STRUCTURAL NONPARAMETRIC COINTEGRATING REGRESSION

By

Qiying Wang and Peter C. B. Phillips

May 2008

COWLES FOUNDATION DISCUSSION PAPER NO. 1657

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

Structural Nonparametric Cointegrating Regression *

Qiying Wang
School of Mathematics and Statistics
The University of Sydney

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

February 13, 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 time series, it is shown here that inverse and ill-posed inverse problems do not arise. Remarkably, nonparametric kernel estimation of a structural nonparametric cointegrating regression is consistent and the limit distribution theory is mixed normal, giving simple useable asymptotics in practical work. The results provide a convenient basis for inference in structural nonparametric regression with nonstationary time series. The methods may be applied to a wide 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, Nonlinear functional, Nonparametric regression, Structural estimation, Unit root.

JEL Classification: C14, C22.

---

*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. In fact, the random wandering nature of stochastically nonstationary time series ensures that the regressor inevitably departs from any particular locality and thereby assists in tracing out (and identifying) the structural function. 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, the present 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.

The paper is organized as follows. Section 2 introduces the model and assumptions. Section 3 provides the main result on the consistency and limit distribution of the kernel estimator in a structural model of nonlinear cointegration. 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 and 7 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 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 $K_h(s) = \frac{1}{h} K(s/h)$, $K(x)$ 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, KMT) investigated $\hat{f}(x)$ in the situation where x_t is a recurrent Markov chain; and Wang and Phillips (2006, 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 even though the rates of convergence are different and typically slower when x_t is nonstationary. However, the strong conditions under which the asymptotic theory of KMT and WP is developed limits its usefulness in applications. It seems particularly important to relax conditions of independence, so that the system is a structural model that allows joint dependence between the regressor and dependent variable in the regression. The goal of the present paper is to remove this assumption of independence and to develop a limit theory for structural functional estimation in the context of nonstationary time series.

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 for which ϵ_1 has a density $d(x)$. The sequence $\{\epsilon_t\}_{t\geq 1}$ is assumed to be independent of another *iid* random sequence $\{\lambda_t\}_{t\geq 1}$. We use the following assumptions in the asymptotic development.

Assumption 1. $x_t = \sum_{j=1}^{t} \eta_j$ where $\eta_j = \sum_{k=0}^{\infty} \phi_k \epsilon_{j-k}$ with $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$ and $\sum_{k=0}^{\infty} k^2 |\phi_k| < \infty$.

Assumption 2. $u_t = u(\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-m_0}, \lambda_t)$ satisfies $Eu_t = 0$ and $Eu_t^4 < \infty$ for $t \geq m_0$, where $u(x_0, x_1, ..., x_{m_0}, y)$ is a real measurable function on R^{m_0+2}. We define $u_t = 0$ for $1 \leq t \leq m_0 - 1$.

Assumption 3. $K(x)$ is a nonnegative bounded three times continuous differentiable function satisfying $\int K(x)dx < \infty$ and $\int |K^{(i)}(x)|dx < \infty$ for $i = 1, 2, 3$.

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

Assumption 1 is standard in a cointegrating regression framework, so that x_t is a partial sum of linear process innovations that satisfy a simple summability condition with long run moving average coefficient $\phi \neq 0$. 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_t, 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.

Assumption 3 places stronger conditions on the kernel function than is usual in kernel estimation, requiring integrable derivatives to the third order. These conditions are needed for technical reasons in the proofs and they are clearly satisfied for many commonly used kernels. 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 $f(x) = |x|^\alpha$ and $f(x) = 1/(1+|x|^\alpha)$ for some $\alpha > 0$ satisfy Assumption 4.

3 Main result and outline of the proof

The limit theory for the conventional kernel regression estimate $\hat{f}(x)$ 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 $nh^6 \to 0$,*

$$\Big(h \sum_{t=1}^{n} K_h(x_t - x)\Big)^{1/2} (\hat{f}(x) - f(x)) \ \to_D \ N(0, \sigma^2), \tag{3.1}$$

where $\sigma^2 = E(u_{m0}^2) \int_{-\infty}^{\infty} K^2(s)ds / \int_{-\infty}^{\infty} K(x)dx$.

Remarks

(a) The proof of (3.1) is given in the Appendix. To outline the essentials of the argument here we split the error of estimation $\hat{f}(x) - f(x)$ as

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

It is readily seen that

$$\Big(h \sum_{t=1}^{n} K_h(x_t - x)\Big)^{1/2} (\hat{f}(x) - f(x)) = \sum_{t=1}^{n} u_t \, Z_{nt} + \Theta_{1n}/\Theta_{2n}, \tag{3.2}$$

where $Z_{nt} = K\big(\frac{x_t - x}{h}\big)/\Theta_{2n}$ with $\Theta_{2n}^2 = \sum_{t=1}^n K\big(\frac{x_t-x}{h}\big)$ and

$$\Theta_{1n} = \sum_{t=1}^{n} \left[f(x_t) - f(x)\right] K\Big(\frac{x_t - x}{h}\Big).$$

It has been proved in WP that $\Theta_{1n}/\Theta_{2n} \to_P 0$, which requires that the "signal" $\Theta_{2n}^2 \to \infty$, in Probab., which in turn requires that $nh^2 \to \infty$. The stated result will then follow if we prove

$$\Big\{(nh^2)^{-1/4}\sum_{k=1}^{[nt]} u_k\, K[(x_k-x)/h]\,,\,(nh^2)^{-1/2}\sum_{k=1}^{n} K[(x_k-x)/h]\,\Big\}$$
$$\to_D \,\big\{c_0\, N\, L^{1/2}(t,0),\, d_0\, L(1,0)\big\}, \tag{3.3}$$

on $D[0,1]^2$, where $c_0^2 = \phi\, E(u_{m0}^2)\int_{-\infty}^{\infty} K^2(s)dt$, $d_0 = \phi\int_{-\infty}^{\infty} K(s)ds$, $L(t,0)$ is the local time process at the origin of a Brownian motion $\{W(t)\}_{t\geq 0}$, and N is a standard normal variate independent of $L(t,0)$. The local time process $L(t,a)$ is defined by

$$L(t,a) = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t I\{|W(r)-a|\leq \epsilon\}dr. \tag{3.4}$$

Indeed, since $P(L(1,0) > 0) = 1$, the required result (3.1) follows by (3.3) and the continuous mapping theorem. It remains to prove (3.3), which is done in the Appendix. In fact, it is clearly sufficient for the required result to show that the finite dimensional distributions converge in (3.3).

(b) Result (3.1) shows that $\hat f(x)$ is consistent and has an asymptotic distribution that is mixed normal 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 (3.3). In finite samples, the performance of the functional estimation procedure will depend on how much time the process x_t spends around the point x and how well the bandwidth concentrates attention on this point. As remarked earlier, consistency depends on $h \to 0$, so that function estimation is localized at a single point x as $n \to \infty$. The conditions $nh^2 \to \infty$ and $nh^6 \to 0$ in the theorem require that h tend to zero faster than $n^{-1/6}$ but not as fast as $n^{-1/2}$.

(c) The bandwidth choice h turns out to be particularly important in structural functional estimation when there is contemporaneous correlation between x_t and u_t. For when h is fixed as $n \to \infty$ the estimate $\hat f(x)$ can be shown to be asymptotically biased and when h tends to zero slowly this bias is manifest even in very large samples. Some illustrative simulations are reported in the next section.

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 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. 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 some 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.

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

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 $(nh^2)^{1/4} = n^{1/12}$, which is far slower than the rate $(nh)^{1/2} = n^{2/5}$ for conventional kernel regression.

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 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 new possibilities for functional regression in empirical research with 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 and partial linear cointegration models may be treated in a similar way to (2.1). There are also issues of specification testing, functional form tests, and cointegration tests, which may now be addressed using the methods of the paper. We plan to report on some of these extensions in later work.

6 Proof of Theorem 3.1

As shown in Remark (a), the proof of the theorem essentially amounts to proving (3.3). 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 (3.3), we use the following lemma.

LEMMA 6.1. *Suppose that $\{\mathcal{F}_t\}_{t\ge0}$ 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 $(q(t), \xi)_{t\ge0}$ 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, page 14) and the proof follows from 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 $s \in R$,

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

which yields the stated result. $\Box$

By virtue of Lemma 6.1, we now obtain the proof of (3.3). Technical details of some subsidiary results that are used in this proof are given in the next section. Set

$$
\begin{aligned}
\xi_n &= \frac{1}{d_0\sqrt{nh^2}}\sum_{k=1}^n K[(x_k-x)/h], \quad \zeta_n(t) = \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_k, \quad \zeta_n'(t) = \frac{1}{\sqrt{n}\phi}\sum_{k=1}^{[nt]}\eta_k, \\
S_n(t) &= \frac{1}{c_0(nh^2)^{1/4}}\sum_{k=1}^{[nt]} u_k K[(x_k-x)/h], \quad \psi_n(t) = \frac{1}{d_1\sqrt{nh^2}}\sum_{k=1}^{[nt]} u_k^2 K^2[(x_k-x)/h],
\end{aligned}
$$

for $0 \le t \le 1$, where c_0 and d_0 are defined as in (3.3), and $d_1 = \phi\, Eu_{m_0}^2 \int_{-\infty}^{\infty} K^2(s)dt$.

We will prove in Propositions 7.1 and 7.2 that $\zeta_n'(t) \Rightarrow W'(t)$, $\xi_n \to_D \psi(1)$ and $\psi_n(t) \Rightarrow \psi(t)$ on $D[0,1]$, where $\psi(t) := L(t,0)$ and $L(t,s)$ is a local time process of the Wiener process $\{W'(t), 0 \le t \le 1\}$ defined by (3.4). Furthermore we will prove in Proposition 7.4 that $\{S_n(t)\}_{n\ge1}$ is tight on $D[0,1]$. These facts imply that $\{S_n(t), \psi_n(t), \zeta_n'(t), \xi_n\}_{n\ge1}$ is tight on $D[0,1]^4$. Hence, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

$$
\big\{S_{n''}(t), \psi_{n''}(t), \zeta_{n''}'(t), \xi_{n''}\big\} \to_d \big\{\eta(t), \psi(t), W'(t), \psi(1)\big\},
$$

on $D[0,1]^4$, where $\eta(t)$ is a process continuous with probability one by noting (7.19) below. By virtue of (7.1), we also have

$$
\big\{S_{n''}(t), \psi_{n''}(t), \zeta_{n''}(t), \xi_{n''}\big\} \to_d \big\{\eta(t), \psi(t), W'(t), \psi(1)\big\}, \tag{6.1}
$$

on $D[0,1]^4$. 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 $0 \le s < t \le 1$,

$$
\begin{aligned}
E\Big(\big[\eta(t)-\eta(s)\big] \mid \mathcal{F}_s\Big) &= 0, \quad a.s., & (6.2)\\
E\Big(\big\{[\eta(t)-\eta(s)]^2 - [\psi(t)-\psi(s)]\big\} \mid \mathcal{F}_s\Big) &= 0, \quad a.s., & (6.3)
\end{aligned}
$$

then it follows from Lemma 6.1 that the finite-dimensional distributions of $(\eta(t), \xi)$ 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 (3.3) 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 the $\{n\}$ 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 the similar arguments as 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

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

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

where $\chi(s) = \chi(x_1, ..., x_s, u_1, ..., u_s)$, a functional of $x_1, ..., x_s, u_1, ..., u_s$.

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

$$\begin{aligned} x_t &= \sum_{j=1}^{t} \sum_{i=-\infty}^{j} \epsilon_i \phi_{j-i} \\ &= \sum_{j=1}^{s} \sum_{i=-\infty}^{j} \epsilon_i \phi_{j-i} + \sum_{j=s+1}^{t} \sum_{i=-\infty}^{j} \epsilon_i \phi_{j-i} \\ &= x_s + \sum_{j=s+1}^{t} \sum_{i=-\infty}^{s} \epsilon_i \phi_{j-i} + \sum_{j=s+1}^{t} \sum_{i=s+1}^{j} \epsilon_i \phi_{j-i} \\ &:= x^*_{s,t} + x'_t, \end{aligned} \tag{6.10}$$

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

$$x'_t = \sum_{j=1}^{t-s} \sum_{i=1}^{j} \epsilon_{i+s} \phi_{j-i} = \sum_{i=1}^{t-s} \epsilon_{i+s} \sum_{j=i}^{t-s} \phi_{j-i} =_d \sum_{i=1}^{t-s} \epsilon_i \sum_{j=0}^{t-s-i} \phi_j,$$

where $=_d$ denotes the same in distribution.

Now, by independence of ϵ_k again and conditional arguments, it suffices to show that, for any $0 \le s < t \le 1$ and any μ,

$$\begin{aligned} &\sup_{y, 1 \le m \le n} E\Big\{ \sum_{k=1}^{m} u_k \, K[(y + x''_k)/h] \, e^{i\mu \sum_{i=1}^m \epsilon_i/\sqrt{n}} \Big\} \\ &= o[(nh^2)^{1/4}], \end{aligned} \tag{6.11}$$

$$\begin{aligned} &\sup_{y, 1 \le m \le n} E\Big(\big\{ \sum_{k=1}^{m} u_k \, K[(y + x''_k)/h] \big\}^2 - \sum_{k=1}^{m} u_k^2 \, K^2[(y + x''_k)/h] \Big) \, e^{i\mu \sum_{i=1}^m \epsilon_i/\sqrt{n}} \\ &= o[(nh^2)^{1/2}], \end{aligned} \tag{6.12}$$

where $x''_k = \sum_{i=1}^{k} \epsilon_i \sum_{j=0}^{k-i} \phi_j$. This follows from Proposition 7.5.

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. *Under an appropriate probability space $\{\Omega, \mathcal{F}, P\}$, there exist a Winner process $W(t)$ such that $\sup_t |\zeta_n(t) - W(t)| = o_P(1)$ and*

$$\sup_{0 \le t \le 1} |\zeta_n(t) - \zeta'_n(t)| = o_P(1) \tag{7.1}$$

(which implies that $\zeta_n(t) \Rightarrow W(t)$ and $\zeta'_n(t) \Rightarrow W(t)$ on $D[0,1]$).

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

$$\frac{1}{\sqrt{nh^2}} \sum_{k=1}^{[nt]} K[(x_k - x)/h] \;\Rightarrow\; d_0\, \psi(t), \tag{7.2}$$

$$\frac{1}{\sqrt{nh^2}} \sum_{k=1}^{[nt]} K^2[(x_k - x)/h]\, u_k^2 \;\Rightarrow\; d_1\, \psi(t), \tag{7.3}$$

on $D[0,1]$, where $d_0 = \phi \int_{-\infty}^{\infty} K(s)dt$ and $d_1 = \phi\, E u_{m_0}^2 \int_{-\infty}^{\infty} K^2(s)dt$.

PROPOSITION 7.3. *For any fixed $0 \le t \le 1$, we have that $S_n(t)$, $S_n^2(t)$ and $\psi_n(t)$, $n \ge 1$, are uniformly integrable.*

PROPOSITION 7.4. *We have that $\{S_n(t)\}_{n\ge 1}$ is tight on $D[0,1]$.*

PROPOSITION 7.5. *We have that, for any $u \in R$,*

$$\sup_{y,1\le m\le n} E\Big\{ \sum_{k=1}^{m} u_k\, K[(y + x_k'')/h]\, e^{i\mu \sum_{i=1}^m \epsilon_i/\sqrt{n}} \Big\}$$
$$= o[(nh^2)^{1/4}], \tag{7.4}$$

$$\sup_{y,1\le m\le n} E\Big(\big\{ \sum_{k=1}^{m} u_k\, K[(y + x_k'')/h] \big\}^2 - \sum_{k=1}^{m} u_k^2\, K^2[(y + x_k'')/h] \Big)\, e^{i\mu \sum_{i=1}^m \epsilon_i/\sqrt{n}}$$
$$= o[(nh^2)^{1/2}]. \tag{7.5}$$

Proposition 7.1 is well-known. In order to prove Proposition 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 $I_{k,l}$ and II_k defined by

$$I_{k,l} = E\Big[r(x_k''/h)\, r_1(x_l''/h)\, g(u_k)\, g_1(u_l)\, \exp\big\{ i\mu \sum_{j=1}^{l} \epsilon_j/\sqrt{n} \big\} \Big],$$

$$II_k = E\Big[r(x_k''/h)\, g(u_k)\, \exp\big\{ i\mu \sum_{j=1}^{k} \epsilon_j/\sqrt{n} \big\} \Big], \tag{7.6}$$

under different settings of $g(x)$ and $g_1(x)$. We have the following lemmas, which will play a core rule in the proof of main results. We always assume $l < k$ 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

$$|II_k| \le C\, h/\sqrt{k}. \tag{7.7}$$

(b) If $Eg(u_k)=0$ and $Eg^2(u_k)<\infty$, then

$$|II_k| \;\leq\; C\,(k^{-2}+h/k). \tag{7.8}$$

LEMMA 7.2. *Suppose that $\int(1+|\lambda|)|\hat r(\lambda)|d\lambda<\infty$ and $\int(1+|\lambda|)|\hat r_1(\lambda)|d\lambda<\infty$, 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 a $n_0>0$ such that, for all $n\geq n_0$ and all $l-k\geq 1$,*

$$|I_{k,l}| \;\leq\; C\left[\epsilon\,(l-k)^{-3/2}+h\,(l-k)^{-1}\right]\Big(k\sum_{j=k/2}^{\infty}|\phi_j|+k^{-2}+h/\sqrt{k}\Big), \tag{7.9}$$

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

We only prove Lemma 7.2. The proof of Lemma 7.1 is the same and hence the details are omitted.

The proof of Lemma 7.2. We have $r(x)=\frac{1}{2\pi}\int e^{-ixt}\hat r(t)dt$ and $r_1(x)=\frac{1}{2\pi}\int e^{-ixt}\hat r_1(t)dt$ as $\int(|r(t)|+|r_1(t)|)dt<\infty$. 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 $l<k$. Since

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

it follows from independence of the ϵ_k's that

$$\begin{aligned} |I_{k,l}| \;\leq\; &\int\Big|E\{e^{iz^{(2)}/h}\}\Big|\Big|E\{e^{iz^{(3)}/h}\,g_1(u_l)\}\Big|\,|\hat r_1(\lambda)| \\ &\qquad\Big(\int\big|E\{e^{iz^{(1)}/h}\,g(u_k)\}\big|\,|\hat r(t)|\,dt\Big)\,d\lambda, \end{aligned} \tag{7.10}$$

where

$$\begin{aligned} z^{(1)} &= \sum_{q=1}^{k}\epsilon_q\big[\lambda\sum_{j=0}^{l-q}\phi_j-t\sum_{j=0}^{k-q}\phi_j+u\,h/\sqrt{n}\big], \\ z^{(2)} &= \sum_{q=k+1}^{l-m_0}\epsilon_q\,\big(\lambda\sum_{j=0}^{l-q}\phi_j+u\,h/\sqrt{n}\big), \\ z^{(3)} &= \sum_{q=l-m_0+1}^{l}\epsilon_q\,\big(\lambda\sum_{j=0}^{l-q}\phi_j+u\,h/\sqrt{n}\big). \end{aligned}$$

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,

$$\begin{aligned}\Lambda(\lambda, k) \;&:=\; \int \big|\, E\{e^{iz^{(1)}/h}\, g(u_k)\,\}\big|\, |\hat{r}(t)|\, dt \\ &\le\; C\,\big(\,|\lambda|\, h^{-1} k \sum_{j=k/2}^{\infty} |\phi_j| + k^{-2} + h/\sqrt{k}\,\big). \end{aligned} \tag{7.11}$$

To estimate $\Lambda(\lambda, k)$, take δ sufficiently large such that $|Ee^{is\epsilon_1}| \le e^{-1/2}$ whenever $|s| \ge \delta|\phi|/2$, where $\phi = \sum_{j=0}^{\infty}\phi_j$. This may be done by using the fact $|Ee^{it\epsilon}| \to 0$, as $t \to \infty$, since $E\epsilon_1 = 0$, $E\epsilon_1^2 = 1$ and ϵ_1 has a density. Furthermore, take k_0 ($k_0 \ge 2m_0$) sufficiently large such that $\sum_{j=k_0/2+1}^{\infty}|\phi_j| \le |\phi|/2$. We claim that, for all $k \ge k_0/2$,

$$\Big|Ee^{i\,\epsilon_1\, t\, \sum_{j=0}^{k}\phi_j}\Big| \;\le\; \begin{cases} e^{-1/2} & \text{if } |t| \ge \delta, \\ e^{-\gamma t^2} & \text{if } |t| \le \delta, \end{cases} \tag{7.12}$$

where $\gamma > 0$ is a constant not depending on k. Indeed, the result (7.12) for $|t| \ge \delta$ follows from the fact that $\big|t\,\sum_{j=0}^{k}\phi_j\big| \ge \delta|\phi|/2$ whenever $k \ge k_0/2$. If $|t| \le \delta$, then $\big|t\,\sum_{j=0}^{k}\phi_j\big| \le t_0 := \delta\sum_{j=0}^{\infty}|\phi_j|$. Since $|Ee^{it_0\epsilon_1}| \le e^{-1/2}$, it follows from Theorem 3 of Petrov (1995) that

$$\Big|Ee^{i\,\epsilon_1\, t\, \sum_{j=0}^{k}\phi_j}\Big| \;\le\; 1 - \frac{1-e^{-1/2}}{8t_0^2}\, t^2\,\big(\sum_{j=0}^{k}\phi_j\big)^2 \le e^{-\gamma t^2},$$

with $\gamma = (1 - e^{-1/2})\phi^2/(32t_0^2) > 0$. This gives (7.12).

The result (7.12) will be used to estimate $\Lambda(\lambda, k)$. To this end, put $\tau_{\lambda,t}^{(q)} = \lambda\sum_{j=0}^{l-q}\phi_j - t\sum_{j=0}^{k-q}\phi_j$, $W^{(1)} = \sum_{q=1}^{k/2}\epsilon_q\,\tau_{\lambda,t}^{(q)}$ and $W^{(2)} = (\lambda - t)\sum_{q=1}^{k/2}\epsilon_q\,\sum_{j=0}^{k-q}\phi_j$. Note that $\tau_{\lambda,t}^{(q)} = (\lambda - t)\sum_{j=0}^{k-q}\phi_j + \lambda\sum_{j=k-q+1}^{l-q}\phi_j$. We have

$$E|W^{(1)} - W^{(2)}| \le |\lambda| \sum_{q=1}^{k/2} E|\epsilon_q| \sum_{j=k-q+1}^{l-q} |\phi_j| \le C\,|\lambda|\, k \sum_{j=k/2}^{\infty} |\phi_j|.$$

This together with (7.12) yields that, for all $k \ge k_0$,

$$\begin{aligned}\big|Ee^{i\,W^{(1)}/h}\big| \;&\le\; E|W^{(1)} - W^{(2)}|/h + \big|Ee^{i\,W^{(2)}/h}\big| \\ &\le\; C\,|\lambda|\, h^{-1}\, k \sum_{j=k/2}^{\infty} |\phi_j| + \begin{cases} e^{-k/4} & \text{if } |t-\lambda| \ge \delta\, h, \\ e^{-\gamma k (t-\lambda)^2/2h^2} & \text{if } |t-\lambda| \le \delta\, h. \end{cases}\end{aligned}$$

Hence, by noting $Z^{(1)} = W^{(1)} + \sum_{q=k/2+1}^{k} \epsilon_q \tau_{\lambda,t}^{(q)}$ and $k/2 \le k - m_0$ (which implies that $W^{(1)}$ is independent of u_k), it follows from the independence of ϵ_k again that

$$\begin{aligned}\Lambda(\lambda,k) &\le \int \big| E\{e^{iW^{(1)}/h}\}\big| \, E\big|g(u_k)\big| \, |\hat{r}(t)| \, dt \\ &\le C\,|\lambda|\,h^{-1}\,k \sum_{j=k/2}^{\infty} |\phi_j| + e^{-k/4} \int_{|t-\lambda|\ge\delta h} |\hat{r}(t)|\,dt + \int_{|t-\lambda|\le\delta h} e^{-\gamma k(t-\lambda)^2/2h^2}\,dt \\ &\le C\big(|\lambda|\,h^{-1}\,k \sum_{j=k/2}^{\infty} |\phi_j| + k^{-2} + h/\sqrt{k}\big).\end{aligned}$$

This proves (7.11) for $k \ge k_0$.

We now turn back to the proof of (7.9). We will estimate $I_{k,l}$ in three separate settings:

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

Case I. $l - k \ge 2k_0$ and $k \ge k_0$. In this case, we note that $|I_{k,l}| \le I_{k,l}^{(1)} + I_{k,l}^{(2)}$, where, $\Lambda(\lambda,k)$ is defined as in (7.11), δ is defined as in (7.12),

$$\begin{aligned}I_{k,l}^{(1)} &= \int_{|\lambda|\le\delta h} \Big| E\big\{e^{iz^{(2)}/h}\big\}\Big| \, \Big| E\big\{e^{iz^{(3)}/h} g_1(u_l)\big\}\Big| \, \Lambda(\lambda,k)\,|\hat{r}_1(\lambda)|\,d\lambda, \\ I_{k,l}^{(2)} &= \int_{|\lambda|>\delta h} \Big| E\big\{e^{iz^{(2)}/h}\big\}\Big| \, \Big| E\big\{e^{iz^{(3)}/h} g_1(u_l)\big\}\Big| \, \Lambda(\lambda,k)\,|\hat{r}_1(\lambda)|\,d\lambda.\end{aligned}$$

First estimate $I_{k,l}^{(1)}$. Since $Eg_1(u_l) = 0$, we have

$$\begin{aligned}\Big| E\Big\{e^{iz^{(3)}/h} g_1(u_l)\Big\}\Big| &= \Big| E\Big\{\big(e^{iz^{(3)}/h} - 1\big)\, g_1(u_l)\Big\}\Big| \\ &\le h^{-1} E\big[|z^{(3)}|\,|g_1(u_l)|\big] \le m_0 (E\epsilon_1^2)^{1/2} (Eg_1^2(u_l))^{1/2}\,|\lambda|\,h^{-1}.\end{aligned}$$

On the other hand, by noting $l - m_0 \ge (l+k)/2$ and $l - q \ge k_0$ for all $k \le q \le (l+k)/2$ since $l - k \ge 2k_0$ and $k_0 \ge 2m_0$, it follows from (7.12) that

$$\big|E\{e^{iz^{(2)}/h}\big| \le \Pi_{q=k}^{(l+k)/2} \big|Ee^{i\epsilon_q \lambda \sum_{j=0}^{l-q} \phi_j / h}\big| \le e^{-\gamma(l-k)\lambda^2/2h^2}.$$

These estimates, together with (7.11), yield that, for $|\lambda| \le \delta h$,

$$\begin{aligned}I_{k,l}^{(1)} &\le C\,h^{-1} \int_{|\lambda|\le\delta h} |\lambda|\, e^{-\gamma(l-k)\lambda^2/h^2}\,\Lambda(\lambda,k)\,d\lambda \\ &\le C\,h\,(l-k)^{-3/2}\,k \sum_{j=k/2}^{\infty} |\phi_j| + C\,h\,(l-k)^{-1}\,(k^{-2} + h/\sqrt{k}).\end{aligned}$$

By using similar arguments, we obtain that $|\,E\{e^{iz^{(3)}/h} g_1(u_l)\}| \le E|g_1(u_l)|$ and $\big|E\{e^{iz^{(2)}/h}\}\big| \le e^{-(l-k)/4}$ when $|\lambda| \ge \delta h$. On the other hand, we also have

$$|\,E\{e^{iz^{(3)}/h} g_1(u_l)\}| \to 0, \quad \text{whenever } \lambda/h \to \infty, \tag{7.13}$$

uniformly for all $l \ge m_0$. Indeed, supposing $\phi_0 \neq 0$ (if $\phi_0 = 0$, we may use ψ_1 and so on), we have $E\{e^{iz^{(3)}/h}\, g_1(u_l)\} = E\big\{e^{i\epsilon_l \phi_0 \lambda/h} g^*(\epsilon_l)\big\}$, where $g^*(\epsilon_l) = E\big[e^{i(z^{(3)} - \epsilon_l\phi_0\lambda)/h} g_1(u_l) \mid \epsilon_l\big]$. By recalling that ϵ_l has a density $d(x)$, it is readily seen that

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

uniformly for all l. The result (7.13) follows from the Riemann-Lebesgue theorem. By virtue of (7.13), for any $\epsilon > 0$, there exists a n_0 (A_0 respectively) such that, for all $n \ge n_0$ ($|\lambda|/h \ge A_0$ respectively), $|\, E\{e^{iz^{(3)}/h}\, g_1(u_l)\}| \le \epsilon$. Hence,

$$\begin{aligned} I_{k,l}^{(2)} &\le e^{-\,(l-k)/4} \Big(\int_{|\lambda| > A_0 h} + \int_{\delta h \le |\lambda| \le A_0 h}\Big) |\, E\{e^{iz^{(3)}/h}\, g_1(u_l)\}|\, \Lambda(\lambda, k)\, |\hat r_1(\lambda)|\, d\lambda \\ &\le C\,(\epsilon + h)\, e^{-(l-k)/4}\, \big[k \sum_{j=k/2}^{\infty} |\phi_j| + k^{-2} + h/\sqrt{k}\big], \end{aligned}$$

where we have used the fact $\int (1 + |\lambda|) |\hat r_1(\lambda)|\, d\lambda < \infty$. Combining the estimates for $I_{k,l}^{(1)}$ and $I_{k,l}^{(2)}$, simple calculations provide the result (7.9) in case I.

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

$$|I_{k,l}| \;\le\; C\,(\epsilon + h)\,\big(h^{-1} k \sum_{j=k/2}^{\infty} |\phi_j| + k^{-2} + h/\sqrt{k}\big). \tag{7.14}$$

In fact, as in (7.10), we have

$$|I_{k,l}| \;\le\; \int\!\!\int |\, E\big\{e^{iz^{(4)}/h}\big\}| \Big|\, E\big\{e^{iz^{(5)}/h}\, g(u_k)\, g_1(u_l)\big\}\Big|\, |\hat r(t)|\, |\hat r_1(\lambda)|\, dt\, d\lambda, \tag{7.15}$$

where

$$\begin{aligned} z^{(4)} &= \sum_{q=1}^{k-m_0} \epsilon_q \big[\lambda \sum_{j=0}^{l-q} \phi_j - t \sum_{j=0}^{k-q} \phi_j + u\, h/\sqrt{n}\big], \\ z^{(5)} &= \sum_{q=k-m_0+1}^{l} \epsilon_q\, \big(\lambda \sum_{j=0}^{l-q} \phi_j + u\, h/\sqrt{n}\big) - \sum_{q=k-m_0+1}^{k} \epsilon_q\, t \sum_{j=0}^{k-q} \phi_j. \end{aligned}$$

Similar arguments as in the proof of (7.11) give that, for all λ and all $k \ge k_0$,

$$\begin{aligned} \Lambda_1(\lambda, k) &:= \int |\, E\big\{e^{iz^{(4)}/h}\big\}|\, |\hat r(t)|\, dt \\ &\le C\big(|\lambda|\, h^{-1} k \sum_{j=k/2}^{\infty} |\phi_j| + k^{-2} + h/\sqrt{k}\big). \end{aligned}$$

Note that

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

For any $\epsilon > 0$, similar to the proof of (7.13), 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

$$\begin{aligned} |I_{k,l}| &\leq \int \Big(\int_{|\lambda|\leq A_0 h} + \int_{|\lambda|>A_0 h}\Big)\big|\, E\big\{e^{iz^{(4)}/h}\big\}\big|\,\Big|\, E\big\{e^{iz^{(5)}/h}\, g(u_k)\, g_1(u_l)\big\}\Big|\, |\hat{r}(t)|\, |\hat{r}_1(\lambda)|\, dt\, d\lambda \\ &\leq C\int_{|\lambda|\leq A_0 h} \Lambda_1(\lambda,k)\, d\lambda + C\,\epsilon \int_{|\lambda|>A_0 h} \Lambda_1(\lambda,k)\, |\hat{r}_1(\lambda)|\, d\lambda \\ &\leq C\,(\epsilon+h)\,\big(h^{-1}k\sum_{j=k/2}^{\infty}|\phi_j| + k^{-2} + h/\sqrt{k}\big). \end{aligned}$$

This proves (7.14) and hence the result (7.9) in case II.

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

$$|I_{k,l}| \;\leq\; C\,\big[\epsilon\,(l-k)^{-3/2} + h\,(l-k)^{-1}\big]. \tag{7.16}$$

In order to prove (7.16), split $l > k$ into $l-k \geq 2k_0$ and $l-k \leq 2k_0$. The result (7.9) then follows from the same arguments as in proofs of cases I and II but replacing the estimate of $\Lambda(\lambda,k)$ in (7.11) 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 it follows from Proposition 17.2.1 of Gasquet and Witomski (1998, page 157) that $\int |r(x)|dx = \int |K(x)|dx < \infty$ and $\int (1+|\lambda|)\, |\hat{r}(\lambda)|d\lambda \leq \int (1+|\lambda|)|\hat{K}(\lambda)|d\lambda < \infty$ 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 $n \geq n_0$,

$$\begin{aligned} \sum_{1\leq k<l\leq n} |I_{k,l}| &\leq C \sum_{1\leq k<l\leq n} \big[\epsilon\,(l-k)^{-3/2} + h\,(l-k)^{-1}\big]\Big(k\sum_{j=k/2}^{\infty}|\phi_j| + k^{-2} + h/\sqrt{k}\Big) \\ &\leq C\,(\epsilon + h\sum_{k=1}^{n} k^{-1})\sum_{k=1}^{n}\big(k\sum_{j=k/2}^{\infty}|\phi_j| + k^{-2} + h/\sqrt{k}\big) \\ &\leq C(\epsilon + h\,\log n)\,(C + \sqrt{n}\,h), \end{aligned}$$

since $\sum_{k=1}^{\infty} k^2|\phi_k| < \infty$. This implies (7.5) since $h\log n \to 0$. The proof of (7.4) is similar and the details are omitted.

Proofs of Proposition 7.3. Let $\psi_n'(t) = \frac{1}{\sqrt{n}\,h} \sum_{k=1}^{[nt]} K^2[(x_k - x)/h] E\mu_k^2$. We first prove

$$\begin{aligned} \sup_{0\le t\le 1} E|\psi_n(t) - \psi_n'(t)|^2 &= o(1), &(7.17)\\ \sup_{0\le t\le 1} |E\psi_n(t) - ES_n^2(t)| &= o(1). &(7.18)\end{aligned}$$

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

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

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

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

The result (7.17) follows from $nh^2 \to \infty$, $h\log n \to 0$ and arbitrary of ϵ.

By noting

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

in a similar argument as above we may prove (7.18). The details are omitted.

By noting that $\psi_n'(t) \Rightarrow L(t, 0)$ on $D[0,1]$ by using Proposition 7.1 and Theorem 2.1 of Wang and Phillips (2006), it follows from (7.17) and (7.18) that

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

for each fixed $0 \le t \le 1$. This yields that $S_n^2(t)$ and $\psi_n(t)$ are uniformly integrable by Theorem 5.4 of Billingsly (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 7.2. The result (7.17) 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 $L(t,0)$, since $\psi_n'(t) \Rightarrow L(t,0)$ on $D[0,1]$. On the other hand, $\psi_n(t)$ is tight on $D[0,1]$ since $\psi_n(t)$ is positive. This proves $\psi_n(t) \Rightarrow L(t,0)$ on $D[0,1]$.

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

$$\max_{1\le k\le n} |u_k K[(x_k-x)/h]| = o_P[(nh^2)^{1/4}], \tag{7.19}$$

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

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

we have

$$P\big[|S_n(t)-S_n(t_m)|\ge\epsilon \mid S_n(t_1),S_n(t_2),...,S_n(t_m)\big]\le\alpha_n(\epsilon,\delta), \quad a.s. \tag{7.20}$$

By noting $\max_{1\le k\le n}|u_kK[(x_k-x)/h]|\le\big\{\sum_{j=1}^n u_j^4K^4[(x_j-x)/h]\big\}^{1/4}$, the result (7.19) follows from $Eu_j^4K^4[(x_j-x)/h]\le C\,h/\sqrt{j}$ by Lemma 7.1, with a simple calculation. As for (7.20), it only needs to show that

$$\sup_{|t-s|\le\delta} P\Big(|\sum_{k=[ns]+1}^{[nt]} u_kK[(x_k-x)/h]|\ge\epsilon\, d_n \mid \epsilon_{[ns]},\epsilon_{[ns]-1},...;\eta_{[ns]},...,\eta_1\Big)\le\alpha_n(\epsilon,\delta). \tag{7.21}$$

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

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

As in the proof of (7.18) with a minor modification, it is clear that, whenever n is large enough,

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

This yields $\lim_{\delta\to0}\limsup_{n\to\infty}\alpha_n(\epsilon,\delta)=0$ for each $\epsilon>0$. The proof of Proposition 7.4 is complete.

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). Lineaer 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).

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.

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

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

Wang, Q. and P.C.B. Phillips (2006). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Cowles Fundation Discussion Paper, No. 1594, Yale University.

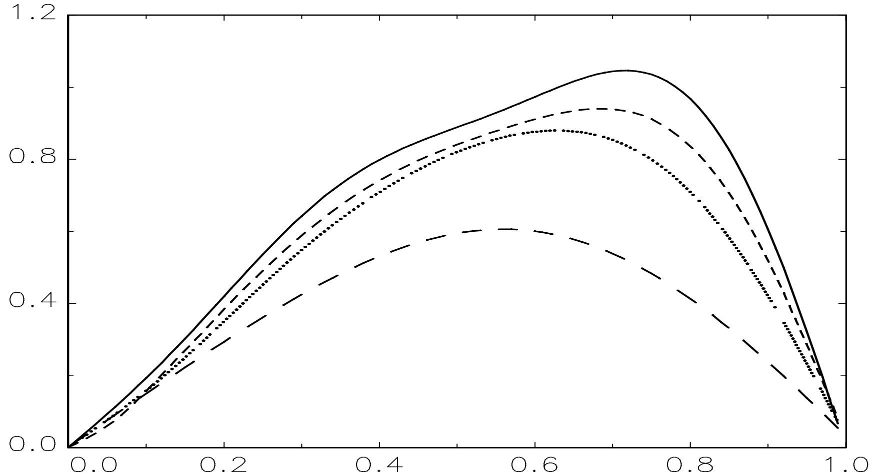

Figure 1: Graphs over the interval $[0, 1]$ of $f_A(x)$ and Monte Carlo estimates of $E\left(\hat{f}_A(x)\right)$ for $h = n^{-1/2}$ (short dashes), $h = n^{-1/3}$ (dotted) and $h = n^{-1/5}$ (long dashes) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

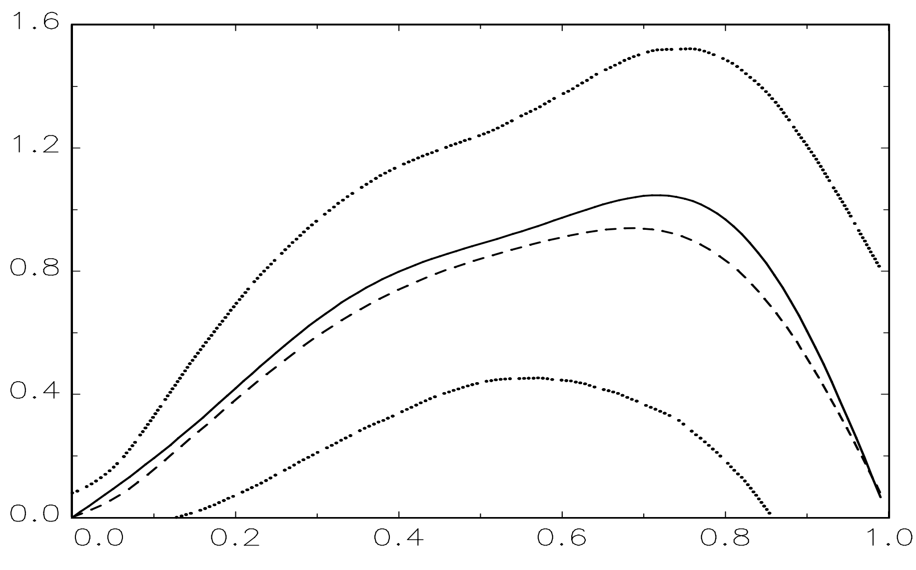

Figure 2: Graphs over the interval $[0, 1]$ of estimation bands for $f_A(x)$ (solid line), the Monte Carlo estimate of $E\left(\hat{f}_A(x)\right)$ for $h = n^{-1/2}$ (short dashes) and 95% estimation bands (dotted) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

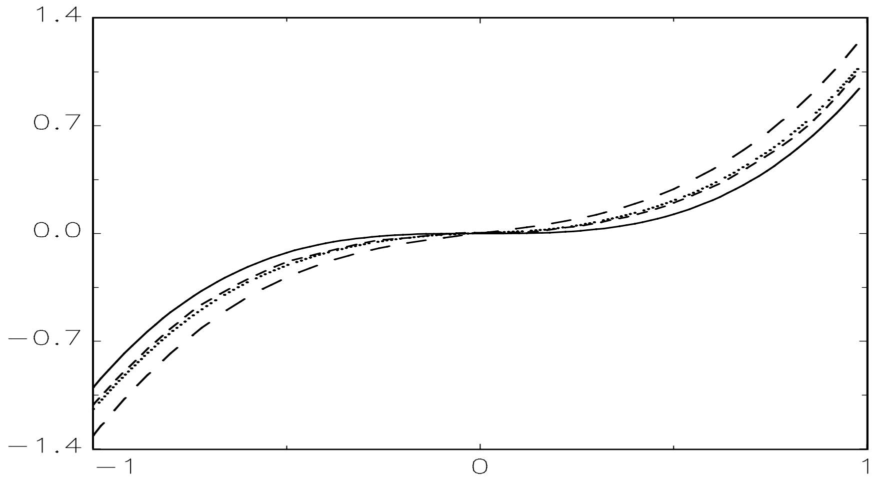

Figure 3: Graphs of $f_B(x)$ and Monte Carlo estimates of $E\left(\hat{f}_B(x)\right)$ for $h = n^{-1/2}$ (short dashes), $h = n^{-1/3}$ (dotted) and $h = n^{-1/5}$ (long dashes) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.

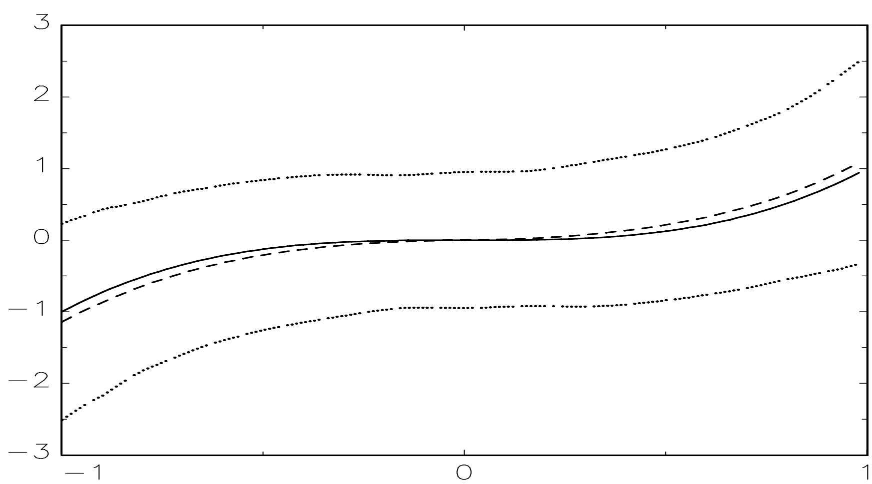

Figure 4: Graphs of estimation bands for $f_B(x)$ (solid line), the Monte Carlo estimate of $E\left(\hat{f}_B(x)\right)$ for $h = n^{-1/3}$ (short dashes) and 95% estimation bands (dotted) with $\theta = 100$, $\sigma = 0.1$ and $n = 500$.