STRUCTURAL NONPARAMETRIC COINTEGRATING REGRESSION

BY

QIYING WANG and PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1289

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

2009

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 77, No. 6 (November, 2009), 1901–1948

STRUCTURAL NONPARAMETRIC COINTEGRATING REGRESSION

QIYING WANG
School of Mathematics and Statistics, The University of Sydney, Sydney, NSW 2006, Australia

PETER C. B. PHILLIPS
Cowles Foundation, Yale University, New Haven, CT 06520-8281, U.S.A. and University of Auckland, Auckland, New Zealand and University of Southampton, Southampton, U.K., and Singapore Management University, Singapore

Econometrica, Vol. 77, No. 6 (November, 2009), 1901–1948

STRUCTURAL NONPARAMETRIC COINTEGRATING REGRESSION

BY QIYING WANG AND PETER C. B. PHILLIPS[1]

Nonparametric estimation of a structural cointegrating regression model is studied. As in the standard linear cointegrating regression model, the regressor and the dependent variable are jointly dependent and contemporaneously correlated. In nonparametric estimation problems, joint dependence is known to be a major complication that affects identification, induces bias in conventional kernel estimates, and frequently leads to ill-posed inverse problems. In functional cointegrating regressions where the regressor is an integrated or near-integrated time series, it is shown here that inverse and ill-posed inverse problems do not arise. Instead, simple nonparametric kernel estimation of a structural nonparametric cointegrating regression is consistent and the limit distribution theory is mixed normal, giving straightforward asymptotics that are useable in practical work. It is further shown that use of augmented regression, as is common in linear cointegration modeling to address endogeneity, does not lead to bias reduction in nonparametric regression, but there is an asymptotic gain in variance reduction. The results provide a convenient basis for inference in structural nonparametric regression with nonstationary time series when there is a single integrated or near-integrated regressor. The methods may be applied to a range of empirical models where functional estimation of cointegrating relations is required.

KEYWORDS: Brownian local time, cointegration, functional regression, Gaussian process, integrated process, kernel estimate, near integration, nonlinear functional, nonparametric regression, structural estimation, unit root.

1. INTRODUCTION

A GOOD DEAL OF RECENT ATTENTION in econometrics has focused on functional estimation in structural econometric models and the inverse problems to which they frequently give rise. A leading example is a structural nonlinear regression where the functional form is the object of primary interest. In such systems, identification and estimation are typically much more challenging than in linear systems because they involve the inversion of integral operator equations which may be ill-posed in the sense that the solutions may not exist, may not be unique, and may not be continuous. Some recent contributions to this field include Newey, Powell, and Vella (1999), Newey and Powell (2003), Ai and Chen (2003), Florens (2003), and Hall and Horowitz (2005). Overviews of the ill-posed inverse literature are given in Florens (2003) and Carrasco, Florens, and Renault (2007). All of this literature has focused on microeconometric and stationary time series settings.

In linear structural systems, problems of inversion from the reduced form are much simpler, and conditions for identification and consistent estimation techniques have been extensively studied. Under linearity, it is also well known

[1]The authors thank a co-editor and two referees for helpful comments. Wang acknowledges partial research support from the Australian Research Council. Phillips acknowledges partial research support from a Kelly Fellowship and NSF Grant SES 06-47086.

 DOI: 10.3982/ECTA7732

that the presence of nonstationary regressors can provide a simplification. In particular, for cointegrated systems involving time series with unit roots, structural relations are actually present in the reduced form (and therefore always identified) because of the unit roots in a subset of the determining equations. In fact, such models can always be written in error correction or reduced rank regression format where the structural relations are immediately evident.

The present paper shows that nonstationarity leads to major simplifications in the context of structural nonlinear functional regression. The primary simplification arises because in nonlinear models with endogenous nonstationary regressors, there is no ill-posed inverse problem. In fact, there is no inverse problem at all in the functional treatment of such systems. Furthermore, identification does not require the existence of instrumental variables that are orthogonal to the equation errors. Finally, and perhaps most importantly for practical work, consistent estimation may be accomplished using standard kernel regression techniques, and inference may be conducted in the usual way and is valid asymptotically under simple regularity conditions. These results for kernel regression in structural nonlinear models of cointegration open up new possibilities for empirical research.

The reason why there is no inverse problem in structural nonlinear nonstationary systems can be explained heuristically as follows. In a nonparametric structural setting, it is conventional to impose on the disturbances a zero conditional mean condition given certain instruments, so as to assist in identifying an infinite-dimensional function. Such conditions lead to an integral equation involving the conditional probability distribution of the regressors and the structural function integrated over the space of the regressor. This equation describes the relation between the structure and reduced form, and its solution, if it exists and is unique, delivers the unknown structural function. But when the endogenous regressor is nonstationary, there is no invariant probability distribution of the regressor, only the local time density of the limiting stochastic process corresponding to a standardized version of the regressor as it sojourns in the neighborhood of a particular spatial value. Accordingly, there is no integral equation relating the structure to the reduced form. In fact, the structural equation itself is locally also a reduced form equation in the neighborhood of this spatial value, for when an endogenous regressor is in the locality of a specific value, the systematic part of the structural equation depends on that specific value and the equation is effectively a reduced form. What is required is that the nonstationary regressor spends enough time in the vicinity of a point in the space to ensure consistent estimation. This in turn requires recurrence, so that the local time of the limit process corresponding to the time series is positive. In addition, the random wandering nature of a stochastically nonstationary regressor such as a unit root process ensures that the regressor inevitably departs from any particular locality and thereby assists in tracing out (and identifying) the structural function over a wide domain. The process is similar to the manner in which instruments may shift the location in which a structural function is observed and in doing so assist in the process of identification when the data are stationary.

Linear cointegrating systems reveal a strong form of this property. As mentioned above, in linear cointegration the inverse problem disappears completely because the structural relations continue to be present in the reduced form. Indeed, they are the same as reduced form equations up to simple time shifts, which are of no importance in linear long run relations. In nonlinear structural cointegration, the same behavior applies locally in the vicinity of a particular spatial value, thereby giving local identification of the structural function and facilitating estimation.

In linear cointegration, the signal strength of a nonstationary regressor ensures that least squares estimation is consistent, although the estimates are well known to have second order bias (Phillips and Durlauf (1986), Stock (1987)) and are therefore seldom used in practical work. Much attention has therefore been given in the time series literature to the development of econometric estimation methods that remove the second order bias and are asymptotically and semiparametrically efficient.

In nonlinear structural functional estimation with a single nonstationary regressor, this paper shows that local kernel regression methods are consistent and that under some regularity conditions they are also asymptotically mixed normally distributed, so that conventional approaches to inference are possible. It is not necessary to use special methods or even an augmented regression equation where the cointegrating model is adjusted for the conditional mean to account for endogeneity, such as the augmented regressions that underlie semiparametric methods like FM regression or dynamic least squares in linear cointegrating models. These results constitute a major simplification in the functional treatment of nonlinear cointegrated systems and they directly open up empirical applications with existing methods.

In related recent work, Karlsen, Myklebust, and Tjøstheim (2007) and Schienle (2008) used Markov chain methods to develop an asymptotic theory of kernel regression that allows for some forms of nonstationarity and endogeneity in the regressor. Schienle also considered additive nonparametric models with many nonstationary regressors and smooth backfitting methods of estimation.

The results in the current paper are obtained using local time convergence techniques, extending those in Wang and Phillips (2009) to the endogenous regressor case and allowing for both integrated and near-integrated regressors with general forms of serial dependence in the generating mechanism and equilibrium error. The validity of the limit theory in the case of near-integrated regressors is important in practice because it is often convenient in empirical work not to insist on unit roots and to allow for roots near unity in the regressors. By contrast, conventional methods of estimation and inference in parametric models of linear cointegration are known to break down when the regressors have roots local to unity.

The paper is organized as follows. Section 2 introduces the model and assumptions. Section 3 provides the main results on the consistency and limit distribution of the kernel estimator in a structural model of nonlinear cointegration and associated methods of inference. Section 4 reports some Monte Carlo simulations that explore the finite sample performance of the kernel estimator and the effects of augmented regression specification. Section 5 concludes and outlines ways in which the present paper may be extended. Proofs and various subsidiary technical results are given in Sections 6–9, which function as appendices to the paper.

2. MODEL AND ASSUMPTIONS

We consider the nonlinear structural model of cointegration

$$y_t = f(x_t) + u_t \qquad (t = 1, 2, \dots, n), \tag{2.1}$$

where u_t is a zero mean stationary equilibrium error, x_t is a jointly dependent nonstationary regressor, and f is an unknown function to be estimated with the observed data $\{y_t, x_t\}_{t=1}^n$. The conventional kernel estimate of $f(x)$ in model (2.1) is given by

$$\hat{f}(x) = \frac{\sum_{t=1}^{n} y_t K_h(x_t - x)}{\sum_{t=1}^{n} K_h(x_t - x)}, \tag{2.2}$$

where $K_h(s) = (1/h)K(s/h)$, $K(x)$ is a nonnegative real function, and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$.

The limit behavior of $\hat{f}(x)$ has been investigated in past work in some special situations, notably where the error process u_t is a martingale difference sequence and there is no contemporaneous correlation between x_t and u_t. These are strong conditions, they are particularly restrictive in relation to the conventional linear cointegrating regression framework, and they are unlikely to be satisfied in econometric applications. However, they do facilitate the development of a limit theory by various methods. In particular, Karlsen, Myklebust, and Tjøstheim (2007) investigated $\hat{f}(x)$ in the situation where x_t is a recurrent Markov chain, allowing for some dependence between x_t and u_t. Under similar conditions and using related Markov chain methods, Schienle (2008) investigated additive nonlinear versions of (2.1) and obtained a limit theory for nonparametric regressions under smooth backfitting. Wang and Phillips (2009, hereafter WP) considered an alternative treatment by making use of local time limit theory and, instead of recurrent Markov chains, worked with partial sum representations of the type $x_t = \sum_{j=1}^t \xi_j$, where ξ_j is a general linear process.

These authors showed that the limit theory for $\hat{f}(x)$ has links to traditional nonparametric asymptotics for stationary models with exogenous regressors even though the rates of convergence are different and typically slower when x_t is nonstationary and the limit theory is mixed normal rather than normal.

In extending this work, it seems particularly important to relax conditions of independence and permit joint determination of x_t and y_t, and to allow for serial dependence in the equilibrium errors u_t and the innovations driving x_t, so that the system is a time series structural model. The goal of the present paper is to do so and to develop a limit theory for structural functional estimation in the context of nonstationary time series that is more in line with the type of assumptions made for parametric linear cointegrated systems.

Throughout the paper we let $\{\epsilon_t\}_{t\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) continuous random variables with $E\epsilon_1 = 0$ and $E\epsilon_1^2 = 1$, and with the characteristic function $\varphi(t)$ of ϵ_1 satisfying $\int_{-\infty}^{\infty} |\varphi(t)|\,dt < \infty$. The sequence $\{\epsilon_t\}_{t\geq 1}$ is assumed to be independent of another i.i.d. random sequence $\{\lambda_t\}_{t\geq 1}$ that enters into the generating mechanism for the equilibrium errors. These two sequences comprise the innovations that drive the time series structure of the model. We use the following assumptions in the asymptotic development.

ASSUMPTION 1: *$x_t = \rho x_{t-1} + \eta_t$, where $x_0 = 0$, $\rho = 1 + \kappa/n$ with κ being 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 $\sum_{k=0}^{\infty} |\phi_k| < \infty$.*

ASSUMPTION 2: *$u_t = u(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m_0+1}, \lambda_t, \lambda_{t-1}, \ldots, \lambda_{t-m_0+1})$ satisfies $Eu_t = 0$ and $Eu_t^4 < \infty$ for $t \geq m_0$, where $u(x_1, \ldots, x_{m_0}, y_1, \ldots, y_{m_0})$ is a real measurable function on R^{2m_0}. We define $u_t = 0$ for $1 \leq t \leq m_0 - 1$.*

ASSUMPTION 3: *$K(x)$ is a nonnegative bounded continuous function satisfying $\int K(x)\,dx < \infty$ and $\int |\hat{K}(x)|\,dx < \infty$, where $\hat{K}(x) = \int e^{ixt} K(t)\,dt$.*

ASSUMPTION 4: *For given x, there exists a real function $f_1(s, x)$ and a $0 < \gamma \leq 1$ such that, when h sufficiently small, $|f(hy + x) - f(x)| \leq h^{\gamma} f_1(y, x)$ for all $y \in R$ and $\int_{-\infty}^{\infty} K(s) f_1(s, x)\,ds < \infty$.*

Assumption 1 allows for both a unit root ($\kappa = 0$) and a near unit root ($\kappa \neq 0$) regressor by virtue of the localizing coefficient κ, and is standard in the near-integrated regression framework (Chan and Wei (1987), Phillips (1987, 1988)). The regressor x_t is then a triangular array formed from a (weighted) partial sum of linear process innovations that satisfy a simple summability condition with long run moving average coefficient $\phi \neq 0$. We remark that in the cointegrating framework, it is conventional to set $\kappa = 0$ so that the regressor is integrated and this turns out to be important in inference. Indeed, in linear parametric cointegration, it is well known (e.g., Elliott (1998)) that near integration ($\kappa \neq 0$) leads to failure of standard cointegration estimation and test procedures. As shown here, no such failures occur under near integration in the nonparametric regression context.

Assumption 2 allows the equation error u_t to be serially dependent and cross-correlated with x_s for $|t-s| < m_0$, thereby inducing endogeneity in the regressor. As a consequence, we may have $\mathrm{cov}(u_t, x_t) \neq 0$. This makes the model in the current paper essentially different from the one investigated in Theorem 3.2 of WP. WP imposed the condition that x_t is adapted to $\mathcal{F}_{t-1}$, where $(u_t, \mathcal{F}_t)$ forms a martingale difference. Hence, under the conditions of WP, one always has $\mathrm{cov}(u_t, x_t) = E[x_t E(u_t \mid \mathcal{F}_{t-1})] = 0$. This difference explains why the proof of the main result in the current paper is so different from that in WP. In WP, we could use a general martingale central limit theorem (CLT) result, but such an approach is not possible in the current framework because the sample covariance function is not a martingale.

In the asymptotic development below, the lag parameter m_0 in Assumption 2 is assumed to be finite, but this could likely be relaxed under some additional conditions and with greater complexity in the proofs, although that is not done here. It is not necessary for u_t to depend on λ_s, in which case there would be only a single innovation sequence. However, in most practical cases involving cointegration between two variables, we can expect that there will be two innovation sequences. While u_t is stationary in Assumption 2, we later discuss some nonstationary cases where the conditional variance of u_t may depend on x_t. Note also that Assumption 2 allows for a nonlinear generating mechanism for the equilibrium error u_t. This seems appropriate in a context where the regression function itself is allowed to take a general nonlinear form.

Assumption 3 places stronger conditions on the kernel function than are usual in kernel estimation, requiring that the Fourier transform of $K(x)$ is integrable. This condition is needed for technical reasons in the proofs and is clearly satisfied for many commonly used kernels, like the normal kernel or kernels that have a compact support.

Assumption 4, which was used in WP, is quite weak and can be verified for various kernels $K(x)$ and regression functions $f(x)$. For instance, if $K(x)$ is a standard normal kernel or has a compact support, a wide range of regression functions $f(x)$ are included. Thus, commonly occurring functions like $f(x) = |x|^\beta$ and $f(x) = 1/(1+|x|^\beta)$ for some $\beta > 0$ satisfy Assumption 4 with $\gamma = \min\{\beta, 1\}$. When $\gamma = 1$, stronger smoothness conditions on $f(x)$ can be used to assist in developing analytic forms for the asymptotic bias function in kernel estimation.

3. MAIN RESULT AND OUTLINE OF THE PROOF

The limit theory for the conventional kernel regression estimate $\hat{f}(x)$ under random normalization turns out to be very simple and is given in the following theorem.

THEOREM 3.1: *For any h satisfying $nh^2 \to \infty$ and $h \to 0$,*

$$\hat{f}(x) \to_p f(x). \tag{3.1}$$

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

$$\left(h \sum_{t=1}^{n} K_h(x_t - x)\right)^{1/2} (\hat{f}(x) - f(x)) \to_D N(0, \sigma^2), \tag{3.2}$$

where $\sigma^2 = E(u_{m_0}^2) \int_{-\infty}^{\infty} K^2(s)\,ds / \int_{-\infty}^{\infty} K(x)\,dx$.

REMARK A: The result (3.1) implies that $\hat{f}(x)$ is a consistent estimate of $f(x)$. Furthermore, as in WP, we may show that

$$\hat{f}(x) - f(x) = o_P\{a_n[h^\gamma + (\sqrt{n}h)^{-1/2}]\}, \tag{3.3}$$

where γ is defined as in Assumption 4 and a_n diverges to infinity as slowly as required. This indicates that a possible "optimal" bandwidth h which yields the best rate in (3.3) or the minimal $E(\hat{f}(x) - f(x))^2$ at least for general γ satisfies

$$h^* \sim a \arg\min_h \{h^\gamma + (\sqrt{n}h)^{-1/2}\} \sim a' n^{-1/[2(1+2\gamma)]},$$

where a and a' are positive constants. In the most common case that $\gamma = 1$, this result suggests a possible optimal bandwidth to be $h^* \sim a' n^{-1/6}$, so that $h = o(n^{-1/6})$ ensures undersmoothing. This is different from nonparametric regression with a stationary regressor, which typically requires $h = o(n^{-1/5})$ for undersmoothing. Under stronger smoothness conditions on $f(x)$, it is possible to develop an explicit expression for the bias function and the weaker condition $h = o(n^{-1/10})$ applies for undersmoothing. Some further discussion and results are given in Remark C and Section 9.

REMARK B: To outline the essentials of the argument in the proof of Theorem 3.1, we split the error of estimation $\hat{f}(x) - f(x)$ as

$$\hat{f}(x) - f(x) = \frac{\sum_{t=1}^{n} u_t K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} + \frac{\sum_{t=1}^{n} [f(x_t) - f(x)] K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]}. \tag{3.4}$$

The result (3.3), which implies (3.1) by letting $a_n = \min\{h^{-\gamma}, (\sqrt{n}h)^{1/2}\}$, will follow if we prove

$$\Theta_{1n} := \sum_{t=1}^{n} u_t K[(x_t - x)/h] = O_P\{(\sqrt{n}h)^{1/2}\}, \tag{3.5}$$

$$\Theta_{2n} := \sum_{t=1}^{n} [f(x_t) - f(x)] K[(x_t - x)/h] = O_P\{\sqrt{n}h^{1+\gamma}\}, \tag{3.6}$$

and if, for any a_n diverging to infinity as slowly as required,

$$\Theta_{3n} := 1 \Big/ \sum_{t=1}^{n} K[(x_t - x)/h] = o_P\{a_n/(\sqrt{n}h)\}. \tag{3.7}$$

On the other hand, it is readily seen that

$$\left(h \sum_{t=1}^{n} K_h(x_t - x)\right)^{1/2} (\hat{f}(x) - f(x)) = \frac{\sum_{t=1}^{n} u_t K[(x_t - x)/h]}{\sqrt{\sum_{t=1}^{n} K[(x_t - x)/h]}} + \Theta_{2n}\sqrt{\Theta_{3n}}.$$

By virtue of (3.6) and (3.7) with $a_n = (nh^{2+4\gamma})^{-1/8}$, we obtain $\Theta_{2n}\sqrt{\Theta_{3n}} \to_P 0$, since $nh^{2+4\gamma} \to 0$. The stated result (3.2) will then follow if we prove

$$\left\{(nh^2)^{-1/4} \sum_{k=1}^{[nt]} u_k K[(x_k - x)/h], (nh^2)^{-1/2} \sum_{k=1}^{n} K[(x_k - x)/h]\right\} \to_D \{d_0 N L^{1/2}(t, 0), d_1 L(1, 0)\} \tag{3.8}$$

on $D[0, 1]^2$, where $d_0^2 = |\phi|^{-1} E(u_{m_0}^2) \int_{-\infty}^{\infty} K^2(s)\,dt$, $d_1 = |\phi|^{-1} \int_{-\infty}^{\infty} K(s)\,ds$, $L(t, 0)$ is the local time process at the origin of the Gaussian diffusion process $\{J_\kappa(t)\}_{t \geq 0}$ defined by

$$J_\kappa(t) = W(t) + \kappa \int_0^t e^{(t-s)\kappa} W(s)\,ds \tag{3.9}$$

and $\{W(t)\}_{t \geq 0}$ being a standard Brownian motion, and where N is a standard normal variate independent of $L(t, 0)$. The local time process $L(t, a)$ is defined by

$$L(t,a)=\lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t I\{|J_\kappa(r)-a|\le\epsilon\}\,dr. \tag{3.10}$$

Indeed, since $P(L(1,0)>0)=1$, the required result (3.2) follows by (3.8) and the continuous mapping theorem. It remains to prove (3.5)–(3.8), which are established in Section 6. As for (3.8), it is clearly sufficient for the required result to show that the finite-dimensional distributions converge in (3.8).

REMARK C: Results (3.2) and (3.8) show that $\hat f(x)$ has an asymptotic distribution that is mixed normal and that this limit theory holds even in the presence of an endogenous regressor. The mixing variate in the limit distribution depends on the local time process $L(1,0)$, as follows from (3.8). Explicitly,

$$(nh^2)^{1/4}(\hat f(x)-f(x))\to_D d_0 d_1^{-1} N L^{-1/2}(1,0) \tag{3.11}$$

whenever $nh^2\to\infty$ and $nh^{2(1+2\gamma)}\to 0$. Again, this is different from nonparametric regression with a stationary regressor. As noticed in WP, in the nonstationary case, the amount of time spent by the process around any particular spatial point is of order $\sqrt{n}$ rather than n, so that the corresponding convergence rate in such regressions is now $\sqrt{\sqrt{n}h}=(nh^2)^{1/4}$, which requires that $nh^2\to\infty$. In effect, the local sample size is $\sqrt{n}h$ in nonstationary regression involving integrated processes, rather than nh as in the case of stationary regression. The condition that $nh^{2(1+2\gamma)}\to 0$ is required to remove bias. This condition can be further relaxed if we add stronger smoothness conditions on $f(x)$ and incorporate an explicit bias term in (3.11). A full development requires further conditions and a very detailed analysis, which we defer to later work. In the simplest case where $\kappa=0$, u_t is a martingale difference sequence with $E(u_t^2)=\sigma_u^2$, u_t is independent of x_t, K satisfies $\int K(y)\,dy=1$, $\int yK(y)\,dy=0$ and has compact support, and f has continuous, bounded third derivatives, it is shown in Section 9 that

$$(nh^2)^{1/4}\left[\hat f(x)-f(x)-\frac{h^2}{2}f''(x)\int_{-\infty}^{\infty}y^2K(y)\,dy\right] \Rightarrow \frac{N\left(0,\sigma_u^2\int_{-\infty}^{\infty}K^2(s)\,ds\right)}{L(1,0)^{1/2}}, \tag{3.12}$$

provided $nh^{14}\to 0$ and $nh^2\to\infty$. Importantly, there is no linear term in the bias function appearing in (3.12), in contrast to the limit theory for local level nonparametric regression for stationary time series.

REMARK D: As is clear from the second member of (3.8), the signal strength in the present kernel regression is $O(\sum_{k=1}^{n} K[(x_k - x)/h]) = O(\sqrt{n}h)$, which gives the local sample size in this case, so consistency requires that the bandwidth h does not pass to zero too fast (viz., $nh^2 \to \infty$). On the other hand, when h tends to zero slowly, estimation bias is manifest even in very large samples. Some illustrative simulations are reported in the next section.

REMARK E: The limiting variance of the (randomly normalized) kernel estimator in (3.2) is simply a scalar multiple of the variance of the equilibrium error, namely $Eu_{m_0}^2$, rather than a conditional variance that depends on $x_t \sim x$, as is commonly the case in kernel regression theory for stationary time series. This difference is explained by the fact that, under Assumption 2, u_t is stationary and, even though u_t is correlated with the shocks $\varepsilon_t, \dots, \varepsilon_{t-m_0+1}$ involved in generating the regressor x_t, the variation of u_t when $x_t \sim x$ is still measured by $Eu_{m_0}^2$ in the limit theory. If Assumption 2 is relaxed to allow for some explicit nonstationarity in the conditional variance of u_t, then this may impact the limit theory. The manner in which the limit theory is affected depends on the form of the conditional variance function. For instance, suppose the equilibrium error is $u_t' = g(x_t)u_t$, where u_t satisfies Assumption 2 and is independent of x_t, and where g is a positive continuous function (e.g., $g(x) = 1/(1 + |x|^\alpha)$ for some $\alpha > 0$). In this case under some additional regularity conditions, modifications to the arguments given in Proposition 7.2 show that the variance of the limit distribution is now given by $\sigma^2(x) = E(u_{m_0}^2)g(x)^2 \int_{-\infty}^{\infty} K^2(s)\,ds / \int_{-\infty}^{\infty} K(x)\,dx$. The limiting variance of the kernel estimator is then simply a scalar multiple of the variance of the equilibrium error, where the scalar depends on $g(x)$.

REMARK F: Theorem 3.1 gives a pointwise result at the value x, while the process x_t itself is recurrent and wanders over the whole real line. For fixed points $x \neq x'$, the kernel cross-product

$$\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right) = o_p(1) \quad \text{for} \quad x \neq x'. \tag{3.13}$$

To show (3.13), note that if $x_t/\sqrt{t}$ has a bounded density $h_t(y)$, as in WP, we have

$$\begin{aligned} &E\left[K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right)\right] \\ &\quad = \int \left[K\left(\frac{\sqrt{t}y - x}{h}\right) K\left(\frac{\sqrt{t}y - x'}{h}\right)\right] h_t(y)\,dy \\ &\quad = ht^{-1/2} \int K(y) K\left[y + \frac{x - x'}{h}\right] h_t\left[\frac{yh + x}{\sqrt{t}}\right] dy \end{aligned}$$

$$\sim ht^{-1/2}h_t(0)\int K(y)K\left[y+\frac{x-x'}{h}\right]dy$$
$$= o\big(ht^{-1/2}\big)$$

whenever $x \neq x'$, $h \to 0$, and $t \to \infty$. Then

$$\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h}\right)K\left(\frac{x_t-x'}{h}\right) = o_p\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{t^{1/2}}\right) = o_p(1).$$

This result and Theorem 2.1 of WP give

$$\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}\begin{bmatrix} K\left(\frac{x_t-x}{h}\right)^2 & K\left(\frac{x_t-x}{h}\right)K\left(\frac{x_t-x'}{h}\right) \\ K\left(\frac{x_t-x}{h}\right)K\left(\frac{x_t-x'}{h}\right) & K\left(\frac{x_t-x'}{h}\right)^2 \end{bmatrix}$$
$$\Rightarrow L(1,0)\begin{bmatrix} \int K(s)^2\,ds & 0 \\ 0 & \int K(s)^2\,ds \end{bmatrix}.$$

Following the same line of argument as in the proof of Theorem 3.2 of WP, it follows that in the special case where u_t is a martingale difference sequence independent of x_t, the regression ordinates $(\hat{f}(x), \hat{f}(x'))$ have a mixed normal limit distribution with diagonal covariance matrix. The ordinates are then asymptotically conditionally independent given the local time $L(1,0)$. Extension of this theory to the general case where u_t and x_t are dependent involves more complex limit theory and is left for later work.

REMARK G: The error variance term $Eu_{m_0}^2$ in the limit distribution (3.2) may be estimated by a localized version of the usual residual based method. Indeed, by letting

$$\hat{\sigma}_n^2 = \frac{\sum_{t=1}^{n}[y_t-\hat{f}(x)]^2K_h(x_t-x)}{\sum_{t=1}^{n}K_h(x_t-x)}, \tag{3.14}$$

we have the following theorem under minor additional conditions.

THEOREM 3.2: *In addition to Assumptions 1–4*, $Eu_{m_0}^8 < \infty$ *and* $\int_{-\infty}^{\infty} K(s) f_1^2(s, x)\,ds < \infty$ *for given* x. *Then, for any* h *satisfying* $nh^2 \to \infty$ *and* $h \to 0$,

$$\hat{\sigma}_n^2 \to_p Eu_{m_0}^2. \tag{3.15}$$

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

$$(nh^2)^{1/4}\big(\hat{\sigma}_n^2 - Eu_{m_0}^2\big) \to_D \sigma_1 N L^{-1/2}(1, 0), \tag{3.16}$$

where N *and* $L(1, 0)$ *are defined as in* (3.8) *and*

$$\sigma_1^2 = E\big(u_{m_0}^2 - Eu_{m_0}^2\big)^2 \int_{-\infty}^{\infty} K^2(s)\,ds \Big/ \int_{-\infty}^{\infty} K(x)\,dx.$$

While the estimator $\hat{\sigma}_n^2$ is constructed from the regression residuals $y_t - \hat{f}(x)$, it is also localized at x because of the action of the kernel function $K_h(x_t - x)$ in (3.14). Note, however, that in the present case the limit theory for $\hat{\sigma}_n^2$ is not localized at x. In particular, the limit of $\hat{\sigma}_n^2$ is the unconditional variance $Eu_{m_0}^2$, not a conditional variance, and the limit distribution of $\hat{\sigma}_n^2$ given in (3.16) depends only on the local time $L(1, 0)$ of the limit process at the origin, not on the precise value of x. The explanation is that conditioning on the neighborhood $x_t \sim x$ is equivalent to $x_t/\sqrt{n} \sim x/\sqrt{n}$ or $x_t/\sqrt{n} \sim 0$, which translates into the local time of the limit process of x_t at the origin irrespective of the given value of x. For the same reason, as discussed in Remark E above, the limit distribution of the kernel regression estimator given in (3.2) depends on the variance $Eu_{m_0}^2$. However, as discussed in Remark E, in the more general context where there is nonstationary conditional heterogeneity, the limit of $\hat{\sigma}_n^2$ may be correspondingly affected. For instance, in the case considered there where $u_t' = g(x_t)u_t$, u_t satisfies Assumption 2, and g is a positive continuous function, we find that $\hat{\sigma}_n^2 \to_p Eu_{m_0}^2 g(x)^2$.

REMARK H: In parametric cointegrating regression, techniques such as FM regression (Phillips and Hansen (1990)) have been developed to eliminate the second order bias effects in the limit theory that arise from endogeneity, thereby improving upon simple least squares regression. These techniques typically augment the regression equation to address the effects of endogeneity by adjusting for the (long run) conditional mean. Interestingly, there is no need to augment the regression equation in this way to achieve bias reduction in nonparametric cointegrating regression. To illustrate, we take the case where the regressor follows the simple model $x_t = x_{t-1} + \epsilon_t$ and $E(u_t \mid \epsilon_t) = \lambda \epsilon_t$.

The augmented regression equation is, say,

$$y_t = f(x_t) + E(u_t|\epsilon_t) + (u_t - E(u_t|\epsilon_t)) = f(x_t) + \lambda \Delta x_t + u_{y.\epsilon,t}, \tag{3.17}$$

which is the nonparametric analogue of the augmented regression used in linear cointegration models. From observations on Δx_t and the residuals $\hat{u}_t = y_t - \hat{f}(x_t)$, where $\hat{f}$ is defined by (2.2), the first stage nonparametric estimate of f in model (2.1), the least squares estimate of λ is given by

$$\hat{\lambda} = \frac{\sum_{t=1}^{n}[y_t - \hat{f}(x_t)]\Delta x_t}{\sum_{t=1}^{n}\Delta x_t^2}.$$

Using $\hat{\lambda}$ in place of λ in (3.17) and conventional kernel regression to estimate f in this equation, we have the following nonparametric augmented regression estimate of f

$$\hat{f}_a(x) = \frac{\sum_{t=1}^{n}(y_t - \hat{\lambda}\Delta x_t)K_h(x_t - x)}{\sum_{t=1}^{n}K_h(x_t - x)}. \tag{3.18}$$

We now show that the limit distribution of $\hat{f}_a(x)$ is the same as that of $\hat{f}(x)$ except for a scale variance effect that arises from the adjusted error term $u_{y.\epsilon,t}$ in (3.17). Hence, there is no bias reduction in the use of the augmented regression equation (3.17), unlike linear parametric cointegration.

Indeed we have the following theorem.

THEOREM 3.3: *In addition to Assumptions 1–4, assume* $E(u_{m_0}^2\epsilon_{m_0}^2) < \infty$, $K(x)$ *has a compact support, and* $f(x)$ *satisfies* $|f(x) - f(y)| \le C|x - y|$ *whenever* $x - y$ *is sufficiently small. Then, for any* h *satisfying* $nh^2 \to \infty$ *and* $h \to 0$,

$$\hat{\lambda} \to_p \lambda. \tag{3.19}$$

Furthermore, for any h *satisfying* $hn^{1/2+\delta_0} \to \infty$ *and* $nh^6 \to 0$, *where* $0 < \delta_0 < 1$,

$$\left(h\sum_{t=1}^{n}K_h(x_t - x)\right)^{1/2}(\hat{f}_a(x) - f(x)) \to_D N(0, \sigma_a^2), \tag{3.20}$$

where $\sigma_a^2 = E(u_{y.\epsilon,m_0}^2)\int_{-\infty}^{\infty}K^2(s)\,ds / \int_{-\infty}^{\infty}K(x)\,dx$.

Thus, the effect of estimating the augmented regression equation in (3.17) is to deliver a scale variance reduction that corresponds to $E(u_{y.\epsilon,t}^2) \le E(u_t^2)$ in the limit theory for the estimator $\hat{f}_a(x)$. The variance reduction results from the inclusion of the stationary regressor Δx_t in (3.17). There is no bias reduction, unlike the case of linear parametric cointegration. The same result and the same limit theory apply for the infeasible kernel estimate

$$\tilde{f}_a(x) = \frac{\sum_{t=1}^{n}(y_t - \lambda \Delta x_t) K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]}, \tag{3.21}$$

where λ is assumed known.

4. SIMULATIONS

This section reports the results of a simulation experiment investigating the finite sample performance of the kernel regression estimator. The generating mechanism follows (2.1) and has the explicit form

$$y_t = f(x_t) + \sigma u_t, \quad \Delta x_t = \epsilon_t,$$
$$u_t = (\lambda_t + \theta \epsilon_t)/(1 + \theta^2)^{1/2},$$

where (ϵ_t, λ_t) are i.i.d. $N(0, I_2)$ and $x_0 = 0$. The following two regression functions were used in the simulations:

$$f_A(x) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \sin(j\pi x)}{j^2}, \quad f_B(x) = x^3.$$

The first function corresponds (up to a scale factor) to the function used in Hall and Horowitz (2005) and is truncated at $j = 4$ for computation. Figures 1 and 2 graph these functions (the solid lines) and the mean simulated kernel estimates (broken lines) over the intervals $[0, 1]$ and $[-1, 1]$ for kernel estimates of f_A and f_B, respectively. Bias, variance, and mean squared error for the estimates were computed on the grid of values $\{x = 0.01k; k = 0, 1, \dots, 100\}$ for $[0, 1]$ and $\{x = -1 + 0.02k; k = 0, 1, \dots, 100\}$ for $[-1, 1]$ based on 20,000 replications. Simulations were performed for $\theta = 0.2$ (weak endogeneity, $\mathrm{corr}(u_t, \epsilon_t) = 0.2$) and $\theta = 2.0$ (strong endogeneity, $\mathrm{corr}(u_t, \epsilon_t) = 0.9$) for $\sigma = 0.2$ and for the sample size $n = 500$. An Epanechnikov kernel was used with bandwidths $h = n^{-10/18}, n^{-1/2}, n^{-1/3}, n^{-1/5}$.

Table I shows the performance of the regression estimates $\hat{f}$, $\hat{f}_a$, and $\tilde{f}_a$ computed for various bandwidths, two of which $(n^{-1/3}, n^{-1/5})$ satisfy and two of which $(n^{-10/18}, n^{-1/2})$ violate the condition $nh^2 \to \infty$ of Theorem 3.1, thereby showing the effects of bandwidth on estimation. While smaller bandwidths may

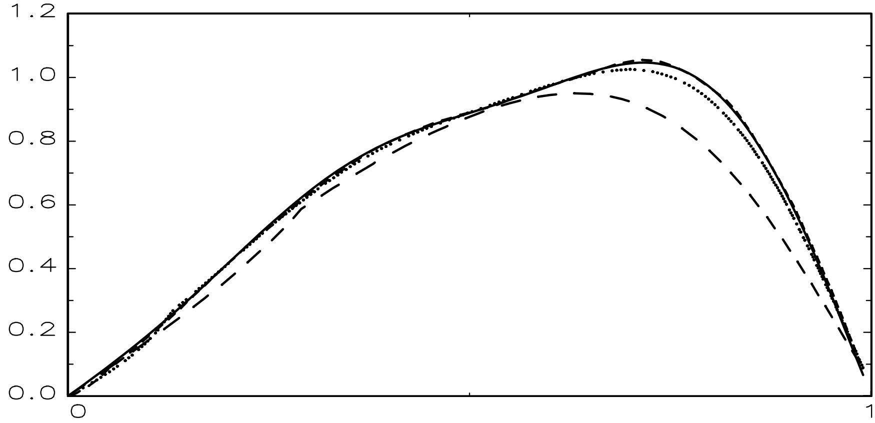

FIGURE 1.—Graphs over the interval [0, 1] of $f_A(x)$ and Monte Carlo estimates of $E(\hat{f}_A(x))$ for $h = n^{-1/2}$ (short dashes), $h = n^{-1/3}$ (dotted), and $h = n^{-1/5}$ (long dashes) with $\theta = 2$, $\sigma = 0.2$, and $n = 500$.

reduce bias, when $h \to 0$ so fast that $nh^2 \not\to \infty$, the "signal" $\sum_{k=1}^{n} K[(x_k - x)/h]$ no longer diverges and the estimate $\hat{f}$ is inconsistent (see (3.8)). Also, since x_t is recurrent and wanders over the real line, some simulations are inevitably thin in subsets of the chosen domains (as in the simulation design

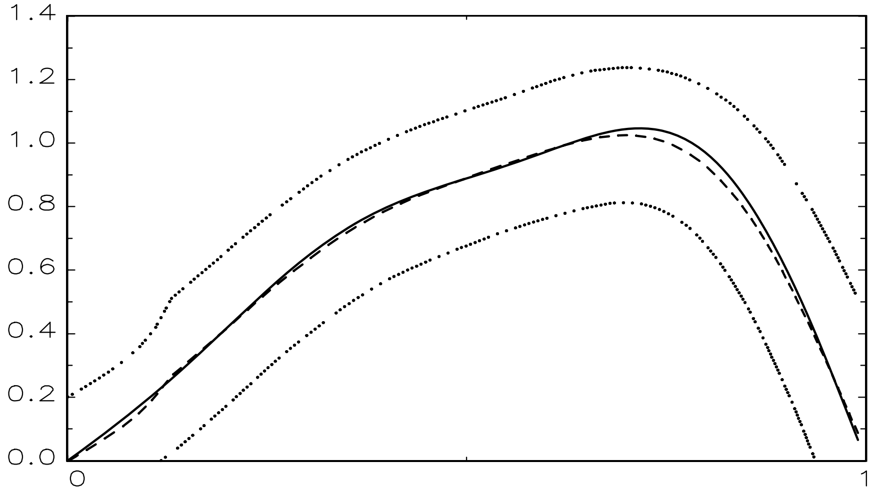

FIGURE 2.—Graphs over the interval [0, 1] of estimation bands for $f_A(x)$ (solid line), the Monte Carlo estimate of $E(f_A(x))$ for $h = n^{1/3}$ (short dashes), and 95% estimation bands (dotted) with $\theta = 2$, $\sigma = 0.2$, and $n = 500$.

TABLE I

SUMMARY COMPARISONS (AVERAGED OVER THE GRIDS DESCRIBED IN THE TEXT) OF LOCAL LEVEL, FEASIBLE AUGMENTED, AND INFEASIBLE AUGMENTED NONPARAMETRIC ESTIMATES

		$\hat{f}(x)$			$\tilde{f}_a(x)$			$\hat{f}_a(x)$		
θ	h	Bias	Std	MSE	Bias	Std	MSE	Bias	Std	MSE
					Model A: $f_A(x) = \sum_{j=1}^4 \frac{(-1)^{j+1}\sin(j\pi x)}{j^2}$					
2	$n^{-10/18}$	−0.005	0.129	0.017	0.000	0.064	0.004	−0.001	0.069	0.005
	$n^{-1/2}$	0.000	0.126	0.016	0.001	0.063	0.004	0.001	0.070	0.005
	$n^{-1/3}$	0.011	0.122	0.016	0.015	0.076	0.008	0.014	0.083	0.009
	$n^{-1/5}$	0.067	0.141	0.032	0.073	0.114	0.027	0.071	0.117	0.027
0.2	$n^{-10/18}$	−0.001	0.135	0.018	−0.000	0.133	0.018	−0.000	0.133	0.018
	$n^{-1/2}$	0.001	0.128	0.017	0.001	0.126	0.016	0.002	0.126	0.016
	$n^{-1/3}$	0.015	0.122	0.016	0.016	0.120	0.015	0.016	0.121	0.016
	$n^{-1/5}$	0.072	0.141	0.033	0.073	0.139	0.033	0.072	0.140	0.033
					Model B: $f_B(x) = x^3$					
2	$n^{-10/18}$	0.000	0.125	0.164	0.000	0.058	0.004	−0.003	0.153	0.028
	$n^{-1/2}$	0.001	0.119	0.015	0.000	0.056	0.003	−0.000	0.171	0.030
	$n^{-1/3}$	0.000	0.104	0.011	0.000	0.055	0.003	−0.001	0.371	0.138
	$n^{-1/5}$	0.000	0.105	0.011	0.004	0.069	0.007	−0.002	0.491	0.244
0.2	$n^{-10/18}$	0.001	0.127	0.017	0.001	0.124	0.016	−0.002	0.188	0.039
	$n^{-1/2}$	−0.001	0.121	0.015	−0.001	0.119	0.014	−0.002	0.203	0.042
	$n^{-1/3}$	0.000	0.104	0.011	0.000	0.102	0.010	−0.001	0.381	0.146
	$n^{-1/5}$	0.001	0.102	0.012	0.000	0.102	0.012	−0.002	0.497	0.249

here) and this inevitably affects performance due to the small "local" sample size.

The results in Table I show that in both models the degree of endogeneity (θ) in the regressor has a negligible effect on the properties of the kernel regression estimate $\hat{f}$, although estimation bias does increase in Model A when the bandwidth is $h = n^{-1/5}$, which is the conventional rate for stationary series. For both models, finite sample performance of $\hat{f}$ in terms of mean squared error (MSE) seems to be optimized for h around $n^{-1/3}$.

Table I also enables a comparison between the local level estimate $\hat{f}$ and the feasible and infeasible augmented regression estimates $\hat{f}_a(x)$ and $\tilde{f}_a(x)$. The infeasible estimate has uniformly smaller variance than $\hat{f}(x)$, as may be expected from asymptotic theory, and it also has smaller variance than $\hat{f}_a(x)$, resulting from the finite sample effects of estimating λ in the latter. For Model A and the case of strong endogeneity, the variance of the feasible estimate $\hat{f}_a(x)$ is considerably smaller than that of $\hat{f}(x)$, so the feasible procedure has a clear advantage in this case. But these gains do not appear under weak endogeneity or under either weak or strong endogeneity for

Model B. The variance of the feasible estimate $\hat{f}_a(x)$ is, in fact, much larger than that of $\hat{f}(x)$ for Model B. Both $\hat{f}_a(x)$ and $\tilde{f}_a(x)$ display negligible bias, just as $\hat{f}(x)$, but there is no apparent gain in terms of bias reduction from the use of the augmented regression estimates, including the infeasible estimate. These results indicate that feasible nonparametric estimation of the augmented regression equation (3.17), leading to $\hat{f}_a(x)$, does not dominate the simple nonparametric regression estimate $\hat{f}(x)$ in terms of bias. Neither does $\hat{f}_a(x)$ always dominate $\hat{f}(x)$ in terms of variance in finite samples, even though asymptotic theory suggests an improvement and the equation error in (3.17) has smaller variance than that of (2.1). This outcome contrasts with linear parametric cointegrating regression, where it is generally beneficial—and near universal empirical practice—to fit the augmented regression equation.

Figures 1 and 2 show results for the Monte Carlo approximations to $E(\hat{f}_A(x))$ and $E(\hat{f}_B(x))$ corresponding to bandwidths $h = n^{-1/2}$ (broken line), $h = n^{-1/3}$ (dotted line), and $h = n^{-1/5}$ (dashed and dotted line) for $\theta = 2$. Figures 3 and 4 show the Monte Carlo approximations to $E(\hat{f}_A(x))$ and $E(\hat{f}_B(x))$ together with a 95% pointwise "estimation band." As in Hall and Horowitz (2005), these bands connect points $f(x_j \pm \delta_j)$, where each δ_j is chosen so that the interval $[f(x_j) - \delta_j, f(x_j) + \delta_j]$ contains 95% of the 10,000 simulated values of $\hat{f}(x_j)$ for Models A and B, respectively. Apparently, the bands can be wide, reflecting the slower rate of convergence of the kernel estimate $\hat{f}(x)$ in the nonstationary case. In particular, since x_t spends only $\sqrt{n}$ of its time in the neighborhood of any specific point, the effective sample

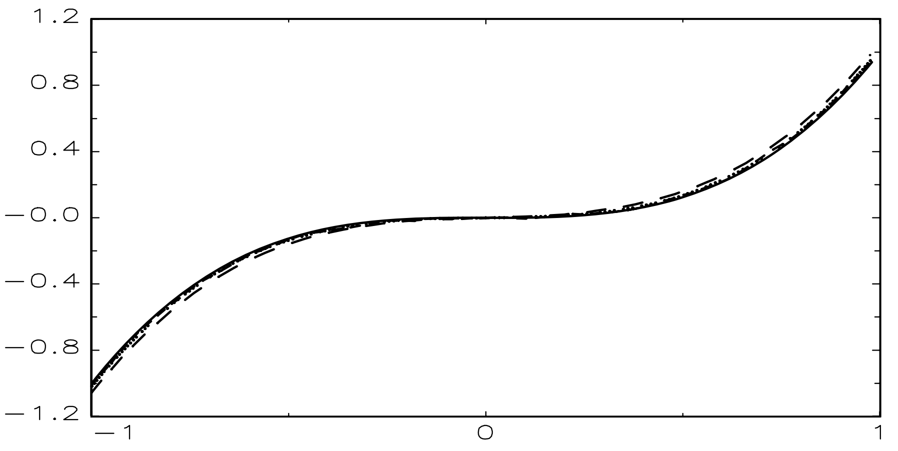

FIGURE 3.—Graphs of $f_B(x)$ and Monte Carlo estimates of $E(\hat{f}_B(x))$ for $h = n^{-1/2}$ (short dashes), $h = n^{-1/3}$ (dotted), and $h = n^{-1/5}$ (long dashes) with $6 = 2$, $\sigma = 0.2$, and $n = 500$.

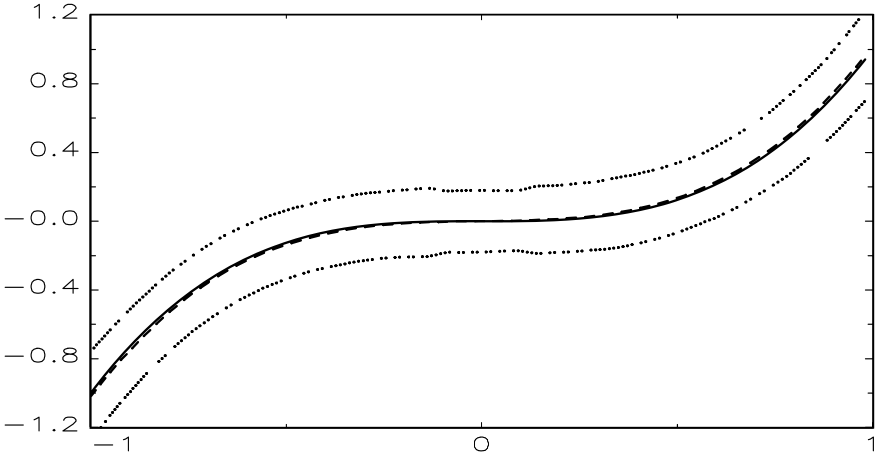

FIGURE 4.—Graphs of estimation bands for $f_B(x)$ (solid line), the Monte Carlo estimate of $E(\hat{f}_B(x))$ for $h = n^{-1/3}$ (short dashes), and 95% estimation bands (dotted) with $6 = 2$, $\sigma = 0.2$, and $n = 500$.

size for pointwise estimation purposes is $\sqrt{500} \sim 22$. When $h = n^{-1/3}$, it follows from Theorem 3.1 that the convergence rate is $(nh^2)^{1/4} = n^{1/12}$, which is much slower than the rate $(nh)^{1/2} = n^{2/5}$ for conventional kernel regression.

Using Theorems 3.1 and 3.2, an asymptotic $100(1-\alpha)\%$ level confidence interval for $f(x)$ is given by

$$\hat{f}(x) \pm z_{\alpha/2}\left(\frac{\hat{\sigma}_n^2 \mu_{K^2}/\mu_K}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right)}\right)^{1/2},$$

where $\mu_{K^2} = \int_{-\infty}^{\infty} K^2(s)\,ds$, $\mu_K = \int_{-\infty}^{\infty} K(s)\,ds$, and $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ using the standard normal cumulative distribution function (c.d.f.) Φ. Figures 5 and 6 show the empirical coverage probabilities of these pointwise asymptotic confidence intervals for f_A and f_B over 100 equispaced points on the domains [0, 1] and [−1, 1], using an Epanechnikov kernel, various bandwidths as shown, and setting $\alpha = 0.05$ and $n = 500$. For both functions, the coverage rates are more uniform over the respective domains for the smaller bandwidth choices, but the undercoverage also increases as the bandwidth gets smaller. For both f_A and f_B there is evidence of substantial undercoverage with around 60% coverage over most of the domain when $h = n^{-1/3}$ and around 70% coverage when $h = n^{-1/4}$. Coverage is higher (around 80%) when $h = n^{-1/5}$, but for function f_A dips to below 60% in the region (around $x \sim 0.7$) where the nonparametric estimator is most biased for larger bandwidths (see Figure 1).

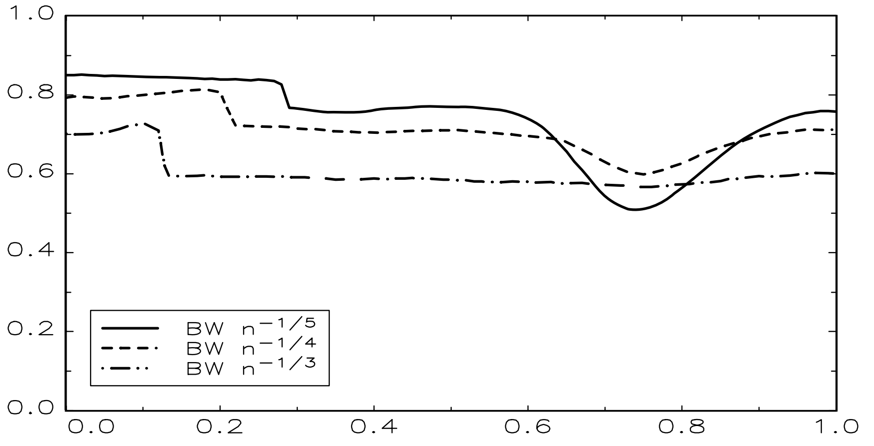

FIGURE 5.—Coverage probabilities of (nominal 95%) confidence intervals for $f_A(x)$ over [0, 1] for different bandwidths.

5. CONCLUSION

The main results in the paper have many implications. First, there is no inverse problem in structural models of nonlinear cointegration of the form (2.1), where the regressor is an endogenously generated integrated or near-integrated process. This result reveals a major simplification in structural non-parametric regression in cointegrating models, avoiding the need for instrumentation and eliminating ill-posed functional equation inversions. Second,

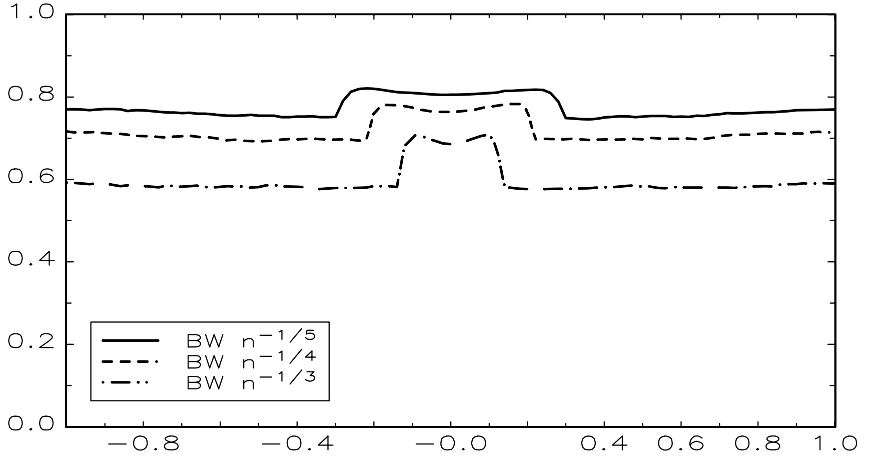

FIGURE 6.—Coverage probabilities of (nominal 95%) confidence intervals for $f_B(x)$ over [0, 1] for different bandwidths.

functional estimation of (2.1) is straightforward in practice and may be accomplished by standard kernel methods with no modification. These methods yield consistent estimates that have a mixed normal limit distribution, thereby validating conventional methods of inference in the nonstationary nonparametric setting. Third, the methods are applicable without change when the regressor is near-integrated with a root local to unity rather than unity, providing a major departure from the parametric case where near integration presents substantial difficulties in inference.

The results open up interesting possibilities for functional regression in empirical research with integrated and near-integrated processes. There are some possible extensions of the ideas presented here to other models involving nonlinear functions of integrated processes. In particular, additive nonlinear cointegration models (cf. Schienle (2008)) and partial linear cointegration models may be treated in a similar way to (2.1). But multiple nonadditive regression models do present difficulties arising from the nonrecurrence of the limit processes in high dimensions (cf. Park and Phillips (2001)). There are issues of specification testing, functional form tests, and cointegration tests, which may now be addressed using these methods. Also, the simulations reported here indicate that there is a need to improve confidence interval coverage probabilities in the use of these nonparametric methods. We hope to report progress on some of these issues in later work.

6. PROOF OF THEOREM 3.1

As shown in Remark B, the proof of the theorem essentially amounts to proving (3.5)–(3.8). To do so, we will make use of various subsidiary results which are proved here and in the next section.

First, it is convenient to introduce the following definitions and notation. If $\alpha_n^{(1)}, \alpha_n^{(2)}, \dots, \alpha_n^{(k)}$ $(1 \le n \le \infty)$ are random elements of $D[0,1]$, we will understand the condition

$$\left(\alpha_n^{(1)}, \alpha_n^{(2)}, \dots, \alpha_n^{(k)}\right) \to_D \left(\alpha_\infty^{(1)}, \alpha_\infty^{(2)}, \dots, \alpha_\infty^{(k)}\right)$$

to mean that for all $\alpha_\infty^{(1)}, \alpha_\infty^{(2)}, \dots, \alpha_\infty^{(k)}$ continuity sets $A_1, A_2, \dots, A_k$,

$$\begin{aligned} &P\left(\alpha_n^{(1)} \in A_1, \alpha_n^{(2)} \in A_1, \dots, \alpha_n^{(k)} \in A_k\right) \\ &\quad \to P\left(\alpha_\infty^{(1)} \in A_1, \alpha_\infty^{(2)} \in A_2, \dots, \alpha_\infty^{(k)} \in A_k\right) \end{aligned}$$

(see Billingsley (1968, Theorem 3.1) or Hall (1977)). $D[0,1]^k$ will be used to denote $D[0,1] \times \cdots \times D[0,1]$, the k-times coordinate product space of $D[0,1]$. We still use $\Rightarrow$ to denote weak convergence on $D[0,1]$.

To prove (3.8), we use the following lemma.

LEMMA 6.1: *Suppose that* $\{\mathcal{F}_t\}_{t\ge 0}$ *is an increasing sequence of* σ*-fields*, $q(t)$ *is a process that is* $\mathcal{F}_t$*-measurable for each* t *and continuous with probability* 1,

$Eq^2(t) < \infty$, and $q(0) = 0$. Let $\psi(t)$, $t \ge 0$, be a process that is nondecreasing and continuous with probability 1, and satisfies $\psi(0) = 0$ and $E\psi^2(t) < \infty$. Let ξ be a random variable which is $\mathcal{F}_t$-measurable for each $t \ge 0$. If, for any $\gamma_j \ge 0$, $j = 1, 2, \dots, r$, and any $0 \le s < t \le t_0 < t_1 < \cdots < t_r < \infty$,

$$E\left(\exp\left(-\sum_{j=1}^{r}\gamma_j[\psi(t_j) - \psi(t_{j-1})]\right)[q(t) - q(s)] \,\Big|\, \mathcal{F}_s\right) = 0 \quad a.s.,$$

$$E\left(\exp\left(-\sum_{j=1}^{r}\gamma_j[\psi(t_j) - \psi(t_{j-1})]\right)\right.$$
$$\left. \times \left\{[q(t) - q(s)]^2 - [\psi(t) - \psi(s)]\right\} \,\Big|\, \mathcal{F}_s\right) = 0 \quad a.s.,$$

then the finite-dimensional distributions of the process $(q(t), \xi)_{t\ge0}$ coincide with those of the process $(W[\psi(t)], \xi)_{t\ge0}$, where $W(s)$ is a standard Brownian motion with $EW^2(s) = s$ independent of $\psi(t)$.

PROOF: This lemma is an extension of Theorem 3.1 of Borodin and Ibragimov (1995, p. 14) and the proof follows the same lines as in their work. Indeed, by using the fact that ξ is $\mathcal{F}_t$-measurable for each $t \ge 0$, it follows from the same arguments as in the proof of Theorem 3.1 of Borodin and Ibragimov (1995) that, for any $t_0 < t_1, \dots, t_r < \infty$, $\alpha_j \in R$ and $s \in R$,

$$E\exp\left(i\sum_{j=1}^{r}\alpha_j[q(t_j) - q(t_{j-1})] + is\xi\right)$$
$$= E\left[\exp\left(i\sum_{j=1}^{r-1}\alpha_j[q(t_j) - q(t_{j-1})] + is\xi\right)\right.$$
$$\left.\times E\big(\exp\big(i\alpha_r[q(t_r) - q(t_{r-1})]\big) \mid \mathcal{F}_{t_{r-1}}\big)\right]$$
$$= E\left[\exp\left(-\frac{\alpha_r^2}{2}[\psi(t_r) - \psi(t_{r-1})]\right)\right.$$
$$\left.\times \exp\left(i\sum_{j=1}^{r-1}\alpha_j[q(t_j) - q(t_{j-1})] + is\xi\right)\right]$$
$$= \cdots = E\exp\left(-\frac{\alpha_r^2}{2}\sum_{j=1}^{r}[\psi(t_j) - \psi(t_{j-1})] + is\xi\right),$$

which yields the stated result. *Q.E.D.*

By virtue of Lemma 6.1, we now obtain the proof of (3.8). Technical details of some subsidiary results that are used in this proof are given in the next section. Set

$$\zeta_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \epsilon_k, \quad \psi'_n(t) = \frac{1}{d_1\sqrt{nh^2}} \sum_{k=1}^{[nt]} K\left[\frac{x_k - x}{h}\right],$$

$$S_n(t) = \frac{1}{d_0(nh^2)^{1/4}} \sum_{k=1}^{[nt]} u_k K\left[\frac{x_k - x}{h}\right],$$

$$\psi_n(t) = \frac{1}{d_0^2\sqrt{nh^2}} \sum_{k=1}^{[nt]} u_k^2 K^2\left[\frac{x_k - x}{h}\right]$$

for $0 \le t \le 1$, where d_0 and d_1 are defined as in (3.8).

We will prove in Propositions 7.1 and 7.2 that $\zeta_n(t) \Rightarrow W(t)$, $\psi'_n(t) \Rightarrow \psi(t)$, and $\psi_n(t) \Rightarrow \psi(t)$ on $D[0,1]$, where $\psi(t) := L(t,0)$. Furthermore we will prove in Proposition 7.4 that $\{S_n(t)\}_{n\ge1}$ is tight on $D[0,1]$. These facts imply that

$$\{S_n(t), \psi_n(t), \psi'_n(t), \zeta_n(t)\}_{n\ge1}$$

is tight on $D[0,1]^4$. Hence, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

$$\{S_{n''}(t), \psi_{n''}(t), \psi'_{n''}(t), \zeta_{n''}(t)\} \to_d \{\eta(t), \psi(t), \psi(t), W(t)\} \tag{6.1}$$

on $D[0,1]^4$, where $\eta(t)$ is a process continuous with probability 1 by noting (7.26) below. Write $\mathcal{F}_s = \sigma\{W(t), 0 \le t \le 1; \eta(t), 0 \le t \le s\}$. It is readily seen that $\mathcal{F}_s \uparrow$ and $\eta(s)$ is $\mathcal{F}_s$-measurable for each $0 \le s \le 1$. Also note that $\psi(t)$ (for any fixed $t \in [0,1]$) is $\mathcal{F}_s$-measurable for each $0 \le s \le 1$. If we prove that for any $0 \le s < t \le 1$,

$$E\big([\eta(t) - \eta(s)] \mid \mathcal{F}_s\big) = 0 \quad \text{a.s.,} \tag{6.2}$$

$$E\big(\big\{[\eta(t) - \eta(s)]^2 - [\psi(t) - \psi(s)]\big\} \mid \mathcal{F}_s\big) = 0 \quad \text{a.s.,} \tag{6.3}$$

then it follows from Lemma 6.1 that the finite-dimensional distributions of $(\eta(t), \psi(1))$ coincide with those of $\{NL^{1/2}(t,0), L(1,0)\}$, where N is normal variate independent of $L(t,0)$. The result (3.8) therefore follows, since $\eta(t)$ does not depend on the choice of the subsequence.

Let $0 \le t_0 < t_2 < \cdots < t_r = 1$, let r be an arbitrary integer, and let $G(\cdots)$ be an arbitrary bounded measurable function. To prove (6.2) and (6.3), it suffices to show that

$$E[\eta(t_j) - \eta(t_{j-1})]G[\eta(t_0), \dots, \eta(t_{j-1}); W(t_0), \dots, W(t_r)] = 0, \tag{6.4}$$

$$E\{[\eta(t_j) - \eta(t_{j-1})]^2 - [\psi(t_j) - \psi(t_{j-1})]\} \times G[\eta(t_0), \dots, \eta(t_{j-1}); W(t_0), \dots, W(t_r)] = 0. \tag{6.5}$$

Recall (6.1). Without loss of generality, we assume the sequence $\{n''\}$ is just $\{n\}$ itself. Since $S_n(t), S_n^2(t)$, and $\psi_n(t)$, for each $0 \le t \le 1$ are uniformly integrable (see Proposition 7.3), the statements (6.4) and (6.5) will follow if we prove

$$E[S_n(t_j) - S_n(t_{j-1})]G[\cdots] \to 0, \tag{6.6}$$

$$E\{[S_n(t_j) - S_n(t_{j-1})]^2 - [\psi_n(t_j) - \psi_n(t_{j-1})]\}G[\cdots] \to 0, \tag{6.7}$$

where $G[\cdots] = G[S_n(t_0), \dots, S_n(t_{j-1}); \zeta_n(t_0), \dots, \zeta_n(t_r)]$ (see, e.g., Theorem 5.4 of Billingsley (1968)). Furthermore, by using similar arguments to those in the proofs of Lemmas 5.4 and 5.5 in Borodin and Ibragimov (1995), we may choose

$$G(y_0, y_1, \dots, y_{j-1}; z_0, z_1, \dots, z_r) = \exp\left\{i\left(\sum_{k=0}^{j-1} \lambda_k y_k + \sum_{k=0}^{r} \mu_k z_k\right)\right\}.$$

Therefore, by independence of ϵ_k, we only need to show that

$$E\left\{\sum_{k=[nt_{j-1}]+1}^{[nt_j]} u_k K[(x_k - x)/h] \exp\bigl(i\mu_j^*[\zeta_n(t_j) - \zeta_n(t_{j-1})] + i\chi(t_{j-1})\bigr)\right\} = o\bigl[(nh^2)^{1/4}\bigr], \tag{6.8}$$

$$E\left\{\left[\sum_{k=[nt_{j-1}]+1}^{[nt_j]} u_k K[(x_k - x)/h]\right]^2 - \sum_{k=[nt_{j-1}]+1}^{[nt_j]} u_k^2 K^2[(x_k - x)/h]\right\} \times \exp\bigl(i\mu_j^*[\zeta_n(t_j) - \zeta_n(t_{j-1})] + i\chi(t_{j-1})\bigr) = o\bigl[(nh^2)^{1/2}\bigr], \tag{6.9}$$

where $\chi(s) = \chi(x_1, \dots, x_s, u_1, \dots, u_s)$, a functional of $x_1, \dots, x_s, u_1, \dots, u_s$, and $\mu_j^* = \sum_{k=j}^{r} \mu_k$.

Note that $\chi(s)$ depends only on $(\ldots, \epsilon_{s-1}, \epsilon_s)$ and $\lambda_1, \ldots, \lambda_s$, and we may write

$$
\begin{aligned}
(6.10)\qquad x_t &= \sum_{j=1}^{t} \rho^{t-j}\eta_j = \sum_{j=1}^{t} \rho^{t-j} \sum_{i=-\infty}^{j} \epsilon_i \phi_{j-i} \\
&= \rho^{t-s} x_s + \sum_{j=s+1}^{t} \rho^{t-j} \sum_{i=-\infty}^{s} \epsilon_i \phi_{j-i} + \sum_{j=s+1}^{t} \rho^{t-j} \sum_{i=s+1}^{j} \epsilon_i \phi_{j-i} \\
&:= x_{s,t}^{*} + x_{s,t}',
\end{aligned}
$$

where $x_{s,t}^{*}$ depends only on $(\ldots, \epsilon_{s-1}, \epsilon_s)$ and

$$
x_{s,t}' = \sum_{j=1}^{t-s} \rho^{t-j-s} \sum_{i=1}^{j} \epsilon_{i+s} \phi_{j-i} = \sum_{i=s+1}^{t} \epsilon_i \sum_{j=0}^{t-i} \rho^{t-j-i} \phi_j .
$$

Now, by independence of ϵ_k again and conditioning arguments, it suffices to show that, for any μ,

$$
\begin{aligned}
(6.11)\qquad & \sup_{y, 0 \le s < m \le n} E\left\{ \sum_{k=s+1}^{m} u_k K[(y + x_{s,k}')/h] \exp\left(i\mu \sum_{i=1}^{m} \epsilon_i/\sqrt{n} \right) \right\} \\
& \quad = o\big[(nh^2)^{1/4}\big],
\end{aligned}
$$

$$
\begin{aligned}
(6.12)\qquad & \sup_{y, 0 \le s < m \le n} E\left(\left\{ \sum_{k=s+1}^{m} u_k K[(y + x_{s,k}')/h] \right\}^2 - \sum_{k=s+1}^{m} u_k^2 K^2[(y + x_{s,k}')/h] \right) \\
& \qquad \times \exp\left(i\mu \sum_{i=1}^{m} \epsilon_i/\sqrt{n} \right) \\
& \quad = o\big[(nh^2)^{1/2}\big].
\end{aligned}
$$

This follows from Proposition 7.5. The proof of (3.8) is now complete.

We next prove (3.5)–(3.7). In fact, it follows from Proposition 7.3 that, uniformly in n, $E\Theta_{1n}^2/(nh^2)^{1/2} = d_0^2 E S_n^2(1) \le C$. This yields (3.5) by the Markov's inequality. It follows from Claim 1 in the proof of Proposition 7.2 that $x_t/\sqrt{n}\phi$ satisfies Assumption 2.3 of WP. The same argument as in the proof of (5.18) in WP yields (3.6). As for (3.7), it follows from Proposition 7.2, together with the fact that $P(L(t, 0) > 0) = 1$. The proof of Theorem 3.1 is now complete. *Q.E.D.*

7. SOME USEFUL SUBSIDIARY PROPOSITIONS

In this section we will prove the following propositions required in the proof of Theorem 3.1. Notation will be the same as in the previous section except when explicitly mentioned.

PROPOSITION 7.1: *We have*

$$\zeta_n(t) \Rightarrow W(t) \quad \text{and} \tag{7.1}$$

$$\zeta_n'(t) := \frac{1}{\sqrt{n}\phi} \sum_{k=1}^{[nt]} \rho^{[nt]-k} \eta_k \Rightarrow J_\kappa(t) \quad \text{on} \quad D[0,1],$$

where $\{W(t), t \geq 0\}$ *is a standard Brownian motion and* $J_\kappa(t)$ *is defined as in* (3.9).

PROOF: The first statement of (7.1) is well known. So that to $\zeta_n'(t) \Rightarrow J_\kappa(t)$, for each fixed $l \geq 1$, put

$$Z_{1j}^{(l)} = \sum_{k=0}^{l} \phi_k \epsilon_{j-k} \quad \text{and} \quad Z_{2j}^{(l)} = \sum_{k=l+1}^{\infty} \phi_k \epsilon_{j-k}.$$

It is readily seen that for any $m \geq 1$,

$$\begin{aligned}
\sum_{j=1}^{m} \rho^{m-j} Z_{1j}^{(l)} &= \sum_{j=1}^{m} \rho^{m-j} \sum_{k=0}^{l} \phi_k \epsilon_{j-k} \\
&= \sum_{k=0}^{l} \rho^{-k} \phi_k \sum_{j=1}^{m} \rho^{m-j} \epsilon_j + \sum_{s=1}^{l} \rho^{m+s-1} \epsilon_{1-s} \sum_{j=s}^{l} \rho^{-j} \phi_j \\
&\quad + \sum_{s=0}^{l-1} \rho^{j} \epsilon_{m-s} \sum_{j=s+1}^{l} \rho^{-j} \phi_j \\
&= \sum_{k=0}^{l} \rho^{-k} \phi_k \sum_{j=1}^{m} \rho^{m-j} \epsilon_j + R(m,l), \quad \text{say.}
\end{aligned}$$

Therefore, for fixed $l \geq 1$,

$$\begin{aligned}
\zeta_n'(t) &= \left(\frac{1}{\phi} \sum_{k=0}^{l} \rho^{-k} \phi_k \right) \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \rho^{[nt]-j} \epsilon_j + \frac{1}{\sqrt{n}\phi} R([nt], l) \\
&\quad + \frac{1}{\sqrt{n}\phi} \sum_{j=1}^{[nt]} \rho^{[nt]-j} Z_{2j}^{(l)}.
\end{aligned} \tag{7.2}$$

Note that $\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\rho^{[nt]-j}\epsilon_j \Rightarrow J_\kappa(t)$ (see Chan and Wei (1987) and Phillips (1987)) and $\sum_{k=0}^{l}\rho^{-k}\phi_k \to \phi$ as $n\to\infty$ first and then $l\to\infty$. By virtue of Theorem 4.1 of Billingsley (1968, p. 25), to prove $\zeta_n'(t)\Rightarrow J_\kappa(t)$, it suffices to show that for any $\delta>0$,

$$\limsup_{n\to\infty} P\Big\{\sup_{0\le t\le 1}\big|R([nt],l)\big|\ge \delta\sqrt{n}\Big\}=0 \tag{7.3}$$

for fixed $l\ge 1$ and

$$\lim_{l\to\infty}\limsup_{n\to\infty} P\left\{\sup_{0\le t\le 1}\left|\sum_{j=1}^{[nt]} Z_{2j}^{(l)}\right|\ge \delta\sqrt{n}\right\}=0. \tag{7.4}$$

Recall $\lim_{n\to\infty}\rho^n=e^\kappa$, which yields $e^{-|\kappa|}/2\le\rho^k\le 2e^{|\kappa|}$ for all $-n\le k\le n$ and n sufficiently large. The result (7.3) holds since $\sum_{k=0}^{\infty}|\phi_k|<\infty$, and hence as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sup_{0\le t\le 1}\big|R([nt],l)\big|\le \frac{1}{\sqrt{n}}\max_{-l\le j\le n}|\epsilon_j|\sum_{s=0}^{l}\left(\sum_{j=s}^{l}|\phi_j|+\sum_{j=s+1}^{l}|\phi_j|\right)\xrightarrow{P} 0.$$

We next prove (7.4). Noting

$$\sum_{j=1}^{m}\rho^{m-j}Z_{2j}^{(l)}=\sum_{k=l+1}^{\infty}\phi_k\sum_{j=1}^{m}\rho^{m-j}\epsilon_{j-k}\quad\text{for any } m\ge 1,$$

by applying the Hölder inequality and the independence of ϵ_k, we have

$$\begin{aligned} E\sup_{0\le t\le 1}\left(\sum_{j=1}^{k_n(t)}Z_{2j}^{(l)}\right)^2 &\le \sum_{k=l+1}^{\infty}|\phi_k|\sum_{k=l+1}^{\infty}|\phi_k| E\max_{1\le m\le n}\left(\sum_{j=1}^{m}\rho^{m-j}\epsilon_{j-k}\right)^2\\ &\le Cn\left(\sum_{k=l+1}^{\infty}|\phi_k|\right)^2. \end{aligned}$$

Result (7.4) now follows immediately from the Markov inequality and $\sum_{k=l+1}^{\infty}|\phi_k|\to 0$ as $l\to\infty$. The proof of Proposition 7.1 is complete. *Q.E.D.*

PROPOSITION 7.2: *For any h satisfying $h\to 0$ and $nh^2\to\infty$, we have*

$$\frac{1}{\sqrt{nh^2}}\sum_{k=1}^{[nt]}K^i\left[\frac{x_k-x}{h}\right]\Rightarrow d_iL(t,0)\qquad (i=1,2), \tag{7.5}$$

$$\frac{1}{\sqrt{nh^2}}\sum_{k=1}^{[nt]}K^2\left[\frac{x_k-x}{h}\right]u_k^2\Rightarrow d_0^2L(t,0) \tag{7.6}$$

on $D[0,1]$, *where* $d_i = |\phi|^{-1}\int_{-\infty}^{\infty} K^i(s)\,ds$, $i = 1, 2$, $d_0^2 = |\phi|^{-1} E u_{m_0}^2 \int_{-\infty}^{\infty} K^2(s)\,ds$, *and* $L(t,s)$ *is the local time process of the Gaussian diffusion process* $\{J_\kappa(t), t \geq 0\}$ *defined by* (3.9), *in which* $\{W(t), t \geq 0\}$ *is a standard Brownian motion*.

PROPOSITION 7.3: *For any fixed* $0 \leq t \leq 1$, $S_n(t)$, $S_n^2(t)$, *and* $\psi_n(t)$, $n \geq 1$, *are uniformly integrable*.

PROPOSITION 7.4: $\{S_n(t)\}_{n\geq 1}$ *is tight on* $D[0,1]$.

PROPOSITION 7.5: *Results* (6.11) *and* (6.12) *hold true for any* $u \in R$.

To prove Propositions 7.2–7.5, we need some preliminaries.

Let $r(x)$ and $r_1(x)$ be bounded functions such that $\int_{-\infty}^{\infty}(|r(x)| + |r_1(x)|)\,dx < \infty$. We first calculate the values of $\mathrm{I}_{k,l}^{(s)}$ and $\mathrm{II}_k^{(s)}$ defined by

$$\mathrm{I}_{k,l}^{(s)} = E\left[r(x'_{s,k}/h) r_1(x'_{s,l}/h) g(u_k) g_1(u_l) \exp\left\{i\mu \sum_{j=1}^{l} \epsilon_j/\sqrt{n}\right\}\right], \tag{7.7}$$

$$\mathrm{II}_k^{(s)} = E\left[r(x'_{s,k}/h) g(u_k) \exp\left\{i\mu \sum_{j=1}^{k} \epsilon_j/\sqrt{n}\right\}\right],$$

under different settings of $g(x)$ and $g_1(x)$, where $x'_{s,k}$ is defined as in (6.10). We have the following lemmas, which will play a core rule in the proof of the main results. We always assume $k < l$ and let C denote a constant not dependent on k, l, and n, which may be different from line to line.

LEMMA 7.1: *Suppose* $\int |\hat{r}(\lambda)|\,d\lambda < \infty$, *where* $\hat{r}(t) = \int e^{itx} r(x)\,dx$.

(a) *If* $E|g(u_k)| < \infty$, *then, for all* $k \geq s+1$,

$$\left|\mathrm{II}_k^{(s)}\right| \leq Ch/\sqrt{k-s}. \tag{7.8}$$

(b) *If* $Eg(u_k) = 0$ *and* $Eg^2(u_k) < \infty$, *then, for all* $k \geq s+1$,

$$\left|\mathrm{II}_k^{(s)}\right| \leq C[(k-s)^{-2} + h/(k-s)]. \tag{7.9}$$

LEMMA 7.2: *Suppose that* $\int |\hat{r}(\lambda)|\,d\lambda < \infty$ *and* $\int |\hat{r}_1(\lambda)|\,d\lambda < \infty$, *where* $\hat{r}(t) = \int e^{itx} r(x)\,dx$ *and* $\hat{r}_1(t) = \int e^{itx} r_1(x)\,dx$. *Suppose that* $Eg(u_l) = Eg_1(u_k) = 0$ *and* $Eg^2(u_{m_0}) + Eg_1^2(u_{m_0}) < \infty$. *Then, for any* $\epsilon > 0$, *there exists an* $n_0 > 0$ *such that, for all* $n \geq n_0$, *all* $l - k \geq 1$, *and all* $k \geq s+1$,

$$\left|\mathrm{I}_{k,l}^{(s)}\right| \leq C[\epsilon(l-k)^{-2} + h(l-k)^{-1}][(k-s)^{-2} + h/\sqrt{k-s}]. \tag{7.10}$$

We only prove Lemma 7.2 with $s = 0$. The proofs of Lemmas 7.1 and 7.2 with $s \neq 0$ are the same and hence the details are omitted.

PROOF OF LEMMA 7.2: Write $x''_k = x'_{0,k}$ and $\mathrm{I}_{k,l} = \mathrm{I}^{(0)}_{k,l}$. As $\int(|\hat{r}(t)| + |\hat{r}_1(t)|)\,dt < \infty$, we have $r(x) = \frac{1}{2\pi}\int e^{-ixt}\hat{r}(t)\,dt$ and $r_1(x) = \frac{1}{2\pi}\int e^{-ixt}\hat{r}_1(t)\,dt$. This yields that

$$\begin{aligned}\mathrm{I}_{k,l} &= E\left[r(x''_k/h)r_1(x''_l/h)g(u_k)g_1(u_l)\exp\left\{i\mu\sum_{j=1}^{l}\epsilon_j/\sqrt{n}\right\}\right] \\ &= \int\int E\left\{\exp(-itx''_k/h)\exp(i\lambda x''_l/h)g(u_k)g_1(u_l)\right. \\ &\qquad \left.\times\exp\left(i\mu\sum_{j=1}^{l}\epsilon_j/\sqrt{n}\right)\right\}\hat{r}(t)\overline{\hat{r}_1(\lambda)}\,dt\,d\lambda.\end{aligned}$$

Define $\sum_{j=k}^{l} = 0$ if $l < k$, and put $\nabla(k) = \sum_{j=0}^{k}\rho^{-j}\phi_j$ and $a_{s,q} = \rho^{l-q}\nabla(s-q)$. Since

$$x''_l = \sum_{q=1}^{l}\epsilon_q\sum_{j=0}^{l-q}\rho^{l-q-j}\phi_j = \left(\sum_{q=1}^{k} + \sum_{q=k+1}^{l-m_0} + \sum_{q=l-m_0+1}^{l}\right)\epsilon_q a_{l,q},$$

it follows from independence of the ϵ_k's that, for $l - k \geq m_0 + 1$,

$$\begin{aligned}|\mathrm{I}_{k,l}| \leq \int &\left|E\left\{e^{iz^{(2)}/h}\right\}\right|\left|E\left\{e^{iz^{(3)}/h}g_1(u_l)\right\}\right||\hat{r}_1(\lambda)| \\ &\times\left(\int\left|E\left\{e^{iz^{(1)}/h}g(u_k)\right\}\right|\left|\hat{r}(t)\right|dt\right)d\lambda,\end{aligned} \tag{7.11}$$

where

$$z^{(1)} = \sum_{q=1}^{k}\epsilon_q(\lambda a_{l,q} - t a_{k,q} + uh/\sqrt{n}),$$

$$z^{(2)} = \sum_{q=k+1}^{l-m_0}\epsilon_q(\lambda a_{l,q} + uh/\sqrt{n}),$$

$$z^{(3)} = \sum_{q=l-m_0+1}^{l}\epsilon_q(\lambda a_{l,q} + uh/\sqrt{n}).$$

We may take n sufficiently large so that $u/\sqrt{n}$ is as small as required. Without loss of generality, we assume $u = 0$ in the following proof for convenience of notation. We first show that, for all k sufficiently large,

$$\Lambda(\lambda, k) := \int \left|E\left\{e^{iz^{(1)}/h} g(u_k)\right\}\right| |\hat{r}(t)|\, dt \le C(k^{-2} + h/\sqrt{k}). \tag{7.12}$$

To estimate $\Lambda(\lambda, k)$, we need some preliminaries. Recall $\rho = 1 + \kappa/n$. For any given s, we have $\lim_{n\to\infty} |\nabla(s)| = |\sum_{j=0}^{s} \phi_j|$. This fact implies that k_0 can be taken sufficiently large such that whenever n is sufficiently large,

$$\sum_{j=k_0/2+1}^{\infty} |\phi_j| \le e^{-|\kappa|}|\phi|/4 \le e^{-|\kappa|}|\nabla(k_0)|, \tag{7.13}$$

and hence for all $k_0 \le s \le n$ and $1 \le q \le s/2$,

$$|a_{s,q}| \ge 2^{-1} e^{-|\kappa|}\left(|\nabla(k_0/2)| - 2e^{|\kappa|} \sum_{j=k_0/2+1}^{\infty} |\phi_j|\right) \ge e^{-|\kappa|}|\phi|/4, \tag{7.14}$$

where we have used the well known fact that $\lim_{n\to\infty} \rho^n = e^{\kappa}$, which yields $e^{-|\kappa|}/2 \le \rho^k \le 2e^{|\kappa|}$ for all $-n \le k \le n$. Further write Ω_1 (Ω_2, respectively) for the set of $1 \le q \le k/2$ such that $|\lambda a_{l,q} - t a_{k,q}| \ge h$ ($|\lambda a_{l,q} - t a_{k,q}| < h$, respectively), and

$$B_1 = \sum_{q\in\Omega_2} a_{k,q}^2, \quad B_2 = \sum_{q\in\Omega_2} a_{l,q} a_{k,q}, \quad B_3 = \sum_{q\in\Omega_2} a_{l,q}^2.$$

By virtue of (7.13), it is readily seen that $B_1 \ge Ck$ whenever $\#(\Omega_1) \le \sqrt{k}$, where $\#(A)$ denotes the number of elements in A. We are now ready to prove (7.12). First notice that there exist constants $\gamma_1 > 0$ and $\gamma_2 > 0$ such that

$$|Ee^{i\epsilon_1 t}| \le \begin{cases} e^{-\gamma_1}, & \text{if } |t| \ge 1, \\ e^{-\gamma_2 t^2}, & \text{if } |t| \le 1, \end{cases} \tag{7.15}$$

since $E\epsilon_1 = 0$, $E\epsilon_1^2 = 1$, and ϵ_1 has a density; see, for example, Chapter 1 of Petrov (1995). Also note that

$$\begin{aligned} \sum_{q\in\Omega_2} (\lambda a_{l,q} - t a_{k,q})^2 &= \lambda^2 B_3 - 2\lambda t B_2 + t^2 B_1 \\ &= B_1(t - \lambda B_2/B_1)^2 + \lambda^2(B_3 - B_2^2/B_1) \\ &\ge B_1(t - \lambda B_2/B_1)^2, \end{aligned}$$

since $B_2^2 \le B_1 B_3$, by Hölder's inequality. It follows from the independence of ϵ_t that, for all $k \ge k_0$,

$$\left|Ee^{iW^{(1)}/h}\right| \le \exp\left\{-\gamma_1 \#(\Omega_1) - \gamma_2 h^{-2} \sum_{q\in\Omega_2} (\lambda a_{l,q} - t a_{k,q})^2\right\}$$
$$\le \exp\{-\gamma_1 \#(\Omega_1) - \gamma_2 B_1 h^{-2} (t - \lambda B_2/B_1)^2\},$$

where $W^{(1)} = \sum_{q=1}^{k/2} \epsilon_q (\lambda a_{l,q} - t a_{k,q})$. This, together with the facts $z^{(1)} = W^{(1)} + \sum_{q=k/2+1}^{k} \epsilon_q (\lambda a_{l,q} - t a_{k,q})$ and $k/2 \le k - m_0$ (which implies that $W^{(1)}$ is independent of u_k) yield that

$$\begin{aligned}\Lambda(\lambda, k) &\le \int \left|E\{\exp(iW^{(1)}/h)\}\right| E|g(u_k)||\hat{r}(t)|\,dt \\ &\le C \int_{\#(\Omega_1)\ge\sqrt{k}} \exp(-\gamma_1 \#(\Omega_1))|\hat{r}(t)|\,dt \\ &\quad + C \int_{\#(\Omega_1)\le\sqrt{k}} \exp(-\gamma_2 B_1 h^{-2} (t - \lambda B_2/B_1)^2)\,dt \\ &\le Ck^{-2} \int |\hat{r}(t)|\,dt + \int \exp(-\gamma_2 B_1 h^{-2} t^2)\,dt \\ &\le C(k^{-2} + h/\sqrt{k}).\end{aligned}$$

This proves (7.12) for $k \ge k_0$.

We now turn back to the proof of (7.10). We will estimate $\mathrm{I}_{k,l}$ in three separate settings:

$$\begin{aligned}&l - k \ge 2k_0 \quad \text{and} \quad k \ge k_0,\\ &l - k \le 2k_0 \quad \text{and} \quad k \ge k_0,\\ &l > k \quad \text{and} \quad k \le k_0,\end{aligned}$$

where, without loss of generality, we assume $k_0 \ge 2m_0$.

CASE I: $l - k \ge 2k_0$ and $k \ge k_0$. We first notice that, for any $\delta > 0$, there exist constants $\gamma_3 > 0$ and $\gamma_4 > 0$ such that, for all $s \ge k_0$ and $q \le s/2$,

$$|E \exp(i\epsilon_1 \lambda a_{s,q}/h)| \le \begin{cases} e^{-\gamma_3}, & \text{if } |\lambda| \ge \delta h, \\ e^{-\gamma_4 \lambda^2/h^2}, & \text{if } |\lambda| \le \delta h. \end{cases}$$

This fact follows from (7.14) and (7.15) with a simple calculation. Hence it follows from the facts $l - m_0 \ge (l+k)/2$ and $l - q \ge k_0$ for all $k \le q \le (l+k)/2$, since $l-k \geq 2k_0$ and $k_0 \geq 2m_0$, that

$$(7.16)\qquad \left|Ee^{iz^{(2)}/h}\right| \leq \prod_{q=k}^{(l+k)/2} |E\exp(i\epsilon_q \lambda a_{l,q}/h)| \leq \begin{cases} \exp(-\gamma_3(l-k)), & \text{if } |\lambda| \geq \delta h, \\ \exp(-\gamma_4(l-k)\lambda^2/h^2), & \text{if } |\lambda| \leq \delta h. \end{cases}$$

On the other hand, since $Eg_1(u_l)=0$, we have

$$(7.17)\qquad \begin{aligned} \left|E\{e^{iz^{(3)}/h}g_1(u_l)\}\right| &= \left|E\{(e^{iz^{(3)}/h}-1)g_1(u_l)\}\right| \\ &\leq h^{-1}E\left[\left|z^{(3)}\right||g_1(u_l)|\right] \\ &\leq C(E\epsilon_1^2)^{1/2}(Eg_1^2(u_l))^{1/2}|\lambda|h^{-1}. \end{aligned}$$

We also have

$$(7.18)\qquad \left|E\{e^{iz^{(3)}/h}g_1(u_l)\}\right| \to 0, \quad \text{whenever} \quad \lambda h^{-1} \to \infty,$$

uniformly for all $l \geq m_0$. Indeed, supposing $\phi_0 \neq 0$ (if $\phi_0 = 0$, we may use ϕ_1 and so on), we have $E\{\exp(iz^{(3)}/h)g_1(u_l)\} = E\{\exp(i\epsilon_l\phi_0\lambda\rho^{n-l}/h)g^*(\epsilon_l)\}$, where $g^*(\epsilon_l) = E[\exp(i(z^{(3)} - \epsilon_l\phi_0\lambda\rho^{n-l})/h)g_1(u_l) \mid \epsilon_l]$. By recalling that ϵ_l has a density $d(x)$, it is readily seen that

$$\int \sup_\lambda |g^*(x)|d(x)\,dx \leq E|g_1(u_l)| < \infty$$

uniformly for all l. The result (7.18) follows from the Riemann–Lebesgue theorem.

By virtue of (7.18), for any $\epsilon > 0$, there exists an n_0 (A_0 respectively) such that, for all $n \geq n_0$ ($|\lambda|/h \geq A_0$, respectively), $|E\{\exp(iz^{(3)}/h)g_1(u_l)\}| \leq \epsilon$. This, together with (7.12) and (7.16) with $\delta = A_0$, yields that

$$\begin{aligned} \mathrm{I}_{k,l}^{(2)} &:= \int_{|\lambda|>A_0 h} \left|E\{e^{iz^{(2)}/h}\}\right|\left|E\{e^{iz^{(3)}/h}g_1(u_l)\}\right| \Lambda(\lambda,k)|\hat{r}_1(\lambda)|\,d\lambda \\ &\leq C\epsilon e^{-\gamma_3(l-k)}(k^{-2}+h/\sqrt{k}) \int_{|\lambda|>A_0 h} |\hat{r}_1(\lambda)|\,d\lambda \\ &\leq C\epsilon(l-k)^{-2}(k^{-2}+h/\sqrt{k}). \end{aligned}$$

Similarly it follows from (7.12), (7.16) with $\delta = A_0$, and (7.17) that

$$\mathrm{I}_{k,l}^{(1)} := \int_{|\lambda|\le A_0 h} \big|E\{e^{iz^{(2)}/h}\}\big|\big|E\{e^{iz^{(3)}/h} g_1(u_l)\}\big| |\Lambda(\lambda,k)|\hat{r}_1(\lambda)|\,d\lambda$$
$$\le C(k^{-2} + h/\sqrt{k})h^{-1} \int_{|\lambda|\le A_0 h} \lambda \exp(-\gamma_4(l-k)\lambda^2/h^2)\,d\lambda$$
$$\le Ch(l-k)^{-1}(k^{-2} + h/\sqrt{k}).$$

The result (7.10) in Case I now follows from

$$\mathrm{I}_{k,l} \le \mathrm{I}_{k,l}^{(1)} + \mathrm{I}_{k,l}^{(2)} \le C[\epsilon(l-k)^{-2} + h(l-k)^{-1}](k^{-2} + h/\sqrt{k}).$$

CASE II: $l - k \le 2k_0$ and $k \ge k_0$. In this case, we only need to show that

$$|\mathrm{I}_{k,l}| \le C(\epsilon + h)(k^{-2} + h/\sqrt{k}). \tag{7.19}$$

In fact, as in (7.11), we have

$$|\mathrm{I}_{k,l}| \le \int\int \big|E\{e^{iz^{(4)}/h}\}\big|\big|E\{e^{iz^{(5)}/h} g(u_k) g_1(u_l)\}\big| |\hat{r}(t)||\hat{r}_1(\lambda)|\,dt\,d\lambda, \tag{7.20}$$

where

$$z^{(4)} = \sum_{q=1}^{k-m_0} \epsilon_q[\lambda a_{l,q} - t a_{k,q}],$$
$$z^{(5)} = \sum_{q=k-m_0+1}^{l} \epsilon_q(\lambda a_{l,q} + uh/\sqrt{n}) - t \sum_{q=k-m_0+1}^{k} \epsilon_q a_{k,q}.$$

Similar arguments to those in the proof of (7.12) give that, for all λ and all $k \ge k_0$,

$$\Lambda_1(\lambda,k) := \int \big|E\{e^{iz^{(4)}/h}\}\big| |\hat{r}(t)|\,dt \le C(k^{-2} + h/\sqrt{k}).$$

Note that

$$E|g(u_k)g_1(u_l)| \le (Eg^2(u_k))^{1/2}(Eg_1^2(u_l))^{1/2} < \infty.$$

For any $\epsilon > 0$, similar to the proof of (7.18), there exists an n_0 (A_0, respectively) such that, for all $n \ge n_0$ ($|\lambda|/h \ge A_0$, respectively), $|E\{e^{iz^{(5)}/h} g(u_k) g_1(u_l)\}| \le \epsilon$.

By virtue of these facts, we have

$$|\mathrm{I}_{k,l}| \le \left(\int_{|\lambda|\le A_0 h} + \int_{|\lambda|>A_0 h}\right) \left|E\left\{e^{iz^{(5)}/h} g(u_k) g_1(u_l)\right\}\right| |\hat{r}_1(\lambda)| \Lambda_1(\lambda, k)\, d\lambda$$
$$\le C\left(\int_{|\lambda|\le A_0 h} d\lambda + \epsilon \int_{|\lambda|>A_0 h} |\hat{r}_1(\lambda)|\, d\lambda\right)(k^{-2} + h/\sqrt{k})$$
$$\le C(\epsilon + h)(k^{-2} + h/\sqrt{k}).$$

This proves (7.19) and hence the result (7.10) in Case II.

CASE III: $l > k$ and $k \le k_0$. In this case, we only need to prove

$$|\mathrm{I}_{k,l}| \le C\left[\epsilon(l-k)^{-3/2} + h(l-k)^{-1}\right]. \tag{7.21}$$

To prove (7.21), split $l > k$ into $l - k \ge 2k_0$ and $l - k \le 2k_0$. The result (7.10) then follows from the same arguments as in the proofs of Cases I and II but replacing the estimate of $\Lambda(\lambda, k)$ in (7.12) by

$$\Lambda(\lambda, k) \le E|g(u_k)| \int |\hat{r}(t)|\, dt \le C.$$

We omit the details. The proof of Lemma 7.2 is now complete. *Q.E.D.*

We are now ready to prove the propositions. We first mention that, under the conditions for $K(t)$, if we let $r(t) = K(y/h + t)$ or $r(t) = K^2(y/h + t)$, then $\int |r(x)|\, dx = \int |K(x)|\, dx < \infty$ and $\int |\hat{r}(\lambda)|\, d\lambda \le \int |\hat{K}(\lambda)|\, d\lambda < \infty$ uniformly for all $y \in R$.

PROOF OF PROPOSITION 7.5: Let $r(t) = r_1(t) = K(y/h + t)$ and $g(x) = g_1(x) = x$. It follows from Lemma 7.2 that for any $\epsilon > 0$, there exists an n_0 such that, whenever $n \ge n_0$,

$$\sum_{1\le k<l\le n} |\mathrm{I}_{k,l}| \le C \sum_{1\le k<l\le n} [\epsilon(l-k)^{-2} + h(l-k)^{-1}](k^{-2} + h/\sqrt{k})$$
$$\le C\left(\epsilon + h\sum_{k=1}^{n} k^{-1}\right)\sum_{k=1}^{n}(k^{-2} + h/\sqrt{k})$$
$$\le C(\epsilon + h\log n)(C + \sqrt{n}h).$$

This implies (6.12) since $h \log n \to 0$ and $nh^2 \to \infty$. The proof of (6.11) is similar and the details are omitted. *Q.E.D.*

PROOF OF PROPOSITION 7.2: We first note that, under a suitable probability space $\{\Omega, \mathcal{F}, P\}$, there exists an equivalent process $\hat{\zeta}'_n(t)$ of $\zeta'_n(t)$ (i.e., $\hat{\zeta}'_n(i/n) =_d \zeta'_n(i/n), 1 \le i \le n$, for each $n \ge 1$) such that

$$\sup_{0 \le t \le 1} |\hat{\zeta}'_n(t) - J_\kappa(t)| = o_P(1) \tag{7.22}$$

by Proposition 7.1 and the Skorohod–Dudley–Wichura representation theorem. Also, we may make the following claim:

CLAIM 1: $x_{j,n} := \hat{\zeta}'_n(j/n)$ or, equivalently, $x_{j,n} = \zeta'_n(j/n)$ satisfies Assumption 2.3 of WP.

The proof of this claim is similar to Corollary 2.2 of WP. Here we only give a outline. Write

$$\zeta'_n(l/n) - \zeta'_n(k/n) = S_{1l} + S_{2l},$$

where $S_{1l} = \frac{1}{\sqrt{n}\phi} \sum_{j=k+1}^{l} \rho^{l-j} \sum_{i=-\infty}^{j} \epsilon_i \phi_{j-i} + (\rho^{l-k} - 1)\zeta'_n(k/n)$ and

$$S_{2l} = \frac{\rho^l}{\sqrt{n}\phi} \sum_{j=k+1}^{l} \rho^{-j} \sum_{i=k+1}^{j} \epsilon_i \phi_{j-i} = \frac{\rho^l}{\sqrt{n}\phi} \sum_{i=k+1}^{l} \rho^{-i}\epsilon_i \sum_{j=0}^{l-i} \rho^{-j}\phi_j.$$

Furthermore let $d_{l,k,n}^2 = (\rho^{2l}/(n\phi^2)) \sum_{i=k+1}^{l} \rho^{-2i} (\sum_{j=0}^{l-i} \rho^{-j}\phi_j)^2$ and $\mathcal{F}_{t,n} = \sigma(\ldots, \epsilon_{t-1}, \epsilon_t)$. Recall (7.14). It is readily seen that $d_{l,k,n}^2 \ge C(l-k)/n$ whenever $l-k$ is sufficiently large. This implies that $d_{l,k,n}$ satisfies Assumption 2.3(i) of WP. On the other hand, by using a similar argument as in the proof of Corollary 2.2 of WP with minor modifications, it may be shown that the standardized sum

$$S_{2l}/d_{l,k,n} = \sum_{i=k+1}^{l} \rho^{-i}\epsilon_i \sum_{j=0}^{l-i} \rho^{-j}\phi_j \Bigg/ \sqrt{\sum_{i=k+1}^{l} \rho^{-2i} \left(\sum_{j=0}^{l-i} \rho^{-j}\phi_j \right)^2}$$

has a bounded density $h_{l,k}(x)$ satisfying

$$\sup_x |h_{l,k}(x) - n(x)| \to 0 \quad \text{as} \quad l-k \ge \delta n \to \infty,$$

where $n(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the standard normal density. Hence, conditional on $\mathcal{F}_{k,n}$, $(x_{l,n} - x_{k,n})/d_{l,k,n} = (S_{1l} + S_{2l})/d_{l,k,n}$ has a density $h_{l,k}(x - S_{1l}/d_{l,k,n})$, which is uniformly bounded by a constant C and

$$\begin{aligned}
&\sup_{l-k\geq\delta n}\sup_{|u|\leq\delta}\left|h_{l,k}\left(\frac{x-S_{1l}}{d_{l,k,n}}\right)-h_{l,k}\left(\frac{-S_{1l}}{d_{l,k,n}}\right)\right| \\
&\quad\leq 2\sup_{l-k\geq\delta n}\left|h_{l,k}(x)-\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right)\right| \\
&\qquad+\frac{1}{\sqrt{2\pi}}\sup_{|u|\leq\delta}\sup_{x}\left|\exp\left(\frac{-(x+u)^2}{2}-e^{-x^2/2}\right)\right| \\
&\quad\to 0
\end{aligned}$$

as $n\to\infty$ first and then $\delta\to 0$. This proves that Assumption 2.3(ii) holds true for $x_{k,n}$, and also completes the proof of Claim 1.

By virtue of all the above facts, it follows from Theorem 2.1 of WP with the settings $c_n=\sqrt{n}|\phi|/h$ and $g(t)=K^i(t-x/h)$, $i=1,2$, that

$$\begin{aligned}
&\sup_{0\leq t\leq 1}\left|\frac{\phi}{\sqrt{nh^2}}\sum_{k=1}^{[nt]}K^i\left[\frac{\sqrt{n}\phi\hat{\zeta}_n'(k/n)-x}{h}\right]-L(t,0)\int_{-\infty}^{\infty}K^i(s)\,ds\right| \\
&\quad\to_P 0.
\end{aligned}$$

This, together with the fact that $\hat{\zeta}_n'(k/n)=_d\zeta_n'(k/n)=x_k/(\sqrt{n}\phi)$, $1\leq k\leq n$ for each $n\geq 1$, implies that the finite-dimensional distributions of $T_{in}(t):=(1/\sqrt{nh^2})\sum_{k=1}^{[nt]}K^i[(x_k-x)/h]$ converge to those of $d_iL(t,0)$. On the other hand, by applying the same argument as in the proof of Proposition 7.4, it is easy to show that $T_{in}(t), n\geq 1$, is tight. Hence $T_{in}(t)\Rightarrow d_iL(t,0)$, $i=1$ or 2, on $D[0,1]$. This proves the result (7.5).

To prove (7.6), write $\psi_n'(t)=\frac{1}{\sqrt{nh}}\sum_{k=1}^{[nt]}K^2[(x_k-x)/h]u_k^2$ and $\psi_n''(t)=\frac{1}{\sqrt{nh}}\sum_{k=1}^{[nt]}K^2[(x_k-x)/h]Eu_k^2$. We first prove

$$\sup_{0\leq t\leq 1}E|\psi_n'(t)-\psi_n''(t)|^2=o(1). \tag{7.23}$$

In fact, by recalling $x_k=x_{0,k}^*+x_{0,k}'$ (see (6.10)), where $x_{0,k}^*$ depends only on $\epsilon_0,\epsilon_{-1},\ldots,$ we have, almost surely,

$$\begin{aligned}
&E\left[|\psi_n'(t)-\psi_n''(t)|^2\mid\epsilon_0,\epsilon_{-1},\ldots\right] \\
&\quad\leq\frac{1}{nh^2}\sup_{y,1\leq m\leq n}E\left[\sum_{k=1}^{m}K^2\left[\frac{y+x_{0,k}'}{h}\right](u_k^2-Eu_k^2)\right]^2
\end{aligned}$$

$$\le \frac{1}{nh^2} \sup_y \Bigg[\sum_{k=1}^{n} Er^2\left(\frac{x'_{0,k}}{h}\right) g^2(u_k) + 2 \sum_{1\le k<l\le n} \left| Er\left(\frac{x'_{0,k}}{h}\right) r\left(\frac{x'_{0,l}}{h}\right) g(u_k) g_1(u_l) \right| \Bigg],$$

where $r(t) = K^2(y/h + t)$, $g(t) = t^2 - Eu_k^2$, and $g_1(t) = t^2 - Eu_l^2$. Again it follows from Lemmas 7.1 and 7.2 that, for any $\epsilon > 0$, there exists an n_0 such that for all $n \ge n_0$,

$$E[|\psi'_n(t) - \psi''_n(t)|^2 \mid \epsilon_0, \epsilon_{-1}, \ldots] \le C \frac{1}{nh} \sum_{k=m_0}^{n} k^{-1/2} + C(\epsilon + h \log n) \le C\left[\epsilon + h \log n + \frac{1}{\sqrt{n}h}\right],$$

almost surely. The result (7.23) follows from $nh^2 \to \infty$, $h \log n \to 0$, and the fact that ϵ is arbitrary.

The result (7.23) means that $\psi'_n(t)$ and $\psi''_n(t)$ have the same finite-dimensional limit distributions. Hence, the finite-dimensional distributions of $\psi'_n(t)$ converge to those of $d_0^2 L(t,0)$, since $\psi''_n(t) \Rightarrow d_0^2 L(t,0)$ on $D[0,1]$ by (7.5) and the fact $Eu_k^2 = Eu_{m_0}^2$ whenever $k \ge m_0$. On the other hand, $\psi'_n(t)$ is tight on $D[0,1]$, which follows from the same argument as in the proof of Proposition 7.4. This proves $\psi'_n(t) \Rightarrow d_0^2 L(t,0)$ on $D[0,1]$, that is, the result (7.6). *Q.E.D.*

PROOF OF PROPOSITION 7.3: We first claim that, for each fixed t,

$$\sup_n E[\psi''_n(t)]^2 < \infty, \tag{7.24}$$

where $\psi''_n(t) = \frac{1}{\sqrt{n}h} \sum_{k=1}^{[nt]} K^2[(x_k - x)/h] E u_k^2$ as above. In fact, by recalling $x_k = x^*_{s,k} + x'_{s,k}$ (see (6.10)), where $x^*_{s,k}$ depends only on $\epsilon_s, \epsilon_{s-1}, \ldots$, it follows from Lemma 7.1 with $r(t) = r_y(t) = K^2(y/h + t)$ and $g(t) = 1$ that, for each fixed t,

$$E|\psi''_n(t)|^2 \le \frac{C}{nh^2} \Bigg[\sum_{k=1}^{n} EK^4\left[\frac{x_k - x}{h}\right] + 2 \sum_{1\le k<l\le n} E\left\{ K^2\left[\frac{x_k - x}{h}\right] K^2\left[\frac{x_l - x}{h}\right]\right\}\Bigg]$$

$$\le \frac{C}{nh^2} \Bigg[\sum_{k=1}^{n} \sup_y Er_y^2\left(\frac{x'_{0,k}}{h}\right)$$

$$+ 2 \sum_{1\le k<l\le n} \sup_y Er_y\left(\frac{x'_{0k}}{h}\right) \sup_y Er_y\left(\frac{x'_{kj}}{h}\right)\Bigg]$$
$$\le \frac{C}{nh^2}\left[\sum_{k=1}^{n} hk^{-1/2} + 2 \sum_{1\le k<l\le n} h^2 k^{-1/2}(l-k)^{-1/2}\right]$$
$$< \infty$$

uniformly on n, as required. The result (7.24), together with (7.23), implies that $\sup_n E[\psi_n(t)]^2 < \infty$ and hence $\psi_n(t)$ is uniformly integrable.

To prove the uniform integrability of $S_n^2(t)$, we first notice that

$$\sup_{0\le t\le 1} E|\psi_n(t) - S_n^2(t)| = o(1). \tag{7.25}$$

This follows from the similar argument as in the proof of (7.23) and the fact that

$$\psi_n(t) - S_n^2(t) = \frac{2d_0^{-2}}{nh^2} \sum_{1\le k<l\le [nt]} u_k u_l K\left[\frac{x_k - x}{h}\right] K\left[\frac{x_l - x}{h}\right].$$

By virtue of (7.25), for any $A > 0$ and fixed t, we have

$$\left|ES_n^2(t) I_{S_n^2(t)\ge A} - E\psi_n(t) I_{S_n^2(t)\ge A}\right| \le \sup_{0\le t\le 1} E|\psi_n(t) - S_n^2(t)| = o(1).$$

This, together with the fact that

$$E\psi_n(t) I_{S_n^2(t)\ge A} \le E\psi_n(t) I_{\psi_n(t)\ge\sqrt{A}} + \sqrt{A} P(S_n^2(t) \ge A)$$
$$\le E\psi_n(t) I_{\psi_n(t)\ge\sqrt{A}} + A^{-1/2} E\psi_n(t) + o(1),$$

implies that

$$\lim_{A\to\infty} \sup_n ES_n^2(t) I_{S_n^2(t)\ge A} \le \lim_{A\to\infty} \sup_n \left[E\psi_n(t) I_{\psi_n(t)\ge\sqrt{A}} + A^{-1/2} E\psi_n(t)\right]$$
$$= 0,$$

where we have used the uniform integrability of $\psi_n(t)$. That is, $S_n^2(t)$ is uniformly integrable. The integrability of $S_n(t)$ follows from that of $S_n^2(t)$. The proof of Proposition 7.3 is now complete. *Q.E.D.*

PROOF OF PROPOSITION 7.4: We will use Theorem 4 of Billingsley (1974) to establish the tightness of $S_n(t)$ on $D[0, 1]$. According to this theorem, we only need to show that

$$\max_{1\le k\le n} \left|u_k K[(x_k - x)/h]\right| = o_P\left[(nh^2)^{1/4}\right] \tag{7.26}$$

and that there exists a sequence of $\alpha_n(\epsilon, \delta)$ satisfying $\lim_{\delta\to 0}\limsup_{n\to\infty} \alpha_n(\epsilon, \delta) = 0$ for each $\epsilon > 0$ such that, for

$$0 \le t_1 \le t_2 \le \cdots \le t_m \le t \le 1, \quad t - t_m \le \delta,$$

we have

$$P\big[|S_n(t) - S_n(t_m)| \ge \epsilon \mid S_n(t_1), S_n(t_2), \ldots, S_n(t_m)\big] \le \alpha_n(\epsilon, \delta) \quad \text{a.s.} \tag{7.27}$$

By noting $\max_{1\le k\le n} |u_k K[(x_k - x)/h]| \le \{\sum_{j=1}^n u_j^4 K^4[(x_j - x)/h]\}^{1/4}$, the result (7.26) follows from $Eu_j^4 K^4[(x_j - x)/h] \le Ch/\sqrt{j}$ by Lemma 7.1, with a simple calculation. As for (7.27), it only needs to show that

$$\begin{aligned}\sup_{|t-s|\le\delta} P\Bigg(\Bigg|\sum_{k=[ns]+1}^{[nt]} u_k K[(x_k - x)/h]\Bigg| \\ \ge \epsilon d_n \Bigg| \epsilon_{[ns]}, \epsilon_{[ns]-1}, \ldots; \eta_{[ns]}, \ldots, \eta_1\Bigg) \\ \le \alpha_n(\epsilon, \delta).\end{aligned} \tag{7.28}$$

In terms of the independence, we may choose $\alpha_n(\epsilon, \delta)$ as

$$\alpha_n(\epsilon, \delta) := \epsilon^{-2}(nh^2)^{-1/2} \sup_{y, 0\le t\le\delta} E\Bigg\{\sum_{k=1}^{[nt]} u_k K[(y + x'_{0,k})/h]\Bigg\}^2.$$

As in the proof of (7.23) with a minor modification, it is clear that whenever n is large enough,

$$\begin{aligned}\alpha_n(\epsilon, \delta) &\le \epsilon^{-2}(nh^2)^{-1/2} \sup_y \sum_{k=1}^{[n\delta]} E\big\{u_k^2 K^2[(y + x'_{0,k})/h]\big\} \\ &\quad + \epsilon^{-2}(nh^2)^{-1/2} \sup_y 2 \sum_{1\le k<l\le[n\delta]} \big|E\big\{u_k u_l K[(y + x'_{0,k})/h] \\ &\quad \times K[(y + x'_{0,l})/h]\big\}\big| \\ &\le C\epsilon^{-2}(nh^2)^{-1/2} \sum_{k=1}^{[n\delta]} h/\sqrt{k}(1 + \epsilon + h\log n).\end{aligned}$$

This yields $\lim_{\delta\to 0}\limsup_{n\to\infty} \alpha_n(\epsilon, \delta) = 0$ for each $\epsilon > 0$. The proof of Proposition 7.4 is complete. *Q.E.D.*

8. PROOF OF THEOREM 3.2

We may write

$$\hat{\sigma}_n^2 - Eu_{m_0}^2 = \Theta_{3n}[\Theta_{4n} + \Theta_{5n} + \Theta_{6n}],$$

where Θ_{3n} is defined as in (3.7),

$$\Theta_{4n} = \sum_{t=1}^{n} \big(u_t^2 - Eu_{m_0}^2\big) K[(x_t - x)/h],$$

$$\Theta_{5n} := 2\sum_{t=1}^{n} [f(x_t) - \hat{f}(x)] u_t K[(x_t - x)/h],$$

$$\Theta_{6n} := \sum_{t=1}^{n} [f(x_t) - \hat{f}(x)]^2 K[(x_t - x)/h].$$

As in the proof of (3.6) with minor modifications, we have $\Theta_{6n} = O_P\{\sqrt{n} h^{1+\gamma}\}$. As in the proof of (3.5), we obtain $\Theta_{4n} = O_P\{(\sqrt{n}h)^{1/2}\}$ and

$$\Theta_{1n}' := \sum_{t=1}^{n} u_t^2 K[(x_t - x)/h] = O_P(\sqrt{n}h).$$

These facts, together with (3.7), imply that

$$\hat{\sigma}_n^2 - Eu_{m_0}^2 = o_P\big\{a_n\big[h^{\gamma/2} + (\sqrt{n}h)^{-1/2}\big]\big\}, \tag{8.1}$$

where a_n diverges to infinity as slowly as required and where we use the fact that by Hölder's inequality,

$$|\Theta_{5n}| \le 2\Theta_{6n}^{1/2}\Theta_{1n}'^{1/2} = O_P\big\{(\sqrt{n}h)h^{\gamma/2}\big\}.$$

Now, result (3.15) follows from (8.1) by choosing $a_n = \min\{h^{-\gamma/4}, (\sqrt{n}h)^{1/4}\}$.

On the other hand, similar to the proof of (3.8), we may prove

$$\begin{aligned} &\left\{(nh^2)^{-1/4} \sum_{k=1}^{[nt]} (u_k^2 - Eu_{m_0}^2) K[(x_k - x)/h],\right. \\ &\qquad \left.(nh^2)^{-1/2} \sum_{k=1}^{n} K[(x_k - x)/h]\right\} \\ &\quad \to_D \left\{d_0' N L^{1/2}(t,0), d_1 L(1,0)\right\} \end{aligned} \tag{8.2}$$

on $D[0,1]^2$, where $d_0'^2 = |\phi|^{-1} E(u_{m_0}^2 - Eu_{m_0}^2)^2 \int_{-\infty}^{\infty} K^2(s)\,dt$, $d_1 = |\phi|^{-1} \times \int_{-\infty}^{\infty} K(s)\,ds$, and N is a standard normal variate independent of $L(1,0)$, as in (3.8). This, together with the fact that $\Theta_{3n}(\Theta_{4n}+\Theta_{5n}) = o_P(a_n h^\gamma)$ for any a_n diverging to infinity as slowly as required, yields

$$\begin{aligned}(nh^2)^{1/4}\big(\hat{\sigma}_n^2 - Eu_{m_0}^2\big) &= (nh^2)^{-1/2}\Theta_{3n}\big[(nh^2)^{-1/4}\Theta_{4n}\big] \\ &\quad + (nh^2)^{1/4}\Theta_{3n}(\Theta_{5n}+\Theta_{6n}) \\ &\to_D \sigma_1 N L(1,0)^{-1/2},\end{aligned}$$

whenever $nh^2 \to \infty$ and $nh^{2+2\gamma} \to 0$. The proof of Theorem 3.2 is now complete. *Q.E.D.*

9. BIAS ANALYSIS

We consider the special case where, in addition to earlier conditions, $\kappa = 0$, u_t is a martingale difference sequence with $E(u_t^2) = \sigma_u^2$, u_t is independent of x_t, K satisfies $\int K(y)\,dy = 1$, $\int yK(y)\,dy = 0$ and has compact support, and f has continuous, bounded third derivatives. It follows from the proof of Theorems 2.1 and 3.1 of WP that, on a suitably enlarged probability space

$$\frac{1}{\sqrt{n}h}\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) \to_P L(1,0), \tag{9.1}$$

and

$$(nh^2)^{1/4}\frac{\sum_{t=1}^{n} u_t K\left(\frac{x_t - x}{h}\right)}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right)} \Rightarrow \frac{N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K^2(s)\,ds\right)}{L(1,0)^{1/2}} \tag{9.2}$$

whenever $nh^2 \to \infty$ and $h \to 0$. The error decomposition is

$$\hat{f}(x_t) - f(x) = \frac{\sum_{t=1}^{n}\{f(x_t) - f(x)\}K\left(\frac{x_t - x}{h_n}\right)}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)} + \frac{\sum_{t=1}^{n} u_t K\left(\frac{x_t - x}{h_n}\right)}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)}. \tag{9.3}$$

The bias term in the numerator of the first term of (9.3) involves

$$\sum_{t=1}^{n}\{f(x_t)-f(x)\}K\left(\frac{x_t-x}{h}\right)=\mathrm{I}_a+\mathrm{I}_b+\mathrm{I}_c, \tag{9.4}$$

where

$$\mathrm{I}_a=f'(x)\sum_{t=1}^{n}(x_t-x)K\left(\frac{x_t-x}{h}\right),$$
$$\mathrm{I}_b=\frac{1}{2}f''(x)\sum_{t=1}^{n}(x_t-x)^2K\left(\frac{x_t-x}{h}\right),$$
$$\mathrm{I}_c=\sum_{t=1}^{n}\left\{f(x_t)-f(x)-f'(x)(x_t-x)-\frac{1}{2}f''(x)(x_t-x)^2\right\}$$
$$\times K\left(\frac{x_t-x}{h}\right).$$

As in (9.1) above, we have

$$\begin{aligned}\frac{\mathrm{I}_b}{\sqrt{n}h^3} &= \frac{1}{2}f''(x)\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}H\left(\frac{x_t-x}{h}\right) \quad (\text{where } H(s):=s^2K(s))\\ &\to_P \frac{1}{2}f''(x)\int_{-\infty}^{\infty}H(y)\,dy\,L(1,0).\end{aligned} \tag{9.5}$$

We show below that the remaining terms of (9.4) have the order

$$\mathrm{I}_a+\mathrm{I}_c=O_P\big((\sqrt{n}h^3)^{1/2}+(\sqrt{n}h^5\log n)^{1/2}+(\sqrt{n}h^4)\big). \tag{9.6}$$

It follows from (9.1), (9.4), and (9.5) that

$$\frac{\sum_{t=1}^{n}\{f(x_t)-f(x)\}K\left(\frac{x_t-x}{h}\right)}{\sum_{t=1}^{n}K\left(\frac{x_t-x}{h}\right)} \tag{9.7}$$
$$=\frac{\frac{1}{\sqrt{n}h}(\mathrm{I}_a+\mathrm{I}_b+\mathrm{I}_c)}{\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h_n}\right)}$$

$$= \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} y^2 K(y)\, dy \{1 + o_P(1)\}$$

$$+ O_P\left(\left(\frac{h}{\sqrt{n}}\right)^{1/2} + \left(\frac{h^3 \log n}{\sqrt{n}}\right)^{1/2} + h^3\right).$$

Then, from (9.3), (9.2), and (9.7),

$$(nh^2)^{1/4}\left[\hat{f}(x_t) - f(x) - \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} y^2 K(y)\, dy\right]$$

$$= \frac{\frac{1}{(nh^2)^{1/4}} \sum_{t=1}^{n} u_t K\left(\frac{x_t - x}{h}\right)}{\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right)}$$

$$+ O_P\left(h + h^2 (\log n)^{1/2} + n^{1/4} h^{7/2}\right)$$

$$\Rightarrow \frac{N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K^2(s)\, ds\right)}{L(1,0)^{1/2}},$$

provided $h^4 \log n + nh^{14} \to 0$, for which $nh^{14} \to 0$ suffices.

It remains to prove (9.6). As shown in the proof of Proposition 7.2, $x_{t,n} = n^{-1/2} x_t$ satisfies Assumption 2.3 of WP, so that for $t > s$, the scaled quantity $\frac{\sqrt{n}}{\sqrt{t-s}}(x_{t,n} - x_{s,n})$ has a uniformly bounded density $h_{t,s,n}(y)$. Furthermore we may prove that $h_{t,s,n}$ is locally Lipschitz in the neighborhood of the origin, that is,

$$|h_{t,s,n}(x) - h_{t,s,n}(0)| \le c|x|. \tag{9.8}$$

Then, for some constant C whose value may change in each occurrence, we have

$$E|\mathrm{I}_c| \le \sum_{t=1}^{n} \int_{-\infty}^{\infty} \left\{ \left| f(\sqrt{t} y) - f(x) - f'(x)(\sqrt{t} y - x) \right.\right. \tag{9.9}$$

$$\left.\left. - \frac{1}{2} f''(x)(\sqrt{t} y - x)^2 \right| K\left(\frac{\sqrt{t}}{h} y - \frac{x}{h}\right) \right\} h_{t,0,n}(y)\, dy$$

$$\le hC \sum_{t=1}^{n} \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} \left| \left\{ f(hy + x) - f(x) - f'(x)(hy) \right.\right.$$

$$- \frac{1}{2} f''(x)(hy)^2 \bigg| K(y) \bigg\} \, dy$$

$$\leq Ch^4 \sum_{t=1}^{n} \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} |s|^3 K(s) \, ds \leq C\sqrt{n} h^4,$$

using the fact that K has compact support. As for I_a, we have

$$E\mathrm{I}_a^2 \leq Ch^2 E \Bigg[\sum_{t=1}^{n} H_1\bigg(\frac{x_t - x}{h} \bigg) \Bigg]^2 \quad (\text{with } H_1(y) := yK(y))$$

$$\leq Ch^2 \Bigg[\sum_{t=1}^{n} EH_1^2\bigg(\frac{x_t - x}{h} \bigg) + \sum_{1 \leq s < t \leq n} EH_1\bigg(\frac{x_s - x}{h} \bigg) H_1\bigg(\frac{x_t - x}{h} \bigg) \Bigg].$$

It is readily seen that

$$EH_1^2\bigg(\frac{x_t - x}{h} \bigg) = \int H_1^2\bigg(\frac{\sqrt{t} y - x}{h} \bigg) h_{t,0,n}(y) \, dy \leq \frac{Ch}{\sqrt{t}}. \tag{9.10}$$

Since $\int H_1(y) \, dy = 0$ and using (9.8), we also have

$$E\bigg\{ H_1\bigg(\frac{x_t - x}{h} \bigg) \,\Big|\, \mathcal{F}_s \bigg\}$$

$$= \int H_1\bigg(\frac{\sqrt{t-s} y + x_s - x}{h} \bigg) h_{t,s,n}(y) \, dy$$

$$= \frac{h}{\sqrt{t-s}} \int H_1\bigg(y + \frac{x_s - x}{h} \bigg) h_{t,s,n}\bigg(\frac{hy}{\sqrt{t-s}} \bigg) \, dy$$

$$\leq \frac{h}{\sqrt{t-s}} \int \bigg| H_1\bigg(y + \frac{x_s - x}{h} \bigg) \bigg| \bigg| h_{t,s,n}\bigg(\frac{hy}{\sqrt{t-s}} \bigg) - h_{t,s,n}(0) \bigg| \, dy$$

$$\leq C \bigg(\frac{h}{\sqrt{t-s}} \bigg)^2 \int \bigg| H_1\bigg(y + \frac{x_s - x}{h} \bigg) \bigg| |y| \, dy,$$

since y is restricted to the compact support of K. Thus,

$$\bigg| EH_1\bigg(\frac{x_s - x}{h} \bigg) H_1\bigg(\frac{x_t - x}{h} \bigg) \bigg| \tag{9.11}$$

$$\leq E\bigg\{ \bigg| H_1\bigg(\frac{x_s - x}{h} \bigg) \bigg| \bigg| E\bigg[H_1\bigg(\frac{x_t - x}{h} \bigg) \,\Big|\, \mathcal{F}_s \bigg] \bigg| \bigg\}$$

$$\leq C\Big(\frac{h}{\sqrt{t-s}}\Big)^2 \int E\Big|H_1\Big(\frac{x_s-x}{h}\Big)\Big|\Big|H_1\Big(y+\frac{x_s-x}{h}\Big)\Big||y|\,dy$$
$$\leq C\Big(\frac{h}{\sqrt{t-s}}\Big)^2\Big(\frac{h}{\sqrt{s}}\Big)\int |H_1(y)||H_1(y+z)||y|\,dz\,dy$$
$$\leq C\Big(\frac{h}{\sqrt{t-s}}\Big)^2\Big(\frac{h}{\sqrt{s}}\Big).$$

Taking the bounds (9.10) and (9.11) in $E\mathrm{I}_a^2$, we get

$$E\mathrm{I}_a^2 \leq Ch^3\sum_{t=1}^{n}\frac{1}{\sqrt{t}} + Ch^5\sum_{1\leq s<t\leq n}\Big(\frac{1}{\sqrt{t-s}}\Big)^2\Big(\frac{1}{\sqrt{s}}\Big) \tag{9.12}$$
$$\leq Ch^3\sqrt{n} + Ch^5\sqrt{n}\log n,$$

using the fact that $\sum_{1\leq s<t\leq n}\frac{1}{t-s}\frac{1}{\sqrt{s}} = 2\sqrt{n}\log n + O(\sqrt{n})$. Combining (9.9) and (9.12) gives (9.6) as required.

10. PROOF OF THEOREM 3.3

We rewrite $\hat{f}_a(x)$ as

$$\hat{f}_a(x) = \frac{\sum_{t=1}^{n}(y_t-\lambda\Delta x_t)K[(x_t-x)/h]}{\sum_{t=1}^{n}K[(x_t-x)/h]} + (\hat{\lambda}-\lambda)\frac{\sum_{t=1}^{n}\Delta x_t K[(x_t-x)/h]}{\sum_{t=1}^{n}K[(x_t-x)/h]}.$$

Recall that $\Delta x_t = \epsilon_t$ and $y_t - \lambda\Delta x_t = f(x_t) + u_{y.\epsilon,t}$. Since $u_{y.\epsilon,t}$ and ϵ_t satisfy Assumption 2, as in the proof of Theorem 3.1 which makes an application of (3.8), Theorem 3.3 will follow if we prove $\hat{\lambda} \to_P \lambda$, that is, we only need to prove the result (3.19).

We may write

$$\hat{\lambda} - \lambda = \frac{\sum_{t=1}^{n}[f(x_t) - \hat{f}(x_t)]\epsilon_t}{\sum_{t=1}^{n}\epsilon_t^2} + \frac{\sum_{t=1}^{n}u_{y.\epsilon,t}\epsilon_t}{\sum_{t=1}^{n}\epsilon_t^2}.$$

Since $\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 \to_{\text{a.s.}} E\epsilon_1^2 < \infty$ and $(\sum_{t=1}^{n}[f(x_t) - \hat{f}(x_t)]\epsilon_t)^2 \le \sum_{t=1}^{n}[f(x_t) - \hat{f}(x_t)]^2 \sum_{t=1}^{n}\epsilon_t^2$ by Hölder's inequality, the result (3.19) will follow if we prove

$$\frac{1}{n}\sum_{t=1}^{n}u_{y.\epsilon,t}\epsilon_t = o_P(1), \tag{10.1}$$

$$\frac{1}{n}\sum_{t=1}^{n}[f(x_t) - \hat{f}(x_t)]^2 = o_P(1). \tag{10.2}$$

Note that $\{u_{y.\epsilon,t}\epsilon_t, \mathcal{F}_t\}_{t\ge1}$, where $\mathcal{F}_t = \sigma(\epsilon_1, \ldots, \epsilon_t)$, form a martingale difference and $E(u_{y.\epsilon,t}\epsilon_t)^2 \le E(u_t^2\epsilon_t^2) < \infty$. Result (10.1) follows straightforwardly.

We next prove (10.2). Throughout the remaining part of the proof, we denote by $C, C_1, \ldots$ constants which may differ from line to line. Recall that $|f(x) - f(y)| \le C|x - y|$ and K has a compact support. We then have $(|x_t - x_j|/h)K[(x_t - x_j)/h] \le C_1K[(x_t - x_j)/h]$ and hence

$$J_{1j} := \frac{\sum_{t=1}^{n}[f(x_t) - f(x_j)]K[(x_t - x_j)/h]}{\sum_{t=1}^{n}K[(x_t - x_j)/h]} \le \frac{C\sum_{t=1}^{n}|x_t - x_j|K[(x_t - x_j)/h]}{\sum_{t=1}^{n}K[(x_t - x_j)/h]} \le C_1h. \tag{10.3}$$

Further, let $Y_j = \sum_{t=j+1}^{n}K[(x_t - x_j)/h]$ and

$$J_{2j} = \frac{\sum_{t=1}^{n}u_tK[(x_t - x_j)/h]}{\sum_{t=1}^{n}K[(x_t - x_j)/h]}.$$

Since $\sum_{t=1}^{n} K[(x_t - x_j)/h] \geq Y_j$, result (10.3) together with (3.4) yields

$$
\begin{aligned}
(10.4) \qquad \frac{1}{n}\sum_{j=1}^{n}[f(x_j) - \hat{f}(x_j)]^2 &\leq \frac{2}{n}\sum_{j=1}^{n}(J_{2j}^2 + J_{1j}^2) \\
&\leq \frac{2}{n}\sum_{j=1}^{n-2} Y_j^{-2}\left\{\sum_{t=1}^{n} u_t K[(x_t - x_j)/h]\right\}^2 + Ch^2 \\
&:= \Lambda_n + Ch^2, \quad \text{say.}
\end{aligned}
$$

Since $x_t - x_j = \sum_{s=j+1}^{t} \epsilon_s$ if $t > j$ and $x_t - x_j = -\sum_{s=t+1}^{j} \epsilon_s$ if $t < j$, by using similar arguments as in the proofs of (6.12) and/or Proposition 7.2, we have

$$
(10.5) \qquad E\left\{\sum_{t=1}^{n} u_t K[(x_t - x_j)/h]\right\}^2 \leq C(nh^2)^{1/2}.
$$

On the other hand, by noting that x_t is a random walk and hence a 1/2-null recurrent Markov chain with Lebesgue measure as an invariance measure (see Section 6 of Karlsen and Tjøstheim (2001)), it follows from Lemmas 3.4 and 3.5 and Theorem 5.1 that, for all $\delta > 0$,

$$
(10.6) \qquad \frac{1}{h\sqrt{n}}\sum_{t=1}^{n} K\left[\left(\sum_{s=1}^{t}\epsilon_s\right)\Big/h\right] \geq Cn^{-\delta} \quad \text{a.s.}
$$

Result (10.6), together with i.i.d. properties of ϵ_i, implies that $\forall\eta > 0$ and $\forall\delta > 0$, there exists an n_0 such that for all $n \geq n_0$,

$$
\begin{aligned}
(10.7) \qquad & P\left(\frac{1}{h\sqrt{n-j}} Y_j \geq Cn^{-\delta}, 1 \leq j \leq n-2\right) \\
&\quad = P\left(\frac{1}{h\sqrt{j}}\sum_{t=1}^{j} K\left[\left(\sum_{s=1}^{t}\epsilon_s\right)\Big/h\right] \geq Cn^{-\delta}, 1 \leq j \leq n-1\right) \\
&\quad \geq 1 - \eta.
\end{aligned}
$$

This, together with (10.5), yields that, for $\forall\eta > 0$ and $\forall\delta > 0$,

$$
\begin{aligned}
P(\Lambda_n \geq n^{-\delta}) &\leq \eta + P\left(\Lambda_n \geq n^{-\delta}, \frac{1}{h\sqrt{n-j}} Y_j \geq Cn^{-\delta}, 1 \leq j \leq n-2\right) \\
&\leq \eta + P\left(\sum_{j=1}^{n-2}\frac{1}{n-j}\left\{\sum_{t=1}^{n} u_t K\left[\frac{x_t - x_j}{h}\right]\right\}^2 \geq Ch^2 n^{1-3\delta}\right)
\end{aligned}
$$

$$\leq \eta + Ch^{-2}n^{3\delta-1}(\sqrt{n}h)\sum_{j=1}^{n-2}\frac{1}{n-j}$$
$$\leq \eta + Ch^{-1}n^{3\delta-1/2}\log n$$

whenever $n \geq n_0$. Taking $\delta = \delta_0/4$ and recalling $hn^{1/2+\delta_0} \to \infty$, we obtain $\Lambda_n = o_P(1)$. This, together with (10.4), proves (10.2) and also completes the proof of Theorem 3.3. *Q.E.D.*

REFERENCES

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843. [1901]

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley. [1920,1923,1926]

——— (1974): "Conditional Distributions and Tightness," *The Annals of Probability*, 2, 480–485. [1937]

BORODIN, A. N., AND I. A. IBRAGIMOV (1995): *Limit Theorems for Functionals of Random Walks*. Proceedings of the Steklov Institute of Mathematics, Vol. 195. Providence, RI: American Mathematical Society. [1921,1923]

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, Vol. 6B, ed. by J. Heckman and E. Leamer. Amsterdam: North Holland. [1901]

CHAN, N. H., AND C. Z. WEI (1987): "Asymptotic Inference for Nearly Nonstationary $AR(1)$ Process," *The Annals of Statistics*, 15, 1050–1063. [1905,1926]

ELLIOTT, G. (1998): "On the Robustness of Cointegration Methods When Regression Almost Have Unit Roots," *Econometrica*, 66, 149–158. [1905]

FLORENS, J.-P. (2003): "Inverse Problems and Structural Econometrics: The Example of Instrumental Variables," in *Advances in Economics and Econometrics: Theory and Applications—Eighth World Congress*. Econometric Society Monographs, Vol. 36, ed. by L. P. Hansen, S. J. Turnovsky, and M. Dewatripont. Cambridge, NY: Cambridge University Press. [1901]

HALL, P. (1977): "Martingale Invariance Principles," *The Annals of Probability*, 5, 875–887. [1920]

HALL, P., AND J. L. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904–2929. [1901,1914,1917]

KARLSEN, H. A., AND D. TJØSTHEIM (2001): "Nonparametric Estimation in Null Recurrent Time Series," *The Annals of Statistics*, 29, 372–416. [1946]

KARLSEN, H. A., T. MYKLEBUST, AND D. TJØSTHEIM (2007): "Nonparametric Estimation in a Nonlinear Cointegration Type Model," *The Annals of Statistics*, 35, 252–299. [1903,1904]

NEWEY, W. K., AND J. J. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. [1901]

NEWEY, W. K., J. L. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603. [1901]

PARK, J. Y., AND P. C. B. PHILLIPS (2001): "Nonlinear Regressions With Integrated Time Series," *Econometrica*, 69, 117–161. [1920]

PETROV, V. V. (1995): *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*. Oxford Studies in Probability, Vol. 4. New York: The Clarendon Press, Oxford University Press. [1929]

PHILLIPS, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547. [1905,1926]

——— (1988): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1044. [1905]

PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series Regression With Integrated Processes," *Review of Economic Studies*, 53, 473–495. [1903]

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression With I(1) Processes," *Review of Economic Studies*, 57, 99–125. [1912]

SCHIENLE, M. (2008): "Nonparametric Nonstationary Regression," Unpublished Ph.D. Thesis, University of Mannheim. [1903,1904,1920]

STOCK, J. H. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegration Vectors," *Econometrica*, 55, 1035–1056. [1903]

WANG, Q., AND P. C. B. PHILLIPS (2009): "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression," *Econometric Theory*, 25, 710–738. [1903,1904]

School of Mathematics and Statistics, The University of Sydney, Sydney, NSW 2006, Australia; qiying@maths.usyd.edu.au

and

Cowles Foundation, Yale University, Box 208281, New Haven, CT 06520-8281, U.S.A. and University of Auckland, Auckland, New Zealand and University of Southampton, Southampton, U.K., and Singapore Management University, Singapore; peter.phillips@yale.edu.

Manuscript received February, 2008; final revision received February, 2009.