Structural Nonparametric Cointegrating Regression *

Qiying Wang
School of Mathematics and Statistics
The University of Sydney

Peter C. B. Phillips
Yale University, University of Auckland
University of York & Singapore Management University

August 25, 2008

Abstract

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 useable in practical work. 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.

Key words and phrases: Brownian Local time, Cointegration, Functional regression, Gaussian process, Integrated process, Kernel estimate, Near integration, Nonlinear functional, Nonparametric regression, Structural estimation, Unit root.

JEL Classification: C14, C22.

*The authors thank the coeditor two referees for helpful comments on the original version. Wang acknowledges partial research support from Australian Research Council. Phillips acknowledges partial research support from a Kelly Fellowship and the NSF under Grant No. SES 06-47086.

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 (2004). Overviews of the ill-posed inverse literature are given in Florens (2003) and Carrasco, Florens and Renault (2006). 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 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 many 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, in order 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 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. 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 allowing for some forms of nonstationarity and endogeneity in the regressor. Schienle also considers 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 (2008) 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 a simulation experiment exploring the finite sample performance of the kernel estimator. 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 as Appendices to the paper.

2 Model and Assumptions

We consider the following nonlinear structural model of cointegration

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

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

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

where $K_h(s) = \frac{1}{h}K(s/h)$, $K(x)$ 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 (2008, 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 (iid) continuous random variables with $E\epsilon_1 = 0$, $E\epsilon_1^2 = 1$, 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 iid 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. $u_t = u(\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-m_0+1}, \lambda_t, \lambda_{t-1}, ..., \lambda_{t-m_0+1})$ satisfies $Eu_t = 0$ and $Eu_t^4 < \infty$ for $t \geq m_0$, where $u(x_1, ..., x_{m_0}, y_1, ..., y_{m_0})$ 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 an $0 < \gamma \leq 1$ 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 near unit root ($\kappa \neq 0$) regressor by virtue of the localizing coefficient κ and is standard in the near integrated regression framework (Phillips, 1987, 1988; Chan and Wei, 1987). 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. In the asymptotic development below, m_0 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 is 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 that Assumption 2 allows for a nonlinear generating mechanisms 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 is 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 having a compact support.

Assumptions 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 occuring 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 $nh^{2(1+2\gamma)} \to 0$,

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

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

Remarks

(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

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

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

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

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 that of 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.

(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

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

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

$$\Theta_{1n} := \sum_{t=1}^n u_t K[(x_t - x)/h] = O_P\{(\sqrt{n}h)^{1/2}\}, \tag{3.4}$$

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

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

$$\Theta_{3n} := 1\Big/\sum_{t=1}^n K[(x_t - x)/h] = o_P\{a_n/(\sqrt{n}h)\}. \tag{3.6}$$

On the other hand, it is readily seen that

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

By virtue of (3.5) and (3.6) 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

$$\Big\{(nh^2)^{-1/4}\sum_{k=1}^{[nt]}u_k\,K[(x_k-x)/h]\,,\,(nh^2)^{-1/2}\sum_{k=1}^{n}K[(x_k-x)/h]\Big\}$$
$$\to_D\ \big\{d_0\,N\,L^{1/2}(t,0),\,d_1\,L(1,0)\big\},\tag{3.7}$$

on $D[0,1]^2$, where $d_0^2=|\phi|^{-1}\,E(u_{m0}^2)\int_{-\infty}^{\infty}K^2(s)dt$, $d_1=|\phi|^{-1}\int_{-\infty}^{\infty}K(s)ds$, $L(t,0)$ is the local time process at the origin of the Gaussian diffusion process $\{J_\kappa(t)\}_{t\ge 0}$ defined by

$$J_\kappa(t)=W(t)+\kappa\int_0^t e^{(t-s)\kappa}W(s)ds\tag{3.8}$$

and $\{W(t)\}_{t\ge0}$ 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

$$L(t,a)=\lim_{\epsilon\to0}\frac{1}{2\epsilon}\int_0^t I\{|J_\kappa(r)-a|\le\epsilon\}dr.\tag{3.9}$$

Indeed, since $P(L(1,0)>0)=1$, the required result (3.2) follows by (3.7) and the continuous mapping theorem. It remains to prove (3.4)–(3.7), which are established in the Appendix. As for (3.7), it is clearly sufficient for the required result to show that the finite dimensional distributions converge in (3.7).

(c) Results (3.2) and (3.7) 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 (3.7). Explicitly,

$$(nh^2)^{1/4}\,(\hat f(x)-f(x))\ \ \to_D\ \ d_0\,d_1^{-1}\,N\,L^{-1/2}(1,0),\tag{3.10}$$

whenever $nh^2\to\infty$ and $nh^{2(1+2\gamma)}\to 0$. Again, this is different from that of 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 (3.10). 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 $E\left(u_t^2\right) = \sigma_u^2$, 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 the Appendix in Section 9 that

$$\left(nh^2\right)^{1/4}\left[\hat{f}\left(x_t\right) - f\left(x\right) - \frac{h^2}{2} f''\left(x\right)\int_{-\infty}^{\infty} y^2 K(y)dy\right] \Rightarrow \frac{N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K^2\left(s\right)ds\right)}{L\left(1,0\right)^{1/2}}, \tag{3.11}$$

provided $nh^{14} \to 0$ and $nh^2 \to \infty$.

(d) As is clear from the second member of (3.7), 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 that 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.

(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, viz., $Eu_{m_0}^2$, 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 $\varepsilon_t, ..., \varepsilon_{t-m_0+1}$ 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 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 $g(x)$.

(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

$$\frac{1}{n^{1/4}h^{1/2}} \sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right) = o_p(1) \quad \text{for } x \neq x'. \tag{3.12}$$

To show (3.12), note that if $x_t/\sqrt{t}$ has a bounded density $h_t(y)$, as in WP, we have

$$\begin{aligned}
& E\Big[K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right)\Big] \\
= & \int \Big[K\left(\frac{\sqrt{t}y - x}{h}\right) K\left(\frac{\sqrt{t}y - x'}{h}\right)\Big] h_t(y)dy \\
= & \; h\, t^{-1/2} \int K(y)\, K\left[y + (x - x')/h\right] h_t[(yh + x)/\sqrt{t}]dy \\
\sim & \; h\, t^{-1/2} h_t(0) \int K(y)\, K\left[y + (x - x')/h\right] dy = o(h\, t^{-1/2}),
\end{aligned}$$

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

$$\frac{1}{n^{1/4}h^{1/2}} \sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right) = o_p\left(\frac{h^{1/2}}{n^{1/4}} \sum_{t=1}^{n} \frac{1}{t^{1/2}}\right) = o_p\left(h^{1/2}\right).$$

This result and theorem 2.1 of WP give

$$\begin{aligned}
& \frac{1}{n^{1/4}h^{1/2}} \sum_{t=1}^{n} \begin{bmatrix} K\left(\frac{x_t - x}{h}\right)^2 & K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right) \\ K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_t - x'}{h}\right) & K\left(\frac{x_t - x'}{h}\right)^2 \end{bmatrix} \\
\Rightarrow & \; L(1, 0) \begin{bmatrix} \int K(s)^2 ds & 0 \\ 0 & \int K(s)^2 ds \end{bmatrix}.
\end{aligned}$$

Following the same line of argument 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 $\left(\hat{f}(x), \hat{f}(x')\right)$ 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.

(g) The error variance term Eu_{m0}^2 in the limit distribution (3.2) may be estimated by a localized version of the usual residual based method. Indeed, by letting

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

we have the following theorem under minor additional conditions.

THEOREM 3.2. *In addition to Assumptions 1-4,* $Eu_{m0}^8 < \infty$ *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* $h \to 0$,

$$\hat{\sigma}_n^2 \to_p Eu_{m0}^2. \tag{3.13}$$

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

$$(nh^2)^{1/4} \left(\hat{\sigma}_n^2 - Eu_{m0}^2\right) \to_D \sigma_1 N L^{-1/2}(1,0), \tag{3.14}$$

where N *and* $L(1,0)$ *are defined as in (3.7) and* $\sigma_1^2 = E(u_{m0}^2 - Eu_{m0}^2)^2 \int_{-\infty}^{\infty} K^2(s) ds / \int_{-\infty}^{\infty} K(x) dx$.

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 (3.13). 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_{m0}^2, not a conditional variance, and the limit distribution of $\hat{\sigma}_n^2$ given in (3.14) 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_{m0}^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 $\hat{\sigma}_n^2 \to_p Eu_{m0}^2 g(x)^2$.

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

$$\begin{aligned} y_t &= f(x_t) + u_t, \quad \Delta x_t = \epsilon_t, \\ u_t &= (\lambda_t + \theta \epsilon_t) / \left(1 + \theta^2\right)^{1/2}, \end{aligned}$$

where (ϵ_t, λ_t) are *iid* $N(0, \sigma^2 I_2)$. 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. Figs. 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 $\{x = 0.01k : k = 0, 1, ..., 100\}$ for $[0, 1]$ and $\{x = -1 + 0.02k; k = 0, 1, ..., 100\}$ for $[-1, 1]$ based on 10,000 replications. Simulations were performed for $\theta = 1$ (weak endogeneity) and $\theta = 100$ (strong endogeneity), with $\sigma = 0.1$, and for the sample size $n = 500$. A Gaussian kernel was used with bandwidths $h = n^{-10/18}, n^{-1/2}, n^{-1/3}, n^{-1/5}$.

Table 1 shows the performance of the regression estimate $\hat{f}$ computed over various bandwidths, h, and endogeneity parameters, θ, for the two models. Since x_t is recurrent and wanders over the real line, some simulations are inevitably thin in subsets of the chosen domains and this inevitably affects performance because the local sample size is small. In both models the degree of endogeneity (θ) in the regressor has a negligible effect on the properties of the kernel regression estimate when h is small. It is also clear that estimation bias can be substantial, particularly for model A with bandwidth $h = n^{-1/5}$, corresponding to the conventional rate for stationary series. Bias is substantially reduced for the smaller bandwidths $h = n^{-1/2}, n^{-1/3}$ at the cost of an increase in dispersion and is further reduced when $h = n^{-10/18}$ although this choice and $h = n^{-1/2}$ violate the condition $nh^2 \to \infty$ of theorem 3.1. The downward bias in the case of $\hat{f}_A$ over the domain $[0, 1]$ appears to be due to the periodic nature of the function f_A and the effects of smoothing over x values for which the function is negative. The bias in $\hat{f}_B$ is similarly towards the origin over the whole domain $[-1, 1]$. The performance characteristics seem to be little affected by the magnitude of the endogeneity parameter θ. For model A, finite sample performance in terms of MSE seems to be optimized for h close to $n^{-1/2}$. For model B, $h = n^{-1/5}$ delivers the best MSE performance largely because of the substantial gains in variance reduction with the larger bandwidth that occur in this case. Thus, bias reduction through choice of a very small bandwidth may be important in overall finite sample performance for some regression functions but much less so for other functions. Of course, if $h \to 0$ so fast that $nh^2 \not\to \infty$ then the "signal" $\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) \not\to \infty$ and the kernel estimate is not consistent.

Figs. 1 and 2 show results for the Monte Carlo approximations to $E\left(\hat{f}_A(x)\right)$ and $E\left(\hat{f}_B(x)\right)$ 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 $\theta = 100$. Figs 3 and 4 show the Monte Carlo approximations to $E\left(\hat{f}_A(x)\right)$ and $E\left(\hat{f}_B(x)\right)$ 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 are quite wide, reflecting the much 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 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 $\left(nh^2\right)^{1/4} = n^{1/12}$, which is far slower than the rate $(nh)^{1/2} = n^{2/5}$ for conventional kernel regression.

Table 1

Model A: $f_A(x) = \sum_{j=1}^{4} \frac{(-1)^{j+1}\sin(j\pi x)}{j^2}$

θ	h	**Bias**	**Std**	**MSE**
100	$n^{-10/18}$	0.056	0.234	0.066
	$n^{-1/2}$	0.059	0.229	0.064
	$n^{-1/3}$	0.106	0.208	0.066
	$n^{-1/5}$	0.274	0.193	0.145
1	$n^{-10/18}$	0.058	0.235	0.067
	$n^{-1/2}$	0.061	0.229	0.065
	$n^{-1/3}$	0.108	0.209	0.067
	$n^{-1/5}$	0.276	0.193	0.145

Model B: $f_B(x) = x^3$

θ	h	**Bias**	**Std**	**MSE**
100	$n^{-10/18}$	0.0005	0.801	0.651
	$n^{-1/2}$	0.0003	0.739	0.556
	$n^{-1/3}$	0.0005	0.541	0.305
	$n^{-1/5}$	0.0021	0.387	0.190
1	$n^{-10/18}$	0.0027	0.802	0.648
	$n^{-1/2}$	0.0027	0.740	0.553
	$n^{-1/3}$	0.0033	0.541	0.302
	$n^{-1/5}$	0.0051	0.395	0.188

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 cdf Φ. Figs 5 and 6 show the empirical coverage probabilities of these pointwise asymptotic confidence interval for f_A and f_B over 100 equispaced points on the domains $[0,1]$ and $[-1,1]$, using a standard normal kernel, various bandwidths as shown, and setting $\alpha = 0.05$ and $n = 500$. For both functions the coverage rates are closer to the nominal level of 95% and more uniform over the respective domains for the smaller bandwidth choices. For function f_A there is evidence of substantial undercoverage in the interior of the $[0, 1]$ interval, where the nonparametric estimator was seen to be biased (Fig. 1) for larger bandwidths. For function f_B, the undercoverage is also substantial for the larger bandwidths but in this case away from the origin, while at the origin there is some evidence of overcoverage for the larger bandwidths. For both functions, the smaller bandwidth choice seems to give more uniform performance, the coverage probability is around 90% and is close enough to the nominal level to be satisfactory.

5 Conclusion

The two main results in the present paper have important implications for applications. 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 nonparametric regression in cointegrating models, avoiding the need for instrumentation and completely eliminating ill-posed functional equation inversions. Second, functional estimation of (2.1) is straightforward in practice and may be accomplished by standard kernel methods. These methods yield consistent estimates that have a mixed normal limit distribution, thereby validating conventional methods of inference in the nonstationary nonparametric setting.

The results open up some interesting possibilities for functional regression in empirical research with integrated and near integrated processes. In addition to many possible empirical applications with the methods, there are some interesting extensions of the ideas presented here to other useful models involving nonlinear functions of integrated processes. In particular, additive nonlinear cointegration models (c.f. Schienle, 2008) and partial linear cointegration models may be treated in a similar way to (2.1), but multiple non-additive regression models present difficulties arising from the nonrecurrence of the limit processes in high dimensions (c.f. Park and Phillips, 2000). There are also issues of specification testing, functional form tests, and cointegration tests, which may now be addressed using these methods. It will also be of interest to consider the properties of instrumental variable procedures in the present nonstationary context. We plan to report on some of these extensions in later work.

6 Proof of Theorem 3.1

As shown in Remark (b), the proof of the theorem essentially amounts to proving (3.4)–(3.7). 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 $\alpha_n^{(1)}$, $\alpha_n^{(2)}$,..., $\alpha_n^{(k)}$ $(1 \le n \le \infty)$ are random elements of $D[0,1]$, we will understand the condition

$$(\alpha_n^{(1)}, \alpha_n^{(2)}, ..., \alpha_n^{(k)}) \to_D (\alpha_\infty^{(1)}, \alpha_\infty^{(2)}, ..., \alpha_\infty^{(k)})$$

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

$$P\big(\alpha_n^{(1)} \in A_1, \alpha_n^{(2)} \in A_1, ..., \alpha_n^{(k)} \in A_k\big) \to P\big(\alpha_\infty^{(1)} \in A_1, \alpha_\infty^{(2)} \in A_2, ..., \alpha_\infty^{(k)} \in A_k\big).$$

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

In order to prove (3.7), 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 $\gamma_j \ge 0, j = 1, 2, ..., r$, and any $0 \le s < t \le t_0 < t_1 < ... < t_r < \infty$,*

$$\begin{aligned} E\Big(e^{-\sum_{j=1}^r \gamma_j[\psi(t_j)-\psi(t_{j-1})]}\big[q(t)-q(s)\big] \mid \mathcal{F}_s\Big) &= 0, \quad a.s., \\ E\Big(e^{-\sum_{j=1}^r \gamma_j[\psi(t_j)-\psi(t_{j-1})]}\big\{[q(t)-q(s)]^2 - [\psi(t)-\psi(s)]\big\} \mid \mathcal{F}_s\Big) &= 0, \quad a.s. \end{aligned}$$

then the finite-dimensional distributions of the process $(q(t), \xi)_{t\ge 0}$ coincide with those of the process $(W[\psi(t)], \xi)_{t\ge 0}$, 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, page 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 $t_0 < t_1, ..., t_r < \infty$, $\alpha_j \in R$

and $s \in R$,

$$
\begin{aligned}
&Ee^{i\sum_{j=1}^{r}\alpha_j[q(t_j)-q(t_{j-1})]+is\xi} \\
&\quad = E\Big[e^{i\sum_{j=1}^{r-1}\alpha_j[q(t_j)-q(t_{j-1})]+is\xi}\,E\Big(e^{i\alpha_r[q(t_r)-q(t_{r-1})]}\mid \mathcal{F}_{t_{r-1}}\Big)\Big] \\
&\quad = E\Big[\,e^{-\frac{\alpha_r^2}{2}[\psi(t_r)-\psi(t_{r-1})]}\,e^{i\sum_{j=1}^{r-1}\alpha_j[q(t_j)-q(t_{j-1})]+is\xi}\,\Big] \\
&\quad = \ldots = Ee^{-\frac{\alpha_r^2}{2}\sum_{j=1}^{r}[\psi(t_j)-\psi(t_{j-1})]+is\xi},
\end{aligned}
$$

which yields the stated result. $\square$

By virtue of Lemma 6.1, we now obtain the proof of (3.7). Technical details of some subsidiary results that are used in this proof are given in the next section. Set

$$
\begin{aligned}
\zeta_n(t) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_k, \quad \psi_n(t) = \frac{1}{d_0^2\sqrt{nh^2}}\sum_{k=1}^{[nt]}u_k^2K^2[(x_k-x)/h], \\
S_n(t) &= \frac{1}{d_0(nh^2)^{1/4}}\sum_{k=1}^{[nt]}u_kK[(x_k-x)/h],
\end{aligned}
$$

for $0 \le t \le 1$, where $d_0$is defined as in (3.7).

We will prove in Propositions 7.1 and 7.2 that $\zeta_n(t) \Rightarrow W(t)$ 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

$$\{S_n(t),\ \psi_n(t),\ \zeta_n(t)\}_{n\ge1}$$

is tight on $D[0,1]^3$. Hence, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

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

on $D[0,1]^5$, where $\eta(t)$ is a process continuous with probability one by noting (7.25) 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 $0 \le s < t \le 1$,

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

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

then it follows from Lemma 6.1 that the finite-dimensional distributions of $(\eta(t), \psi(1))$ coincide with those of $\{N\,L^{1/2}(t,0), L(1,0)\}$, where N is normal variate independent of

$L(t,0)$. The result (3.7) therefore follows, since $\eta(t)$ does not depend on the choice of the subsequence.

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

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

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

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 prove

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

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

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

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

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

$$\begin{aligned} & E\Big\{\sum_{k=[nt_{j-1}]+1}^{[nt_j]} u_k\, K[(x_k - x)/h] e^{i\mu_j^*[\zeta_n(t_j) - \zeta_n(t_{j-1})] + i\chi(t_{j-1})}\Big\} \\ &= o[(nh^2)^{1/4}], \quad (6.8) \\ & E\Big\{\big[\sum_{k=[nt_{j-1}]+1}^{[nt_j]} u_k\, K[(x_k - x)/h]\big]^2 - \sum_{k=[nt_{j-1}]+1}^{[nt_j]} u_k^2\, K^2[(x_k - x)/h]\Big\}\, e^{i\mu_j^*[\zeta_n(t_j) - \zeta_n(t_{j-1})] + i\chi(t_{j-1})} \\ &= o[(nh^2)^{1/2}], \quad (6.9) \end{aligned}$$

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

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

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

where $x^*_{s,t}$ depends only on $(..., \epsilon_{s-1}, \epsilon_s)$ 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,

$$\begin{aligned} &\sup_{y,0\le s<m\le n} E\Big\{ \sum_{k=s+1}^{m} u_k\, K[(y+x'_{s,k})/h]\, e^{i\mu \sum_{i=1}^m \epsilon_i/\sqrt{n}} \Big\} \\ &\quad = o[(nh^2)^{1/4}], \qquad (6.11) \\ &\sup_{y,0\le s<m\le n} E\Big(\Big\{ \sum_{k=s+1}^{m} u_k\, K[(y+x'_{s,k})/h] \Big\}^2 - \sum_{k=s+1}^{m} u_k^2\, K^2[(y+x'_{s,k})/h] \Big)\, e^{i\mu \sum_{i=1}^m \epsilon_i/\sqrt{n}} \\ &\quad = o[(nh^2)^{1/2}]. \qquad (6.12) \end{aligned}$$

This follows from Proposition 7.5. The proof of (3.7) is now complete.

We next prove (3.4)-(3.6). In fact, it follows from Proposition 7.3 that, uniformly in n, $E\Theta^2_{1n}/(nh^2)^{1/2} = d_0^2\, ES_n^2(1) \le C$. This yields (3.4) 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 proof of (5.18) in WP yields (3.5). As for (3.6), 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.

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 same as in the previous section except when explicitly mentioned.

PROPOSITION 7.1. *We have*

$$\zeta_n(t) \Rightarrow W(t) \quad \text{and} \quad \zeta'_n(t) := \frac{1}{\sqrt{n}\phi} \sum_{k=1}^{[nt]} \rho^{[nt]-k}\, \eta_k \Rightarrow J_\kappa(t) \quad \text{on } D[0,1], \qquad (7.1)$$

where $\{W(t), t\ge 0\}$ is a standard Brownian motion and $J_\kappa(t)$ is defined as in (3.8).

Proof. The first statement of (7.1) is well-known. In order to $\zeta'_n(t) \Rightarrow J_\kappa(t)$, for each fixed $l \ge 1$, put

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

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

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

Therefore, for fixed $l \geq 1$,

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

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, page 25), to prove $\zeta_n'(t) \Rightarrow J_\kappa(t)$, it suffices to show that for any $\delta > 0$,

$$\limsup_{n\to\infty} P\left\{ \sup_{0\le t\le 1} |R([nt], l)| \geq \delta \sqrt{n} \right\} \;=\; 0, \quad (7.3)$$

for fixed $l \geq 1$ and

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

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

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

We next prove (7.4). Noting

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

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

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

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.

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

$$\frac{1}{\sqrt{nh^2}} \sum_{k=1}^{[nt]} K^i[(x_k - x)/h] \;\Rightarrow\; d_i\, L(t,0), \quad i = 1,\ 2, \tag{7.5}$$

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

on $D[0,1]$, where $d_i = |\phi|^{-1} \int_{-\infty}^{\infty} K^i(s)ds, i = 1,\ 2$ and $d_0^2 = |\phi|^{-1}\, Eu_{m_0}^2 \int_{-\infty}^{\infty} K^2(s)ds$, and $L(t,s)$ is the local time process of the Gaussian diffusion process $\{J_\kappa(t), t \geq 0\}$ defined by (3.8), 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 $u \in R$.*

In order 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 $I_{k,l}^{(s)}$ and $II_k^{(s)}$ defined by

$$\begin{aligned} I_{k,l}^{(s)} &= E\Big[r(x'_{s,k}/h)\, r_1(x'_{s,l}/h)\, g(u_k)\, g_1(u_l)\, \exp\big\{i\mu \sum_{j=1}^{l} \epsilon_j/\sqrt{n}\big\}\Big], \\ II_k^{(s)} &= E\Big[r(x'_{s,k}/h)\, g(u_k)\, \exp\big\{i\mu \sum_{j=1}^{k} \epsilon_j/\sqrt{n}\big\}\Big], \end{aligned} \tag{7.7}$$

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 depending 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 $k \geq s+1$,

$$|II_k^{(s)}| \;\leq\; C\, h/\sqrt{k-s}. \tag{7.8}$$

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

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

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 $k \geq s+1$,*

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

where we define $\sum_{j=t/2}^{\infty} = \sum_{j \geq t/2}$.

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

The proof of Lemma 7.2. Write $x_k'' = x_{0,k}'$ and $I_{k,l} = I_{k,l}^{(0)}$. 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

$$\begin{aligned} I_{k,l} &= E\Big[r(x_k''/h)\, r_1(x_l''/h)\, g(u_k)\, g_1(u_l)\, \exp\{i\mu \sum_{j=1}^{l} \epsilon_j/\sqrt{n}\}\Big] \\ &= \int\!\!\int E\Big\{e^{-it\, x_k''/h}\, e^{i\lambda\, x_l''/h}\, g(u_k)\, g_1(u_l)\, e^{i\mu \sum_{j=1}^{l} \epsilon_j/\sqrt{n}}\Big\}\, \hat{r}(t)\, \overline{\hat{r}_1(\lambda)}\, dt\, d\lambda. \end{aligned}$$

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

$$x_l'' \;\; = \;\; \sum_{q=1}^{l} \epsilon_q \sum_{j=0}^{l-q} \rho^{l-q-j}\, \phi_j = \Big(\sum_{q=1}^{k} + \sum_{q=k+1}^{l-m_0} + \sum_{q=l-m_0+1}^{l}\Big) \epsilon_q\, a_{l,q},$$

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

$$\begin{aligned} |I_{k,l}| \;\; \leq \;\; & \int \Big| E\{e^{iz^{(2)}/h}\}\Big| \Big| E\{e^{iz^{(3)}/h}\, g_1(u_l)\}\Big|\, |\hat{r}_1(\lambda)| \\ & \qquad \Big(\int \big| E\{e^{iz^{(1)}/h}\, g(u_k)\}\big|\, |\hat{r}(t)|\, dt\Big)\, d\lambda, \end{aligned} \tag{7.11}$$

where

$$\begin{aligned} z^{(1)} &= \sum_{q=1}^{k} \epsilon_q \big(\lambda\, a_{l,q} - t\, a_{k,q} + u\, h/\sqrt{n}\big), \\ z^{(2)} &= \sum_{q=k+1}^{l-m_0} \epsilon_q\, \big(\lambda\; a_{l,q} + u\, h/\sqrt{n}\big), \\ z^{(3)} &= \sum_{q=l-m_0+1}^{l} \epsilon_q\, \big(\lambda\; a_{l,q} + u\, h/\sqrt{n}\big). \end{aligned}$$

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,

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

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 sufficiently large,

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

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

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

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

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

By virtue of (7.13), it is readily seen that $B_1 \geq C\,k$, whenever $\#(\Omega_2) \leq \sqrt{k}$, 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

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

since $E\epsilon_1 = 0$, $E\epsilon_1^2 = 1$ and ϵ_1 has a density. See, e.g., Chapter 1 of Petrov (1995). Also note that

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

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

$$\begin{aligned} \big|Ee^{i\,W^{(1)}/h}\big| \;\; &\leq \;\; \exp\big\{ -\gamma_1\#(\Omega_{1n}) - \gamma_2\, h^{-2} \sum_{q\in\Omega_2} (\lambda\, a_{l,q} - t a_{k,q})^2 \big\} \\ &\leq \;\; \exp\big\{ -\gamma_1\#(\Omega_{1n}) - \gamma_2\, B_1\, h^{-2}(t - \lambda B_2/B_1)^2 \big\} \end{aligned}$$

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

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

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

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

$$l - k \ge 2k_0 \text{ and } k \ge k_0; \quad l - k \le 2k_0 \text{ and } k \ge k_0; \quad l > k \text{ and } k \le k_0,$$

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 $q \le s/2$,

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

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 $l - k \ge 2k_0$ and $k_0 \ge 2m_0$, that

$$\big|Ee^{iz^{(2)}/h}\big| \le \Pi_{q=k}^{(l+k)/2}\big|Ee^{i\epsilon_q\,\lambda\, a_{l,q}/h}\big| \le \begin{cases} e^{-\gamma_3(l-k)} & \text{if } |\lambda| \ge \delta\, h, \\ e^{-\gamma_4\,(l-k)\lambda^2/h^2} & \text{if } |\lambda| \le \delta\, h. \end{cases} \tag{7.16}$$

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

$$\begin{aligned}
\Big|\,E\Big\{e^{iz^{(3)}/h}\, g_1(u_l)\Big\}\Big| &= \Big|\,E\Big\{\big(e^{iz^{(3)}/h} - 1\big)\, g_1(u_l)\Big\}\Big| \le h^{-1}\, E\big[|z^{(3)}|\,|g_1(u_l)|\big] \\
&\le C\,(E\epsilon_1^2)^{1/2}(Eg_1^2(u_l))^{1/2}\,|\lambda|\, h^{-1}.
\end{aligned} \tag{7.17}$$

We also have

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

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 $E\{e^{iz^{(3)}/h}\, g_1(u_l)\} = E\big\{e^{i\epsilon_l \phi_0 \lambda\, \rho^{n-l}/h} g^*(\epsilon_l)\big\}$, where $g^*(\epsilon_l) = E\big[e^{i(z^{(3)} - \epsilon_l \phi_0 \lambda\, \rho^{n-l})/h} g_1(u_l) \mid \epsilon_l\big]$. 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 a n_0 (A_0 respectively) such that, for all $n \geq n_0$ ($|\lambda|/h \geq A_0$ respectively), $|\, E\{e^{iz^{(3)}/h}\, g_1(u_l)\}| \leq \epsilon$. This, together with (7.12) and (7.16) with $\delta = A_0$, yields that

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

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

$$\begin{aligned} I_{k,l}^{(1)} &:= \int_{|\lambda| \leq A_0 h} \Big| E\big\{e^{iz^{(2)}/h}\big\}\Big| \Big| E\big\{e^{iz^{(3)}/h}\, g_1(u_l)\big\}\Big| \Lambda(\lambda, k)\, |\hat{r}_1(\lambda)|\, d\lambda \\ &\leq C\, (k^{-2} + h/\sqrt{k})\, h^{-1} \int_{|\lambda| \leq A_0 h} \lambda\, e^{-\gamma_4\, (l-k) \lambda^2/h^2}\, d\lambda \\ &\leq C\, h\, (l-k)^{-1}\, (k^{-2} + h/\sqrt{k}). \end{aligned}$$

The result (7.10) in **Case I** now follows from

$$I_{k,l} \leq I_{k,l}^{(1)} + I_{k,l}^{(2)} \leq C\, \big[\epsilon\, (l-k)^{-2} + h\, (l-k)^{-1}\big]\, (k^{-2} + h/\sqrt{k}).$$

Case II. $l - k \leq 2k_0$ and $k \geq k_0$. In this case, we only need to show that

$$|I_{k,l}| \;\leq\; C\, (\epsilon + h)\, \big(k^{-2} + h/\sqrt{k}\big). \tag{7.19}$$

In fact, as in (7.11), we have

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

where

$$\begin{aligned} z^{(4)} &= \sum_{q=1}^{k-m_0} \epsilon_q \big[\lambda\, a_{l,q} - t\, a_{k,q}\big], \\ z^{(5)} &= \sum_{q=k-m_0+1}^{l} \epsilon_q\, \big(\lambda\, a_{l,q} + u\, h/\sqrt{n}\big) - t \sum_{q=k-m_0+1}^{k} \epsilon_q\, a_{k,q}. \end{aligned}$$

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

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

Note that

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

For any $\epsilon > 0$, similar to the proof of (7.18), there exists a n_0 (A_0 respectively) such that, for all $n \geq n_0$ ($|\lambda|/h \geq A_0$ respectively), $|\, E\{e^{iz^{(5)}/h}\, g(u_k)\, g_1(u_l)\}| \leq \epsilon$. By virtue of these facts, we have

$$\begin{aligned} |I_{k,l}| \;\; &\leq \;\; \Big(\int_{|\lambda| \leq A_0 h} + \int_{|\lambda| > A_0 h}\Big) \Big|\, E\big\{e^{iz^{(5)}/h}\, g(u_k)\, g_1(u_l)\big\}\Big| \, |\hat{r}_1(\lambda)| \, \Lambda_1(\lambda, k) \, d\lambda \\ &\leq \;\; C\Big(\int_{|\lambda| \leq A_0 h} d\lambda + \epsilon \int_{|\lambda| > A_0 h} |\hat{r}_1(\lambda)| \, d\lambda\Big) \, (k^{-2} + h/\sqrt{k}) \\ &\leq \;\; C\,(\epsilon + h)\, \big(k^{-2} + h/\sqrt{k}\big). \end{aligned}$$

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

Case III. $l > k$ and $k \leq k_0$. In this case, we only need to prove

$$|I_{k,l}| \;\; \leq \;\; C\,\big[\epsilon\,(l-k)^{-3/2} + h\,(l-k)^{-1}\big]. \tag{7.21}$$

In order to prove (7.21), split $l > k$ into $l - k \geq 2k_0$ and $l - k \leq 2k_0$. 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

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

We omit the details. The proof of Lemma 7.2 is now complete.

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 $\int |\hat{r}(\lambda)| d\lambda \leq \int |\hat{K}(\lambda)| d\lambda < \infty$ 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 a n_0 such that, whenever $n \geq n_0$,

$$\begin{aligned} \sum_{1 \leq k < l \leq n} |I_{k,l}| \;\; &\leq \;\; C \sum_{1 \leq k < l \leq n} \big[\epsilon\,(l-k)^{-2} + h\,(l-k)^{-1}\big] \Big(k^{-2} + h/\sqrt{k}\Big) \\ &\leq \;\; C\,(\epsilon + h \sum_{k=1}^{n} k^{-1}) \sum_{k=1}^{n} \Big(k^{-2} + h/\sqrt{k}\Big) \\ &\leq \;\; C(\epsilon + h\, \log n)\, (C + \sqrt{n}\, h). \end{aligned}$$

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.

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

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

by Proposition 7.1 and the Skorohod-Dudley-Wichura representation theorem. Also, we may claim that

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

$$\zeta_n'(l/n) - \zeta_n'(k/n) \quad = \quad S_{1l} + S_{2l}$$

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 $d_{l,k,n}^2 = \frac{\rho^{2l}}{n\phi^2} \sum_{i=k+1}^{l} \rho^{-2i} (\sum_{j=0}^{l-i} \rho^{-j}\phi_j)^2$ and $\mathcal{F}_{t,n} = \sigma(..., \epsilon_{t-1}, \epsilon_t)$. 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

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

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

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

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

$$\begin{aligned} & \sup_{l-k\ge\delta n} \sup_{|u|\le\delta} |h_{l,k}(x - S_{1l}/d_{l,k,n}) - h_{l,k}(-S_{1l}/d_{l,k,n})| \\ \le\ & 2 \sup_{l-k\ge\delta n} |h_{l,k}(x) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2})| + \frac{1}{\sqrt{2\pi}} \sup_{|u|\le\delta} \sup_x |e^{-(x+u)^2/2 - e^{-x^2/2}}| \\ \to\ & 0, \end{aligned}$$

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

$$\sup_{0\le t\le 1}\Big|\frac{\phi}{\sqrt{nh^2}}\sum_{k=1}^{[nt]} K^i\big[(\sqrt{n}\,\phi\,\hat{\zeta}_n'(k/n) - x)/h\big] - L(t,0)\int_{-\infty}^{\infty} K^i(s)ds\Big| \to_P 0.$$

This, together with the fact that $\hat{\zeta}_n'(k/n) =_d \zeta_n'(k/n) = x_k/(\sqrt{n}\phi), 1 \le k \le n$ for each $n \ge 1$, implies that the finite dimensional distributions of $T_{in}(t) := \frac{1}{\sqrt{nh^2}}\sum_{k=1}^{[nt]} K^i\big[(x_k - x)/h\big]$ converge to those of $d_i\, L(t, 0)$. On the other hand, by applying for Theorem 15.2 of Billingsley (1968), it is easy to show that $T_{in}(t), n \ge 1$ is tight since $K(.)$ is positive. Hence $T_{in}(t) \Rightarrow d_i\, L(t, 0)$, $i = 1$ or 2 on $D[0, 1]$. This proves the result (7.5).

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

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

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

$$\begin{aligned}
E\Big[|\psi_n'(t) - \psi_n''(t)|^2 \mid \epsilon_0, \epsilon_{-1}, ...\Big] &\le \frac{1}{nh^2}\sup_{y, 1\le m\le n} E\Big[\sum_{k=1}^m K^2[(y + x'_{0,k})/h](u_k^2 - E u_k^2)\Big]^2 \\
&\le \frac{1}{nh^2}\sup_y \Big[\sum_{k=1}^n E r^2(x'_{0,k}/h) g^2(u_k) \\
&\qquad + 2\sum_{1\le k<l\le n} |E r(x'_{0,k}) r(x'_{0,l}) g(u_k) g_1(u_l)|\Big],
\end{aligned}$$

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

$$\begin{aligned}
E\Big[|\psi_n'(t) - \psi_n''(t)|^2 \mid \epsilon_0, \epsilon_{-1}, ...\Big] &\le C\,\frac{1}{nh}\sum_{k=m_0}^n k^{-1/2} + C(\epsilon + h\,\log n) \\
&\le C[\epsilon + h\log n + 1/(\sqrt{n}h)],
\end{aligned}$$

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 $k \geq m_0$. On the other hand, $\psi_n'(t)$ is tight on $D[0,1]$ since $\psi_n'(t)$ is positive. This proves $\psi_n'(t) \Rightarrow d_0^2\, L(t,0)$ on $D[0,1]$, that is, the result (7.6).

Proofs of Proposition 7.3. By noting

$$E\psi_n(t) - ES_n^2(t) = \frac{2d_0^{-2}}{nh^2} \sum_{1\leq k<l\leq [nt]} E\big\{u_k\, u_l\, K[(x_k-x)/h]\, K[(x_l-x)/h]\big\},$$

in a similar argument as in the proof of (7.23), we may obtain

$$\sup_{0\leq t\leq 1} |E\psi_n(t) - ES_n^2(t)| \quad = \quad o(1). \tag{7.24}$$

Recall $\psi_n(t) = \psi_n'(t)/d_0^2$. It follows from (7.23) and (7.24) that

$$E\psi_n(t) \to EL(t,0) \quad \text{and} \quad ES_n^2(t) \to EL(t,0).$$

for each fixed $0 \leq t \leq 1$. This yields that $S_n^2(t)$ and $\psi_n(t)$ are uniformly integrable by Theorem 5.4 of Billingsley (1968), since both $S_n^2(t)$ and $\psi_n(t)$ are positive and integrable random variables. The integrability of $S_n(t)$ follows from that of $S_n^2(t)$. The proof of Proposition 7.3 is now complete.

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

$$\max_{1\leq k\leq n} |u_k K[(x_k-x)/h]| = o_P[(nh^2)^{1/4}], \tag{7.25}$$

and 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

$$0 \leq t_1 \leq t_2 \leq ... \leq t_m \leq t \leq 1, \quad t - t_m \leq \delta,$$

we have

$$P\big[|S_n(t) - S_n(t_m)| \geq \epsilon \mid S_n(t_1), S_n(t_2), ..., S_n(t_m)\big] \leq \alpha_n(\epsilon,\delta), \quad a.s. \tag{7.26}$$

By noting $\max_{1\leq k\leq n} |u_k K[(x_k-x)/h]| \leq \big\{\sum_{j=1}^n u_j^4 K^4[(x_j-x)/h]\big\}^{1/4}$, the result (7.25) follows from $Eu_j^4 K^4[(x_j-x)/h] \leq C\, h/\sqrt{j}$ by Lemma 7.1, with a simple calculation. As for (7.26), it only needs to show that

$$\sup_{|t-s|\leq\delta} P\Big(|\sum_{k=[ns]+1}^{[nt]} u_k K[(x_k-x)/h]| \geq \epsilon\, d_n \mid \epsilon_{[ns]}, \epsilon_{[ns]-1}, ...; \eta_{[ns]}, ..., \eta_1\Big) \leq \alpha_n(\epsilon,\delta). \tag{7.27}$$

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

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

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

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

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.

8 Proof of Theorem 3.2

We may write

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

where Θ_{3n} is defined as in (3.6),

$$\begin{aligned}
\Theta_{4n} &= \sum_{t=1}^n (u_t^2 - Eu_{m_0}^2) K[(x_t-x)/h], \\
\Theta_{5n} &:= 2\sum_{t=1}^n \big[f(x_t)-\hat f(x)\big]\, u_t\, K[(x_t-x)/h], \\
\Theta_{6n} &:= \sum_{t=1}^n \big[f(x_t)-\hat f(x)\big]^2\, K[(x_t-x)/h].
\end{aligned}$$

As in the proof of (3.5) with minor modifications, we have $\Theta_{6n} = O_P\{\sqrt{n}h^{1+\gamma}\}$. As in the proof of (3.4), 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 (3.6), imply that

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

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

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

Now, result (3.13) 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 (3.7), we may prove

$$\Big\{(nh^2)^{-1/4}\sum_{k=1}^{[nt]}(u_k^2 - Eu_{m_0}^2)\,K[(x_k-x)/h]\,,\,(nh^2)^{-1/2}\sum_{k=1}^{n}K[(x_k-x)/h]\Big\}$$
$$\to_D\ \big\{d_0'\,N\,L^{1/2}(t,0),\ d_1\,L(1,0)\big\}, \tag{8.2}$$

on $D[0,1]^2$, where $d_0'^2 = |\phi|^{-1}\,E\big(u_{m_0}^2 - Eu_{m_0}^2\big)^2\int_{-\infty}^{\infty}K^2(s)dt$, $d_1 = |\phi|^{-1}\int_{-\infty}^{\infty}K(s)ds$, and where N is a standard normal variate independent of $L(1,0)$, as in (3.7). This, together with the fact that $\Theta_{3n}(\Theta_{4n}+\Theta_{5n}) = o_P(a_nh^\gamma)$ for any a_n diverging to infinity as slowly as required, yields

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

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

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 enlarge probability space,

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

and

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

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

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

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

$$\sum_{t=1}^{n}\{f(x_t)-f(x)\}K\left(\frac{x_t-x}{h}\right)=I_a+I_b+I_c, \tag{9.4}$$

where

$$\begin{aligned} I_a &= f'(x)\sum_{t=1}^{n}(x_t-x)\,K\left(\frac{x_t-x}{h}\right), \\ I_b &= \frac{1}{2}f''(x)\sum_{t=1}^{n}(x_t-x)^2\,K\left(\frac{x_t-x}{h}\right), \\ I_c &= \sum_{t=1}^{n}\left\{f(x_t)-f(x)-f'(x)(x_t-x)-\frac{1}{2}f''(x)(x_t-x)^2\right\}K\left(\frac{x_t-x}{h}\right). \end{aligned}$$

As in (9.1) above, we have

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

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

$$I_a+I_c=O_P\left(\left(\sqrt{n}h^3\right)^{1/2}+(\sqrt{n}h^5\log n)^{1/2}+\left(\sqrt{n}h^4\right)\right). \tag{9.6}$$

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

$$\begin{aligned} \frac{\sum_{t=1}^{n}\{f(x_t)-f(x)\}K\left(\frac{x_t-x}{h}\right)}{\sum_{t=1}^{n}K\left(\frac{x_t-x}{h}\right)} &= \frac{\frac{1}{\sqrt{n}h}(I_a+I_b+I_c)}{\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h_n}\right)} \\ = \frac{h^2}{2}f''(x)\int_{-\infty}^{\infty}y^2K(y)dy\,\{1+o_p(1)\} \\ +O_P\left(\left(\frac{h}{\sqrt{n}}\right)^{1/2}+\left(\frac{h^3\log n}{\sqrt{n}}\right)^{1/2}+h^3\right). \end{aligned} \tag{9.7}$$

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

$$\begin{aligned} &\left(nh^2\right)^{1/4}\left[\hat{f}(x_t)-f(x)-\frac{h^2}{2}f''(x)\int_{-\infty}^{\infty}y^2K(y)dy\right] \\ &= \frac{\frac{1}{(nh^2)^{1/4}}\sum_{t=1}^{n}u_tK\left(\frac{x_t-x}{h}\right)}{\frac{1}{\sqrt{n}h}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h}\right)}+O_P\left(h+h^2(\log n)^{1/2}+n^{1/4}h^{7/2}\right) \\ &\Rightarrow \frac{N\left(0,\sigma_u^2\int_{-\infty}^{\infty}K^2(s)\,ds\right)}{L(1,0)^{1/2}}, \end{aligned}$$

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, i.e.,

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

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

$$\begin{aligned}
E|I_c| &\le \sum_{t=1}^{n} \int_{-\infty}^{\infty} \Big\{ \Big| f(\sqrt{t}y) - f(x) - f'(x)(\sqrt{t}\, y - x) \\
&\qquad\qquad - \frac{1}{2} f''(x)(\sqrt{t}\, y - x)^2 \Big| \, K\left(\frac{\sqrt{t}}{h} y - \frac{x}{h}\right) \Big\} h_{t,0,n}(y)\, dy \\
&\le h\, C \sum_{t=1}^{n} \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} |\{ f(h\, y + x) - f(x) - f'(x)(hy) \\
&\qquad\qquad - \frac{1}{2} f''(x)(h\, y)^2 \Big| \, K(y) \Big\} \, dy \\
&\le C\, h^4 \sum_{t=1}^{n} \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} |s|^3 K(s)\, ds \le C\sqrt{n}\, h^4,
\end{aligned} \tag{9.9}$$

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

$$\begin{aligned}
EI_a^2 &\le C\, h^2 E\left[\sum_{t=1}^{n} H_1\left(\frac{x_t - x}{h}\right)\right]^2 \quad \text{with} \quad H_1(y) := yK(y) \\
&\le C\, h^2 \left[\sum_{t=1}^{n} EH_1^2\left(\frac{x_t - x}{h}\right) + \sum_{1 \le s < t \le n} EH_1\left(\frac{x_s - x}{h}\right) H_1\left(\frac{x_t - x}{h}\right)\right].
\end{aligned}$$

It is readily seen that

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

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

$$\begin{aligned}
&E\left\{H_1\left(\frac{x_t - x}{h}\right) \mid \mathcal{F}_s\right\} \\
&= \int H_1\left(\frac{\sqrt{t-s}y + x_s - x}{h}\right) h_{t,s,n}(y) dy \\
&= \frac{h}{\sqrt{t-s}} \int H_1\left(y + \frac{x_s - x}{h}\right) h_{t,s,n}\left(\frac{hy}{\sqrt{t-s}}\right) dy \\
&\le \frac{h}{\sqrt{t-s}} \int \left| H_1\left(y + \frac{x_s - x}{h}\right)\right| \left| h_{t,s,n}\left(\frac{hy}{\sqrt{t-s}}\right) - h_{t,s,n}(0)\right| dy \\
&\le C\left(\frac{h}{\sqrt{t-s}}\right)^2 \int \left| H_1\left(y + \frac{x_s - x}{h}\right)\right| |y| dy,
\end{aligned}$$

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

$$\begin{aligned}
& \left| EH_1\left(\frac{x_s-x}{h}\right) H_1\left(\frac{x_t-x}{h}\right)\right| \\
\leq\ & E\left\{\left|H_1\left(\frac{x_s-x}{h}\right)\right|\left|E\left[H_1\left(\frac{x_t-x}{h}\right) \mid \mathcal{F}_s\right]\right|\right\} \\
\leq\ & C\left(\frac{h}{\sqrt{t-s}}\right)^2 \int E\left|H_1\left(\frac{x_s-x}{h}\right)\right|\left|H_1\left(y+\frac{x_s-x}{h}\right)\right| |y| dy \\
\leq\ & C\left(\frac{h}{\sqrt{t-s}}\right)^2 \left(\frac{h}{\sqrt{s}}\right) \int |H_1(y)|\ |H_1(y+z)|\ |y| dz dy \\
\leq\ & C\left(\frac{h}{\sqrt{t-s}}\right)^2 \left(\frac{h}{\sqrt{s}}\right). \qquad (9.11)
\end{aligned}$$

Taking the bounds (9.10) and (9.11) into EI_a^2, we get

$$\begin{aligned}
EI_a^2 &\leq C\, h^3 \sum_{t=1}^{n} \frac{1}{\sqrt{t}} + C\, h^5 \sum_{1\leq s<t\leq n} \left(\frac{1}{\sqrt{t-s}}\right)^2 \left(\frac{1}{\sqrt{s}}\right) \\
&\leq C\, h^3 \sqrt{n} + C\, h^5 \sqrt{n} \log n, \qquad (9.12)
\end{aligned}$$

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

REFERENCES

Ai, C. and X. Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica,* **71**, 1795-1843.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

Billingsley, P. (1974). Conditional distributions and tightness. *Annals of Probability,* **2**, 480-485.

Borodin, A. N. and Ibragimov, I. A. (1995). *Limit theorems for functionals of random walks.* Proc. Steklov Inst. Math., no. 2.

Carasco, M., J.-P. Florens and E. Renault (2006). Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. in *Handbook of Econometrics* ed. by J. Heckman and E. Leamer, Vol. 6, North Holland (to appear).

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary $AR(1)$ process. *Annals of Statistics,* **15**, 1050-1063.

Elliott, G. (1998). "On the Robustness of Cointegration Methods When Regression Almost Have Unit Roots." *Econometrica*, **66**, 149-158.

Florens, J.-P. (2003). Inverse problems and structural econometrics: The example of instrumental variables. in *Advances in Economics and Econometrics: theory and Applications - Eigth World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Vol. 36 of *Econometric Society Monographs.* Cambridge University Press.

Gasquet, C. and Witomski, P. (1998). *Fourier Analysis and Applications.* Springer.

Hall, P. (1977). Martingale invariance principles. *Annals of Probability,* **5**, 875-887.

Hall, P. and J. L. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics,* **33**, 2904-2929.

Karlesn, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model, *Annals of Statistics*, **35**, 252-299.

Newey, W. K. and J. J. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica,* **71**, 1565-1578.

Newey, W.K., J.L. Powell, and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica,* **67**, 565-603.

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

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

Phillips, P. C. B. (1988), "Regression Theory for Near-Integrated Time Series," *Econometrica*, **56**, 1021-1044.

Phillips, P. C. B. and S. N. Durlauf (1986). Multiple Time Series Regression with Integrated Processes. *Review of Economic Studies*, **53**, 473-495.

Stock, J. H. (1987) Asymptotic Properties of Least Squares Estimators of Cointegration Vectors. *Econometrica,* **55**, 1035-1056.

Wang, Q. and P. C. B. Phillips (2008). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory,* forthcoming. Cowles Fundation Discussion Paper, No. 1594, Yale University.

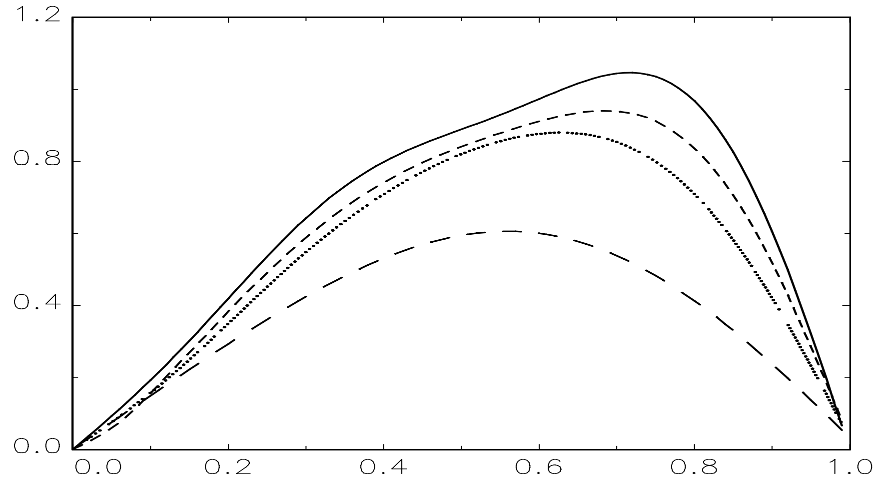

Fig. 1 Graphs over the interval $[0,1]$ of $f_A(x)$ and Monte Carlo estimates of $E\left(\hat{f}_A(x)\right)$ for $h = n^{-1/2}$ (short dashes), $h = n^{-1/3}$ (dotted) and $h = n^{-1/5}$ (long dashes) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

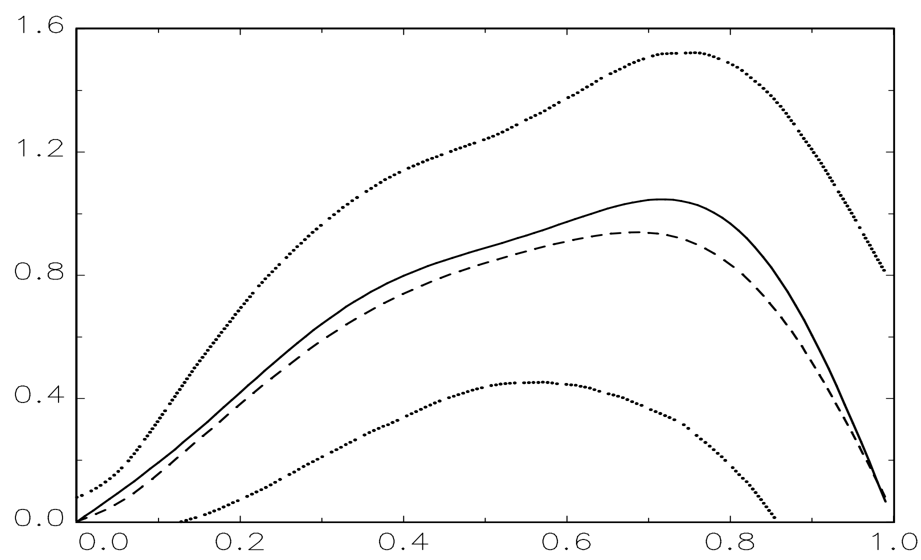

Fig 2. Graphs over the interval $[0,1]$ of estimation bands for $f_A(x)$ (solid line), the Monte Carlo estimate of $E\left(\hat{f}_A(x)\right)$ for $h = n^{-1/2}$ (short dashes) and 95% estimation bands (dotted) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

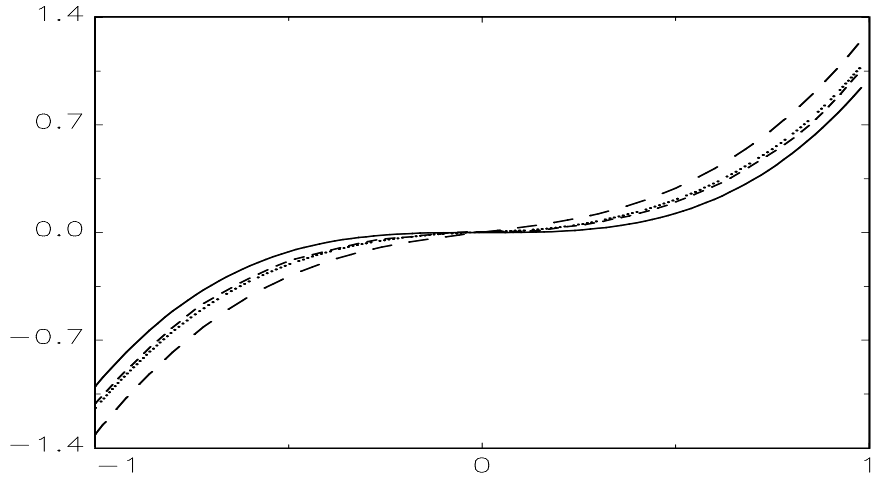

Fig. 3: Graphs of $f_B(x)$ and Monte Carlo estimates of $E\left(\hat{f}_B(x)\right)$ for $h = n^{-1/2}$ (short dashes), $h = n^{-1/3}$ (dotted) and $h = n^{-1/5}$ (long dashes) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

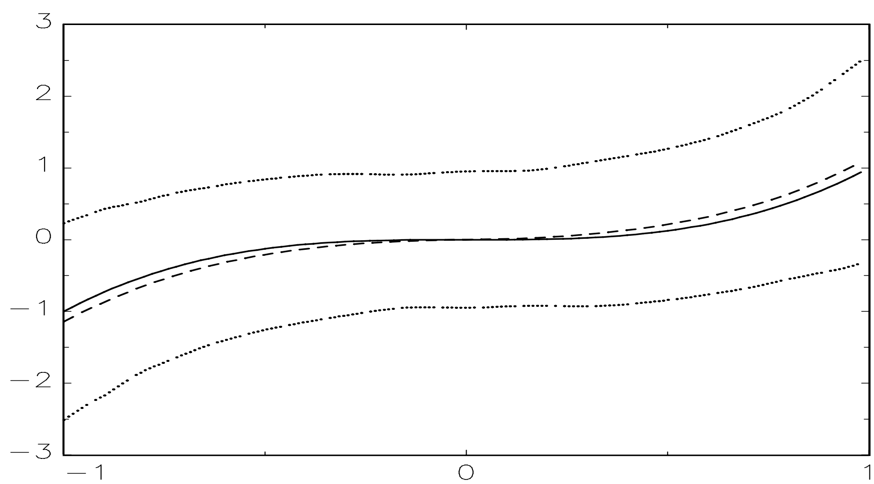

Fig. 4: Graphs of estimation bands for $f_B(x)$ (solid line), the Monte Carlo estimate of $E\left(\hat{f}_B(x)\right)$ for $h = n^{-1/3}$ (short dashes) and 95% estimation bands (dotted) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

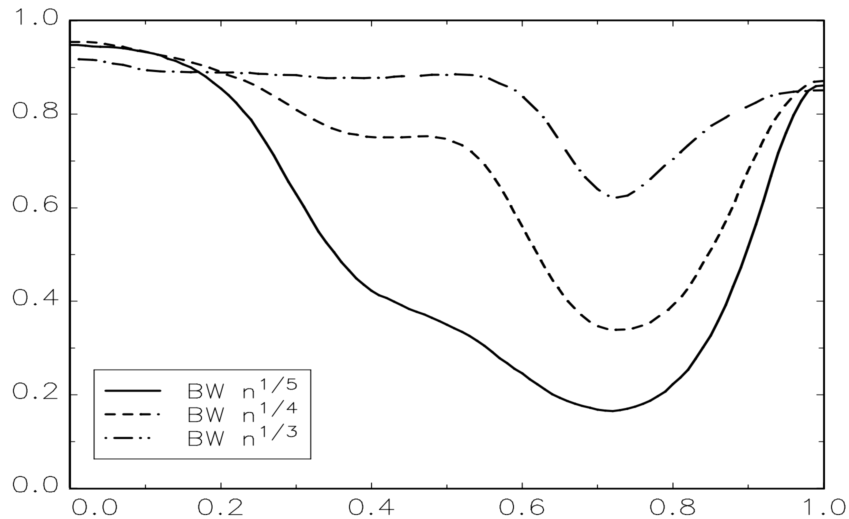

Fig. 5: Coverage probabilities of (nominal 95%) confidence intervals for $f_A(x)$ over $[0, 1]$ for different bandwidths.

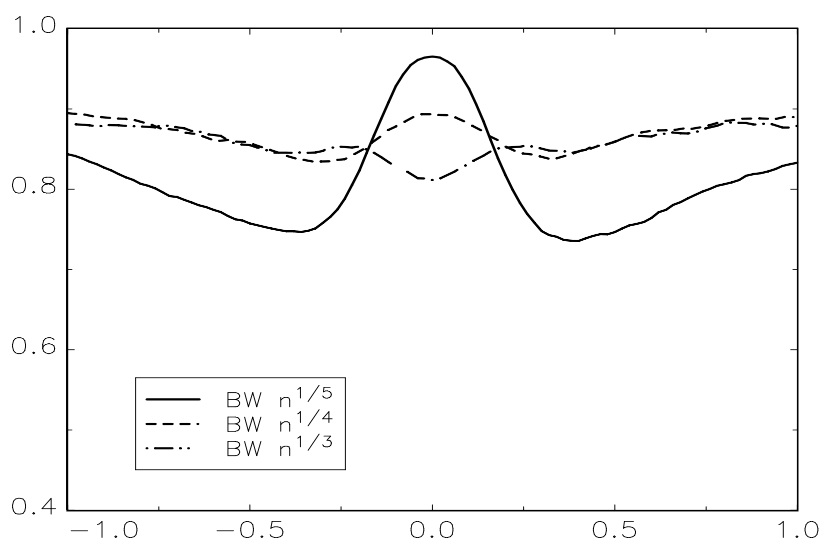

Fig. 6: Coverage probabilities of (nominal 95%) confidence intervals for $f_B(x)$ over $[0, 1]$ for different bandwidths.