$

where Z is standard $N(0, 1)$. It is clear that even in this simple case the limit distribution is driven by the behavior of the ratio multiplying Z, whose components involve sums that are of the same form as those in (1.2). This explains why results such as (1.3) and Theorem 2.1 are so useful in studying the behavior of the kernel estimator $\hat{f}(x)$.

Our first theorem assumes that the x_t are independent of u_t. We relax this independence condition in the second theorem. Throughout the section we make use of the following assumptions.

Assumption 3.1. The kernel K satisfies that $\int_{-\infty}^{\infty} K(s)\,ds = 1$ and $\sup_s K(s) < \infty$.

Assumption 3.2. For given x, there exists a real function $f_1(s,x)$ and is $0 < \gamma \leq 1$ such that, when h sufficiently small, $|f(hy+x) - f(x)| \leq h^{\gamma} f_1(y,x)$ for all $y \in R$ and $\int_{-\infty}^{\infty} K(s) f_1(s,x)\,ds < \infty$.

Assumption 3.3. $(u_t, \mathcal{F}_t, 1 \leq t \leq n)$ is a martingale difference with $\mathrm{E}(u_t^2|F_{t-1}) \to_{a.s.} \sigma^2 > 0$ as $t \to \infty$ and $\sup_{1\leq t\leq n} \mathrm{E}(|u_t|^q \mid \mathcal{F}_{t-1}) < \infty$ a.s. for some $q > 2$.

Assumption 3.4. There exists a sequence $0 < d_n \to \infty$ for which $d_n = o(n)$ and such that $x_{i,n} = x_i/d_n$, $1 \leq i \leq n, n \geq 1$, satisfies Assumption 2.3.

Assumption 3.5. There exists a continuous Gaussian process $G(t)$ having a continuous local time $L_G(t,s)$ such that $x_{[nt],n} \Rightarrow G(t)$, on $D[0,1]$, where d_n and $x_{i,n} = x_i/d_n$ are defined as in Assumption 3.4 and weak convergence is understood with respect to the Skorohod topology on the space $D[0,1]$.

Our first result on the limit theory for nonparametric cointegrating regression is as follows.

THEOREM 3.1. *Suppose Assumptions 3.1–3.5 hold and* $(x_t)_1^n$ *is independent of* $(u_t)_1^n$. *Then, for any h satisfying* $nh/d_n \to \infty$ *and* $h \to 0$,

$$\hat{f}(x) \to_p f(x). \tag{3.5}$$

Furthermore, for any h satisfying $nh/d_n \to \infty$ *and* $nh^{1+2\gamma}/d_n \to 0$,

$$\left(h \sum_{t=1}^{n} K_h(x_t - x) \right)^{1/2} (\hat{f}(x) - f(x)) \to_D N\left(0, \sigma_1^2\right), \tag{3.6}$$

where $\sigma_1^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(s)\,dt$.

Remarks 3.1. The conditions in Assumptions 3.1 and 3.2 are quite weak and are simply verified for various kernels $K(x)$ and regression functions $f(x)$. For instance, if $K(x)$ is a standard normal kernel or has a compact support as in Karlsen et al. (2007), a wide range of regression functions $f(x)$ are included. Thus, commonly occurring functions such as $f(x) = |x|^{\beta}$ and $f(x) = 1/(1 + |x|^{\beta})$ for some $\beta > 0$ satisfy Assumption 3.2 with $\gamma = \min\{\beta, 1\}$. Assumption 3.3 is a standard error condition in correctly specified stationary models. If we add more restrictions on d_n and h as in Karlsen et al. (2007), this assumption may be replaced by a stationary linear process condition, so the martingale difference condition is not necessary. Further, the independence of x_t and u_t may be partly relaxed, as shown subsequently. Finally, by noting that fractional Brownian motion $W_\beta(t)$ is a continuous Gaussian process, the processes $x_t = \sum_{j=1}^{t} \xi_j$, where ξ_t is the Gaussian process defined in Corollary 2.1 $[d_n^2 = n^\beta h(n), 0 < \beta < 2]$, the linear process defined in Corollary 2.2 $\left[d_n^2 = c_\mu n^{3-2\mu} h^2(n), 1/2 < \mu < 1\right]$, and the fractional process Z_t^* (in this case, x_t is an $I(d+1)$ process) defined in Remark 2.4 $\left[d_n^2 = C_0^2 n^{1+2d}, 0 \le d < 1/2\right]$ all satisfy Assumptions 3.4 and 3.5. Thus, Theorem 3.1 and result (3.6) have a wide range of application for nonstationary series. As an example of their theory, Karlsen et al. (2007) established (3.6) for an $I(1)$ time series x_t. However, unlike the present approach, it seems that their method cannot be extended to fractional processes, such as the $I(d+1)$ process defined in Remark 2.4.

Remarks 3.2. The result (3.5) implies that $\hat{f}(x)$ is a consistent estimate of $f(x)$. In fact, as shown in the proof of Theorem 3.1 in Section 5, we may obtain

$$\hat{f}(x) - f(x) = o_P\left\{a_n\left[h^\gamma + \sqrt{d_n/(nh)}\right]\right\}, \tag{3.7}$$

where γ is defined in Assumption 3.2 and a_n diverges to infinity as slowly as required. This indicates that a possible "optimal" bandwidth h that yields the best rate in (3.7) or the minimal $\mathrm{E}(\hat{f}(x) - f(x))^2$ satisfies

$$h^* \sim a \ \mathrm{argmin}_h \left\{h^\gamma + \sqrt{d_n/(nh)}\right\} \sim a' (d_n/n)^{1/(1+2\gamma)},$$

where a and a' are positive constants. The choice of the optimal bandwidth h in the present context requires a very detailed analysis of the asymptotic bias, and we leave developments in this direction for later work.

Remarks 3.3. It is interesting to notice that the bandwidth h needs to satisfy certain rate conditions to ensure that the stated asymptotic normality applies. For instance, in the most common situation where $d_n = \sqrt{n}$ and $\gamma = 1$ (e.g., when the ϵ_t are i.i.d. random variables), we require $nh^2 \to \infty$ and $nh^6 \to 0$. This can be explained as follows. In stationary nonparametric models the convergence rate of a kernel regression estimate is $\sqrt{nh}$, requiring that $nh \to \infty$. Undersmoothing in such regressions to avoid bias typically requires that $h = o(n^{-1/5})$. In the nonstationary case, the amount of time spent by the process around any particular spatial point is of order $\sqrt{n}$ rather than n, so that the corresponding rate in such regressions is now $\sqrt{\sqrt{n}h}$, which requires that $nh^2 \to \infty$. Undersmoothing to remove asymptotic bias in this situation typically requires a rate smaller than that in the stationary case. Here we find that the rate $h = o(n^{-1/6})$ is sufficient for undersmoothing. Also note that the choice of bandwidth h is related to the nature of the nonstationary regressor. For instance, in the situation of practical interest where $\gamma = 1$ and the nonstationary regressor x_t is the $I(d+1)$ process Z_t defined in Remark 2.4, we require $n^{1/2-d}h \to \infty$ and $n^{1/2-d}h^3 \to 0$. The bandwidth h is then related to the fractional differentiation index $d \in [0, 1/2)$.

Our next theorem considers the effect of some relaxation of the restriction on the independence between x_t and u_t. To do so, denote the stochastic processes U_n and V_n on $D[0,1]$ by

$$U_n(r) = x_{[nr],n} \quad \text{and} \quad V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t,$$

where d_n and $x_{i,n} = x_i/d_n$ are defined as in Assumption 3.4.

THEOREM 3.2. *Suppose Assumptions 3.1–3.4 hold. Suppose that, for each $n \geq 2$, $x_{i,n}$ is adapted to $\mathcal{F}_{i-1}$, $2 \leq i \leq n$, and $(U_n, V_n) \Rightarrow_D (U, V)$ on $D[0,1]^2$ as $n \to \infty$, where (U, V) is correlated vector Brownian motion. Then (3.3) still holds true for any h satisfying $nh/d_n \to \infty$ and $nh^{1+2\gamma}/d_n \to 0$.*

Remarks 3.4. The preceding result applies to nonparametric cointegration models such as

$$y_t = f(x_{t-\ell}) + u_t, \qquad t = 1, 2, \ldots, n, \tag{3.8}$$

defined for some integer lag $\ell \geq 1$ and where u_t and x_t are as in (3.1). Models of this type arise, for example, in the study of inefficiencies in asset pricing where there may be delayed responses between nonstationary macroeconomic fundamentals or certain financial series and asset return variables. In such cases, the regressor is predetermined, and the error process is still a martingale difference.

Remarks 3.5. Theorem 3.2 can also be used to construct a limit theory for a nonparametric kernel estimate of $m(x)$ in the unit root autoregressive model

$$y_t = m(y_{t-1}) + u_t, \qquad m(y_{t-1}) = \alpha\, y_{t-1}, \quad a.s.$$

with $\alpha = 1$ and $y_0 = 0$. This model is an instance of a simple cointegrated system where there is (trivial) parametric linear cointegration between y_t and y_{t-1} but that is fitted by nonparametric regression because the form of the autoregression is unknown. To illustrate how the theory is applied, let u_t be a sequence of i.i.d. random variables with $\mathrm{E}u_0 = 1$, $\mathrm{E}u_0^2 = 1$, $\mathrm{E}|u_0|^q < \infty$ for some $q > 2$ and the characteristic function $\varphi(t)$ of u_0 satisfying $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$. As in (3.2), the conventional kernel estimate of $m(x)$ is given as follows:

$$\hat{m}(x) = \frac{\sum_{t=1}^{n} y_t K_h(y_{t-1} - x)}{\sum_{t=1}^{n} K_h(y_{t-1} - x)}.$$

In this case, $x_{i,n} = y_{i-1} = \sum_{t=1}^{i-1} u_t$, and the stochastic processes U_n and V_n on $D[0,1]$ are defined by

$$U_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]-1} u_t \quad \text{and} \quad V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t.$$

By letting $\mathcal{F}_i = \sigma\{u_1, u_2, \ldots, u_i\}$, it is easy to check that the $x_{i,n}$ are $\mathcal{F}_{i-1}$ measurable and $(U_n, V_n) \Rightarrow_D (W, W)$ on $D[0,1]^2$ because $V_n(r) \Rightarrow_D W(r)$ on $D[0,1]$ and $\sup_r |U_n(r) - V_n(r)| \le \sup_r |u_{[nr]}|/\sqrt{n} \to_P 0$. It therefore follows from Theorem 3.2 that

$$\left(h \sum_{t=1}^{n} K_h(y_{t-1} - x)\right)^{1/2} (\hat{m}(x) - m(x)) \to_D N(0, \sigma_1^2), \tag{3.9}$$

where $\sigma_1^2 = \int_{-\infty}^{\infty} K^2(s)\, dt$. Result (3.9) provides a simple demonstration that kernel autogression in the case of a unit root is asymptotically normal upon standardization in the usual way. However, the implied convergence rate is slower than that in stationary nonparametric autoregession and much slower than the parametric rate in the unit root case, as found in Phillips and Park (1998) and Guerre (2004).

4. CONCLUSION

The main advantage of the approach adopted here is its simplicity. Just as sample averages of a kernel function of a strictly stationary time series inform us about the probability density of the time series at some locality, the same sample averages of an integrated process provide local spatial density information about the trajectories of the process. The fact that the rates of convergence differ between the two cases simply reflects the fact that integrated time series wander over the entire sample space and spend only $O(\sqrt{n})$ of the sample time in the vicinity of particular points such as the origin. The proofs of the results given here on local time density estimation and nonparametric cointegrating regression take advantage of these characteristics. In some respects, such as with the randomly normalized error in the nonparametric regression estimate (3.6), the results appear to relate closely to conventional nonparametric asymptotics. In other respects, such as the nature of the limit process (2.6) and the rate of convergence of the cointegrating regression estimate (3.6), the results are quite different from those of conventional kernel estimates and kernel regression.

The nonparametric formulation of cointegrating relations seems important in many different empirical applications, especially in view of the fact that economic variables are frequently considered to be driven by fundamentals that have random wandering characteristics. Nonparametric treatment of such relations is appealing because the nature of the functional dependence on fundamentals is seldom specified. The limit distribution theory of Karlsen et al. (2007) and the present paper on the kernel estimation of such relations provides a foundation for empirical work in this context. Further work seems desirable on many different econometric aspects of this central problem, such as dealing with endogeneous regressor issues and rules for bandwidth selection.

5. PROOFS OF THEOREMS

This section provides proofs of the main results. The proof of Theorem 2.1 is simple and mainly uses conventional asymptotic arguments.

Proof of Theorem 2.1. Write

$$L_n^{(r)} = \frac{c_n}{n}\sum_{k=1}^{[nr]} g\big(c_n\, x_{k,n}\big), \qquad L_{n,\epsilon}^{(r)} = \frac{c_n}{n}\sum_{k=1}^{[nr]} \int_{-\infty}^{\infty} g\big[c_n\,(x_{k,n}+z\epsilon)\big]\phi(z)\,dz,$$

where $\phi(x)=\phi_1(x)$ with $\phi_\epsilon(x)=\big(1/\epsilon\sqrt{2\pi}\big)\exp\big\{-(x^2/2\epsilon^2)\big\}$. By a similar argument to the proof of Lemma 7 of Jeganathan (2004), we have that, for any $\epsilon>0$,

$$L_{n,\epsilon}^{(r)} - \frac{\tau}{n}\sum_{k=1}^{[nr]}\phi_\epsilon(x_{k,n}) = o_P(1), \tag{5.1}$$

uniformly in $r\in[0,1]$. Now Theorem 2.1 will follow if we prove that

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{0\le r\le 1} \mathrm{E}|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0. \tag{5.2}$$

Indeed it follows from the continuous mapping theorem that, for $\forall\epsilon>0$ and any $r\in[0,1]$,

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{[nr]}\phi_\epsilon(x_{k,n}) &= \int_0^r \phi_\epsilon(x_{[nt],n})\,dt - \frac{1}{n}\phi_\epsilon(0) + \frac{1}{n}\phi_\epsilon(x_{n,[nr],n})\\ &\to_D \int_0^r \phi_\epsilon(G(t))\,dt.\end{aligned} \tag{5.3}$$

Furthermore, by recalling that $L(t,s)$ is a continuous local time process satisfying (2.1),

$$\int_0^r \phi_\epsilon(G(t))\,dt = \int_{-\infty}^{\infty}\phi(x)L(r,\epsilon x)\,dx = L(r,0)+o_{a.s.}(1), \tag{5.4}$$

as $\epsilon\to 0$. By (5.1)–(5.4), we obtain (2.6). The proof of (2.7) is the same except that we replace (5.3) by

$$\begin{aligned}&\sup_{0\le r\le 1}\Big|\frac{1}{n}\sum_{k=1}^{[nr]}\phi_\epsilon(x_{k,n}) - \int_0^r \phi_\epsilon(G(t))\,dt\Big|\\ &\quad\le \int_0^1 \big|\phi_\epsilon(x_{[nt],n}) - \phi_\epsilon(G(t))\big|\,dt + \frac{2}{n}\\ &\quad\le A(\epsilon)\sup_{0\le t\le 1}|x_{[nt],n}-G(t)| + 2/n \;\to_P\; 0,\end{aligned}$$

as $n\to\infty$.

We next prove (5.2). Write $Y_{k,n}(z) = g[c_n x_{k,n}] - g[c_n(x_{k,n} + z\epsilon)]$. Because $\int_{-\infty}^{\infty} \phi(x)\,dx = 1$, it is readily seen that

$$\sup_{0\le r\le 1} \mathrm{E}|L_n^{(r)} - L_{n,\epsilon}^{(r)}| \le \int_{-\infty}^{\infty} \frac{c_n}{n} \sup_{0\le r\le 1} \mathrm{E}\Big|\sum_{k=1}^{[nr]} Y_{k,n}(z)\Big| \phi(z)\,dz. \tag{5.5}$$

Recall that $x_{k,n}/d_{k,0,n}$ has a density $h_{k,0,n}(x)$ that is bounded by a constant K for all x, $1 \le k \le n$ and $n \ge 1$. For all $z \in R$ and $1 \le k \le n$, we have

$$\begin{aligned} c_n \mathrm{E}\big|Y_{k,n}(z)\big| &= c_n \int_{-\infty}^{\infty} \Big| g\left[c_n\left(d_{k,0,n}\, x + z\epsilon\right)\right] - g\left(c_n\, d_{k,0,n}\, x\right) \Big| h_{k,0,n}(x)\,dx \\ &\le \frac{A}{d_{k,0,n}} \int_{-\infty}^{\infty} \big|g(x + c_n z\epsilon) - g(x)\big|dx \le 2A \int_{-\infty}^{\infty} \big|g(x)\big|\,dx/d_{k,0,n}. \end{aligned} \tag{5.6}$$

Hence, for each $z \in R$, $\frac{c_n}{n} \sup_{0\le r\le 1} \mathrm{E}|\sum_{k=1}^{[nr]} Y_{k,n}(z)| \le A_1 \frac{1}{n} \sum_{k=1}^{n} 1/d_{k,0,n} < \infty$, by (2.4). This, together with (5.5) and the dominated convergence theorem, implies that, to prove (5.2), it suffices to show that, for each fixed z,

$$\Lambda_n(\epsilon) \equiv \frac{c_n^2}{n^2} \sup_{0\le r\le 1} \mathrm{E}\Big[\sum_{k=1}^{[nr]} Y_{k,n}(z)\Big]^2 \to 0, \tag{5.7}$$

when $n \to \infty$ first and then $\epsilon \to 0$. We may represent $\Lambda_n(\epsilon)$ as

$$\begin{aligned} \Lambda_n(\epsilon) &\le \frac{c_n^2}{n^2} \sum_{k=1}^{n} \mathrm{E}\,Y_{k,n}^2(z) + \frac{2c_n^2}{n^2} \sum_{k=1}^{n} \sum_{l=k+1}^{n} \big|\mathrm{E}\{Y_{k,n}(z)\,Y_{l,n}(z)\}\big| \\ &= \Lambda_{1n}(\epsilon) + \Lambda_{2n}(\epsilon), \quad \text{say.} \end{aligned}$$

Because $g^2(x)$ is integrable, by a similar argument as in the proof of (5.6), we have

$$\Lambda_{1n}(\epsilon) \le \frac{Ac_n}{n^2} \sum 1/d_{k,0,n} \le A_1 c_n/n \to 0.$$

We next prove $\lim_{\epsilon\to 0} \lim_{n\to\infty} \Lambda_{2n}(\epsilon) = 0$, and then (5.7) follows accordingly. Write $\Omega_n = \Omega_n\left(\epsilon^{1/(2m_0)}\right)$. Recall that by Assumption 2.3 the $x_{k,n}$ are adapted to $\mathcal{F}_{k,n}$ and, conditional on $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a density $h_{l,k,n}(x)$ that is bounded by a constant K. We obtain

$$\begin{aligned} &c_n d_{l,k,n} \left|\mathrm{E}(Y_{l,n} \mid \mathcal{F}_{k,n})\right| \\ &\quad = c_n d_{l,k,n} \Big| \int_{-\infty}^{\infty} \Big(g\left[c_n x_{k,n} + c_n d_{l,k,n}\, y\right] \\ &\qquad\qquad - g\left[c_n(x_{k,n} + z\epsilon) + c_n d_{l,k,n}\, y\right]\Big) h_{l,k,n}(y)\,dy \Big| \end{aligned}$$

$$\le \int_{-\infty}^{\infty} |g(y)| \left| V(y, c_n x_{k,n}) \right| dy$$

$$\le \begin{cases} A, & \text{if } (l,k) \notin \Omega_n \\ A \int_{|y|\ge\sqrt{c_n}} |g(y)| dy + \int_{|y|\le\sqrt{c_n}} |g(y)| |V(y, c_n x_{k,n})| dy, & \text{if } (l,k) \in \Omega_n, \end{cases}$$

where $V(y,t) = h_{l,k,n}\left(\frac{y-t}{c_n d_{l,k,n}}\right) - h_{l,k,n}\left(\frac{y-t-c_n z\epsilon}{c_n d_{l,k,n}}\right)$. Note that $\inf_{(l,k)\in\Omega_n} d_{l,k,n} \ge \epsilon^{1/2}/C$ and for any given $\epsilon > 0$, $c_n \to \infty$ implies that $c_n \ge 1/\epsilon$ when n is large enough. We further have

$$|V(y,t)| \le \left| h_{l,k,n}\left(\frac{y-t}{c_n d_{l,k,n}}\right) - h_{l,k,n}(0) \right| + \left| h_{l,k,n}\left(\frac{y-t-c_n z\epsilon}{c_n d_{l,k,n}}\right) - h_{l,k,n}(0) \right|$$

$$\le 2 \sup_{|u|\le C(1+z)\epsilon^{1/2}} |h_{l,k,n}(u) - h_{l,k,n}(0)|,$$

whenever $|y| \le \sqrt{c_n}$ and $|t| \le \sqrt{c_n} + c_n z\epsilon$. Now, as in the proof of (5.6), whenever $|y| \le \sqrt{c_n}$, n is large enough, and $(l,k) \in \Omega_n$,

$$\mathrm{E}\,|Y_{k,n}(z)|\,|V(y, c_n x_{k,n})|$$

$$= \int_{-\infty}^{\infty} \left| g\left[c_n (d_{k,0,n} x + z\epsilon)\right] - g\left(c_n d_{k,0,n} x\right) \right| |V(y, c_n d_{k,0,n} x)| h_{k,0,n}(x)\, dx$$

$$\le \frac{A}{c_n d_{k,0,n}} \int_{-\infty}^{\infty} \left| g(x + c_n z\epsilon) - g(x) \right| |V(y,x)|\, dx$$

$$\le \frac{A}{c_n d_{k,0,n}} \int_{-\infty}^{\infty} \left| g(x) \right| \left[|V(y,x)| + |V(y, x - c_n z\epsilon)| \right] dx$$

$$\le \frac{A}{c_n d_{k,0,n}} \left[\int_{|x|\ge\sqrt{c_n}} |g(x)|\, dx + \sup_{|u|\le C(1+z)\epsilon^{1/2}} |h_{l,k,n}(u) - h_{l,k,n}(0)| \right],$$

where we have used the fact that $V(y,t)$ is bounded. In view of these facts, together with (5.6), we obtain that, if $(l,k) \notin \Omega_n$,

$$\left| \mathrm{E}\, Y_{k,n}(z)\, Y_{l,n}(z) \right| = \left| \mathrm{E}\, Y_{k,n}(z)\, \mathrm{E}\big(Y_{l,n}(z) \mid \mathcal{F}_{k,n}\big) \right|$$

$$\le A\,(c_n d_{l,k,n})^{-1}\, \mathrm{E}\,|Y_{k,n}(z)| \le A_1\,(c_n^2 d_{l,k,n} d_{k,0,n})^{-1}, \qquad \textbf{(5.8)}$$

and if $(l,k) \in \Omega_n$,

$$\left| \mathrm{E}\, Y_{k,n}(z)\, Y_{l,n}(z) \right|$$

$$\le A\,(c_n d_{l,k,n})^{-1}\, \mathrm{E}\,|Y_{k,n}(z)| \int_{|y|\ge\sqrt{c_n}} |g(y)|\, dy$$

$$+A\,(c_n\,d_{l,k,n})^{-1}\int_{|y|\le\sqrt{c_n}}|g(y)|\mathrm{E}|Y_{k,n}(z)||V(y,c_n\,x_{k,n})|\,dy$$

$$\le A\left(c_n^2\,d_{l,k,n}\,d_{k,0,n}\right)^{-1}\Big(\int_{|y|\ge\sqrt{c_n}}|g(y)|\,dy+\sup_{|u|\le C(1+z)\epsilon^{1/2}}|h_{l,k,n}(u)$$

$$-h_{l,k,n}(0)|\Big). \tag{5.9}$$

It follows from (5.8) and (5.9) and (2.2)–(2.5) that, with $\eta=\epsilon^{1/2}/C$ given subsequently,

$$|\Lambda_{2n}(\epsilon)| \le \frac{2c_n^2}{n^2}\left(\sum_{l>k,(l,k)\notin\Omega_n}+\sum_{(l,k)\in\Omega_n}\right)\left|\mathrm{E}\left[Y_{k,n}(z)\,Y_{l,n}(z)\right]\right|$$

$$\le \frac{A}{n^2}\sum_{k=(1-\eta)n}^{n}(d_{k,0,n})^{-1}\max_{1\le k\le n-1}\sum_{l=k+1}^{n}(d_{l,k,n})^{-1}$$

$$+\frac{A}{n^2}\max_{0\le k\le n-1}\sum_{l=k+1}^{n}(d_{l,k,n})^{-1}\max_{0\le k\le(1-\eta)n}\sum_{l=k+1}^{k+\eta n}(d_{l,k,n})^{-1}$$

$$+\frac{A}{n^2}\sum_{k=1}^{n}(d_{k,0,n})^{-1}\max_{1\le k\le n-1}\sum_{l=k+1}^{n}(d_{l,k,n})^{-1}$$

$$\times\left[\int_{|y|\ge c_n^{1/2}}|g(y)|\,dy+\sup_{(l,k)\in\Omega_n}\sup_{|u|\le Cz\epsilon}|h_{l,k,n}(u)-h_{l,k,n}(0)|\right]$$

$$\to 0,$$

as $n\to\infty$ first and then $\epsilon\to 0$, as required. The proof of Theorem 2.1 is now complete.

Proof of Corollary 2.1. It suffices to show that, for any $c_n'\to\infty$, $c_n'/n\to 0$, and $r\in[0,1]$,

$$\frac{c_n'}{n}\sum_{k=1}^{[nr]}g\left(c_n'\,x_{k,n}\right)\to_D \tau\,L_{W_{\alpha/2}}(r,0), \tag{5.10}$$

where $x_{k,n}=S_k/d_n$. Note that $d_n^2=\mathrm{E}S_n^2$. It follows from Lemma 5.1 in Taqqu (1975) that $x_{[nt],n}\Rightarrow W_{\alpha/2}(t), 0\le t\le 1$, on $D[0,1]$, where $W_\beta(t)$ is a fractional Brownian motion having a continuous local time $L_{W_\beta}(t,s)$ with regard to both coordinates (t,s) in $[0,\infty)\times R$. Therefore $x_{i,n}$ satisfies Assumption 2.2. We next show that $x_{i,n}$ also satisfies Assumption 2.3 and then (5.10) follows from Theorem 2.1 accordingly.

To check Assumption 2.3, let $\mathcal{F}_{t,n}=\sigma\{\xi_1,\xi_2,\dots,\xi_t\}$ and n_0 be so large that $|\tilde\gamma_{l,k}|\le\lambda\,d_k\,d_{l-k}$ for all $\min\{k,l-k\}\ge n_0$. The choice of n_0 is possible because of the second part of condition (2.8). For any $0\le k<l\le n$, let

$$d_{l,k,n}=\begin{cases} d_{l,k}^*/d_n, & \text{if } \min\{k,l-k\}\ge n_0,\\ d_l/d_n, & \text{otherwise,}\end{cases}$$

where $d^*_{l,k} = \left[d^2_{l-k} - \tilde{\gamma}^2_{l,k}/d^2_k\right]^{1/2}$. Recall that $d_n^2 \sim n^\alpha h(n)$ and note that $d^{-1}_{l,k,n} \le d_n/d_l + \left(1-\lambda^2\right)^{-1/2} d_n/d_{l-k}$. It is readily seen that, as $n \to \infty$,

$$\inf_{(l,k)\in\Omega_n(\eta)} d_{l,k,n} \ge (1-\lambda^2)^{1/2} \inf_{(l,k)\in\Omega_n(\eta)} d_{l-k}/d_n \ge C\,(1-\lambda^2)^{1/2}\,\eta^{\alpha/2},$$

and $d_{l,k,n}$ satisfy (2.2)–(2.4). On the other hand, by noting that $(S_k, S_l - S_k) \sim N(0,\Sigma)$, where $\Sigma = \begin{pmatrix} d_k^2 & \tilde{\gamma}_{l,k} \\ \tilde{\gamma}_{l,k} & d^2_{l-k} \end{pmatrix}$, the conditional distribution of $S_l - S_k$ given S_k is $N(\tilde{\gamma}_{l,k} S_k/d_k^2,\ d^{*2}_{l,k})$. This implies that, conditional on $\mathcal{F}_{t,n}$,

$$(x_{l,n} - x_{k,n})/d_{l,k,n} = (S_l - S_k)/d^*_{l,k} \ \sim\ N\big(\tilde{\gamma}_{l,k}\, S_k/(d_k^2\, d^*_{l,k}),\ \ 1\big),$$

for $\min\{k, l-k\} \ge n_0$, and

$$(x_{l,n} - x_{k,n})/d_{l,k,n} = (S_l - S_k)/d_l \ \sim\ N\big(\tilde{\gamma}_{l,k}\, S_k/(d_k^2\, d_l),\ \ d^{*2}_{l,k}/d_l^2\big),$$

in other cases. Therefore $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a bounded density $h_{j,k,n}(x)$. The $h_{j,k,n}(x)$ satisfy (2.5) because, whenever $\min\{k, l-k\} \ge n_0$,

$$\sup_x \left|h_{j,k,n}(x) - h_{j,k,n}(x+u)\right| \le \frac{1}{\sqrt{2\pi}} \sup_x |e^{-(u+x)^2/2} - e^{-x^2/2}| \le A\,|u|.$$

This proves that Assumption 2.3 holds true for x_{in} and also completes the proof of Corollary 2.1.

Proof of Corollary 2.2. First we prove part (i). We need some preliminaries. Write $\tilde{\psi}_i = \sum_{j=0}^i \psi_j$, $\tilde{S}_n = \sum_{i=0}^n \tilde{\psi}_i \epsilon_i$, and $\Lambda_n^2 = \sum_{i=0}^n (\tilde{\psi}_i)^2$. Also let $f_n(t) = Ee^{it\tilde{S}_n/\Lambda_n}$. Recalling the definitions of ψ_j, simple calculations show that $\tilde{\psi}_i \sim 1/(1-\mu) i^{1-\mu} h(i)$ and $\Lambda_n^2 \sim 1/((1-\mu)^2(3-2\mu))n^{3-2\mu}h^2(n)$. This, together with the facts that $\mathrm{E}\epsilon_0 = 0, \mathrm{E}\epsilon_0^2 = 1$, and $\mathrm{E}\tilde{S}_n^2 = \Lambda_n^2$, implies that $\tilde{S}_n/\Lambda_n \to_D N(0,1)$. Furthermore, we may prove the following results.

(a) for each $n \ge 1$, if not all $\tilde{\psi}_i = 0, 0 \le i \le n$, then $\tilde{S}_n/\Lambda_n$ has a density $h_n(x)$ that is uniformly bounded by a constant K;

(b) as $n \to \infty$, the density function $h_n(x)$ satisfies that

$$\sup_x |h_n(x) - n(x)| \le \frac{1}{2\pi} \int_{-\infty}^{\infty} |f_n(t) - e^{-t^2/2}|\,dt \to 0, \qquad \textbf{(5.11)}$$

where $n(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the density of a standard normal.

In order to prove (a) and (b), we need the following facts:

(I) For n sufficiently large, there exist $0 < c_1 < c_2 < \infty$ such that $c_1\sqrt{n} < \tilde{\psi}_i/\Lambda_n \le c_2\sqrt{n}$ for $n/2 \le i \le n$.

(II) For some $\delta_0 > 0$, there exists a $0 < \eta < 1$ such that

$$|\varphi(t)| = |\mathrm{E}e^{it\epsilon_0}| \le \begin{cases} e^{-t^2/4}, & \text{for } |t| \le \delta_0, \\ \eta, & \text{for } |t| \ge \delta_0. \end{cases}$$

Fact (I) follows immediately from the estimates of $\tilde{\psi}_i$ and Λ_n. Recalling $\mathrm{E}\epsilon_0 = 0, \mathrm{E}\epsilon_0^2 = 1$, and $\int_{-\infty}^{\infty} |\varphi(t)|\,dt < \infty$, fact (II) follows from (5.6) and the proof of Theorem 5.2 in Feller (1971, Chap. 8, p. 489). In view of (I) and (II), $\forall\epsilon > 0$, by choosing $\delta = \delta_0/c_2$ we have

$$\int_{-\infty}^{\infty} |f_n(t)|\,dt \le \left(\int_{|t|\le\delta\sqrt{n}} + \int_{|t|>\delta\sqrt{n}}\right) \prod_{j=[n/2]+1}^{n} |\mathrm{E}e^{it\tilde{\psi}_j\epsilon_j/\Lambda_n}|\,dt$$
$$\le \int e^{-t^2/8}\,dt + C\eta^{n/2-1}\int |\mathrm{E}e^{it\epsilon_0}|\,dt < \infty.$$

This yields result (a) (see, e.g., Lukács, 1970, Thm. 3.2.2).

The left inequality of (5.11) is obvious. In order to prove the convergence in (5.11), we split the integral into $I_{1n} + I_{2n}$, where

$$I_{1n} = \int_{|t|\le A} |f_n(t) - e^{-t^2/2}|\,dt \quad \text{and} \quad I_{2n} = \int_{|t|\ge A} |f_n(t) - e^{-t^2/2}|\,dt.$$

It is clear that $I_{1n} \to 0$ for each $A > 0$ since $\tilde{S}_n/\Lambda_n \to_D N(0,1)$. On the other hand, $\forall\epsilon > 0$, by choosing A sufficiently large such that $\int_{|t|\ge A} e^{-t^2/8}\,dt < \epsilon/2$, we have

$$I_{2n} \le \left(\int_{A<|t|\le\delta\sqrt{n}} + \int_{|t|>\delta\sqrt{n}}\right) \prod_{j=[n/2]+1}^{n} |\mathrm{E}e^{it\tilde{\psi}_j\epsilon_j/\Lambda_n}|\,dt + \int_{|t|\ge A} e^{-t^2/2}\,dt$$
$$\le 2\int_{|t|\ge A} e^{-t^2/8}\,dt + C\eta^{n/2-1} < 2\epsilon,$$

whenever n is sufficiently large since $0 < \eta < 1$. Combining these facts proves the convergence of (5.11) and completes the proof of results (b).

We are now ready to prove part (i). The fact that $d_n^2 = \mathrm{E}S_n^2 \sim c_\mu n^{3-2\mu} h^2(n)$ with $c_\mu = 1/((1-\mu)(3-2\mu))\int_0^\infty x^{-\mu}(x+1)^{-\mu}\,dx$ can be found in Proposition 2.1 of Wang, Ling, and Gulati (2003a). To prove (2.10), it suffices to show that, for any $c_n' \to \infty$, $c_n'/n \to 0$, and $r \in [0,1]$,

$$\frac{c_n'}{n}\sum_{k=1}^{[nr]} g\left(c_n' x_{k,n}\right) \to_D \tau\, L_{W_{3/2-\mu}}(r,0), \tag{5.12}$$

where $x_{k,n} = S_k/d_n$. The result (5.12) may be proved by checking the conditions of Theorem 2.1. Indeed, it follows from Gorodetskiĭ(1977) (also see Wang, Ling, and Gulati, 2003b) that $x_{[nt],n} \Rightarrow W_\beta(t), 0 \le t \le 1$, on $D[0,1]$, where $\beta = (3-2\mu)/2$ and $W_\beta(t)$ is a fractional Brownian motion having a continuous local time $L_{W_\beta}(t,s)$ with regard to (t,s) in $[0,\infty) \times R$. This implies that $x_{i,n}$ satisfies Assumption 2.2.

We next show that $x_{i,n}$ also satisfies Assumption 2.3. To do this, let $\mathcal{F}_{t,n} = \sigma\{\ldots, \epsilon_{t-1}, \epsilon_t\}$ and $d_{l,k,n} = \Lambda_{l-k}/d_n$. Recall that $d_n^2/\Lambda_n^2 \sim (1-\mu)\int_0^1 x^{-\mu}(x+1)^{-\mu}\,dx$. It is readily seen that, as $n \to \infty$,

$$\inf_{(l,k)\in\Omega_n(\eta)} d_{l,k,n} = \inf_{(l,k)\in\Omega_n(\eta)} \Lambda_{l-k}/d_n \geq C\,\eta^{(3-2\mu)/2},$$

for some constant $C > 0$ and $d_{l,k,n}$ satisfy (2.2)–(2.4). On the other hand, by noting that

$$\begin{aligned} S_l &= \sum_{j=1}^{l}\sum_{i=-\infty}^{j} \epsilon_i\,\psi_{j-i} \\ &= \sum_{j=1}^{k}\sum_{i=-\infty}^{j} \epsilon_i\,\psi_{j-i} + \sum_{j=k+1}^{l}\sum_{i=-\infty}^{j} \epsilon_i\,\psi_{j-i} \\ &= S_k + \sum_{j=k+1}^{l}\sum_{i=-\infty}^{k} \epsilon_i\,\psi_{j-i} + \sum_{j=k+1}^{l}\sum_{i=k+1}^{j} \epsilon_i\,\psi_{j-i} \\ &:= S_k + S_{1l} + S_{2l}, \end{aligned}$$

it follows from the independence of ϵ_i, results (a) and (b) given earlier, and the fact that $S_{2l} =_d \tilde{S}_{l-k}$ (where $=_d$ denotes equivalence in distribution) that, conditional on $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n} = (S_{1l} + S_{2l})/\Lambda_{l-k}$ has a density $h_{l-k}(x - S_{1l}/\Lambda_{l-k})$ that is uniformly bounded by a constant K for all $n \geq 1$ and

$$\begin{aligned} &\sup_{(l,k)\in\Omega_n[\delta^{1/\alpha}]}\ \sup_{|u|\leq\delta} \left|h_{l-k}(u - S_{1l}/\Lambda_{l-k}) - h_{l-k}(-S_{1l}/\Lambda_{l-k})\right| \\ &\quad\leq 2 \sup_{(l,k)\in\Omega_n[\delta^{1/\alpha}]}\ \sup_x |h_{l-k}(x) - \frac{1}{\sqrt{2\pi}}e^{-x^2/2}| \\ &\qquad + \frac{1}{\sqrt{2\pi}} \sup_{|u|\leq\delta}\ \sup_x |e^{-(x+u)^2/2} - e^{-x^2/2}| \\ &\quad\to 0, \end{aligned}$$

as $n \to \infty$ first and then $\delta \to 0$, because of (5.11). This proves that Assumption 2.3 holds true for $x_{i,n}$. Combining the preceding facts, the result (5.12) follows from Theorem 2.1. The proof of part (i) is now complete.

As for part (ii), the fact that $d_n^2 \sim \psi^2 n$ is well known. To prove (2.11), it suffices to show that, for any $c_n' \to \infty$, $c_n'/n \to 0$, and $r \in [0,1]$,

$$\frac{c_n'}{n}\sum_{k=1}^{[nr]} g\left(c_n' x_{k,n}\right) \to_D \tau\, L_W(r,0), \tag{5.13}$$

where $x_{k,n} = S_k/d_n$. By noting that $\Lambda_n^2 \sim d_n^2 \sim \psi^2 n$ if $\sum_{k=0}^{\infty}|\psi_k| < \infty$ and $\psi \equiv \sum_{k=0}^{\infty}\psi_k \neq 0$, the result may be proved by a similar argument as in the proof of (5.12) except that the weak convergence in Gorodetskiĭ (1977) is replaced by Hannan (1979). We omit the details. The proof of Corollary 2.2 is now complete.

Proof of Theorem 3.1. We first note that, under a suitable probability space $\{\Omega, \mathcal{F}, P\}$, there exists an equivalent process x_i^* of x_i (i.e., $x_i =_d x_i^*, 1 \le i \le n, n \ge 1$) such that

$$\sup_{0 \le t \le 1} \left| x^*_{[nt],n} - G(t) \right| = o_P(1), \tag{5.14}$$

where $x^*_{i,n} = x_i^*/d_n$, by Assumption 3.5 and the Skorohod–Dudley–Wichura representation theorem. Without loss of generality we assume that x_i satisfies (5.14) (hence $x_{i,n}$ satisfies Assumption 2.2*), and x_t and u_t, $1 \le t \le n$, are defined on the same probability space $\{\Omega, \mathcal{F}, P\}$. If it were not so, it can be easily arranged because the result to be proved in Theorem 3.1 involves only weak convergence.

We first prove (3.7). The consistency result (3.5) will then follow by choosing $a_n = \min\{h^{-\gamma}, (nh/d_n)^{1/2}\}$. To prove (3.7), we split $\hat{f}(x) - f(x)$ as

$$\hat{f}(x) - f(x) = \frac{\sum_{t=1}^n u_t K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)} + \frac{\sum_{t=1}^n \left[f(x_t) - f(x)\right] K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)}. \tag{5.15}$$

Because $x_{i,n} = x_i/d_n$ satisfies Assumptions 2.2* and 2.3, and for any $\lambda \ge 1$, $g(s) = K^\lambda(s)$ satisfies Assumption 2.1 because of Assumption 3.1, it follows from Theorem 2.1 and Remark 2.1 that, for any $\lambda \ge 1$ and h satisfying $h \to 0$ and $nh/d_n \to \infty$,

$$\frac{d_n}{nh} \sum_{t=1}^n K^\lambda \left(\frac{d_n}{h} x_{t,n} - \frac{x}{h} \right) \to_P L_G(1,0) \int_{-\infty}^{\infty} K^\lambda(s)\, ds. \tag{5.16}$$

Note that $P(L_G(1,0) > 0) = 1$. The result (5.16) implies that, for any a_n diverging to infinity as slowly as required, $1/\sum_{t=1}^n K_h(x_t - x) = o_P\left\{a_n\, d_n/n\right\}$. Now, the result (3.7) will follow if we prove

$$\Theta_{1n} := \sum_{t=1}^n u_t K_h(x_t - x) = O_P\left\{\sqrt{n/(d_n h)}\right\}, \tag{5.17}$$

$$\Theta_{2n} := \sum_{t=1}^n \left[f(x_t) - f(x)\right] K_h(x_t - x) = O_P\left(nh^\gamma/d_n\right). \tag{5.18}$$

In fact, by recalling that $x_{t,n}/d_{t,0,n}$ has a density $h_{t,0,n}(x)$ (in the notation of Assumption 2.3(b) as a result of $x_{i,n}$ satisfying Assumption 2.3) that is uniformly bounded by a constant K, it follows from $\mathrm{E}u_t = 0$ and the independence between u_t and x_t that

$$\begin{aligned}
\mathrm{E}\Theta_{1n}^2 &= \sigma^2 h^{-2} \sum_{t=1}^{n} \mathrm{E}K^2\left(\frac{d_n}{h}x_{t,n} - \frac{x}{h}\right) \\
&= \sigma^2 h^{-2} \sum_{t=1}^{n} \int_{-\infty}^{\infty} K^2\left(\frac{d_n\, d_{t,0,n}}{h}y - \frac{x}{h}\right) h_{t,0,n}(y)\,dy \\
&\le \sigma^2 K \int K^2(y)\,dy \frac{1}{n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\frac{n}{d_n h} \le A n/(d_n h),
\end{aligned} \tag{5.19}$$

because $d_{t,0,n}$ satisfies (2.4). This proves (5.17). Similarly, we have

$$\begin{aligned}
\mathrm{E}|\Theta_{2n}| &\le h^{-1}\sum_{t=1}^{n}\mathrm{E}\Big\{\big|f(d_n x_{t,n}) - f(x)\big|\, K\big(\frac{d_n}{h}x_{t,n} - \frac{x}{h}\big)\Big\} \\
&= h^{-1}\sum_{t=1}^{n}\int_{-\infty}^{\infty}\Big\{\big|f(d_n d_{t,0,n}\,y) - f(x)\big|\, K\left(\frac{d_n\, d_{t,0,n}}{h}y - \frac{x}{h}\right)\Big\} h_{t,0,n}(y)\,dy \\
&\le \frac{1}{d_n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}\Big\{\big|f(hy + x) - f(x)\big|\, K\big(y\big)\Big\}\,dy \\
&\le \frac{nh^{\gamma}}{d_n}\frac{1}{n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}K\big(s\big)\, f_1(s,x)\,ds \le A n h^{\gamma}/d_n,
\end{aligned}$$

which implies (5.18). This completes the proof of (3.7).

We next prove (3.6). It follows from (5.15) that

$$\left(h\sum_{t=1}^{n}K_h(x_t - x)\right)^{1/2}(\hat{f}(x) - f(x)) = \sum_{t=1}^{n}u_t Z_{nt} + \left(\frac{d_n h}{n}\right)^{1/2}\Theta_{2n}/\Theta_{3n}, \tag{5.20}$$

where $Z_{nt} = \big(\frac{d_n}{nh}\big)^{1/2} K\big((d_n/h)x_{t,n} - (x/h)\big)/\Theta_{3n}$ with $\Theta_{3n}^2 = \frac{d_n}{nh}\sum_{t=1}^{n} K\big((d_n/h)\, x_{t,n} - (x/h)\big)$ and Θ_{2n} is defined as in (5.18). It is readily seen from (5.16) and (5.18) that $\big(\frac{d_n h}{n}\big)^{1/2}\Theta_{2n}/\Theta_{3n} = o_P(1)$, because $nh^{1+2\gamma}/d_n \to 0$. Therefore, to prove (3.6), it suffices to show that, for any h satisfying $nh^{1+2\gamma}/d_n \to 0$ and $nh/d_n \to \infty$,

$$V_n \equiv \frac{1}{\Lambda_n}\sum_{t=1}^{n}u_t Z_{nt} \to_D N(0,\sigma^2), \tag{5.21}$$

where $\Lambda_n^2 = \Theta_{3n}^{-2}(d_n/nh)\sum_{t=1}^{n}K^2\big((d_n/h)x_{t,n} - (x/h)\big)$, because $\Lambda_n^2 \to_P \int_{-\infty}^{\infty} K^2(s)\,ds$ by (5.16) and Assumption 3.1.

To prove (5.21), first note that, given $\{x_1, x_2, \ldots, x_t\}$, the sequence $(Z_{nt}u_t,\ t = 1, 2, \ldots, n)$ is a martingale difference because x_t is independent of u_t. It then follows from Theorem 3.9 ((3.75) there) in Hall and Heyde (1980) with $\delta = q/2 - 1$ that

$$\sup_x \big|P(V_n \le x\sigma \mid x_1, x_2, \ldots, x_n) - \Phi(x)\big| \le A(\delta)\,\mathcal{L}_n^{1/(1+q)}, \quad \text{a.s.},$$

where $A(\delta)$ is a constant depending only on δ, $q > 2$ by Assumption 3.3, and

$$\mathcal{L}_n = \frac{1}{\sigma^q \Lambda_n^q} \sum_{k=1}^{n} |Z_{nk}|^q \mathrm{E}|u_k|^q + \mathrm{E}\Big| \frac{1}{\sigma^2 \Lambda_n^2} \sum_{k=1}^{n} Z_{nk}^2 \big[\mathrm{E}(u_k^2|\mathcal{F}_{k-1}) - \sigma^2\big]\Big|^{q/2}.$$

Recall from Assumption 3.1 that $K(x)$ is uniformly bounded and by definition of Z_{nt}, we have $\Lambda_n^2 = \sum_{t=1}^n Z_{nt}^2$. Routine calculations show that

$$\mathcal{L}_n \le \frac{A}{\sigma^q \Lambda_n^{q-2}} \left(\frac{d_n}{nh}\right)^{(q-2)/2} + o_P(1) = o_P(1),$$

because $q > 2$, $nh/d_n \to \infty$, and $\Lambda_n^2 \to_P \int_{-\infty}^{\infty} K^2(s)\, ds$ by (5.16). Therefore, we obtain

$$\sup_x \big|P(V_n \le x\sigma) - \Phi(x)\big| \le \mathrm{E}\Big[\sup_x \big|P(V_n \le x\sigma \mid x_1, x_2, \ldots, x_n) - \Phi(x)\big|\Big] \to 0.$$

This proves (5.21) and also completes the proof of Theorem 3.1.

Proof of Theorem 3.2. The idea of this theorem is similar to Park and Phillips (2001). First notice that, under the assumption $(U_n, V_n) \Rightarrow_D (U, V)$, it follows from the so-called Skorohod–Dudley–Wichura representation theorem that there is a common probability space $(\Omega, \mathcal{F}, P)$ supporting (U_n^0, V_n^0) and (U, V) such that

$$(U_n, V_n) =_d \left(U_n^0, V_n^0\right) \quad \text{and} \quad \left(U_n^0, V_n^0\right) \to_{a.s.} (U, V) \tag{5.22}$$

in $D[0,1]^2$ with the uniform topology. Moreover, as in the proof of Lemma 2.1 in Park and Phillips (2001), $\left(U_n^0, V_n^0\right)$ can be chosen such that for each $n \ge 1$

$$U_n^0(k/n) =_d U_n(k/n) \quad \text{and} \quad V_n^0(k/n) =_d V(\tau_{nk}/n), \quad k = 1, 2, \ldots, n, \tag{5.23}$$

where $\tau_{n,[ns]}, 0 \le s \le 1$, are stopping times in $(\Omega, \mathcal{F}, P)$ with $\tau_{n,0} = 0$ satisfying

$$\sup_{0 \le t \le 1} \left| \frac{\tau_{n,[ns]} - s}{n^\delta} \right| \to_{a.s.} 0 \tag{5.24}$$

as $n \to \infty$ for any $\delta > \max(1/2, 2/q)$. These facts, together with (5.20), yield that, under the extended probability space,

$$\left(h \sum_{t=1}^{n} K_h(x_t - x)\right)^{1/2} (\hat{f}(x) - f(x))$$
$$=_d \frac{1}{\Theta_{2n}^*} \sum_{t=1}^{n} Y_{nt} \left[V\left(\frac{\tau_{n,t}}{n}\right) - V\left(\frac{\tau_{n,t-1}}{n}\right)\right] + \frac{\Theta_{1n}^*}{\Theta_{2n}^*}, \tag{5.25}$$

where $Y_{nt} = \left(\frac{d_n}{h}\right)^{1/2} K\left[(d_n/h)\, U_n^0\left(\frac{t}{n}\right) - (x/h)\right]$, $\Theta_{2n}^{*2} = \frac{d_n}{nh} \sum_{t=1}^n K\left[(d_n/h)\, U_n^0\right.$ $\left.\left(\frac{t}{n}\right) - (x/h)\right]$ and

$$\Theta_{1n}^* = \left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} \left[f\left(d_n U_n^0\left(\frac{t}{n}\right)\right) - f(x)\right] K\left[\frac{d_n}{h} U_n^0\left(\frac{t}{n}\right) - \frac{x}{h}\right].$$

Because (5.22) implies that Assumption 2.2* holds true for $U_n^0(t/n)$ with $G(t)$ being a Brownian motion (i.e., $G(t) = U(t)$), it follows from a similar argument to the proofs of (5.16) and (5.18) that, for any $\lambda \geq 1$,

$$\frac{d_n}{nh}\sum_{t=1}^{[nr]} K^\lambda\left[\frac{d_n}{h}U_n^0\left(\frac{t}{n}\right) - \frac{x}{h}\right] \to_P L_U(r,0)\int_{-\infty}^{\infty} K^\lambda(s)\,ds, \tag{5.26}$$

uniformly in $r \in [0,1]$ and $\Theta_{1n}^* = o_P(1)$. We mention that (5.26) also implies, for any $\lambda \geq 1$, uniformly in $r \in [0,1]$,

$$\begin{aligned}\frac{d_n}{h}\int_0^r K^\lambda\left[\frac{d_n}{h}U_n^0(s) - \frac{x}{h}\right] ds &= \frac{d_n}{nh}\sum_{t=1}^{[nr]} K^\lambda\left[\frac{d_n}{h}U_n^0\left(\frac{t}{n}\right) - \frac{x}{h}\right] + \frac{d_n}{nh}K^\lambda\left(-\frac{x}{h}\right)\\ &\quad + K^\lambda\left(\frac{d_n}{h}U_n^0\left(\frac{[nr]}{n}\right) - \frac{x}{h}\right)(nr - [nr])\\ &\to_P L_U(r,0)\int_{-\infty}^{\infty} K^\lambda(s)\,ds,\end{aligned} \tag{5.27}$$

because $K(x)$ is uniformly bounded and $\frac{d_n}{nh} \to 0$.

By virtue of (5.26) and $\Theta_{1n}^* = o_P(1)$, we have $\Theta_{2n}^* \to_P L_U^{1/2}(1,0)$ and $\Theta_{1n}^*/\Theta_{2n}^* \to_P 0$. These facts, together with (5.25) and (5.26), imply that (3.6) will follow if we prove

$$\left\{\sum_{t=1}^n Y_{nt}\left[V\left(\frac{\tau_{n,t}}{n}\right) - V\left(\frac{\tau_{n,t-1}}{n}\right)\right], \sum_{t=1}^n Y_{nt}^2\right\} \to_D (\eta N, \eta^2), \tag{5.28}$$

where $\eta^2 = \sigma^2\int_{-\infty}^{\infty} K^2(s)\,ds\, L_U(1,0)$ and N is a standard normal variable independent of η.

To prove (5.28), write

$$M_n(r) = \sum_{t=1}^{j-1} Y_{nt}\left[V\left(\frac{\tau_{n,t}}{n}\right) - V\left(\frac{\tau_{n,t-1}}{n}\right)\right] + Y_{n,j-1}\left[V\left(\frac{r}{n}\right) - V\left(\frac{\tau_{n,j-1}}{n}\right)\right], \tag{5.29}$$

for $\tau_{n,j-1}/n < r \leq \tau_{n,j}/n$, $j = 1,2,\ldots,k$. Under the conditions of Theorem 3.2, it is readily seen that M_n is a continuous martingale with the quadratic variation process $[M_n]$ given by

$$\begin{aligned}[M_n]_r &= \sum_{k=1}^{j-1} Y_{nt}^2\left(\frac{\tau_{n,k}}{n} - \frac{\tau_{n,k-1}}{n}\right) + Y_{n,j-1}^2\left(r - \frac{\tau_{n,j-1}}{n}\right)\\ &= \frac{d_n\sigma^2}{h}\int_0^r K^2\left[\frac{d_n}{h}U_n^0(s) - \frac{x}{h}\right] ds\,\left[1 + o_P(1)\right]\\ &\to_P L_U(r,0)\,\sigma^2\int_{-\infty}^{\infty} K^2(s)\,ds,\end{aligned} \tag{5.30}$$

uniformly in $r \in [0,1]$, in view of (5.24) and (5.27) with $\lambda = 2$. For the covariance process $[M_n, U]$ of M_n and U, we also have

$$\begin{aligned}[M_n, U]_r &= \sum_{k=1}^{j-1} Y_{nk}\left(\frac{\tau_{n,k}}{n} - \frac{\tau_{n,k-1}}{n}\right)\sigma_{uv} + Y_{n,k-1}\left(r - \frac{\tau_{n,j-1}}{n}\right)\sigma_{uv} \\ &= \sigma_{uv}\left(\frac{d_n}{h}\right)^{1/2}\int_0^r K\left[\frac{d_n}{h}U_n^0(s) - \frac{x}{h}\right]ds\left[1 + o_P(1)\right] \\ &\to_P 0, \end{aligned} \tag{5.31}$$

where $\sigma_{uv} = \mathrm{cov}(V, U)$, because $(h/d_n)^{1/2} \to 0$ and in view of (5.27) with $\lambda = 1$. It follows easily from (5.31) that

$$[M_n, U]_{\rho_n(r)} \to_P 0, \tag{5.32}$$

where $\rho_n(r) = \inf\{s \in [0,1] : [M_n]_s > r\}$ is a sequence of time changes. If we call B^n (i.e., $B^n(r) = M_n\{\rho_n(r)\}$) the DDS (Dambis, Dubins-Schwarz) Brownian motion (see, e.g., Revuz and Yor, 1999, p. 181) of the continuous martingale M_n defined by (5.29), it follows from Theorem 2.3 of Revuz and Yor (1999, p. 524) that B^n converges in distribution to a Wiener process W in view of (5.32). Now, by using (5.30) and noting that $M_n(r)$ is equal to $B^n([M_n]_r)$, it is plain that $M_n(1) \to_D \eta N$, where N is a normal variate independent of η. On the other hand, we have

$$\begin{aligned} &\mathrm{E}\left[\max_{1\le t\le n} |Y_{nt}|\left|V\left(\frac{\tau_{n,t}}{n}\right) - V\left(\frac{\tau_{n,t-1}}{n}\right)\right|\right] \\ &\quad= \left(\frac{d_n}{nh}\right)^{1/2}\mathrm{E}\left[\max_{1\le t\le n} K\left(\frac{x_t}{h} - \frac{x}{h}\right)|u_t|\right] \\ &\quad\le \left(\frac{d_n}{nh}\right)^{1/2}\mathrm{E}\left[\sum_{t=1}^{n} K^q\left(\frac{x_t}{h} - \frac{x}{h}\right)|u_t|^q\right]^{1/q} \\ &\quad\le \left(\frac{d_n}{nh}\right)^{1/2}\left[\max_{1\le t\le n}\mathrm{E}(|u_t|^q|\mathcal{F}_{t-1})\sum_{t=1}^{n}\mathrm{E}\left\{K^q\left(\frac{x_t}{h} - \frac{x}{h}\right)\right\}\right]^{1/q} \\ &\quad\le A\left(\frac{d_n}{nh}\right)^{1/2-1/q} \to 0, \end{aligned} \tag{5.33}$$

because $nh/d_n \to 0$ and $q > 2$, where we have used Assumption 3.3 and calculations similar to those in (5.19) which yield $\sum_{t=1}^{n}\mathrm{E}\big\{K^q\big((x_t/h) - (x/h)\big)\big\} \le Anh/d_n$. These facts, together with (5.30), imply that

$$(M_n(1), [M_n]_1) \to_D (\eta N, \eta^2), \tag{5.34}$$

by Corollary 6.30 of Jacod and Shiryaev (2003, p. 385). Now, the result (5.28) follows from (5.34) and (5.33) by some routine calculations. The proof of Theorem 3.2 is complete.

REFERENCES

Akonom, J. (1993) Comportement asymptotique du temps d'occupation du processus des sommes partielles. *Ann. Inst. H. Poincaré Probab. Statist.* 29, 57–81.

Bandi, F. (2004) On Persistence and Nonparametric Estimation (with an Application to Stock Return Predictability). Manuscript, Graduate School of Business, Chicago.

Berkes, I. & L. Horváth (2006) Convergence of integral functionals of stochastic processes. *Econometric Theory* 22, 304–322.

Borodin, A.N. & I.A. Ibragimov (1995) *Limit Theorems for Functionals of Random Walks.* Proc. Steklov Inst. Math. 2.

de Jong, R. (2004) Addendum to: "Asymptotics for nonlinear. transformations of integrated time series." *Econometric Theory* 20, 627–635.

de Jong, R. & C.-H. Wang (2005) Further results on the asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 21, 413–430.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, vol. II, 2nd ed. Wiley.

Geman, D., & J. Horowitz (1980) Occupation densities. *Annals of Probability* 8, 1–67.

Gorodetskiĭ, V.V. (1977) Convergence to semistable Gaussian processes. *Teor. Verojatnost. i Primenen.* 22, 513–522.

Guerre, E. (2004) Design-Adaptive Pointwise Nonparametric Regression Estimation for Recurrent Markov Time Series. Manuscript, Queen Mary College, London.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic.

Hannan, E.J. (1979) The central limit theorem for time series regression. *Stochastic Process. Appl.* 9, 281–289.

Hu, L. & P.C.B. Phillips (2004) Dynamics of the federal funds target rate: a nonstationary discrete choice approach. *Journal of Applied Econometrics* 19, 851–867.

Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes*, 2nd ed. Springer-Verlag.

Jeganathan, P. (2004) Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *Annals of Probability* 32, 1771–1795.

Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.

Karlsen, H.A., T. Myklebust & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration model. *Annals of Statistics* 35, 252–299.

Lukács, E. (1970) *Characteristic Functions*. Hafner.

Park, J.Y. & P.C.B. Phillips (1999) Asymptotics for nonlinear transformation of integrated time series. *Econometric Theory*, 15, 269–298.

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.

Phillips, P.C.B. (2001) Descriptive econometrics for non-stationary time series with empirical applications. *Journal of Applied Econometrics* 16, 389–413.

Phillips, P.C.B. (2005) Econometric analysis of Fisher's equation. *American Journal of Economics and Sociology* 64, 125–168.

Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion paper 1181.

Pötscher, B.M. (2004) Nonlinear functions and convergence to Brownian motion: Beyond the continuous mapping theorem. *Econometric Theory* 20, 1–22.

Revuz, D. & M. Yor (1999) *Continuous Martingales and Brownian Motion.* Fundamental Principles of Mathematical Sciences 293. Springer-Verlag.

Shorack, G.R. & J.A. Wellner (1986) *Empirical Processes with Applications to Statistics.* Wiley.

Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichskeitstheorie and Verwandte Gebiete* 31, 287–302.

Wang, Q., Y.-X. Lin, & C.M. Gulati (2003a) Strong approximation for long memory processes with applications. *Journal of Theoretical Probability* 16, 377–389.

Wang, Q., Y.-X. Lin, & C.M. Gulati (2003b) Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* 19, 143–164.