A SPECIFICATION TEST FOR NONLINEAR NONSTATIONARY MODELS

BY QIYING WANG\textsuperscript{1} AND PETER C. B. PHILLIPS\textsuperscript{2}

University of Sydney, and Yale University, University of Auckland, University of Southampton, Singapore Management University

We provide a limit theory for a general class of kernel smoothed U-statistics that may be used for specification testing in time series regression with nonstationary data. The test framework allows for linear and nonlinear models with endogenous regressors that have autoregressive unit roots or near unit roots. The limit theory for the specification test depends on the self-intersection local time of a Gaussian process. A new weak convergence result is developed for certain partial sums of functions involving nonstationary time series that converges to the intersection local time process. This result is of independent interest and is useful in other applications. Simulations examine the finite sample performance of the test.

1. Introduction. One of the advantages of nonparametric modeling is the opportunity for specification testing of particular parametric models against general alternatives. The past three decades have witnessed many developments in such specification tests involving nonparametric and semiparametric techniques that allow for independent, short memory and long-range dependent data. Recent research on the nonparametric modeling of nonstationary data opens up some new possibilities that seem relevant to applications in many fields, including nonlinear diffusion models in continuous time [Bandi and Phillips (2003, 2007)] and cointegration models in economics and finance.

Cointegration models were originally developed in a linear parametric framework that has been widely used in econometric applications. That
framework was extended in Park and Phillips (1999, 2001) to allow for nonlinear parametric formulations under certain restrictions on the function nonlinearity. While considerably broadening the class of allowable nonstationary models, the potential for parametric misspecification in these models is still present and is important to test in applied work.

The hypothesis of linear cointegration is of particular interest in this context, given the vast empirical literature. Recent papers by Karlsen, Myklebust and Tjøstheim (2007), Wang and Phillips (2009a, 2009b, 2011) and Schienle (2008) have developed asymptotic theory for nonparametric kernel regression of nonlinear nonstationary systems. This work facilitates the comparison of various parametric specifications against a more general nonparametric nonlinear alternative. Such comparisons may be based on weighted sums of squared differences between the parametric and nonparametric estimates of the system or on a kernel-based U-statistic test which uses a smoothed version of the parametric estimator in its construction [e.g., Gao (2007), Chapter 3].

A major obstacle in the development of such specification tests is the technical difficulty of developing a limit theory for these weighted sums which typically involve kernel functions with multiple nonstationary regressor arguments. Few results are currently available, and because of this shortage, attempts to develop specification tests for nonlinear regression models with nonstationarity have been highly specific and do not involve nonparametric alternatives or kernel methods. Some examples of recent work in parametric models include Choi and Saikonnen (2004, 2010), Marmer (2008), Hong and Phillips (2010) and Kasparis and Phillips (2012). An exception is the recent work for testing linearity in autoregression and parametric time series regression by Gao et al. (2009a, 2009b) who obtained a limit distribution theory for a kernel based specification test in a setting that involves martingale difference errors and random walk regressors.

The present paper makes a related contribution and seeks to provide a general theory of specification tests that is applicable for a wider class of nonstationary regressors that includes both unit root and near unit root processes. The latter are important in practical work where a unit root restriction is deemed too restrictive. The paper contributes to this emerging literature in two ways. First, we provide a limit theory for a general class of kernel-based specification tests of parametric nonlinear regression models that allows for near unit root processes driven by short memory (linear process) errors. This limit theory should be widely applicable to specification testing in nonlinear cointegrated systems.

Second, the limit theory of the specification test involves the self-intersection local time of a Gaussian limit process. The result requires establishing weak convergence to this self-intersection local time process, which is of independent interest, and a feasible central limit theorem involving an empirical
estimator of the intersection local time that can be used to construct the test statistic. Thus, the results provide some new theories for intersection local time, weak convergence and specification test asymptotics that are relevant in applications.

The paper is organized as follows. Section 2 lays out the nonparametric and parametric models and assumptions. Section 3 gives the main results on specification test limit theory. Section 4 reports some simulation evidence on test performance. Section 5 gives the weak convergence theory for intersection local time. Section 6 gives proofs of the main theorems in Section 3.

The proofs of the local time limit theory in Section 5 and some supplemental technical results in Section 6 can be found in the supplementary material [Wang and Phillips (2012)].

2. Model and assumptions. We consider the nonlinear cointegrating regression model

\[ y_{t+1} = f(x_t) + u_{t+1}, \quad t = 1, 2, \ldots, n, \]

(2.1)

where \( u_t \) is a stationary error process, and \( x_t \) is a nonstationary regressor. We are interested in testing the null hypothesis

\[ H_0: f(x) = f(x, \theta), \quad \theta \in \Omega_0, \]

for \( x \in R \), where \( f(x, \theta) \) is a given real function indexed by a vector \( \theta \) of unknown parameters which lie in the parameter space \( \Omega_0 \).

To test \( H_0 \) we make use of the following kernel-smoothed test statistic:

\[ S_n = \sum_{s,t=1, s \neq t}^{n} \hat{u}_{t+1} \hat{u}_{s+1} K[(x_t - x_s)/h], \]

(2.2)

involving the parametric regression residuals \( \hat{u}_{t+1} = y_{t+1} - f(x_t, \hat{\theta}) \), where \( K(x) \) is a nonnegative real kernel function, \( h \) is a bandwidth satisfying \( h \equiv h_n \to 0 \) as the sample size \( n \to \infty \) and \( \hat{\theta} \) is a parametric estimator of \( \theta \) under the null \( H_0 \), that is consistent whenever \( \theta \in \Omega_0 \).

The statistic \( S_n \) in (2.2) has commonly been applied to test parametric specifications in stationary time series regression [see Gao (2007)] and was used by Gao et al. (2009a, 2009b) to test for linearity in autoregression and a parametric conditional mean function in time series regression involving a random walk regressor. \( S_n \) is a weighted U-statistic with kernel weights that depend on standardized differentials \( (x_t - x_s)/h \) of the regressor. The weights focus attention in the statistic on those components in the sum where the nonstationary regressor \( x_t \) nearly intersects itself. This smoothing scheme gives prominence to product components \( \hat{u}_{t+1} \hat{u}_{s+1} \) in the sum where \( s \) and \( t \) may differ considerably but for which the corresponding regressor process takes similar values (i.e., \( x_t, x_s \approx x \) for some \( x \)), thereby enabling a test of \( H_0 \).
The difficulty in the development of an asymptotic theory for $S_n$ stems from the presence of the kernel weights $K((x_t - x_s)/h)$. The behavior of these weights depends on the self intersection properties of $x_t$ in the sample, and, as $n \to \infty$, this translates into the corresponding properties of the stochastic process to which a standardized version of $x_t$ converges. To establish asymptotics for $S_n$, we need to account for this limit behavior, which leads to a new limit theory involving the self-intersection local time of a Gaussian process (i.e., the local time for which the process intersects itself).

We use the following assumptions in our development.

**Assumption 1.** (i) $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) continuous random variables with $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$, and with the characteristic function $\varphi(t)$ of $\epsilon_0$ satisfying $|t||\varphi(t)| \to 0$, as $|t| \to \infty$.

(ii) $x_t = \rho x_{t-1} + \eta_t$, $x_0 = 0$, $\rho = 1 + \kappa/n$, $1 \leq t \leq n$,

where $\kappa$ is a constant and $\eta_t = \sum_{k=0}^{\infty} \phi_k \epsilon_{t-k}$ with $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$ and $\sum_{k=0}^{\infty} k^{1+\delta} |\phi_k| < \infty$ for some $\delta > 0$.

**Assumption 2.** (i) $\{u_t, F_t\}_{t \geq 1}$, where $F_t$ is a sequence of increasing $\sigma$-fields which is independent of $\epsilon_k, k \geq t + 1$, forms a martingale difference satisfying $E(u_{t+1}^2 | F_t) \to a.s., \sigma^2 > 0$ as $t \to \infty$ and $\sup_{t \geq 1} E(|u_{t+1}|^4 | F_t) < \infty$.

(ii) $x_t$ is adapted to $F_t$, and there exists a correlated vector Brownian motion $(W, V)$ such that

\[ (2.4) \quad \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_j, \frac{1}{\sqrt{n\sigma}} \sum_{j=1}^{[nt]} u_{j+1} \right) \Rightarrow_D (W(t), V(t)) \]

on $D[0,1]^2$ as $n \to \infty$.

**Assumption 3.** $K(x)$ is a nonnegative real function satisfying $\sup_x K(x) < \infty$ and $\int K(x) dx < \infty$.

**Assumption 4.** (i) There is a sequence of positive real numbers $\delta_n$ satisfying $\delta_n \to 0$ as $n \to \infty$ such that $\sup_{\theta \in \Theta_0} ||\hat{\theta} - \theta|| = o_P(\delta_n)$, where $|| \cdot ||$ denotes the Euclidean norm.

(ii) There exists some $\varepsilon_0 > 0$ such that $\frac{\partial^2 f(x,t)}{\partial \theta^2}$ is continuous in both $x \in R$ and $t \in \Theta_0$, where $\Theta_0 = \{t : ||t - \theta|| \leq \varepsilon_0, \theta \in \Omega_0\}$.

(iii) Uniformly for $\theta \in \Omega_0$,

\[ \left| \frac{\partial f(x,t)}{\partial \theta} \right|_{t=\theta} + \left| \frac{\partial^2 f(x,t)}{\partial \theta^2} \right|_{t=\theta} \leq C(1 + |x|^\beta) \]

for some constants $\beta \geq 0$ and $C > 0$. 

(iv) Uniformly for \( \theta \in \Omega_0 \), there exist \( 0 < \gamma' \leq 1 \) and \( \max\{0, 3/4 - 2\beta\} < \gamma \leq 1 \) such that

\[ |g(x + y, \theta) - g(x, \theta)| \leq C|y|^\gamma \left\{ \begin{array}{ll}
1 + |x|^{\beta - 1} + |y|^\beta, & \text{if } \beta > 0, \\
1 + |x|^\gamma - 1, & \text{if } \beta = 0,
\end{array} \right. \]

for any \( x, y \in R \), where \( g(x, t) = \frac{\partial f(x, t)}{\partial t} \).

**Assumption 5.** \( nh^2 \to \infty \), \( \delta_n^2 n^{1+\beta} \sqrt{h} \to 0 \) and \( nh^4 \log^2 n \to 0 \), where \( \beta \) and \( \delta_n^2 \) are defined as in Assumption 4. Also, \( \int (1 + |x|^{2\beta + 1}) K(x) \, dx < \infty \) and \( E|\epsilon_0|^{4\beta + 2} < \infty \).

Assumption 1 allows for both a unit root (\( \kappa = 0 \)) and a near unit root (\( \kappa \neq 0 \)) regressor by virtue of the localizing coefficient \( \kappa \) and is standard in the near integrated regression framework [Phillips (1987, 1988), Chan and Wei (1987)]. Compared to the estimation theory developed in Wang and Phillips (2009a, 2009b) and for technical convenience in the present work, we impose the stronger summability condition \( \sum_{k=0}^{\infty} k^{1+\delta} |\phi_k| < \infty \) for some \( \delta > 0 \) on the coefficients of the linear process \( \eta_t = \sum_{k=0}^{\infty} \phi_k \epsilon_{t-k} \) driving the regressor \( x_t \). Under these conditions, it is well known that the standardized process \( x_{[nt]} = x_{[nt]} \sqrt{n} \) converges weakly to the Gaussian process \( G(t) = \int_0^t e^{\kappa(t-s)} \, dW(s) \), where \( W(t) \) is a standard Brownian motion. See (5.2) below or Phillips and Solo (1992).

Assumption 2(i) is a standard martingale difference condition on the equation innovations \( u_t \), so that \( \text{cov}(u_{t+1}, x_t) = E[x_t E(u_{t+1} \mid \mathcal{F}_t)] = 0 \). Wang and Phillips (2009b) allowed for endogeneity in their nonparametric structure, so the equation error could be serially dependent and cross-correlated with \( x_t \) for \( |t - s| \leq m_0 \) for some finite \( m_0 \). It is not clear at the moment if the results of the present paper on testing extend to the more general error structure considered in Wang and Phillips (2009b), but simulation results suggest that this may be so. Assumption 2(ii) is a standard functional law for partial sum processes [e.g., Park and Phillips (2001)].

Assumption 3 is a standard condition on \( K(x) \) as in the stationary situation. The integrability condition is weaker than the common alternative requirement that \( K(x) \) has compact support.

As seen in Assumption 5, the sequence \( \delta_n \) in Assumption 4(i) may be chosen as \( \delta_n^2 = n^{-(1+\beta)/2} h^{-1/8} \). As \( h \to 0 \) and \( \kappa = 0 \) in (2.3), Assumption 4(i) holds under very general conditions, such as those of Theorem 5.2 in Park and Phillips (2001). Indeed, by Park and Phillips (2001), we may choose \( \hat{\theta} \) such that \( \sup_{\theta \in \Theta_0} \| \hat{\theta} - \theta \| = O_P(n^{-1+\beta/2}) \), under our Assumption 4(ii)–(iv). Assumption 4(ii)–(iv) is quite weak and includes a wide class of functions. Typical examples include polynomial forms like \( f(x, \theta) = \theta_1 + \theta_2 x + \cdots + \theta_k x^{k-1} \), where \( \theta = (\theta_1, \ldots, \theta_k) \), power functions like \( f(x, a, b, c) = a + bx^c \),
shift functions like $f(x, \theta) = x(1 + \theta x)I(x \geq 0)$ and weighted exponentials such as $f(x, a, b) = (a + be^x)/(1 + e^x)$. However, Assumption 4 excludes models where $f(x, \theta)$ is integrable, because parametric rates of convergence are known to be $O(n^{1/4})$ in this case [see Park and Phillips (2001)]. It seems that cases with integrable $f(x, \theta)$ require different techniques and these are left for future investigation.

As in estimation limit theory, the condition in Assumption 5 that the bandwidth $h$ satisfies $nh^2 \to \infty$ is necessary. The further condition that $nh^4 \log^2 n \to 0$ restricts the choice of $h$ and, at least with the techniques used here, seems difficult to relax in the general case studied in the present work, although it may be substantially relaxed in less general models as discussed later in the paper. The condition that $\delta_n^2 n^{1+\beta} \sqrt{h} \to 0$ holds automatically if $\sup_{\theta \in \Omega_0} \| \theta - \theta \| = O_P(n^{-(1+\beta)/2})$. As explained above, the latter condition holds true under very general settings such as Assumption 4(ii)–(iv). We also impose a higher moment condition on the innovation $\epsilon_0$ in Assumption 5 which helps in the development of the limit theory.

3. Main results on specification. The limit distribution of $S_n$ under standardization involves nuisance parameters $\sigma$ and $\phi$, which are the limit of $E\epsilon^2_t$ as $t \to \infty$ and the sum of coefficients of the linear process appearing in Assumption 1; see Corollary 3.1 below. While convenient, this formulation obviously restricts direct use of the result in applications. The dependence on the nuisance parameters can be simply removed by self-normalization. Indeed, by defining

$$V_n^2 = \sum_{s,t=1, s \neq t}^n \hat{u}_{t+1}^2 \hat{u}_{s+1}^2 K^2[(x_t - x_s)/h],$$

we have the following main result.

**Theorem 3.1.** Under Assumptions 1–5 and the null hypothesis, we have

$$\frac{S_n}{\sqrt{2V_n}} \to_D N,$$

where $N$ is a standard normal variate.

The limit in Theorem 3.1 is normal and does not depend on any nuisance parameters. As a test statistic, $Z_n = S_n/\sqrt{2V_n}$ has a big advantage in applications. In order to investigate the asymptotic power of the test, we consider the local alternative models

$$H_1: f(x) = f(x, \theta) + \rho_n m(x),$$

where $\theta \in \Omega_0$, $\rho_n$ is a sequence of constants, and $m(x)$ is a real function. This kind of local alternative model is commonly used in the theory of non-parametric inference involving stationary data; see, for instance, Horowitz and Spokoiny (2001).
Assumption 6. There exists a $\nu \geq 0$ such that

$$0 < \inf_{|x| \geq 1} \frac{|m(x)|}{|x|^\nu} \leq C \sup_{x \in \mathbb{R}} \frac{|m(x)|}{1 + |x|^\nu} < \infty,$$

and there exist $0 < \gamma' \leq 1$ and $\max\{0, 3/4 - 2\nu\} < \gamma \leq 1$ such that

$$|m(x + y) - m(x)| \leq C |y|^\gamma \left\{ \begin{array}{ll} 1 + |x|^{\nu - 1} + |y|^{\nu}, & \text{if } \nu > 0, \\ 1 + |x|^{-1} |y|^{\gamma'}, & \text{if } \nu = 0, \end{array} \right.$$  

for any $x, y \in \mathbb{R}$ and for some constant $C > 0$.

Assumption 6 is quite weak which is satisfied by a large class of real functions such as $m(x) = a_1 + a_2 x + \cdots + a_k x^{k-1}$, $m(x) = a + bx^c$ and $m(x) = (a + be^x)/(1 + e^x)$. If $m(x)$ is positive (or negative) on $\mathbb{R}$, condition (3.3) is not necessary.

Theorem 3.2. In addition to Assumptions 1–6, $\int (1 + |x|^{2\nu + 2}) K(x) \, dx < \infty$ and $E|\xi_0|^{4\nu + 2} < \infty$. Then, under $H_1$, we have

$$\lim_{n \to \infty} P\left( \frac{S_n}{\sqrt{2V_n}} \geq t_\alpha \right) = 1$$

for any $\rho_n$ satisfying $\rho_n^2 n^{1/2 + \nu} h^{1/2} \to \infty$, and for any $0 < \alpha < 1$, where $\Phi(t_\alpha) = 1 - \alpha$ and $\Phi$ is the standard normal c.d.f.

Theorem 3.2 shows that our test has nontrivial power against the local alternative whenever $\rho_n \to 0$ at a rate that is slower than $n^{-1/8 - \nu/2}$, as $nh^2 \to \infty$. This is different from the stationary situation where in general a test has a nontrivial power if only $\rho_n \to 0$ at a rate that is slower than $n^{-1/2}$. It is interesting to notice that the rate is related to the magnitude of $m(x)$ and the bandwidth $h$. The test has stronger discriminatory power the larger the value of $\nu$. The reason is that the nonlinear shape characteristics in $m(x)$ are magnified over a wide domain and this property is exploited by the test because the nonstationary regressor is recurrent.

Theorem 3.2 seems to be new to the literature. Under very strict restrictions (namely that $x_t$ is a random walk and $x_t$ is independent of $u_t$), the result in Theorem 3.1 has been considered in Gao et al. (2009a). Not only the generalization of our result, but the techniques used in this paper are quite different from Gao et al. (2009a, 2009b). To outline the essentials of the argument in the proof of Theorem 3.1, under the null hypothesis, we split $S_n$ as

$$S_n = 2 \sum_{t=2}^n u_{t+1} Y_{nt} + 2 \sum_{i,t=1}^n u_{i+1} [f(x_t, \theta) - f(x_t, \hat{\theta})] K[(x_t - x_i)/h]$$
\begin{align*}
&\sum_{i,t=1}^{\infty} \left[ f(x_i, \theta) - f(x_i, \hat{\theta}) \right] \left[ f(x_t, \theta) - f(x_t, \hat{\theta}) \right] K[(x_t - x_i)/h] \\
&= 2S_{1n} + 2S_{2n} + S_{3n} \quad \text{say},
\end{align*}

where \( Y_{nt} = \sum_{i=1}^{t-1} u_{i+1} K[(x_t - x_i)/h] \). It will be proved in Section 6.1 that terms \( S_{2n} \) and \( S_{3n} \) are negligible in comparison with \( S_{1n} \). Furthermore it will be proved that, under the null hypothesis,

\begin{align*}
V_n^2 &= \sigma^4 \sum_{t,s=1}^{\infty} K^2[(x_t - x_s)/h] + o_P(n^{3/2}h) \\
&= 2\sigma^2 \sum_{i=2}^{n} Y_{nt}^2 + o_P(n^{3/2}h).
\end{align*}

By virtue of these facts, Theorem 3.1 follows from the following theorem, giving a joint convergence result for \( S_{1n} \) and its conditional variance \( \sum_{t=2}^{\infty} Y_{nt}^2 \). This result, along with the following Corollary 3.1, is of some independent interest.

**Theorem 3.3.** Under Assumptions 1–3, \( nh^2 \to \infty \) and \( nh^4 \log^2 n \to 0 \), we have

\begin{equation}
\left( \frac{1}{\sigma d_n} \sum_{i=2}^{n} u_{i+1} Y_{nt}, \frac{1}{d_n} \sum_{i=2}^{n} Y_{nt}^2 \right) \to_D (\eta N, \eta^2),
\end{equation}

where \( d_n^2 = (2\phi)^{-1}\sigma^2 n^{3/2}h \int_{-\infty}^{\infty} K^2(x) dx \), \( \eta \) is the self intersection local time generated by the process \( G = \int_0^t e^{\kappa(t-s)} dW(s) \), and \( N \) is a standard normal variate which is independent of \( \eta^2 \).

**Corollary 3.1.** Under Assumptions 1–5, we have

\[ \frac{S_n}{\tau_n} \to_D \eta N, \]

where \( \tau_n^2 = (8\phi)^{-1}\sigma^4 n^{3/2}h \int_{-\infty}^{\infty} K^2(x) dx \), \( \eta^2 \) and \( N \) are defined as in Theorem 3.3.

Here and below, we define

\begin{equation}
L_G(t, u) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \int_0^t 1[(G(x) - G(y)) - u] < \varepsilon \] dx dy

= \int_0^t \int_0^t \delta_u [G(x) - G(y)] dx dy,
\end{equation}
where $\delta_u$ is the dirac function. $L_G(t, u)$ characterizes the amount of time over the interval $[0, t]$ that the process $G(t)$ spends at a distance $u$ from itself, and is well defined, as shown in Section 5. When $u = 0$, $L_G(t, 0)$ describes the self-intersection time of the process $G(t)$. Using the definition of the dirac function, the extended occupation times formula [e.g., Revuz and Yor (1999), page 232], and integration by parts with the local time measure, we may write

$$L_G(t, 0) = 2 \int_0^t \int_0^y \delta_0[G(x) - G(y)] \, dx \, dy$$

$$= 2 \int_0^t \ell_G(s, G(s)) \, ds$$

$$= 2 \int_{-\infty}^\infty \int_0^t \ell_G(s, a) \, d\ell_G(s, a) \, da$$

$$= \int_{-\infty}^\infty \ell_G(t, a)^2 \, da,$$

where $\ell_G(s, a)$ is the local time spent by the process $G$ at $a$ over the time interval $[0, t]$, namely,

$$\ell_G(t, a) = \int_0^t \delta_a[G(s)] \, ds = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1[|G(s) - a| < \varepsilon] \, ds.$$

The process $\ell_G(s, G(s))$ is the local time that the process $G$ has spent at its current position $G(s)$ over the time interval $[0, s]$. It appears in the limit theory for nonparametric nonstationary spurious regression [Phillips (2009)]. Aldous (1986) gave (3.9) for the case of Brownian motion.

It is interesting to note that $S_{1n}$ is a martingale sequence with conditional variance $\sum_{t=2}^n Y_{nt}^2$, suggesting that some version of the martingale central limit theorem [e.g., Hall and Heyde (1980), Chapter 3] may be applicable. However, the problem is complicated by the U-statistic structure and the weak convergence of the conditional variance, and use of existing limit theory seems difficult. To investigate the asymptotics of $S_{1n}$, we therefore develop our own approach. As part of this development, in Section 5, we provide a general weak convergence theory to intersection local time, which is of independent interest and useful in other applications. The conditions required for this development are weaker than those in establishing Theorem 3.3 and that section may be read separately.

We finally remark that the restrictive condition on the bandwidth $h$ in Theorems 3.1–3.3 (i.e., $nh^4 \log^2 n \to 0$) is mainly used to offset the impact of the error terms in (3.5) and (3.6). It seems difficult to relax this condition.
under the prevailing Assumption 2, which allows for endogeneity in the regressor $x_t$. See, for instance, the proof of Proposition 6.4 given in the supplementary material [Wang and Phillips (2012)]. The restriction $nh^4 \log^2 n \to 0$ on $h$ in Theorems 3.1–3.3, however, can be reduced to the minimal requirement $h \to 0$, if Assumption 2 is replaced by the following Assumption 2*.

**Assumption 2*. For each $n \geq 1$, \( \{u_t, \mathcal{F}_{t,n}\}_{1 \leq t \leq n} \) forms a martingale difference satisfying $\lim_{t \to \infty} \sup_{n \geq t} |E(u_{t+1}^2 | \mathcal{F}_{t,n}) - \sigma^2| = 0$, a.s. and

$$\sup_{n \geq t \geq 1} E(|u_{t+1}|^4 | \mathcal{F}_{t,n}) < \infty,$$

where

$$\mathcal{F}_{t,n} = \sigma(u_1, \ldots, u_t; x_1, \ldots, x_n), \quad t = 1, 2, \ldots, n; n \geq 1.$$

Note that Assumption 2* holds true if $x_t$ is independent of $u_t$, and $\{u_t, \mathcal{F}_{t,n}\}_{t \geq 1}$ forms a martingale difference satisfying $E(u_{t+1}^2 | \mathcal{F}_t) \to a.s. \sigma^2 > 0$ as $t \to \infty$ and $\sup_{n \geq t \geq 1} E(|u_{t+1}|^4 | \mathcal{F}_t) < \infty$, where $\mathcal{F}_t$ is a sequence of increasing $\sigma$-fields. The independence assumption was used in Gao et al. (2009a) to establish a similar version of Theorem 3.1.

4. Simulations. Simulations were conducted to evaluate the finite sample performance of the statistic $Z_n = S_n / \sqrt{2V_n}$ under the null and some local alternatives under various assumptions about the generating mechanism. The results are summarized here, and more detailed findings are reported in the supplementary material [Wang and Phillips (2012)]. The model followed (2.1) with $y_{t+1} = f(x_t) + u_{t+1}, x_t = x_{t-1} + \eta_t$, $x_0 = 0$, and $\eta_t$ generated by an AR(1) process $\eta = \lambda \eta_{t-1} + \epsilon_t$ or an MA(1) process $\eta = \epsilon_t + \lambda \epsilon_{t-1}$ with $(u_t, \epsilon_t) \sim i.i.d. \ N(0, (1, \lambda \lambda')^T)$. A linear null hypothesis $H_0 : f(x) = \theta_0 + \theta_1 x$ was used together with polynomial local alternatives $H_1 : f(x) = \theta_0 + \theta_1 x + \rho_n x^p$, with $\rho_n = 1/(n^{1/4} + \rho^{1/3} h^{1/4})$. The parameter settings were $\theta_0 = 0, \theta_1 = 1$, $\nu \in \{0.5, 1.5, 2, 3\}$ and $r \in \{0, \pm 0.5, \pm 0.75\}$. Results are reported for sample sizes $n \in \{100, 200, 500\}$ and bandwidth settings $h = n^{-p}$ for $p \in \{1/4, 1/3, 1/2\}$. Note that $h = n^{-1/4}$ satisfies Assumption 2* but not Assumption 2. The number of replications was 5000.

Table 1 shows the actual size of the test for various $n$ and bandwidth choices $h$ and for both exogenous ($r = 0$) and endogenous ($r = \pm 0.5$) regressor cases with serially uncorrelated errors ($\lambda = 0$). Table 2 shows the corresponding results for AR errors with $\lambda = \pm 0.4$. Size results for MA errors are similar and are given in the supplementary material [Wang and Phillips (2012)]. Under i.i.d. errors the test is somewhat undersized for $n = 100, 200$ but is close to the nominal for $n = 500$ and for all bandwidth choices. There is some mild oversizing under serially dependent $\eta$ when $\lambda = -0.4$ for bandwidth $h = n^{-1/4}$, but size seems satisfactory for $\lambda = 0.4$ and for the smaller
Table 1
Size: η_t = ε_t

|    | Nominal size 5% | Nominal size 1% |
|----|----------------|----------------|
| 100 | h = n^{-1/4} | n^{-1/3} | n^{-1/2.5} |
|    | r = 0         | r = 0.5 | r = -0.5 |
| 100 | 0.028         | 0.030  | 0.031  |
| 200 | 0.034         | 0.038  | 0.036  |
| 500 | 0.044         | 0.041  | 0.046  |

bandwidths h = n^{-1/3}, n^{-1/2.5}. Since negative λ reduces the long run moving average coefficient φ [φ = 1/(1 − λ) for AR η_t] these results suggest that the strength of the long run signal in x_t (measured by the long-run variance of η_t) affects the performance of the test. On the other hand, endogeneity at the correlation level r = ±0.5 appears to have little effect on performance, which mirrors results for estimation in the nonlinear nonstationary case [Wang and Phillips (2009b)]. Higher levels of correlation (r = ±0.75) produce some size distortion when there is serial dependence, but not when the errors are independent; see Table 3.

Table 4–6 show test power against the local alternative H_1 for polynomial alternatives (cubic ν = 3, quadratic ν = 2 and three halves ν = 1.5). Results for the case ν = 0.5 are given in the supplementary material [Wang and Phillips (2012)]. Again, there is little difference between the exogenous and endogenous cases, so only the endogenous case is reported here. As may be expected, there is greater local discriminatory power for cubic (ν = 3) than quadratic (ν = 2) or three halves (ν = 1.5) alternatives. For n = 100 (500) power is greater than 69% (90%) for a nominal 1% test and greater than 74% (92%) for a nominal 5% test when ν = 3 under AR errors with λ = 0.4 (Table 4). The corresponding results when ν = 2 and n = 100 (500) are 15% (38%) for a nominal 1% test and 23% (46%) for a nominal 5% test (Table 5). Serial dependence affects power, which is higher for λ = 0.4 than for λ = −0.4 in all cases. So lower long-run signal strength in the regressor tends to reduce discriminatory power. For ν = 1.5 and λ = −0.4, power is low even for n = 500 (2% for a 1% test and 7% for a 5% test, Table 6). Low
Table 2

Size: \( \eta_t = \lambda \eta_{t-1} + \varepsilon_t, \ r = \pm 0.5 \)

| Nominal size 5% |               |               |               |               |               |
|-----------------|---------------|---------------|---------------|---------------|
| \( n \)         | \( h = n^{-1/4} \) | \( n^{-1/3} \) | \( n^{-1/2.5} \) | \( n^{-1/4} \) | \( n^{-1/3} \) | \( n^{-1/2.5} \) |
| 100             | 0.034         | 0.038         | 0.041         | 0.002         | 0.004         | 0.005         |
| 200             | 0.044         | 0.044         | 0.047         | 0.007         | 0.008         | 0.009         |
| 500             | 0.058         | 0.058         | 0.057         | 0.026         | 0.029         | 0.031         |

| r = 0.5, \( \lambda = 0.4 \) |               |               |               |               |               |
|-----------------|---------------|---------------|---------------|---------------|
| \( n = 100 \)  | 0.038         | 0.042         | 0.046         |               |               |
| 200             | 0.051         | 0.051         | 0.051         |               |               |
| 500             | 0.070         | 0.061         | 0.057         |               |               |

| Nominal size 1% |               |               |               |               |               |
|-----------------|---------------|---------------|---------------|---------------|
| \( n \)         | \( h = n^{-1/4} \) | \( n^{-1/3} \) | \( n^{-1/2.5} \) | \( n^{-1/4} \) | \( n^{-1/3} \) | \( n^{-1/2.5} \) |
| 100             | 0.035         | 0.040         | 0.043         | 0.012         | 0.012         | 0.012         |
| 200             | 0.050         | 0.049         | 0.050         | 0.018         | 0.015         | 0.013         |
| 500             | 0.073         | 0.064         | 0.056         | 0.026         | 0.018         | 0.016         |

power also occurs against the local alternative with \( \nu = 0.5 \) [see Wang and Phillips (2012)], which also reduces signal strength in the regressor function. Thus, discriminatory power is dependent on the specific alternative and, as asymptotic theory suggests, is sensitive to the magnitude rate (\( \nu \)) of \( m(x) \) as \( |x| \to \infty \).

Overall, the finite sample results reflect the asymptotic theory and seem reasonable for practical use in testing when there is some endogeneity in nonparametric nonstationary regression, especially if smaller bandwidth choices than usual are employed. In cases of serial dependence when the long-run signal strength in the regressor \( x_t \) is reduced, finite sample adjustments for the test critical values may be useful in correcting size, as has been found for i.i.d. and stationary regressors [Li and Wang (1998)].

In practice, the exact \( \alpha \)-level critical value \( \ell_\alpha(h) \) (0 < \( \alpha \) < 1) of the finite sample distribution of \( S_n/\sqrt{2V_n} \) depends on all the unknown parameters and functions in the model. The development of a rigorous theory of approximation for \( \ell_\alpha(h) \) and the choice of an optimal bandwidth for use in testing are challenging problems in the nonstationary setting. Gao et al. (2009a) provided an approximate value of \( \ell_\alpha(h) \) by using the bootstrap and considered numerical solutions for a bandwidth \( h \) that optimizes the power function, both under the assumption that \( x_t \) and \( u_t \) are independent. It is
not clear at the moment whether similar techniques can be rigorously justified in the current general model and there is presently no optimal approach to bandwidth selection. The investigation of such finite sample adjustments and selection criteria is therefore left for later research. Earlier analysis of the restrictions on the bandwidth in Theorems 3.1–3.3, in conjunction with the simulation evidence, indicates that smaller bandwidths than usual for stationary regression are likely to be more reliable in practical work for specification testing of nonlinear nonstationary regression.

5. Convergence to intersection local time. Consider a linear process \( \{ \eta_j, j \geq 1 \} \) defined by \( \eta_j = \sum_{k=0}^{\infty} \phi_k \epsilon_{j-k} \), where \( \{ \epsilon_j, j \in \mathbb{Z} \} \) is a sequence of i.i.d. random variables with \( E \epsilon_0 = 0 \) and \( E \epsilon_0^2 = 1 \), and the coefficients \( \phi_k, k \geq 0 \).
Local power: \( \nu = 3, \eta_t = \lambda \eta_{t-1} + \varepsilon_t, r = \pm 0.5 \)

| Nominal size 5% | Nominal size 1% |
|-----------------|-----------------|
| \( n \)        | \( h = n^{-1/4} \) | \( h = n^{-1/3} \) | \( h = n^{-1/2.5} \) | \( h = n^{-1/4} \) | \( h = n^{-1/3} \) | \( h = n^{-1/2.5} \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 100             | 0.819           | 0.779           | 0.743           | 0.787           | 0.739           | 0.693           |
| 200             | 0.906           | 0.878           | 0.845           | 0.892           | 0.849           | 0.811           |
| 500             | 0.971           | 0.950           | 0.923           | 0.963           | 0.935           | 0.901           |

- \( r = 0.5, \lambda = 0.4 \)
- \( r = 0.5, \lambda = -0.4 \)
- \( r = -0.5, \lambda = 0.4 \)
- \( r = -0.5, \lambda = -0.4 \)

are assumed to satisfy \( \sum_{k=0}^{\infty} |\phi_k| < \infty \) and \( \phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0 \). Let

\[
y_{k,n} = \rho y_{k-1,n} + \eta_k, \quad y_{0,n} = 0, \quad \rho = 1 + \kappa/n,
\]

where \( \kappa \) is a constant. The array \( y_{k,n}, k \geq 0 \) is known as a nearly unstable process or, in the econometric literature, as a near-integrated time series.

Write \( x_{k,n} = y_{k,n}/\sqrt{n} \phi \). The classical invariance principle gives

\[
x_{[nt],n} \Rightarrow G(t) := \int_0^t e^{\kappa(t-s)} dW(s) = W(t) + \kappa \int_0^t e^{\kappa(t-s)} W(s) ds
\]

on \( D[0,1] \), where \( W(t) \) is a standard Brownian motion [e.g., Phillips (1987), Buchmann and Chan (2007), Wang and Phillips (2009b)]. Furthermore, \( \{\varepsilon_j, j \in \mathbb{Z}\} \) can be redefined on a richer probability space which also contains a standard Brownian motion \( W_1(t) \) such that

\[
\sup_{0 \leq t \leq 1} |x_{[nt],n} - G_1(t)| = o_P(1),
\]

where \( G_1(t) = W_1(t) + \kappa \int_0^t e^{\kappa(t-s)} W_1(s) ds \). Indeed, by noting on the richer space that

\[
\sup_{0 \leq t \leq 1} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \varepsilon_j - W_1(t) \right| = o_P(1)
\]
**Table 5**  
*Local power: $\nu = 2$, $\eta_t = \lambda\eta_{t-1} + \varepsilon_t$, $r = \pm 0.5$*

| $n$ | $h = n^{-1/4}$ | $n^{-1/3}$ | $n^{-1/2.5}$ | $n^{-1/4}$ | $n^{-1/3}$ | $n^{-1/2.5}$ |
|-----|----------------|------------|---------------|------------|------------|---------------|
|     | $r = 0.5, \lambda = 0.4$ |            |               | $r = 0.5, \lambda = -0.4$ |            |               | $r = -0.5, \lambda = 0.4$ |            |               | $r = -0.5, \lambda = -0.4$ |            |               |
| 100 | 0.357          | 0.282      | 0.228         | 0.357      | 0.282      | 0.228         | 0.058         | 0.054       | 0.053         | 0.065       | 0.066       | 0.067         | 0.114         | 0.123       | 0.128         | 0.056       | 0.050       | 0.046         |
| 200 | 0.484          | 0.389      | 0.315         | 0.484      | 0.389      | 0.315         | 0.103         | 0.083       | 0.068         | 0.157       | 0.159       | 0.160         | 0.226         | 0.235       | 0.244         | 0.437       | 0.457       | 0.462         |
| 500 | 0.682          | 0.557      | 0.458         | 0.682      | 0.557      | 0.458         | 0.169         | 0.118       | 0.094         | 0.098       | 0.057       | 0.036         | 0.114         | 0.123       | 0.096         | 0.173       | 0.123       | 0.096         |

[see, e.g., Csörgő and Révész (1981)], and using this result in place of the fact that $\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_j \to W(t)$ on $D[0,1]$, the same technique as in the proof of Phillips (1987) [see also Chan and Wei (1987)] yields

$$\sup_{0 \leq t \leq 1} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \rho^{[nt]-j} \epsilon_j - G_1(t) \right| = o_P(1).$$

The result (5.3) can now be obtained by the same argument, with minor modifications, as in the proof of Proposition 7.1 in Wang and Phillips (2009b).

The aim of this section is to investigate the asymptotic behavior of a functional $S_{[nr]}$ of the $x_{k,n}$, defined by

$$S_{[nr]} = \frac{c_n}{n^r} \sum_{k,j=1}^{[nr]} g[c_n(x_{k,n} - x_{j,n})],$$

(5.5)

where $g$ is a real function on $R$, and $c_n$ is a certain sequence of positive constants. Under certain conditions on $g(x)$, $\epsilon_0$ and $c_n$, it is established that, for each fixed $0 < r \leq 1$, $S_{[nr]}$ converges to an intersection local time process of $G(t)$. Explicitly, we have the following main result.
Table 6

Local power: $\nu = 1.5$, $\eta_t = \lambda \eta_{t-1} + \varepsilon_t$, $r = \pm 0.5$

|      | Nominal size 5% |         | Nominal size 1% |         |
|------|-----------------|---------|-----------------|---------|
|      | $n^{-1/4}$      | $n^{-1/3}$ | $n^{-1/2.5}$    | $n^{-1/4}$ | $n^{-1/3}$ | $n^{-1/2.5}$ |
|      | $r = 0.5$, $\lambda = 0.4$ |         | $r = 0.5$, $\lambda = -0.4$ |         |
| 100  | 0.058           | 0.051   | 0.045           | 0.021   | 0.012   | 0.010   |
| 200  | 0.087           | 0.065   | 0.057           | 0.040   | 0.022   | 0.015   |
| 500  | 0.158           | 0.103   | 0.077           | 0.096   | 0.046   | 0.024   |
|      | $r = -0.5$, $\lambda = 0.4$ |         | $r = -0.5$, $\lambda = -0.4$ |         |
| 100  | 0.043           | 0.040   | 0.041           | 0.016   | 0.014   | 0.012   |
| 200  | 0.061           | 0.058   | 0.055           | 0.024   | 0.019   | 0.015   |
| 500  | 0.096           | 0.074   | 0.070           | 0.038   | 0.031   | 0.023   |
|      | $r = -0.5$, $\lambda = -0.4$ |         |         |         |
| 100  | 0.049           | 0.049   | 0.049           | 0.018   | 0.017   | 0.013   |
| 200  | 0.063           | 0.058   | 0.059           | 0.024   | 0.021   | 0.017   |
| 500  | 0.092           | 0.074   | 0.064           | 0.037   | 0.029   | 0.021   |

Theorem 5.1. Suppose that $\int_{-\infty}^{\infty} |g(x)| \, dx < \infty$, $\omega \equiv \int_{-\infty}^{\infty} g(x) \, dx \neq 0$ and $\int_{-\infty}^{\infty} |Ee^{it\varepsilon_0}| \, dt < \infty$. Then, for any $c_n \to \infty$, $n/c_n \to \infty$ and fixed $r \in (0, 1]$, \begin{equation}
S_{[nr]} \xrightarrow{D} \omega L_G(r, 0),
\end{equation}
where $L_G(t, u)$ is the intersection local time of $G(t)$ defined in (3.8). Furthermore, under the same probability space for which (5.3) holds, we have that, for any $c_n \to \infty$ and $n/c_n \to \infty$,
\begin{equation}
\sup_{0 \leq r \leq 1} |S_{[nr]} - \omega L_G_1(r, 0)| \to p 0.
\end{equation}

The integrability condition on the characteristic function of $\varepsilon_0$ can be weakened if we place further restrictions on $g(x)$. Indeed, we have the following theorem.

Theorem 5.2. Theorem 5.1 still holds if $\int_{-\infty}^{\infty} |Ee^{it\varepsilon_0}| \, dt < \infty$ is replaced by the Cramér condition, that is, $\lim \sup_{|t| \to \infty} |Ee^{it\varepsilon_0}| < 1$, and, in addition to the stated conditions already on $g(x)$, we have $|g(x)| \leq M/(1 + |x|^{1+b})$ for some $b > 0$, where $M$ is a constant.
It is interesting to notice that the additional condition on \( g(x) \) in Theorem 5.2 cannot be reduced without further restriction on \( \epsilon_0 \) like that in Theorem 5.1. This claim can be explained as in Example 4.2.2 of Borodin and Ibragimov (1994) with some minor modifications. On the other hand, the asymptotic behavior of \( S_{[nr]} \) when \( c_n = 1 \) is quite different, as seen in the following theorem.

**Theorem 5.3.** Suppose that \( g(x) \) is Borel measurable function satisfying

\[
\lim_{h \to 0} \int_{-K}^{K} |x|^{\alpha - 1} \sup_{|u| \leq h} |g(x + u) - g(x)| \, dx = 0
\]

for all \( K > 0 \) and some \( 0 < \alpha \leq 1 \). Then, under the same probability space for which (5.3) holds, we have

\[
\sup_{0 \leq r \leq 1} \left| \frac{1}{n^r} \sum_{k,j=1}^{[nr]} g(x_{k,n} - x_{j,n}) - \int_0^r \int_0^r g(G_1(u) - G_1(v)) \, du \, dv \right| = o_P(1).
\]

We mention that condition (5.8) is quite weak. Indeed, example 2.8 and the discussion following Theorem 2.3 in Berkes and Horváth (2006) shows that (5.8) cannot be replaced by

\[
\lim_{h \to 0} \int_{-K}^{K} |x|^{\alpha - 1} |g(x + u) - g(x)| \, dx = 0
\]

for all \( K > 0 \) and some \( 0 < \alpha \leq 1 \).

Local time has figured in much recent work on parametric and nonparametric estimation with nonstationary data. Motivated by nonlinear regression with integrated time series [Park and Phillips (1999, 2001)] and nonparametric estimation of nonlinear cointegration models, many authors [Phillips and Park (1998), Karlsen and Tjøstheim (2001), Karlsen, Myklebust and Tjøstheim (2007), Wang and Phillips (2009a)] have used or proved weak convergence to the local time of a stochastic process, including results of the following type: under certain conditions on the function \( g \), the limiting stochastic process \( G(t) \), a sequence \( c_n \to \infty \), and normalized data \( x_{k,n} \)

\[
\frac{c_n}{n} \sum_{k=1}^{[nr]} g(c_n x_{k,n}) \to_D \omega_{G}(1, 0),
\]

where \( \ell_G(t, s) \) is the local time of the process \( G(t) \) at the spatial point \( s \). We refer to Borodin and Ibragimov (1994) (and their references for related work) for the particular situation where \( c_n x_{k,n} \) is a partial sum of i.i.d. random variables, and to Akonom (1993), Phillips and Park (1998), Jeganathan (2004) and de Jong and Wang (2005) for the case where \( c_n x_{k,n} \) is a partial
sum of a linear process. Wang and Phillips [(2009a), Theorem 2.1] generalized these results to include not only linear process partial sums but also cases where \( c_n x_{k,n} \) is a partial sum of a Gaussian process, including fractionally integrated time series.

Our present research on the statistic \( S_{[nr]} \) in (5.5) has a similar motivation to this earlier work on convergence to a local time process. However, the statistic \( S_{[nr]} \) has a much more complex U-statistic form, and the technical difficulties of establishing weak convergence are greater. The approach of Wang and Phillips [(2009a), Theorem 2.1] remains useful, however, and is implemented in the proofs of Theorems 3.1–3.3.

Finally we mention some earlier work investigating the intersection local time process and weak convergence for certain specialized situations. This work restricts the function \( g \) in (5.5) to the indicator function and the discrete process \( y_{k,n} \) in (5.1) to a lattice random walk taking integer values; see, for instance, Aldous (1986), van der Hofstad, den Hollander and König (1997), van der Hofstad and König (2001) and van der Hofstad, den Hollander and König (2003). The present paper seems to the first to consider weak convergence to intersection local time for a general linear process and a general function \( g \).

The proofs of Theorems 5.1–5.3 are given in the supplementary material [Wang and Phillips (2012)].

6. Proofs of Theorems 3.1–3.3. We start with several propositions. Their proofs are given in the supplementary material [Wang and Phillips (2012)]. Throughout the section, we let \( C, C_1, C_2, \ldots \) be constants which may differ at each appearance.

**Proposition 6.1.** Suppose Assumptions 1 and 2 hold. For any \( \alpha_1, \alpha_2 \geq 0 \), if \( \sup_x |p(x)| < \infty, \int (1 + |x|^{\max\{\alpha_1, \alpha_2\} + 1}) |p(x)| \, dx < \infty \) and \( E|\epsilon_0|^{\alpha_1 + \alpha_2 + 2} < \infty \), then

\[
\Lambda_n := \sum_{s,t=1 \atop s \neq t}^n g(u_{s+1})g_1(u_{t+1})(1 + |x_s|^{\alpha_1})(1 + |x_t|^{\alpha_2})p[(x_t - x_s)/h]
\]

(6.1)

\[= O_P(n^{3/2 + \alpha_1/2 + \alpha_2/2} h),\]

where \( g(x) \) and \( g_1(x) \) are real functions such that

\[\sup_{s \geq 1} E\{[g^2(u_{s+1}) + g_1^2(u_{s+1})] \mid F_s\} < \infty.\]

If additionally \( \alpha_1 > 0 \), then

\[\tilde{\Lambda}_n := \sum_{1 \leq s < t \leq n} g(u_{s+1})(1 + |x_s|^{\alpha_1 - 1})p[(x_t - x_s)/h]
\]

(6.2)

\[= O_P(n^{\max\{3/2, 1 + \alpha_1/2\}} h).\]
Proposition 6.2. Suppose Assumptions 1–3 hold. Then, for any \(g(x, \theta)\) satisfying (2.5) and \(|g(x, \theta)| \leq C(1 + |x|^\beta)\), where \(\theta \in \Omega_0\), we have

\[
(6.3) \quad \Delta_n := \sum_{s, t=1}^{n} u_{s+1}g(x_t, \theta)K[(x_t - x_s)/h] = O_P(n^{5/4+\beta/2}h^{3/4}),
\]
provided that \(nh^2 \to \infty\), \(nh^4 \to 0\), \(\int (1 + |x|^\beta + 1)K(x)dx < \infty\) and \(E|\epsilon_0|^\beta + 2 < \infty\). Similarly, (6.3) holds true if we replace \(g(x, \theta)\) and \(\beta\) by \(m(x)\) and \(\nu\), respectively, where \(m(x)\) is defined as in Assumption 6.

Proposition 6.3. Suppose Assumptions 1–3 hold and \(nh^2 \to \infty\). Then, for any real function \(g(x)\) satisfying \(\sup_s \geq 1 E\{g^2(u_{s+1}) | F_s\} < \infty\), we have

\[
(6.4) \quad \Gamma_n := \sum_{s, t=1}^{n} g(u_{s+1})(u_{t+1}^2 - \sigma^2)K^2[(x_t - x_s)/h] = o_P(n^{3/2}h).
\]

Proposition 6.4. In addition to Assumptions 1–3, we have \(|u_j| \leq A\) and \(nh^2 \to \infty\). Then,

\[
(6.5) \quad R_n := \sum_{t=1}^{n} \sum_{i, j=1}^{t-1} u_{i+1}u_{j+1}K[(x_t - x_i)/h]K[(x_t - x_j)/h] = o_P(n^{3/2}h).
\]

Proposition 6.5. Under Assumptions 1–3 and \(h \log^2 n \to 0\), we have

\[
(6.6) \quad EZ_{tkr}^2 \leq C \max_{1 \leq i,j \leq n} E[|u_i|(1 + |u_j|)](1 + h\sqrt{t} - r - k)
\]
for \(1 \leq k \leq t - r\) and \(r \geq 1\), where \(Z_{tkr} = \sum_{i=k}^{t-r} u_iK[(x_t - x_i)/h]\). Similarly,

\[
(6.7) \quad E \left\{ \sum_{i=1}^{t-1} u_{i+1}^2 - E(u_{i+1}^2 | F_j) \right\} K^2[(x_t - x_i)/h] \right\}^2 \leq C(1 + h\sqrt{t}).
\]

If in addition \(|u_j| \leq A\), where \(A\) is a constant, then

\[
(6.8) \quad EZ_{tkr}^2 \leq Ch^3 t^{3/2},
\]
and for any \(1 \leq m \leq t/2\),

\[
(6.9) \quad EZ_{tm}^2 \leq \frac{Ch^2 t^2}{m^{3/2}} + \frac{Ch^2 t \log(t - m)}{\sqrt{m}} + \frac{Ch^2 t}{m},
\]
where \(Z_{tm} = \sum_{i=1}^{t-m-1} u_{i+1}E(K[(x_t - x_i)/h] | F_{t-m})\).
6.1. Proof of Theorem 3.1. By virtue of (3.5) and Theorem 3.3, it suffices to verify (3.6) and show that

\[(6.10)\]

\[S_{2n} = o_P(n^{3/4} \sqrt{h}) \quad \text{and} \quad S_{3n} = o_P(n^{3/4} \sqrt{h}).\]

To this end, for \(\delta > 0\), let \(\Omega_n = \{\hat{\theta} : \|\hat{\theta} - \theta\| \leq \delta \delta_n, \theta \in \Omega_0\}\), where \(\delta_n\) is given in Assumption 4(i).

We first prove (6.10). Note that \(\Omega_n \subset \Theta_0\) for all \(n\) sufficiently large. Under Assumption 4, it follows by Taylor's expansion that, whenever \(n\) is sufficiently large and \(\hat{\theta} \in \Omega_n\),

\[(6.11)\]

\[S_{2n} = (\theta - \hat{\theta}) \sum_{i,t=1}^{n} u_{i+1} \frac{\partial f(x_t, \theta)}{\partial \theta} K([x_t - x_i]/h) + S_{2n1},\]

where

\[S_{2n1} \leq C |\hat{\theta} - \theta|^2 \sum_{i,t=1}^{n} |u_{i+1}| (1 + |x_t|^\beta) K([x_t - x_i]/h).\]

By Proposition 6.2 with \(g(x, \theta) = \frac{\partial f(x, \theta)}{\partial \theta}\) and \(\delta_n^2 n^{1+\beta} \sqrt{h} \to 0\), the first term in the decomposition of \(S_{1n}\) is equal to

\[O_P(\delta_n n^{5/4+\beta/2} h^{3/4}) = o_P(n^{3/4} \sqrt{h}).\]

On the other hand, by Proposition 6.1 and \(nh^2 \to \infty\), we get

\[S_{2n1} = O_P(\delta_n^2 n^{3/2+\beta/2} h) = o_P(n^{3/4} \sqrt{h}).\]

These facts imply, for any \(\delta > 0\),

\[(6.12)\]

\[P(|S_{2n}| \geq \delta n^{3/4} \sqrt{h}) \leq P(|S_{2n}| \geq \delta n^{3/4} \sqrt{h}, \hat{\theta} \in \Omega_n) + P(\|\hat{\theta} - \theta\| \geq \delta \delta_n) \to 0 \quad \text{as} \quad n \to \infty.\]

Similarly, by using Proposition 6.1 and noting

\[|S_{3n}| \leq C |\hat{\theta} - \theta|^2 \sum_{i,t=1}^{n} \left| \frac{\partial f(x_t, \theta)}{\partial \theta} \right| \left| \frac{\partial f(x_t, \theta)}{\partial \theta} \right| K([x_t - x_i]/h)\]

\[(6.13)\]

\[\leq C \delta_n^2 \sum_{i,t=1}^{n} (1 + |x_i|^\beta) (1 + |x_t|^\beta) K([x_t - x_i]/h)\]

\[= O_P(\delta_n^2 n^{3/2+\beta} h) = o_P(n^{3/4} \sqrt{h}),\]
whenever $\hat{\theta} \in \Omega_n$, we obtain, for any $\delta > 0$,

$$P(|S_{3n}| \geq \delta n^{3/4} \sqrt{h})$$

(6.14) \[ \leq P(|S_{3n}| \geq \delta n^{3/4} \sqrt{h}, \hat{\theta} \in \Omega_n) + P(|\hat{\theta} - \theta| \geq \delta \delta_n) \rightarrow 0 \quad \text{as } n \to \infty. \]

Combining (6.12) and (6.14), we obtain (6.10).

We next prove (3.6). We may write

$$V_n^2 = \sum_{s,t=1}^{n} u_{s+1}^2 u_{t+1}^2 K^2[(x_t - x_s)/h]$$

$$+ \sum_{s,t=1}^{n} (u_{s+1}^2 - u_{s+1}^2) u_{t+1}^2 K^2[(x_t - x_s)/h]$$

(6.15)

$$+ \sum_{s,t=1}^{n} u_{s+1}^2 (u_{t+1}^2 - u_{t+1}^2) K^2[(x_t - x_s)/h]$$

$$:= V_{1n} + V_{2n} + V_{3n}.$$  

Recall $|f(x_s, \theta) - f(x_s, \hat{\theta})| \leq C \delta_n (1 + |x_s|^{\beta})$ whenever $\hat{\theta} \in \Omega_n$ and $|\hat{u}_{t+1}^2 - u_{t+1}^2| = 2|u_{t+1}||f(x_s, \theta) - f(x_s, \hat{\theta})| + |f(x_s, \theta) - f(x_s, \hat{\theta})|^2$. It is readily seen from Proposition 6.1 that, given $\hat{\theta} \in \Omega_n$,

$$|V_{2n}| + |V_{3n}| \leq C \delta_n \sum_{s,t=1}^{n} |u_{s+1}| u_{t+1}^2 (1 + |x_s|^{\beta}) K[(x_t - x_s)/h]$$

$$+ C \delta_n \sum_{s,t=1}^{n} u_{t+1}^2 (1 + |x_s|^{2\beta}) K[(x_t - x_s)/h]$$

$$+ C \delta_n \sum_{s,t=1}^{n} |u_{s+1}| (1 + |x_s|^{\beta})(1 + |x_t|^{2\beta}) K[(x_t - x_s)/h]$$

$$+ C \delta_n \sum_{s,t=1}^{n} (1 + |x_s|^{2\beta})(1 + |x_t|^{2\beta}) K[(x_t - x_s)/h]$$

$$= O_P(n^{3/2}h)(\delta_n n^{\beta/2} + \delta_n^2 n^{\beta} + \delta_n^3 n^{3\beta/2} + \delta_n^4 n^{2\beta})$$

$$= o_P(n^{3/2}h),$$
since $nh^2 \to \infty$ and $\sigma_n^2 n^{1+\beta}/\sqrt{h} \to 0$. As for $V_{1n}$, by Proposition 6.3, we have

$$V_{1n} = \sigma^4 \sum_{s,t=1}^{n} K^2[(x_t - x_s)/h] + \sum_{s,t=1}^{n} (u_{t+1}^2 + \sigma^2)(u_{s+1}^2 - \sigma^2)K^2[(x_t - x_s)/h]$$

$$= \sigma^4 \sum_{s,t=1}^{n} K^2[(x_t - x_s)/h] + o_P(n^{3/2}h).$$

Taking these estimates into (6.15), we get the first part of (3.6).

In order to prove the second part of (3.6), we first assume $|u_j| \leq A$. In this case, simple calculations together with Propositions 6.3 and 6.4 yield that

$$\sum_{t=2}^{n} V_{nt}^2 = \sum_{t=2}^{n} \sum_{s=1}^{t-1} u_{s+1}^2 K^2[(x_t - x_s)/h]$$

$$+ \sum_{t=1}^{n} \sum_{i,j=1}^{t-1} u_{i+1} u_{j+1} K[(x_t - x_i)/h] K[(x_t - x_j)/h]$$

$$= \frac{\sigma^2}{2} \sum_{s,t=1}^{n} K^2[(x_t - x_s)/h] + o_P(n^{3/2}h)$$

as required. The idea to remove the restriction $|u_j| \leq A$ is the same as in the proof of Theorem 3.3. We omit the details. The proof of Theorem 3.1 is now complete.

6.2. Proof of Theorem 3.2. Put $\hat{u}_{t+1} = u_{t+1} + f(x_t, \theta) - f(x_t, \hat{\theta})$. Under $H_1$, we may write

$$S_n = S_{1n} + 2S_{2n} + S_{3n} - S_{4n} + S_{5n},$$

where $S_{1n}, S_{2n}, S_{3n}$ are defined as in (3.5), and

$$S_{4n} = 2\rho_n \sum_{i,t=1}^{n} m(x_i) \hat{u}_{t+1}^i K[(x_t - x_i)/h],$$

$$S_{5n} = \rho_n^2 \sum_{i,t=1}^{n} m(x_i) m(x_t) K[(x_t - x_i)/h].$$
Thus (3.4) will follow if we prove

\begin{align}
S_{jn} &= O_P(n^{3/4}h^{1/2}), \quad j = 1, 2, 3, \tag{6.18} \\
S_{4n} &= O_P(\rho_n n^{5/4+\nu/2}h^{3/4}), \tag{6.19} \\
V_n^2 &= O_P(n^{3/2}h + \rho_n^4 n^{3/2+2\nu}h) \quad \text{under } H_1, \tag{6.20}
\end{align}

and for any \(\epsilon_n \to 0,\)

\begin{equation}
S_{5n} \geq \epsilon_n \rho_n^2 n^{3/2+\nu}h \quad \text{in Probab.} \tag{6.21}
\end{equation}

Here and below, the notation \(A_n \geq B_n,\) in Probab. means that \(\lim_{n \to \infty} P(A_n \geq B_n) = 1,\) as \(n \to \infty.\) Indeed, by choosing \(\epsilon_n^{-2} = \min\{\rho_n^2 n^{1+\nu} \sqrt{\ln n}, n^{3/2} \sqrt{\ln n}\},\)

it is readily seen that \(\epsilon_n \to 0,\) \(|S_{jn}| = O_P(\epsilon_n S_{5n}) = o_P(S_{5n})\) for \(j = 1, 2, 3, 4\) and \(S_{5n}/V_n \geq \epsilon_n^{-1},\) in Probab. Hence \(S_n/V_n \geq \epsilon_n^{-1}/2,\) in Probab., which yields (3.4).

We next prove (6.19)–(6.21). The proof of (6.18) for \(j = 2, 3\) is given in (6.10), and the result for \(j = 1\) is simple by martingale properties and Proposition 6.5.

Equation (6.21) first. We may write

\begin{equation}
S_{5n} = S_{5n1} + S_{5n2}, \tag{6.22}
\end{equation}

where \(S_{5n1} = 2\rho_n^2 \sum_{1 \leq i < t \leq n} m^2(x_i)K[(x_t - x_i)/h]\) and

\[|S_{5n2}| \leq 2\rho_n^2 \sum_{1 \leq i < t \leq n} |m(x_i)||m(x_t) - m(x_i)|K[(x_t - x_i)/h].\]

Let \(\nu' = \nu\) if \(\nu > 0\) and \(\nu' = \gamma'\) if \(\nu = 0.\) It follows from (3.3) and Proposition 6.1 that

\[|S_{5n2}| \leq Ch^\gamma \rho_n^2 \sum_{1 \leq i < t \leq n} (1 + |x_i|^{\nu})(1 + |x_i|^{\nu'-1} + |x_t - x_i|^{\nu'})K_{\gamma}[(x_t - x_i)/h]\]

\[\leq Ch^\gamma \rho_n^2 \sum_{1 \leq i < t \leq n} \{(1 + |x_i|^{\nu'-1} + |x_i|^{\nu' + \nu'-1})K_{\gamma}[(x_t - x_i)/h]\]

\begin{equation}
= O_P(h^{1+\gamma} \rho_n^2 (n^{\max\{3/2,1+(\nu+\nu')/2\}} + n^{3/2+\nu/2})) \tag{6.23}
\end{equation}

where \(K_u(x) = |x|^u K(x), u > 0\) and we have used the fact that \(\sup_x |K_u(x)| < \infty\) whenever \(\int K_u(x) dx < \infty\) [recall \(\sup_x |K(x)| < \infty\). Since \(h \to 0\) and \(0 < \gamma \leq 1,\) to prove (6.21), it only needs to show that, for any \(h^{\gamma/2} \leq \epsilon_n \to 0,\)

\begin{equation}
S_{5n1} \geq \epsilon_n \rho_n^2 n^{3/2+\nu} \quad \text{in Probab.} \tag{6.24}
\end{equation}
In fact, by (5.3) and letting $x_{[ns],n} = x_{[ns]}/(\sqrt{n}\phi)$,

$$\inf_{n/2 \leq j \leq n} |x_j| \geq \sqrt{n}\phi\left(\inf_{1/2 \leq s \leq 1} |G_1(s)| - \sup_{1/2 \leq s \leq 1} |x_{[ns],n} - G_1(s)|\right)$$

\begin{equation}
\geq \epsilon_n^{1/4}\sqrt{n} \quad \text{in Probab.}
\end{equation}

Similarly, by using (5.7) in Theorem 5.1, we have

\begin{equation}
\sum_{n \geq t > i \geq n/2} K[(x_t - x_i)/h] \geq \epsilon_n^{1/4} n^{3/2} h \quad \text{in Probab.}
\end{equation}

Combining (3.2), (6.25) and (6.26), we obtain that

$$S_{5n1} \geq \epsilon_n^{1/2} \rho_n^2 \sum_{n \geq t > i \geq n/2} |x_i|^{2\nu} |f(|x_i| \geq 1)K[(x_t - x_i)/h]$$

$$\geq \epsilon_n^{3/4} \rho_n^2 n^{\nu} \sum_{n \geq t > i \geq n/2} K[(x_t - x_i)/h]$$

$$\geq \epsilon_n \rho_n^2 n^{3/2 + \nu} h \quad \text{in Probab.}$$

This provides (6.24) and also completes the proof of (6.21).

Next prove (6.19). We have

$$S_{4n} = 2\rho_n \sum_{i,t=1}^n m(x_i) u_{t+1} K[(x_t - x_i)/h] + S_{4n1},$$

where, by recalling $|f(x_t, \theta) - f(x_t, \hat{\theta})| \leq C||\hat{\theta} - \theta||(1 + |x_t|^\beta)$ by Assumption 4, it follows from Proposition 6.1 that

$$|S_{4n1}| \leq \sum_{i,t=1}^n (1 + |x_i|^\nu)|f(x_t, \theta) - f(x_t, \hat{\theta})|K[(x_t - x_i)/h]$$

$$\leq C \rho_n ||\hat{\theta} - \theta|| \sum_{i,t=1}^n (1 + |x_i|^\nu)(1 + |x_t|^\beta)K[(x_t - x_i)/h]$$

$$= O_P(\rho_n \delta_n n^{3/2 + \nu/2 + \beta/2} h).$$

This, together with Proposition 6.2, yields that

$$S_{4n} = O_P(\rho_n \delta_n n^{3/2 + \nu/2 + \beta/2} h) + O_P(\rho_n n^{5/4 + \nu/2} h^{3/4})$$

$$= O_P(\rho_n n^{5/4 + \nu/2} h^{3/4}),$$

since $\delta_n^2 n^{1+\beta} \rightarrow 0$. The result (6.19) is proved.
Finally, we prove (6.20). Under $H_1$, we have

$$ V_n^2 = \sum_{s,t=1, s \neq t}^{n} [\hat{u}_{t+1}^s + \rho_n m(x_t)]^2 [\hat{u}_{s+1}^t + \rho_n m(x_s)]^2 K^2[(x_t - x_s)/h] $$

(6.27)

$$ \leq 2V_{6n} + 4V_{7n} + 2V_{8n}, $$

where

$$ V_{6n} = \sum_{s,t=1}^{n} \hat{u}_{t+1}^s \hat{u}_{s+1}^t K^2[(x_t - x_s)/h], $$

$$ V_{7n} = \rho_n^2 \sum_{s,t=1}^{n} \hat{u}_{t+1}^s m^2(x_s) K^2[(x_t - x_s)/h], $$

$$ V_{8n} = \rho_n^4 \sum_{s,t=1}^{n} m^2(x_t) m^2(x_s) K^2[(x_t - x_s)/h]. $$

By recalling $|m(x)| \leq C|x|^\nu$ and

$$ \hat{u}_{t+1}^s \leq 2(u_{t+1}^s + |f(x_t, \theta) - f(x_t, \hat{\theta})|^2) \leq C[u_{t+1}^s + O_P(\delta_n^2)(1 + |x_t|^{2\beta})], $$

it following repeatedly from Proposition 6.1 and $\delta_n^2 n^{1+\beta} \sqrt{n} \to 0$ that

$$ V_{6n} \leq C \sum_{s,t=1}^{n} [u_{s+1}^2 + O_P(\delta_n^2)(1 + |x_s|^{2\beta})] [u_{t+1}^2 + O_P(\delta_n^2)(1 + |x_t|^{2\beta})] \times K^2[(x_t - x_s)/h] $$

$$ = O_P(n^{3/2} h) + O_P(\delta_n^2 n^{3/2+\beta} h) + O_P(\delta_n^4 n^{3/2+2\beta} h) $$

$$ = O_P(n^{3/2} h). $$

Similarly, we have

$$ V_{7n} \leq C \rho_n^2 \sum_{s,t=1}^{n} [u_{s+1}^2 + O_P(\delta_n^2)(1 + |x_s|^{2\beta})] [u_{t+1}^2 + O_P(\delta_n^2)(1 + |x_t|^{2\beta})] K^2[(x_t - x_s)/h] $$

$$ = O_P(\rho_n^2 n^{3/2+\nu} h) + O_P(\rho_n^2 \delta_n^2 n^{3/2+\beta+\nu} h) = O_P(\rho_n^2 n^{3/2+\nu} h), $$
V_{8n} \leq \rho_n^4 \sum_{s,t=1 \atop s \neq t}^{n} (1 + |x_t|^{2\nu}) (1 + |x_s|^{2\nu}) K^2 ([x_t - x_s]/h) \\
= O_P(\rho_n^4 n^{3/2 + 2\nu} h).

Combining all these estimates, we obtain
\[ V_2^2 = O_P(n^{3/2} h) + O_P(\rho_n^2 n^{3/2 + \nu} h) + O_P(\rho_n^4 n^{3/2 + 2\nu} h) \]
\[ = O_P(n^{3/2} h + \rho_n^4 n^{3/2 + 2\nu} h) \]
as required. The proof of Theorem 3.3 is complete.

6.3. Proof of Theorem 3.3. We first assume |u_t| \leq A, where A is a constant. This restriction will be removed later. Write G_n(t) = x_{[nt]}/\sqrt{n \delta} and V_n(t) = \sum_{j=1}^{[nt]} u_{j+1}/\sqrt{n \delta}. Under Assumptions 1 and 2, the same arguments as those in Buchmann and Chan (2007) or Wang and Phillips (2009b), with minor modifications, show that
\[ (G_n, V_n) \Rightarrow_D (G, V) \]
on D[0,1]^2, where G(t) = W(t) + \kappa \int_0^t e^{\epsilon(t-s)} W(s) ds. By virtue of (6.28), it follows from the so-called Skorohod–Dudley–Wichura representation theorem that there is a common probability space (\Omega, \mathcal{F}, P) supporting (G_n^0, V_n^0) and (G, V) such that
\[ (G_n, V_n) \Rightarrow d (G_n^0, V_n^0) \text{ and } (G_n^0, V_n^0) \Rightarrow_{\text{a.s.}} (G, V) \]
in D[0,1]^2 with the uniform topology. Moreover, as in the proof of Lemma 2.1 in Park and Phillips (2001), \( V_n^0 \) can be chosen such that, for each \( n \geq 1 \),
\[ V_n^0(k/n) = V(\tau_{nk}/n), \quad k = 1, 2, \ldots, n, \]
where \( \tau_{nk}, 1 \leq k \leq n \), are stopping times with respect to \( \mathcal{F}_{n,k}^0 \) in (\( \Omega, \mathcal{F}, P \)) with
\[ \mathcal{F}_{n,k}^0 = \sigma\{V(r), r \leq \tau_{nk}/n; G_n^0(s/n), s = 1, \ldots, k + 1\}, \]
satisfying \( \tau_{n,0} = 0 \),
\[ \sup_{1 \leq k \leq n} \left| \frac{\tau_{nk} - k}{n^\delta} \right| \Rightarrow_{\text{a.s.}} 0 \]
as \( n \to \infty \) for any \( 1/2 < \delta < 1 \), and
\[ E[(\tau_{nk} - \tau_{nk-1}) | \mathcal{F}_{n,k-1}^0] = \sigma^{-2} E[u_{k+1}^2 | \mathcal{F}_k] \text{ and } \\
E[(\tau_{nk} - \tau_{nk-1})^2 | \mathcal{F}_{n,k-1}^0] \leq C \sigma^{-4m} E[u_{k+1}^{4m} | \mathcal{F}_k], \quad m \geq 1, \text{ a.s.} \]
for some constant \( C > 0 \). We mention that result (6.32) does not explicitly appear in Lemma 2.1 of Park and Phillips (2001); however, it can be obtained by a construction along the same lines as Theorem A1 of Hall and Heyde (1980).
It follows from (6.30) that, under the extended probability space,

\[
\left( \frac{1}{\sigma d_n} \sum_{t=2}^{n} u_{t+1} Y_{nt}, \frac{1}{d_n^2} \sum_{t=2}^{n} Y_{nt}^2 \right)
\]

\[= d \left( \sum_{t=2}^{n} [V(\tau_{nt}/n) - V(\tau_{nt-1}/n)] Y_{nt}^* + \frac{1}{n} \sum_{t=2}^{n} Y_{nt}^2 \right),\]

where, with \( c_n = \sqrt{n\phi/h} \),

\[Y_{nt}^* = \frac{n\sigma}{d_n} \sum_{i=1}^{t-1} [V(\tau_{ni}/n) - V(\tau_{ni-1}/n)] K \{ c_n [G_n^0(\tau_{ni}/n) - G_n^0(\tau_{ni-1}/n)]/n \}.\]

To establish our main result, we extend \( \sum_{i=2}^{n} [V(\tau_{nt}/n) - V(\tau_{nt-1}/n)] Y_{nt}^* \) to a continuous martingale. This can be done by defining

\[M_n(r) = \sum_{t=2}^{j-1} Y_{nt}^* \left[ V\left( \frac{\tau_{nt}}{n} \right) - V\left( \frac{\tau_{nt-1}}{n} \right) \right] + Y_{n,j}^* \left[ V(r) - V\left( \frac{\tau_{nj-1}}{n} \right) \right]\]

for \( \tau_{nj-1}/n < r \leq \tau_{nj}/n, j = 1, 2, \ldots, n \), and

\[M_n(r) = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} Y_{nt}^* \left[ V\left( \frac{\tau_{nt}}{n} \right) - V\left( \frac{\tau_{nt-1}}{n} \right) \right] + \frac{1}{\sqrt{n}} \left[ V(r) - V\left( \frac{\tau_{n,n}}{n} \right) \right]\]

for \( r \geq \tau_{n,n}/n \). It is readily seen that \( M_n \) is a continuous martingale with quadratic variation process \([M_n]\) given by

\[\lfloor M_n \rfloor = \sum_{t=2}^{j-1} Y_{nt}^2 \left( \frac{\tau_{nt}}{n} - \frac{\tau_{nt-1}}{n} \right) + Y_{n,j}^2 \left( r - \frac{\tau_{nj-1}}{n} \right)\]

for \( \tau_{nj-1}/n < r \leq \tau_{nj}/n, j = 1, 2, \ldots, n \), and

\[\lfloor M_n \rfloor = \frac{1}{n} \left( r - \frac{\tau_{n,n}}{n} \right)\]

for \( r \geq \tau_{n,n}/n \). Similarly, the covariance process \([M_n, V]\) of \( M_n \) and \( V \) is given by

\[\lfloor M_n, V \rfloor = \sum_{t=2}^{j-1} Y_{nt}^* \left( \frac{\tau_{nt}}{n} - \frac{\tau_{nt-1}}{n} \right) + Y_{n,j}^* \left( r - \frac{\tau_{nj-1}}{n} \right)\]

for \( \tau_{nj-1}/n < r \leq \tau_{nj}/n, j = 1, 2, \ldots, n \), and

\[\lfloor M_n, V \rfloor = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} Y_{nt}^* \left( \frac{\tau_{nt}}{n} - \frac{\tau_{nt-1}}{n} \right) + \frac{1}{\sqrt{n}} \left( r - \frac{\tau_{n,n}}{n} \right)\]

for \( r \geq \tau_{n,n}/n \).
Write $\rho_n(t) = \inf \{ s : [M_n]_s > t \}$, a sequence of time changes. Note that $[M_n]_\infty = \infty$ for every $n \geq 1$ and

$$[M_n, V]_{\rho_n(t)} \to P 0 \quad \text{as } n \to \infty \quad (6.40)$$

for every $t \in R$, by (6.42) in Proposition 6.6 below. Theorem 2.3 of Revuz and Yor [(1999), page 524] yields that, if we call $B^n$ [i.e., $B^n(r) = M_n\{\rho_n(r)\}$] the DDS Brownian motion [see, e.g., Revuz and Yor (1999), page 181] of the continuous martingale $M_n$ defined by (6.34) and (6.35), then $B^n$ converges in distribution to a Wiener process $W$. Since the law of the processes $B^n$ are all given by Wiener measure, it is plain that $B^n(r) \Rightarrow W(r)$ (mixing), where the concept of mixing can be found in Hall and Heyde (1980), page 56. This, together with (6.43) in Proposition 6.7 below, yields that $(B^n(r), [M_n]_1) \Rightarrow (W(r), \eta^2)$, where $W$ is independent of $\eta^2 = L_G(1,0)$, defined as in (3.8).

Now, by noting that $M_n(1)$ is equal to $B^n([M_n]_1)$, the continuous mapping theorem implies that

$$[M_n(1), [M_n]_1] \to D (\eta N, \eta^2), \quad (6.41)$$

where $N$ is a normal variate independent of $\eta$.

By virtue of (6.33) and (6.41), the required result of the theorem follows (6.44) and (6.45) in Proposition 6.7 and Proposition 6.8 below.

It remains to show the following Propositions 6.6–6.8, whose proofs are given in the supplementary material [Wang and Phillips (2012)]. The proof of Theorem 3.3 under $|u_j| \leq A$ is now complete.

**Proposition 6.6.** In addition to Assumptions 1–3, assume that $|u_j| \leq A$, $nh^2 \to \infty$ and $h \log^2 n \to 0$. Then, as $n \to \infty$,

$$[M_n, V]_r \to 0 \quad \text{in Probab.} \quad (6.42)$$

uniformly on $r \in [0, T]$, where $T$ is an arbitrary given constant.

**Proposition 6.7.** In addition to Assumptions 1–3, assume that $|u_j| \leq A$, $nh^2 \to \infty$ and $nh^4 \log^2 n \to 0$. Under the extended probability space used in (6.29), we have

$$[M_n]_1 \to P \eta^2, \quad (6.43)$$

where $\eta^2 = L_G(1,0)$ is defined as in (3.8), and

$$[M_n]_1 - \frac{1}{n} \sum_{t=1}^{n} Y^*_{nt} = o_P(1). \quad (6.44)$$

**Proposition 6.8.** In addition to Assumptions 1–3, assume that $|u_j| \leq A$, $nh^2 \to \infty$ and $nh^4 \log^2 n \to 0$. Then,

$$M_n(1) - \sum_{t=2}^{n} Y^*_{nt} \left[ V\left(\frac{\tau_{nt}}{n}\right) - V\left(\frac{\tau_{nt-1}}{n}\right) \right] = o_P(1). \quad (6.45)$$
We next remove the restriction $|u_j| \leq A$. To this end, let

$$
\begin{align*}
    u_{1j} &= u_j I(|u_j| \leq A/2) - E[u_j I(|u_j| \leq A/2) | \mathcal{F}_{j-1}], \\
    u_{2j} &= u_j I(|u_j| > A/2) - E[u_j I(|u_j| > A/2) | \mathcal{F}_{j-1}]
\end{align*}
$$

and

$$
Y_{1nt} = \sum_{i=1}^{t-1} u_{1,i+1} K[(x_t - x_i)/h], \quad Y_{2nt} = \sum_{i=1}^{t-1} u_{2,i+1} K[(x_t - x_i)/h].
$$

With this notation, we may write

$$
\begin{align*}
    \frac{1}{d_n} \sum_{t=2}^{n} u_{t+1} Y_{nt} &= \frac{1}{d_n} \sum_{t=2}^{n} u_{1,t+1} Y_{1nt} + \frac{1}{d_n} \sum_{t=2}^{n} u_{1,t+1} Y_{2nt} + \frac{1}{d_n} \sum_{t=2}^{n} u_{2,t+1} Y_{nt} \\
    &= \frac{1}{d_n} \sum_{t=2}^{n} u_{1,t+1} Y_{1nt} + \Lambda_{1n} + \Lambda_{2n}, \\
    \frac{1}{d^2_n} \sum_{t=2}^{n} Y_{2nt} &= \frac{1}{d^2_n} \sum_{t=2}^{n} Y_{1nt}^2 + \frac{2}{d^2_n} \sum_{t=2}^{n} Y_{1nt} Y_{2nt} + \frac{1}{d^2_n} \sum_{t=2}^{n} Y_{2nt}^2 \\
    &= \frac{1}{d^2_n} \sum_{t=2}^{n} Y_{1nt}^2 + \Lambda_{3n} + \Lambda_{4n}.
\end{align*}
$$

Recall that $|u_{1j}| \leq A$, and $u_{1j}$ is a martingale difference satisfying

$$
E(u_{1t}^2 | \mathcal{F}_{t-1}) = E(u_{1t}^2 I(|u_t| \leq A) | \mathcal{F}_{t-1})
$$

$$
- [E(u_t I(|u_t| \leq A) | \mathcal{F}_{t-1})]^2
\rightarrow \sigma^2 \quad \text{a.s.}
$$

as $j, A \rightarrow \infty$. It follows from the proof of (3.7) under $|u_j| \leq A$ that, when $n \rightarrow \infty$ first, and then $A \rightarrow \infty$,

$$
\left(\frac{1}{\sigma d_n} \sum_{t=2}^{n} u_{1,t+1} Y_{1nt}, \frac{1}{d^2_n} \sum_{t=2}^{n} Y_{1nt}^2\right) \rightarrow_D (\eta N, \eta^2).
$$

Now it is readily seen that the required result will follow if we prove

$$
\Lambda_{in} \rightarrow 0, \quad i = 1, 2, 3, 4,
$$

as $n \rightarrow \infty$ first, and then $A \rightarrow \infty$. In fact, by virtue of (6.6) in Proposition 6.5,

$$
\sup_{1 \leq i \leq n} E u_i^2 \leq \sup_{1 \leq i \leq n} (E u_i^4)^{1/4} < \infty
$$
and supₓ K(x) < ∞, we have, for 1 ≤ t ≤ n,

\[ EY_{nt}^2 \leq 2 \supₓ K(x) EU_t^2 + 2E \left( \sum_{i=1}^{t-2} u_{i+1} K[(x_t - x_i)/h] \right)^2 \]

\[ \leq C \sup_{1 \leq i \leq n} EU_i^2 (1 + h^2 \sqrt{t} \log t + h \sqrt{t}) \leq C_1 h \sqrt{n}, \]

since h log n → 0 and nh^2 → ∞. Similarly,

\[ EY_{1nt}^2 \leq C \sup_{1 \leq i \leq n} EU_i^2 I(|u_i| \leq A) (1 + h^2 \sqrt{t} \log t + h \sqrt{t}) \leq C_1 A^{-2} h \sqrt{n}, \]

\[ EY_{2nt}^2 \leq C \sup_{1 \leq i \leq n} EU_i^2 I(|u_i| > A) (1 + h^2 \sqrt{t} \log t + h \sqrt{t}) \leq C_1 A^{-2} h \sqrt{n}. \]

These results, together with the fact that u_{1j} and u_{2j} both are martingale difference satisfying

\[ \sup_j E(u_{1,j+1}^2 | \mathcal{F}_j) \leq \sup_j [E(u_j^4 | \mathcal{F}_j)]^{1/2} \leq C, \]

\[ \sup_j E(u_{2,j+1}^2 | \mathcal{F}_j) \leq \sup_j E(u_j^2 I_{|u_j| > A} | \mathcal{F}_j) \]

\[ \leq A^{-2} \sup_j E(u_j^4 | \mathcal{F}_j) \leq CA^{-2}, \]

yield that, as n → ∞ first, and then A → ∞,

\[ E\Lambda_{1n}^2 \leq \frac{C}{n^{3/2} h} \sum_{t=2}^{n} EY_{1nt}^2 \leq CA^{-2} \rightarrow 0, \]

\[ E\Lambda_{2n}^2 \leq \frac{CA^{-2}}{n^{3/2} h} \sum_{t=2}^{n} EY_{nt}^2 \leq CA^{-2} \rightarrow 0, \]

\[ E\Lambda_{4n} \leq \frac{C}{n^{3/2} h} \sum_{t=2}^{n} EY_{2nt}^2 \leq CA^{-2} \rightarrow 0, \]

\[ E|\Lambda_{3n}| \leq \frac{C}{n^{3/2} h} \sum_{t=2}^{n} (EY_{1nt}^2)^{1/2} (EY_{2nt}^2)^{1/2} \leq CA^{-1} \rightarrow 0. \]

This proves (6.49), and hence the proof of Theorem 3.3 is complete.

**Acknowledgments.** Our thanks to the Editor, Associate Editor and referees for helpful comments on earlier versions.

**SUPPLEMENTARY MATERIAL**

**Supplement to "A specification test for nonlinear nonstationary models"** (DOI: 10.1214/12-AOS975SUPP; .pdf). Further details on the derivations in the present paper and supporting lemmas and proofs of the main results.
on convergence to intersection local time are contained in the supplement to the paper, Wang and Phillips (2012).

REFERENCES

AKONOM, J. (1993). Comportement asymptotique du temps d’occupation du processus des sommes partielles. Ann. Inst. Henri Poincaré Probab. Stat. 29 57–81. MR1204518

ALDOUS, D. J. (1986). Self-intersections of 1-dimensional random walks. Probab. Theory Related Fields 72 559–587. MR0847386

BANDI, F. M. and PHILLIPS, P. C. B. (2003). Fully nonparametric estimation of scalar diffusion models. Econometrica 71 241–283. MR1956859

BANDI, F. M. and PHILLIPS, P. C. B. (2007). A simple approach to the parametric estimation of potentially nonstationary diffusions. J. Econometrics 137 354–395. MR2354949

BERKES, I. and HORVÁTH, L. (2006). Convergence of integral functionals of stochastic processes. Econometric Theory 22 304–322. MR2290391

BORODIN, A. N. and IBRAGIMOV, I. A. (1994). Limit theorems for functionals of random walks. Tr. Mat. Inst. Steklova 195 286. MR1368394

BUCHMANN, B. and CHAN, N. H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. Ann. Statist. 35 2001–2017. MR2363961

CHAN, N. H. and WEI, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. Ann. Statist. 15 1050–1063. MR0902245

CHOI, I. and SAIKKONEN, P. (2004). Testing linearity in cointegrating smooth transition regressions. Econom. J. 7 341–365. MR2103419

CHOI, I. and SAIKKONEN, P. (2010). Tests for nonlinear cointegration. Econometric Theory 26 682–709. MR2646476

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). Strong Approximations in Probability and Statistics. Academic Press, New York. MR0666546

DE JONG, R. and WANG, C.-H. (2005). Further results on the asymptotics for nonlinear transformations of integrated time series. Econometric Theory 21 413–430. MR2179544

GAO, J. (2007). Nonlinear Time Series: Semiparametric and Nonparametric Methods. Monographs on Statistics and Applied Probability 108. Chapman & Hall/CRC, Boca Raton, FL. MR2297190

GAO, J., KING, M., LU, Z. and TJÓSTHEIM, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. Econometric Theory 25 1869–1892. MR2557585

GAO, J., KING, M., LU, Z. and TJÓSTHEIM, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. Ann. Statist. 37 3893–3928. MR2572447

HALL, P. and HEYDE, C. C. (1980). Martingale Limit Theory and Its Application. Academic Press, New York. MR0624435

HONG, S. H. and PHILLIPS, P. C. B. (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. J. Bus. Econom. Statist. 28 96–114. MR2650603

HOROWITZ, J. L. and SPOKOINY, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69 599–631. MR1828537

JEGANATHAN, P. (2004). Convergence of functionals of sums of r.v.s to local times of fractional stable motions. Ann. Probab. 32 1771–1795. MR2073177

KARLSEN, H. A., MYKLEBUST, T. and TJÓSTHEIM, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. Ann. Statist. 35 252–299. MR2332276
KARLSEN, H. A. and TJOSTHEIM, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29** 372–416. MR1863963

KASPARIS, I. and PHILLIPS, P. C. B. (2012). Dynamic misspecification in nonparametric cointegrating regression. *J. Econom.* To appear.

LI, Q. and WANG, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econom.* **87** 145–165. MR1648892

MARMER, V. (2008). Nonlinearity, nonstationarity, and spurious forecasts. *J. Econometrics* **142** 1–27. MR2408730

PARK, J. Y. and PHILLIPS, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* **15** 269–298. MR1704225

PARK, J. Y. and PHILLIPS, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69** 117–161. MR1806536

PHILLIPS, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547. MR0909357

PHILLIPS, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica* **56** 1021–1043. MR0961417

PHILLIPS, P. C. B. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory* **25** 1466–1497. MR2557571

PHILLIPS, P. C. B. and PARK, J. Y. (1998). Nonstationary density estimation and kernel autoregression. Discussion Paper 1181, Cowles Foundation, Yale Univ.

PHILLIPS, P. C. B. and SOLO, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20** 971–1001. MR1165602

REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **293**. Springer, Berlin. MR1725357

SCHIENLE, M. (2008). Nonparametric nonstationary regression. Doctoral thesis, Mannheim Univ.

VAN DER HOFSTAD, R., DEN HOLLANDER, F. and KÖNIG, W. (1997). Central limit theorem for the Edwards model. *Ann. Probab.* **25** 573–597. MR1434119

VAN DER HOFSTAD, R., DEN HOLLANDER, F. and KÖNIG, W. (2003). Weak interaction limits for one-dimensional random polymers. *Probab. Theory Related Fields* **125** 483–521. MR1974412

VAN DER HOFSTAD, R. and KÖNIG, W. (2001). A survey of one-dimensional random polymers. *J. Stat. Phys.* **103** 915–944. MR1851362

WANG, Q. and PHILLIPS, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25** 710–738. MR2507529

WANG, Q. and PHILLIPS, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica* **77** 1901–1948. MR2573873

WANG, Q. and PHILLIPS, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory* **27** 235–259. MR2782038

WANG, Q. and PHILLIPS, P. C. B. (2012). Supplement to “A specification test for nonlinear nonstationary models.” DOI:10.1214/12-AOS975SUPP.