T E C H N I C A L R E P O R T

0242

LOCAL LINEAR SPATIAL REGRESSION

M. HALLIN, Z. LU and L. T. TRAN

I A P S T A T I S T I C S N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Local Linear Spatial Regression *

Marc Hallin,
ISRO, ECARES, and Département de Mathématique
Université Libre de Bruxelles, Belgium

Zudi Lu
Institute of Systems Science
Academy of Mathematics and Systems Sciences
Chinese Academy of Sciences, Beijing, China

and

Lanh T. Tran
Department of Mathematics
Indiana University
Bloomington, U.S.A.

November 20, 2002

Abstract

A local linear kernel estimator of the regression function $\mathbf{x} \mapsto g(\mathbf{x}) := \mathrm{E}[Y_{\mathbf{i}}|\mathbf{X_i} = \mathbf{x}]$, $\mathbf{x} \in \mathbb{R}^d$ of a stationary $(d+1)$-dimensional spatial processes $\{(Y_{\mathbf{i}}, \mathbf{X_i}), \mathbf{i} \in \mathbb{Z}^N\}$ observed over a rectangular domain of the form $\mathcal{I}_{\mathbf{n}} := \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N | 1 \leq i_k \leq n_k, \, k = 1, \ldots, N\}$, $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{Z}^N$ is proposed and investigated. Under mild regularity assumptions, asymptotic normality of the estimators of $g(\mathbf{x})$ and its derivatives is established. Appropriate choices of the bandwidths are proposed. The spatial process is assumed to satisfy some very general mixing conditions, generalizing classical time-series strong mixing concepts. The size of the rectangular domain $\mathcal{I}_{\mathbf{n}}$ is allowed to tend to infinity at different rates depending on the direction in $\mathbb{Z}^N$.

AMS 1980 subject classification : Primary: 62G05; Secondary: 60J25, 62J02.

Key words and phrases : mixing random field, bandwidth choice, local linear kernel estimate, spatial regression, asymptotic normality.

1 Introduction

Spatial data arise in a variety of fields, including econometrics, epidemiology, environmental science, image analysis, oceanography, and many others. The statistical treatment of such

*Research supported by a PAI contract of the Belgian Federal Government, an Action de Recherche Concertée of the Communauté Française de Belgique, the Natural Science Foundation of China, and NSF grant DMS-9403718.

data is the subject of an abundant literature, which cannot be reviewed here; for background reading, we refer the reader to the monographs by Anselin and Florax (1995), Cressie (1991), Guyon (1995), Possolo (1991), or Ripley (1981).

Let $\mathbb{Z}^N$, $N \geq 1$, denote the integer lattice points in the N-dimensional Euclidean space. A point $\mathbf{i} = (i_1, \ldots, i_N)$ in $\mathbb{Z}^N$ will be referred to as a *site*. Spatial data are modelled as finite realizations of vector stochastic processes indexed by $\mathbf{i} \in \mathbb{Z}^N$: *random fields*. In this paper, we will consider strictly stationary $(d+1)$-dimensional random fields, of the form

$$\Big\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})\ ;\ \mathbf{i} \in \mathbb{Z}^N\Big\}, \tag{1.1}$$

where $Y_{\mathbf{i}}$, with values in $\mathbb{R}$, and $\mathbf{X}_{\mathbf{i}}$, with values in $\mathbb{R}^d$, are defined over some probability space $(\Omega, \mathcal{F}, \mathrm{P})$.

A crucial problem for a number of applications is the problem of *spatial regression*, where the influence of a vector $\mathbf{X}_{\mathbf{i}}$ of covariates on some response variable $Y_{\mathbf{i}}$ is to be studied in a context of complex spatial dependence. More specifically, assuming that $Y_{\mathbf{i}}$ has finite expectation, the quantity under study in such problems is the *spatial regression function*

$$g\ :\ \mathbf{x} \mapsto g(\mathbf{x}) := \mathrm{E}\Big[Y_{\mathbf{i}} | \mathbf{X}_{\mathbf{i}} = \mathbf{x}\Big].$$

The spatial dependence structure in this context plays the role of a nuisance, and remains unspecified. Although g of course is only defined up to a P-null set of values of $\mathbf{x}$ (being a class of P-a.s. mutually equal functions rather than a function), we will treat it, for the sake of simplicity, as a well-defined real-valued $\mathbf{x}$-measurable function, which has no implication on the probabilistic statements of this paper. In the particular case under which $\mathbf{X}_{\mathbf{i}}$ itself is measurable with respect to a subset of $Y_{\mathbf{j}}$'s, with $\mathbf{j}$ ranging over some neighborhood of $\mathbf{i}$, g is called a *spatial* auto*regression function*. Such spatial autoregression models were considered as early as 1954, in the particular case of a linear autoregression function g, by Whittle (1954, 1963); see Besag (1974) for further developments in this context.

In this paper, we are concerned with estimating the spatial regression (autoregression) function $g : \mathbf{x} \mapsto g(\mathbf{x})$; contrary to Whittle (1954), we adopt a nonparametric point of view, avoiding any parametric specification of the possibly extremely complex spatial dependence structure of the data.

For $N = 1$, this problem reduces to the classical problem of (auto)regression for serially dependent observations, which has received extensive attention in the literature: see, for instance, Roussas (1969, 1988), Masry (1983, 1986), Robinson (1983, 1987), Ioannides and Roussas (1987), Masry and Györfi (1987), Yakowitz (1987), Boente and Fraiman (1988), Bosq (1989), Györfi, Härdle, Sarda and Vieu (1989), Tran (1989), Masry and Tjøstheim (1995), Hallin and Tran (1996), Lu and Cheng (1997), Lu (2001), Wu and Mielniczuk (2002), to quote only a few. Quite surprisingly, despite its importance for applications, the spatial version ($N > 1$) of the same problem remains essentially unexplored. Several recent papers (among which Tran 1990, Tran and Yakowitz 1993, Carbon, Hallin, and Tran 1996, Hallin, Lu, and Tran 2001 and 2002) are dealing with the related problem of estimating the density f of a random field of the form $\{\mathbf{X}_{\mathbf{i}}\ ;\ \mathbf{i} \in \mathbb{Z}^N\}$, but, to the best of our knowledge, the only results available on the estimation of spatial regression functions are those by Lu (2000), who investigates the properties of a Nadaraya-Watson kernel estimator for g.

Though the Nadaraya-Watson method is central in most nonparametric regression method in the traditional serial case ($N = 1$), it has been well documented (see, for instance, Fan and

Gijbels 1996) that this approach suffers from several severe drawbacks, such as poor boundary performances, excessive bias and low efficiency, and that the local polynomial fitting methods developed by Stone (1977) and Cleveland (1979) are generally preferable. Local polynomial fitting, and particularly its special case—local linear fitting —recently have become increasingly popular in the light of recent work by Cleveland and Loader (1996), Fan (1992), Fan and Gijbels (1992, 1995), Hastie and Loader (1993), Ruppert and Wand (1994), and several others. In this paper, we extend this approach to the context of spatial regression ($N > 1$) by defining an estimator of g based on local linear fitting and establishing its asymptotic properties.

The paper is organized as follows. In Section 2.1 we provide the notation and main assumptions. Section 2.2 introduces the main ideas underlying local linear regression in the context of random fields, and sketches the main steps of the proofs to be developed in the sequel. Section 2.3 is devoted to some preliminary results. Section 3 is the main section of the paper, where asymptotic normality is proved under various types of asymptotics and various mixing assumptions. Section 4 provides some numerical illustrations. Proofs and technical lemmas are concentrated in Section 5.

2 Local linear estimation of spatial regression.

2.1 Notation and main assumptions.

For the sake of convenience, we are summarizing here the main assumptions we are making on the random field (1.1) and the kernel K to be used in the estimation method. Assumptions (A1)-(A4) are related to the random field itself.

(A1) The random field (1.1) is strictly stationary. For all $\mathbf{i}$ and $\mathbf{j}$ in $\mathbb{Z}^N$, the vectors $\mathbf{X_i}$ and $\mathbf{X_j}$ admit a joint density $f_{\mathbf{ij}}$; moreover, $|f_{\mathbf{ij}}(\mathbf{x}', \mathbf{x}'') - f(\mathbf{x}')f(\mathbf{x}'')| \leq C$ for all $\mathbf{i}, \mathbf{j} \in \mathbb{Z}^N$, all $\mathbf{x}', \mathbf{x}'' \in \mathbb{R}^d$, where $C > 0$ is some constant, and f denotes the marginal density of $\mathbf{X_i}$.

(A2) The random variable $Y_{\mathbf{i}}$ has finite absolute moment of order $(2+\delta)$, that is, $\mathrm{E}\left[|Y_{\mathbf{i}}|^{2+\delta}\right] < \infty$ for some $\delta > 0$.

(A3) The spatial regression function g is twice differentiable. Denoting by $g'(\mathbf{x})$ and $g''(\mathbf{x})$ its gradient and the matrix of its second derivatives (at $\mathbf{x}$), respectively, $\mathbf{x} \mapsto g''(\mathbf{x})$ is continuous at all $\mathbf{x}$.

Assumption (A1) has been used by Masry (1986) in the serial case $N = 1$, and by Tran (1990) in the spatial context ($N > 1$).

Assumption (A4) is an assumption of spatial mixing taking two distinct forms (either (A4) and (A4$'$) or (A4) and (A4$''$)). For any collection of sites $\mathcal{S} \subset \mathbb{Z}^N$, denote by $\mathcal{B}(\mathcal{S})$ the Borel σ-field generated by $\{(Y_{\mathbf{i}}, \mathbf{X_i}) |\ \mathbf{i} \in \mathcal{S}\}$; for each couple $\mathcal{S}', \mathcal{S}''$, let $d(\mathcal{S}', \mathcal{S}'') := \min\{\|\mathbf{i}' - \mathbf{i}''\| \,|\, \mathbf{i}' \in \mathcal{S}',\ , \mathbf{i}'' \in \mathcal{S}''\}$ be the distance between $\mathcal{S}'$ and $\mathcal{S}''$, where $\|\mathbf{i}\| := (i_1^2 + \ldots + i_N^2)^{1/2}$ stands for the Euclidean norm. Finally, write $\mathrm{Card}(\mathcal{S})$ for the cardinality of $\mathcal{S}$.

(A4) There exist a function φ such that $\varphi(t) \downarrow 0$ as $t \to \infty$, and a function $\psi : \mathbb{N}^2 \to \mathbb{R}^+$ symmetric and decreasing in each of its two arguments, such that the random field (1.1) is mixing, with spatial mixing coefficients α satisfying

$$\begin{aligned} \alpha(\mathcal{B}(\mathcal{S}'), \mathcal{B}(\mathcal{S}'')) &:= \sup\{|\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)|, A \in \mathcal{B}(\mathcal{S}'),\ B \in \mathcal{B}(\mathcal{S}'')\} \\ &\leq \psi(\mathrm{Card}(\mathcal{S}'), \mathrm{Card}(\mathcal{S}''))\varphi(d(\mathcal{S}', \mathcal{S}'')). \end{aligned} \tag{2.1}$$

for any $\mathcal{S}', \mathcal{S}'' \subset \mathbb{Z}^N$. The function φ moreover is such that

$$\lim_{m\to\infty} m^a \sum_{j=m}^{\infty} j^{N-1}\{\varphi(j)\}^{\delta/(2+\delta)} = 0 \quad \text{for some constant } a > (4+\delta)N/(2+\delta).$$

The assumptions we are making on the function ψ are either

(A4′) $\psi(n', n'') \leq \min(n', n'')$

or

(A4″) $\psi(n', n'') \leq C(n' + n'' + 1)^{\kappa}$ for some $C > 0$ and $\kappa > 1$.

In case (2.1) holds with $\psi \equiv 1$, the random field $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})\}$ is called *strongly mixing.*

In the serial case ($N = 1$), many stochastic processes and time series are known to be strongly mixing. Withers (1981) has obtained various conditions for linear processes to be strongly mixing. Under certain weak assumptions, autoregressive and more general nonlinear time series models are strongly mixing with exponential mixing rates : see Pham and Tran (1985), Pham (1986), Tjøstheim (1990), and Lu (1998). Guyon (1987) has shown that the results of Withers under certain conditions extend to linear random fields, of the form $X_{\mathbf{n}} = \sum_{\mathbf{j}\in\mathbb{Z}^N} g_{\mathbf{j}} Z_{\mathbf{n}-\mathbf{j}}$, where the $Z_{\mathbf{j}}$'s are independent random variables. Assumptions (A4′) and (A4″) are the same as the mixing conditions used by Neaderhouser (1980) and Takahata (1983), respectively, and are weaker than the uniform strong mixing condition considered by Nakhapetyan (1980). They are satisfied by many spatial models, as shown by Neaderhouser (1980), Rosenblatt (1985), and Guyon (1987).

Throughout, we assume that the random field (1.1) is observed over a rectangular region of the form $\mathcal{I}_{\mathbf{n}} := \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N | \, 1 \leq i_k \leq n_k, \ k = 1, \ldots, N\}$, for $\mathbf{n} = (n_1, \cdots, n_N) \in \mathbb{Z}^N$ with strictly positive coordinates $n_1, \ldots, n_N$. The total sample size is thus $\widehat{\mathbf{n}} := \prod_{k=1}^N n_k$. We write $\mathbf{n} \to \infty$ as soon as $\min_{1\leq k\leq N}\{n_k\} \to \infty$. A more demanding way for $\mathbf{n}$ to tend to infinity is the one considered in Tran (1990): we use the notation $\mathbf{n} \implies \infty$ if $\mathbf{n} \to \infty$ and moreover $|n_j/n_k| < C$ for some $0 < C < \infty$, $1 \leq j, \, k \leq N$. In this latter case, all components of $\mathbf{n}$ are required to tend to infinity at the same rate.

Assumption (A5) deals with the kernel function $K : \mathbb{R}^N \to \mathbb{R}$ to be used in the estimation method. For any $\mathbf{c} := (c_0, \mathbf{c}_1^\tau)^\tau \in \mathbb{R}^{d+1}$, define

$$K_{\mathbf{c}}(\mathbf{u}) := (c_0 + \mathbf{c}_1^\tau \mathbf{u})K(\mathbf{u}). \tag{2.2}$$

(A5)(i) For any $\mathbf{c} \in \mathbb{R}^{d+1}$, $|K_{\mathbf{c}}(\mathbf{u})|$ is uniformly bounded by some constant $K_{\mathbf{c}}^+$, and is integrable: $\int_{\mathbb{R}^{d+1}} |K_{\mathbf{c}}(\mathbf{x})| d\mathbf{x} < \infty$.

(ii) For any $\mathbf{c} \in \mathbb{R}^{d+1}$, $|K_{\mathbf{c}}|$ has an integrable second order radial majorant, that is, $Q_{\mathbf{c}}^K(\mathbf{x}) := \sup_{\|\mathbf{y}\|\geq\|\mathbf{x}\|}[\|\mathbf{y}\|^2 K_{\mathbf{c}}(\mathbf{y})]$ is integrable.

Finally, for convenient reference, we are listing here some conditions on the asymptotic behavior, as $\mathbf{n} \to \infty$, of the bandwidth $b_{\mathbf{n}}$ that will be used in the sequel.

(B1) The bandwith $b_{\mathbf{n}}$ tends to zero in such a way that $\widehat{\mathbf{n}} b_{\mathbf{n}}^d \to \infty$ as $\mathbf{n} \to \infty$.

(B2) There exist two sequences of positive integer vectors, $\mathbf{p} = \mathbf{p}_{\mathbf{n}} := (p_1, \ldots, p_N) \in \mathbb{Z}^N$ and $\mathbf{q} = \mathbf{q}_{\mathbf{n}} := (q, \cdots, q) \in \mathbb{Z}^N$, with $q = q_{\mathbf{n}} \to \infty$ such that $p = p_{\mathbf{n}} := \widehat{\mathbf{p}} = o((\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2})$, $q/p_k \to 0$ and $n_k/p_k \to \infty$ for all $k = 1, \cdots, N$, and $\widehat{\mathbf{n}}\varphi(q) \to 0$.

(B2′) Same as (B2), but last condition replaced with $(\widehat{\mathbf{n}}^{\kappa+1}/p)\,\varphi(q) \to 0$, where κ is the constant appearing in (A4″).

(B3) $b_{\mathbf{n}}$ tends to zero in such a manner that $qb_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1$ and

$$b_{\mathbf{n}}^{-\delta d/(2+\delta)} \sum_{t=q}^{\infty} t^{N-1}\{\varphi(t)\}^{\delta/(2+\delta)} \to 0 \quad \text{as } \mathbf{n} \to \infty. \tag{2.3}$$

2.2 Local linear fitting.

The idea of local linear fitting consists in approximating, in a neighborhood of $\mathbf{x}$, the unknown function g by a linear function. Under (A3), we have

$$g(\mathbf{z}) \approx g(\mathbf{x}) + (\mathbf{g}'(\mathbf{x}))^{\tau}(\mathbf{z}-\mathbf{x}) := a_0 + \mathbf{a}_1^{\tau}(\mathbf{z}-\mathbf{x}).$$

Locally, this suggests estimating $(a_0, \mathbf{a}_1^{\tau}) = (g(\mathbf{x}), g'(\mathbf{x}))$, hence constructing an estimator of g from

$$\begin{pmatrix} g_{\mathbf{n}}(\mathbf{x}) \\ g'_{\mathbf{n}}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \hat{a}_0 \\ \hat{\mathbf{a}}_1 \end{pmatrix} := \arg \min_{(a_0,\,\mathbf{a}_1)\in\mathbb{R}^{d+1}} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} (Y_{\mathbf{j}} - a_0 - \mathbf{a}_1^{\tau}(\mathbf{X}_{\mathbf{j}}-\mathbf{x}))^2 K\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right), \tag{2.4}$$

where $b_{\mathbf{n}}$ is a sequence of bandwiths tending to zero at appropriate rate as $\mathbf{n}$ tends to infinity, and $K(\cdot)$ is a (bounded) kernel with values in $\mathbb{R}^+$.

In the classical serial case ($N=1$; we write i and n instead of $\mathbf{i}$ and $\mathbf{n}$), the solution of the minimization problem (2.4) is easily shown to be $(\mathbf{X}^{\tau}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\tau}\mathbf{W}\mathbf{Y}$, where $\mathbf{X}$ is an $n\times(d+1)$ matrix with i-th row $(1,\, b_n^{-1}(\mathbf{X}_i-\mathbf{x})^{\tau})$, $\mathbf{W} = b_n^{-1}\mathrm{diag}\left(K\left(\frac{\mathbf{X}_1-\mathbf{x}}{b_n}\right), \ldots, K\left(\frac{\mathbf{X}_n-\mathbf{x}}{b_n}\right)\right)$, and $\mathbf{Y} = (Y_1, \cdots, Y_n)^{\tau}$ (see, e.g., Fan and Gijbels 1996). In the spatial case, things are not as simple, and we rather write the solution to (2.4) as

$$\begin{pmatrix} \hat{a}_0 \\ \hat{\mathbf{a}}_1 b_{\mathbf{n}} \end{pmatrix} = U_{\mathbf{n}}^{-1}V_{\mathbf{n}}, \quad \text{where} \quad \mathbf{V}_{\mathbf{n}} := \begin{pmatrix} v_{\mathbf{n}0} \\ \mathbf{v}_{\mathbf{n}1} \end{pmatrix} \quad \text{and} \quad \mathbf{U}_{\mathbf{n}} := \begin{pmatrix} u_{\mathbf{n}00} & \mathbf{u}_{\mathbf{n}01} \\ \mathbf{u}_{\mathbf{n}10} & \mathbf{u}_{\mathbf{n}11} \end{pmatrix},$$

with (letting $\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)_0 := 1$)

$$(\mathbf{V}_{\mathbf{n}})_i := (\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{j}} \left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)_i K\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right), \quad i = 0, \ldots, d,$$

and

$$(\mathbf{U}_{\mathbf{n}})_{i\ell} := (\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)_i \left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)_{\ell} K\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right), \quad i,\, \ell = 0, \ldots, d.$$

It follows that

$$\mathbf{H}_{\mathbf{n}} := \begin{pmatrix} \hat{a}_0 - a_0 \\ \hat{\mathbf{a}}_1 b_{\mathbf{n}} - \mathbf{a}_1 b_{\mathbf{n}} \end{pmatrix} = \begin{pmatrix} g_{\mathbf{n}}(\mathbf{x}) - g(\mathbf{x}) \\ (g'_{\mathbf{n}}(\mathbf{x}) - g'(\mathbf{x}))b_{\mathbf{n}} \end{pmatrix} = \mathbf{U}_{\mathbf{n}}^{-1}\left\{\mathbf{V}_{\mathbf{n}} - \mathbf{U}_{\mathbf{n}}\begin{pmatrix} a_0 \\ \mathbf{a}_1 b_{\mathbf{n}} \end{pmatrix}\right\} := \mathbf{U}_{\mathbf{n}}^{-1}\mathbf{W}_{\mathbf{n}}, \tag{2.5}$$

where

$$\mathbf{W}_{\mathbf{n}} := \begin{pmatrix} w_{\mathbf{n}0} \\ \mathbf{w}_{\mathbf{n}1} \end{pmatrix}, \quad (\mathbf{W}_{\mathbf{n}})_i := (\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} Z_{\mathbf{j}} \left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)_i K\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right), \quad i = 0, \ldots, d, \tag{2.6}$$

and $Z_{\mathbf{j}} := Y_{\mathbf{j}} - a_0 - \mathbf{a}_1^{\tau}(\mathbf{X}_{\mathbf{j}}-\mathbf{x})$.

The organization of the paper is as follows. If, under adequate conditions, we are able to show that

(C1) $(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}(\mathbf{W_n} - \mathrm{E}\mathbf{W_n})$ is asymptotically normal,

(C2) $(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}\mathrm{E}\mathbf{W_n} \longrightarrow \mathbf{0}$ and $\mathrm{Var}\left((\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}\mathbf{W_n}\right) \longrightarrow \mathbf{\Sigma}$, and

(C3) $\mathbf{U_n} \xrightarrow{\mathrm{P}} \mathbf{U}$,

then (2.5) and Slutsky's classical argument imply that, for all $\mathbf{x}$ (all quantities involved indeed depend on $\mathbf{x}$)

$$(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2} \begin{pmatrix} g_{\mathbf{n}}(\mathbf{x}) - g(\mathbf{x}) \\ (g'_{\mathbf{n}}(\mathbf{x}) - g'(\mathbf{x}))b_{\mathbf{n}} \end{pmatrix} = (\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}\mathbf{H_n} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0},\, \mathbf{U}^{-1}\mathbf{\Sigma}\mathbf{U}^{-1\tau}).$$

This asymptotic normality result (with explicit values of $\mathbf{\Sigma}$ and $\mathbf{U}$), under various forms (depending on the mixing assumptions ((A4′) or (A4″)), the choice of the bandwidth $b_{\mathbf{n}}$, the way $\mathbf{n}$ tends to infinity, etc.), is the main contribution of this paper; see Theorems 3.1-3.5. Subsection 2.3 is dealing with (C2) and (C3) under $\mathbf{n} \to \infty$ (hence also under the stronger assumption that $\mathbf{n} \implies \infty$), Subsections 3.1 and 3.2 with (C1) under $\mathbf{n} \implies \infty$ and $\mathbf{n} \to \infty$, respectively.

2.3 Preliminaries.

Claim (C3) is easily established from the following lemma, the proof of which is similar to that of Lemma 2.2 below, and is therefore omitted.

Lemma 2.1 *Assume that Assumptions (A1), (A4), and (A5) hold, that $b_{\mathbf{n}}$ satisfies Assumption (B1), and that $n_k b_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1$ as $\mathbf{n} \to \infty$. Then, for all $\mathbf{x}$,*

$$\mathbf{U_n} \xrightarrow{\mathrm{P}} \mathbf{U} := \begin{pmatrix} f(\mathbf{x})\int K(\mathbf{u})\,d\mathbf{u} & f(\mathbf{x})\int \mathbf{u}^\tau K(\mathbf{u})\,d\mathbf{u} \\ f(\mathbf{x})\int \mathbf{u}K(\mathbf{u})\,d\mathbf{u} & f(\mathbf{x})\int \mathbf{u}\mathbf{u}^\tau K(\mathbf{u})\,d\mathbf{u} \end{pmatrix}$$

as $\mathbf{n} \to \infty$.

The remainder of this section is devoted to claim (C2). The usual Cramér-Wold device will be adopted. For all $\mathbf{c} := (c_0, \mathbf{c}_1^\tau)^\tau \in \mathbb{R}^{1+d}$, let

$$A_{\mathbf{n}} := (\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}\mathbf{c}^\tau\mathbf{W_n} = (\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{-1/2}\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} Z_{\mathbf{j}}\; K_{\mathbf{c}}\left(\frac{\mathbf{X_j} - \mathbf{x}}{b_{\mathbf{n}}}\right),$$

with $K_{\mathbf{c}}(\mathbf{u})$ defined in (2.2). The following lemma provides the asymptotic variance of $A_{\mathbf{n}}$ for all $\mathbf{c}$, hence that of $(\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}\mathbf{W_n}$.

Lemma 2.2 *Assume that Assummptions (A1), (A2), (A4), and (A5) hold, that $b_{\mathbf{n}}$ satisfies Assumption (B1), and that $n_k b_{\mathbf{n}}^{\delta d/[(2+\delta)a]} > 1$ for all $k = 1, \cdots, N$, as $\mathbf{n} \to \infty$. Then,*

$$\lim_{\mathbf{n}\to\infty} \mathrm{Var}[A_{\mathbf{n}}] = \mathrm{Var}(Y_{\mathbf{j}}|\mathbf{X_j} = \mathbf{x})f(\mathbf{x})\int_{\mathbb{R}^d} K_{\mathbf{c}}^2(\mathbf{u})d\mathbf{u} = \mathbf{c}^\tau\mathbf{\Sigma}\mathbf{c}, \tag{2.7}$$

where

$$\mathbf{\Sigma} := \mathrm{Var}(Y_{\mathbf{j}}|\mathbf{X_j} = \mathbf{x})f(\mathbf{x}) \begin{pmatrix} \int K^2(\mathbf{u})d\mathbf{u} & \int \mathbf{u}^\tau K^2(\mathbf{u})d\mathbf{u} \\ \int \mathbf{u}K^2(\mathbf{u})d\mathbf{u} & \int \mathbf{u}\mathbf{u}^\tau K^2(\mathbf{u})d\mathbf{u} \end{pmatrix}.$$

Hence, $\lim_{\mathbf{n}\to\infty} \mathrm{Var}\left((\hat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2}\mathbf{W_n}\right) = \mathbf{\Sigma}$.

Proof. See Section 5.1 $\square$

Next, we consider the asymptotic behavior of $\mathrm{E}[A_{\mathbf{n}}]$.

Lemma 2.3 *Under Assumptions (A3) and (A5),*

$$\begin{aligned}\mathrm{E}[A_{\mathbf{n}}] &= \sqrt{\widehat{\mathbf{n}}b_{\mathbf{n}}^d}b_{\mathbf{n}}^2\frac{1}{2}f(\mathbf{x})\ \mathrm{tr}\left[g''(\mathbf{x})\int \mathbf{u}\mathbf{u}^\tau K_{\mathbf{c}}(\mathbf{u})d\mathbf{u}\right] + o\left(\sqrt{\widehat{\mathbf{n}}b_{\mathbf{n}}^d}b_{\mathbf{n}}^2\right)\\ &= \sqrt{\widehat{\mathbf{n}}b_{\mathbf{n}}^d}b_{\mathbf{n}}^2\left[c_0B_0(\mathbf{x}) + \mathbf{c}_1^\tau\mathbf{B}_1(\mathbf{x})\right] + o\left(\sqrt{\widehat{\mathbf{n}}b_{\mathbf{n}}^d}b_{\mathbf{n}}^2\right),\end{aligned} \tag{2.8}$$

where

$$B_0(\mathbf{x}) := \frac{1}{2}f(\mathbf{x})\sum_{i=1}^{d}\sum_{j=1}^{d}g_{ij}(\mathbf{x})\int u_ju_iK(\mathbf{u})d\mathbf{u}, \quad \mathbf{B}_1(\mathbf{x}) := \frac{1}{2}f(\mathbf{x})\sum_{i=1}^{d}\sum_{j=1}^{d}g_{ij}(\mathbf{x})\int u_ju_i\mathbf{u}K(\mathbf{u})d\mathbf{u},$$

$g_{ij}(\mathbf{x}) = \partial^2 g(\mathbf{x})/\partial x_i\partial x_j$, *$i$, $j = 1, \ldots, d$, and* $\mathbf{u} := (u_1, \ldots u_d)^\tau \in \mathbb{R}^d$.

Proof. See Section 5.2. $\square$

3 Asymptotic Normality.

3.1 Asymptotic Normality under mixing Assumption (A4′)

The asymptotic normality of our estimators relies in a crucial manner on the following lemma (see (2.6) for the definition of $\mathbf{W}_{\mathbf{n}}(x)$).

Lemma 3.1 *Suppose that Assumptions (A1), (A2), (A4)-(A4′), and (A5) hold, and that the bandwidth $b_{\mathbf{n}}$ satisfies conditions (B1)-(B3). Denote by σ^2 the asymptotic variance (2.7). Then $(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}(\mathbf{c}^\tau[\mathbf{W}_{\mathbf{n}}(\mathbf{x}) - \mathrm{E}\mathbf{W}_{\mathbf{n}}(\mathbf{x})]/\sigma)$ is asymptotically standard normal as $\mathbf{n} \to \infty$.*

Proof. Putting

$$\eta_{\mathbf{j}}(\mathbf{x}) := Z_{\mathbf{j}}K_{\mathbf{c}}(\mathbf{x} - \mathbf{X}_{\mathbf{j}}) \quad \text{and} \quad \Delta_{\mathbf{j}}(\mathbf{x}) := \eta_{\mathbf{j}}(\mathbf{x}) - \mathrm{E}\eta_{\mathbf{j}}(\mathbf{x}), \tag{3.1}$$

define $\zeta_{\mathbf{nj}} := b_{\mathbf{n}}^{-d/2}\Delta_{\mathbf{j}}$, and let $S_{\mathbf{n}} := \sum_{\substack{j_k=1\\ k=1,\ldots,N}}^{n_k} \zeta_{\mathbf{nj}}$. Then,

$$\widehat{\mathbf{n}}^{-1/2}S_{\mathbf{n}} = (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\mathbf{c}^\tau(W_{\mathbf{n}}(\mathbf{x}) - \mathrm{E}W_{\mathbf{n}}(\mathbf{x})) = A_{\mathbf{n}} - \mathrm{E}A_{\mathbf{n}}.$$

Now, let us decompose $\widehat{\mathbf{n}}^{-1/2}S_{\mathbf{n}}$ into smaller pieces involving "large" and "small" blocks. More specifically, consider (all sums are running over $\mathbf{i} := (i_1, \ldots, i_N)$)

$$U(1, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\substack{i_k = j_k(p_k+q)+1 \\ k=1,\ldots,N}}^{j_k(p_k+q)+p_k} \zeta_{\mathbf{ni}(\mathbf{x})} \,,$$

$$U(2, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\substack{i_k = j_k(p_k+q)+1 \\ k=1,\ldots,N-1}}^{j_k(p_k+q)+p_k} \sum_{i_N = j_N(p_N+q)+p_N+1}^{(j_N+1)(p_N+q)} \zeta_{\mathbf{ni}}(\mathbf{x}) \,,$$

$$U(3, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\substack{i_k = j_k(p_k+q)+1 \\ k=1,\ldots,N-2}}^{j_k(p_k+q)+p_k} \sum_{i_{N-1} = j_{N-1}(p_{N-1}+q)+p_{N-1}+1}^{(j_{N-1}+1)(p_{N-1}+q)} \sum_{i_N = j_N(p_N+q)+1}^{j_N(p_N+q)+p_N} \zeta_{\mathbf{ni}}(\mathbf{x}) \,,$$

$$U(4, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\substack{i_k = j_k(p_k+q)+1 \\ k=1,\ldots,N-2}}^{j_k(p_k+q)+p_k} \sum_{i_{N-1} = j_{N-1}(p_{N-1}+q)+p_{N-1}+1}^{(j_{N-1}+1)(p_{N-1}+q)} \sum_{i_N = j_N(p_N+q)+p_N+1}^{(j_N+1)(p_N+q)} \zeta_{\mathbf{ni}}(\mathbf{x}) \,,$$

and so on. Note that

$$U(2^N - 1, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\substack{i_k = j_k(p_k+q)+p_k+1 \\ k=1,\ldots,N-1}}^{(j_k+1)(p_k+q)} \sum_{i_N = j_N(p_N+q)+1}^{j_N(p_N+q)+p_N} \zeta_{\mathbf{ni}}(\mathbf{x}),$$

and

$$U(2^N, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\substack{i_k = j_k(p_k+q)+p_k+1 \\ k=1,\ldots,N}}^{(j_k+1)(p_k+q)} \zeta_{\mathbf{ni}}(\mathbf{x}).$$

Without loss of generality, assume that, for some integers $r_1, \ldots, r_N$, $\mathbf{n} = (n_1, \ldots, n_N)$ is such that $n_1 = r_1(p_1 + q), \ldots, n_N = r_N(p_N + q)$, with $r_k \to \infty$ for all $k = 1, \cdots, N$. For each integer $1 \le i \le 2^N$, define

$$T(\mathbf{n}, \mathbf{x}, i) := \sum_{\substack{j_k = 0 \\ k=1,\ldots,N}}^{r_k - 1} U(i, \mathbf{n}, \mathbf{x}, \mathbf{j}).$$

Clearly $S_{\mathbf{n}} = \sum_{i=1}^{2^N} T(\mathbf{n}, \mathbf{x}, i)$. Note that $T(\mathbf{n}, \mathbf{x}, 1)$ is the sum of the random variables $\zeta_{\mathbf{ni}}$ over "large" blocks, whereas $T(\mathbf{n}, \mathbf{x}, i),\ 2 \le i \le 2^N$ are sums over "small" blocks. If it is not the case that $n_1 = r_1(p_1 + q), \ldots, n_N = r_N(p_N + q)$ for some integers $r_1, \ldots r_N$, then an additional term $T(\mathbf{n}, \mathbf{x}, 2^N + 1)$, say, , containing all the $\zeta_{\mathbf{nj}}$'s that are not included in the big or small blocks, can be considered. This term will not change the proof much. The general approach consists in showing that, as $\mathbf{n} \to \infty$,

$$Q_1 := \left| \mathrm{E}\exp[iuT(\mathbf{n},\mathbf{x},1)] - \prod_{\substack{j_k=0\\k=1,\ldots,N}}^{r_k-1} \mathrm{E}\exp[iuU(1,\mathbf{n},\mathbf{x},\mathbf{j})] \right| \longrightarrow 0, \tag{3.2}$$

$$Q_2 \equiv \widehat{\mathbf{n}}^{-1}\mathrm{E}\left(\sum_{i=2}^{2^N} T(\mathbf{n},\mathbf{x},i)\right)^2 \longrightarrow 0, \tag{3.3}$$

$$Q_3 := \widehat{\mathbf{n}}^{-1} \sum_{\substack{j_k=0\\k=1,\ldots,N}}^{r_k-1} \mathrm{E}[U(1,\mathbf{n},\mathbf{x},\mathbf{j})]^2 \longrightarrow \sigma^2, \tag{3.4}$$

$$Q_4 \equiv \widehat{\mathbf{n}}^{-1} \sum_{\substack{j_k=0\\k=1,\ldots,N}}^{r_k-1} \mathrm{E}[(U(1,\mathbf{n},\mathbf{x},\mathbf{j}))^2 I\{|U(1,\mathbf{n},\mathbf{x},\mathbf{j})| > \varepsilon\sigma\widehat{\mathbf{n}}^{1/2}\}] \longrightarrow 0 \tag{3.5}$$

for every $\varepsilon > 0$. Note that

$$\begin{aligned}[][A_{\mathbf{n}} - \mathrm{E}A_{\mathbf{n}}]/\sigma &= (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\mathbf{c}^\tau[W_{\mathbf{n}}(\mathbf{x}) - \mathrm{E}W_{\mathbf{n}}(\mathbf{x})]/\sigma = S_{\mathbf{n}}/(\sigma\widehat{\mathbf{n}}^{1/2}) \\ &= T(\mathbf{n},\mathbf{x},1)/(\sigma\widehat{\mathbf{n}}^{1/2}) + \sum_{i=2}^{2^N} T(\mathbf{n},\mathbf{x},i)/(\sigma\widehat{\mathbf{n}}^{1/2}).\end{aligned}$$

The term $\sum_{i=2}^{2^N} T(\mathbf{n},\mathbf{x},i)/(\sigma\widehat{\mathbf{n}}^{1/2})$ is asymptotically negligible by (3.3). The random variables $U(1,\mathbf{n},\mathbf{x},\mathbf{j})$ are asymptotically mutually independent by (3.2). The asymptotic normality of $T(\mathbf{n},\mathbf{x},1)/(\sigma\widehat{\mathbf{n}}^{1/2})$ follows from (3.4) and the Lindeberg-Feller condition (3.5). The lemma thus follows if we can prove (3.2)-(3.5). This proof is given in Section 5.3. The arguments there are reminiscent of those used by Masry (1986) and Nakhapetyan (1987).

We now turn to the main consistency and asymptotic normality results. First, we consider the case where the sample size tends to ∞ in the manner of Tran (1990), that is, $\mathbf{n} \implies \infty$.

Theorem 3.1 *Let Assumptions (A1), (A2), (A3), (A4′), and (A5) hold, with $\varphi(x) = O(x^{-\mu})$ for some $\mu > 2(3+\delta)N/\delta$. Suppose that there exists a sequence of positive integers $q = q_{\mathbf{n}} \to \infty$ such that $q_{\mathbf{n}} = o((\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/(2N)})$ and $\widehat{\mathbf{n}}q^{-\mu} \to 0$ as $\mathbf{n} \implies \infty$, and that the bandwidth $b_{\mathbf{n}}$ tends to zero in such a manner that*

$$q b_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1 \quad \textit{for some } (4+\delta)N/(2+\delta) < a < \mu\delta/(2+\delta) - N \tag{3.6}$$

as $\mathbf{n} \implies \infty$. Then,

$$(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\left[\begin{pmatrix} g_{\mathbf{n}}(\mathbf{x}) - g(\mathbf{x}) \\ b_{\mathbf{n}}(g'_{\mathbf{n}}(\mathbf{x}) - g'(\mathbf{x})) \end{pmatrix} - \mathbf{U}^{-1}\begin{pmatrix} B_0(\mathbf{x}) \\ \mathbf{B}_1(\mathbf{x}) \end{pmatrix} b_{\mathbf{n}}^2\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0},\, \mathbf{U}^{-1}\mathbf{\Sigma}(\mathbf{U}^{-1})^\tau\right), \tag{3.7}$$

as $\mathbf{n} \implies \infty$, where $\mathbf{U}$, $\mathbf{\Sigma}$, $B_0(\mathbf{x})$ and $\mathbf{B}_1(\mathbf{x})$ are defined in Lemmas 2.1, 2.2, and 2.3, respectively. If furthermore the kernel $K(\cdot)$ is a symmetric density function, then (3.7) can be reinforced into

$$\begin{pmatrix} (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\left[g_{\mathbf{n}}(\mathbf{x}) - g(\mathbf{x}) - B_g(\mathbf{x})b_{\mathbf{n}}^2\right] \\ (\widehat{\mathbf{n}}b_{\mathbf{n}}^{d+2})^{1/2}[g'_{\mathbf{n}}(\mathbf{x}) - g'(\mathbf{x})] \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_0^2(\mathbf{x}) & 0 \\ 0 & \boldsymbol{\sigma}_1^2(\mathbf{x}) \end{pmatrix}\right)$$

(so that $g_{\mathbf{n}}(\mathbf{x})$ and $g'_{\mathbf{n}}(\mathbf{x})$ are asymptotically independent), where

$$B_g(\mathbf{x}) := \frac{1}{2}\sum_{i=1}^{d} g_{ii}(\mathbf{x})\int(\mathbf{u})_i^2 K(\mathbf{u})d\mathbf{u}, \quad \sigma_0^2(\mathbf{x}) := \frac{\mathrm{Var}(Y_{\mathbf{j}}|\mathbf{X}_{\mathbf{j}}=\mathbf{x})\int K^2(\mathbf{u})d\mathbf{u}}{f(\mathbf{x})},$$

and

$$\boldsymbol{\sigma}_1^2(\mathbf{x}) := \frac{\mathrm{Var}(Y_{\mathbf{j}}|\mathbf{X}_{\mathbf{j}}=\mathbf{x})}{f(\mathbf{x})}\left[\int \mathbf{u}\mathbf{u}^\tau K(\mathbf{u})d\mathbf{u}\right]^{-1}\left[\int \mathbf{u}\mathbf{u}^\tau K^2(\mathbf{u})d\mathbf{u}\right]\left[\int \mathbf{u}\mathbf{u}^\tau K(\mathbf{u})d\mathbf{u}\right]^{-1}.$$

Proof. Since q is $o((\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/(2N)})$, there exists $s_{\mathbf{n}} \to 0$ such that $q = (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/(2N)}s_{\mathbf{n}}$. Take $p_k := (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/(2N)}s_{\mathbf{n}}^{1/2}$, $k = 1, \ldots, N$. Then $q/p_k = s_{\mathbf{n}}^{1/2} \to 0$, $\widehat{\mathbf{p}} = (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}s_{\mathbf{n}}^{N/2} = o((\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2})$, and $\widehat{\mathbf{n}}\varphi(q) = \widehat{\mathbf{n}}q^{-\mu} \to 0$. As $\mathbf{n} \implies \infty$, $p := \widehat{\mathbf{p}} < (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}$ for large $\widehat{\mathbf{n}}$. It follows that $\widehat{\mathbf{n}}/p > (\widehat{\mathbf{n}}b_{\mathbf{n}}^{-d})^{1/2} \to \infty$, hence $n_k/p_k \to \infty$ for all k. Thus, condition (B2) is satisfied.

Because $\varphi(j) = Cj^{-\mu}$,

$$\begin{aligned} m^a \sum_{j=m}^{\infty} j^{N-1}\{\varphi(j)\}^{\delta/(2+\delta)} &= Cm^a \sum_{j=m}^{\infty} j^{N-1}j^{-\mu\delta/(2+\delta)} \\ &\le Cm^a m^{N-\mu\delta/(2+\delta)} = m^{-[\mu\delta/(2+\delta)-a-N]} \end{aligned}$$

a quantity that tends to 0 as $m \to \infty$ since $(4+\delta)N/(2+\delta) < a < \mu\delta/(2+\delta) - N$, hence $\mu\delta/(2+\delta) > a+N$. Assumption (A4) and the fact that $qb_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1$ imply that $b_{\mathbf{n}}^{-\delta d/(2+\delta)} < q^a$ and that (2.3) holds. Now,

$$\mathbf{H}_{\mathbf{n}} - \mathbf{U}^{-1}\mathrm{E}\mathbf{W}_{\mathbf{n}} = \mathbf{U}_{\mathbf{n}}^{-1}(\mathbf{W}_{\mathbf{n}} - \mathrm{E}\mathbf{W}_{\mathbf{n}}) + (\mathbf{U}_{\mathbf{n}}^{-1} - \mathbf{U}^{-1})\mathrm{E}\mathbf{W}_{\mathbf{n}}.$$

The theorem thus follows from Lemmas 2.1, 2.3, and 3.1. □

In the important particular case under which $\varphi(x)$ tends to zero at exponential rate, the same results are obtained under milder conditions.

Theorem 3.2 *Let Assumptions (A1), (A2), (A3), (A4′), and (A5) hold, with $\varphi(x) = O(e^{-\xi x})$ for some $\xi > 0$. Then, if $b_{\mathbf{n}}$ tends to zero as $\mathbf{n} \implies \infty$ in such a manner that*

$$(\widehat{\mathbf{n}}b_{\mathbf{n}}^{d(1+2N\delta/[a(2+\delta)])})^{1/(2N)}(\log\widehat{\mathbf{n}})^{-1} \to \infty \text{ for some } a > (4+\delta)N/(2+\delta), \tag{3.8}$$

the conclusions of Theorem 3.1 still hold.

Proof. By (3.8), there exists a monotone positive function $\mathbf{n} \mapsto g(\mathbf{n})$ such that $g(\mathbf{n}) \to \infty$ and $(\widehat{\mathbf{n}}b_{\mathbf{n}}^{d(1+2N\delta/[a(2+\delta)])})^{(1/2N)}(g(\mathbf{n})\log\widehat{\mathbf{n}})^{-1} \to \infty$ as $\mathbf{n} \implies \infty$. Let $q := (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{(1/2N)}(g(\mathbf{n}))^{-1}$, and $p_k := (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/(2N)}g^{-1/2}(\mathbf{n})$. Then $q/p_k = g^{-1/2}(\mathbf{n}) \to 0$, $\widehat{\mathbf{p}} = (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}g^{-N/2}(\mathbf{n}) = o((\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2})$, and $n_k/p_k \to \infty$ as $\mathbf{n} \implies \infty$. For arbitrary $C > 0$, $q \ge C\log\widehat{\mathbf{n}}$ for sufficiently large $\widehat{\mathbf{n}}$. Thus

$$\widehat{\mathbf{n}}\varphi(q) \le C\widehat{\mathbf{n}}e^{-\xi q} \le C\widehat{\mathbf{n}}\exp(-C\xi\log\widehat{\mathbf{n}}) = C\widehat{\mathbf{n}}^{-C\xi+1},$$

which tends to zero if we choose $C > 1/\xi$. Hence, condition (B2) is satisfied. Next, for $0 < \xi' < \xi$,

$$q^a\sum_{i=q}^{\infty} i^{N-1}\varphi(i)^{\delta/(2+\delta)} \le Cq^a\sum_{i=q}^{\infty} i^{N-1}e^{-\xi i\delta/(2+\delta)} \le Cq^a\sum_{i=q}^{\infty} e^{-\xi' i\delta/(2+\delta)} \le Cq^a e^{-\xi' q\delta/(2+\delta)}.$$

Note that $b_{\mathbf{n}}^d \geq C\widehat{\mathbf{n}}^{-1}$ and $q > C \log \widehat{\mathbf{n}}$, so that Assumption (A4) holds. In addition,

$$qb_{\mathbf{n}}^{\delta d/[a(2+\delta)]} = (\widehat{\mathbf{n}} b_{\mathbf{n}}^{d+2N\delta d/[a(2+\delta)]})^{(1/2N)}(g(\mathbf{n}))^{-1} > 1$$

for $\widehat{\mathbf{n}}$ large enough. It is easily verified that this implies that condition (B3) is satisfied. The theorem follows. $\square$

Note that, in the one-dimensional case $N = 1$, and for "large" values of a, the condition (3.8) is "close" to the condition that $nb_n^d \to \infty$, which is usual in the classical case of independent observations.

Next, we consider the situation under which the sample size tends to ∞ in the "weak" sense (that is, $\mathbf{n} \to \infty$ instead of $\mathbf{n} \implies \infty$).

Theorem 3.3 *Let Assumptions (A1), (A2), (A3), (A4′), and (A5) hold, with* $\varphi(x) = O(x^{-\mu})$ *for some* $\mu > 2(3+\delta)N/\delta$. *Let the sequence of positive integers* $q = q_{\mathbf{n}} \to \infty$ *and a bandwidth* $b_{\mathbf{n}}$ *factorizing into* $b_{\mathbf{n}} := \prod_{i=1}^N b_{n_i}$, *such that* $\widehat{\mathbf{n}} q^{-\mu} \to 0$, $q = o((\min_{1\leq k\leq N}(n_k b_{n_k}^d))^{1/2})$, *and*

$$qb_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1 \quad \textit{for some } (4+\delta)N/(2+\delta) < a < \mu\delta/(2+\delta) - N.$$

Then the conclusions of Theorem 3.1 hold as $\mathbf{n} \to \infty$.

Proof. Since $q = o(\min_{1\leq k\leq N}(n_k b_{n_k}^d)^{1/2})$, there exists a sequence $s_{n_k} \to 0$ such that

$$q = \min_{1\leq k\leq N}((n_k b_{n_k}^d)^{1/2} s_{n_k}) \quad \text{as } \mathbf{n} \to \infty.$$

Take $p_k = (n_k b_{n_k}^d)^{1/2} s_{n_k}^{1/2}$. Then $q/p_k \leq s_{n_k}^{1/2} \to 0$, $\widehat{\mathbf{p}} = (\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2} \prod_{k=1}^N s_{n_k}^{1/2} = o((\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2})$, and $\widehat{\mathbf{n}}\varphi(q) = \widehat{\mathbf{n}} q^{-\mu} \to 0$. As $\mathbf{n} \to \infty$, $p_k < (n_k b_{n_k}^d)^{1/2}$, hence $n_k/p_k > (n_k b_{n_k}^{-d})^{1/2} \to \infty$. Thus condition (B2) is satisfied. The end of the proof is entirely similar to that of Theorem 3.1. $\square$

In the important case that $\varphi(x)$ tends to zero at an exponential rate, we have the following result, which parallels Theorem 3.2.

Theorem 3.4 *Let Assumptions (A1), (A2), (A3), (A4′), and (A5) hold, with* $\varphi(x) = O(e^{-\xi x})$ *for some* $\xi > 0$. *Let the bandwidth* $b_{\mathbf{n}}$ *factorize into* $b_{\mathbf{n}} := \prod_{i=1}^N b_{n_i}$ *in such a way that, as* $\mathbf{n} \to \infty$,

$$\min_{1\leq k\leq N}\{(n_k b_{n_k}^d)^{1/2}\} b_{\mathbf{n}}^{d\delta/[a(2+\delta)]}(\log \widehat{\mathbf{n}})^{-1} \to \infty \quad \textit{for some } a > (4+\delta)N/(2+\delta). \tag{3.9}$$

Then the conclusions of Theorem 3.1 hold as $\mathbf{n} \to \infty$.

Proof. By (3.9), there exist positive sequences indexed by n_k such that $g_{n_k} \uparrow \infty$ as $n_k \to \infty$ and

$$\min_{1\leq k\leq N}\{(n_k b_{n_k}^d)^{1/2} g_{n_k}^{-1}\} b_{\mathbf{n}}^{d\delta/[a(2+\delta)]}(\log \widehat{\mathbf{n}})^{-1} \to \infty$$

as $\mathbf{n} \to \infty$. Let $q := \min_{1\leq k\leq N}\{(n_k b_{n_k}^d)^{1/2}(g_{n_k})^{-1}\}$, and $p_k := (n_k b_{n_k}^d)^{1/2} g_{n_k}^{-1/2}$. Then $q/p_k \leq g_{n_k}^{-1/2} \to 0$, $\widehat{\mathbf{p}} = (\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2} \prod_{k=1}^N g_{n_k}^{-1/2} = o((\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2})$, and $n_k/p_k = (n_k b_{n_k}^{-d})^{1/2} g_{n_k}^{1/2} \to \infty$ as $\mathbf{n} \to \infty$. For arbitrary $C > 0$, $q \geq C \log \widehat{\mathbf{n}}$ for suficiently large $\widehat{\mathbf{n}}$. Thus,

$$\widehat{\mathbf{n}}\varphi(q) \leq C\widehat{\mathbf{n}} e^{-\xi q} \leq C\widehat{\mathbf{n}} \exp(-C\xi \log \widehat{\mathbf{n}}) = C\widehat{\mathbf{n}}^{-C\xi+1},$$

which tends to zero for $C > 1/\xi$. Hence, condition (B2) is satisfied. Next, for $0 < \xi' < \xi$,

$$q^a \sum_{i=q}^{\infty} i^{N-1} \varphi(i)^{\delta/(2+\delta)} \le C q^a \sum_{i=q}^{\infty} i^{N-1} e^{-\xi i \delta/(2+\delta)} \le C q^a \sum_{i=q}^{\infty} e^{-\xi' i \delta/(2+\delta)} \le C q^a e^{-\xi' q \delta/(2+\delta)}.$$

Note that $q > C \log \widehat{\mathbf{n}}$. Assumption (A4$'$) and (3.6) imply that $q b_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1$ for $\mathbf{n}$ large enough. This in turn implies that condition (B3) is satisfied. The theorem follows. □

3.2 Asymptotic Normality under mixing Assumption (A4$''$).

We start with an equivalent, under (A4$''$), of Lemma 3.1

Lemma 3.2 *Suppose that Assumptions (A1), (A2), (A4)-(A4$''$), and (A5) hold, and that the bandwidth $b_{\mathbf{n}}$ satisfies conditions (B1), (B2$'$), and (B3). Then, the conclusions of Lemma 3.1 still hold as $\mathbf{n} \to \infty$.*

Proof. The proof is a slight variation of the argument of Lemma 3.1, and we only describe it shortly. The only significant difference is in the checking of (3.2). Let $\widetilde{U}_1, \ldots, \widetilde{U}_M$ be as in Lemma 3.1. By Lemma 5.3 and Assumption (A4$''$),

$$Q_1 \le C \sum_{i=1}^{M} [\widehat{\mathbf{p}} + (M-i)\widehat{\mathbf{p}} + 1]^{\kappa} \varphi(q) \le C \widehat{\mathbf{p}}^{\kappa} M^{\kappa+1} \varphi(q) \le C(\widehat{\mathbf{n}}^{(\kappa+1)}/\widehat{\mathbf{p}})\varphi(q),$$

which tends to zero by condition (B2$'$); (3.2) follows. □

We then have the following counterpart of Theorem 3.1.

Theorem 3.5 *Let Assumptions (A1), (A2), (A3), (A4$''$), and (A5) hold, with $\varphi(x) = O(x^{-\mu})$ for some $\mu > 2(3+\delta)N/\delta$. Suppose that there exists a sequence of positive integers $q = q_{\mathbf{n}} \to \infty$ such that $q_{\mathbf{n}} = o((\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/(2N)})$ and $\widehat{\mathbf{n}}^{\kappa+1} q^{-\mu-N} \to 0$ as $\mathbf{n} \implies \infty$, and that the bandwidth $b_{\mathbf{n}}$ tends to zero in such a manner that (3.6) is satisfied as $\mathbf{n} \implies \infty$. Then the conclusions of Theorem 3.1 hold.*

Proof. Choose the same values for $p_1, \ldots, p_N$, and q as in the proof of Theorem 3.1. Note that because $\widehat{\mathbf{p}} > q^N$ and $\widehat{\mathbf{n}}^{\kappa+1} q^{-\mu-N} = o(1)$,

$$(\widehat{\mathbf{n}}^{\kappa+1}/\widehat{\mathbf{p}})\varphi(q) \le C \widehat{\mathbf{n}}^{\kappa+1} q^{-N} q^{-\mu} = \widehat{\mathbf{n}}^{\kappa+1} q^{-\mu-N} \to 0$$

as $\mathbf{n} \implies \infty$. The end of the proof is entirely similar to that of Theorem 3.1, with Lemma 3.2 instead of Lemma 3.1. □

Analogues of Theorems 3.2, 3.3, and 3.4 can also be obtained under Assumption (A4$''$); details are omitted for the sake of brevity.

4 Numerical results

In this section, we report the results of a brief Monte Carlo study of the method described in this paper. We mainly consider two models, both in a two-dimensional space ($N = 2$) (writing (i, j) instead of (i_1, i_2) for the sites $\mathbf{i} \in \mathbb{Z}^2$). For the sake of simplicity, $\mathbf{X}$ (written as X) is univariate ($d = 1$).

(a) *Model 1.* Denoting by $\{u_{i,j}, (i,j) \in \mathbb{Z}^2\}$ and $\{e_{i,j}, (i,j) \in \mathbb{Z}^2\}$ two mutually independent *i.i.d.* $\mathcal{N}(0,1)$ white noise processes, let

$$Y_{i,j} = g(X_{i,j}) + u_{i,j}, \qquad \text{with} \qquad g(x) := \frac{1}{3}e^x + \frac{2}{3}e^{-x},$$

where $\{X_{i,j}, (i,j) \in \mathbb{Z}^2\}$ is generated by the spatial autoregression

$$X_{i,j} = \sin(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}) + e_{i,j}.$$

(b) *Model 2.* Denoting again by $\{e_{i,j}, (i,j) \in \mathbb{Z}^2\}$ an *i.i.d.* $\mathcal{N}(0,1)$ white noise process, let $\{Y_{i,j}, (i,j) \in \mathbb{Z}^2\}$ be generated by

$$Y_{i,j} = \sin(Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1}) + e_{i,j},$$

and set

$$X^0_{i,j} := Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1}. \tag{4.1}$$

Then the prediction function $x \mapsto g(x) := E\left[Y_{i,j} | X^0_{i,j} = x\right]$ provides the optimal prediction of $Y_{i,j}$ based on $X^0_{i,j}$ in the sense of minimal mean squared prediction error. Note that, in the spatial context, this optimal prediction function $g(\cdot)$ generally differs from the spatial autoregression function itself (here, $\sin(\cdot)$); see Whittle (1954) for details. Beyond a simple estimation of g, we also will investigate the impact on prediction performance of including additional spatial lags of $Y_{i,j}$ into the definition of $X_{i,j}$.

Data from these two models were simulated over a grid of the form $\{(i,j),\ i = 1, \ldots, 150+m,\ j = 1, \ldots, 150+n\}$. Initial values were set to zero, and the values obtained for i or j less than 150 were discarded, thus allowing for a *warming up period.* For the remaining $m \times n$ data set, we estimated the spatial regression/prediction function using the local linear approach described in this paper. A data-driven choice of the bandwidth in this context wo uld be highly desirable. In view of the lack of theoretical results on this point, we uniformly chose a bandwidth of 0.5 in all our simulations. The simulation results, each with 10 replications, are displayed in Figures 1 and 2 for Models 1 and 2, respectively.

Model 1 is a spatial regression model, with the covariates $X_{i,j}$ forming a nonlinear autoregressive process. Inspection of Figure 1 shows that the estimation of the regression function $g(\cdot)$ is quite good and stable, even for sample sizes as small as $m = 10$ and $n = 20$.

Model 2 is in a spatial autoregressive model, where $Y_{i,j}$ forms a process with nonlinear spatial autoregression function $\sin(\cdot)$. Various definitions of $X_{i,j}$, involving different spatial lags of $Y_{i,j}$, yield various prediction functions, which are shown in Figures 2(a) through 2(f). The results in Figures 2(a) and (b), correspond to $X_{i,j} = X^0_{i,j} := Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1}$, that is, the lags of order ± 1 of $Y_{i,j}$ which also appear in the generating process (4.1). In Figure 2(a), the sample sizes $m = 10$ and $n = 20$ are the same as in Figure 1, but the results (still, for 10 replications) are more dispersed. In Figure 2(b), the sample sizes ($m = 30$ and $n = 40$) are slightly larger, and the results (over 10 replications) seem much more stable. These sample sizes therefore were maintained throughout all subsequent simulations. In Figure 2(c), we chose

$$X^c_{i,j} := +Y_{i,j-2} + Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1} + Y_{i+2,j} + Y_{i,j+2},$$

thus including lagged values of of $Y_{i,j}$ up to order ± 2, in an isotropic way. Nonisotropic choices of $X_{i,j}$ were made in the simulations reported in Figures 2(d) through 2(f): $X^d_{i,j} := Y_{i-1,j} + Y_{i,j-1}$

in Figure 2(d), $X^e_{i,j} := Y_{i+1,j} + Y_{i,j+1}$ in Figure 2(e), and $X^f_{i,j} := Y_{i-2,j} + Y_{i,j-2} + Y_{i-1,j} + Y_{i,j-1}$ in Figure 2(f), respectively.

A more systematic simulation study certainly would be desirable. However, it seems that even in very small samples (see Figure 1), the performance of our method is excellent in pure spatial regression problems (with spatially correlated covariates). Larger samples apparently are required, though, in spatial autoregression models.

5 Appendix: proofs.

5.1 Proof of Lemma 2.2.

The proof of Lemma 2.2 relies on two intermediate results. The first one is a lemma borrowed from Ibragimov and Linnik (1971) or Deo (1973), where we refer to for a proof.

Lemma 5.1 *(i) Suppose that (A1) holds. Let $\mathcal{L}_r(\mathcal{F})$ denote the class of $\mathcal{F}$-measurable random variables ξ satisfying $\|\xi\|_r := (\mathrm{E}|\xi|^r)^{1/r} < \infty$. Let $\mathbf{X} \in \mathcal{L}_r(\mathcal{B}(\mathcal{S}))$ and $Y \in \mathcal{L}_S(\mathcal{B}(\mathcal{S}'))$. Then, for any $1 \leq r, s, h < \infty$ such that $r^{-1} + s^{-1} + h^{-1} = 1$,*

$$\Big|\mathrm{E}[\mathbf{X}Y] - \mathrm{E}[\mathbf{X}]\mathrm{E}[Y]\Big| \leq C\|\mathbf{X}\|_r\|Y\|_s[\alpha(\mathcal{S}, \mathcal{S}')]^{1/h}. \tag{5.1}$$

(ii) If moreover $|\mathbf{X}|$ and $|Y|$ are P*-a.s. bounded, the right-hand side of (5.1) can be replaced with $C\alpha(\mathcal{S}, \mathcal{S}')$.*

The second one is a lemma of independent interest, which plays a crucial role here and in the subsequent sections. For the sake of generality, and in order for this lemma to apply beyond the specific context of this paper, we do not necessarily assume that the mixing coefficient α take the form imposed in Assumption (A4).

Before stating the lemma, let us first introduce some further notation. Let

$$A_{\mathbf{n}} = (\widehat{\mathbf{n}} b^d_{\mathbf{n}})^{-1/2} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \eta_{\mathbf{j}}(x), \qquad u_{\mathbf{n}i\ell} = (\widehat{\mathbf{n}} b^d_{\mathbf{n}})^{-1} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \eta_{i\ell\mathbf{j}}(\mathbf{x}),$$

and

$$\begin{aligned}\mathrm{Var}(A_{\mathbf{n}}) &= (\widehat{\mathbf{n}} b^d_{\mathbf{n}})^{-1} \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \mathrm{E}\left[\Delta^2_{\mathbf{j}}(\mathbf{x})\right] + (\widehat{\mathbf{n}} b^d_{\mathbf{n}})^{-1} \sum\sum_{\{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}} | \exists k\,:\, i_k\neq j_k\}} \mathrm{E}\left[\Delta_{\mathbf{i}}(\mathbf{x})\Delta_{\mathbf{j}}(\mathbf{x})\right] \\ &:= \tilde{I}(\mathbf{x}) + \tilde{R}(\mathbf{x}), \quad \text{say}\end{aligned}$$

(see (3.1) for a definition of $\eta_{\mathbf{j}}$ and $\Delta_{\mathbf{j}}$). For any $\mathbf{c_n} := (c_{\mathbf{n}1}, \cdots, c_{\mathbf{n}N}) \in \mathbb{Z}^N$ with $1 < c_{\mathbf{n}k} < n_k$ for all $k = 1, \cdots, N$, define $\tilde{J}_1(\mathbf{x}) := b_{\mathbf{n}}^{\delta d/(4+\delta)+d} \prod_{k=1}^N (n_k c_{\mathbf{n}k})$ and

$$\tilde{J}_2(\mathbf{x}) := b_{\mathbf{n}}^{2d/(2+\delta)} \widehat{\mathbf{n}} \sum_{k=1}^N \left(\sum_{\substack{|j_s|=1 \\ s=1,\cdots,k-1}}^{n_s} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{\substack{|j_s|=1 \\ s=k+1,\cdots,N}}^{n_s} \{\varphi(j_1, \cdots, j_N)\}^{\delta/(2+\delta)} \right).$$

Lemma 5.2 *Let $\{(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}}); \mathbf{j} \in \mathbb{Z}^N\}$ denote a stationary spatial process with general mixing coefficient*

$$\varphi(\mathbf{j}) = \varphi(j_1, \ldots, j_N) := \sup\Big\{|\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)| : A \in \mathcal{B}(\{Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}\}), B \in \mathcal{B}(\{Y_{\mathbf{i+j}}, \mathbf{X}_{\mathbf{i+j}})\}\Big\},$$

and assume that Assumptions (A1), (A2), and (A5) hold. Then,

$$|\tilde{R}(\mathbf{x})| \le C(\hat{\mathbf{n}} b_{\mathbf{n}}^{d})^{-1} \left[\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x}) \right]. \tag{5.2}$$

If furthermore $\varphi(j_1, \ldots, j_N)$ takes the form $\varphi(\|\mathbf{j}\|)$, then

$$\tilde{J}_2(\mathbf{x}) \le C b_{\mathbf{n}}^{2d/(2+\delta)} \hat{\mathbf{n}} \sum_{k=1}^{N} \left(\sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)} \right). \tag{5.3}$$

Proof. Set $L = L_{\mathbf{n}} = b_{\mathbf{n}}^{-2d/(4+\delta)}$. Defining $Z_{1\mathbf{j}} := Z_{\mathbf{j}} I_{\{|Z_{\mathbf{j}}| \le L\}}$ and $Z_{2\mathbf{j}} := Z_{\mathbf{j}} I_{\{|Z_{\mathbf{j}}| > L\}}$, let

$$\eta_{i\mathbf{j}}(\mathbf{x}) := Z_{i\mathbf{j}} K_c(\mathbf{x} - \mathbf{X}_{\mathbf{j}}) \quad \text{and} \quad \Delta_{i\mathbf{j}}(\mathbf{x}) := \eta_{i\mathbf{j}}(\mathbf{x}) - E\eta_{i\mathbf{j}}(\mathbf{x}), \quad i = 1, 2.$$

Then, $Z_{\mathbf{j}} = Z_{1\mathbf{j}} + Z_{2\mathbf{j}}$, $\quad \Delta_{\mathbf{j}}(\mathbf{x}) = \Delta_{1\mathbf{j}}(\mathbf{x}) + \Delta_{2\mathbf{j}}(\mathbf{x})$, and hence

$$\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x}) = \mathrm{E}\Delta_{1\mathbf{j}}(\mathbf{x})\Delta_{1\mathbf{i}}(\mathbf{x}) + \mathrm{E}\Delta_{1\mathbf{j}}(\mathbf{x})\Delta_{2\mathbf{i}}(\mathbf{x}) + \mathrm{E}\Delta_{2\mathbf{j}}(\mathbf{x})\Delta_{1\mathbf{i}}(\mathbf{x}) + \mathrm{E}\Delta_{2\mathbf{j}}(\mathbf{x})\Delta_{2\mathbf{i}}(\mathbf{x}). \tag{5.4}$$

Firstly, we note that

$$\begin{aligned}
b_{\mathbf{n}}^{-d} |\mathrm{E}\Delta_{1\mathbf{j}}(\mathbf{x})\Delta_{2\mathbf{i}}(\mathbf{x})| &\le \{b_{\mathbf{n}}^{-d}\mathrm{E}\eta_{1\mathbf{j}}^2(\mathbf{x})\}^{1/2}\{b_{\mathbf{n}}^{-d}\mathrm{E}\eta_{2\mathbf{i}}^2(\mathbf{x})\}^{1/2} \\
&\le \{b_{\mathbf{n}}^{-d}\mathrm{E}Z_{1\mathbf{j}}^2 K_{\mathbf{c}}^2((\mathbf{x} - \mathbf{X}_{\mathbf{j}})/b_{\mathbf{n}})\}^{1/2}\{b_{\mathbf{n}}^{-d}\mathrm{E}Z_{2\mathbf{i}}^2 K_{\mathbf{c}}^2((\mathbf{x} - \mathbf{X}_{\mathbf{j}})/b_{\mathbf{n}})\}^{1/2} \\
&\le C\{b_{\mathbf{n}}^{-d}\mathrm{E}|Z_{\mathbf{i}}|^2 I_{\{|Z_{\mathbf{i}}| > L\}} K_{\mathbf{c}}((\mathbf{x} - \mathbf{X_1})/b_{\mathbf{n}})\}^{1/2} \\
&\le C\{L^{-\delta} b_{\mathbf{n}}^{-d}\mathrm{E}|Z_{\mathbf{j}}|^{2+\delta} I_{\{|Z_{\mathbf{j}}| > L\}} K_{\mathbf{c}}((\mathbf{x} - \mathbf{X_1})/b_{\mathbf{n}})\}^{1/2} \\
&\le C L_{\mathbf{n}}^{-\delta/2} = C b_{\mathbf{n}}^{\delta d/(4+\delta)}.
\end{aligned}$$

Similarly,

$$b_{\mathbf{n}}^{-d}|\mathrm{E}\Delta_{2\mathbf{j}}(\mathbf{x})\Delta_{1\mathbf{i}}(\mathbf{x})| \le C L_{\mathbf{n}}^{-\delta/2} = C b_{\mathbf{n}}^{\delta d/(4+\delta)} \quad \text{and} \quad b_{\mathbf{n}}^{-d}|\mathrm{E}\Delta_{2\mathbf{j}}(\mathbf{x})\Delta_{2\mathbf{i}}(\mathbf{x})| \le C b_{\mathbf{n}}^{2\delta d/(4+\delta)}.$$

Next, for $\mathbf{i} = \mathbf{j}$, letting $K_{\mathbf{n}}(\mathbf{x}) := (1/b_{\mathbf{n}}^d) K(\mathbf{x}/b_{\mathbf{n}})$ and $K_{\mathbf{cn}}(\mathbf{x}) := (1/b_{\mathbf{n}}^d) K_{\mathbf{c}}(\mathbf{x}/b_{\mathbf{n}})$,,

$$\begin{aligned}
&b_{\mathbf{n}}^{-d}\mathrm{E}\Delta_{1\mathbf{j}}(\mathbf{x})\Delta_{1\mathbf{i}}(\mathbf{x}) \\
&\quad = b_{\mathbf{n}}^{d}\{\mathrm{E}Z_{1\mathbf{i}}Z_{1\mathbf{j}}K_{\mathbf{cn}}(\mathbf{x} - \mathbf{X}_{\mathbf{i}})K_{\mathbf{cn}}(\mathbf{x} - \mathbf{X}_{\mathbf{j}}) - \mathrm{E}Z_{1\mathbf{i}}K_{\mathbf{cn}}(\mathbf{x} - \mathbf{X}_{\mathbf{i}})\mathrm{E}Z_{1\mathbf{j}}K_{\mathbf{cn}}(\mathbf{x} - \mathbf{X}_{\mathbf{j}})\} \\
&\quad = b_{\mathbf{n}}^{d}\int\!\!\int K_{\mathbf{cn}}(\mathbf{x} - \mathbf{u})K_{\mathbf{cn}}(\mathbf{x} - \mathbf{v})\{g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v})f(\mathbf{u})f(\mathbf{v})\}d\mathbf{u}d\mathbf{v},
\end{aligned}$$

where $g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v}) := \mathrm{E}(Z_{1\mathbf{i}}Z_{1\mathbf{j}}|\mathbf{X}_{\mathbf{i}} = \mathbf{u}, \mathbf{X}_{\mathbf{j}} = \mathbf{v})$, and $g_1^{(1)}(\mathbf{u}) := \mathrm{E}(Z_{1\mathbf{i}}|\mathbf{X}_{\mathbf{i}} = \mathbf{u})$. Since, by definition, $|Z_{1\mathbf{i}}| \le L_{\mathbf{n}}$, we have that $|g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})| \le L_{\mathbf{n}}^2$ and $|g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v})| \le L_{\mathbf{n}}^2$. Thus,

$$\begin{aligned}
&|g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v})f(\mathbf{u})f(\mathbf{v})| \\
&\qquad \le |g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})(f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - f(\mathbf{u})f(\mathbf{v}))| + |(g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v}) - g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v}))f(\mathbf{u})f(\mathbf{v})| \\
&\qquad \le L_{\mathbf{n}}^2|f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - f(\mathbf{u})f(\mathbf{v})| + 2L_{\mathbf{n}}^2 f(\mathbf{u})f(\mathbf{v}).
\end{aligned}$$

It then follows from (A1) and the Lebesgue density theorem (see Chapter 2 of Devroye and Györfi 1985) that

$$\begin{aligned}
b_{\mathbf{n}}^{-d}|\mathrm{E}\Delta_{1\mathbf{j}}(\mathbf{x})\Delta_{1\mathbf{i}}(\mathbf{x})| &\le b_{\mathbf{n}}^{d}\int\int K_{c\mathbf{n}}(\mathbf{x} - \mathbf{u})K_{c\mathbf{n}}(\mathbf{x} - \mathbf{v})L_{\mathbf{n}}^2|f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - f(\mathbf{u})f(\mathbf{v})|\, d\mathbf{u}d\mathbf{v} \\
&\qquad\qquad + b_{\mathbf{n}}^{d}\int\int 2L_{\mathbf{n}}^2 f(\mathbf{u})f(\mathbf{v})\}\, d\mathbf{u}d\mathbf{v} \\
&\le C b_{\mathbf{n}}^{d}\left(L_{\mathbf{n}}^2\{\int K_{\mathbf{cn}}(\mathbf{x} - \mathbf{u})\, d\mathbf{u}\}^2 + 2L_{\mathbf{n}}^2\{\int K_{\mathbf{n}}(\mathbf{x} - \mathbf{u})f(\mathbf{u})\, d\mathbf{u}\}^2\right) \\
&\le C b_{\mathbf{n}}^{d} L_{\mathbf{n}}^2 = C b_{\mathbf{n}}^{\delta d/(4+\delta)}.
\end{aligned} \tag{5.5}$$

Thus, by (5.4)–(5.5),

$$b_{\mathbf{n}}^{-d}|\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})| \le CL_{\mathbf{n}}^{-\delta/2} + Cb_{\mathbf{n}}^{d}L_{\mathbf{n}}^{2} = Cb_{\mathbf{n}}^{\delta d/(4+\delta)}. \tag{5.6}$$

Let $c_{\mathbf{n}} = (c_{\mathbf{n}1}, \cdots, c_{\mathbf{n}N}) \in \mathbb{R}^N$ be a sequence of vectors with positive components. Define

$$\mathcal{S}_1 := \{\mathbf{i} \neq \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : |j_k - i_k| \le c_{\mathbf{n}k}, \text{ for all } k = 1, \cdots, N\},$$

and

$$\mathcal{S}_2 := \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : |j_k - i_k| > c_{\mathbf{n}k}, \text{ for some } k = 1, \cdots, N\}.$$

Clearly, $\mathrm{Card}(\mathcal{S}_1) \le 2^N \hat{\mathbf{n}} \prod_{k=1}^{N} c_{\mathbf{n}k}$. Splitting $\tilde{R}(\mathbf{x})$ into $(\hat{\mathbf{n}} b_{\mathbf{n}}^{d})^{-1}(J_1 + J_2)$, with

$$J_\ell := \sum_{\mathbf{i},\mathbf{j}} \sum_{\in \mathcal{S}_\ell} \mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x}), \quad \ell = 1, 2,$$

it follows from (5.6) that

$$|J_1| \le Cb_{\mathbf{n}}^{\delta d/(4+\delta)+d}\mathrm{Card}(\mathcal{S}_1) \le 2^N Cb_{\mathbf{n}}^{\delta d/(4+\delta)+d}\hat{\mathbf{n}} \prod_{k=1}^{N} c_{\mathbf{n}k}. \tag{5.7}$$

Turning to J_2, we have $|J_2| \le \sum_{\mathbf{i},\mathbf{j}} \sum_{\in \mathcal{S}_2} |\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})|$. Lemma 5.1, with $r = s = 2 + \delta$ and $h = (2+\delta)/\delta$ yields

$$\begin{aligned}
|\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})| &\le C(\mathrm{E}|Z_{\mathbf{i}}K_{\mathbf{c}}((\mathbf{x} - \mathbf{X}_{\mathbf{i}})/b_{\mathbf{n}})|^{2+\delta})^{2/(2+\delta)}\{\varphi(\mathbf{j} - \mathbf{i})\}^{\delta/(2+\delta)} \\
&\le Cb_{\mathbf{n}}^{2d/(2+\delta)}(b_{\mathbf{n}}^{-d}\mathrm{E}|Z_{\mathbf{i}}K_{\mathbf{c}}((\mathbf{x} - \mathbf{X}_{\mathbf{i}})/b_{\mathbf{n}})|^{2+\delta})^{2/(2+\delta)}\{\varphi(\mathbf{j} - \mathbf{i})\}^{\delta/(2+\delta)} \\
&\le Cb_{\mathbf{n}}^{2d/(2+\delta)}\{\varphi(\mathbf{j} - \mathbf{i})\}^{\delta/(2+\delta)}.
\end{aligned} \tag{5.8}$$

Hence,

$$|J_2| \le Cb_{\mathbf{n}}^{2d/(2+\delta)} \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}_2} \sum \{\varphi(\mathbf{j} - \mathbf{i})\}^{\delta/(2+\delta)} := Cb_{\mathbf{n}}^{2d/(2+\delta)}\Sigma_2, \quad \text{say.} \tag{5.9}$$

We now analyze the quantity Σ_2 in detail. For any N-tuple $\mathbf{0} \neq \boldsymbol{\ell} = (\ell_1, \ldots, \ell_N) \in \{0,1\}^N$, set

$\mathcal{S}(\ell_1, \ldots, \ell_N) := \{\mathbf{i}, \mathbf{j} \in I_{\mathbf{n}} : |j_k - i_k| > c_{\mathbf{n}k}$ if $\ell_k = 1$ and $|j_k - i_k| \le c_{\mathbf{n}k}$ if $\ell_k = 0$, $k = 1, \cdots, N\}$,

and

$$V(\ell_1, \ldots, \ell_N) := \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}(\ell_1, \ldots, \ell_N)} \sum \{\varphi(\mathbf{j} - \mathbf{i})\}^{\delta/(2+\delta)}.$$

Then,

$$\Sigma_2 = \sum_{\mathbf{i},\mathbf{j}} \sum_{\in \mathcal{S}_2} \{\varphi(\mathbf{j} - \mathbf{i})\}^{\delta/(2+\delta)} = \sum_{\mathbf{0} \neq \boldsymbol{\ell} \in \{0,1\}^N} V(\ell_1, \ldots, \ell_N). \tag{5.10}$$

Without loss of generality, consider $V(1, 0, \ldots, 0)$. Because $\sum_{|i_k - j_k| > c_{\mathbf{n}k}} \ldots$ decomposes into $\sum_{i_k=1}^{n_k - c_{\mathbf{n}k} - 1} \sum_{j_k = i_k + c_{\mathbf{n}k} + 1}^{n_k} \ldots + \sum_{j_k=1}^{n_k - c_{\mathbf{n}k} - 1} \sum_{i_k = j_k + c_{\mathbf{n}k} + 1}^{n_k} \ldots$, and $\sum_{|i_k - j_k| \le c_{\mathbf{n}k}} \ldots$ into $\sum_{i_k=1}^{n_k - c_{\mathbf{n}k}} \sum_{j_k = i_k + 1}^{i_k + c_{\mathbf{n}k}} \ldots + \sum_{j_k=1}^{n_k - c_{\mathbf{n}k}} \sum_{i_k = j_k + 1}^{j_k + c_{\mathbf{n}k}} \ldots$, we have

$$
\begin{aligned}
V(1,0,\ldots,0) &= \sum_{|i_1-j_1|>c_{\mathbf{n}1}} \sum_{|i_2-j_2|\le c_{\mathbf{n}2}} \cdots \sum_{|i_N-j_N|\le c_{\mathbf{n}N}} \{\varphi(j_1-i_1,\ldots,j_N-i_N)\}^{\delta/(2+\delta)} \\
&\le \widehat{\mathbf{n}} \left\{ \sum_{j_1=c_{\mathbf{n}1}}^{n_1} + \sum_{-j_1=c_{\mathbf{n}1}}^{n_1} \right\} \left\{ \sum_{j_2=1}^{c_{\mathbf{n}2}} + \sum_{-j_2=1}^{c_{\mathbf{n}2}} \right\} \cdots \left\{ \sum_{j_N=1}^{c_{\mathbf{n}N}} + \sum_{-j_N=1}^{c_{\mathbf{n}N}} \right\} \{\varphi(j_1,\ldots,j_N)\}^{\delta/(2+\delta)} \\
&\le \widehat{\mathbf{n}} \sum_{|j_1|=c_{\mathbf{n}1}}^{n_1} \sum_{|j_2|=1}^{c_{\mathbf{n}2}} \cdots \sum_{|j_N|=1}^{c_{\mathbf{n}N}} \{\varphi(j_1,\ldots,j_N)\}^{\delta/(2+\delta)} \\
&\le \widehat{\mathbf{n}} \sum_{|j_1|=c_{\mathbf{n}1}}^{n_1} \sum_{|j_2|=1}^{n_2} \cdots \sum_{|j_N|=1}^{n_N} \{\varphi(j_1,\ldots,j_N)\}^{\delta/(2+\delta)}.
\end{aligned}
$$

More generally,

$$
V(\ell_1,\ell_2,\ldots,\ell_N) \le \widehat{\mathbf{n}} \sum_{|j_1|} \cdots \sum_{|j_k|} \cdots \sum_{|j_N|} \{\varphi(j_1,\ldots,j_N)\}^{\delta/(2+\delta)}, \quad , \tag{5.11}
$$

where the sums $\sum_{|j_k|}$ run over all values of j_k such that $1 \le |j_k| \le n_k$ if $\ell_1 = 0$, such that $c_{\mathbf{n}1} \le |j_k| \le n_k$ if $\ell_1 = 1$. Since the summands are nonnegative, for $1 \le c_{\mathbf{n}k} \le n_k$, we have $\sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \cdots \le \sum_{|j_k|=1}^{n_k} \cdots$, and (5.9), (5.10), and (5.11) imply

$$
|J_2| \le C b_{\mathbf{n}}^{2d/(2+\delta)} \widehat{\mathbf{n}} \sum_{k=1}^{N} \left(\sum_{|j_1|=1}^{n_1} \cdots \sum_{|j_{k-1}|=1}^{n_{k-1}} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{|j_{k+1}|=1}^{n_{k+1}} \cdots \sum_{|j_N|=1}^{n_N} \{\varphi(j_1,\ldots,j_N)\}^{\delta/(2+\delta)} \right). \tag{5.12}
$$

Thus (5.2) is a consequence of (5.7) and (5.12). If furthermore $\varphi(j_1,\ldots,j_N)$ depends on $\|\mathbf{j}\|$ only, then,

$$
\begin{aligned}
&\sum_{|j_1|=1}^{n_1} \cdots \sum_{|j_{k-1}|=1}^{n_{k-1}} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{|j_{k+1}|=1}^{n_{k+1}} \cdots \sum_{|j_N|=1}^{n_N} \{\varphi(\|\mathbf{j}\|)\}^{\delta/(2+\delta)} \\
&\le \sum_{|j_1|=1}^{n_1} \cdots \sum_{|j_{k-1}|=1}^{n_{k-1}} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{|j_{k+1}|=1}^{n_{k+1}} \cdots \sum_{|j_{N-1}|=1}^{n_{N-1}} \sum_{t^2=j_1^2+\cdots+j_{N-1}^2+1}^{j_1^2+\cdots+j_{N-1}^2+n_N^2} \{\varphi(t)\}^{\delta/(2+\delta)} \\
&\le \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} \sum_{|j_1|=1}^{t} \cdots \sum_{|j_{N-1}|=1}^{t} \{\varphi(t)\}^{\delta/(2+\delta)} \le \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)};
\end{aligned}
$$

(5.3) follows. □

Proof of Lemma 2.2. Observe that

$$
\begin{aligned}
\tilde{I}(\mathbf{x}) &= b_{\mathbf{n}}^{-d} \mathrm{E}\Delta_{\mathbf{j}}^2(\mathbf{x}) = b_{\mathbf{n}}^{-d}[\mathrm{E}\eta_{\mathbf{j}}^2 - (\mathrm{E}\eta_{\mathbf{j}})^2] \\
&= b_{\mathbf{n}}^{-d} \left[\mathrm{E}Z_{\mathbf{j}}^2 K_{\mathbf{c}}^2((\mathbf{x}-\mathbf{X}_{\mathbf{j}})/b_{\mathbf{n}}) - \{\mathrm{E}Z_{\mathbf{j}} K_{\mathbf{c}}((\mathbf{x}-\mathbf{X}_{\mathbf{j}})/b_{\mathbf{n}})\}^2 \right].
\end{aligned} \tag{5.13}
$$

Under Assumption (A5), by the Lebesgue density Theorem,

$$
\lim_{\mathbf{n}\to\infty} \int_{\mathbb{R}^d} b_{\mathbf{n}}^{-d} \mathrm{E}[Z_{\mathbf{j}}^2 | \mathbf{X}_{\mathbf{j}} = \mathbf{u}] K_{\mathbf{c}}^2((\mathbf{x}-\mathbf{u})/b_{\mathbf{n}}) f(\mathbf{u})\, d\mathbf{u} = g^{(2)}(\mathbf{x}) f(\mathbf{x}) \int_{\mathbb{R}^d} K_{\mathbf{c}}^2(\mathbf{u})\, d\mathbf{u},
$$

$$\lim_{\mathbf{n}\to\infty}\int_{\mathbb{R}^d} b_{\mathbf{n}}^{-d}\mathrm{E}[Z_{\mathbf{j}}|\mathbf{X}_{\mathbf{j}}=\mathbf{u}]K_{\mathbf{c}}((\mathbf{x}-\mathbf{u})/b_{\mathbf{n}})f(\mathbf{u})\,d\mathbf{u} = g^{(1)}(\mathbf{x})f(\mathbf{x})\int_{\mathbb{R}^d} K(\mathbf{u})\,d\mathbf{u},$$

where $g^{(i)}(\mathbf{x}) := \mathrm{E}[Z_{\mathbf{j}}^i|\mathbf{X}_{\mathbf{j}}=\mathbf{x}]$ for $i = 1, 2$. It is easily seen that $b_{\mathbf{n}}^{-d}\{\mathrm{E}Z_{\mathbf{j}}K_{\mathbf{c}}((\mathbf{x}-\mathbf{X}_{\mathbf{j}})/b_{\mathbf{n}})\}^2 \to 0$. Thus, from (5.13),

$$\lim_{\mathbf{n}\to\infty}\tilde{I}(\mathbf{x}) = g^{(2)}(\mathbf{x})f(\mathbf{x})\int_{\mathbb{R}^d} K_{\mathbf{c}}^2(\mathbf{u})\,d\mathbf{u}, \tag{5.14}$$

where $g^{(2)}(\mathbf{x}) = \mathrm{E}\{Z_{\mathbf{j}}^2|\mathbf{X}_{\mathbf{j}}=\mathbf{x}\} = E\{(Y_{\mathbf{j}}-g(\mathbf{x}))^2|\mathbf{X}_{\mathbf{j}}=\mathbf{x}\} = \mathrm{Var}\{Y_{\mathbf{j}}|\mathbf{X}_{\mathbf{j}}=\mathbf{x}\}$.

Let $c_{\mathbf{n}k}^a := b_{\mathbf{n}}^{-\delta d/(2+\delta)} \to \infty$. Clearly, $c_{\mathbf{n}k} < n_k$ because $n_k b_{\mathbf{n}}^{\delta d/[(2+\delta)a]} > 1$ for all k. Apply Lemma 5.2. Since $N/[(2+\delta)a] < 1/(4+\delta)$ due to the fact that $a > (4+\delta)N/(2+\delta)$, and

$$(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1}\tilde{J}_2 \le C\sum_{k=1}^N\left(c_{\mathbf{n}k}^a\sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1}\{\varphi(t)\}^{\delta/(2+\delta)}\right) \to 0 \tag{5.15}$$

because $c_{\mathbf{n}k}\to\infty$, (5.3) and Assumption (A4) imply that

$$(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1}\tilde{J}_1 \le Cb_{\mathbf{n}}^{\delta d/(4+\delta)}c_{\mathbf{n}1}\dots c_{\mathbf{n}N} = Cb_{\mathbf{n}}^{\delta d/(4+\delta)}b_{\mathbf{n}}^{-\delta dN/[(2+\delta)a]} \to 0,$$

hence, by (5.2), that

$$|\tilde{R}(x)| = (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1}|\tilde{J}(x)| \le C(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1}(\tilde{J}_1+\tilde{J}_2) \to 0. \tag{5.16}$$

Finally, (2.7) follows from (5.14) and (5.16), which completes the proof of Lemma 3.2. □

5.2 Proof of Lemma 2.3.

From (2.5) and the definition of $A_{\mathbf{n}}$ (recall that $a_0 = g(\mathbf{x})$, $\mathbf{a}_1 = g'(\mathbf{x})$),

$$\begin{aligned}
\mathrm{E}[A_{\mathbf{n}}] &= (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\ b_{\mathbf{n}}^{-d}\mathrm{E}Z_{\mathbf{j}}K_{\mathbf{c}}\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right) = (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\ b_{\mathbf{n}}^{-d}\mathrm{E}(Y_{\mathbf{j}}-a_0-\mathbf{a}_1^\tau(\mathbf{X}_{\mathbf{j}}-\mathbf{x}))K_{\mathbf{c}}\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)\\
&= (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\ b_{\mathbf{n}}^{-d}\mathrm{E}(g(\mathbf{X}_{\mathbf{j}})-a_0-\mathbf{a}_1^\tau(\mathbf{X}_{\mathbf{j}}-\mathbf{x}))K_{\mathbf{c}}\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)\\
&= (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}\ b_{\mathbf{n}}^{-d}\mathrm{E}(\mathbf{X}_{\mathbf{j}}-\mathbf{x})^\tau g''(\mathbf{x}+\boldsymbol{\xi}(\mathbf{X}_{\mathbf{j}}-\mathbf{x}))(\mathbf{X}_{\mathbf{j}}-\mathbf{x})K_{\mathbf{c}}\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right) \quad (\text{where } |\boldsymbol{\xi}|<1)\\
&= (\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}b_{\mathbf{n}}^2\ b_{\mathbf{n}}^{-d}\ \mathrm{tr}\mathrm{E}\left[g''(\mathbf{x}+\boldsymbol{\xi}(\mathbf{X}_{\mathbf{j}}-\mathbf{x}))\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right)^\tau\right]K_{\mathbf{c}}\left(\frac{\mathbf{X}_{\mathbf{j}}-\mathbf{x}}{b_{\mathbf{n}}}\right);
\end{aligned}$$

the lemma follows via Assumption (A3). □

5.3 Proof of Lemma 3.1.

Before turning to the end of the proof of Lemma 3.1, we establish the following preliminary lemma, which significantly reinforces Lemma 3.1 in Tran (1990).

Lemma 5.3 *Let the spatial process $\{Y_{\mathbf{i}},\mathbf{X}_{\mathbf{i}}\}$ satisfy the mixing property (2.1), and denote by $\widetilde{U}_j$, $j = 1,\dots,M$, an M-tuple of measurable functions such that $\widetilde{U}_j$ be measurable with respect to $\{(Y_{\mathbf{i}},\mathbf{X}_{\mathbf{i}}), \mathbf{i}\in\tilde{\mathcal{I}}_j\}$, where $\tilde{\mathcal{I}}_j \subset \mathcal{I}_{\mathbf{n}}$. If $\mathrm{Card}(\tilde{\mathcal{I}}_j) \le p$ and $d(\tilde{\mathcal{I}}_\ell,\tilde{\mathcal{I}}_j) \ge q$ for any $\ell \ne j$, then*

$$\left|\mathrm{E}\left[\exp\{iu\sum_{j=1}^M\widetilde{U}_j\}\right] - \prod_{j=1}^M\mathrm{E}\left[\exp\{iu\widetilde{U}_j\}\right]\right| \le C\sum_{j=1}^{M-1}\psi(p,\,(M-j)p)\varphi(q)$$

where $i = \sqrt{-1}$.

Proof. Let $a_j := \exp\{iu\widetilde{U}_j\}$. Then

$$\begin{aligned}\mathrm{E}\,[a_1...a_M] - \mathrm{E}\,[a_1]\cdots\mathrm{E}\,[a_M] &= \mathrm{E}\,[a_1\cdots a_M] - \mathrm{E}\,[a_1]\,\mathrm{E}\,[a_2\cdots a_M] \\ &+\mathrm{E}\,[a_1]\,\{\mathrm{E}\,[a_2\cdots a_M] - \mathrm{E}\,[a_2]\,\mathrm{E}\,[a_3\cdots a_M]\} + \cdots \\ &+\mathrm{E}\,[a_1]\,\mathrm{E}\,[a_2]\cdots\mathrm{E}\,[a_{M-2}]\,\{\mathrm{E}\,[a_{M-1}a_M] - \mathrm{E}\,[a_{M-1}]\,\mathrm{E}\,[a_M]\}\,.\end{aligned}$$

Since $|\mathrm{E}\,[a_i]\,| \le 1$,

$$\begin{aligned}|\mathrm{E}\,[a_1...a_M] - \mathrm{E}\,[a_1]\cdots\mathrm{E}\,[a_M]\,| &\le |\mathrm{E}\,[a_1\cdots a_M] - \mathrm{E}\,[a_1]\,\mathrm{E}\,[a_2\cdots a_M]\,| \\ &+|\mathrm{E}\,[a_2\cdots a_M] - \mathrm{E}\,[a_2]\,\mathrm{E}\,[a_3\cdots a_M]\,| + \cdots \\ &+|\mathrm{E}\,[a_{M-1}a_M] - \mathrm{E}\,[a_{M-1}]\,\mathrm{E}\,[a_M]\,|.\end{aligned}$$

Note that $d(I_\ell, I_j) \ge q$ for any $\ell \ne j$. The lemma then follows by applying Lemma 5.1(ii) to each term on the right hand side. $\square$

Proof of Lemma 3.1 (continued). In order to complete the proof of Lemma 3.1, we still have to prove (3.2)-(3.5).

Proof of (3.2). Ranking the random variables $U(1, \mathbf{n}, \mathbf{x}, \mathbf{j})$ in an arbitrary manner, refer to them as $\widetilde{U}_1, \dots, \widetilde{U}_M$. Note that $M = \prod_{k=1}^N r_k = \widehat{\mathbf{n}}\{\prod_{k=1}^N (p_k + q)\}^{-1} \le \widehat{\mathbf{n}}/p$, where $p = \prod_{k=1}^N p_k$. Let

$$\mathcal{I}(1, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \{\mathbf{i} : j_k(p_k + q) + 1 \le i_k \le j_k(p_k + q) + p_k,\ k = 1, \cdots, N\}\,.$$

The distance between two distinct sets $\mathcal{I}(1, \mathbf{n}, \mathbf{x}, \mathbf{j})$ and $\mathcal{I}(1, \mathbf{n}, \mathbf{x}, \mathbf{j}')$ is at least q. Clearly, $\mathcal{I}(1, \mathbf{n}, \mathbf{x}, \mathbf{j})$ is the set of sites involved in $U(1, \mathbf{n}, \mathbf{x}, \mathbf{j})$. As for the set of sites $\widetilde{\mathcal{I}}_j$ associated with $\widetilde{U}_j$, it contains p sites. Hence, in view of Lemma 5.3 and Assumption (A4$'$),

$$Q_1 \le C \sum_{k=1}^{M-1} \min\{p,\ (M-k)p\}\ \varphi(q) \le CMp\ \varphi(q) \le C\widehat{\mathbf{n}}\ \varphi(q),$$

which tends to zero by condition (B2).

Proof of (3.3). In order to prove (3.3), it is enough to show that

$$\widehat{\mathbf{n}}^{-1}\mathrm{E}[T^2(\mathbf{n}, \mathbf{x}, i)] \to 0 \quad \text{for any } 2 \le i \le 2^N.$$

Without loss of generality, consider $\mathrm{E}[T^2(\mathbf{n}, \mathbf{x}, 2)]$. Ranking the random variables $U(2, \mathbf{n}, \mathbf{x}, \mathbf{j})$ in an arbitrary manner, refer to them as $\widehat{U}_1, \dots \widehat{U}_M$. We have

$$\mathrm{E}[T^2(\mathbf{n}, \mathbf{x}, 2)] = \sum_{i=1}^M \mathrm{Var}(\widehat{U}_i) + 2 \sum_{1 \le i < j \le M} \mathrm{Cov}(\widehat{U}_i, \widehat{U}_j) := \widehat{V}_1 + \widehat{V}_2, \quad \text{say.} \tag{5.17}$$

Since $\mathbf{X}_n$ is stationary (recall that $\zeta_{\mathbf{nj}}(\mathbf{x}) := b_{\mathbf{n}}^{-d/2}\Delta_{\mathbf{j}}(\mathbf{x})$),

$$\mathrm{Var}(\widehat{U}_i) = \mathrm{E}\left[\left(\sum_{\substack{i_k=1 \\ k=1,\dots,N-1}}^{p_k} \sum_{i_N=1}^{q} \zeta_{\mathbf{ni}}(\mathbf{x})\right)^2\right] + \sum_{\mathbf{i} \ne \mathbf{j} \in \mathcal{J}} \mathrm{E}\,[\zeta_{\mathbf{nj}}(\mathbf{x})\zeta_{\mathbf{ni}}(\mathbf{x})] := \widehat{V}_{11} + \widehat{V}_{12},$$

where $\mathcal{J} = \mathcal{J}(\mathbf{p}, q) := \{\mathbf{i}, \mathbf{j} : 1 \le i_k, j_k \le p_k, k = 1, \ldots, N-1, \text{ and } 1 \le i_N, j_N \le q\}$. From (5.13) and the Lebesgue density theorem (see Chapter 2 of Devroye and Györfi 1985),

$$\widehat{V}_{11} = \left(\prod_{k=1}^{N-1} p_k\right) q \ \mathrm{Var}\{\zeta_{\mathbf{ni}}(\mathbf{x})\} = \left(\prod_{k=1}^{N-1} p_k\right) q \ \{b_{\mathbf{n}}^{-d} \mathrm{E}\Delta_{\mathbf{i}}^2(\mathbf{x})\} \le C \left(\prod_{k=1}^{N-1} p_k\right) q \ .$$

Thus, applying Lemma 5.2 with $n_k = p_k$, $k = 1, \cdots, N-1$, and $n_N = q$ yields

$$\begin{aligned}
\widehat{V}_{12} &= b_{\mathbf{n}}^{-d} \sum_{\mathbf{i} \ne \mathbf{j} \in \mathcal{J}} \mathrm{E}\left[\Delta_{\mathbf{j}}(\mathbf{x}) \Delta_{\mathbf{i}}(\mathbf{x})\right] \\
&\le C b_{\mathbf{n}}^{-d} \left[b_{\mathbf{n}}^{\delta d/(4+\delta)+d} \left(\prod_{k=1}^{N-1} p_k c_{\mathbf{n}k}\right) q c_{\mathbf{n}N} + b_{\mathbf{n}}^{2d/(2+\delta)} \left(\prod_{k=1}^{N-1} p_k\right) q \sum_{k=1}^{N} \left(\sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)}\right)\right] \\
&= C \left(\prod_{k=1}^{N-1} p_k\right) q \left[b_{\mathbf{n}}^{\delta d/(4+\delta)} \left(\prod_{k=1}^{N} c_{\mathbf{n}k}\right) + b_{\mathbf{n}}^{-\delta d/(2+\delta)} \sum_{k=1}^{N} \left(\sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)}\right)\right] \\
&:= C \left(\prod_{k=1}^{N-1} p_k\right) q \ \pi_{\mathbf{n}}.
\end{aligned}$$

It follows that

$$\widehat{\mathbf{n}}^{-1} \widehat{V}_1 = \widehat{\mathbf{n}}^{-1} M \left(\hat{V}_{11} + \hat{V}_{12}\right) \le \widehat{\mathbf{n}}^{-1} MC \left(\prod_{k=1}^{N-1} p_k\right) q [1 + \pi_{\mathbf{n}}] \le C(q/p_N)[1 + \pi_{\mathbf{n}}]. \tag{5.18}$$

Set

$$\begin{aligned}
\mathcal{I}(2, n, \mathbf{x}, \mathbf{j}) := \{\mathbf{i} \ : & j_k(p_k + q) + 1 \le i_k \le j_k(p_k + q) + p_k, \ 1 \le k \le N-1, \\
& j_N(p_N + q) + p_N + 1 \le i_N \le (j_N + 1)(p_N + q)\}.
\end{aligned}$$

Then $U(2, \mathbf{n}, \mathbf{x}, \mathbf{j}) = \sum_{\mathbf{i} \in \mathcal{I}(2,n,\mathbf{x},\mathbf{j})} \zeta_{\mathbf{ni}}$. Since $p_k > q$, if $\mathbf{j}$ and $\mathbf{j}'$ belong to two distinct sets $I(2, \mathbf{n}, \mathbf{x}, \mathbf{j})$ and $I(2, \mathbf{n}, \mathbf{x}, \mathbf{j}')$, then $||\mathbf{j} - \mathbf{j}'|| > q$. In view of (5.8) and (5.17), we obtain

$$\begin{aligned}
|\widehat{V}_2| &\le C \sum_{\{\mathbf{i},\mathbf{j}: \|\mathbf{i}-\mathbf{j}\| \ge q, \ 1 \le i_k, j_k \le n_k\}} \sum |\mathrm{E}\left[\zeta_{\mathbf{ni}}(\mathbf{x}) \zeta_{\mathbf{nj}}(\mathbf{x})\right]| \\
&\le C b_{\mathbf{n}}^{-d} \sum_{\{\mathbf{i},\mathbf{j}: \|\mathbf{i}-\mathbf{j}\| \ge q, \ 1 \le i_k, j_k \le n_k\}} \sum |\mathrm{E}\left[\Delta_{\mathbf{ni}}(\mathbf{x}) \Delta_{\mathbf{nj}}(\mathbf{x})\right]| \\
&\le C b_{\mathbf{n}}^{-d} \sum_{\{\mathbf{i},\mathbf{j}: \|\mathbf{i}-\mathbf{j}\| \ge q, \ 1 \le i_k, j_k \le n_k\}} \sum b_{\mathbf{n}}^{2d/(2+\delta)} \ \{\varphi(\|\mathbf{j} - \mathbf{i}\|)\}^{\delta/(2+\delta)} \\
&\le C b_{\mathbf{n}}^{-\delta d/(2+\delta)} \left(\prod_{k=1}^{N} n_k\right) \left(\sum_{t=q}^{\|\mathbf{n}\|} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)}\right).
\end{aligned} \tag{5.19}$$

Take $c_{\mathbf{n}k}^a = b_{\mathbf{n}}^{-\delta d/(2+\delta)} \to \infty$. Condition (B3) implies that $q b_{\mathbf{n}}^{\delta d/[a(2+\delta)]} > 1$, so that $c_{\mathbf{n}k} < q \le p_k$. Then, as proved in (5.15) and (5.16), it follows from Assumption (A4) that $\pi_{\mathbf{n}} \to 0$. Thus, from (5.17), (5.18), and (5.19),

$$\widehat{\mathbf{n}}^{-1} \mathrm{E}[T^2(\mathbf{n}, \mathbf{x}, 2)] \le C(q/p_N)[1 + \pi_{\mathbf{n}}] + C b_{\mathbf{n}}^{-\delta d/(2+\delta)} \left(\sum_{t=q}^{\infty} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)}\right),$$

which tends to zero by $q/p_N \to 0$ and condition (B3); (3.3) follows.

Proof of (3.4). Let $S'_{\mathbf{n}} := T(\mathbf{n}, \mathbf{x}, 1)$ and $S''_{\mathbf{n}} := \sum_{i=2}^{2^N} T(\mathbf{n}, \mathbf{x}, i)$. Then $S'_{\mathbf{n}}$ is a sum of $Y_{\mathbf{j}}$'s over the "large" blocks, $S''_{\mathbf{n}}$ over the "small" ones. Lemma 3.2 implies $\widehat{\mathbf{n}}^{-1}\mathrm{E}\left[|S_{\mathbf{n}}|^2\right] \to \sigma^2$. This, combined with (3.3), entails $\widehat{\mathbf{n}}^{-1}\mathrm{E}\left[|S'_{\mathbf{n}}|^2\right] \to \sigma^2$. Now,

$$\widehat{\mathbf{n}}^{-1}\mathrm{E}\left[|S'_{\mathbf{n}}|^2\right] = \widehat{\mathbf{n}}^{-1} \sum_{\substack{j_k=0 \\ k=1,\ldots,N}}^{r_k-1} \mathrm{E}[U^2(1, \mathbf{n}, \mathbf{x}, \mathbf{j})] + \widehat{\mathbf{n}}^{-1} \sum_{\mathbf{i} \neq \mathbf{j} \in \mathcal{J}^*} \mathrm{Cov}\left(U(1, \mathbf{n}, \mathbf{x}, \mathbf{j}), U(1, \mathbf{n}, \mathbf{x}, \mathbf{i})\right). \quad (5.20)$$

where $\mathcal{J}^* = \mathcal{J}^*(\mathbf{p}, q) := \{\mathbf{i}, \mathbf{j} : 1 \le i_k, j_k \le r_k - 1, k = 1, \ldots, N\}$. Observe that (3.4) follows from (5.20) if the last sum in the right-hand side of (5.20) tends to zero as $\mathbf{n} \to \infty$. Using the same argument as in the derivation of the bound (5.17) for $\widehat{V}_2$, this sum can be bounded by

$$C b_{\mathbf{n}}^{-\delta d/(2+\delta)} \sum_{\|\mathbf{i}\|>q} \sum_{\substack{i_k=1 \\ k=1,\ldots,N}}^{n_k-1} \{\varphi(\|\mathbf{i}\|)\}^{\delta/(2+\delta)} \le C b_{\mathbf{n}}^{-\delta d/(2+\delta)} \left(\sum_{t=q}^{\infty} t^{N-1} \{\varphi(t)\}^{\delta/(2+\delta)} \right),$$

which tends to zero by condition (B3).

Proof of (3.5). We need a trucation argument because $Z_{\mathbf{i}}$ is not necessarily bounded. Set $Z_{\mathbf{i}}^L := Z_{\mathbf{i}} I_{\{|Z_{\mathbf{i}}| \le L\}}$, $\eta_{\mathbf{i}}^L := Z_{\mathbf{i}}^L K_{\mathbf{c}}((\mathbf{X_i} - \mathbf{x})/b_{\mathbf{n}})$, $\Delta_{\mathbf{i}}^L := \eta_{\mathbf{i}}^L - \mathrm{E}\eta_{\mathbf{i}}^L$, $\zeta_{\mathbf{ni}}^L := b_{\mathbf{n}}^{-d/2} \Delta_{\mathbf{i}}^L$, where L is a fixed positive constant, and define $U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j}) := \sum_{\mathbf{i} \in \mathcal{I}(1,\mathbf{n},\mathbf{x},\mathbf{j})} \zeta_{\mathbf{ni}}^L$. Put

$$Q_4^L := \widehat{\mathbf{n}}^{-1} \sum_{\substack{j_k=0 \\ k=1,\ldots,N}}^{r_k-1} \mathrm{E}\left[(U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j}))^2 I\{|U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j})| > \varepsilon\sigma\widehat{\mathbf{n}}^{1/2}\}\right].$$

Clearly, $|\zeta_{\mathbf{ni}}^L| \le CLb_{\mathbf{n}}^{-d/2}$. Therefore $|U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j})| < CLpb_{\mathbf{n}}^{-d/2}$. Hence

$$Q_4^L \le C\widehat{\mathbf{p}}^2 b_{\mathbf{n}}^{-d} \widehat{\mathbf{n}}^{-1} \sum_{\substack{j_k=0 \\ k=1,\ldots,N}}^{r_k-1} \mathrm{P}[U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j}) > \varepsilon\sigma\widehat{\mathbf{n}}^{1/2}].$$

Now, $U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j})/(\sigma\widehat{\mathbf{n}}^{1/2}) \le C\widehat{\mathbf{p}}(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{-1/2} \to 0$, since $\widehat{\mathbf{p}} = [(\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2}/s_{\mathbf{n}}]$, where $s_{\mathbf{n}} \to \infty$. Thus $\mathrm{P}[U^L(1, \mathbf{n}, \mathbf{x}, \mathbf{j}) > \varepsilon\sigma\widehat{\mathbf{n}}^{1/2}] = 0$ at all $\mathbf{j}$ for sufficiently large $\widehat{\mathbf{n}}$. Thus $Q_4^L = 0$ for large $\widehat{\mathbf{n}}$, and (3.5) holds for the truncated variables. Hence,

$$\widehat{\mathbf{n}}^{-1/2} S_{\mathbf{n}}^L := \widehat{\mathbf{n}}^{-1/2} \sum_{\substack{j_k=1 \\ k=1,\ldots,N}}^{n_k} \zeta_{\mathbf{nj}}^L \xrightarrow{\mathcal{L}} N(0, \sigma_L^2), \quad (5.21)$$

where $\sigma_L^2 := \mathrm{Var}(Z_{\mathbf{i}}^L | \mathbf{X_i} = \mathbf{x}) f(\mathbf{x}) \int K_{\mathbf{c}}^2(\mathbf{u}) d\mathbf{u}$.

Defining $S_{\mathbf{n}}^{L*} := \sum_{\substack{j_k=1 \\ k=1,\ldots,N}}^{n_k} (\zeta_{\mathbf{nj}} - \zeta_{\mathbf{nj}}^L)$, we have $S_{\mathbf{n}} = S_{\mathbf{n}}^L + S_{\mathbf{n}}^{L*}$. Note that

$$\begin{aligned} \left|\mathrm{E}\left[\exp(iuS_{\mathbf{n}}/\widehat{\mathbf{n}}^{1/2})\right] - \exp(-u^2\sigma^2/2)\right| &\le \left|\mathrm{E}[\exp(iuS_{\mathbf{n}}^L/\widehat{\mathbf{n}}^{1/2}) - \exp(-u^2\sigma_L^2/2)]\ \exp(iuS_{\mathbf{n}}^{L*}/\widehat{\mathbf{n}}^{1/2})\right| \\ &+ \left|\mathrm{E}[\exp(iuS_{\mathbf{n}}^{L*}/\widehat{\mathbf{n}}^{1/2}) - 1]\ \exp(-u^2\sigma_L^2/2)\right| + \left|\exp(-u^2\sigma_L^2/2) - \exp(-u^2\sigma^2/2)\right| \\ &= E_1 + E_2 + E_3, \quad \text{say.} \end{aligned}$$

Letting $\mathbf{n} \to \infty$, E_1 tends to 0 by (5.21) and the dominated convergence theorem. Letting L go to infinity, the dominated convergence theorem also implies that $\sigma_L^2 := \mathrm{Var}(Z_{\mathbf{i}}^L|\mathbf{X_i} = \mathbf{x})f(\mathbf{x})\int K_{\mathbf{c}}^2(\mathbf{u})d\mathbf{u}$ converges to

$$\mathrm{Var}(Z_{\mathbf{i}}|\mathbf{X_i} = \mathbf{x})f(\mathbf{x})\int K_{\mathbf{c}}^2(\mathbf{u})d\mathbf{u} = \mathrm{Var}(Y_{\mathbf{i}}|\mathbf{X_i} = \mathbf{x})f(\mathbf{x})\int K_{\mathbf{c}}^2(\mathbf{u})d\mathbf{u} := \sigma^2,$$

and hence that E_3 tends to 0. Finally, in order to prove that E_2 also tends to 0, it suffices to show that $S_{\mathbf{n}}^{L*}/\widehat{\mathbf{n}}^{1/2} \to 0$ in probability as first $\mathbf{n} \to \infty$ and then $L \to \infty$, which in turn would follow if we could show that

$$\mathrm{E}\left[(S_{\mathbf{n}}^{L*}/\widehat{\mathbf{n}}^{1/2})^2\right] \to \mathrm{Var}(|Z_{\mathbf{i}}|I_{\{|Z_{\mathbf{i}}|>L\}}|\mathbf{X_i} = \mathbf{x})f(\mathbf{x})\int K_{\mathbf{c}}^2(\mathbf{u})d\mathbf{u} \quad \text{as } \mathbf{n} \to \infty.$$

This follows along the same lines as Lemma 3.2. The proof of Lemma 3.1 is thus complete. □

References

[1] Anselin, L. and Florax, R.J.G.M. (1995). *New Directions in Spatial Econometrics.* Berlin: Springer.

[2] Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B* **36**, 192-225.

[3] Boente, G. and Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *Journal of Multivariate Analysis* **25**, 90–99.

[4] Bosq, D. (1989). Estimation et prévision nonparamétrique d'un processus stationnaire. *Comptes Rendus de l'Académie des Sciences de Paris* I **308**, 453–456.

[5] Carbon, M., Hallin, M., and Tran, L.T. (1996). Kernel density estimation for random fields: the L_1 theory, *Journal of Nonparametric Statistics* **6**, 157–170.

[6] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836.

[7] Cleveland, W.S. and Loader, C. (1996). Smoothing by local regression: principles and methods. In: *Statistical Theory and Computational Aspects of Smoothing*, W. Härdle and M.G. Schimek, Eds. Heidelberg: Physica-Verlag.

[8] Cressie, N.A.C. (1991). *Statistics for Spatial Data.* New York: John Wiley.

[9] Deo, C. M. (1973). A note on empirical processes of strong mixing sequences. *Annals of Probability* **1**, 870–875.

[10] Devroye, L., and Györfi, L. (1985). *Nonparametric Density Estimation: The L1 View.* New York: John Wiley.

[11] Fan J (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998–1004.

[12] Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics* **20**, 2008–2036.

[13] Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *Journal of the Royal Statistical Society, Series B* **57**, 371–394.

[14] Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications.* London: Chapman and Hall.

[15] Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblable conditionelle: étude asymptotique et application au cas Markovien. In J.-J. Droesbeke et al., Eds, *Spatial Processes and Spatial Time Series Analysis*, *Proceedings of the 6th Franco-Belgian Meeting of Statisticians.* Brussels: FUSL.

[16] Guyon, X. (1995). *Random Fields on a Network.* New York: Springer Verlag.

[17] Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1990). *Nonparametric Curve Estimation From Time Series*, Lecture Notes in Statistics 60. New York: Springer Verlag.

[18] Hallin, M. and Tran, L. T. (1996). Kernel density estimation for linear processes: asymptotic normality and optimal bandwidth derivation. *Annals of the Institute of Statistical Mathematics* **48**, 429—449.

[19] Hallin, M., Lu, Z., and Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7**, 657-668.

[20] Hallin, M., Lu, Z., and Tran, L. T. (2002). Density estimation for spatial processes: the L_1 theory. Submitted.

[21] Hastie, T. and Loader, C. (1993). Local regression: automatic kernel carpentry. *Statistical Science* **8**, 120–143.

[22] Ibragimov, I. A., and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters-Noordhoff.

[23] Ioannides, D. A., and Roussas, G. G. (1987). Note on the uniform convergence of density estimates for mixing random variables. *Satistics & Probability Letters* **5**, 279-285.

[24] Lu, Z. (1998). On geometric ergodicity of a non-linear autoregressive (AR) model with an autoregressive conditional heteroscedastic (ARCH) term. *Statistica Sinica* **8**, 1205–1217.

[25] Lu, Z. (2001). Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics* **53**, 447-468.

[26] Lu, Z. (2000). Spatial nonparametric regression estimation. *Research Report AMSS-2000-061*, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing.

[27] Lu, Z. and Cheng, P. (1997). Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series. *Journal of Statistical Planning & Inference* **65**, 67–86.

[28] Masry, E. (1983). Probability density estimation from sampled data. *IEEE Trans. Inform. Theory* **IT-29** 696-709.

[29] Masry, E. (1986). Recursive probability density estimation from weakly dependent processes. *IEEE Transactions on Information Theory* **IT-32**, 254-267.

[30] Masry, E. and Györfi, L. (1987). Strong consistency and rates for recursive density estimators for stationary mixing processes. *Journal of Multivariate Analysis* **22**, 79–93.

[31] Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong consistency and asymptotic normality. *Econometric Theory* **11**, 258–289.

[32] Nakhapetyan, B.S. (1980). The central limit theorem for random fields with mixing conditions. In R. L. Dobrushin and Ya. G. Sinai, Eds., *Advances in Probability* **6**, *Multicomponent Systems*, 531-548. New York: Marcel Dekker.

[33] Nakhapetyan, B. S. (1987). An approach to proving limit theorems for dependent random variables. *Theory of Probability and its Applications* **32**, 535–539.

[34] Neaderhouser, C. C. (1980). Convergence of blocks spins defined on random fields. *Journal of Statistical Physics* **22**, 673–684.

[35] Pham, T. D. (1986). The mixing properties of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and their Applications* **23**, 291–300.

[36] Pham, D. T., and Tran, L. T. (1985). Some strong mixing properties of time series models. *Stochastic Processes and their Applications* **19** 297–303.

[37] Possolo, A. (1991). *Spatial Statistics and Imaging.* IMS Lecture Notes-Monograph Series. Hayward: Institute of Mathematical Statistics.

[38] Ripley, B. (1981). *Spatial Statistics.* New York: Wiley.

[39] Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* **4**, 185–207.

[40] Robinson, P.M. (1987). Time series residuals with application to probability density estimation. *Journal of Time Series Analysis* **8**, 329–344.

[41] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Boston: Birkhauser.

[42] Roussas, G. G. (1969). Nonparametric estimation of the distribution function of a Markov process. *Annals of Mathematical Statistics* **40**, 1386–1400.

[43] Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *Journal of Statistical Planning & Inference* **18**, 135–149.

[44] Ruppert D, and Wand, M P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics* **22**, 1346–1370.

[45] Stone, C.J. (1977). Consistent nonparametric regression. *Annals of Statistics* **5**, 595–620.

[46] Takahata, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64** 445–456.

[47] Tjøstheim, D. (1990) Nonlinear time series and Markov chains. *Advances in Applied Probability* **22**, 587–611.

[48] Tran, L. T. (1989). The L_1 convergence of kernel density estimates under dependence. *Canadian Journal of Statistics* **17**, 197–208.

[49] Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis* **34**, 37–53.

[50] Tran, L. T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *Journal of Multivariate Analysis* **44**, 23–46.

[51] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434–449.

[52] Whittle, P. (1963). Stochastic process in several dimensions. *Bulletin of the International Statistical Institute* **40**, 974–985.

[53] Withers, C. S. (1981). Conditions for linear processes to be strong mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **57**, 477–480.

[54] Wu, W.B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Annals of Statistics* **30**, 1441-1459.

[55] Yakowitz, S. (1987). Nearest-neighbor methods for time series analysis. *Journal of Time Series Analysis* **8**, 235–247.

Figure 1. The true (dotted line) and estimated (solid lines) regression functions in ten replications of Model 1, with sample sizes $m = 10, n = 20$.

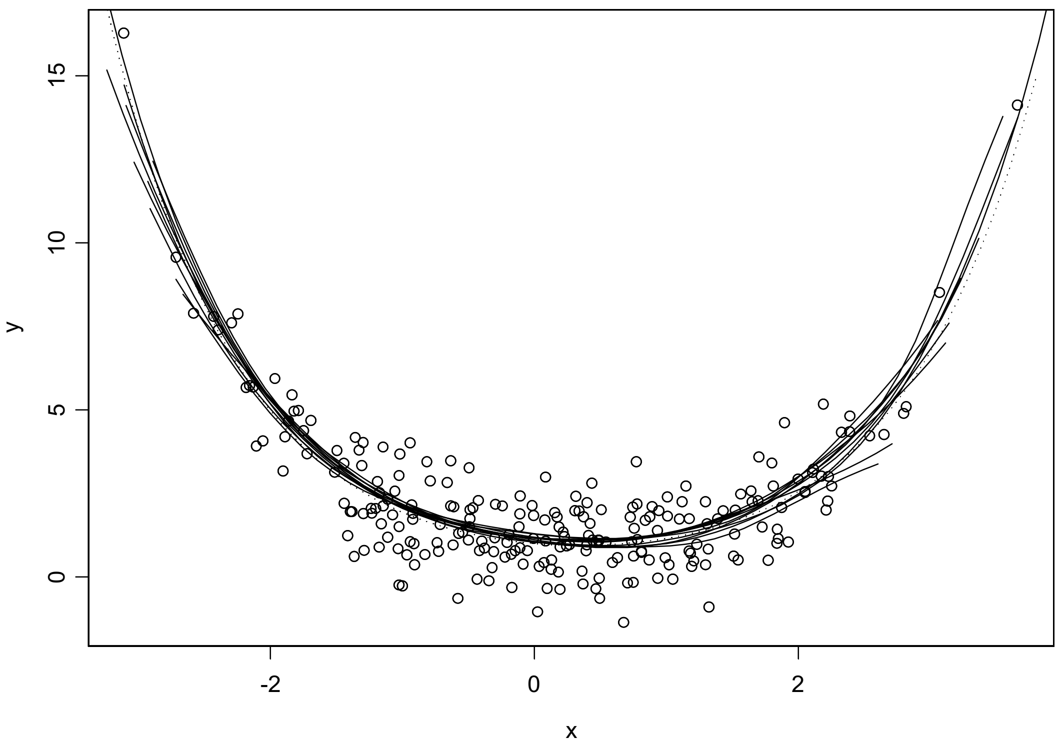

Figure 2(a). The estimated (solid lines) regression (prediction) functions in ten replications of Model 2, with $X_{i,j} = X^0_{i,j} := Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1}$ and sample sizes $m = 10, n = 20$.

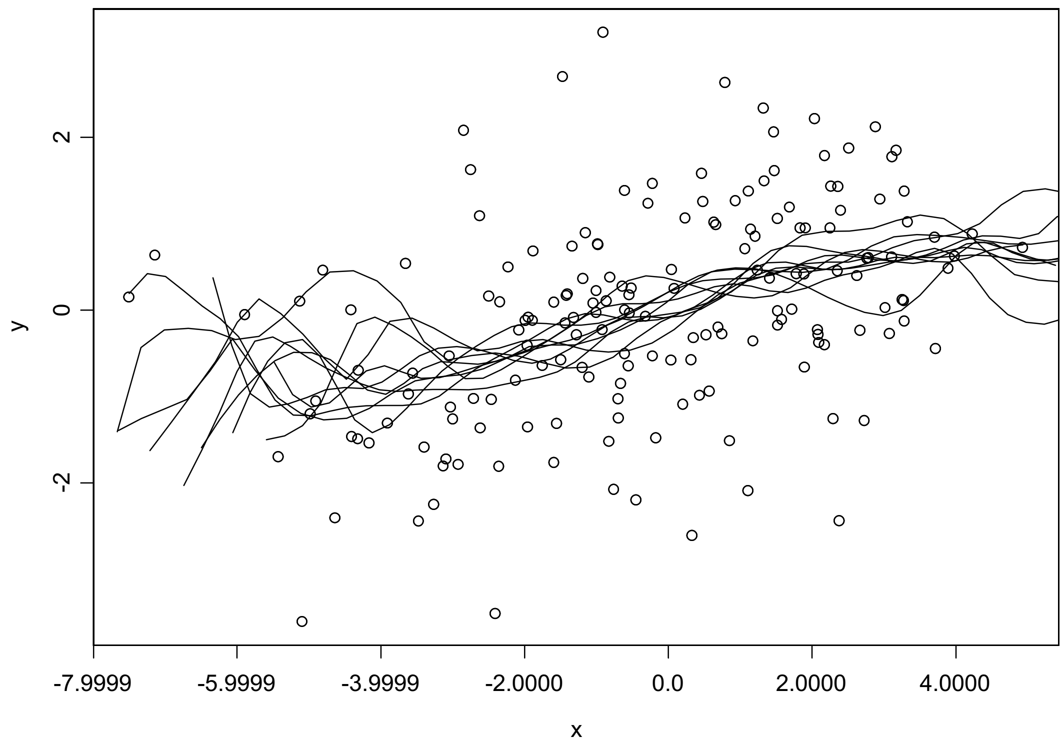

Figure 2(b). The estimated (solid lines) regression (prediction) functions in ten replications of Model 2, with $X_{i,j} = X^0_{i,j} := Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1}$ and sample sizes $m = 30, n = 40$.

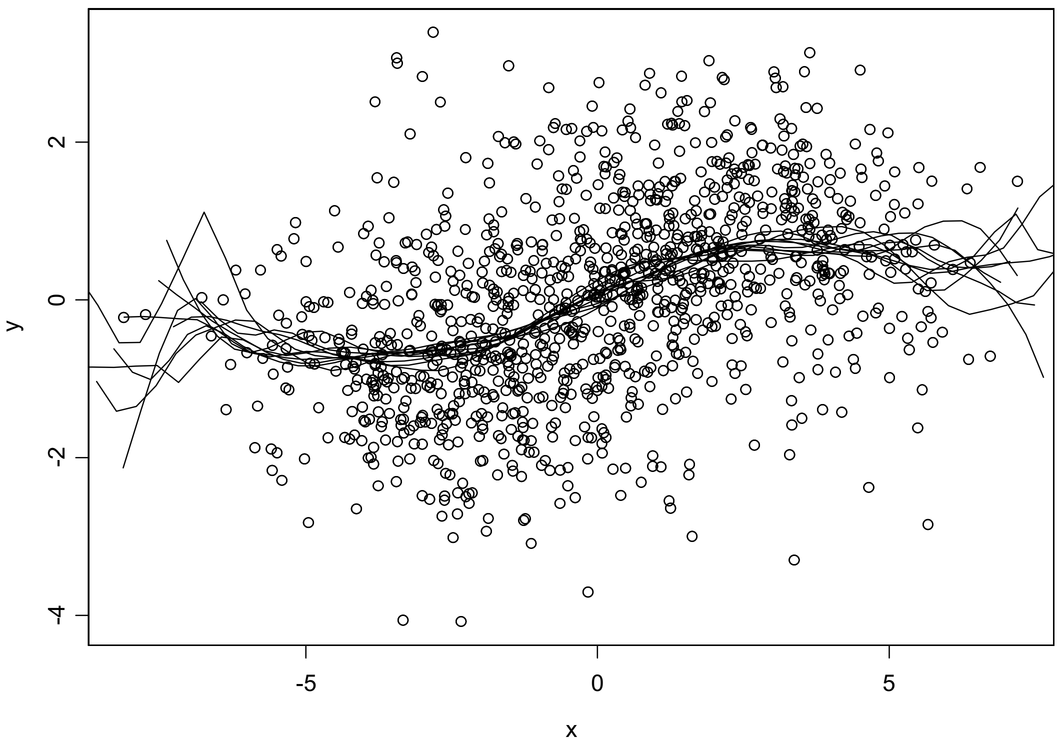

Figure 2(c). The estimated (solid lines) regression (prediction) functions in ten replications of Model 2, with $X_{i,j} = X_{i,j}^{c} := Y_{i-2,j} + Y_{i,j-2} + Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1} + Y_{i+2,j} + Y_{i,j+2}$ and sample sizes $m = 30, n = 40$.

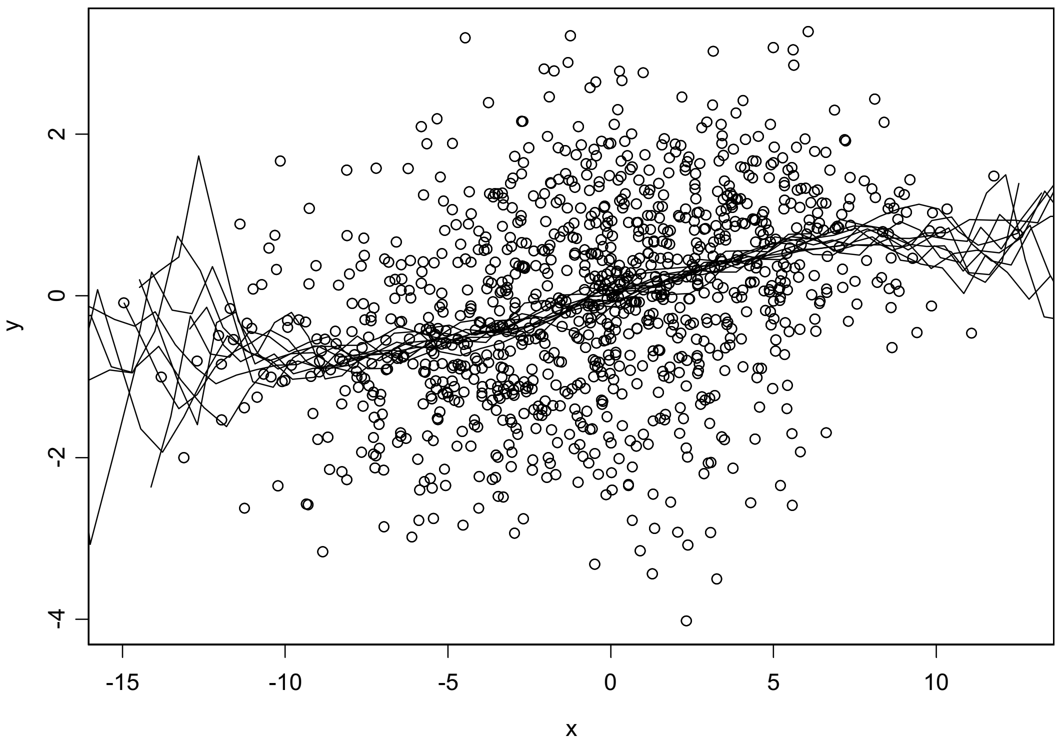

Figure 2(d). The estimated (solid lines) regression (prediction) functions in ten replications of Model 2, with $X_{i,j} = X_{i,j}^{d} := Y_{i-1,j} + Y_{i,j-1}$ and sample sizes $m = 30, n = 40$.

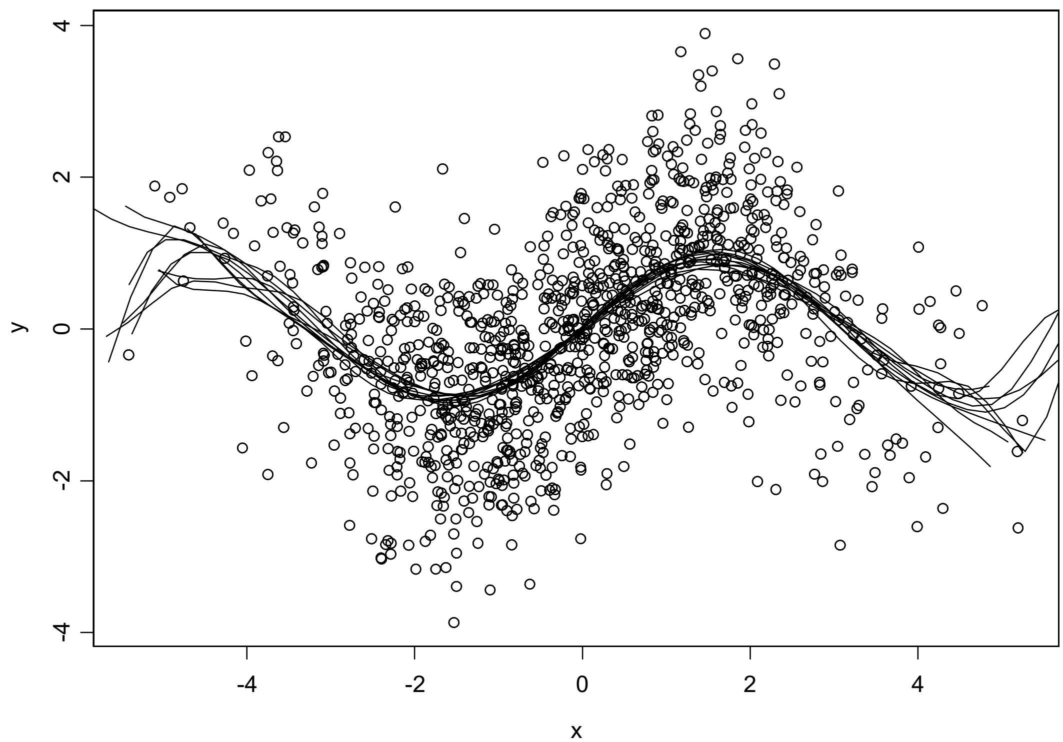

Figure 2(e). The estimated (solid lines) regression (prediction) functions in ten replications of Model 2, with $X_{i,j} = X^{e}_{i,j} := Y_{i+1,j} + Y_{i,j+1}$ and sample sizes $m = 30, n = 40$.

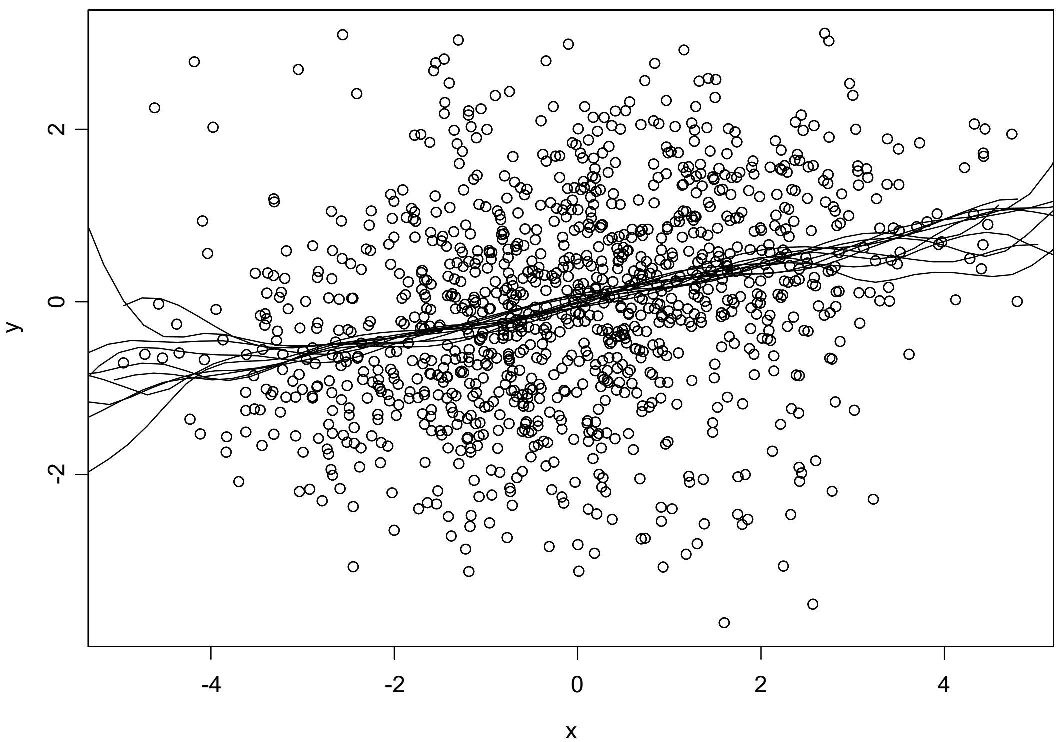

Figure 2(f). The estimated (solid lines) regression (prediction) functions in ten replications of Model 2, with $X_{i,j} = X_{i,j}^{f} := Y_{i-1,j} + Y_{i,j-1} + Y_{i-2,j} + Y_{i,j-2}$ and sample sizes $m = 30, n = 40$.