Journal of Econometrics 167 (2012) 224–239

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Nonparametric spatial regression under near-epoch dependence

Nazgul Jenish *

Department of Economics, New York University, 19 West 4th Street, New York, NY 10012, United States

ARTICLE INFO

Article history:
Received 13 July 2010
Received in revised form
24 October 2011
Accepted 16 November 2011
Available online 3 December 2011

JEL classification:
C13
C14
C21

Keywords:
Nonparametric regression
Local linear estimator
Near-epoch dependent spatial processes
Asymptotic normality

ABSTRACT

This paper establishes asymptotic normality and uniform consistency with convergence rates of the local linear estimator for spatial near-epoch dependent (NED) processes. The class of the NED spatial processes covers important spatial processes, including nonlinear autoregressive and infinite moving average random fields, which generally do not satisfy mixing conditions. Apart from accommodating a larger class of dependent processes, the proposed asymptotic theory allows for triangular arrays of heterogeneous random fields located on unevenly spaced lattices and sampled over regions of arbitrary configuration. All these features make the results applicable in a wide range of empirical settings.

1. Introduction

Over the last years, there has been a considerable interest in modeling and estimation of spatial dependence. The econometrics literature has largely focused on parametric spatial regression models,[1] while there has been only a limited research on nonparametric estimation of spatial models. However, the functional form of the regression is often unknown, and nonparametric estimation may therefore be desirable, especially given that large spatial datasets are becoming increasingly available. Recently, Robinson (2011) has developed an asymptotic theory of the Nadaraya–Watson estimator for spatially dependent data. To model spatial dependence, Robinson (2011) assumes that the errors are, up to a random scalar, generated as a linear process in independent innovations that are independent of the regressors. This structure allows for both short- and long-range dependence.

The present paper develops an asymptotic theory of the local linear estimator for an alternative class of spatially dependent processes. Specifically, we consider random fields that are near-epoch dependent (NED) on a strongly mixing input process as defined in Jenish and Prucha (2010), who adapted this concept from the time series literature.

The main motivation for the NED condition is that many important dependent processes are not mixing, e.g., autoregressive (AR) and infinite moving average (MA (∞)) under general conditions. Sufficient conditions for the α-mixing property of linear processes with independent innovations were established by Gorodetskii (1977), and further extended by a number of authors. These conditions include: (i) smoothness of the density functions of innovations, (ii) sufficiently fast rates of decay of the coefficients, and (iii) invertibility of the linear process. Failure of these conditions may cause an AR or MA (∞) process to be nonmixing. First, Andrews (1984) showed that AR(1) processes with independent Bernoulli innovations are not α-mixing. Thus, mixing may not hold in the case of discrete innovations. Second, Gorodetskii (1977) demonstrated that the strong mixing property may fail even in the case of continuously distributed innovations when the coefficients of the linear process do not fall off sufficiently fast. At the same time, the NED property is satisfied in both examples of Andrews (1984) and Gorodetskii (1977), by Proposition 1 of this paper and similar results in Jenish and Prucha (2010).

Against this background, the advantage of the NED condition relative to mixing is threefold. First, the NED property is satisfied by linear AR and MA (∞) with discrete innovations that do not satisfy the strong mixing property. Second, as shown in this paper, the NED property is also satisfied by nonlinear AR and MA (∞) random fields under mild conditions, while such conditions are not readily available for mixing. Thus, the class of NED random fields is strictly larger than that of mixing random fields. Third, the NED condition

* Tel.: +1 212 998 3891.
E-mail address: nazgul.jenish@nyu.edu.

[1] For recent contributions, see, e.g., Robinson (2008, 2009, 2010), Lee (2004, 2007), Kelejian and Prucha (2010) and Chen and Conley (2001).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.11.008

is easier to check than mixing. As the above examples indicate, AR and MA (∞) random fields with the NED property generate a fairly rich dependence structure (which is weaker than mixing) even with independent innovations.

Spatial NED processes have also been considered by Hallin et al. (2001, 2004a) in the context of density estimation. Yet, to our knowledge, there have been no results on the nonparametric spatial regression for near-epoch dependent random fields. The asymptotic properties of kernel estimators for a spatial regression have been investigated under strong mixing conditions: by Lu and Chen (2002, 2004) in the case of Nadaraya–Watson estimator, and Hallin et al. (2004b) in the case of local linear estimator. Gao et al. (2006) study a semiparametric spatial regression, again, under strong mixing conditions.

In this paper, we focus on the local linear estimator of a regression function. This estimator has recently become popular due to a number of its desirable properties. The advantages of the local linear estimator over the Nadaraya–Watson estimator include reduced bias, design adaptability, better boundary properties and mini-max efficiency, see, e.g., Fan and Yao (2003). In the spatial context, Hallin et al. (2004b) show that the local linear estimator for stationary strong mixing random fields retains the same limiting distribution as in the i.i.d. case. We extend the latter result to a larger class of dependent random fields, and also allow for heterogeneity of the data-generating process. In addition, we establish uniform consistency of the local linear estimator with convergence rates, which may be useful in semiparametric estimation of spatial models. Our asymptotic normality result also generalizes Lu and Linton (2007) who prove asymptotic normality of this estimator for NED time series processes. For the special case of time series processes, the paper improves on the bandwidth, NED and mixing conditions used in Lu and Linton (2007).

The asymptotic results in this paper have a number of convenient features that facilitate their practical applications. First, we allow for triangular arrays, heterogeneous and unevenly spaced data—features essential for econometric applications, see Robinson (2011) for insightful discussion. Furthermore, we do not impose any restrictions on the configuration of the sample region and allow the sample to grow at different rates in different directions, i.e., allow for *nonisotropic divergence* in the terminology of Hallin et al. (2004b). Simulations suggest that the local linear estimator performs well in moderate and large samples.

The rest of the paper is organized as follows. Section 2 describes the data-generating process and estimation method. Section 3 proves asymptotic normality of the local linear estimator and gives a consistent estimator of its covariance matrix. Section 4 establishes the rates of uniform convergence of the local linear estimator. Section 5 contains a Monte Carlo study. All proofs are collected in the Appendices.

2. DGP and estimation procedure

Let $\{(Y_{in}, X_{in}^\tau)^\tau, i \in \Gamma_n\}$ be the data-generating process (DGP), where Y_{in} and X_{in} are, respectively, $\mathbb{R}$ and $\mathbb{R}^p$-valued random fields, a^τ denotes the transpose of the vector a, and $\Gamma_n \subset \Gamma$ is a finite sample region on the lattice Γ, which is equipped with the metric $\rho(i,j) = \max_{1\le l\le d} |j_l - i_l|$. Throughout, we maintain the following assumption on Γ:

Assumption 1. The lattice $\Gamma \subset \mathbb{R}^d$, $d \ge 1$, is infinite countable. All points in Γ are located at distances of at least 1 from each other.

The assumption of a minimum distance ensures the growth of the sample size via expansion of the sample region Γ_n, i.e. increasing domain asymptotics, and rules out infill asymptotics. This condition allows data located on irregularly spaced lattices, nonrectangular graphs, and trees. As such, it should be applicable to a wide range of empirical situations.

In this paper, we are interested in estimating the conditional mean function

$$E(Y_{in}|X_{in} = x) = g(x), \quad x \in \mathbb{R}^p,$$

where $g(x)$ does not depend on i and n, and is twice continuously differentiable in a neighborhood of x. We will use the local linear estimation method.

Suppose that $g(z)$ can be approximated in a neighborhood of x as:

$$g(z) \approx g(x) + g'(x)^\tau (z - x)$$

where $g'(x)$ is the derivative of $g(x)$. This suggests the following local estimator of the regression function $g(x)$ and its derivative $g'(x)$:

$$\begin{pmatrix} g_n(x) \\ g_n'(x) \end{pmatrix} = \arg \min_{(a_0,a_1)\in\mathbb{R}^{p+1}} \sum_{i\in\Gamma_n} \left(Y_{in} - a_0 - a_1^\tau (X_{in} - x)\right)^2 \times K\left((X_{in} - x)/b_n\right)$$

where b_n is a sequence of bandwidths tending to zero as $n \to \infty$, and $K(\cdot)$ is a kernel function.

To simplify notation, we will suppress dependence of random variables on x, and following Lu and Linton (2007), use the following representation:

$$\begin{pmatrix} g_n(x) \\ g_n'(x) b_n \end{pmatrix} = U_n^{-1} V_n$$

where the elements of the matrices U_n and V_n are given by

$$(V_n)_k = (\widehat{\mathbf{n}} b_n^p)^{-1} \sum_{i\in\Gamma_n} Y_{in} \left(\frac{X_{in} - x}{b_n}\right)_k K\left(\frac{X_{in} - x}{b_n}\right), \quad k = 0, 1, \ldots, p \tag{1}$$

$$(U_n)_{kl} = (\widehat{\mathbf{n}} b_n^p)^{-1} \sum_{i\in\Gamma_n} \left(\frac{X_{in} - x}{b_n}\right)_k \left(\frac{X_{in} - x}{b_n}\right)_l \times K\left(\frac{X_{in} - x}{b_n}\right), \quad k, l = 0, 1, \ldots, p \tag{2}$$

with $((X_{in} - x)/b_n)_0 := 1$. This representation yields the following expression that is convenient for derivation of the estimator's asymptotic properties:

$$\begin{pmatrix} g_n(x) - g(x) \\ (g_n'(x) - g'(x))\, b_n \end{pmatrix} = U_n^{-1}\left\{V_n - U_n \begin{pmatrix} g(x) \\ g'(x) b_n \end{pmatrix}\right\} = U_n^{-1} W_n \tag{3}$$

with

$$(W_n)_k = (\widehat{\mathbf{n}} b_n^p)^{-1} \sum_{i\in\Gamma_n} Z_{in} \left(\frac{X_{in} - x}{b_n}\right)_k \times K\left(\frac{X_{in} - x}{b_n}\right), \quad k = 0, \ldots, p \tag{4}$$

$$Z_{in} = Y_{in} - g(x) - g'(x)^\tau (X_{in} - x). \tag{5}$$

To establish the asymptotic properties of this spatial local linear estimator, we need to make some assumptions about the dependence structure of the data-generating process, $\xi_{in} = (Y_{in}, X_{in}^\tau)^\tau$. Let $\xi = \{\xi_{in}, i \in \Gamma_n, n \ge 1\}$ and $\varepsilon = \{\varepsilon_{in}, i \in \Gamma, n \ge 1\}$ be arrays of vector-valued random fields defined on a probability space $(\Omega, \mathfrak{F}, P)$ and taking their values in $\mathbb{R}^{p_x}$ and $\mathbb{R}^{p_\varepsilon}$, respectively. We assume that the probability space $(\Omega, \mathfrak{F}, P)$ is sufficiently rich to accommodate a sequence of i.i.d. random variables $\{u_i, i \in \Gamma\}$, uniformly distributed on the unit interval and independent of $\{\varepsilon_{in}, i \in \Gamma\}$.

We consider data-generating processes that are L_2-NED on some strong mixing process. For ease of reference, we give these definitions below. In the following, $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^p$, $\|\cdot\|_q = \left[E\|\cdot\|^q\right]^{1/q}$ – the L_q-norm, and $\|A\| = [\mathrm{trace}(A^\tau A)]^{1/2}$ – the norm of a nonrandom real matrix A.

Definition 1. The random field $\xi = \{\xi_{in}, i \in \Gamma_n, n \geq 1\}$, $\|\xi_{in}\|_q < \infty$, $q \geq 1$, is *said to be* L_q*-NED on* $\varepsilon = \{\varepsilon_{in}, i \in \Gamma, n \geq 1\}$ if

$$\sup_{n,i\in\Gamma_n} \|\xi_{in} - E(\xi_{in}|\mathfrak{F}_{in}(s))\|_q \leq \psi(s) \tag{6}$$

for $\psi(s)$ such that $\lim_{s\to\infty}\psi(s) = 0$, where $\mathfrak{F}_{in}(s) = \sigma(\varepsilon_{jn}; j \in \Gamma : \rho(i,j) \leq s)$.

Clearly, if a vector-valued process is NED, so is each of its components, and vice versa. The NED concept dates back to Ibragimov (1962), and was further developed by Billingsley (1968) and McLeish (1975).

Some examples of spatial processes that satisfy the NED property are given in Jenish and Prucha (2010). They show that linear AR random fields, $X_i = \sum_{l=1}^{k} a_l X_{i-v_l} + \varepsilon_i$, and nonlinear MA (∞) random fields, $X_{in} = H_{in}((\varepsilon_{jn})_{j\in\Gamma})$, are L_2-NED under some mild conditions. In particular, $X_{in} = H_{in}((\varepsilon_{jn})_{j\in\Gamma})$, where $H_{in} : \Xi^{\Gamma} \to \mathbb{R}^{p_x}$, $\Xi = \mathbb{R}^{p_\varepsilon}$ are measurable functions, is L_2-NED on $\{\varepsilon_{in}\}$ with the NED coefficients $\psi(s) = \|\varepsilon\|_2 \sup_{n,i\in\Gamma} \sum_{j\in\Gamma:\rho(i,j)>s} w_{ijn}$ if for all $e, e^\bullet \in \Xi^{\Gamma}$

$$\|H_{in}(e) - H_{in}(e^\bullet)\| \leq \sum_{j\in\Gamma} w_{ijn} \left\|e_j - e_j^\bullet\right\| \tag{7}$$

and

$$\lim_{s\to\infty}\sup_{n,i\in\Gamma}\sum_{j\in\Gamma:\rho(i,j)>s} w_{ijn} = 0, \quad \text{and} \quad \|\varepsilon\|_2 = \sup_{n,i\in\Gamma}\|\varepsilon_{in}\|_2 < \infty. \tag{8}$$

We now provide an additional example of NED random fields.

Example (*Nonlinear Autoregressive Random Fields*)**.** Consider the following nonlinear spatial autoregressive model on $\mathbb{Z}^d$:

$$X_i = F\left(\left(X_{i-j}\right)_{j\in\mathbb{Z}^d, 0<|j|\leq r}; \varepsilon_i\right) \tag{9}$$

where $r < \infty$ is fixed, $\varepsilon = \{\varepsilon_i, i \in \mathbb{Z}^d\}$ is i.i.d. and $F : \mathcal{X}^{N(r)} \times \mathbb{R}^{p_\varepsilon} \to \mathbb{R}^{p_x}$ with $\mathcal{X} = \mathbb{R}^{p_x}$, and $N(r) = \left\{j \in \mathbb{Z}^d : 0 < |j| \leq r\right\}$, is a measurable map satisfying for some $a_j \geq 0$ and all $x, x^\bullet \in \mathcal{X}^{N(r)}$:

$$\|F(0;\varepsilon_0)\|_q < \infty \tag{10}$$

$$\|F(x;\varepsilon_0) - F(x^\bullet;\varepsilon_0)\| \leq \sum_{j\in N(r)} a_j \left\|x_j - x_j^\bullet\right\|, \qquad a = \sum_{j\in N(r)} a_j < 1. \tag{11}$$

Under these conditions, by Theorem 1 of Doukhan and Truquet (2007), there exists a unique stationary solution of Eq. (9) that can be represented as $X_i = H\left(\left(\varepsilon_{i-j}\right)_{j\in\mathbb{Z}^d}\right)$. Moreover,

Proposition 1. *Under conditions (10)–(11), the random field (9) is* L_q*-NED on* $\{\varepsilon_i, i \in \mathbb{Z}^d\}$ *with the NED coefficients* $\psi(s) = 2\|X_0\|_q\, a^{[s/r]}$.

As for the input process $\varepsilon = \{\varepsilon_{in}, i \in \Gamma, n \geq 1\}$, we assume that it is mixing with the mixing coefficients defined as follows.

Definition 2. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two σ-algebras of $\mathfrak{F}$, and let

$\alpha(\mathfrak{A},\mathfrak{B}) = \sup(|P(AB) - P(A)P(B)|, A \in \mathfrak{A}, B \in \mathfrak{B})$.

For $K, L \subseteq \Gamma_n$, let $\sigma_n(K) = \sigma(\varepsilon_{in}; i \in K)$. Then, the mixing coefficients of ε are defined as: $\alpha(k,l,r) = \sup_{n,K,L}(\alpha(\sigma_n(K), \sigma_n(L)), |K| \leq k, |L| \leq l, \rho(K,L) \geq r)$, where $|K|$ denotes the cardinality of K and $\rho(K,L) = \inf\{\rho(i,j) : i \in K, j \in L\}$.

In contrast to standard mixing numbers for time-series processes, the mixing coefficients for random fields depend not only on the distance between two datasets but also their sizes. To account explicitly for such dependence, it is assumed that $\alpha(k,l,r) \leq \varphi(k,l)\widehat{\alpha}(r)$, where $\varphi(k,l)$ is nondecreasing in each argument, and $\widehat{\alpha}(r)$ is a nonincreasing function of the distance. The idea is to account separately for the two different aspects of dependence: (i) decay of dependence with the distance, and (ii) accumulation of dependence with the growth of the sample size. The common choice of $\varphi(k,l)$ in the random fields literature is $\varphi(k,l) = (k+l)^\nu$, $\nu \geq 0$. These mixing coefficients have been used widely in the random fields literature by Bulinskii (1989), Bradley (1993), Hallin et al. (2004b) and Jenish and Prucha (2010), among others. Bradley (1993) provides examples of stationary random fields satisfying these conditions with $k = l$ and $\nu = 1$. Furthermore, Bulinskii (1989) constructs linear random fields satisfying the same conditions with $\nu = 1$ for any given decay rate of $\widehat{\alpha}(r)$. Clearly, the standard mixing coefficients for time series satisfy the above conditions with $\nu = 0$.

3. Asymptotic normality and covariance matrix estimation

We establish asymptotic normality of the local linear estimator under the following set of assumptions:

Assumption 2. The data-generating process $\{\xi_{in} = \left(Y_{in}, X_{in}^\tau\right)^\tau, i \in \Gamma_n, n \geq 1\}$, $\xi_{in} = H_{in}\left(\left(\varepsilon_{jn}\right)_{j\in\Gamma}\right)$, satisfies:

(a) ξ_{in} is uniformly $L_{2+\delta}$-bounded for $\delta > 0$, i.e., $\sup_{n,i\in\Gamma_n} E|\xi_{in}|^{2+\delta} < \infty$.
(b) ξ_{in} is L_2-NED on $\{\varepsilon_{in}, i \in \Gamma, n \geq 1\}$ with the NED coefficients $\psi(m) = O(m^{-\gamma})$ with $\gamma > a(1+2/\delta)\left(p^{-1}+2^{-1}+\delta^{-1} + 2(\delta p)^{-1} + da^{-1}\right)$ for $a > \delta d/(2+\delta)$.
(c) The α-mixing coefficients of $\{\varepsilon_{in}\}$ satisfy $\alpha(k,l,r) \leq (k+l)^\nu\widehat{\alpha}(r)$ for some $\nu \geq 0$ and $\widehat{\alpha}(r) = O(r^{-\mu})$, $\mu > d\nu + (d+a)(1+2/\delta)$ for a in part (b).
(d) $\{\Gamma_n\}$ is sequence of finite subsets of Γ such that $\widehat{\mathbf{n}} = |\Gamma_n| \to \infty$. The bandwidth is such that $b_n \to 0$ and $\widehat{\mathbf{n}}b_n^p \to \infty$.

Assumption 3. (a) $g(x) = E(Y_{in}|X_{in} = x)$ is twice continuously differentiable in an open neighborhood of x.
(b) The marginal densities of $X_{in}, f_{in}(u)$, are continuous at x uniformly over $i \in \Gamma_n, n \in \mathbb{N}$, $\sup_{n,i\in\Gamma_n} f_{in}(x) < \infty$, and $\lim_{n\to\infty}\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x) = \overline{f}(x)$.
(c) The functions $G_{in}(u) = E\left(Y_{in}^2|X_{in} = u\right)$ are continuous at x uniformly over $i \in \Gamma_n, n \in \mathbb{N}$, and $\lim_{n\to\infty}\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\sigma_{in}^2(x) f_{in}(x) = \overline{\omega}(x)$, where $\sigma_{in}^2(u) = \mathrm{Var}(Y_{in}|X_{in} = u)$.
(d) There exists $\theta > 0$ such that

$$\beta_1(x,\theta) = \sup_{n,i\in\Gamma_n}\sup_{u:|u-x|<\theta} E\left(|Z_{in}|^{2+\delta}|X_{in} = u\right) f_{in}(u) < \infty$$

$$\beta_2(x,\theta) = \sup_{n,i,j\in\Gamma_n}\sup_{|u-x|<\theta,|v-x|<\theta} R_{ijn}(u,v) < \infty$$

$$\beta_3(x,\theta) = \sup_{n,i,j\in\Gamma_n}\sup_{|u-x|<\theta,|v-x|<\theta} f_{ijn}(u,v) < \infty$$

where $f_{ijn}(u,v)$ is the joint density of X_{in} and X_{jn} and

$$R_{ijn}(u,v) = E\left(\left|Z_{in}Z_{jn}\right| |X_{in} = u, X_{jn} = v\right) f_{ijn}(u,v).$$

Assumption 4. The kernel $K : \mathbb{R}^p \to \mathbb{R}$ satisfies:

(a) $K(\cdot)$ is a symmetric kernel such that $\int_{\mathbb{R}^p} K(u)du = 1$ and $\int uu^\tau K(u)du$ is nonsingular.

(b) For any $c = \left(c_0, c_1^\tau\right)^\tau \in \mathbb{R}^{p+1}$, the function $K_c(u) = \left(c_0 + c_1^\tau u\right) K(u)$ satisfies: (i) $\sup_{u\in\mathbb{R}^p} |K_c(u)| \le \overline{K}_c$ and (ii) $|K_c(u)| = O\left(|u|^{-\kappa}\right)$, with $\kappa \ge 2p$.

(c) For any $c \in \mathbb{R}^{p+1}$, $|K_c(u) - K_c(v)| \le C|u - v|$ for $u, v \in \mathbb{R}^p$ and $C < \infty$.

Assumption 2 specifies the dependence structure and moment conditions for the data-generating process as well as the bandwidth conditions. In particular, it requires uniform $L_{2+\delta}$ boundedness of the data-generating process, as in Lu and Linton (2007) and Hallin et al. (2004b). However, in contrast to these authors, we allow for nonstationary processes. Assumption 3 places some smoothness and boundedness conditions on the moment and density functions. They also restrict the degree of heterogeneity of the data-generating process. These conditions are mainly required to establish the asymptotic variance and bias of the estimator. In the stationary case, Assumptions 3(a)–(c) are standard in the literature. Assumption 3(d) is analogous to Assumption A9 of Robinson (2011).

Finally, Assumption 4 describes the class of kernels. Part (a) only simplifies the formulas for the variance and bias, and can be dropped. Parts (b)–(c) are similar to assumptions of Lu and Linton (2007). The condition on the rate of the decay of the kernel has also been used by Robinson (2011). It ensures that $\int uu^\tau K(u)du < \infty$.

Toward the asymptotic normality result, we prove a series of lemmata.

Lemma 1. *Under Assumptions 2(d), 3(b), and 4,*

$$U_n \xrightarrow{p} U = \overline{f}(x)\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \int uu^\tau K(u)du \end{pmatrix}.$$

Next lemma establishes the asymptotic variance of the estimator.

Lemma 2. *Under Assumptions 2–4, $\widehat{\mathbf{n}}b_n^p \mathrm{Var}(W_n) \to \Sigma$, where*

$$\Sigma = \overline{\varpi}(x)\begin{pmatrix} \int K^2(u)du & \mathbf{0} \\ \mathbf{0} & \int uu^\tau K^2(u)du \end{pmatrix}$$

and $\overline{\varpi}(x) = \lim_{n\to\infty} \widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} \sigma_{in}^2(x) f_{in}(x)$.

Finally, Lemma 3 provides an approximation to the asymptotic bias of the estimator. The asymptotic bias does not depend on the dependence structure of the process, and is of order of b_n^2 as in the classical i.i.d. case.

Lemma 3. *Under Assumptions 2(d), 3(a)–(b), and 4,*

$$\begin{pmatrix} Eg_n(x) - g(x) \\ E\left(g_n'(x) - g'(x)\right) b_n \end{pmatrix} = \begin{pmatrix} B_g(x)b_n^2 \\ \mathbf{0} \end{pmatrix} + o(b_n^2)$$

where

$$B_g(x) = \frac{1}{2}\sum_{k=1}^{p} g_{kk}(x)\int u_k^2 K(u)du, \tag{12}$$

$$g_{kk}(x) = \partial^2 g(x)/\partial x_k^2, \quad k = 1, \dots, p, \text{ and } x = (x_1, \dots, x_p)^\tau.$$

Based on these lemmata, one can show asymptotic normality of the local linear estimator.

Theorem 1. *Suppose Assumptions 2–4 and hold with $\overline{f}(x) > 0$. Then,*

$$\begin{pmatrix} \left(\widehat{\mathbf{n}}b_n^p\right)^{1/2}\left(g_n(x) - g(x) - B_g(x)b_n^2\right) \\ \left(\widehat{\mathbf{n}}b_n^{p+2}\right)^{1/2}\left(g_n'(x) - g'(x)\right) \end{pmatrix} \Longrightarrow N\left(\mathbf{0}, \begin{pmatrix} \sigma_0^2(x) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}_1^2(x) \end{pmatrix}\right)$$

where $B_g(x)$ is given in (12) and

$$\sigma_0^2(x) = \frac{\overline{\varpi}(x)}{\overline{f}(x)^2}\int K^2(u)du$$

$$\boldsymbol{\sigma}_1^2(x) = \frac{\overline{\varpi}(x)}{\overline{f}(x)^2}\left[\int uu^\tau K(u)du\right]^{-1} \times \int uu^\tau K^2(u)du\left[\int uu^\tau K(u)du\right]^{-1}.$$

Theorem 1 can be easily extended to joint asymptotic normality for any finite collection of points $(x_1, \dots, x_m)$ using the Cramér–Wold device. The theorem does not impose any restrictions on the configuration and growth behavior of the sample regions. The sample can grow at different rates in different directions in space. Furthermore, one can eliminate the asymptotic bias by placing the additional restriction $\left(\widehat{\mathbf{n}}b_n^p\right)^{1/2} b_n^2 \to 0$ on the bandwidth.

In general, there is a trade-off between the bandwidth and dependence conditions in asymptotic normality results for nonparametric estimators. In Theorem 1, most of restrictions are placed on the rate of decay of the NED coefficients, rather than on the bandwidth rate. This makes the bandwidth choice flexible. The latter is a more important practical problem than the rate of decay the NED coefficients. At the same time, the polynomial decay rates of the NED numbers in Theorem 1 are not restrictive as in many applications including autoregressive and moving average random fields, the NED coefficients decay at geometric rate, see Proposition 1. Given no additional restrictions on the bandwidth, the optimal bandwidth $b_n = c\widehat{\mathbf{n}}^{-1/(4+p)}$ is attainable.

In the case $d = 1$, Theorem 1 relaxes the bandwidth, mixing and NED conditions of Lu and Linton (2007, Theorem 3.1). First, we relax restrictions on the bandwidth. Second, we rely on L_2-NED instead of $L_{2+\delta/2}$-NED conditions. Third, we improve on the rate of the decay of the mixing coefficients from $\mu > (a+1)(1+4/\delta)$ to $\mu > (a+1)(1+2/\delta)$, see Assumption 2(c). For the optimal bandwidth size, the decay rate of L_2-NED numbers in Theorem 1 is even weaker that of $L_{2+\delta/2}$-NED numbers in Corollary 3.1 of Lu and Linton (2007).

Finally, Theorem 1 neither dominates nor is dominated by the asymptotic normality result of Robinson (2011). Compared to Robinson (2011), Theorem 1 allows for dependence between the regressors and errors, and permits error processes generated as nonlinear functionals of some input process. On the other hand, Robinson (2011) allows for both short- and long-range dependence, while Theorem 1 accommodates only short-range dependence.

We now turn to estimation of the covariance matrix of the local linear estimator. To this end, we need to construct estimators for $\overline{\varpi}(x)$ and $\overline{f}(x)$. In the stationary case, $\overline{f}(x)$ reduces to the common density function $f(x)$, and $\overline{\varpi}(x) = \left[E\left(Y_i^2|X_i = x\right) - g^2(x)\right]f(x)$ so that one can use the standard kernel estimators for $f(x)$, $E\left(Y_i^2|X_i = x\right)$, and $g(x)$.

In the nonstationary case, we propose the following estimators

$$\widehat{f}_n(x) = \frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n} K\left(\frac{X_{in} - x}{b_n}\right)$$

and

$$\widehat{\omega}_n(x) = \frac{1}{\widehat{\mathbf{n}} b_n^p} \sum_{i \in \Gamma_n} \widehat{e}_{in}^2(x) K\left(\frac{X_{in} - x}{b_n}\right) \text{ with } \widehat{e}_{in}(x) = Y_{in} - g_n(x).$$

Theorem 2. *Under Assumptions 2–4, $p\lim \widehat{f}_n(x) = \bar{f}(x)$, and $p\lim \widehat{\omega}_n(x) = \bar{\omega}(x)$.*

4. Uniform consistency

In many applications, it is important to obtain uniform consistency of a nonparametric estimator rather than just pointwise consistency, which follows immediately from Theorem 1. For instance, in the two-step semiparametric estimation, uniform consistency of the first-step nonparametric estimator is key to proving consistency and asymptotic normality of the second-step parametric estimator. In this section, we therefore establish uniform consistency of the local linear estimator.

We maintain the general set-up of the previous sections and Assumptions 1–2. In addition, we need the following assumptions.

Assumption 5. (a) The distribution of X_{in} is absolutely continuous with respect to Lebesgue measure with density $f_{in}(x)$ for all $i \in \Gamma_n$, $n \in \mathbb{N}$.
(b) The average densities $\bar{f}_n(x) = \widehat{\mathbf{n}}^{-1} \sum_{i \in \Gamma_n} f_{in}(x)$ are uniformly bounded, i.e., $\sup_{n \geq 1} \sup_{x \in \mathbb{R}^p} \bar{f}_n(x) < \infty$, and satisfy for $B_f < \infty$ and all $x, x^{\bullet} \in \mathbb{R}^p$

$$\sup_{n \geq 1} \left| \bar{f}_n(x) - \bar{f}_n(x^{\bullet}) \right| \leq B_f \left\| x - x^{\bullet} \right\|.$$

(c) $g(x) = E(Y_{in} | X_{in} = x)$ is twice continuously differentiable on $\mathbb{R}^p$, and $\sup_{x \in \mathbb{R}^p} \left\| g''(x) \right\| < \infty$, where $g''(x)$ is the Hessian of $g(x)$.

Assumption 6. Kernel $K : \mathbb{R}^p \to \mathbb{R}$ satisfies Assumption 4(a), $\int \|x\|^3 |K(x)|\, dx < \infty$. For $k, l = 0, 1, \ldots, p$, the functions $\widetilde{K}_{kl}(x) = x_k x_l K(x)$, where $x_0 = 1$ and $x = (x_1, \ldots, x_p)^{\tau} \in \mathbb{R}^p$, have Fourier transforms

$$\Psi_{kl}(r) = (2\pi)^p \int_{\mathbb{R}^p} \exp(\mathbf{i} r^{\tau} x) \widetilde{K}_{kl}(x) dx, \tag{13}$$

that satisfy $\int (1 + \|r\|) \|\Psi(r)\|\, dr < \infty$, where $\mathbf{i} = \sqrt{-1}$ and $\Psi(r)$ is the $(p+1) \times (p+1)$ matrix with the elements, $\Psi_{kl}(r)$, defined in (13).

Theorem 3. *Let $\mathbf{X} = \left\{x : \widehat{\mathbf{n}}^{-1} \sum_{i \in \Gamma_n} f_{in}(x) \geq M > 0\right\}$. Suppose Assumption 5–6 and 2 hold with $\mu > d(\nu + 1 + 2/\delta)$. Then,*

$$\sup_{x \in \mathbf{X}} |g_n(x) - g(x)| = O_p(\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)} b_n^{-p-d/(2\gamma+d)}) + O_p\left(b_n^2\right)$$

$$\sup_{x \in \mathbf{X}} \left| g_n'(x) - g'(x) \right| = O_p(\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)} b_n^{-p-1-d/(2\gamma+d)}) + O_p\left(b_n\right)$$

provided that the right-hand sides of both expressions are $o_p(1)$.

Theorem 3 obtains the rates of uniform convergence for the estimators of both regression function and its derivative. This result could be useful in proving asymptotic normality of semiparametric estimators when the local linear estimator is used as a first-step estimator of some unknown function. In the time series context, uniform consistency of the Nadaraya–Watson estimator has been established by Bierens (1983) and Andrews (1995). In the special case $d = 1$, the rates of convergence in Theorem 3 are consistent with those in Theorem 1(b) of Andrews (1995).

5. Monte Carlo results

In this section, we report the results of a Monte Carlo study of the local linear estimator for spatial NED processes. We consider two dependent spatial processes which are *not* strong mixing, but satisfy the NED property under mild conditions.

Model 1. *Regular lattice.* Let $\left\{e_{i,j}, (i,j) \in \mathbb{Z}^2\right\}$ and $\{\varepsilon_{i,j}, (i,j) \in \mathbb{Z}^2\}$ be two independent processes on $\mathbb{Z}^2$, where $e_{i,j} \sim$ i.i.d. $N(0,1)$ and $\varepsilon_{i,j} \sim$ i.i.d. Bernoulli $(1/2)$. Consider

$$Y_{i,j} = g(X_{i,j}) + u_{i,j} \quad \text{with } g(x) = \sin(2x) + 2\exp(-16x^2) \tag{14}$$

where $\left\{u_{i,j}\right\}$ and $\left\{X_{i,j}\right\}$ satisfy the spatial autoregressions

$$u_{i,j} = \frac{1}{4}\left(u_{i-1,j} + u_{i,j-1}\right) + \varepsilon_{i,j}, \qquad X_{i,j} = \frac{1}{8}\left(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}\right) + e_{i,j}.$$

By Proposition 1 of this paper and Proposition 1 of Jenish and Prucha (2010), the process $\left\{\left(Y_{i,j}, X_{i,j}\right)^{\tau}\right\}$ is L_2-NED on $\left\{\left(e_{i,j}, \varepsilon_{i,j}\right)^{\tau}\right\}$. However, $\left\{u_{i,j}\right\}$ and hence $\left\{Y_{i,j}\right\}$ are *not* mixing, which can be shown using arguments similar to Andrews (1984).

Data are simulated over a rectangular grid of $(m + 300) \times (n + 300)$ locations. To achieve stationarity, we iterated on the data generating process for 150 iterations, discarding the first 149 samples, and used sample of size mn from the final iteration (by discarding the 300 outer boundary points along each of the axes). We use samples of size 200, 600 and 1200.

Model 2. *Irregular lattice.* To demonstrate robustness of our results to irregularly spaced data, we consider an unevenly spaced lattice Γ depicted in Fig. 2(d). The lattice is generated by removing, at random, nodes of the regular lattice $\mathbb{Z}^2$.

As before, let $\left\{e_{i,j}, (i,j) \in \Gamma\right\}$ and $\left\{\varepsilon_{i,j}, (i,j) \in \Gamma\right\}$ be two independent processes, where $e_{i,j} \sim$ i.i.d. $N(0,1)$ and $\varepsilon_{i,j} \sim$ i.i.d. Bernoulli $(1/2)$, and let $\left\{Y_{i,j}, (i,j) \in \Gamma\right\}$ be generated according to (14) with

$$u_{i,j} = \sum_{(k,l) \in \Gamma} 2^{-|k|-|l|} \varepsilon_{i-k,j-l}, \qquad X_{i,j} = \sum_{(k,l) \in \Gamma} 2^{-|k|-|l|} e_{i-k,j-l}.$$

While not mixing, $\left\{\left(Y_{i,j}, X_{i,j}\right)^{\tau}\right\}$ is L_2-NED on $\left\{\left(e_{i,j}, \varepsilon_{i,j}\right)^{\tau}\right\}$ in light of (7)–(8).

The box plots of the estimator based on 100 replications as well as the estimated curves are depicted in Figs. 1 and 2 for Models 1 and 2, respectively. In all simulations, we used the Epanechnikov kernel and the bandwidth $b = 0.2$ chosen according to the cross-validation rule of Craven and Wahba (1979).

Inspection of Figs. 1 and 2 reveals tat the local linear estimator performs well in medium and large samples both for regularly and irregular spaced data. Uneven spacing of data, which is prevalent in economic applications, does not seem to affect adversely the properties of the local linear estimator, except it requires larger sample regions for sparse data to get a desirable sample size. Overall, the simulations results agree well with the asymptotic theory of the previous sections: the estimated curves fit the true line more closely and become more stable as the sample size increases.

Acknowledgments

I am grateful to the Editor P.M. Robinson, Associate Editor and two anonymous referees for their valuable comments that led to considerable improvement of the paper. I owe special thanks to the referee for pointing to some errors in the earlier version of the paper. I would also like to thank Robert de Jong and the participants of the Econometrics seminar at the OSU for their helpful comments.

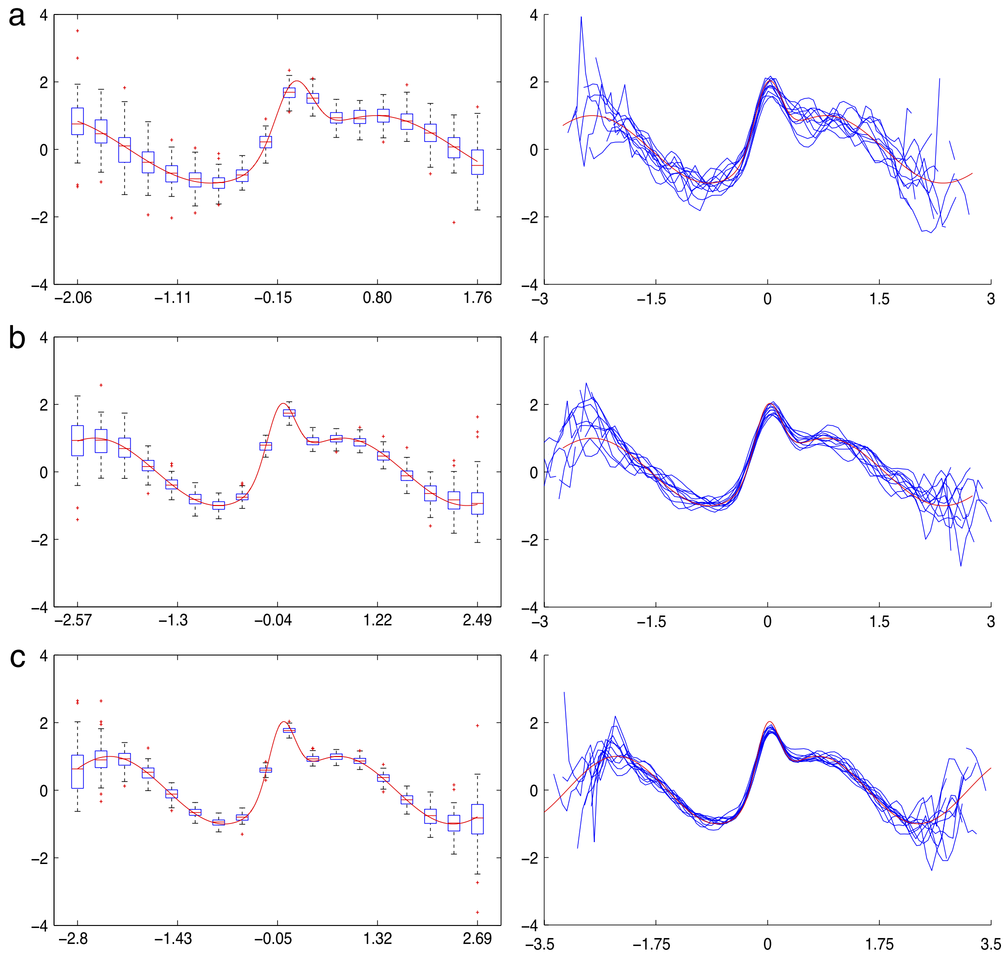

Fig. 1. Regular lattice $\mathbb{Z}^2$: box plots and estimated curves for model: $Y_{ij} = \sin(2X_{ij}) + 2\exp(-16X_{ij}^2) + u_{ij}$, where $X_{ij} = \frac{1}{8}(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}) + e_{ij}$, $e_{ij} \sim$ i.i.d. $N(0,1)$, $u_{ij} = \frac{1}{4}(u_{i,j-1} + u_{i-1,j}) + \varepsilon_{ij}$, $\varepsilon_{ij} \sim$ i.i.d. Bernoulli $\left(\frac{1}{2}\right)$ (a) sample size $= 200$ (b) sample size $= 600$ (c) sample size $= 1200$.

Appendix A. Proofs for Section 2

The proofs of our results in this and subsequent sections makes use of the following lemma given in Dudley and Philipp (1983), which is a variant of the classical Skorohod's coupling lemma.

Lemma A.1 (*Dudley and Philipp, 1983*)**.** *Suppose $(\Omega, \mathfrak{F}, P)$ is a probability space, S_1 and S_2 are Polish spaces, and Q is a probability measure on $S_1 \times S_2$. Furthermore, suppose that there are independent random variables X and U, where U is uniformly distributed on $[0, 1]$, and $X : \Omega \to S_1$ has the distribution S_1, the first marginal of Q. Then, there exists a function $h : [0, 1] \times S_1 \to S_2$ such that $(X, h(U, X))$ has the distribution Q.*

Proof of Proposition 1. By Theorem 1 of Doukhan and Truquet (2007), there exists a unique stationary solution of Eq. (9) that can be represented as $X_i = H\left(\left(\varepsilon_{i-j}\right)_{j\in\mathbb{Z}^d}\right)$. For any $s \in \mathbb{N}$, let $\eta^{(s)} = (\varepsilon_i)_{i\in\mathbb{Z}^d:|i|\le s}$, and $\zeta^{(s)} = (\varepsilon_i)_{i\in\mathbb{Z}^d:|i|>s}$. By Lemma A.1, there exists a function $h(U, \eta^{(s)})$ such that the process $\varepsilon^{(s)} = \left(\varepsilon_i^{(s)}\right)_{i\in\mathbb{Z}^d} = \left(\eta^{(s)}; h(U, \eta^{(s)})\right)$ has the same distribution as $\varepsilon = (\varepsilon_i)_{i\in\mathbb{Z}^d} = \left(\eta^{(s)}, \zeta^{(s)}\right)$. In other words, $\varepsilon^{(s)}$ is a copy of ε with $\varepsilon_i^{(s)} = \varepsilon_i$ for $|i| \le s$.

Now, define the following process $X_i^{(s)} \equiv H\left(\left(\varepsilon_{i-j}^{(s)}\right)_{j\in\mathbb{Z}^d}\right)$. Since $\varepsilon^{(s)}$ has the same distribution as ε, we have for all $s \in \mathbb{N}$ and $i \in \mathbb{Z}^d$

$$X_i^{(s)} = F\left(\left(X_{i-j}^{(s)}\right)_{j\in\mathbb{Z}^d, 0<|j|\le r}; \varepsilon_i^{(s)}\right).$$

If $|i| \le s$, then $\varepsilon_i^{(s)} = \varepsilon_i$ and by condition (11), we have

$$\left\| X_i - X_i^{(s)} \right\|_q \le \sum_{j\in N(r)} a_j \left\| X_{i-j} - X_{i-j}^{(s)} \right\|_q. \tag{A.1}$$

Consider two cases: $s \ge r$ and $s < r$. If $s \ge r$, recursive use of (A.1) gives

$$\begin{aligned}
&\left\| X_0 - X_0^{(s)} \right\|_q \\
&\quad\le \sum_{j_1\in N(r)} a_{j_1} \left\| X_{-j_1} - X_{-j_1}^{(s)} \right\|_q \\
&\quad\le \sum_{j_1\in N(r)} a_{j_1} \sum_{j_2\in N(r)} a_{j_2} \left\| X_{-j_1-j_2} - X_{-j_1-j_2}^{(s)} \right\|_q \\
&\quad\le \cdots \le \sum_{j_1\in N(r)} a_{j_1} \ldots \sum_{j_k\in N(r)} a_{j_k} \\
&\qquad\times \left\| X_{-j_1-j_2-\cdots-j_k} - X_{-j_1-j_2-\cdots-j_k}^{(s)} \right\|_q \le 2\left\| X_0 \right\|_q a^k
\end{aligned}$$

where $k = [s/r]$ is the integer of part of s/r. If $s < r$, then

$$\begin{aligned}
\left\| X_0 - X_0^{(s)} \right\|_q &\le \sum_{j_1\in N(r)} a_{j_1} \left\| X_{-j_1} - X_{-j_1}^{(s)} \right\|_q \\
&\le 2\left\| X_0 \right\|_q a \le 2\left\| X_0 \right\|_q a^k
\end{aligned}$$

since $a < 1$ and $k = [s/r] < 1$.

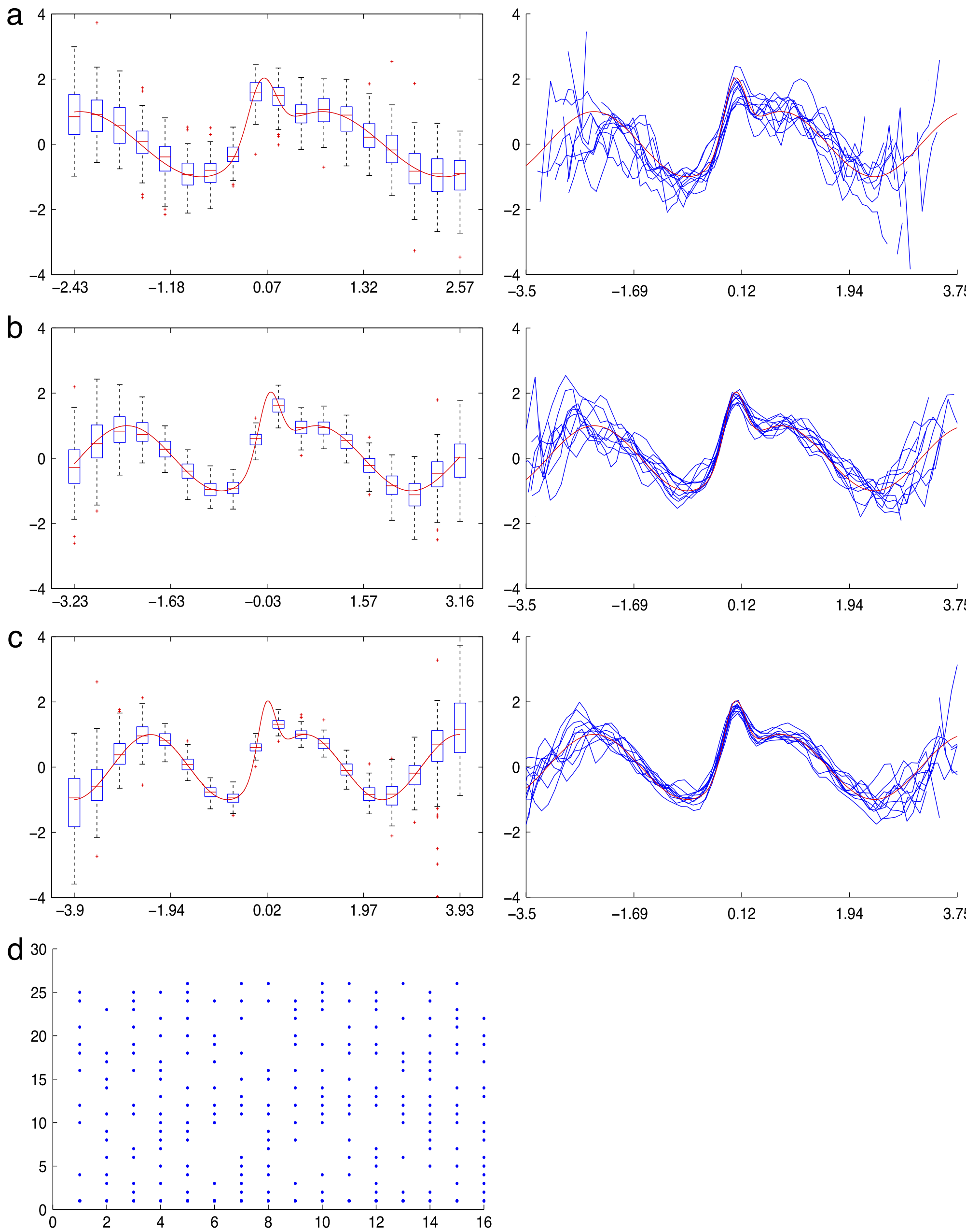

Fig. 2. Irregular lattice Γ: box plots and estimated curves for model: $Y_{ij} = \sin(2X_{ij}) + 2\exp(-16X_{ij}^2) + u_{ij}$, $X_{ij} = \sum_{(k,l)\in\Gamma} 2^{-(|k|+|l|)} e_{i-k,j-l}$, with $e_{ij} \sim$ i.i.d. $N(0,1)$, $u_{ij} = \sum_{(k,l)\in\Gamma} 2^{-(|k|+|l|)} \varepsilon_{i-k,j-l}$, $\varepsilon_{ij} \sim$ i.i.d. Bernoulli $\left(\frac{1}{2}\right)$ (a) sample size = 200 (b) sample size = 600 (c) sample size = 1200 (d) lattice Γ.

We now verify the NED condition $\|X_0 - E[X_0|\mathfrak{F}_0(s)]\|_q \to 0$. Since

$$X_0 = H\left(\left(\varepsilon_j\right)_{j\in\mathbb{Z}^d}\right) = H\left(\eta^{(s)}, \zeta^{(s)}\right) \quad \text{and}$$

$$X_0^{(s)} = H\left(\left(\varepsilon_j^{(s)}\right)_{j\in\mathbb{Z}^d}\right) = H\left(\eta^{(s)}, h(U, \eta^{(s)})\right)$$

have the same distribution, $\mathfrak{F}_0(s) = \sigma\left((\varepsilon_i)_{i\in\mathbb{Z}^d:|i|\le s}\right) = \sigma(\eta^{(s)})$ and $\varepsilon_i^{(s)} = \varepsilon_i$ for all $|i| \le s$, we have $E[X_0|\mathfrak{F}_0(s)] = E[X_0^{(s)}|\mathfrak{F}_0(s)] = \int_0^1 H\left(\eta^{(s)}, h(u, \eta^{(s)})\right) du$.

Then, by the Jensen inequality

$$\begin{aligned}
&\|X_0 - E[X_0|\mathfrak{F}_0(s)]\|_q \\
&= \left\| H\left(\eta^{(s)}, \zeta^{(s)}\right) - \int_0^1 H\left(\eta^{(s)}, h(u, \eta^{(s)})\right) du \right\|_q \\
&\le \left\{ E\int_0^1 \left| H\left(\eta^{(s)}, \zeta^{(s)}\right) - H\left(\eta^{(s)}, h(u, \eta^{(s)})\right)\right|^q du \right\}^{1/q} \\
&= \left\{ E\left| H\left(\eta^{(s)}, \zeta^{(s)}\right) - H\left(\eta^{(s)}, h(U, \eta^{(s)})\right)\right|^q \right\}^{1/q} \\
&= \left\| H\left(\eta^{(s)}, \zeta^{(s)}\right) - H\left(\eta^{(s)}, h(U, \eta^{(s)})\right)\right\|_q \\
&= \left\| X_0 - X_0^{(s)} \right\|_q \le \psi(s) = 2\|X_0\|_q a^{[s/r]} \to 0
\end{aligned}$$

which completes the proof of the proposition. □

Appendix B. Technical lemmata

In this section, we establish a series of technical lemmata that will be used in the proof of our main results. Let $N_i(m) = \{j \in \Gamma : \rho(i,j) \le m\}$ denote the m-neighborhood of point $i \in \Gamma$, and $\mathfrak{F}_{in}(m) = \sigma(\varepsilon_{jn}; j \in N_i(m))$ denote the σ-algebra generated by the ε_{jn} located in $N_i(m)$. Throughout, C denotes a generic positive constant that does not depend on n and may vary from line to line.

The proof of asymptotic normality relies on the approximation of the data-generating process $\xi_{in} = \left(Y_{in}, X_{in}^{\tau}\right)^{\tau}$ by the process $\xi_{in}^{(m)}$ defined in the following lemma.

Lemma B.2. *Let $\xi_{in} = H_{in}\left(\left(\varepsilon_{jn}\right)_{j\in\Gamma}\right)$ be L_q-NED on $\varepsilon = \{\varepsilon_{in}, i \in \Gamma\}$ with the NED coefficients $\{\psi(m)\}$. Suppose that there exists on $(\Omega, \mathfrak{F}, P)$ a sequence of i.i.d. random variables $\{U_i, i \in \Gamma\}$, $U_i \sim$ Uniform $[0,1]$, that is independent of $\{\varepsilon_{in}\}$. Then, for any fixed $m \in \mathbb{N}$, there exists a process $\xi_{in}^{(m)}$ satisfying:*

(a) *$\xi_{in}^{(m)}$ has the same distribution as ξ_{in} for all $i \in \Gamma_n$, $n \ge 1$.*

(b) $\sup_{n,i\in\Gamma_n} \left\| \xi_{in} - \xi_{in}^{(m)} \right\|_q \le 2\psi(m)$.

(c) *If the α-mixing coefficients of ε satisfy $\alpha_\varepsilon(k,l,r) \le (k+l)^\nu \widehat{\alpha}(r)$ for some $\nu \ge 0$ and $\widehat{\alpha}(r)$, then the α-mixing coefficients of $\xi_{in}^{(m)}$, $\alpha_{\xi,m}(1,1,r)$, satisfy*

$$\alpha_{\xi,m}(1,1,r) \le \begin{cases} 1, & r \le 2m \\ Cm^{d\nu}\widehat{\alpha}(r-2m), & r > 2m \end{cases} \tag{B.2}$$

for some constant $C < \infty$ that does not depend on m and n.

Proof of Lemma B.2. *Part* (a). For any $i \in \Gamma_n$, let $\eta_{in}^{(m)} = \left(\varepsilon_{jn}\right)_{j\in N_i(m)}$ and $\zeta_{in}^{(m)} = \left(\varepsilon_{jn}\right)_{j\in\Gamma\setminus N_i(m)}$. By Lemma A.1, there exists a function $h(U_i, \eta_{in}^{(m)})$ s. t. $\varepsilon^{(m)} = \left(\eta_{in}^{(m)}, h(U_i, \eta_{in}^{(m)})\right)$ has the same distribution as $\varepsilon = \left(\eta_{in}^{(m)}, \zeta_{in}^{(m)}\right)$. Then, the process $\xi_{in}^{(m)} \equiv H_{in}(\eta_{in}^{(m)}, h(U_i, \eta_{in}^{(m)}))$ has the same distribution as ξ_{in}.

Part (b). Since $\xi_{in}^{(m)}$ has the same distribution as ξ_{in} and $\varepsilon_{in}^{(m)} = \varepsilon_{in}$ for all $|i| \le m$, $E\left(\xi_{in}^{(m)} | \mathfrak{F}_{in}(m)\right) = E\left(\xi_{in} | \mathfrak{F}_{in}(m)\right)$. Then, for all $n \ge 1$, $i \in \Gamma_n$

$$\left\| \xi_{in} - \xi_{in}^{(m)} \right\|_q \le \| \xi_{in} - E\left(\xi_{in} | \mathfrak{F}_{in}(m)\right) \|_q + \left\| \xi_{in}^{(m)} - E\left(\xi_{in}^{(m)} | \mathfrak{F}_{in}(m)\right) \right\|_q \le 2\psi(m).$$

Part (c). To simplify notation, we suppress dependence on n and m, and let $\xi_i \equiv \xi_{in}^{(m)}$, $\eta_i \equiv \eta_{in}^{(m)}$ and $\xi_i \equiv \Phi_i(\eta_i, U_i) = H_i(\eta_i, h(U_i, \eta_i))$. We first show that

$$\alpha_{\xi,m}(1,1,r) \le \alpha_{\eta,m}(1,1,r). \tag{B.3}$$

For any $i, j \in \Gamma$ such that $\rho(i,j) \ge r$, define the sets

$$A = \{\omega : \xi_i(\omega) \in D\} \in \sigma(\xi_i) \quad \text{for } D \in \mathfrak{B}(\mathbb{R}^{p+1})$$
$$B = \left\{\omega : \xi_j(\omega) \in G\right\} \in \sigma(\xi_j) \quad \text{for } G \in \mathfrak{B}(\mathbb{R}^{p+1})$$

where $\mathfrak{B}(\mathbb{R}^p)$ is the Borel-field on $\mathbb{R}^p$. Also, define

$$D|u = \{\eta_i : \Phi_i(\eta_i, u) \in D\} \in \mathfrak{B}(\mathbb{R}^{p_\eta}),$$
$$A^* = \{\omega : \eta_i(\omega) \in D|u\} \in \sigma(\eta_i)$$
$$G|u' = \left\{\eta_j : \Phi_j(\eta_j, u') \in G\right\} \in \mathfrak{B}(\mathbb{R}^{p_\eta}),$$
$$B^* = \left\{\omega : \eta_j(\omega) \in G|u'\right\} \in \sigma(\eta_j).$$

Given $\mathfrak{F}/\mathfrak{B}(\mathbb{R}^{p+1})$ measurability of ξ_i, to prove (B.3), it suffices to show that for any $i, j \in \Gamma$ such that $\rho(i,j) \ge r$ and any $G, D \in \mathfrak{B}(\mathbb{R}^{p+1})$

$$|P(AB) - P(A)P(B)| \le \sup_{A^*\in\sigma(\eta_i), B^*\in\sigma(\eta_j)} |P(A^*B^*) - P(A^*)P(B^*)|. \tag{B.4}$$

First, note that by independence of $\{U_i\}$ and $\{\eta_i\}$

$$P(A) = \int_0^1 \left(\int_{D|u} dF_{\eta_i} \right) du = \int_0^1 P(D|u)\, du$$

$$P(AB) = \int_0^1 \int_0^1 \left(\int_{D|u\cap G|u'} dF_{\eta_i\eta_j} \right) du du' = \int_0^1 \int_0^1 P\left(D|u \cap G|u\right) du du'.$$

Then,

$$|P(AB) - P(A)P(B)| \le \int_0^1 \int_0^1 \Big| P\left(D|u \cap G|u'\right) - P\left(D|u\right) P\left(G|u'\right) \Big| du du' \le \sup_{A^*\in\sigma(\eta_i), B^*\in\sigma(\eta_j)} |P(A^*B^*) - P(A^*)P(B^*)|$$

which verifies (B.4). Now, taking sup over all $A \in \sigma(\xi_i)$, $B \in \sigma(\xi_j)$, and then over all $i, j \in \Gamma$ s.t. $\rho(i,j) \ge r$ in the left-hand-side of (B.4) gives

$$\alpha_{\xi,m}(1,1,r) \le \alpha_{\eta,m}(1,1,r).$$

To prove (B.2), note that $\sigma(\eta_{in}^{(m)}) = \mathfrak{F}_{in}(m)$ and hence the mixing coefficients of $\eta_{in}^{(m)}$ satisfy

$$\alpha_{\eta,m}(1,1,r) \le \begin{cases} 1, r \le 2m \\ \alpha(Cm^d, Cm^d, r-2m), r > 2m \end{cases} \le \begin{cases} 1, & r \le 2m \\ Cm^{d\nu}\widehat{\alpha}(r-2m), & r > 2m \end{cases}$$

where the $\alpha(\cdot,\cdot,r-2m)$ are the mixing coefficients of the input process ε. Here, we used the fact that the m-neighborhood of any point on Γ contains at most Cm^d points of Γ for some constant C that does not depend on m, as shown in Lemma A.1 of Jenish and Prucha (2009). The proof of the lemma is now complete. □

Next lemma establishes various moment conditions for the data-generating process.

Lemma B.3. *Let*

$$K_{cin} = K_c\left((x - X_{in})/b_n\right) \quad \text{and} \quad \Delta_{in} = Z_{in}K_{cin} - EZ_{in}K_{cin}.$$

Under Assumptions 2–4,

(a) *For $q = 1, 2, 2+\delta$, $\sup_{i\in\Gamma_n} E\left|Z_{in}K_{cin}\right|^q = O\left(b_n^p\right)$.*

(b) $\lim_{n\to\infty} \left(\widehat{\mathbf{n}}b_n^p\right)^{-1} \sum_{i\in\Gamma_n} EZ_{in}^2K_{cin}^2 = \overline{\omega}(x)\int_{\mathbb{R}^p} K_c^2(u)du$, *where $\overline{\omega}(x)$ defined in Assumption 3(d).*

(c) $\lim_{n\to\infty} \left(\widehat{\mathbf{n}}b_n^p\right)^{-1} \sum_{i\in\Gamma_n} \sum_{j\ne i, j\in\Gamma_n} E\left(\Delta_{in}\Delta_{jn}\right) = 0$.

Proof of Lemma B.3. *Part* (a). Let $M_{in}(u) = E\left(|Z_{in}|^{2+\delta} | X_{in} = u\right)$. Using the arguments analogous to those in Robinson (2009), we have

$$E\left|Z_{in}K_{cin}\right|^{2+\delta} = b_n^p \int M_{in}(x - b_nu) f_{in}(x - b_nu) \left|K_c(u)\right|^{2+\delta} du = \left(\int_{u:|b_nu|<\theta} + \int_{u:|b_nu|\ge\theta} \right) \times b_n^p M_{in}(x - b_nu) f_{in}(x - b_nu) \left|K_c(u)\right|^{2+\delta} du.$$

By Assumptions 3(d) and 4, the first integral on the r.h.s. of the last equality is bounded by $b_n^p\beta_1(x,\theta)\int|K_c(u)|^{2+\delta}du = O(b_n^p)$ and the second integral by $\sup_{|u|\geq\theta/b_n}|K_c(u)|^{2+\delta}\sup_{n,i\in\Gamma_n}E|Z_{in}|^{2+\delta} = O\left(b_n^{\kappa(2+\delta)}\right) = O(b_n^p)$, since $|K_c(u)| = O\left(|u|^{-\kappa}\right)$ with $\kappa\geq 2p$ and hence $\sup_{|u|\geq\theta/b_n}|K_c(u)| = O(b_n^{\kappa})$. Thus, $\sup_{i\in\Gamma_n}E|Z_{in}K_{cin}|^{2+\delta} = O\left(b_n^p\right)$.

Similarly, using Assumptions 3(b) and (d), one can show that

$$\sup_{i\in\Gamma_n}E|Z_{in}K_{cin}| = O\left(b_n^p\right) \quad\text{and}\quad \sup_{i\in\Gamma_n}EZ_{in}^2K_{cin}^2 = O\left(b_n^p\right).$$

Part (b). Observe that

$$\begin{aligned}&\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}EZ_{in}^2K_{cin}^2\\ &\quad=\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}E\left\{K_{cin}^2\left[E\left(Y_{in}^2|X_{in}\right)-g^2(x)\right]\right\}\\ &\quad=\int\left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\left[G_{in}(x-b_nu)-g^2(x)\right]f_{in}(x-b_nu)\right\}K_c^2(u)\,du.\end{aligned}$$

By Assumptions 3(b)–(c), for any $\epsilon>0$ there exists $N_1(\epsilon)$ such that for all $n\geq N_1$ and all $i\in\Gamma_n$

$$\begin{aligned}&\left|\left(G_{in}(x-b_nu)-g^2(x)\right)f_{in}(x-b_nu)\right.\\ &\quad\left.-\left(G_{in}(x)-g^2(x)\right)f_{in}(x)\right|<\epsilon/2\end{aligned}$$

where $G_{in}(x)-g^2(x) = E\left(Y_{in}^2|X_{in}=x\right) - E^2(Y_{in}|X_{in}=x) = \sigma_{in}^2(x)$.

Since the last inequality holds for all $i\in\Gamma_n$, we have

$$\begin{aligned}&\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\left|\left(G_{in}(x-b_nu)-g^2(x)\right)f_{in}(x-b_nu)-\sigma_{in}^2(x)f_{in}(x)\right|\\ &\quad<\epsilon/2.\end{aligned}$$

Furthermore, by Assumption 3(c), for the same ϵ there exists $N_2(\epsilon)$ such that for all $n\geq N_2$

$$\left|\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\sigma_{in}^2(x)f_{in}(x)-\overline{\omega}(x)\right|<\epsilon/2.$$

Hence, $\lim_{n\to\infty}\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\left[G_{in}(x-b_nu)-g^2(x)\right]f_{in}(x-b_nu) = \overline{\omega}(x)$. Then, by the Lebesgue density theorem

$$\lim_{n\to\infty}\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}EZ_{in}^2K_{cin}^2 = \overline{\omega}(x)\int_{\mathbb{R}^p}K_c^2(u)du.$$

Part (c). Define the following sets on the lattice:

$$\begin{aligned}&I_1 = \{i\neq j\in\Gamma_n:\rho(i,j)\leq 3m\}\\ &\text{and}\quad I_2 = \{i,j\in\Gamma_n:\rho(i,j)>3m\}.\end{aligned}\tag{B.5}$$

By Lemma A.1 of Jenish and Prucha (2009), the cardinality of the set I_1 satisfies $|I_1|\leq C\,\widehat{\mathbf{n}}m^d$. Now, decompose the sum as:

$$\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\sum_{j\neq i,j\in\Gamma_n}E\Delta_{in}\Delta_{jn} = Q_{1n}+Q_{2n}$$

where

$$Q_{1n} = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i,j\in I_1}\sum E\Delta_{in}\Delta_{jn}\quad\text{and}$$

$$Q_{2n} = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i,j\in I_2}\sum E\Delta_{in}\Delta_{jn}.$$

We need to show that $\lim_{n\to\infty}(Q_{1n}+Q_{2n}) = 0$. We first bound Q_{1n}. Observe that $E\Delta_{in}\Delta_{jn} = E\left(Z_{in}K_{cin}Z_{jn}K_{cjn}\right) - E\left(Z_{in}K_{cin}\right)E\left(Z_{jn}K_{cjn}\right)$. It follows from part (a) that

$$\sup_{i,j\in\Gamma_n}E\left(Z_{in}K_{cin}\right)E\left(Z_{jn}K_{cjn}\right) = O(b_n^{2p}).\tag{B.6}$$

We now show that

$$\sup_{i,j\in\Gamma_n}E\left(Z_{in}K_{cin}Z_{jn}K_{cjn}\right)\leq\sup_{i,j\in\Gamma_n}E\left|Z_{in}K_{cin}Z_{jn}K_{cjn}\right| = O(b_n^{2p}).\tag{B.7}$$

Following Robinson (2011), consider the sets $J_{1n}(\theta) = \left\{u,v:|b_nu|<\theta,|b_nv|<\theta\right\}$, $J_{2n}(\theta) = \{u,v:|b_nu|<\theta,|b_nv|\geq\theta\}$, and $J_{3n}(\theta) = \{u,v:|b_nu|\geq\theta,|b_nv|\geq\theta\}$ and write

$$\begin{aligned}&E\left|Z_{in}Z_{jn}K_{cin}K_{cjn}\right|\\ &\quad=b_n^{2p}\iint R_{ijn}(x-b_nu,x-b_nv)\left|K_c(u)K_c(v)\right|dudv\\ &\quad=\left(\int_{J_{1n}(\theta)}+2\int_{J_{2n}(\theta)}+\int_{J_{3n}(\theta)}\right)b_n^{2p}R_{ijn}(x-b_nu,x-b_nv)\\ &\qquad\times\left|K_c(u)K_c(v)\right|dudv.\end{aligned}$$

By Assumption 3(d), the first integral is bounded by $b_n^{2p}\beta_2(x,\theta)\left[\int|K_c(u)|du\right]^2 = O(b_n^{2p})$. By Assumption 4(b), the second integral is bounded by

$$\sup_{u\in\mathbb{R}^p}|K_c(u)|\sup_{|v|\geq\theta/b_n}|K_c(v)|E\left|Z_{in}Z_{jn}\right| = O(b_n^{\kappa}) = O(b_n^{2p}),$$

since $\kappa\geq 2p$. Similarly, we have the following bound on the third integral

$$\sup_{|u|\geq\theta/b_n}|K_c(u)|\sup_{|v|\geq\theta/b_n}|K_c(v)|E\left|Z_{in}Z_{jn}\right| = O\left(b_n^{2\kappa}\right) = O(b_n^{2p}),$$

which verifies (B.7). Thus, $\sup_{i,j\in\Gamma_n}E\Delta_{in}\Delta_{jn} = O(b_n^{2p})$.

Setting $m = b_n^{-p\delta/a(2+\delta)}$ with $a>\delta d/(2+\delta)$ gives

$$\begin{aligned}Q_{1n} &= \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i,j\in I_1}\sum E\Delta_{in}\Delta_{jn}\\ &\leq C\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\widehat{\mathbf{n}}m^db_n^{2p} = Cm^db_n^p\to 0.\end{aligned}\tag{B.8}$$

Next, we establish a bound on Q_{2n}. For any $i,j\in I_2$ with $h = \rho(i,j)>3m$,

$$\begin{aligned}E\Delta_{in}\Delta_{jn} &= E\Delta_{in}^{[h/3]}\Delta_{jn}^{[h/3]}+E\Delta_{in}\\ &\quad\times\left(\Delta_{jn}-\Delta_{jn}^{[h/3]}\right)+E\left(\Delta_{in}-\Delta_{in}^{[h/3]}\right)\Delta_{jn}^{[h/3]}\end{aligned}\tag{B.9}$$

where $[h]$ denotes the integer part of h and

$$\Delta_{in}^{[h/3]} = Z_{in}^{[h/3]}K_{in}^{[h/3]}-EZ_{in}^{[h/3]}K_{in}^{[h/3]},\tag{B.10}$$

$$Z_{in}^{[h/3]} = Y_{in}^{[h/3]}-g(x)-g'(x)^{\tau}\left(X_{in}^{[h/3]}-x\right)$$

$$K_{in}^{[h/3]} = K_c\left(\left(x-X_{in}^{[h/3]}\right)/b_n\right)$$

$$\xi_{in}^{[h/3]} = \left(Y_{in}^{[h/3]},X_{in}^{[h/3]\tau}\right)^{\tau}\quad\text{is as defined in Lemma B.2.}$$

By Lemma B.2, $\left(Y_{in}^{[h/3]},X_{in}^{[h/3]\tau}\right)^{\tau}$ has the same distribution as $\left(Y_{in},X_{in}^{\tau}\right)^{\tau}$ and satisfies the mixing condition (B.2). By the well-known strong mixing inequality, see, Ibragimov (1962, Corollary A.2), and Lemma B.2(c), we have the following bound on the first term in (B.9) for large n:

$$\begin{aligned}\left|E\Delta_{in}^{[h/3]}\Delta_{jn}^{[h/3]}\right| &\leq C\left\|\Delta_{in}^{[h/3]}\right\|_{2+\delta}\left\|\Delta_{jn}^{[h/3]}\right\|_{2+\delta}\\ &\quad\times[h/3]^{dv_*}\,\widehat{\alpha}^{\delta/(2+\delta)}(h-2[h/3])\\ &\leq C\left\|\Delta_{in}\right\|_{2+\delta}\left\|\Delta_{jn}\right\|_{2+\delta}[h/3]^{dv_*}\\ &\quad\times\widehat{\alpha}^{\delta/(2+\delta)}([h/3])\\ &\leq Cb_n^{2p/(2+\delta)}[h/3]^{dv_*}\,\widehat{\alpha}^{\delta/(2+\delta)}([h/3])\end{aligned}\tag{B.11}$$

where $\nu_* = \nu\delta/(2+\delta)$. We now bound the second term in (B.9). For some $L > 0$, let $Z_{iL} = Z_{in}\mathbf{1}\{|Z_{in}| \le L\}$ and $\widetilde{Z}_{iL} = Z_{in}\mathbf{1}\{|Z_{in}| > L\}$.

By Assumption 2(a),

$$E\widetilde{Z}_{iL}^2 \le L^{-\delta} \sup_{n,i\in\Gamma_n} E|Z_{in}|^{2+\delta}\mathbf{1}\{|Z_{in}| > L\} \le L^{-\delta}\sup_{n,i\in\Gamma_n} E|Z_{in}|^{2+\delta} = O(L^{-\delta}).$$

By Lemma B.2(b), boundedness and Lipschitz continuity of the kernel function, we have for large n

$$\begin{aligned}
&\left\|\Delta_{jn} - \Delta_{jn}^{[h/3]}\right\|_2 \\
&= \left\|Z_{jn}K_{jn} - Z_{jn}^{[h/3]}K_{jn}^{[h/3]}\right\|_2 \\
&\le \left\|K_{jn}^{[h/3]}\left(Z_{jn} - Z_{jn}^{[h/3]}\right) + Z_{jn}\left(K_{jn} - K_{jn}^{[h/3]}\right)\right\|_2 \\
&\le \left\|K_{jn}^{[h/3]}\left(Z_{jn} - Z_{jn}^{[h/3]}\right)\right\|_2 \\
&\quad + \left\|Z_{jL}\left(K_{jn} - K_{jn}^{[h/3]}\right)\right\|_2 + \left\|\widetilde{Z}_{jL}\left(K_{jn} - K_{jn}^{[h/3]}\right)\right\|_2 \\
&\le C\left\{\psi([h/3]) + Lb_n^{-1}\psi([h/3]) + L^{-\delta/2}\right\} \\
&\le C\left\{Lb_n^{-1}\psi([h/3]) + L^{-\delta/2}\right\}.
\end{aligned} \tag{B.12}$$

Setting $L = \psi^{-2/(2+\delta)}([h/3])$ in the last inequality gives

$$\left\|\Delta_{jn} - \Delta_{jn}^{[h/3]}\right\|_2 \le C\left\{b_n^{-1}\psi^{\delta/(2+\delta)}([h/3]) + \psi^{\delta/(2+\delta)}([h/3])\right\} \le Cb_n^{-1}\psi^{\delta/(2+\delta)}([h/3]).$$

Using the last inequality yields the following bound on the second term in (B.9)

$$E\Delta_{in}\left(\Delta_{jn} - \Delta_{jn}^{[h/3]}\right) \le \|\Delta_{in}\|_2\left\|\Delta_{jn} - \Delta_{jn}^{[h/3]}\right\|_2 \le Cb_n^{-1+p/2}\psi^{\delta/(2+\delta)}([h/3]).$$

Similarly,

$$E\left(\Delta_{in} - \Delta_{in}^{[h/3]}\right)\Delta_{jn}^{[h/3]} \le Cb_n^{-1+p/2}\psi^{\delta/(2+\delta)}([h/3]) \tag{B.13}$$

since $\Delta_{jn}^{[h/3]}$ and Δ_{jn} have the same marginal distribution.

Collecting inequalities (B.11)–(B.13) gives

$$\begin{aligned}
Q_{2n} \le\ & C_1\widehat{\mathbf{n}}^{-1}b_n^{-p\delta/(2+\delta)}\sum\sum_{i,j\in I_2}[\rho(i,j)/3]^{d\nu_*}\,\widehat{\alpha}^{\delta/(2+\delta)}([\rho(i,j)/3]) \\
&+ C_2\widehat{\mathbf{n}}^{-1}b_n^{-1-p/2}\sum\sum_{i,j\in I_2}\psi^{\delta/(2+\delta)}([\rho(i,j)/3]) \\
&:= \varrho_{1n} + \varrho_{2n}.
\end{aligned}$$

By Lemma A.1 of Jenish and Prucha (2009), for any $i \in \Gamma$

$$|\{j \in \Gamma : r \le \rho(i,j) < r+1\}| \le Cr^{d-1}. \tag{B.14}$$

Using this fact yields

$$\begin{aligned}
\varrho_{1n} &\le C_1\widehat{\mathbf{n}}^{-1}b_n^{-p\delta/(2+\delta)}\sum_{i\in\Gamma_n}\sum_{r=3m}^{\infty}\sum_{j\in\Gamma_n:\rho(i,j)\in[r,r+1)}[\rho(i,j)/3]^{d\nu_*} \\
&\quad\times\widehat{\alpha}^{\delta/(2+\delta)}([\rho(i,j)/3]) \\
&\le C\,\widehat{\mathbf{n}}^{-1}b_n^{-p\delta/(2+\delta)}\sum_{i\in\Gamma_n}\sum_{r=3m}^{\infty}r^{d-1}[r/3]^{d\nu_*}\,\widehat{\alpha}^{\delta/(2+\delta)}([r/3]) \\
&\le Cb_n^{-p\delta/(2+\delta)}\sum_{r=m}^{\infty}r^{d(\nu_*+1)-1}\widehat{\alpha}^{\delta/(2+\delta)}(r).
\end{aligned} \tag{B.15}$$

Since $m = b_n^{-p\delta/[a(2+\delta)]}$, $a > \delta d/(2+\delta)$, Assumption 2(c) implies:

$$\varrho_{1n} \le Cm^a\sum_{r=m}^{\infty}r^{d(\nu_*+1)-1}\widehat{\alpha}^{\delta/(2+\delta)}(r) \to 0 \quad \text{as } n\to\infty. \tag{B.16}$$

By similar arguments, we have by Assumption 2(b):

$$\begin{aligned}
\varrho_{2n} &\le Cb_n^{-1-p/2}\sum_{r=m}^{\infty}r^{d-1}\psi^{\delta/(2+\delta)}(r) \\
&\le Cb_n^{-1-p/2}\sum_{r=m}^{\infty}r^{d-\delta\gamma/(2+\delta)-1} \le Cb_n^{-1-p/2}m^{d-\delta\gamma/(2+\delta)} \\
&= Cb_n^{-1-p/2}b_n^{-p\delta(d-\delta\gamma/(2+\delta))/[a(2+\delta)]} \\
&\to 0 \quad \text{as } n\to\infty.
\end{aligned} \tag{B.17}$$

Finally, collecting (B.8), (B.16) and (B.17) gives $0 \le Q_{1n} + Q_{2n} \to 0$ as $n \to \infty$, which completes the proof of the lemma. □

Appendix C. Proofs for Section 3

C.1. Proof of Lemma 1

The lemma will follow if we show that

$$E\|U_n - EU_n\| \to 0 \quad \text{and} \quad EU_n \to U. \tag{C.1}$$

To prove the first result, it suffices to show that for any $c = (c_0, c_1^\tau)^\tau \in \mathbb{R}^{p+1}$,

$$E|c^\tau U_n - Ec^\tau U_n| \to 0. \tag{C.2}$$

Note that $c^\tau U_n$ is the $(p+1)$-vector with the k-th element given by

$$(c^\tau U_n)_k = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\left(\frac{X_{in}-x}{b_n}\right)_k K_c\left(\frac{X_{in}-x}{b_n}\right), \quad k = 0, 1, \ldots, p$$

with $K_c(u)$ defined in Assumption 4. To prove (C.2), we show that for each $k = 0, 1, \ldots, p$, $E[(c^\tau U_n)_k - E(c^\tau U_n)_k]^2 \to 0$. Let

$$\begin{aligned}
\chi_{in} &= b_n^{-1}(X_{in}-x)_k K_{cin} - b_n^{-1}E(X_{in}-x)_k K_{cin}, \\
K_{cin} &= K_c((x - X_{in})/b_n).
\end{aligned}$$

Then,

$$\begin{aligned}
E[(c^\tau U_n)_k - E(c^\tau U_n)_k]^2 &= \left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}E\chi_{in}^2 + \left(\widehat{\mathbf{n}}b_n^p\right)^{-2} \\
&\quad\times\sum_{i\in\Gamma_n}\sum_{i\neq j\in\Gamma_n}E\chi_{in}\chi_{jn} \\
&:= J_{1n} + J_{2n}.
\end{aligned}$$

By Lemma B.3(a), $\sup_{i\in\Gamma_n}E\chi_{in}^2 = O\left(b_n^p\right)$ and hence

$$J_{1n} = \left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}E\chi_{in}^2 \le C\left(\widehat{\mathbf{n}}b_n^p\right)^{-1} \to 0.$$

Furthermore, Lemma B.3(c) with $Z_{in} := (X_{in}-x)_k/b_n$ gives

$$\lim_{n\to\infty}J_{2n} = \lim_{n\to\infty}\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}\sum_{j\neq i,j\in\Gamma_n}E\chi_{in}\chi_{jn} = 0,$$

which proves the first statement in (C.1).

We now prove the second result in (C.1). We will show that for all k, $l = 0, 1, 2, \ldots, p$, $E(U_n)_{kl} \to \overline{f}(x)\int u_k u_l K(u)du$. Using arguments similar to those in the proof of Lemma B.3(b), we have by Assumptions 3–4

$$E\,(U_n)_{kl} = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\int f_{in}(u)\left((u-x)/b_n\right)_k \times\left((u-x)/b_n\right)_l K\left((x-u)/b_n\right)du$$
$$= \int\left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x-b_n u)\right\}u_k u_l K(u)\,du$$
$$\to \bar{f}(x)\int u_k u_l K(u)du.$$

Thus,

$$EU_n \to U = \bar{f}(x)\begin{pmatrix}\int K(u)du & \int u^\tau K(u)du\\ \int uK(u)du & \int uu^\tau K(u)du\end{pmatrix}$$
$$= \bar{f}(x)\begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \int uu^\tau K(u)du\end{pmatrix}$$

since $K(u)$ is a symmetric density function. □

C.2. Proof of Lemma 2

To prove the lemma, it suffices to show that for any $c = \left(c_0, c_1^\tau\right)^\tau \in \mathbb{R}^{p+1}$,

$$\mathrm{Var}\,(A_n) \to \sigma^2 = c^\tau\Sigma c = \overline{\omega}(x)\int K_c^2(u)du$$

where

$$A_n = \left(\widehat{\mathbf{n}}b_n^p\right)^{1/2} c^\tau W_n = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1/2}\sum_{i\in\Gamma_n} Z_{in}K_c\left(\frac{X_{in}-x}{b_n}\right).$$

Write

$$\mathrm{Var}(A_n) = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n} E\Delta_{in}^2 + \left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\sum_{j\neq i, j\in\Gamma_n} E\Delta_{in}\Delta_{jn} \quad (C.3)$$

where $\Delta_{in} = Z_{in}K_{in} - EZ_{in}K_{in}$. By Lemma B.3(c), the second term in (C.3) converges to zero. Then, by Lemma B.3(b)

$$\lim_{n\to\infty}\mathrm{Var}(A_n) = \lim_{n\to\infty}\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n} E\Delta_{in}^2$$
$$= \overline{\omega}(x)\int_{\mathbb{R}^p} K_c^2(u)du = c^\tau\Sigma c,$$

where

$$\Sigma = \overline{\omega}(x)\begin{pmatrix}\int K^2(u)du & \int u^\tau K^2(u)du\\ \int uK^2(u)du & \int uu^\tau K^2(u)du\end{pmatrix}$$
$$= \overline{\omega}(x)\begin{pmatrix}\int K^2(u)du & \mathbf{0}\\ \mathbf{0} & \int uu^\tau K^2(u)du\end{pmatrix}$$

by symmetry of the kernel function. □

C.3. Proof of Lemma 3

The lemma will follow if we show that for any $c = \left(c_0, c_1^\tau\right)^\tau \in \mathbb{R}^{p+1}$,

$$E\left(c'W_n\right) = b_n^2\frac{1}{2}\bar{f}(x)\mathrm{tr}\left[g''(x)\int uu^\tau K_c^2(u)du\right] + o(b_n^2)$$
$$= b_n^2 c_0\bar{f}(x)B_g(x) + o(b_n^2).$$

A second-order Taylor expansion of $g(u)$ about x gives

$$E(Y_{in}|X_{in}=u) = g(x) + g'(x)^\tau(u-x) + \frac{1}{2}(u-x)^\tau g'' \times(x+\vartheta(u-x))(u-x)$$

for some scalar $|\vartheta| < 1$, where u is in an open neighborhood of x. Then,

$$EZ_{in}K_c\left((x-X_{in})/b_n\right)$$
$$= \int f_{in}(u)\left[g(u)-g(x)-g'(x)^\tau(u-x)\right]K_c\left((x-u)/b_n\right)du$$
$$= \frac{1}{2}\int f_{in}(u)(u-x)^\tau g''(x+\vartheta(u-x))(u-x) \times K_c\left((x-u)/b_n\right)du$$
$$= \frac{b_n^p b_n^2}{2}\int f_{in}(x-ub_n)u^\tau g''(x-\vartheta ub_n)uK_c(u)\,du$$

and hence

$$E\left(c^\tau W_n\right) = \frac{b_n^2}{2}\int\left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x-ub_n)\right\} \times u^\tau g''(x-\vartheta ub_n)uK_c(u)\,du$$
$$= \frac{b_n^2}{2}\mathrm{tr}\left[\int\left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x-ub_n)\right\} \times g''(x-\vartheta b_n u)uu^\tau K_c(u)\,du\right].$$

Since $g''(u)$ and $f_{in}(u)$ are continuous at x uniformly over $i\in\Gamma_n$, $n\in\mathbb{N}$, we have by Assumption 3(a)–(b)

$$\left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x-ub_n)\right\}g''(x-\vartheta b_n u) = \bar{f}(x)g''(x) + o(1).$$

Hence,

$$E\left(c'W_n\right) = \frac{b_n^2}{2}\bar{f}(x)\mathrm{tr}\left[g''(x)\int uu^\tau K_c(u)\,du\right] + o(b_n^2)$$
$$= b_n^2\left[c_0B_0(x) + c_1^\tau B_1(x)\right] + o(b_n^2)$$

where $B_0(x) = \frac{1}{2}\bar{f}(x)\sum_{k=1}^p\sum_{l=1}^p[g_{kl}(x)\int u_k u_l K(u)du]$, $g_{kl}(x) = \partial^2 g(x)/\partial x_k\partial x_l$, $k, l = 1,\ldots,p$, and $B_1(x) = \frac{1}{2}\bar{f}(x)\sum_{k=1}^p\sum_{l=1}^p[g_{kl}(x)\int u_k u_l uK(u)du]$.

Since $K(u)$ is a symmetric density function,

$$B_0(x) = \frac{1}{2}\bar{f}(x)\sum_{k=1}^p g_{kk}(x)\int u_k^2K(u)du, \quad \text{and} \quad B_1(x) = \mathbf{0}. \quad □$$

C.4. Proof of Theorem 1

We will employ the Cramer–Wold device. For any $c = \left(c_0, c_1^\tau\right)^\tau \in \mathbb{R}^{p+1}$, define

$$A_n = \left(\widehat{\mathbf{n}}b_n^p\right)^{1/2} c^\tau W_n = \left(\widehat{\mathbf{n}}b_n^p\right)^{-1/2}\sum_{i\in\Gamma_n} Z_{in}K_c\left(\frac{X_{in}-x}{b_n}\right).$$

In light of Lemma 1, by the Slutsky theorem, it suffices to show that for any $c = \left(c_0, c_1^\tau\right)^\tau \in \mathbb{R}^{p+1}$

$$A_n - EA_n \Longrightarrow N(0,\sigma^2) \quad (C.4)$$
$$\mathrm{Var}\,(A_n) \to \sigma^2 = c^\tau\Sigma c. \quad (C.5)$$

Then, (C.4)–(C.5), Lemma 3 and symmetry of the kernel function will imply the conclusion of the theorem. Result (C.5) follows immediately from Lemma 2. So, it remains to prove only (C.4). To this end, we will use the dependent Lindeberg CLT of Bardet et al. (2008, Theorem 1).

We re-order the sample in the following way. Let $\phi : [1, \widehat{\mathbf{n}}] \cap \mathbb{N} \to \Gamma_n$ be an one-to-one map from $[1, \widehat{\mathbf{n}}] \cap \mathbb{N}$ to the sample region Γ_n, where $\widehat{\mathbf{n}} = |\Gamma_n|$ is the sample size, such that for all $1 \le k < l \le \widehat{\mathbf{n}}$: $\phi(k) <_{lex} \phi(l)$, where the lexicographic order is defined as follows: if $(i_1, .., i_d)$ and $(j_1, .., j_d)$ are distinct elements of Γ, then $i <_{lex} j$ means that either $i_1 < j_1$ or for some $r \in \{2, 3, .., d\}$, $i_r < j_r$ and $i_q = j_q$ for $1 \le q \le r$. Existence and uniqueness of such mapping is evident. Using this mapping and the data-generating process $\{\xi_{in}\}$, define

$$S_n = A_n - EA_n = \sum_{i \in \Gamma_n} \varsigma_{in} = \sum_{l=1}^{\widehat{\mathbf{n}}} \varsigma_{\phi(l),n}$$

where

$$\varsigma_{in} = \varsigma(\xi_{in}) = \left(\widehat{\mathbf{n}} b_n^p\right)^{-1/2} \Delta_{in} = \left(\widehat{\mathbf{n}} b_n^p\right)^{-1/2} \times \{Z_{in} K_{in} - E Z_{in} K_{in}\}.$$

Also, define

$$S_{k,n} = \sum_{l=1}^{k} \varsigma_{\phi(l),n}, \quad k = 1, \ldots, \widehat{\mathbf{n}}, \qquad S_{0,n} = 0.$$

To invoke Theorem 1 of Bardet et al. (2008), we need to verify that

A1. $E(S_n)^2 \to \sigma^2$, as $n \to \infty$;

A2. $M_n = \sum_{k=1}^{\widehat{\mathbf{n}}} E\left|\varsigma_{\phi(k),n}\right|^{2+\delta} \to 0$ as $n \to \infty$;

A3. $T_n = \sum_{k=1}^{\widehat{\mathbf{n}}} \left|\mathrm{Cov}\left[\exp\left(\mathbf{i} t S_{k-1,n}\right); \exp\left(\mathbf{i} t \varsigma_{\phi(k),n}\right)\right]\right| \to 0$, as $n \to \infty$ for any $t \in \mathbb{R}$, where $\mathbf{i} = \sqrt{-1}$.

Part A1 follows from Lemma 2. Furthermore, by Lemma B.3(a)

$$M_n = \left(\widehat{\mathbf{n}} b_n^p\right)^{-(2+\delta)/2} \sum_{k=1}^{\widehat{\mathbf{n}}} E\left|Z_{in} K_{in} - E Z_{in} K_{in}\right|^{2+\delta} \le C \left(\widehat{\mathbf{n}} b_n^p\right)^{-(2+\delta)/2} \sum_{k=1}^{\widehat{\mathbf{n}}} E\left|Z_{in} K_{in}\right|^{2+\delta} \le C \left(\widehat{\mathbf{n}} b_n^p\right)^{-\delta/2} \to 0 \quad \text{as } n \to \infty$$

which verifies A2. To prove A3, note that $\exp\left(\mathbf{i} t S_{k-1,n}\right) = \sum_{l=1}^{k-1} \left[\exp\left(\mathbf{i} t S_{l,n}\right) - \exp\left(\mathbf{i} t S_{l-1,n}\right)\right]$, and consequently,

$$\mathrm{Cov}\left(\exp\left(\mathbf{i} t S_{k-1,n}\right); \exp\left(\mathbf{i} t \varsigma_{\phi(k),n}\right)\right) = \sum_{l=1}^{k-1} \mathrm{Cov}\left(\exp\left(\mathbf{i} t S_{l,n}\right) - \exp\left(\mathbf{i} t S_{l-1,n}\right); \exp\left(\mathbf{i} t \varsigma_{\phi(k),n}\right)\right).$$

Let $\left\{\varepsilon_{in}^*\right\}$ be an independent copy of $\{\varepsilon_{in}\}$. Define $\xi_{in}^* = \left(Y_{in}^*, X_{in}^{*\tau}\right)^\tau = H_{in}\left(\left(\varepsilon_{jn}^*\right)_{j \in \Gamma}\right)$ so that $\varsigma_{in}^* = \varsigma\left(\xi_{in}^*\right)$ is an independent copy of $\{\varsigma_{in}\}$. Then,

$$\begin{aligned} &\left|\mathrm{Cov}\left[\exp\left(\mathbf{i} t S_{l,n}\right) - \exp\left(\mathbf{i} t S_{l-1,n}\right); \exp\left(\mathbf{i} t \varsigma_{\phi(k),n}\right)\right]\right| \\ &= \left|E\left(\left[\exp\left(\mathbf{i} t S_{l,n}\right) - \exp\left(\mathbf{i} t S_{l-1,n}\right)\right] \times \left[\exp\left(\mathbf{i} t \varsigma_{\phi(k),n}\right) - \exp\left(\mathbf{i} t \varsigma_{\phi(k),n}^*\right)\right]\right)\right| \\ &\le t^2 E\left|\varsigma_{\phi(l),n}\right| \left|\varsigma_{\phi(k),n} - \varsigma_{\phi(k),n}^*\right| \end{aligned}$$

where the last line follows from the elementary inequality $\left|e^{\mathbf{i}a} - e^{\mathbf{i}b}\right| \le |a - b|$. Thus,

$$0 \le T_n \le t^2 \sum_{k=1}^{\widehat{\mathbf{n}}} \sum_{l=1}^{k-1} \left|\varsigma_{\phi(l),n}\right| \left|\varsigma_{\phi(k),n} - \varsigma_{\phi(k),n}^*\right| \le t^2 \left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \sum_{i \in \Gamma_n} \sum_{j \ne i, j \in \Gamma_n} E\left|\Delta_{in}\right| \left|\Delta_{jn} - \Delta_{jn}^*\right|,$$

and hence to prove A3, it suffices to show convergence of the right hand side of the last inequality to zero.

The proof the latter relies on the same argument as the proof of Lemma B.3(c). Decompose

$$\begin{aligned} &\left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \sum_{i \in \Gamma_n} \sum_{j \ne i, j \in \Gamma_n} E\left|\Delta_{in}\right| \left|\Delta_{jn} - \Delta_{jn}^*\right| \\ &= \left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \sum\sum_{i,j \in I_1} E\left|\Delta_{in}\right| \left|\Delta_{jn} - \Delta_{jn}^*\right| + \left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \sum\sum_{i,j \in I_2} E\left|\Delta_{in}\right| \left|\Delta_{jn} - \Delta_{jn}^*\right| \end{aligned}$$

where the sets I_1 and I_2 are as in (B.5). Observe that $E\left|\Delta_{in}\right| \left|\Delta_{jn} - \Delta_{jn}^*\right| \le E\left|\Delta_{in}\right| \left|\Delta_{jn}\right| + E\left|\Delta_{in}\right| E\left|\Delta_{jn}^*\right|$.

It follows from Lemma B.3(a) that $\sup_{i,j \in \Gamma_n} E\left|\Delta_{in}\right| E\left|\Delta_{jn}^*\right| = O(b_n^{2p})$. As established in the proof of Lemma B.3(b), see (B.6)–(B.7), $\sup_{i,j \in \Gamma_n} E\left|\Delta_{in}\right| \left|\Delta_{jn}\right| = O(b_n^{2p})$. Setting, as before, $m = b_n^{-p\delta/a(2+\delta)}$ with $a > \delta d/(2 + \delta)$ gives

$$\left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \sum\sum_{i,j \in I_1} E\left|\Delta_{in}\right| \left|\Delta_{jn} - \Delta_{jn}^*\right| \le C \left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \widehat{\mathbf{n}} m^d b_n^{2p} = C m^d b_n^p \to 0.$$

To bound the second sum, let $\zeta_{jn} = \Delta_{jn} - \Delta_{jn}^* = Z_{jn} K_{jn} - Z_{jn}^* K_{jn}^*$ and use decomposition as in (B.9)–(B.10) for any $i, j \in I_2$ with $h = \rho(i, j) > 3m$:

$$E\left|\Delta_{in}\right| \left|\zeta_{jn}\right| = E\left|\Delta_{in}^{[h/3]}\right| \zeta_{jn}^{[h/3]} + E\left(\left|\Delta_{in}\right| - \left|\Delta_{in}^{[h/3]}\right|\right) \left|\zeta_{jn}\right| - E\left|\Delta_{in}^{[h/3]}\right| \left(\zeta_{jn}^{[h/3]} - \left|\zeta_{jn}\right|\right).$$

Note that $E\zeta_{jn} = E\zeta_{jn}^{[h/3]} = 0$ so that the first term in the last expression can be bounded as

$$\begin{aligned} E\left|\Delta_{in}^{[h/3]}\right| \zeta_{jn}^{[h/3]} &\le C \left\|\Delta_{in}^{[h/3]}\right\|_{2+\delta} \left\|\zeta_{jn}^{[h/3]}\right\|_{2+\delta} \times [h/3]^{d\nu_*} \widehat{\alpha}^{\delta/(2+\delta)}(h - 2[h/3]) \\ &\le C b_n^{2p/(2+\delta)} [h/3]^{d\nu_*} \widehat{\alpha}^{\delta/(2+\delta)}([h/3]) \end{aligned}$$

where $\nu_* = \nu\delta/(2+\delta)$. By the same arguments as in (B.12)–(B.13), the second and third terms can be bounded as:

$$\begin{aligned} E\left(\left|\Delta_{in}\right| - \left|\Delta_{in}^{[h/3]}\right|\right) \left|\zeta_{jn}\right| &\le \left\|\left|\Delta_{in}\right| - \left|\Delta_{in}^{[h/3]}\right|\right\|_2 \left\|\zeta_{jn}\right\|_2 \\ &\le 2 \left\|\Delta_{in} - \Delta_{in}^{[h/3]}\right\|_2 \left\|Z_{jn} K_{jn}\right\|_2 \\ &\le C b_n^{-1+p/2} \psi^{\delta/(2+\delta)}([h/3]) \end{aligned}$$

and

$$\begin{aligned} E\left|\Delta_{in}^{[h/3]}\right| \left(\zeta_{jn}^{[h/3]} - \left|\zeta_{jn}\right|\right) &\le \left\|\Delta_{in}^{[h/3]}\right\|_2 \left\|\zeta_{jn}^{[h/3]} - \left|\zeta_{jn}\right|\right\|_2 \\ &\le \left\|\Delta_{in}\right\|_2 \left\|\zeta_{jn}^{[h/3]} - \zeta_{jn}\right\|_2 \\ &\le C b_n^{-1+p/2} \psi^{\delta/(2+\delta)}([h/3]). \end{aligned}$$

Now, using steps (B.15)–(B.17) of the proof of Lemma B.3(c) completes the proof of the theorem. □

C.5. Proof of Theorem 2

Step 1. *Consistency of* $\widehat{f}_n(x)$. It suffices to show that

$$p \lim_{n \to \infty} \left|\widehat{f}_n(x) - E\widehat{f}_n(x)\right| = 0 \quad \text{and} \quad \lim_{n \to \infty} E\widehat{f}_n(x) = \bar{f}(x).$$

First, we show $\lim_{n \to \infty} E\widehat{f}_n(x) = \bar{f}(x)$. Using arguments similar to those in the proof of Lemma 1, we have by Assumption 3(b)

$$\lim_{n \to \infty} E\widehat{f}_n(x) = \lim_{n \to \infty} \int \left[\frac{1}{\widehat{\mathbf{n}}} \sum_{i \in \Gamma_n} f_{in}(x - u b_n)\right] K(u)\, du = \bar{f}(x).$$

We next show that

$$p\lim_{n\to\infty}\left|\widehat{f}_n(x)-E\widehat{f}_n(x)\right|=0. \tag{C.6}$$

Let $K_{in}=K\left((x-X_{in})/b_n\right)$. Then, $\widehat{f}_n(x)-E\widehat{f}_n(x)=\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}\varkappa_{in}$, with $\varkappa_{in}=K_{in}-EK_{in}$. Note that

$$E\left(\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}\varkappa_{in}\right)^2=\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}E\varkappa_{in}^2+\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\times\sum_{i\in\Gamma_n}\sum_{i\neq j\in\Gamma_n}E\varkappa_{in}\varkappa_{jn}:=P_{1n}+P_{2n}.$$

Since $\sup_{i\in\Gamma_n}E\varkappa_{in}^2=O\left(b_n^p\right)$, we have $P_{1n}=\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}E\varkappa_{in}^2\leq C\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\to 0$. Setting $Z_{in}:=1$ in Lemma B.3(c) gives

$$P_{2n}=\lim_{n\to\infty}\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\sum_{j\neq i,j\in\Gamma_n}E\varkappa_{in}\varkappa_{jn}=0,$$

which proves (C.6).

Step 2. *Consistency of* $\widehat{\omega}_n(x)$. Let

$$\omega_n(x)=\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}e_{in}^2(x)K\left((x-X_{in})/b_n\right)$$

with $e_{in}(x)=Y_{in}-g(x)$.

Consistency of $\widehat{\omega}_n(x)$ is implied by

$$p\lim_{n\to\infty}\left|\widehat{\omega}_n(x)-\omega_n(x)\right|=0, \tag{C.7}$$

$$p\lim_{n\to\infty}\left|\omega_n(x)-E\omega_n(x)\right|=0 \tag{C.8}$$

$$\lim_{n\to\infty}E\omega_n(x)=\overline{\omega}(x). \tag{C.9}$$

Proof of (C.9) is similar to that of Lemma B.3(b), and is therefore omitted. Note that $\widehat{e}_{in}^2(x)-e_{in}^2(x)=2\left[g(x)-g_n(x)\right]Y_{in}+\left[g_n^2(x)-g^2(x)\right]$.
Then,

$$\left|\widehat{\omega}_n(x)-\omega_n(x)\right|\leq\left|g_n^2(x)-g^2(x)\right|\left|\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}K_{in}\right|+2\left|g(x)-g_n(x)\right|\left|\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}Y_{in}K_{in}\right|.$$

By consistency of $g_n(x)$, $|g(x)-g_n(x)|=o_p(1)$ and $\left|g_n^2(x)-g^2(x)\right|=o_p(1)$. Moreover, by Lemma B.3(a) with $q=1$

$$\left|\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}Y_{in}K_{in}\right|=O_p(1)\quad\text{and}\quad\left|\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}K_{in}\right|=O_p(1)$$

which verifies (C.7). We now prove (C.8). For $L>0$, let

$$\phi(x)=\begin{cases}x, & \text{if } |x|\leq L\\ L, & \text{if } x>L\\ -L, & \text{if } x<-L.\end{cases}$$

Note that $\phi(x)$ is bounded, and satisfies the Lipschitz condition: $|\phi(x_1)-\phi(x_2)|\leq\min\{2L,|x_1-x_2|\}$. Then, by Theorem 17.3 of Davidson (1994, pp. 270–271), $\phi\left(e_{in}^2\right)$ is L_2 NED with the NED coefficients $C\psi(m)$. Now, define

$$e_{in,L}^2=\phi\left(e_{in}^2\right),\ \widetilde{e}_{in,L}^2=e_{in}^2-e_{in,L}^2$$

and $\quad\varsigma_{in}=e_{in,L}^2K_{in}-Ee_{in,L}^2K_{in},\qquad\widetilde{\varsigma}_{in}=\widetilde{e}_{in,L}^2K_{in}-E\widetilde{e}_{in,L}^2K_{in}.$

Then, $\omega_n(x)-E\omega_n(x)=\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}\varsigma_{in}+\frac{1}{\widehat{\mathbf{n}}b_n^p}\sum_{i\in\Gamma_n}\widetilde{\varsigma}_{in}:=S_{1n}+S_{2n}$.

We need to show that $p\lim_{L\to\infty}p\lim_{n\to\infty}|S_{1n}|=0$ and $p\lim_{L\to\infty}p\lim_{n\to\infty}|S_{2n}|=0$. We first prove that $p\lim_{L\to\infty}p\lim_{n\to\infty}|S_{2n}|=0$. Note that

$$\left|\widetilde{e}_{in,L}^2\right|=\begin{cases}0, & \text{if } e_{in}^2\leq L\\ e_{in}^2-L, & \text{if } e_{in}^2>L\end{cases}\leq e_{in}^2\mathbf{1}\left\{\left|e_{in}^2\right|>L\right\}$$

$$E\left|\widetilde{\varsigma}_{in}\right|\leq 2E\left|\widetilde{e}_{in,L}^2\right|\left|K_{in}\right|\leq 2E\left|e_{in}^2\mathbf{1}\left\{\left|e_{in}^2\right|>L\right\}\right|\times\left|K_{in}\right|\leq 2L^{-\delta}E\left|e_{in}\right|^{2+\delta}\left|K_{in}\right|.$$

Furthermore, by Lemma B.3(a), $\sup_{i\in\Gamma_n}E\left|e_{in}\right|^{2+\delta}\left|K_{in}\right|=O(b_n^p)$. Then,

$$E\left|S_{2n}\right|\leq\frac{2}{\widehat{\mathbf{n}}b_n^p}L^{-\delta}\sum_{i\in\Gamma_n}E\left|e_{in}\right|^{2+\delta}\left|K_{in}\right|\leq CL^{-\delta}\to 0$$

and hence $p\lim_{L\to\infty}p\lim_{n\to\infty}|S_{2n}|=0$.

We now show that for each fixed $L>0$, $p\lim_{n\to\infty}|S_{1n}|=0$ and hence $p\lim_{L\to\infty}p\lim_{n\to\infty}|S_{1n}|=0$. Write

$$ES_{1n}^2=\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}E\varsigma_{in}^2+\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{j\neq i,j\in\Gamma_n}E\varsigma_{in}\varsigma_{jn}.$$

Since $\sup_{i\in\Gamma_n}E\varsigma_{in}^2=O\left(b_n^p\right)$, $0\leq\left(\widehat{\mathbf{n}}b_n^p\right)^{-2}\sum_{i\in\Gamma_n}E\varsigma_{in}^2\leq C\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\to 0$.

Finally, note that $e_{in,L}^2=\phi\left(e_{in}\right)$ satisfies the same conditions as Z_{in}: it has the requisite moments, and is L_2 NED of the same size as Z_{in}. Thus, ς_{in} satisfies the same conditions as Δ_{in} and the conclusion of Lemma B.3(b) applies, i.e.,

$$0\leq\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{j\neq i,j\in\Gamma_n}E\varsigma_{in}\varsigma_{jn}\to 0.\quad\square$$

Appendix D. Proofs for Section 4

The proof of Theorem 3 uses the following lemmata.

Lemma D.1. *Under assumptions of Theorem 3,*

(a) *For each $k=0,1,\ldots,p$ and $(V_n)_k$ defined in (1),*

$$\sup_{x\in\mathbb{R}^p}\left|(V_n(x))_k-E\left(V_n(x)\right)_k\right|=O_p(\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)}b_n^{-p-d/(2\gamma+d)})$$

provided that the right-hand side is $o_p(1)$.

(b) *For each $k,l=0,1,\ldots,p$ and $(U_n)_{kl}$ defined in (2),*

$$\sup_{x\in\mathbb{R}^p}\left|(U_n(x))_{kl}-E\left(U_n(x)\right)_{kl}\right|=O_p(\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)}b_n^{-p-d/(2\gamma+d)})$$

provided that the right-hand side is $o_p(1)$.

Lemma D.2. *Under assumptions of Theorem 3, for $k=0,1,\ldots,p$ and $(W_n)_k$ defined in (4), $\sup_{x\in\mathbb{R}^p}\left|E\left(W_n(x)\right)_k\right|=O\left(b_n^2\right)$.*

Lemma D.3. *Under assumptions of Theorem 3, $\sup_{x\in\mathbf{X}}\left\|U_n^{-1}(x)\right\|=O_p(1)$.*

Proof of Theorem 3. Given that

$$\begin{pmatrix}g_n(x)-g(x)\\ \left[g_n'(x)-g'(x)\right]b_n\end{pmatrix}=U_n^{-1}(x)W_n(x),$$

the theorem will follow from Lemmata D.2–D.3 if we show that for each $k=0,1,\ldots,p$, $\sup_{x\in\mathbf{X}}\left|(W_n(x))_k-E(W_n(x))_k\right|=O_p(\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)}b_n^{-p-d/(2\gamma+d)})$. By the definition of $(W_n(x))_k$,

$$\sup_{x\in\mathbf{X}}\left|(W_n(x))_k-E\left(W_n(x)\right)_k\right|$$

$$\leq\sup_{x\in\mathbf{X}}\left|\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}Y_{in}\left(\frac{X_{in}-x}{b_n}\right)_kK\left(\frac{X_{in}-x}{b_n}\right)\right.$$

$$- EY_{in}\left(\frac{X_{in}-x}{b_n}\right)_k K\left(\frac{X_{in}-x}{b_n}\right)\Bigg|$$

$$+ \sup_{x\in\mathbf{X}}|g(x)|\sup_{x\in\mathbf{X}}\Bigg|\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\left(\frac{X_{in}-x}{b_n}\right)_k K\left(\frac{X_{in}-x}{b_n}\right)$$

$$- E\left(\frac{X_{in}-x}{b_n}\right)_k K\left(\frac{X_{in}-x}{b_n}\right)\Bigg| + \sup_{x\in\mathbf{X}}\left\|g'(x)\right\|\sup_{x\in\mathbf{X}}\Bigg\|\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}$$

$$\times\sum_{i\in\Gamma_n}(X_{in}-x)\left(\frac{X_{in}-x}{b_n}\right)_k K\left(\frac{X_{in}-x}{b_n}\right)$$

$$- E\,(X_{in}-x)\left(\frac{X_{in}-x}{b_n}\right)_k K\left(\frac{X_{in}-x}{b_n}\right)\Bigg\|.$$

Thus, it suffices to show that each of the three terms on the r.h.s. of the last inequality is $O_p(\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)}b_n^{-p-d/(2\gamma+d)})$. The desired result for the first term follows immediately from Lemma D.1(a), for the second and third terms—from part (b) Lemma D.1, observing that by compactness of $\mathbf{X}$ and continuity of $g(x)$ and $g'(x)$, $\sup_{x\in\mathbf{X}}|g(x)|<\infty$ and $\sup_{x\in\mathbf{X}}\left\|g'(x)\right\|<\infty$. □

Proof of Lemma D.1. The proof of the lemma relies on the method of Parzen (1962) which involves using Fourier transforms of the functions $\widetilde{K}_{kl}(x)$, and is similar to those of Bierens (1983) and Andrews (1995).

Part (a). Using the representation $\widetilde{K}_{k0}(x) = \int_{\mathbb{R}^p}\exp(-\mathbf{i}r^\tau x)\Psi_{k0}(r)dr$ gives

$$\sup_{x\in\mathbb{R}^p}|(V_n(x))_k - E\,(V_n(x))_k|$$

$$= \sup_{x\in\mathbb{R}^p}\Bigg|\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\Bigg[Y_{in}\int\exp\left(-\mathbf{i}r^\tau\left(X_{in}-x\right)/b_n\right)\Psi_{k0}(r)dr$$

$$-EY_{in}\int\exp(-\mathbf{i}r^\tau\left(X_{in}-x\right)/b_n)\Psi_{k0}(r)dr\Bigg]\Bigg|$$

$$\le \int\Bigg|\left(\widehat{\mathbf{n}}b_n^p\right)^{-1}\sum_{i\in\Gamma_n}\Bigg[Y_{in}\exp(-\mathbf{i}r^\tau X_{in}/b_n)$$

$$- EY_{in}\exp(-\mathbf{i}r^\tau X_{in}/b_n)\Bigg]\Bigg|\sup_{x\in\mathbb{R}^p}|\exp(\mathbf{i}r^\tau x/b_n)|\,|\Psi_{k0}(r)|\,dr$$

$$\le \int\Bigg|\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\Bigg[Y_{in}\exp(-\mathbf{i}r^\tau X_{in})$$

$$- EY_{in}\exp(-\mathbf{i}r^\tau X_{in})\Bigg]\Bigg|\,|\Psi_{k0}(rb_n)|\,dr.$$

Let $X_{in}^{(m)} = E\,(X_{in}|\mathfrak{F}_{in}(m))$ and note that for any $m\in\mathbb{N}$

$$Y_{in}\cos\left(r^\tau X_{in}\right) - EY_{in}\cos\left(r^\tau X_{in}\right)$$

$$= \left[Y_{in}\cos\left(r^\tau X_{in}\right) - Y_{in}^{(m)}\cos\left(r^\tau X_{in}\right)\right]$$

$$+ \left[Y_{in}^{(m)}\cos\left(r^\tau X_{in}\right) - Y_{in}^{(m)}\cos\left(r^\tau X_{in}^{(m)}\right)\right]$$

$$+ \left[Y_{in}^{(m)}\cos(r^\tau X_{in}^{(m)}) - EY_{in}^{(m)}\cos(r^\tau X_{in}^{(m)})\right]$$

$$+ \left[EY_{in}^{(m)}\cos(r^\tau X_{in}^{(m)}) - EY_{in}^{(m)}\cos(r^\tau X_{in})\right]$$

$$+ \left[EY_{in}^{(m)}\cos(r^\tau X_{in}) - EY_{in}\cos(r^\tau X_{in})\right].$$

Using this decomposition, we have

$$E\left|\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\left[Y_{in}\cos\left(r^\tau X_{in}\right) - EY_{in}\cos\left(r^\tau X_{in}\right)\right]\right|$$

$$\le 2\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}E\left|Y_{in}-Y_{in}^{(m)}\right|\left|\cos\left(r^\tau X_{in}\right)\right| + 2\widehat{\mathbf{n}}^{-1}$$

$$\times\sum_{i\in\Gamma_n}E\left|Y_{in}^{(m)}\right|\left|\cos\left(r^\tau X_{in}\right) - \cos\left(r^\tau X_{in}^{(m)}\right)\right|$$

$$+ \mathrm{Var}^{1/2}\left[\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}Y_{in}^{(m)}\cos\left(r^\tau X_{in}^{(m)}\right)\right].$$

By the Cauchy–Schwarz inequality,

$$\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}E\left|Y_{in}-Y_{in}^{(m)}\right|\left|\cos(r^\tau X_{in})\right|$$

$$\le \sup_{n,i\in\Gamma_n}\left\|Y_{in}-Y_{in}^{(m)}\right\|_2 \le \psi(m) \tag{D.1}$$

and

$$\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}E\left|Y_{in}^{(m)}\right|\left|\cos\left(r^\tau X_{in}\right) - \cos\left(r^\tau X_{in}^{(m)}\right)\right|$$

$$\le \sup_{n,i\in\Gamma_n}\|Y_{in}\|_2\left\|r^\tau\left(X_{in}-X_{in}^{(m)}\right)\sin\widetilde{x}\right\|_2 \le C\,\|r\|\,\psi(m) \tag{D.2}$$

where $\widetilde{x}$ lies between $r^\tau X_{in}$ and $r^\tau X_{in}^{(m)}$. Since the m-neighborhood of any point on Γ contains at most Cm^d points of Γ for some constant C that does not depend on m and n, see Lemma A.1 of Jenish and Prucha (2009),

$$\alpha_m(1,1,\rho(i,j)) \le \begin{cases}1, & \rho(i,j)\le 2m\\ Cm^{d\nu_*}\widehat{\alpha}(\rho(i,j)-2m), & \rho(i,j)>2m\end{cases}$$

where $\nu_* = \nu\delta/(2+\delta)$ and $\widehat{\alpha}(r)$ is the mixing coefficient of the input process ε.

By the strong mixing inequality, Ibragimov (1962),

$$\left|\mathrm{Cov}\left(Y_{in}^{(m)}\cos(r^\tau X_{in}^{(m)});\,X_{jn}^{(m)}\cos(r^\tau X_{jn}^{(m)})\right)\right|$$

$$\le 8\sup_{n,i\in\Gamma_n}\|Y_{in}\|_{2+\delta}^2\,\alpha_m^{\delta/((2+\delta))}(1,1,\rho(i,j)).$$

Using the last inequality and (B.14) in the proof of Lemma B.3 gives

$$\mathrm{Var}\left[\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}Y_{in}^{(m)}\cos(r^\tau X_{in}^{(m)})\right]$$

$$= \widehat{\mathbf{n}}^{-2}\sum_{i,j\in\Gamma_n}\mathrm{Cov}\left(Y_{in}^{(m)}\cos(r^\tau X_{in}^{(m)});\,Y_{jn}^{(m)}\cos(r^\tau X_{jn}^{(m)})\right)$$

$$\le C\,\widehat{\mathbf{n}}^{-2}\sum_{i\in\Gamma_n}\left[\sum_{j:\rho(i,j)\le 2m}1 + \sum_{j:\rho(i,j)>2m}m^{d\nu_*}\widehat{\alpha}(\rho(i,j)-2m)\right]$$

$$\le C_1\widehat{\mathbf{n}}^{-1}m^d + C_2\widehat{\mathbf{n}}^{-2}\sum_{i\in\Gamma_n}\sum_{r=2m}^{\infty}\sum_{j\in\Gamma_n:\rho(i,j)\in[r,r+1)}m^{d\nu_*}\widehat{\alpha}^{\delta/(2+\delta)}$$

$$\times([\rho(i,j)-2m])$$

$$\le C_1\widehat{\mathbf{n}}^{-1}m^d + C_2\widehat{\mathbf{n}}^{-1}\sum_{r=1}^{\infty}r^{d(\nu_*+1)-1}\widehat{\alpha}^{\delta/(2+\delta)}(r) \le C\,\widehat{\mathbf{n}}^{-1}m^d \tag{D.3}$$

as $\widehat{\alpha}(r) = O(r^{-\mu})$, $\mu > d\,(\nu+1+2/\delta)$, and $\sum_{r=1}^{\infty}r^{d(\nu_*+1)-1}\widehat{\alpha}^{\delta/(2+\delta)}(r)<\infty$.

Collecting (D.1)–(D.3) yields

$$E\left|\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n}\left[Y_{in}\cos(r^\tau X_{in}) - EY_{in}\cos(r^\tau X_{in})\right]\right|$$

$$\le C\left[(1+\|r\|)\,\psi(m) + \left(m^d/\widehat{\mathbf{n}}\right)^{1/2}\right].$$

The same inequality holds with $\cos(\cdot)$ replaced by $\sin(\cdot)$. Hence,

$$\sup_{x\in\mathbb{R}^p} |(V_n(x))_k - E\,(V_n(x))_k| = O_p\left(\int \left[(1+\|r\|)\,\psi(m) + \left(m^d/\widehat{\mathbf{n}}\right)^{1/2}\right] |\Psi_{k0}(rb_n)|\,dr\right).$$

By a change of variables,

$$\int \left[(1+\|r\|)\,\psi(m) + \left(m^d/\widehat{\mathbf{n}}\right)^{1/2}\right] |\Psi_{k0}(rb_n)|\,dr = b_n^{-p}\int \left[(1+\|r/b_n\|)\,\psi(m) + \left(m^d/\widehat{\mathbf{n}}\right)^{1/2}\right] |\Psi_{k0}(r)|\,dr.$$

Now, choose m to minimize the order of magnitude of the expression in the square brackets. This is achieved by setting $m = \widehat{\mathbf{n}}^{1/(2\gamma+d)} b_n^{-2/(2\gamma+d)}$. For this choice of m, we have by Assumption 6(b)

$$\begin{aligned} b_n^{-p}\int &\left[(1+\|r/b_n\|)\,\psi(m) + \left(m^d/\widehat{\mathbf{n}}\right)^{1/2}\right] |\Psi_{k0}(r)|\,dr \\ &\le C\,\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)} b_n^{-p-d/(2\gamma+d)} \int (\|r\|+1)\,|\Psi_{k0}(r)|\,dr \\ &\le C\,\widehat{\mathbf{n}}^{-\gamma/(2\gamma+d)} b_n^{-p-d/(2\gamma+d)} \end{aligned}$$

which completes the proof of part (a).

Part (b) follows by the same arguments as in part (a) by setting $Y_{in} := 1$ and using the Fourier transforms of the functions $\widetilde{K}_{kl}(x)$ instead of $\widetilde{K}_{k0}(x)$. □

Proof of Lemma D.2. As before, let $\widetilde{K}_{k0}(x) = x_k K(x)$. Using a second-order Taylor expansion of $g(u)$ about x in the proof of Lemma 3 gives

$$\begin{aligned} E\,|(W_n(x))_k| &= \left(\widehat{\mathbf{n}} b_n^p\right)^{-1} \left|\sum_{i\in\Gamma_n} E\Big[Y_{in} - g(x) - g'(x)^\tau \right. \\ &\qquad \left.\times\,(X_{in}-x)\Big]\widetilde{K}_{k0}\left(\frac{X_{in}-x}{b_n}\right)\right| \\ &= \frac{b_n^2}{2}\left|\int \left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x-ub_n)\right\}\right. \\ &\qquad \left.\times\, u^\tau g''(x-\vartheta u b_n) u \widetilde{K}_{k0}(u)du\right| \\ &\le \frac{b_n^2}{2}\sup_{n\ge1}\sup_{x\in\mathbb{R}^p}\left\{\widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x)\right\} \\ &\qquad \times \sup_{x\in\mathbb{R}^p}\left\|g''(x)\right\| \int \|u\|^3 K(u)du \le Cb_n^2 \end{aligned}$$

by Assumptions 5–6. □

Proof of Lemma D.3. Let $G_n = \sup_{x\in\mathbf{X}}\left\|U_n^{-1}(x)\right\|$. To prove the lemma, it suffices to show that there exists $B < \infty$ such that $\lim_{n\to\infty} P\,(G_n \le B) = 1$, which we will shortly denote as

$$G_n \le B \text{ w.p. } \to 1. \tag{D.4}$$

Since $U_n^{-1}(x)$ is symmetric, and $\left\|U_n^{-1}(x)\right\| \le \sqrt{(p+1)}\max_{\|c\|=1}\left(c^\tau U_n^{-2}(x)c\right)^{1/2} = \left|\lambda_{\max}\left(U_n^{-1}(x)\right)\right| = 1/\left|\lambda_{\min}\left(U_n(x)\right)\right|$, to prove (D.4) it in turn suffices to show that

$$1/\left[\inf_{x\in\mathbf{X}}|\lambda_{\min}\left(U_n(x)\right)|\right] \le B \text{ w.p. } \to 1 \tag{D.5}$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote, respectively, the largest and smallest eigenvalues of a matrix.

Inequality (D.5) will in turn follow if we show that there exists $L > 0$ s.t.

$$0 < L \le \inf_{x\in\mathbf{X}}|\lambda_{\min}\left(U_n(x)\right)| \text{ w.p. } \to 1. \tag{D.6}$$

Let

$$\Lambda = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \int uu^\tau K(u)du \end{pmatrix}$$

$$F_n(x) = \bar{f}_n(x)\Lambda \quad \text{with } \bar{f}_n(x) = \widehat{\mathbf{n}}^{-1}\sum_{i\in\Gamma_n} f_{in}(x).$$

We first show that

$$\sup_{x\in\mathbf{X}}\|U_n(x) - F_n(x)\| \xrightarrow{p} 0. \tag{D.7}$$

Note that by Lemma D.1(b), $\sup_{x\in\mathbf{X}}\|U_n(x) - EU_n(x)\| = o_p(1)$. Next, for each $k, l = 0, 1, \ldots, p$, we have by Assumption 5(b)

$$\begin{aligned} \sup_{x\in\mathbf{X}} &|E\,(U_n(x))_{kl} - (F_n(x))_{kl}| \\ &= \sup_{x\in\mathbf{X}}\left(\widehat{\mathbf{n}} b_n^p\right)^{-1}\left|\sum_{i\in\Gamma_n} E\left(\frac{X_{in}-x}{b_n}\right)_k\left(\frac{X_{in}-x}{b_n}\right)_l\right. \\ &\qquad \left.\times\,K\left((x-X_{in})/b_n\right) - \bar{f}_n(x)\Lambda_{kl}\right| \\ &= \sup_{x\in\mathbf{X}}\left|\int \bar{f}_n(x-b_nu)u_ku_lK(u)\,du - \bar{f}_n(x)\int u_ku_lK(u)\,du\right| \\ &\le \sup_{x\in\mathbf{X}}\int\left|\bar{f}_n(x-b_nu) - \bar{f}_n(x)\right||u_ku_lK(u)|\,du \\ &\le B_f b_n\int \|u\|^3|K(u)|\,du \le Cb_n \to 0 \end{aligned}$$

which proves (D.7). By Theorem 6 of Lax (2007, p. 130), (D.7) implies

$$\sup_{x\in\mathbf{X}}|\lambda_{\min}\left(F_n(x)\right) - \lambda_{\min}\left(U_n(x)\right)| = o_p(1).$$

It then follows that

$$\sup_{x\in\mathbf{X}}|\lambda_{\min}\left(F_n(x)\right)| \le \inf_{x\in\mathbf{X}}|\lambda_{\min}\left(U_n(x)\right)| \text{ w.p. } \to 1$$

and hence, by Assumption 6

$$\begin{aligned} 0 < M\,|\lambda_{\min}(\Lambda)| &\le \sup_{x\in\mathbf{X}}|\lambda_{\min}\left(F_n(x)\right)| \\ &\le \inf_{x\in\mathbf{X}}|\lambda_{\min}\left(U_n(x)\right)| \text{ w.p. } \to 1. \quad \square \end{aligned}$$

References

Andrews, D.W.K., 1984. Non-strong mixing autoregressive processes. Journal of Applied Probability 21, 930–934.

Andrews, D.W.K., 1995. Nonparametric kernel estimation of semiparametric models. Econometric Theory 11, 560–596.

Bardet, J.-M., Doukhan, P., Lang, G., Ragache, N., 2008. Dependent Lindeberg central limit theorem and some applications. ESAIM: Probability and Statistics 12, 154–172.

Bierens, H.J., 1983. Uniform consistency of kernel estimators of a regression function under generalized conditions. Journal of the American Statistical Association 77, 699–707.

Billingsley, P., 1968. Convergence of Probability Measures. John Wiley and Sons, New York.

Bradley, R., 1993. Some examples of mixing random fields. Rocky Mountain Journal of Mathematics 23 (2), 495–519.

Bulinskii, A.V., 1989. Limit Theorems Under Weak Dependence Conditions. Moscow University Press.

Chen, X., Conley, T., 2001. A new semiparametric spatial model for panel time series. Journal of Econometrics 105, 59–83.

Craven, P., Wahba, G., 1979. Smoothing noise data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerical Mathematics 31, 377–403.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

Doukhan, P., Truquet, L., 2007. Weakly dependent random fields with infinite interactions. ALEA: Latin American Journal of Probability and Mathematical Statistics 3, 111–132.

Dudley, R.M., Philipp, W., 1983. Invariance principles for sums of Banach space valued random elements and empirical processes. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 62, 509–552.

Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Mmethods. Springer, New York.

Gao, J., Lu, Z., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. Annals of Statistics 34, 1395–1435.

Gorodetskii, V.V., 1977. On the strong mixing property for linear sequences. Theory of Probability and Its Applications 22, 411–413.

Hallin, M., Lu, Z., Tran, L.T., 2001. Density estimation for spatial linear processes. Bernoulli 7, 657–668.

Hallin, M., Lu, Z., Tran, L.T., 2004a. Kernel density estimation for spatial processes: the L_1 theory. Multivariate Analysis 88, 61–75.

Hallin, M., Lu, Z., Tran, L.T., 2004b. Local linear spatial regression. Annals of Statistics 32, 2469–2500.

Ibragimov, I.A., 1962. Some limit theorems for stationary processes. Theory of Probability and Its Applications 7, 349–382.

Jenish, N., Prucha, I.R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86–98.

Jenish, N., Prucha, I.R., 2010. On spatial processes and asymptotic inference under near-epoch dependence. Working Paper. New York University.

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Lax, P.D., 2007. Linear Algebra and its Applications. Wiley-Interscience.

Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L.-F., 2007. GMM and 2SLS for mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

Lu, Z., Chen, X., 2002. Spatial nonparametric regression estimation: non-isotropic case. Acta Mathematicae Applicatae Sinica, English Series 18, 641–656.

Lu, Z., Chen, X., 2004. Spatial kernel regression: weak consistency. Statistics and Probability Letters 68, 125–136.

Lu, Z., Linton, O., 2007. Local linear fitting under near epoch dependence. Econometric Theory 23, 37–70.

McLeish, D.L., 1975. Invariance principles for dependent variables. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 32, 165–178.

Parzen, E., 1962. On estimation of a probability density function and mode. Annals of Mathematical Statistics 33, 1065–1076.

Robinson, P.M., 2008. Developments in the analysis of spatial data. Journal of the Japan Statistical Society 38, 87–96 (issue in honour of H. Akaike).

Robinson, P.M., 2009. Large-sample inference on spatial dependence. The Econometrics Journal 12, S68–S82 (Tenth Anniversary special issue).

Robinson, P.M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6–17.

Robinson, P.M., 2011. Asymptotic theory for nonparametric regression with spatial data. Journal of Econometrics 165, 5–19.