Econometric Theory, **25**, 2009, 1466–1497. Printed in the United States of America.
doi:10.1017/S0266466609990223

LOCAL LIMIT THEORY AND SPURIOUS NONPARAMETRIC REGRESSION

PETER C.B. PHILLIPS
Yale University
University of Auckland
University of York

and

Singapore Management University

A local limit theorem is proved for sample covariances of nonstationary time series and integrable functions of such time series that involve a bandwidth sequence. The resulting theory enables an asymptotic development of nonparametric regression with integrated or fractionally integrated processes that includes the important practical case of spurious regressions. Some local regression diagnostics are suggested for forensic analysis of such regresssions, including a local R^2 and a local Durbin–Watson (DW) ratio, and their asymptotic behavior is investigated. The most immediate findings extend the earlier work on linear spurious regression (Phillips, 1986, *Journal of Econometrics* 33, 311–340) showing that the key behavioral characteristics of statistical significance, low DW ratios and moderate to high R^2 continue to apply locally in nonparametric spurious regression. Some further applications of the limit theory to models of nonlinear functional relations and cointegrating regressions are given. The methods are also shown to be applicable in partial linear semiparametric nonstationary regression.

1. INTRODUCTION

In a now-famous simulation experiment involving linear regressions of independent random walks and integrated processes, Granger and Newbold (1974) showed some of the key features of a spurious regression—spuriously significant coefficients, moderate to high R^2, and low Durbin–Watson ratios—and argued that such phenomena were widespread in applied economics. Of course, concerns in economics over the potential for spurious and nonsense correlations

Some of the results given here were first presented at the Conference in Honour of Paul Newbold at Nottingham University, September 2007. Thanks go to three anonymous referees, Jiti Gao, and Yixiao Sun for helpful comments and to the NSF for partial research support under grant SES 06-47086. Address correspondence to Peter C.B. Phillips, Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520–8268, USA; e-mail: peterphillips@yale.edu.

  0266-4666/09 $15.00

in empirical work go back much further, at least to Hooker (1905), "Student" (1914), Yule (1921), and Fisher (1907, 1930). The first systematic study of spurious relations in time series was undertaken by Yule (1926) in an important contribution that revealed some of the dangers of regressing variables with trends. But, following Yule's study, the subject fell into dormancy for some decades with no further attempts at formal analysis.

The Granger and Newbold paper brought the subject to life again with some startling simulation findings on stochastic trends that quickly attracted attention. The paper itself was only 10 pages long and soon became accepted as a cautionary tale in econometrics, warning against the uncritical use of level regressions for trending economic variables. Looking back now, the simulation experiment reported in the paper seems tiny by modern standards, with only 100 replications and two small tabulations of results. In addition to its simulations, the paper contained some recommendations for and warnings to applied researchers concerning the conduct of empirical research with time series data and the use of formulations in differences rather than levels for such regressions.[1] These recommendations were taken seriously in applied work and were incorporated into econometrics teaching, at least until the mid-1980s, when the concept of cointegration and the methodology of unit root/cointegration testing exploded conventional thinking in the profession about time series regressions in levels and led to formal analytical procedures for evaluating the presence of levels and differences in time series regression equations.

Phillips (1986) initiated the asymptotic analysis of spurious regressions by utilizing function space central limit theory, giving the first implementation of that limit theory to regression problems in econometrics and providing a formal apparatus of analysis. The approach revealed the limit behavior of the regression coefficients, significance tests, and regression diagnostics and confirmed that the simulation findings in Granger and Newbold (1974) accorded well with the new limit theory. A later paper, Phillips (1998), gave a deeper explanation of the limit theory and simulation findings, proving that the fitted regressions estimated (and in the limit accurately reproduced) a finite number of terms in the formal mathematical series representation of the limit process to which the (suitably normalized) dependent variable converged. This result validated a formal interpretation of such fitted (spurious) regressions as coordinate regression systems that capture the trending behavior of one variable in terms of the trends that appear in other variables. The coordinate approach was investigated more systematically in Phillips (2005a).

The present paper extends the asymptotic analysis of Phillips (1986) to a nonparametric regression setting. To develop nonparametric regression asymptotics, a local limit theorem is provided for sample cross moments of a nonstationary time series and integrable functions of another such time series. The theory allows for the presence of kernel functions and bandwidth parameter sequences. The approach taken in this local limit theory draws on recent work of Wang and Phillips (2009a) dealing with nonparametric cointegrating regression, although the results here relate to spurious regression phenomena and are therefore different in character and involve some technical modification of the methods.

The linear spurious regression asymptotics in Phillips (1986) have the simple interpretation of an L_2 regression involving the trajectories of the limiting stochastic processes corresponding to the variables in the original regression, at least after some suitable standardization. A similar interpretation is shown here to apply in the case of nonparametric regression. In the present case, the limiting form of the nonparametric regression at some point x (in the space of the regressor) is simply a weighted average of the trajectory of the limiting stochastic process corresponding to the dependent variable, where the average of the dependent variable is taken only over those time points for which the limiting stochastic process of the regressor variable happens to be in the immediate locality of x. Accordingly, the limit theory in the present paper integrally involves the concept of the local time of a stochastic process, a quantity that directly measures the time spent by a process around a particular value. As is shown here, nonparametric spurious regression asymptotics correspond to a weighted L_2 regression of the limit process of the dependent variable with weights delivered by the local time of the regressor in the locality of x. In effect, the limit is just a continuous time nonparametric kernel regression.

Figure 1a shows a cross plot of (y_t, x_t) coordinates corresponding to 500 observations of two independently drawn Gaussian random walks shown in Figure 1b for y_t and x_t originating at the origin and having standard normal increments. The cloud of points in the figure shows a pattern where y appears to increase for some values of x and decrease for others. Patterns in the data are typical in such cases when a finite number of draws of independent random walks are taken. The specific pattern depends, of course, on the actual time series evolution of the processes. The particular pattern shown in Figure 1a is much more sympathetic to broken trend modeling and nonparametric fitting than it is to linear regression, as the kernel regression fit shown in the figure indicates. Again, this is fairly typical with random walk data. Accordingly, the potential opportunities for spurious trend break regression and nonparametric fitting with such unrelated time series are considerable. One object of the present paper is to explore such phenomena and provide new analytic machinery for studying such nonparametric regressions with nonstationary data.

The rest of the paper is organized as follows. Section 2 provides some heuristic analysis and formal discussion using array limits that avoid some of the main technical difficulties of the limit theory while revealing the main results. Sections 3, 4, and 5 give the main results on the limit theory, its application to nonparametric spurious regression, and some asymptotics for local regression diagnostics. The latter include some new theory on local R^2 and local Durbin–Watson statistics. Section 6 concludes and outlines some further uses of the limit theory and approach given here. Proofs and related technical results are provided in the Appendix.

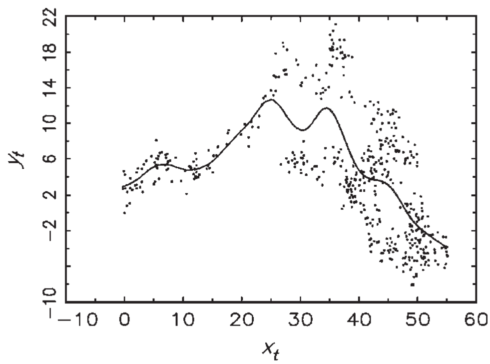

FIGURE 1a. Scatter plot and Nadaraya–Watson nonparametric regression.

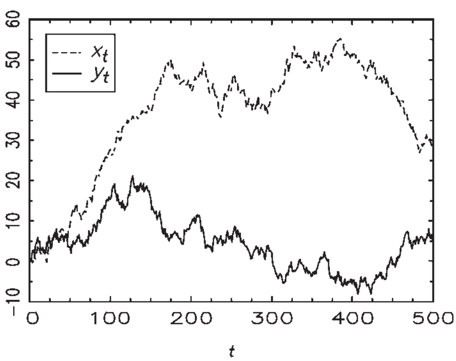

FIGURE 1b. Two independent random walks.

2. HEURISTICS

To motivate and interpret some key results in the paper, this section provides heuristic explanations of the limit theory. The simple derivations given here involve sequential limit arguments that avoid many of the technical complications dealt with later that arise in kernel asymptotics for nonstationary time series.

The object of interest in nonparametric regression typically involves two triangular arrays $(y_{k,n}, x_{k,n}), 1 \le k \le n, n \ge 1$ constructed by standardizing some underlying time series. We assume that there are continuous limiting Gaussian processes $(G_y(t), G_x(t)), 0 \le t \le 1$, for which we have the joint convergence

$$(y_{[nt],n}, x_{[nt],n}) \Rightarrow \left(G_y(t), G_x(t)\right), \tag{1}$$

where $[a]$ denotes the integer part of a and $\Rightarrow$ denotes weak convergence. This framework will include most nonstationary data cases of interest, including integrated and fractionally integrated time series. The main functional of interest, S_n, in the present paper is defined by the sample covariance

$$S_n = \frac{c_n}{n} \sum_{k=1}^{n} y_{t,n} g(c_n x_{k,n}), \tag{2}$$

where c_n is a certain sequence of positive constants and g is a real function on R. The limit behavior of S_n when both $c_n \to \infty$ and $n/c_n \to \infty$ is particularly interesting and important for practical applications as it provides a setting where the sample function depends on both a primary sequence (n) and a secondary sequence (c_n) that both tend to infinity. This formulation is particularly convenient in situations like kernel regression where a bandwidth parameter (h_n) is involved and whose asymptotic behavior ($h_n \to 0$) needs to be accounted for in the analysis. The form of S_n in (2) accommodates a sufficiently wide range of bandwidth choices to be relevant for nonparametric kernel estimation. In most applications the bandwidth arises in a very simple manner and is embedded in the secondary parameter sequence c_n, for instance, as in $c_n = \sqrt{n}/h_n$.

Accordingly, 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 processes $y_{t,n}$ and $x_{t,n}$

$$S_n \Rightarrow \int_{-\infty}^{\infty} g(s)\,ds \int_0^1 G_y(p)\,dL_{G_x}(p,0), \tag{3}$$

where $L_{G_x}(p,s)$ is the local time (defined in Section 3 in (9)) of the process $G_x(t)$ at the spatial point s. When the function g is a kernel density, the "energy" functional $\int_{-\infty}^{\infty} g(s)\,ds = 1$, and the limit (3) is then an average of the limit process G_y taken with respect to the local time measure of G_x at the origin. This result relates to work by Jeganathan (2004) and Wang and Phillips (2009a), who investigated the asymptotic form of similar sample mean functionals involving only a single array $x_{k,n}$. Some other related works that involve limit theory with local time limits can be found in Akonom (1993), Borodin and Ibragimov (1995), Phillips and Park (1998), and Park and Phillips (1999, 2000). Another approach to developing a limit theory for sample functions involving kernel densities has been developed by Karlsen, Myklebust, and Tjøstheim (2007) using null recurrent Markov chain methods. Most recently, Wang and Phillips (2009b) have used local time limit theory techniques to study structural nonparametric cointegrating regression.

A typical example of S_n in the econometric applications that we consider later has the form of a sample cross moment of one variable (y_t) with a kernel function ($K(\cdot)$) of another variable (x_t). This sample moment may be written in standardized form corresponding to (2) as

$$S_n = \frac{1}{nh_n}\sum_{t=1}^{n} y_t K\left(\frac{x_t - x}{h_n}\right) = \frac{1}{\sqrt{n}h_n}\sum_{t=1}^{n}\frac{y_t}{\sqrt{n}}K\left(\frac{\sqrt{n}\left(\frac{x_t}{\sqrt{n}} - \frac{x}{\sqrt{n}}\right)}{h_n}\right)$$

$$= \frac{c_n}{n}\sum_{t=1}^{n} y_{t,n} K\left(c_n\left(x_{t,n} - \frac{x}{\sqrt{n}}\right)\right),$$

with $c_n = \sqrt{n}/h_n$, $y_{t,n} = y_t/\sqrt{n}$, $x_{t,n} = x_t/\sqrt{n}$, and where h_n is the bandwidth parameter. When $(y_{[n\cdot],n}, x_{[n\cdot],n}) \Rightarrow (B_y(\cdot), B_x(\cdot))$, so the limit processes are Brownian motions, and when $\int_{-\infty}^{\infty} K(s)\,ds = 1$, the limit behavior of S_n for fixed x is given by

$$S_n \Rightarrow \int_0^1 B_y(p)\,dL_{B_x}(p,0). \tag{4}$$

This limit is simply the average value of the trajectory of the limit Brownian motion $B_y(p)$ taken over time points $p \in [0,1]$ where the limit process B_x sojourns around the origin. Result (4) and its various extensions turn out to play an important role in kernel regression asymptotics with nonstationary series.

The limit distribution of S_n in the situation where c_n is fixed as $n \to \infty$ is very different from that when $c_n \to \infty$ and $n/c_n \to \infty$. For example, when $c_n = 1$, it is well known that

$$\frac{1}{n}\sum_{k=1}^{n} y_{k,n} g(x_{k,n}) \Rightarrow \int_0^1 G_y(t) g(G_x(t))\,dt \tag{5}$$

by virtue of weak convergence and continuous mapping under rather weak conditions on the function g. Various results related to (5) are well known—see Park and Phillips (1999), de Jong (2004), Pötscher (2004), de Jong and Wang (2005), and Berkes and Horváth (2006). However, when $c_n \to \infty$, not only is the limit result different, but the rate of convergence is affected, the limit theory is much harder to prove, and the final result no longer has a form that is directly associated with a continuous map.

The following heuristic arguments help to reveal the nature of these differences. Note first that by virtue of the extended occupation times formula (see (11) in Section 3), limits of the form given in (5) may also be written as

$$\int_0^1 G_y(p) g\left(G_x(p)\right) dp = \int_{-\infty}^{\infty} g(a) \int_0^1 G_y(p)\,dL_{G_x}(p,a), \tag{6}$$

where $L_{G_x}(p,a)$ is the local time at a of the limit process G_x over the time interval $[0, p]$, as discussed in Section 3. Because the process $L_{G_x}(p,a)$ is continuous and increasing in the argument p, the integral $\int_0^r g(a)\,dL_{G_x}(p,a)$ is a conventional Lebesgue–Stieltjes integral with respect to the local time measure $dL_{G_x}(p,a)$.

Next, rewrite the average S_n so that it is indexed by twin sequences c_m and n, defining

$$S_{m,n} = \frac{c_m}{n} \sum_{k=1}^{n} y_{t,n} g(c_m x_{k,n}) \tag{7}$$

and noting that $S_{m,n} = S_n$ when $m = n$. Thus, the limit of S_n is the diagonal limit of the multidimensional sequence $S_{m,n}$. The limit may also be obtained in a simple manner using sequential convergence methods. In particular, if we first hold c_m fixed as $n \to \infty$ and then pass $m \to \infty$, we have from (5)–(6)

$$\begin{aligned} S_{m,n} &\Rightarrow c_m \int_0^1 G_y(t) g(c_m G_x(t))\, dt \quad \text{as } n \to \infty, \\ &= c_m \int_{-\infty}^{\infty} g(c_m s) \int_0^1 G_y(p)\, dL_{G_x}(p,s)\, ds \\ &= \int_{-\infty}^{\infty} g(a) \int_0^1 G_y(p)\, dL_{G_x}\left(p, \frac{a}{c_m}\right) da := S_{m,\infty} \\ &\Rightarrow \int_{-\infty}^{\infty} g(a)\, da \int_0^1 G_y(p)\, dL_{G_x}(p,0), \quad \text{as } m \to \infty. \end{aligned} \tag{8}$$

It follows that (8) may be regarded as a certain limiting version of S_n in terms of the sequential limits $S_{m,n} \Rightarrow S_{m,\infty} \Rightarrow S_{\infty,\infty}$. The goal is to turn this sequential argument into a joint limit argument so that c_n may play an active role as a sequence involving a bandwidth parameter, thereby including functionals that arise in density estimation and kernel regression.

Observe that the limit (8) involves the local time process $L_{G_x}(p,0)$ where the origin is the relevant spatial point. An extended version of (8) involving different localities for G_x arises for functionals of the type

$$S_{m,n} = \frac{c_m}{n} \sum_{t=1}^{n} y_{t,n} g(c_m (x_{t,n} - a)),$$

where the sequence $x_{t,n}$ is recentered about a. Correspondingly, we then have in the same manner as (8)

$$\begin{aligned} S_{m,n} &\Rightarrow c_m \int_0^1 G_y(p) g\left(c_m (G_x(p) - a)\right) dp \\ &= c_m \int_{-\infty}^{\infty} g\left(c_m (b-a)\right) \int_0^1 G_y(p)\, dL_{G_x}(p,b)\, db \\ &= \int_{-\infty}^{\infty} g(s) \int_0^1 G_y(p)\, dL_{G_x}\left(p, \frac{s}{c_m} + a\right) ds, \quad \text{using } s = c_m(b-a) \\ &\Rightarrow \int_{-\infty}^{\infty} g(s)\, ds \int_0^1 G_y(p)\, dL_{G_x}(p,a), \end{aligned}$$

where the local time process $L_{G_x}(p,a)$ is now evaluated at the spatial point a.

We next proceed to make these results rigorous in terms of direct limits as $n \to \infty$, corresponding to the diagonal sequence $S_n = S_{n,n}$.

3. LOCAL LIMIT THEORY

The local time $\{L_\zeta(t,s), t \geq 0, s \in R\}$ of a measurable stochastic process $\{\zeta(t), t \geq 0\}$ is defined as

$$L_\zeta(t,s) = \lim_{\varepsilon \to 0} (1/2\varepsilon) \int_0^t 1\{|\zeta(r) - s| < \varepsilon\}\, dr. \tag{9}$$

The two-dimensional process $L_\zeta(t,s)$ is a spatial density that records the relative time that the process $\zeta(t)$ sojourns at the spatial point s over the time interval $[0,t]$. For any locally integrable function $T(x)$, the equation

$$\int_0^t T[\zeta(r)]\, dr = \int_{-\infty}^{\infty} T(s) L_\zeta(t,s)\, ds \tag{10}$$

holds with probability one and is known as the occupation times formula. An extended version of the occupation times formula (10) that is useful in our development in this paper takes the form

$$\int_0^t T[r, \zeta(r)]\, dr = \int_{-\infty}^{\infty} ds \int_0^t T(r,s)\, dL_\zeta(r,s) \tag{11}$$

(see Revuz and Yor, 1999, p. 232). For further discussion, existence theorems, and properties of local time processes we refer to Geman and Horowitz (1980), Karatzas and Shreve (1991), and Revuz and Yor (1999). Phillips (2001, 2005b), Park and Phillips (2001), and Park (2006) provide various economic applications and empirical implementations of local time and associated hazard functions.

As in Section 2, let $x_{t,n}$ and $y_{t,n}$ for $0 \leq t \leq n, n \geq 1$ (define $x_{0,n} \equiv 0$ and $y_{0,n} \equiv 0$) be random triangular arrays and let $g(x)$ be a real measurable function on R. We make the following assumptions and use the notation

$$\Omega_n(\eta) = \{(l,k) : \eta n \leq k \leq (1-\eta)n,\ k + \eta n \leq l \leq n\},$$

where $0 < \eta < 1$, following Wang and Phillips (2009a).

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

Assumption 2.2. There exist stochastic processes $(G_x(t), G_y(t))$ for which the weak convergence $\left(x_{[nt],n}, y_{[nt],n}\right) \Rightarrow \left(G_x(t), G_y(t)\right)$ holds with respect to the Skorokhod topology on $D[0,1]^2$. The process $G_x(t)$ has continuous local time $L_G(t,s)$.

Assumption 2.2*. On a suitable probability space $(\Omega, \mathcal{F}, P)$ there exists a stochastic process $G(t)$ for which $\sup_{0 \leq t \leq 1} \left| z_{[nt],n} - G(t) \right| = o_p(1)$ where $z_{t,n} = (x_{t,n}, y_{t,n})$ and $G(t) = (G_x(t), G_y(t))$.

Assumption 2.3. For all $0 \leq k < l \leq n, n \geq 1$, there exist a sequence of (re)standardizing 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 $C > 0$, $\inf_{(l,k)\in\Omega_n(\eta)} d_{l,k,n} \geq \eta^{m_0}/C$ as $n \to \infty$;

(b) $x_{k,n}$ are adapted to $\mathcal{F}_{k,n}$ and, conditional on $y_{k,n}$ and $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a density $h^x_{l,k,n}(x|y)$ that is bounded by a constant for all x and y, $1 \leq k < l \leq n$ and $n \geq 1$, and

$$\sup_{(l,k)\in\Omega_n[\delta^{1/(2m_0)}]} \sup_{|u|\leq\delta} \sup_{y} \left| h^x_{l,k,n}(u|y) - h^x_{l,k,n}(0|y) \right| = o_P(1), \tag{12}$$

when $n \to \infty$ first and then $\delta \to 0$.

(c) $y_{k,n}/d_{k,0,n}$ has a density $h^y_{k,0,n}(y)$ that satisfies $|h^y_{k,0,n}(y)| \leq h(y)$ for some $h(y)$ for which $\int_{-\infty}^{\infty} y^2 h(y)\,dy < \infty$.

As discussed in Wang and Phillips (2009a), Assumptions 2.1 and 2.2 are quite weak and likely to be close to necessary conditions for this kind of problem. Assumption 2.2 involves a joint convergence condition on the process $z_{t,n}$, whereas Wang and Phillips (2009a) place the convergence condition solely on $x_{t,n}$ because in the context of an explicit (cointegrating) regression model the properties of the other variable follow directly from the model.

As for Assumption 2.3, we may choose $\mathcal{F}_{k,n} = \sigma(x_{1,n}, \ldots, x_{k,n})$, the natural filtration for $x_{t,n}$, and the numerical sequence $d_{l,k,n}$ is typically chosen as a standardizing sequence so that, conditional on $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a limit distribution as $l - k \to \infty$. For instance, if $x_{i,n} = \sum_{j=1}^{i} \epsilon_j/\sqrt{n}$, where ϵ_j are independent and identically distributed 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 $d_{l,k,n} = \sqrt{l-k}/\sqrt{n}$. Assumption 2.3(b) requires the existence and boundedness of the conditional densities $h^x_{l,k,n}(x|y)$. This assumption is very convenient in technical arguments. As shown in Corollary 2.2 of Wang and Phillips (2009a), Assumption 2.3(b) holds when $x_{k,n}$ is a standardized partial sum of a linear process under weak summability conditions on the coefficients and with i.i.d. innovations whose characteristic function is integrable, without conditioning on a secondary sequence $y_{k,n}$. Obviously, Assumption 2.3(b) holds in precisely the same framework when $y_{k,n}$ is an independent process, and extension of those conditions to a multivariate linear process seems relatively innocuous. On the other hand, assuming the existence of the conditional densities of $(x_{l,n} - x_{k,n})/d_{l,k,n}$ rules out cases where the constituent variables are discrete. Although it seems likely that the results may hold under a weakening of the assumption to allow for such cases, this has not been proved. Assumption 2.3(c) is a simple dominating second moment condition on the density of $y_{k,n}/d_{k,0,n}$. Again, this seems like a reasonably mild requirement and may be strengthened further when higher order sample moments are involved, such as in

$$S_n = \frac{c_n}{n} \sum_{k=1}^{n} y_{t,n}^p g(c_n x_{k,n}),$$

for some integer $p > 2$.

Assumption 2.3 presumes that the same (re)standardizing sequence $d_{l,k,n}$ applies for both $x_{t,n}$ and $y_{t,n}$, which helps to simplify the conditions and the proof of our main results. This assumption will be sufficient for our purpose in the present paper and will often be satisfied because the observed time series x_t and y_t have similar generating mechanisms. However, we can also allow for individual specific (re)standardizing constants (say, $d_{l,k,n}^x, d_{l,k,n}^y$). With some modification of the statement and proof of the result and under some further conditions on $(d_{l,k,n}^x, d_{l,k,n}^y)$, the limit theory given subsequently in Theorem 1 can be shown to continue to hold. But a full extension along these lines is not needed for the present paper.

The main result needed to develop a regression theory in the present case involves sample covariances between $y_{t,n}$ and integrable functions of the scaled versions $c_n x_{t,n}$ of $x_{t,n}$. The latter are designed to include kernel functions whose bandwidth sequences are embodied in the sequence c_n.

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

$$\frac{c_n}{n} \sum_{t=1}^{[nr]} y_{t,n} g\left(c_n x_{t,n}\right) \Rightarrow \tau \int_0^r G_y(p)\, dL_{G_x}(p, 0). \tag{13}$$

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

$$\sup_{0 \le r \le 1} \left| \frac{c_n}{n} \sum_{t=1}^{[nr]} y_{t,n} g\left(c_n x_{t,n}\right) - \tau \int_0^r G_y(p)\, dL_{G_x}(p, 0) \right| \to_P 0, \tag{14}$$

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

Remarks.

(a) For $a \neq 0$, we have the following useful extension of (13):

$$\frac{c_n}{n} \sum_{t=1}^{[nr]} y_{t,n} g\left(c_n \left(x_{t,n} - a\right)\right) \to_D \tau \int_0^r G_y(p)\, dL_{G_x}(p, a), \tag{15}$$

which gives the limit behavior of the sample moment when $x_{t,n}$ is in the neighborhood of some point a. The limit (15) is expressed in terms of an integral of G_y with respect to the local time measure of the limit process G_x around a. The proof of (15) follows in precisely the same way as (13).

(b) Higher order sample moments have similar limit behavior under suitable integrability conditions in place of Assumption 2.3(c). For instance,

$$\frac{c_n}{n}\sum_{t=1}^{[nr]} y_{t,n}^2 g\left(c_n\left(x_{t,n}-a\right)\right) \Rightarrow \tau\int_0^r G_y^2(p)\,dL_{G_x}(p,a), \tag{16}$$

and, more generally, for any locally integrable function f

$$\frac{c_n}{n}\sum_{t=1}^{[nr]} f\left(y_{t,n}\right) g\left(c_n\left(x_{t,n}-a\right)\right) \Rightarrow \tau\int_0^r f\left(G_y(p)\right) dL_{G_x}(p,a). \tag{17}$$

Then, for the constant function $f\left(y_{t,n}\right)=1$, we have the scaled local time result

$$\frac{c_n}{n}\sum_{t=1}^{[nr]} g\left(c_n\left(x_{t,n}-a\right)\right) \to \tau\int_0^r 1\,dL_{G_x}(p,a)=\tau L_{G_x}(r,a), \tag{18}$$

given earlier in Wang and Phillips (2009a). Again, these results may be established in the same way as (13).

(c) Theorem 1 has quite extensive applications in econometrics that include spurious nonparametric regressions, nonparametric cointegrated regression models, and parametric cointegrated regressions. The next section provides a detailed study of the spurious nonparametric regression application, and later work will consider other applications. Also included in the range of applications are cases where a functional relationship may exist between the limit processes, such as $G_y(t)=f\left(G_x(t)\right)$. We may then write the limit in (13) as

$$\tau\int_0^r G_y(p)\,dL_{G_x}(p,0)=\tau\int_0^r f\left(G_x(p)\right) dL_{G_x}(p,0)=\tau f(0)\,L_{G_x}(r,0).$$

When $x_{t,n}\sim a$ as in (15), we end up with the corresponding limit

$$\tau\int_0^r G_y(p)\,dL_{G_x}(p,a)=\tau\int_0^r f\left(G_x(p)\right) dL_{G_x}(p,a)=\tau f(a)L_{G_x}(r,a).$$

Of course, when x_t and y_t are cointegrated $I(1)$ or $I(d)$ processes, we have a simple linear relationship between the Gaussian limit processes of the form $G_y(t)=\beta G_x(t)$ for some fixed parameter β. In that case, the limit result (15) gives

$$\tau\int_0^r G_y(p)\,dL_{G_x}(p,a)=\tau\beta a L_{G_x}(r,a). \tag{19}$$

Combining (19) with (18), we get the following limit of the nonparametric (cointegrating) regression function:

$$\frac{\tau\beta a L_{G_x}(r,a)}{\tau L_{G_x}(r,a)}=\beta a,$$

which reproduces the linear cointegrating relationship in neighborhoods where $G_x(t) \sim a$. That case was studied by Wang and Phillips (2009a), who also provided the limit distribution of the kernel estimate.

4. NONPARAMETRIC SPURIOUS REGRESSION

Suppose (y_t, x_t) satisfy Assumptions 2.2–2.3 and the nonparametric regression

$$y_t = \hat{g}(x_t) + \hat{v}_t \tag{20}$$

is performed, where $\hat{g}$ is the Nadaraya–Watson kernel estimate

$$\hat{g}(x) = \frac{\sum_{t=1}^{n} y_t K\left(\frac{x_t - x}{h_n}\right)}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)} = \arg\min_{g} \sum_{t=1}^{n} (y_t - g)^2 K_{h_n}(x_t - x)$$

for some kernel function K, with $K_h(s) = (1/h)K(\frac{s}{h})$ and with bandwidth parameter $h = h_n$. We assume that K satisfies the following condition and $h_n \to 0$ as $n \to \infty$.

Assumption 3.1. The kernel K is a nonnegative real function for which $\int_{-\infty}^{\infty} K(s)\,ds = 1$, $\int_{-\infty}^{\infty} K(s)^2\,ds < \infty$, and $\sup_s K(s) < \infty$.

Let d_n be a standardizing sequence for which $d_n \to \infty$ as $n \to \infty$ and $y_{t,n} = d_n^{-1} y_t$ and $x_{t,n} = d_n^{-1} x_t$. For example, when both x_t and y_t are $I(1)$ time series we have $d_n = \sqrt{n}$. Set $c_n = d_n/h_n$ and assume that $c_n/n \to 0$, which requires that $nh_n/d_n \to \infty$, so that h_n should not go to zero too fast. Also, $c_n \to \infty$ requires that h_n be of lower order than d_n.

THEOREM 2. *Suppose Assumptions 2.2, 2.3, and 3.1 hold. Let d_n be a standardizing sequence for which $d_n \to \infty$ as $n \to \infty$ and for which $y_{t,n} = d_n^{-1} y_t$ and $x_{t,n} = d_n^{-1} x_t$ satisfy Assumption 2.2. Then, for any h_n satisfying $nh_n/d_n \to \infty$ and $d_n/h_n \to \infty$,*

$$d_n^{-1}\hat{g}(x) \Rightarrow \begin{cases} \frac{\int_0^1 G_y(p)\,dL_{G_x}(p,0)}{L_{G_x}(1,0)} & \textit{for fixed } x \\ \frac{\int_0^1 G_y(p)\,dL_{G_x}(p,a)}{L_{G_x}(1,a)} & \textit{for } x = d_n a, \textit{ with a fixed} \end{cases}. \tag{21}$$

Remarks.

(d) The limit (21) is the local weighted average of $G_y(p)$ taken over values of $p \in [0,1]$ where $G_x(p)$ sojourns at a. The limit may be expressed as the mean local level from a continuous time weighted regression, namely,

$$\frac{\int_0^1 G_y(p)\,dL_{G_x}(p,a)}{L_{G_x}(1,a)} = \arg\min_{\alpha} \int_0^1 \{G_y(p) - \alpha\}^2\,dL_{G_x}(p,a). \tag{22}$$

Here, the locality is determined by the weight in the spatial measure $dL_{G_x}(p,a)$, which confines attention to the locality $G_x \sim a$. The latter simply serves as the relevant timing device for the measurement of the average and may be interpreted as a continuous time kernel function.

(e) Theorem 2 implies that $\hat{g}(x) = O_p(d_n)$, and so the local level regression coefficient diverges at the rate d_n. When y_t and x_t are $I(1)$, this means that $\hat{g}(x) = O_p\left(\sqrt{n}\right)$, which corresponds to the order of magnitude of the intercept divergence in a linear spurious regression (Phillips, 1986). Thus, there is a correspondence in the limit behavior between linear and nonparametric spurious regression. This is explained by the fact that whatever regression line is fitted, a recurrent time series like y_t visits every point in the space an infinite number of times, so that the order of magnitude of the level (or intercept in a regression) is the same as that of $y_{[n\cdot]}$, namely, $O_p\left(\sqrt{n}\right)$ for an $I(1)$ series. Thus, the heuristic reasoning for divergent behavior in the regression is the same in both cases. In effect, when x_t sojourns around some level $x = d_n a$ for some a (behavior that is mimicked by the limit process $G_x(p)$ sojourning around a), y_t may be taking any value in the space (because x_t and y_t are not cointegrated), and because $y_{[n\cdot]} = O_p(d_n)$ the corresponding level of y_t is $O_p(d_n)$ also, thereby producing a local regression level that has this order asymptotically. More explicitly, suppose x_t and y_t are not cointegrated and satisfy $y_t = \beta x_t + u_t$ where $u_{t,n} = d_n^{-1} u_t$ satisfies $(x_{[n\cdot],n}, u_{[n\cdot],n}) \Rightarrow (G_x(\cdot), G_u(\cdot))$ for some nontrivial Gaussian limit process G_u. Then, as $n \to \infty$, Theorem 2 implies that

$$d_n^{-1}\hat{g}(d_n a) \Rightarrow \beta a + \frac{\int_0^1 G_u(p)\, dL_{G_x}(p,a)}{L_{G_x}(1,a)},$$

whereas in the cointegrated case $d_n^{-1}\hat{g}(d_n a) \to_p \beta a$ because $G_u(p) \equiv 0$, giving a constant limit that reproduces the local form of the cointegrating relation.

(f) If, in place of (20), we run the linear spurious regression

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{v}_t,$$

then the corresponding intercept limit theory is

$$d_n^{-1}\hat{\alpha} \Rightarrow \int_0^1 G_y - \xi \int_0^1 G_y := \eta, \tag{23}$$

where $\xi = \int_0^1 G_y G_x / \int_0^1 G_x^2$. The limit (23) is simply the intercept in a global continuous time regression of G_y on G_x over $[0,1]$, so that η and ξ satisfy

$$(\eta, \xi) = \arg\min_{a,b} \int_0^1 \left\{ G_y(p) - a - bG_x(p) \right\}^2 dp. \tag{24}$$

Thus, (22) gives a local level limit version of the global regression limit result (24).

(g) Theorem 2 does not require that $h_n \to 0$. Instead, h_n simply needs to be of lower order than d_n so that $c_n = d_n/h_n \to \infty$, and h_n should not go to zero too fast so that $n/c_n = nh_n/d_n \to \infty$, thereby satisfying the conditions of Theorem 1.

(h) Theorem 2 also covers the case where there is a functional relationship between the limit processes G_y and G_x, as discussed in Remark (c). Suppose, for example, that $G_y(p) = f(G_x(p))$ for some locally integrable function f. If the standardizing sequence for y_t is $d_{y,n}$ so that $y_{t,n} = d_{y,n}^{-1} y_t$, then as indicated earlier we have the same limit behavior as in (21), and this becomes

$$d_{y,n}^{-1}\hat{g}(x) \Rightarrow \frac{\int_0^1 G_y(p)\, dL_{G_x}(p,a)}{L_{G_x}(1,a)} = \frac{\int_0^1 f(G_x(p))\, dL_{G_x}(p,a)}{L_{G_x}(1,a)} = f(a).$$

Thus, $d_{y,n}^{-1}\hat{g}(x) \to_p f(a)$, and the nonparametric regression function correctly reproduces the functional relation between the limit processes at the spatial point a. When y_t is linearly cointegrated with x_t and of the same (possibly fractional) order, nonparametric regression at $x = d_n a$ produces $d_n^{-1}\hat{g}(x) \to_p \beta a = \beta G_x|_{G_x=a}$, thereby giving the local form of the cointegrating relationship when $G_x = a$, as already noted in Remarks (c) and (e).

5. TESTING AND DIAGNOSTICS

We start by introducing the concept of local residuals from the nonparametric regression (20). These are the residuals $\hat{\upsilon}_t$ in (20) that occur around certain points such as those where x_t is in the vicinity of $d_n a$. Local residuals are useful in developing local versions of significance tests and residual diagnostics. The latter can be used to monitor the local behavioral characteristics of a nonparametric regression. At present, there seems to be no literature on local nonparametric regression diagnostics even for stationary regression models.[2] We therefore introduce two new diagnostic statistics here: a local R^2 to measure fit and a local Durbin–Watson ratio to assess specification. These correspond to the diagnostics considered in Granger and Newbold (1974) and Phillips (1986).

Local residuals may be written as

$$\hat{\upsilon}_t\big|_{x_t\sim d_n a} = \left(y_t - \hat{g}(x_t)\right)\big|_{x_t\sim d_n a},$$

where the affix $x_t \sim d_n a$ signifies that the residuals to be taken are those that arise when x_t is in the vicinity of $d_n a$. The localization may be accomplished in the practical construction of statistics by the use of a kernel. More precisely, we define a local residual sum of squares as

$$s^2_{(d_n a)} = \frac{1}{n}\sum_{t=1}^{n} \hat{v}_t^2 K_{h_n}(x_t - d_n a), \tag{25}$$

where the kernel function now performs the localizing operation. The following lemma gives some preliminary limit theory for $s^2_{(d_n a)}$.

LEMMA A. *Under the conditions of Theorem 2 as $n \to \infty$*

$$\frac{s^2_{(d_n a)}}{d_n} \Rightarrow \int_0^1 C_a^2(r)\, dL_{G_x}(r,a), \tag{26}$$

where $C_a(r) = G_y(r) - \int_0^1 G_y(p)\, dL_{G_x}(p,a)/L_{G_x}(1,a)$.

Remarks.

(i) The limit process $C_a(r)$ in (26) is the limiting form of the standardized localized residual process $\hat{v}_{[nr]}/d_n\big|_{x_{[nr]}\sim d_n a}$. As is apparent in its form, $C_a(r)$ is simply a demeaned version of the process G_y where the mean extracted is the average level of G_y when $G_x \sim a$.

(j) The limit (26) may be rewritten as follows:

$$\begin{aligned}\int_0^1 C_a^2(r)\, dL_{G_x}(r,a) &= \int_0^1 G_y^2(r)\, dL_{G_x}(r,a) - \frac{\left\{\int_0^1 G_y(p)\, dL_{G_x}(p,a)\right\}^2}{L_{G_x}(1,a)} \\ &= \int_0^1 G_y^2(r)\, dL_{G_x}(r,a) - \int_0^1 \hat{G}_y^2(r,a)\, dL_{G_x}(r,a),\end{aligned} \tag{27}$$

which provides a decomposition of the local residual variation (or sum of squares) $\int_0^1 C_a^2(r)\, dL_{G_x}(r,a)$ into the local total variation $\int_0^1 G_y^2(r)\, dL_{G_x}(r,a)$ minus the explained local variation

$$\int_0^1 \hat{G}_y^2(r,a)\, dL_{G_x}(r,a) = \frac{\left\{\int_0^1 G_y(p)\, dL_{G_x}(p,a)\right\}^2}{L_{G_x}(1,a)}, \tag{28}$$

where $\hat{G}_y(r,a) = \int_0^1 G_y(p)\, dL_{G_x}(p,a)/L_{G_x}(1,a)$. The expression for the explained local variation $\int_0^1 \hat{G}_y^2(r,a)\, dL_{G_x}(r,a)$ is based on the continuous time local regression (22). Thus, (27) is a continuous time localized version of the usual least squares decomposition, where the localizing effect is generated through the local time measure $dL_{G_x}(r,a)$. Observe that the fitted local mean $\hat{G}_y(r,a)$ is constant in r and depends only on the spatial point a. In effect, $\hat{G}_y(r,a)$ is the predicted level of $G_y(r)$ delivered from the continuous time regression (22) when $G_x \sim a$, and this "mean" level does not depend on r.

To compute a "standard error" for $\hat{g}(x)$ at $x = \sqrt{n}a$, we assume empirical usage of the standard asymptotic variance formula in a nonparametric regression (based on stationarity assumptions). This has the usual form (e.g., Härdle and Linton, 1994):

$$s_{\hat{g}(x)}^2 = \frac{s_{(x)}^2 \mu_{K^2}}{\sum_{t=1}^n K\left(\frac{x_t - x_t}{h_n}\right)}, \qquad \mu_{K^2} = \int_{-\infty}^{\infty} K(s)^2 \, ds. \tag{29}$$

To assess local "significance" we use in (29) the local residual variance estimate $s_{(x)}^2 = n^{-1} \sum_{t=1}^n \hat{v}_t^2 K_h(x_t - x)$ in place of the sample residual second moment $n^{-1} \sum_{t=1}^n \hat{v}_t^2$. Then, as in conventional linear regression, we assess local statistical significance in terms of the t-ratio $t_{\hat{g}(x)} = \hat{g}(x)/s_{\hat{g}(x)}$.

Local regression diagnostics for (20) may be developed in a similar way. Primarily, we shall consider local R^2 and local Durbin–Watson (*DW*) ratio statistics, corresponding to the analysis of global versions of these diagnostics in linear spurious regression in Phillips (1986). Local versions of R^2 and *DW* may be defined as follows:

$$R_n^2(d_n a) = 1 - \frac{\sum_{t=1}^n \left(y_t - \hat{g}(x_t)\right)^2 K_{h_n}(x_t - d_n a)}{\sum_{t=1}^n y_t^2 K_{h_n}(x_t - d_n a)}, \tag{30}$$

$$DW_n(d_n a) = \frac{\sum_{t=1}^n \left(\Delta \hat{v}_t\right)^2 K_{h_n}(x_t - d_n a)}{\sum_{t=1}^n \hat{v}_t^2 K_{h_n}(x_t - d_n a)}. \tag{31}$$

The statistic $R_n^2(d_n a)$ measures the goodness of fit of the nonparametric regression locally around $x \sim d_n a$. The statistic $DW_n(d_n a)$ is a local variance ratio measuring the extent of local serial correlation in the residuals measured around $x \sim d_n a$. The limit theory for these local diagnostic statistics and the local nonparametric regression t-test is given in the next result.

THEOREM 3. *Under the conditions of Theorem 2, as $n \to \infty$,*

$$\frac{1}{\sqrt{nh_n}} t_{\hat{g}(d_n a)} \Rightarrow \frac{\int_0^1 G_y(p) \, dL_{G_x}(p, a)}{\left\{L_{G_x}(1, a)\right\}^{1/2} \left\{\int_0^1 C_a^2(r) \, dL_{G_x}(r, a) \mu_{K^2}\right\}^{1/2}}, \tag{32}$$

$$R_n^2(d_n a) \Rightarrow \frac{\left(\int_0^1 G_y(p) \, dL_{G_x}(p, a)\right)^2}{L_{G_x}(1, a) \int_0^1 G_y(p)^2 \, dL_{G_x}(p, a)}, \tag{33}$$

$$DW_n(d_n a) \to_p 0. \tag{34}$$

Remarks.

(k) Evidently from (32) the t-ratio diverges, so the nonparametric regression coefficient $\hat{g}(d_n a)$ will inevitably be deemed significant as $n \to \infty$, just as in the linear regression case. The divergence rate is $\sqrt{nh_n}$, which (at least when $h_n \to 0$) is slower than the divergence rate ($\sqrt{n}$) of the regression t-statistic in linear spurious regression (Phillips, 1986).

(l) The local coefficient of determination $R_n^2(d_n a)$ converges weakly to a positive random variable distributed on the interval [0, 1] because

$$\left(\int_0^1 G_y(p)\, dL_{G_x}(p,a)\right)^2 \le L_{G_x}(1,a)\int_0^1 G_y(p)^2\, dL_{G_x}(p,a),$$

by Cauchy–Schwarz. Using (27) and (28), the limit of $R_n^2(d_n a)$ can also be written in the simple format

$$\frac{\int_0^1 \hat{G}_y(p)^2\, dL_{G_x}(p,a)}{\int_0^1 G_y(p)^2\, dL_{G_x}(p,a)}$$

of the ratio of the explained local variation $\int_0^1 \hat{G}_y(p)^2\, dL_{G_x}(p,a)$ to the total local variation $\int_0^1 G_y(p)^2\, dL_{G_x}(p,a)$. This limiting ratio is the local R^2 associated with the continuous time weighted regression (22). By comparison, in a linear spurious regression of y_t on x_t the limiting form of the R^2 statistic is

$$R_n^2 \Rightarrow \frac{\left(\int_0^1 \underline{G}_y \underline{G}_x\right)^2}{\left(\int_0^1 \underline{G}_y^2\right)\left(\int_0^1 \underline{G}_x^2\right)} = \frac{\int_0^1 \hat{\underline{G}}_y^2}{\int_0^1 \underline{G}_y^2}, \tag{35}$$

where $\hat{\underline{G}}_y = \left\{(\int_0^1 \underline{G}_y \underline{G}_x)/(\int_0^1 \underline{G}_x^2)\right\}\underline{G}_x$ and $\underline{G}_y = G_y - \int_0^1 G_y$ and $\underline{G}_y = G_x - \int_0^1 G_x$ are demeaned versions of G_y and G_x. The limit (35) is the R^2 associated with the continuous time global regression (24) and is the ratio of the global explained variation to total variation in G_y.

(m) The Durbin–Watson statistic tends to zero, just as in linear spurious regression. In the present case, this behavior indicates that the serial correlation in the residuals of (20) has a dominating effect in the vicinity of every spatial realization of x_t. As in the case of linear regression, we might expect this behavior to be helpful in diagnostic analysis of the regression.

6. CONCLUSIONS AND EXTENSIONS

The present paper extends the analysis of Phillips (1986) to nonparametric regression fitting. The results show that all the usual characteristics of linear spurious regression are manifest in the context of local level regression, including divergent significance tests, local goodness of fit, and Durbin–Watson ratios converging to zero. There is therefore a need for local diagnostic procedures to assist in validating nonparametric regressions of this type. Some global tests for nonlinear cointegration have recently been developed for parametric models. For example, Hong and Phillips (2006) developed a regression equation specification error test (RESET) test for nonlinearity in cointegrating relations and Kasparis (2006) developed a cumulative sum (CUSUM) test for functional form misspecification in cointegration.

To complement procedures of this type, it would be useful to have tests for local (possibly nonlinear) cointegration in a nonparametric context. The test possibilities are vast, as in the case of linear cointegration, and deserve extensive study. For instance, local versions of the residual-based test statistics that are in common use for testing the null of no cointegration may be constructed, one example being a suitably designed modification of the local residual variance statistic (25) whose asymptotic behavior will differ according to the presence or absence of local cointegration. The detailed study of procedures such as these for validating local cointegrating behavior is left for later research.

The local limit theory given in (13) has various applications beyond those presented here. For example, in nonlinear regression models and partial linear regression models where the data are nonstationary, the limit behavior of various sample cross moments must be evaluated. These moments often involve certain integrable functions of nonstationary series and other nonstationary series, such as $n^{-1}\sum_{t=1}^{n} f(x_t)y_t$ where f satsifies Assumption 2.1 and both x_t and y_t are nonstationary. Suppose x_t and y_t are standardized by $d_n \to \infty$, so that $\left(x_{[n\cdot],n}, y_{[n\cdot],n}\right) = d_n^{-1}\left(x_{[n\cdot]}, y_{[n\cdot]}\right)$ satisfies Assumption 2.2. Then, setting $c_n = d_n$ in (13), Theorem 1 yields the following limit behavior:

$$n^{-1}\sum_{t=1}^{[nr]} f(x_t)y_t = \frac{c_n}{n}\sum_{t=1}^{[nr]} f\left(c_n x_{t,n}\right) y_{t,n} \Rightarrow \int_{-\infty}^{\infty} f(a)\,da \int_0^r G_y(p)\,dL_{G_x}(p,0). \qquad \textbf{(36)}$$

Results of this type are particularly useful in considering parametric nonlinear regressions and in developing a limit theory for partial linear cointegrating regressions.

To illustrate, suppose x_t and z_t are $I(1)$ processes and y_t is generated by the partial linear system

$$y_t = \beta' x_t + g(z_t) + v_t, \qquad \textbf{(37)}$$

where g satisfies Assumption 2.1. Model (37) is a semiparametric cointegrated regression. The behavior of y_t is dominated by the linear component $\beta' x_t$ in (37) because g attentuates the effects of large z_t. Accordingly, (37) may be regarded as a linear cointegrated system that is systematically perturbed by the presence of $g(z_t)$. The nonparametric element $g(z_t)$ has an important influence on estimation and inference even though its contribution is of a smaller order than the linear component. In particular, if we ignore the nonparametric component in (37), the least squares estimate $\hat{\beta}$ of β obtained by regressing y_t on x_t is easily seen to have the following limit theory under conventional regularity conditions (e.g., Phillips, 1988; Phillips and Solo, 1992) and using (36):

$$n\left(\hat{\beta}-\beta\right) \Rightarrow \left(\int_0^1 B_x B_x'\right)^{-1} \left\{\int_{-\infty}^{\infty} g(a)\,da \int_0^1 B_x(p)\,dL_{B_z}(p,0) + \int_0^1 B_x dB_v + \Delta_{xv}\right\}, \tag{38}$$

where B_v is the limit Brownian motion of the standardized partial sum $n^{-1/2} \sum_{t=1}^{[n\cdot]} v_t$ and $\Delta_{xv} = \sum_{h=0}^{\infty} E\left(\Delta x_t v_{t+h}\right)$. Thus, the conventional second-order bias of least squares regression is augmented by an additional bias term arising from the presence of the nonparametric element in (37) via the sample covariance $n^{-1}\sum_{t=1}^{n} g\left(z_t\right) x_t$, whose limiting form is delivered by (36). Standard cointegrating procedures like fully modified regression (Phillips and Hansen, 1990) suffer from the same second-order bias effect. These difficulties are resolved by appropriate nonparametric treatment of g in the cointegrated system (37). The details are currently under investigation and will be reported in subsequent work.

7. NOTATION

$:=$	definitional equality	$[\cdot]$	integer part
$\mathbf{1}\{\cdot\}$	indicator function	$\longrightarrow_{a.s.}$	almost sure convergence
$o_p(1)$	tends to zero in probability	$\longrightarrow_{p}$	convergence in probability
$o_{a.s.}(1)$	tends to zero almost surely	$\Longrightarrow, \rightarrow_D$	weak convergence

NOTES

1. This recommendation on the use of differences has a long history. To control for secular trend effects, Hooker (1905) originally suggested examining correlations of time differences after having earlier suggested the use of deviations from average trend (Hooker, 1901). Persons (1910) looked at various methods, including regressions in differences. "Student" (1914) made a more elaborate suggestion on the use of differences and higher order differences, which corresponds more closely to the ideas in Box and Jenkins (1970) and Box and Pierce (1970), which motivated the Granger and Newbold (1974) recommendations. To cite Yule (1921): "'Student' therefore introduces quite a new idea that is not found in any of the writers previously cited. He desires to find the correlation between x and y when every component in each of the variables is eliminated which can well be called a function of the time, and nothing is left but residuals such that the residual of a given year is uncorrelated with those that precede it or follow it." Interestingly, Yule (1921) disagreed with this particular suggestion, while at the same time being acutely aware of the spurious correlation problem.

2. For instance, standard econometric treatments (e.g., Horowitz, 1998; Pagan and Ullah, 1999; Yatchew, 2003; Li and Racine, 2007; Gao, 2007) make no mention of the idea of local diagnostic testing. After this paper was written the author discovered that Huang and Chen (2008) defined a local R^2 in a similar way to (30).

REFERENCES

Akonom, J. (1993) Comportement asymptotique du temps d'occupation du processus des sommes partielles. *Annals of the Institute of Henri Poincaré* 29, 57–81.

Berkes, I. & L. Horváth (2006) Convergence of integral functionals of stochastic processes. *Econometric Theory* 20, 627–635.

Borodin, A.N. & I.A. Ibragimov (1995) *Limit Theorems for Functionals of Random Walks*. Proceedings of the Steklov Institute of Mathematics, vol. 195, no. 2. American Mathematical Society.

Box, G.E.P. & G.M. Jenkins (1970) *Time Series Analysis, Forecasting and Control.* Holden-Day.

Box, G.E.P. & D.A. Pierce (1970) The distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.

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

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

Fisher, I. (1907) *The Rate of Interest.* Macmillan.

Fisher, I. (1930) *The Theory of Interest.* Macmillan.

Gao, J. (2007) *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman and Hall.

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

Granger, C.W.J. & P. Newbold (1974) Spurious regressions in econometrics. *Journal of Econometrics* 74, 111–120.

Härdle, W. & O. Linton (1994) Applied nonparametric methods. In D.F. McFadden & R.F. Engle III, eds., *The Handbook of Econometrics*, vol. 4, pp. 2295–2339. North–Holland.

Hong, S.-H., & P.C.B. Phillips (2006) Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity. Working paper, Yale University.

Hooker, R.H. (1901) Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society* 64, 485–492.

Hooker, R.H. (1905) On the correlation of successive observations. *Journal of the Royal Statistical Society* 68, 696–703.

Horowitz, J.L. (1998) *Semiparametric Methods in Econometrics.* Springer-Verlag.

Huang, L.-S. & J. Chen (2008) Analysis of variance, coefficient of determination, and F-test for local polynomial regression. *Annals of Statistics* 36, 2085–2109.

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.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer-Verlag.

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

Kasparis, I. (2006) Detection of Functional Form Misspecification in Cointegrating Relations: Working paper, University of Nottingham.

Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.

Park, J. (2006) Nonstationary nonlinearity: An outlook for new opportunities. In D. Corbae, S.N. Durlauf, & B.E. Hansen, eds., *Econometric Theory and Practice*, pp. 178–211. Cambridge University Press.

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

Park, J.Y. & P.C.B. Phillips (2000) Nonstationary binary choice. *Econometrica* 68, 1249–1280.

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

Persons, W.M. (1910) The correlation of economic statistics. *Journal of the American Statistical Association* 12, 287–333.

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.

Phillips, P.C.B. (1988) Multiple regression with integrated processes. In N.U. Prabhu, ed., *Statistical Inference from Stochastic Processes*. Contemporary Mathematics 80, pp. 79–106.

Phillips, P.C.B. (1998) New tools for understanding spurious regressions. *Econometrica* 66, 1299–1326.

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

Phillips, P.C.B. (2005a) Challenges of trending time series econometrics. *Mathematics and Computers in Simulation* 68, 401–416.

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

Phillips, P.C.B. & B.E. Hansen (1990) Standard inference in instrumental variable regressions with I(1) processes. *Review of Economic Studies* 57, 99–125.

Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression, Cowles Foundation Discussion paper 1182, Yale University.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

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 Martingale and Brownian Motion*, 3rd ed. Springer-Verlag.

"Student." (1914) The elimination of spurious correlation due to position in time or space. *Biometrika* 10, 179–180.

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

Wang, Q. & P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica*, forthcoming.

Yatchew, A. (2003) *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press.

Yule, G.U. (1921) On the time-correlation problem, with especial reference to the variate-difference correlation method. *Journal of the Royal Statistical Society* 84, 497–537.

Yule, G.U. (1926) Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. *Journal of the Royal Statistical Society* 89, 1–69.

APPENDIX: Technical Results and Proofs

We start with the following result.

LEMMA B. *Let*

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

$$\phi_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\epsilon^2}\right\},$$

and $\phi(z) = \phi_1(z)$*. Then, for each* $\epsilon > 0$

$$L_{n,\epsilon}^{(r)}(x) - \left(\int_{-\infty}^{\infty} g(s)\, ds\right) \frac{1}{n} \sum_{t=1}^{[nr]} y_{t,n} \phi_\epsilon \left(x_{t,n} + x\right) = o_p(1) \tag{A.1}$$

uniformly in $r \in [0,1]$ *and* x *as* $n \to \infty$ *and* $c_n \to \infty$.

Proof of Lemma B. The proof follows Lemma 7 of Jeganathan (2004). We can set $\epsilon = 1$ and consider

$$L_{n,1}^{(r)}(x) = \frac{c_n}{n} \sum_{t=1}^{[nr]} y_{t,n} \int_{-\infty}^{\infty} g\left(c_n\left(x_{t,n} + x + z\right)\right) \phi(z)\, dz$$

$$= \frac{c_n}{n} \sum_{t=1}^{[nr]} y_{t,n} \int_{-\infty}^{\infty} g(c_n s)\, \phi\left(s - x_{t,n} - x\right)\, ds.$$

Define $G_n(x) = \int_{-\infty}^{x} c_n g(c_n u)\, du = \int_{-\infty}^{x c_n} g(s)\, ds$ so that $dG_n(x) = c_n g(c_n x)\, dx$. Further, define $G(x) = \int_{-\infty}^{\infty} g(s)\, ds$ for $x \geq 0$ and $G(x) = 0$ for $x < 0$. Then, $G_n(x) \to G(x)$ at all continuity points of G as $n \to \infty$, and $G(b) - G(a) = 0$ if $0 \notin (a,b]$.

Note that

$$L_{n,1}^{(r)}(x) = \frac{1}{n} \sum_{t=1}^{[nr]} y_{t,n} \int_{-\infty}^{\infty} \phi\left(s - x_{t,n} - x\right) dG_n(s).$$

Hence for any $\upsilon > 0$ we have

$$L_{n,1}^{(r)}(x) - \frac{1}{n} \sum_{t=1}^{[nr]} y_{t,n} \int_{|s| \leq \upsilon} \phi\left(s - x_{t,n} - x\right) dG_n(s)$$

$$= \frac{1}{n} \sum_{t=1}^{[nr]} y_{t,n} \int_{|s| > \upsilon} \phi\left(s - x_{t,n} - x\right) dG_n(s).$$

Now

$$\sup_{r \in [0,1]} \left| \frac{1}{n} \sum_{t=1}^{[nr]} y_{t,n} \int_{|s| > \upsilon} \phi\left(s - x_{t,n} - x\right) dG_n(s) \right|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \left|y_{t,n}\right| \left| \int_{|s| > \upsilon} \phi\left(s - x_{t,n} - x\right) dG_n(s) \right|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \left|y_{t,n}\right| \int_{|s| > \upsilon} \phi\left(s - x_{t,n} - x\right) c_n \left|g(c_n s)\right| ds$$

$$= \frac{1}{n} \sum_{t=1}^{n} \left|y_{t,n}\right| \int_{|u| > c_n \upsilon} \phi\left(\frac{u}{c_n} - x_{t,n} - x\right) \left|g(u)\right| du$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \left|y_{t,n}\right| \int_{|u| > c_n \upsilon} \left|g(u)\right| du = o_p(1),$$

because $c_n \to \infty$ and ϕ is bounded. Thus

$$L_{n,1}^{(r)}(x) - \frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n} \int_{|s|\le \upsilon} \phi\left(s - x_{t,n} - x\right) dG_n(s) = o_p(1)$$

uniformly in $r \in [0,1]$. Next, by partitioning the interval $[-\upsilon, \upsilon]$ with a grid $\{s_{m,i} : i = -m, \ldots, m\}$ such that $\sup_i \left|s_{m,i} - s_{m,i-1}\right| \le \frac{2\upsilon}{m} \to 0$ as $m \to \infty$, and

$$s_{m,-m} = -\upsilon < s_{m,-m+1} < \cdots < s_{m,m-1} < s_{m,m} = \upsilon,$$

we have

$$\left|\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n} \int_{|s|\le \upsilon} \phi(s - x_{t,n} - x)\, dG_n(s) - \sum_{i=-m}^{m}\left\{\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n}\phi(s_{m,i} - x_{t,n} - x)\int_{s_{m,i}}^{s_{m,i+1}} dG_n(s)\right\}\right|$$

$$\le C\frac{\upsilon}{m}\frac{1}{n}\sum_{t=1}^{[nr]}\left|y_{t,n}\right|\left|\int_{|s|\le\upsilon} dG_n(s)\right|.$$

Here and subsequently C is a constant whose value may change in each usage. Also,

$$\left|\sum_{i=-m}^{m}\left\{\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n}\phi\left(s_{m,i} - x_{t,n} - x\right)\int_{s_{m,i}}^{s_{m,i+1}} dG_n(s)\right\} - \sum_{i=-m}^{m}\left\{\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n}\phi\left(s_{m,i} - x_{t,n} - x\right)\int_{s_{m,i}}^{s_{m,i+1}} dG(s)\right\}\right|$$

$$\le C\sum_{i=-m}^{m}\frac{1}{n}\sum_{t=1}^{[nr]}\left|y_{t,n}\right|\left|\int_{s_{m,i}}^{s_{m,i+1}} d\left(G_n(s) - G(s)\right)\right|.$$

Observe that $\int_{s_{m,i}}^{s_{m,i+1}} dG(s) = 0$ for all $0 < s_{m,i} < s_{m,i+1}$ and $s_{m,i} < s_{m,i+1} < 0$. Hence, as $m \to \infty$, we have

$$\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n}\sum_{i=-m}^{m}\left\{\phi\left(s_{m,i} - x_{t,n} - x\right)\int_{s_{m,i}}^{s_{m,i+1}} dG(s)\right\}$$

$$= \left(\int_{-\infty}^{\infty} g(s)\, ds\right)\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n}\phi\left(x_{t,n} + x\right) + o_p(1)$$

uniformly in $r \in [0,1]$. It follows that

$$\sup_{r,x}\left|L_{n,1}^{(r)}(x) - \left(\int_{-\infty}^{\infty} g(s)\, ds\right)\frac{1}{n}\sum_{t=1}^{[nr]} y_{t,n}\phi\left(x_{t,n} + x\right)\right|$$

$$\le C\frac{1}{n}\sum_{t=1}^{n}\left|y_{t,n}\right| Q\left(\upsilon, m, n\right),$$

where

$$Q(v,m,n) = \int_{|u|>c_n v} |g(u)|du + \frac{v}{m}\left|\int_{|s|\le v} dG_n(s)\right| + \sum_{i=-m}^{m}\left|\int_{y_{m,i}}^{y_{m,i+1}} d(G_n(s)-G(s))\right|.$$

Evidently,

$$\lim_{v\to\infty}\lim_{m\to\infty}\lim_{n\to\infty} Q(v,m,n) = 0,$$

thereby giving the stated result. ■

Proof of Theorem 1. Define

$$L_n^{(r)} = \frac{c_n}{n}\sum_{k=1}^{[nr]} y_{k,n} g(c_n x_{k,n}), \qquad L_{n,\epsilon}^{(r)} = \frac{c_n}{n}\sum_{k=1}^{[nr]} y_{k,n}\int_{-\infty}^{\infty} g\left[c_n (x_{k,n}+z\epsilon)\right]\phi(z)dz, \tag{A.2}$$

where $\phi(x)=\phi_1(x)$. The term $L_{n,\epsilon}^{(r)}$ may be regarded as a locally smoothed version of $L_n^{(r)}$ using the normal density $\phi(x)$. This version is useful because it leads to a further approximation that is amenable to the use of a continuous mapping.

It follows from Lemma B that, for any $\epsilon > 0$,

$$L_{n,\epsilon}^{(r)} - \frac{\tau}{n}\sum_{k=1}^{[nr]} y_{k,n}\phi_\epsilon(x_{k,n}) = o_P(1) \tag{A.3}$$

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

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{0\le r\le 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0. \tag{A.4}$$

Indeed, it follows from the continuous mapping theorem that, for $\forall\epsilon > 0$,

$$\frac{1}{n}\sum_{k=1}^{[nr]} y_{k,n}\phi_\epsilon(x_{k,n}) = \int_0^r y_{[np],n}\phi_\epsilon(x_{[np],n})\,dp - \frac{1}{n}y_{0,n}\phi_\epsilon(0) + \frac{1}{n}y_{[nr],n}\phi_\epsilon(x_{n,[nr]}) \to_D \int_0^r G_y(p)\phi_\epsilon(G_x(p))\,dp. \tag{A.5}$$

Next, using the extended occupation times formula (11) and the fact that the local time process $L_{G_x}(t,s)$ is almost surely continuous, we find that

$$\int_0^r G_y(p)\phi_\epsilon(G_x(p))\,dp = \int_{-\infty}^{\infty} ds \int_0^r G_y(p)\phi_\epsilon(s)\,dL_{G_x}(p,s)$$
$$= \int_{-\infty}^{\infty} ds \int_0^r G_y(p)\frac{1}{\varepsilon}\phi(\frac{s}{\varepsilon})\,dL_{G_x}(p,s)$$
$$= \int_{-\infty}^{\infty} da \int_0^r G_y(p)\phi(a)\,dL_{G_x}(p,\varepsilon a)$$
$$= \int_{-\infty}^{\infty} \phi(a)\,da \int_0^r G_y(p)\,dL_{G_x}(p,0) + o_{a.s.}(1)$$
$$= \int_0^r G_y(p)\,dL_{G_x}(p,0) + o_{a.s.}(1), \qquad \textbf{(A.6)}$$

as $\epsilon \to 0$. Combining (A.6) and (A.3) we have

$$L_{n,\epsilon}^{(r)} \to_D \ \tau \int_0^r G_y(p)\,dL_{G_x}(p,0),$$

as $n \to \infty$ and $\epsilon \to 0$.

Thus, it remains to show (A.4). To do so, we use an argument similar to that of the proof of Theorem 2.1 of Wang and Phillips (2009a). Define

$$X_{k,n}^{\varepsilon}(z) = \left\{g[c_n x_{k,n}] - g[c_n(x_{k,n} + z\epsilon)]\right\},$$

$$Y_{k,n}^{\varepsilon}(z) = y_{k,n} X_{k,n}^{\varepsilon}(z).$$

By definition (A.2) and because $\int_{-\infty}^{\infty}\phi(x)\,dx = 1$ we have

$$L_n^{(r)} - L_{n,\epsilon}^{(r)} = \int_{-\infty}^{\infty} \frac{c_n}{n}\sum_{k=1}^{[nr]} y_{k,n} X_{k,n}^{\varepsilon}(z)\,\phi(z)\,dz = \int_{-\infty}^{\infty} \frac{c_n}{n}\sum_{k=1}^{[nr]} Y_{k,n}^{\varepsilon}(z)\,\phi(z)\,dz,$$

and we may proceed as in the paper by Wang and Phillips (2009a), who prove the result without the variable $y_{k,n}$, that is, with $Y_{k,n}^{\varepsilon}(z)$ replaced by $X_{k,n}^{\varepsilon}(z)$. We have

$$\sup_{0\le r\le 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| \le \int_{-\infty}^{\infty} \frac{c_n}{n} \sup_{0\le r\le 1} E\left|\sum_{k=1}^{[nr]} Y_{k,n}^{\varepsilon}(z)\right| \phi(z)dz. \qquad \textbf{(A.7)}$$

Recall that $x_{k,n}/d_{k,0,n}$ has a conditional density $h_{k,0,n}^x(x|y)$ and $y_{k,n}/d_{k,0,n}$ has a density $h_{k,0,n}^y(y)$, both of which are bounded by a constant for all x and y, $1 \le k \le n$ and $n \ge 1$. It follows that, for all $z \in R$ and $1 \le k \le n$, and a generic constant C,

$$c_n E\left|Y_{k,n}^{\varepsilon}(z)\right| = c_n \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g\left(c_n d_{k,0,n} x\right)$$
$$-g\left[c_n\left(d_{k,0,n} x + z\epsilon\right)\right]\left| h_{k,0,n}^x(x|y)\,dx \right| d_{k,0,n} y \left| h_{k,0,n}^y(y)\,dy\right.$$
$$\le C \int_{-\infty}^{\infty} |g(u) - g(u + c_n z\epsilon)|\,du \int_{-\infty}^{\infty} |y|\,h_{k,0,n}^y(y)\,dy$$
$$\le 2C \int_{-\infty}^{\infty} |g(u)|\,du \int_{-\infty}^{\infty} |y|\,h(y)\,dy < \infty \qquad \textbf{(A.8)}$$

because $\left|h^{y}_{k,0,n}(y)\right| \le h(y)$, $\int_{-\infty}^{\infty} |y|\, h(y) dy < \infty$, and $\int_{-\infty}^{\infty} |g(u)|\, du < \infty$. Thus,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{c_n}{n} \sup_{0\le r\le 1} E\left|\sum_{k=1}^{[nr]} Y^{\epsilon}_{k,n}(z)\right| \phi(z) dz \\
&\le \int_{-\infty}^{\infty} \frac{c_n}{n} \sup_{0\le r\le 1} \sum_{k=1}^{[nr]} E\left|Y^{\epsilon}_{k,n}(z)\right| \phi(z)\, dz \\
&= \int_{-\infty}^{\infty} \frac{c_n}{n} \sum_{k=1}^{n} E\left|Y^{\epsilon}_{k,n}(z)\right| \phi(z)\, dz \\
&\le 2C \int_{-\infty}^{\infty} |g(u)|\, du \int_{-\infty}^{\infty} |y|\, h(y)\, dy < \infty. \qquad \textbf{(A.9)}
\end{aligned}$$

This, together with (A.7) and the dominated convergence theorem, implies that to prove (A.4) it suffices to show for each fixed z that

$$\Lambda_n(\epsilon) \equiv \frac{c_n^2}{n^2} \sup_{0\le r\le 1} E\left[\sum_{k=1}^{[nr]} Y^{\epsilon}_{k,n}(z)\right]^2 \to 0 \qquad \textbf{(A.10)}$$

when $n \to \infty$ first and then $\epsilon \to 0$. With some modifications, the proof of (A.10) follows the proof of Theorem 2.1 in Wang and Phillips (2009a). We rewrite Λ_n as

$$\begin{aligned}
\Lambda_n(\epsilon) &= \frac{c_n^2}{n^2} \sup_{0\le r\le 1} \sum_{k=1}^{[nr]} E\left[Y^{\epsilon}_{k,n}(z)\right]^2 + \frac{2c_n^2}{n^2} \sup_{0\le r\le 1} \sum_{k=1}^{[nr]} \sum_{l=k+1}^{[nr]} E\left[Y^{\epsilon}_{k,n}(z)\, Y^{\epsilon}_{l,n}(z)\right] \\
&= \Lambda_{1n}(\epsilon) + \Lambda_{2n}(\epsilon), \quad \text{say.}
\end{aligned}$$

First, because $g^2(x)$ is integrable, by an argument similar to that leading to (A.9), we have

$$\begin{aligned}
\Lambda_{1n}(\epsilon) &= \frac{c_n^2}{n^2} \sup_{0\le r\le 1} \sum_{k=1}^{[nr]} E\left[Y^{\epsilon}_{k,n}(z)\right]^2 = \frac{c_n^2}{n^2} \sum_{k=1}^{n} E\left[Y^{\epsilon}_{k,n}(z)\right]^2 \\
&= \frac{c_n^2}{n^2} \sum_{k=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| g\left(c_n\, d_{k,0,n}\, x\right) \right. \\
&\quad \left. - g\left[c_n\left(d_{k,0,n}\, x + z\epsilon\right)\right] \right|^2 h^{x}_{k,0,n}(x|y)\, dx \left|d_{k,0,n}\, y\right|^2 h^{y}_{k,0,n}(y)\, dy \\
&\le A \frac{c_n d_{k,0,n}}{n} \int_{-\infty}^{\infty} |g(u) - g(u + c_n z\epsilon)|^2\, du \int_{-\infty}^{\infty} |y|^2\, h^{y}_{k,0,n}(y)\, dy \\
&\le \frac{4Ac_n d_{k,0,n}}{n} \int_{-\infty}^{\infty} |g(u)|^2\, du \int_{-\infty}^{\infty} |y|^2\, h(y)\, dy \to 0
\end{aligned}$$

as $n \to \infty$ because $\int_{-\infty}^{\infty} |g(u)|^2 du \int_{-\infty}^{\infty} |y|^2 h(y) dy < \infty$ and $\frac{c_n d_{k,0,n}}{n} \le \frac{c_n}{n} \to 0$.

We next prove that $\lim_{\epsilon \to 0} \lim_{n \to \infty} \Lambda_{2n}(\epsilon) = 0$, and then the required result (A.10) follows. First, note that

$$E\left[Y_{k,n}^{\varepsilon}(z)\, Y_{l,n}^{\varepsilon}(z)\right] = E\left[Y_{k,n}^{\varepsilon}(z)\, E\left\{Y_{l,n}^{\varepsilon}(z) | \mathcal{F}_{k,n}\right\}\right]. \tag{A.11}$$

Let $\Omega_n = \Omega_n(\epsilon^{1/(2m_0)})$ where, as defined earlier,

$$\Omega_n(\eta) = \{(l,k) : \eta n \le k \le (1-\eta) n, \qquad k + \eta n \le l \le n\}, \quad \text{for } 0 < \eta < 1.$$

Recall that $x_{k,n}$ are adapted to $\mathcal{F}_{k,n}$ and conditional on $\mathcal{F}_{k,n}$ and $y_{l,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a bounded density $h_{l,k,n}(x|y)$. We have

$$\begin{aligned}
& c_n \left| E(y_{l,n}^{\varepsilon} \mid \mathcal{F}_{k,n}) \right| \\
&\quad = c_n \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(g\left[c_n x_{k,n} + c_n d_{l,k,n} x\right] - g\left[c_n (x_{k,n} + z\epsilon) + c_n d_{l,k,n} x\right]\right) \right. \\
&\qquad \left. \times h_{l,k,n}^{x}(x|y)\, dx \left| d_{l,k,n} y \right| h_{l,k,n}^{y}(y)\, dy \right. \\
&\quad = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(g[c_n x_{k,n} + u] \right. \right. \\
&\qquad \left. \left. - g\left[c_n (x_{k,n} + z\epsilon) + u\right]\right) h_{l,k,n}^{x}\left(\frac{u}{c_n d_{l,k,n}} |y\right) du\, |y|\, h_{l,k,n}^{y}(y)\, dy \right| \\
&\quad \le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(v)| \left| V(v, c_n x_{k,n} | y) \right| dv\, |y|\, h_{l,k,n}^{y}(y)\, dy \\
&\quad \le \begin{cases} C, & \text{if } (l,k) \notin \Omega_n \\ C\left\{\int_{-\infty}^{\infty} \left(\int_{|v| \ge \sqrt{c_n}} |g(v)| dv\right) |y|\, h_{l,k,n}^{y}(y)\, dy \right. & \\ \left. + \int_{-\infty}^{\infty} \left(\int_{|v| \le \sqrt{c_n}} |g(v)|\, |V(v, c_n x_{k,n} | y)| dv\right) |y|\, h_{l,k,n}^{y}(y)\, dy\right\}, & \text{if } (l,k) \in \Omega_n, \end{cases}
\end{aligned} \tag{A.12}$$

where

$$V(v, t | y) = h_{l,k,n}^{x}\left(\frac{v - t}{c_n d_{l,k,n}} |y\right) - h_{l,k,n}^{x}\left(\frac{v - t - c_n z\epsilon}{c_n d_{l,k,n}} |y\right).$$

Furthermore, as in the proof of (A.8), whenever $|v| \le \sqrt{c_n}$, n is large enough, and $(l,k) \in \Omega_n$,

$$
\begin{aligned}
E & \left[\left|Y_{k,n}^{\varepsilon}(z)\right|\left|V(v, c_n x_{k,n}|y)\right|\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|g\left[c_n\left(d_{k,0,n}x + z\epsilon\right)\right] - g\left(c_n d_{k,0,n}x\right)\right| |V(v, c_n d_{k,0,n}x)| \\
&\qquad \times h_{k,0,n}^{x}(x|y)\,dx\ \left|d_{k,0,n}y\right| h_{k,0,n}^{y}(y)\,dy \\
&\le \frac{C}{c_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|g(x + c_n z\epsilon) - g(x)|\,|V(v, x|y)|\,dx\ |y|\,h_{k,0,n}^{y}(y)\,dy \\
&\le \frac{C}{c_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|g(x)|\{|V(v, x|y)| + |V(v, x - c_n z\epsilon|y)|\}\,dx\,|y|\,h_{k,0,n}^{y}(y)\,dy \\
&\le \frac{C}{c_n}\left(\int_{|x|\ge\sqrt{c_n}}|g(x)|\,dx + \sup_{|u|\le C\epsilon^{1/2}}|h_{l,k,n}^{x}(u|y) - h_{l,k,n}^{x}(0|y)|\right), \qquad \textbf{(A.13)}
\end{aligned}
$$

where we have used the facts that $\inf_{(l,k)\in\Omega_n} d_{l,k,n} \ge \epsilon^{1/2}/C$, $c_n \to \infty$, $V(v, t|y)$ is bounded, and $\int_{-\infty}^{\infty}|y|\,h_{k,0,n}^{y}(y)dy < \infty$. In particular, the second term in parentheses in (A.13) occurs because $|v| \le \sqrt{c_n}$ and $|x| \le \sqrt{c_n}$, so that

$$
\begin{aligned}
&\left|h_{l,k,n}^{x}\left(\frac{v - x + c_n z\epsilon}{c_n d_{l,k,n}}|y\right) - h_{l,k,n}^{x}\left(\frac{v - x}{c_n d_{l,k,n}}|y\right)\right| \\
&\quad = \left|h_{l,k,n}^{x}\left(\frac{c_n z\epsilon + O\left(\sqrt{c_n}\right)}{c_n d_{l,k,n}}|y\right) - h_{l,k,n}^{x}\left(O\left(\frac{1}{\sqrt{c_n}d_{l,k,n}}\right)|y\right)\right| \\
&\quad = \left|h_{l,k,n}^{x}\left(\frac{z\epsilon}{d_{l,k,n}}\left[1 + O\left(\frac{1}{\sqrt{c_n}}\right)\right]|y\right) - h_{l,k,n}^{x}(o(1)|y)\right| \\
&\quad \le \sup_{|u|\le Cz\epsilon^{1/2}}|h_{l,k,n}^{x}(u|y) - h_{l,k,n}^{x}(0|y)|,
\end{aligned}
$$

because $\inf_{(l,k)\in\Omega_n} d_{l,k,n} \ge \epsilon^{1/2}/C$.

In view of these results, together with (A.8) and (A.11), we find that for $(l,k) \notin \Omega_n$,

$$
\begin{aligned}
\left|E\left[Y_{k,n}^{\epsilon}(z)Y_{l,n}^{\epsilon}(z)\right]\right| &= \left|E\left[Y_{k,n}^{\epsilon}(z)\,E\left\{Y_{l,n}^{\epsilon}(z)|\mathcal{F}_{k,n}\right\}\right]\right| \\
&\le \frac{C}{c_n}\,E\,|Y_{k,n}(z)| \le \frac{C}{c_n^2}, \qquad \textbf{(A.14)}
\end{aligned}
$$

and, if $(l,k) \in \Omega_n$, using (A.12) and (A.13)

$$\left|E\left[Y_{k,n}^{\epsilon}(z)Y_{l,n}^{\epsilon}(z)\right]\right|$$

$$\le \frac{A}{c_n} E\left|Y_{k,n}^{\epsilon}(z)\right| \int_{|v|\ge\sqrt{c_n}} |g(v)|dv$$

$$+\frac{A}{c_n}\int_{-\infty}^{\infty}\int_{|v|\le\sqrt{c_n}} |g(v)|\, E\left\{\left|Y_{k,n}^{\epsilon}(z)\right| \left|V(v, c_n x_{k,n}|y)\right|\right\} dv\, |y|\, h_{l,k,n}^{y}(y)\, dy$$

$$\le \frac{A}{c_n^2}\int_{|v|\ge\sqrt{c_n}} |g(v)|dv$$

$$+\frac{A}{c_n^2}\int_{|v|\le\sqrt{c_n}} |g(v)|dv \int_{-\infty}^{\infty} \sup_{|u|\le Cz\epsilon^{1/2}} |h_{l,k,n}^{x}(u|y) - h_{l,k,n}^{x}(0|y)|\,|y|\, h_{l,k,n}^{y}(y)\, dy. \quad \textbf{(A.15)}$$

It follows from (A.14)–(A.15) that, with $\eta = \epsilon^{1/2}/C$ in what follows,

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

$$\le \frac{C}{n^2}\sum_{l-k\le\eta n} + \frac{C}{n^2}\sum_{(l,k)\in\Omega_n}\int_{|v|\ge\sqrt{c_n}} |g(v)|dv$$

$$+\frac{C}{n^2}\sum_{(l,k)\in\Omega_n}\sup_{|u|\le Cz\epsilon^{1/2}}\sup_{y}|h_{l,k,n}^{x}(u|y) - h_{l,k,n}^{x}(0|y)|$$

$$\le C\eta^2 + C\int_{|v|\ge\sqrt{c_n}} |g(v)|dv + C\sup_{|u|\le Cz\epsilon^{1/2}}\sup_{y}|h_{l,k,n}^{x}(u|y) - h_{l,k,n}^{x}(0|y)|$$

$$\to 0,$$

as $n \to \infty$ and $\epsilon \to 0$, as required. The proof of Theorem 1 is now complete. ■

Proof of Theorem 2. Set $c_n = d_n/h_n$. By virtue of Theorem 1 and Assumption 3.1, we have

$$\frac{1}{nh_n}\sum_{t=1}^{n} y_t K\left(\frac{x_t - x}{h_n}\right) = \frac{d_n}{nh_n}\sum_{t=1}^{n}\frac{y_t}{d_n} K\left(\frac{d_n\left(\frac{x_t}{d_n} - \frac{x}{d_n}\right)}{h_n}\right)$$

$$= \frac{c_n}{n}\sum_{t=1}^{n} y_{t,n} K\left(c_n\left(x_{t,n} - \frac{x}{d_n}\right)\right)$$

$$\Rightarrow \begin{cases} \int_0^1 G_y(p)\, dL_{G_x}(p,0) & \text{for fixed } x \\ \int_0^1 G_y(p)\, dL_{G_x}(p,a) & \text{for } x = d_n a \text{ with } a \text{ fixed} \end{cases} \quad \textbf{(A.16)}$$

and

$$\frac{d_n}{nh_n}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h_n}\right)=\frac{d_n}{nh_n}\sum_{t=1}^{n}K\left(\frac{d_n\left(\frac{x_t}{d_n}-\frac{x}{d_n}\right)}{h_n}\right)$$
$$=\frac{c_n}{n}\sum_{t=1}^{n}K\left(c_n\left(x_{t,n}-\frac{x}{d_n}\right)\right)$$
$$\Rightarrow\begin{cases}L_{G_x}(1,0) & \text{for fixed } x\\ L_{G_x}(1,a) & \text{for } x=d_n a \text{ with } a \text{ fixed.}\end{cases} \tag{A.17}$$

Joint convergence of (A.16) and (A.17) holds in view of Assumption 2.2. It follows that

$$d_n^{-1}\hat{g}(x)=\frac{\frac{1}{nh_n}\sum_{t=1}^{n}y_t K\left(\frac{x_t-x}{h_n}\right)}{\frac{d_n}{nh_n}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h_n}\right)}$$
$$\Rightarrow\begin{cases}\frac{\int_0^1 G_y(p)\,dL_{G_x}(p,0)}{L_{G_x}(1,0)} & \text{for fixed } x\\ \frac{\int_0^1 G_y(p)\,dL_{G_x}(p,a)}{L_{G_x}(1,a)} & \text{for } x=d_n a \text{ with } a \text{ fixed}\end{cases},$$

giving the stated result. ■

Proof of Lemma A. Write $s^2_{(d_n a)}$ in standardized form as

$$\frac{s^2_{(d_n a)}}{d_n}=\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\hat{v}_t}{d_n}\right)^2 K_h\left(x_t-d_n a\right). \tag{A.18}$$

The standardized residuals $\hat{v}_t/d_n$ have the following local form for $x_t\sim d_n a$:

$$\left.\frac{\hat{v}_t}{d_n}\right|_{x_t\sim d_n a}=\left.\left(\frac{y_t}{d_n}-\frac{\hat{g}(x_t)}{d_n}\right)\right|_{x_t\sim d_n a},$$

whose limit behavior is given by

$$\left.\frac{\hat{v}_{[nr]}}{d_n}\right|_{x_{[nr]}\sim d_n a}\Rightarrow G_y(r)-\frac{\int_0^1 G_y(p)\,dL_{G_x}(p,a)}{L_{G_x}(1,a)}:=C_a(r),$$

which follows from Assumption 2.2 and Theorem 2.

The limit of $s^2_{(d_n a)}/d_n$ is then a simple consequence of (a second moment version of) Theorem 1, namely,

$$\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\hat{v}_t}{d_n}\right)^2 K_h(x_t-d_n a)=\frac{d_n}{nh_n}\sum_{t=1}^{n}\left(\frac{\hat{v}_t}{d_n}\right)^2 K\left(\frac{d_n\left(x_{t,n}-a\right)}{h}\right)$$
$$=\frac{c_n}{n}\sum_{t=1}^{n}\left(\frac{\hat{v}_t}{d_n}\right)^2 K\left(c_n\left(x_{t,n}-a\right)\right)$$
$$\Rightarrow\int_0^1 C_a^2(r)\,dL_{G_x}(r,a), \tag{A.19}$$

giving the stated result. ■

Proof of Theorem 3. From Lemma A, (18), and Assumption 2.2, we have

$$\frac{nh_n}{d_n^2}s^2_{\hat{g}(d_na)} = \frac{d_n^{-1}s^2_{(d_na)}\mu_{K^2}}{\frac{d_n}{nh_n}\sum_{t=1}^n K\left(\frac{x_t-d_na}{h_n}\right)}$$

$$\Rightarrow \frac{\int_0^1 C_a^2(r)\,dL_{G_x}(r,a)}{L_{G_x}(1,a)}. \tag{A.20}$$

Then, using Assumption 2.2, (21), (26), and (29), we have

$$\frac{1}{\sqrt{nh_n}}t_{\hat{g}(d_na)} = \frac{1}{\sqrt{nh_n}}\frac{\hat{g}(d_na)}{s_{\hat{g}(d_na)}} = \frac{1}{\sqrt{nh_n}}\frac{d_n}{\left(\frac{d_n^2}{nh_n}\right)^{1/2}}\frac{d_n^{-1}\hat{g}(d_na)}{\left\{\frac{nh_n}{d_n^2}s^2_{\hat{g}(d_na)}\right\}^{1/2}}$$

$$= \frac{d_n^{-1}\hat{g}(d_na)}{\left\{\frac{nh_n}{d_n^2}s^2_{\hat{g}(d_na)}\right\}^{1/2}}$$

$$\Rightarrow \frac{\frac{\int_0^1 G_y(p)\,dL_{G_x}(p,a)}{L_{G_x}(1,a)}}{\left\{\frac{\int_0^1 C_a^2(r)\,dL_{G_x}(r,a)\mu_{K^2}}{L_{G_x}(1,a)}\right\}^{1/2}}$$

$$= \frac{\int_0^1 G_y(p)\,dL_{G_x}(p,a)}{\left\{L_{G_x}(1,a)\int_0^1 C_a^2(r)\,dL_{G_x}(r,a)\,\mu_{K^2}\right\}^{1/2}},$$

giving the stated result.

Next, using (16), Lemma A, (27), and Assumption 2.2, the local R^2 coefficient is

$$R_n^2(d_na) \Rightarrow \frac{\sum_{t=1}^n y_t^2 K_{h_n}(x_t-d_na) - \sum_{t=1}^n \hat{v}_t^2 K_{h_n}(x_t-d_na)}{\sum_{t=1}^n y_t^2 K_{h_n}(x_t-d_na)}$$

$$= \frac{\frac{d_n}{n}\sum_{t=1}^n y_{t,n}^2 K_{h_n}(x_t-d_na) - \frac{d_n}{n}\sum_{t=1}^n \left(\frac{\hat{v}_t}{d_n}\right)^2 K_{h_n}(x_t-d_na)}{\frac{d_n}{n}\sum_{t=1}^n y_{t,n}^2 K_{h_n}(x_t-d_na)}$$

$$\Rightarrow \frac{\int_0^1 G_y(p)^2 dL_{G_x}(p,a) - \int_0^1 C_a^2(r)\,dL_{G_x}(r,a)}{\int_0^1 G_y(p)^2 dL_{G_x}(p,a)}$$

$$= \frac{\int_0^1 \hat{G}_y^2(r) dL_{G_x}(r,a)}{\int_0^1 G_y(p)^2 dL_{G_x}(p,a)} = \frac{\left(\int_0^1 G_y(p) dL_{G_x}(p,a)\right)^2}{L_{G_x}(1,a)\int_0^1 G_y(p)^2 dL_{G_x}(p,a)},$$

as required.

Finally,

$$DW_n\left(\sqrt{n}a\right) = \frac{\sum_{t=1}^{n}\left(\Delta\hat{v}_t\right)^2 K_{h_n}\left(x_t - \sqrt{n}a\right)}{\sum_{t=1}^{n}\hat{v}_t^2 K_{h_n}\left(x_t - \sqrt{n}a\right)}$$

$$= \frac{\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\Delta\hat{v}_t}{d_n}\right)^2 K_{h_n}\left(x_t - \sqrt{n}a\right)}{\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\hat{v}_t}{d_n}\right)^2 K_{h_n}\left(x_t - d_n a\right)}, \tag{A.21}$$

and we need to consider

$$\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\Delta\hat{v}_t}{d_n}\right)^2 K_{h_n}\left(x_t - \sqrt{n}a\right).$$

Now

$$\frac{\Delta\hat{v}_t}{d_n} = \left(\frac{y_t}{d_n} - \frac{y_{t-1}}{d_n}\right) - \left(\frac{\hat{g}(x_t)}{d_n} - \frac{\hat{g}\left(x_{t-1}\right)}{d_n}\right),$$

and so, by Assumption 2.2 and Theorem 2, $\frac{\Delta\hat{v}_{[nr]}}{d_n} \to_p 0$ for all $r \in [0,1]$. Thus,

$$\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\Delta\hat{v}_t}{d_n}\right)^2 K_{h_n}\left(x_t - \sqrt{n}a\right) \to_p 0. \tag{A.22}$$

Because $\frac{d_n}{n}\sum_{t=1}^{n}\left(\frac{\hat{v}_t}{d_n}\right)^2 K_{h_n}(x_t - d_n a) \Rightarrow \int_0^1 C_a^2(r)\, dL_{G_x}(r,a) > 0$, it follows from (A.21) and (A.22) that $DW_n\left(\sqrt{n}a\right) \to_p 0$, as required. ■