Bernoulli **15**(3), 2009, 659–686
DOI: 10.3150/08-BEJ168

Local linear spatial quantile regression

MARC HALLIN[1,2], ZUDI LU[3] and KEMING YU[4]

[1]*Institut de Recherche en Statistique, E.C.A.R.E.S., and Département de Mathématique, Université Libre de Bruxelles, Campus de la Plaine CP 210, B-1050, Brussels, Belgium. E-mail: mhallin@ulb.ac.be*
[2]*Royal Academy of Belgium*
[3]*Department of Mathematics and Statistics, Curtin University of Technology, GPO Box U1987, Perth, WA6845, Australia and School of Mathematical Sciences, University of Adelaide, SA5005, Australia. E-mail: Z.Lu@curtin.edu.au*
[4]*Department of Mathematical Sciences, Brunel University, Uxbridge, West London UB8 3PH, U.K. E-mail: Keming.Yu@brunel.ac.uk*

Let $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^N\}$ be a stationary real-valued $(d+1)$-dimensional spatial processes. Denote by $\mathbf{x} \mapsto q_p(\mathbf{x})$, $p \in (0,1)$, $\mathbf{x} \in \mathbb{R}^d$, the spatial quantile regression function of order p, characterized by $\mathrm{P}\{Y_{\mathbf{i}} \leq q_p(\mathbf{x}) | \mathbf{X}_{\mathbf{i}} = \mathbf{x}\} = p$. Assume that the process has been observed over an N-dimensional rectangular domain of the form $\mathcal{I}_{\mathbf{n}} := \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N | 1 \leq i_k \leq n_k, k = 1, \ldots, N\}$, with $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{Z}^N$. We propose a local linear estimator of q_p. That estimator extends to random fields with unspecified and possibly highly complex spatial dependence structure, the quantile regression methods considered in the context of independent samples or time series. Under mild regularity assumptions, we obtain a Bahadur representation for the estimators of q_p and its first-order derivatives, from which we establish consistency and asymptotic normality. The spatial process is assumed to satisfy general mixing conditions, generalizing classical time series 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 $\mathbb{Z}^N$ (non-isotropic asymptotics). The method provides much richer information than the mean regression approach considered in most spatial modelling techniques.

Keywords: Bahadur representation; local linear estimation; random fields; quantile regression

1. Introduction

Since the pathbreaking paper by Koenker and Basset [29], quantile regression and autoregression methods have attracted considerable interest in all domains of statistics, ranging from time series to survival analysis and growth charts; see [28] for a review. Most surprisingly, they seldom have been considered in a spatial context, although their potential applications to spatial data clearly are without number. Very recently, Koenker and Mizera [30] made a first step towards a spatial quantile-based analysis by proposing, under the name of *penalized triograms*, a penalized spline method based on adaptively selected triangulations of the plane that allows for computing conditional quantiles over a two-dimensional domain. Their method, however, does not incorporate covariates, and hence is a spatial smoothing technique rather than a spatial (auto)regression one.

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*, but also may include a time component. Spatial data are modelled as finite realizations of vector stochastic processes indexed by $\mathbf{i} \in \mathbb{Z}^N$, also

1350-7265

called *random fields*. In this paper, we will consider strictly stationary $(d+1)$-dimensional real random fields of the form

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

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 $(\Omega, \mathcal{F}, \mathrm{P})$. Such spatial data arise in a variety of fields, including econometrics, environmental sciences, image analysis, oceanography, geostatistics and many others. The statistical treatment of such data is the subject of an abundant literature that cannot be reviewed here; for background reading, we refer the reader to the monographs [2,11,21,42,43].

In a number of applications, a crucial problem consists in describing and analyzing the influence of a vector $\mathbf{X}_{\mathbf{i}}$ of covariates on some real-valued response $Y_{\mathbf{i}}$. In the spatial context, this study is particularly difficult due to the possibly highly complex spatial dependence among the various sites – a dependence that typically has to be treated as a nuisance. The traditional approach to this problem consists in assuming that $Y_{\mathbf{i}}$ has finite expectation, so that the *spatial mean regression function* $g : \mathbf{x} \mapsto g(\mathbf{x}) := \mathrm{E}[Y_{\mathbf{i}} | \mathbf{X}_{\mathbf{i}} = \mathbf{x}]$ is well defined and clearly carries relevant information on the dependence of Y on $\mathbf{X}$. This approach has been successfully considered in several papers, among which are [19,24]. However, (conditional) expectations may not exist. Even when they do, they only carry limited information on the dependence under study. In most practical cases, we would expect different structural relationships for the higher- (lower-) order quantiles than for the central ones, and the conditional distribution of Y (asymmetry, spread, ...) is likely to depend on $\mathbf{X}$ as well. A regression analysis based on conditional means ignores such essential features of the dependence of Y on $\mathbf{X}$, which can be taken care of by Koenker and Bassett's more general conditional quantile analysis only.

In this paper, instead of spatial mean regression, we thus consider the *spatial quantile regression functions* $q_p : \mathbf{x} \mapsto q_p(\mathbf{x})$, $0 < p < 1$, characterized by $\mathrm{P}\{Y_{\mathbf{i}} \le q_p(\mathbf{x}) | \mathbf{X}_{\mathbf{i}} = \mathbf{x}\} = p$. Although q_p (just as g) is only defined up to a P-null set of $\mathbf{x}$ values, we 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 neighbourhood of $\mathbf{i}$, q_p is called a *spatial quantile autoregression function*. Parametric (linear) spatial *mean* autoregression models were considered as early as 1954 by [51]; see [3,52] for further developments in this approach. Contrary to [51], we adopt a nonparametric point of view, as in [24], avoiding any parametric specification, both for q_p as for the possibly extremely complex spatial dependence structure of the data.

For $N = 1$, our problem reduces to the classical one of quantile (auto)regression for independent or serially dependent observations and has received extensive attention in the literature; see, for instance, [7,14,17,27,29,31–33,41,50,53–55]. Quite surprisingly, despite its obvious importance for applications, the spatial version $(N > 1)$ of the same problem remains essentially unexplored. Several recent papers (among which [8,22,23,48,49] deal with the related problem of estimating the density f of a random field of the form $\{\mathbf{X}_{\mathbf{i}}; \mathbf{i} \in \mathbb{Z}^N\}$, whereas [19,24,35,36] consider the estimation of spatial mean regression functions). To the best of our knowledge, no attempt has been made so far to estimate spatial quantile regression functions.

Being of a nonparametric nature, our estimators of spatial quantile regression functions naturally involve some smoothing techniques. With the functions q_p to be estimated being defined over the d-dimensional space of covariates, smoothing naturally is over the $\mathbf{X}$ values, not (as in spatial smoothing methods) over the sites $\mathbf{i}$. Among all smoothing techniques, the Nadaraya–Watson method, in the traditional serial case ($N = 1$), is probably the most standard one; it has been well documented, however – see, for instance, [16] – that it suffers from several severe drawbacks, such as poor boundary performances, excessive bias and low efficiency, and that the local polynomial fitting methods developed in [10,46] are generally preferable. Such local polynomial methods, and more particularly local linear fitting, have become increasingly popular in the light of recent works; see [15,16,34,45,53,54]. For $N = 1$, [27,37] delineate the asymptotics of local polynomial fitting for quantile regression under general mixing conditions. In this paper, we extend this approach to the context of spatial quantile regression ($N > 1$) by defining an estimator of q_p based on local linear regression quantiles.

Extending classical time series asymptotics ($N = 1$) to spatial asymptotics ($N > 1$), however, is far from trivial. Due to the absence of a canonical ordering in the space, there is no obvious definition of tail σ-fields, ergodicity, mixing and other traditional time-domain concepts. Little seems to exist about this in the literature, where only central limit results are well documented; see, for example, [5,39]. Even the simple idea of a sample size $\mathbf{n}$ going to infinity (the sample size here is a domain in $\mathbb{Z}^N$) has to be clarified in this setting. Assumptions (A3), (A3$'$) and (A3$''$) are reasonable and flexible generalizations of traditional time series concepts.

The stationary assumption we are making throughout plays a fundamental role. Its main consequence is that conditional densities (of $Y_{\mathbf{i}}$ conditional on $\mathbf{X}_{\mathbf{i}} = \mathbf{x}$) – hence the conditional quantile functions q_p – only depend on $\mathbf{x}$, not on $\mathbf{i}$. The regressors $\mathbf{X}_{\mathbf{i}}$ and $\mathbf{X}_{\mathbf{j}}$ may be strongly dependent (at neighbouring sites $\mathbf{i} \sim \mathbf{j}$) or nearly independent (at distant sites $\mathbf{i}$ and $\mathbf{j}$). If they take similar values, they will yield similar conditional Y-quantiles: $q_p(\mathbf{X}_{\mathbf{i}}) \sim q_p(\mathbf{X}_{\mathbf{j}})$. *Local* linear fitting here means *local* in the regressor's space. Note, however, that when the regressors $\mathbf{X}_{\mathbf{i}}$ contain neighbouring values of $Y_{\mathbf{i}}$ (quantile *auto*regression), the analysis automatically recovers some spatial smoothing flavor.

Depending on the context, all assumptions can be criticized, and so can the assumption of spatial stationarity – no more so, however, than the time series assumption of stationarity over time. In the time series context, whenever stationarity definitely cannot be assumed, two major remedies are considered (still, in a nonparametric perspective). The most sophisticated one is based on Dahlhaus' idea of *locally stationary* processes [12], and relies on an *infill asymptotics* scheme. Extending this approach to quantiles and the spatial context, however, is well beyond the scope of this paper, and should be left for future research.

A much simpler and less formal method, which is of daily practice in time series analysis, consists in a preliminary *detrending* of the observations. Transposed to a spatial setting, this idea implicitly relies on a model of the form $(\tilde{Y}_{\mathbf{i}}, \tilde{\mathbf{X}}'_{\mathbf{i}})' = (\mu_Y(\mathbf{s}_{\mathbf{i}}), \boldsymbol{\mu}'_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}))' + (Y_{\mathbf{i}}, \mathbf{X}'_{\mathbf{i}})'$, where $\mathbf{s}_{\mathbf{i}} := (i_1/n_1, \ldots, i_N/n_N)$, $1 \le i_k \le n_k$, $k = 1, 2, \ldots, N$. Here $\mathbf{s} \mapsto (\mu_Y(\mathbf{s}), \boldsymbol{\mu}'_{\mathbf{X}}(\mathbf{s}))'$ is an unspecified non-random spatial *trend* (defined over $(0, 1)^N$) and $(Y_{\mathbf{i}}, \mathbf{X}'_{\mathbf{i}})'$ is an unobservable stationary random field with unconditional mean or median zero. The analysis then proceeds in two steps. First (*detrending*), an estimation $(\hat{\mu}_Y(\mathbf{s}_{\mathbf{i}}), \hat{\boldsymbol{\mu}}'_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}})')$ of the spatial trend is removed via some adequate spatial smoothing method (that is, smoothing with respect to $\mathbf{s}_{\mathbf{i}}$); see [1,6,56] for recent discussion. In the second step, the detrended data is supposed to satisfy the stationarity assumption and subjected to the estimation method proposed, yielding for the detrended Y's and detrended

$\mathbf{X}$'s an estimated conditional p-quantile function $\check{q}_p$. At site $\mathbf{i}$, the estimated p-quantile of $\tilde{Y}_{\mathbf{i}}$ conditional on $\tilde{\mathbf{X}}_{\mathbf{i}}$ then is obtained as $\hat{\mu}_Y(\mathbf{s_i}) + \check{q}_p(\tilde{\mathbf{X}}_{\mathbf{i}} - \hat{\boldsymbol{\mu}}_{\mathbf{X}}(\mathbf{s_i}))$.

In most references, including the traditional time series ones, no formal justification is given for this two-step strategy. A more formal approach is developed in Section 3, where we show that, under adequate assumptions, preliminary detrending does not affect the asymptotic results of Section 2.

The paper is organized as follows: In Section 2.1 we provide the notation and main assumptions. Sections 2.2 and 2.3 introduce the main ideas underlying local linear regression in the context of random fields and sketch the main steps of the proofs to be developed in the sequel. Section 2.4, where asymptotic normality is stated under various types of asymptotics and various mixing assumptions, is the main theoretical section of the paper. Section 3 extends those results to the case of a random field with spatial trend. In Section 4, the method is applied to an environmental data set. Proofs and technical lemmas are concentrated in an Appendix.

2. Local linear estimation of spatial quantile 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)–(A3) are related to the random field itself.

(A1) (Densities) The process (1.1) is strictly stationary; $(Y_{\mathbf{i}}, \mathbf{X_i})$ has density f and, denoting by $f_{\mathbf{X}}$ the marginal density of $\mathbf{X}$, by $f_{Y|\mathbf{X}=\mathbf{x}}$ the density of Y conditional on $\mathbf{X} = \mathbf{x}$ and by $f_{\mathbf{i},\mathbf{j}}(\mathbf{x}, \tilde{\mathbf{x}})$ the joint density of $(\mathbf{X_i}, \mathbf{X_j})$ at $(\mathbf{x}, \tilde{\mathbf{x}})$,
  - (i) $\mathbf{x} \mapsto f_{\mathbf{X}}(\mathbf{x})$ is strictly positive and continuous for all $\mathbf{x}$;
  - (ii) for all $\mathbf{x}$, there exist a neighbourhood B of $y = q_p(\mathbf{x})$ and a neighbourhood $\mathbf{B}$ of $\mathbf{x}$ such that $y \mapsto f_{Y|\mathbf{X}=\mathbf{x}}(y) > 0$ is continuous over B, uniformly in $\mathbf{x} \in \mathbf{B}$, while $\mathbf{x} \mapsto f_{Y|\mathbf{X}=\mathbf{x}}(y)$ is continuous over $\mathbf{B}$ for all $y \in B$;
  - (iii) $\sup_{\mathbf{i},\mathbf{j} \in \mathbb{Z}^N} \sup_{\mathbf{x},\tilde{\mathbf{x}} \in \mathbb{R}^d} f_{\mathbf{i},\mathbf{j}}(\mathbf{x}, \tilde{\mathbf{x}}) \le C$ for some $C > 0$.

(A2) (Spatial quantile regression functions) The function $\mathbf{x} \mapsto q_p(\mathbf{x})$ is twice continuously differentiable. Denoting by $\dot{q}_p(\mathbf{x})$ its gradient and by $\ddot{q}_p(\mathbf{x})$ the matrix of its second derivatives (at $\mathbf{x}$), $\mathbf{x} \mapsto \ddot{q}_p(\mathbf{x})$ is continuous at all $\mathbf{x}$.

Conditions similar to Assumption (A1) have been considered in the literature, in the *i.i.d.* setting (cf. [17]). Assumption (A2) is standard.

Besides Assumptions (A1) and (A2), we need some appropriate assumption of *spatial mixing*. For any collection $\mathcal{S} \subset \mathbb{Z}^N$ of sites, denote by $\mathcal{B}(\mathcal{S})$ the σ-field generated by $\{(Y_{\mathbf{i}}, \mathbf{X_i}) | \mathbf{i} \in \mathcal{S}\}$. 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 $\|\mathbf{i}\| := (i_1^2 + \cdots + i_N^2)^{1/2}$ stands for the Euclidean norm. Finally, write $\mathrm{Card}(\mathcal{S})$ for the cardinality of $\mathcal{S}$. As in [24], two distinct forms (either (A3) and (A3$'$) or (A3) and (A3$''$)) of spatial mixing are considered.

(A3) (Spatial mixing) There exist two functions, $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\varphi(t) \downarrow_{t\to\infty} 0$, and $\psi : \mathbb{N}^2 \to \mathbb{R}^+$ symmetric and decreasing in its two arguments, such that

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

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

$$\lim_{m\to\infty} m^a \sum_{j=m}^{\infty} j^{N-1}\{\varphi(j)\} = 0 \qquad \text{for some constant } a > N.$$

The assumptions we are making on the function ψ are either

(A3′) $\psi(n', n'') \leq \min(n', n'')$

or (throughout we denote by C a generic positive constant, the value of which may vary according to the context)

(A3″) $\psi(n', n'') \leq C(n' + n'' + 1)^\kappa$ for some $C > 0$ and $\kappa > 1$.

In case (2.1) holds with $\psi \equiv 1$, the random field $\{(Y_\mathbf{i}, \mathbf{X_i})\}$ is called *strongly mixing*. In the serial case ($N = 1$), many stochastic processes and time series are known to be strongly mixing; cf. [18]. It is shown in [20] that, under certain conditions, linear random fields of the form $\mathbf{X_n} = \sum_{\mathbf{j}\in\mathbb{Z}^N} \mathbf{g_j Z_{n-j}}$, where the $\mathbf{Z_j}$'s are independent random variables, are strongly mixing. Assumptions (A3′) and (A3″) are the same as the mixing conditions used in [40,47], respectively, and are weaker than the uniform strong mixing condition considered in [39]. Such assumptions are the price to be paid for the presence of an unspecified spatial dependence structure.

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 $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{Z}^N$ 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 following one, where all components of $\mathbf{n}$ tend to infinity at the same rate. As in [48], we write $\mathbf{n} \Longrightarrow \infty$ if $\mathbf{n} \to \infty$ and $|n_j/n_k| < C$ for some $0 < C < \infty$, $1 \leq j, k \leq N$.

Assumption (A4) deals with the kernel function $K : \mathbb{R}^d \to \mathbb{R}$, and Assumptions (B1)–(B2) with the bandwidth $h_\mathbf{n}$ to be used in the estimation method. For any $\mathbf{c} := (c_0, \mathbf{c}_1')' \in \mathbb{R}^{d+1}$, define $K_\mathbf{c}(\mathbf{u}) := (c_0 + \mathbf{c}_1'\mathbf{u})K(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^d$.

(A4) (Kernels) (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, that is, $\int_{\mathbb{R}^{d+1}} |K_\mathbf{c}(\mathbf{x})|\,\mathrm{d}\mathbf{x} < \infty$;
 (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;
 (iii) The kernel function K is a continuously differentiable and bounded density function with compact support $C_K \subset \mathbb{R}^d$ such that $\int \mathbf{u}K(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbf{0}$ and $\int \mathbf{uu}'K(\mathbf{u})\,\mathrm{d}\mathbf{u}$ is positive definite.

(B1) (Bandwidths) The bandwith is such that $\lim_{\mathbf{n}\to\infty} h_\mathbf{n} = 0$ and $\lim_{\mathbf{n}\to\infty} \widehat{\mathbf{n}} h_\mathbf{n}^d = \infty$.

(B2) (Bandwidths) Same as (B1), but $\hat{\mathbf{n}} h_\mathbf{n}^{4+d} = \mathrm{O}(1)$ as $\mathbf{n} \to \infty$.

Finally, write $f_{Y|\mathbf{X}}(y|\mathbf{x})$ for $f_{Y|\mathbf{X}=\mathbf{x}}(y)$ and $F_{Y|\mathbf{X}}(y|\mathbf{x}) := \mathrm{P}(Y_{\mathbf{i}} < y|\mathbf{X}_{\mathbf{i}} = \mathbf{x})$ for the corresponding conditional distribution function. Primes denote transposes.

2.2. Local linear fitting of the spatial quantile regression function

In this section we extend traditional local linear fitting ideas to the context of spatial quantile regression. Write $\dot{q}_p(\mathbf{x}) = (\partial q_p(\mathbf{x})/\partial x_1, \ldots, \partial q_p(\mathbf{x})/\partial x_d)'$ for the gradient at $\mathbf{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d$ of $\mathbf{x} \mapsto q_p(\mathbf{x})$. The basic idea of local linear fitting (see [16,17,34,54]) consists in approximating in a neighbourhood of $\mathbf{x}$ the unknown quantile regression function $q_p(\mathbf{z})$ by a linear function:

$$q_p(\mathbf{z}) \approx q_p(\mathbf{x}) + (\dot{q}_p(\mathbf{x}))'(\mathbf{z} - \mathbf{x}) =: a_0 + \mathbf{a}_1'(\mathbf{z} - \mathbf{x}). \tag{2.2}$$

Therefore, estimating $(q_p(\mathbf{x}), \dot{q}_p(\mathbf{x}))$ is locally equivalent to estimating $(a_0, \mathbf{a}_1) = (a_0(\mathbf{x}), \mathbf{a}_1(\mathbf{x}))$. The classical theory of quantile regression suggests the estimators

$$(\widehat{a}_0, \widehat{\mathbf{a}}_1) := \arg \min_{(a_0, \mathbf{a}_1)} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \rho_p\big(Y_{\mathbf{i}} - a_0 - \mathbf{a}_1'(\mathbf{X}_{\mathbf{i}} - \mathbf{x})\big) K_h(\mathbf{X}_{\mathbf{i}} - \mathbf{x}), \tag{2.3}$$

where $\rho_p(y) := y(p - I_{\{y<0\}})$ stands for the traditional check function, I_A is the indicator function of set A and $K_h(\mathbf{x}) := h_{\mathbf{n}}^{-d} K(\mathbf{x}/h_{\mathbf{n}})$, with a kernel function K defined on $\mathbb{R}^d$ and a bandwidth $h = h_{\mathbf{n}} > 0$ tending to 0 as $\mathbf{n} \to \infty$. This motivates the choice of $\widehat{q}_p(\mathbf{x}) := \widehat{a}_0$ and $\widehat{\dot{q}}_p(\mathbf{x}) := \widehat{\mathbf{a}}_1$ as estimators of $q_p(\mathbf{x})$ and $\dot{q}_p(\mathbf{x})$, respectively. Note that (2.3) does not require the regular grid structure we are assuming throughout. It seems intuitively clear that "nearly regular grids" will not harm the results of this paper. However, the asymptotic treatment of irregular grids (essentially, a definition of a "nearly regular grid") is a delicate and problematic issue that we will not consider here.

2.3. Bahadur representation

The definition (2.3) looks simple, but unlike the local linear fitting estimator for spatial mean regression proposed in [24], it does not allow for an explicit solution, which creates additional difficulties in developing the asymptotic theory. We overcome these difficulties by obtaining a Bahadur representation for $\widehat{q}_p$ and $\widehat{\dot{q}}_p$.

Since the first Bahadur representation for regression quantiles was obtained in [44] (under i.i.d. errors), several results of that type have been proposed in the literature; see [4,9,26,31]. The result by Chaudhuri [9], who establishes a Bahadur representation for quantile regression functions and their derivatives of arbitrary orders, is particularly remarkable. The context, however, is a nonparametric regression model of the form $Y_i = \theta(X_i) + \varepsilon_i$, where the errors ε_i are i.i.d. and independent of the regressors X_i; the influence on quantiles of the X_i's thus is limited to conditional shifts, which precludes all forms of conditional heteroskedasticity. Our result is more general, as it allows for complex spatial dependencies, and does not put any restriction on the influence of regressors on the conditional distribution of Y – as long as mixing assumptions are satisfied. On the other hand, our Bahadur representation is a *weak* one (with o_{P} remainder –

which is all we need for asymptotic normality), whereas Chaudhuri's is a *strong* one (with a.s. convergence), addressing first-order derivatives only. The proof of Theorem 2.1 is postponed to Section A.2.

Theorem 2.1 (Bahadur representation). *Let Assumptions* (A1), (A3), (A4) *and* (B1) *hold, and assume that* $\mathbf{x} \mapsto q_p(\mathbf{x})$ *is continuously differentiable at* $\mathbf{x}$, *with gradient* $\dot{q}_p(\mathbf{x})$. *Then,*

$$(\widehat{\mathbf{n}}h_{\mathbf{n}}^d)^{1/2}\begin{pmatrix} \widehat{q}_p(\mathbf{x}) - q_p(\mathbf{x}) \\ h_{\mathbf{n}}\big(\widehat{\dot{q}}_p(\mathbf{x}) - \dot{q}_p(\mathbf{x})\big)\end{pmatrix} = \frac{\eta_p(\mathbf{x})}{\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\psi_p(Y_{\mathbf{i}}^*)\begin{pmatrix} 1 \\ \frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\end{pmatrix} K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right) + o_P(1),$$

as $\mathbf{n} \to \infty$, *where* $\psi_p(y) := p - I_{\{y<0\}}$, $Y_{\mathbf{i}}^* := Y_{\mathbf{i}}^*(p) := Y_{\mathbf{i}} - q_p(\mathbf{x}) - (\dot{q}_p(\mathbf{x}))'(\mathbf{X_i} - \mathbf{x})$, *and* $\eta_p(\mathbf{x}) := (f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}))^{-1}$.

2.4. Asymptotic normality

Using the powerful tool of the Bahadur representation, we can establish the consistency and derive the asymptotic distribution of the local linear quantile regression estimates under weak conditions. First, we consider the case where the sample size tends to ∞ in the manner of [48], that is, $\mathbf{n} \Longrightarrow \infty$. Assuming now that (A2) holds, so that $\mathbf{x} \mapsto q_p(\mathbf{x})$ is twice differentiable, let

$$B_0(\mathbf{x}) := \{f_{\mathbf{X}}(\mathbf{x})\}^{-1}\operatorname{tr}\left[\ddot{q}_p(\mathbf{x})\int \mathbf{u}\mathbf{u}' K(\mathbf{u})\,\mathrm{d}\mathbf{u}\right] \quad \text{and} \quad \mathbf{B}_1(\mathbf{x}) := (B_{11}(\mathbf{x}), \dots, B_{1d}(\mathbf{x}))',$$

with

$$B_{1j}(\mathbf{x}) := f_{\mathbf{X}}^{-1}(\mathbf{x})\operatorname{tr}\left[\ddot{q}_p(\mathbf{x})\int \mathbf{u}\mathbf{u}' u_j K(\mathbf{u})\,\mathrm{d}\mathbf{u}\right], \qquad j = 1, \dots, d,$$

$$\sigma_0^2(\mathbf{x}) := \eta^*(\mathbf{x})\int K^2(\mathbf{u})\,\mathrm{d}\mathbf{u}, \quad \text{and} \quad \boldsymbol{\sigma}_1^2(\mathbf{x}) := \eta^*(\mathbf{x})\int \mathbf{u}\mathbf{u}' K^2(\mathbf{u})\,\mathrm{d}\mathbf{u},$$

where $\eta^*(\mathbf{x}) := \eta_p^2(\mathbf{x})p(1-p)f_{\mathbf{X}}(\mathbf{x}) = p(1-p)/f_{\mathbf{X}}(\mathbf{x})f_{Y|\mathbf{X}}^2(q_p(\mathbf{x})|\mathbf{x})$.

Theorem 2.2. *Let Assumptions* (A1), (A2), (A3′), (A4) (*with* $\varphi(x) = \mathrm{O}(x^{-\mu})$ *as* $x \to \infty$ *for some* $\mu > 2N$) *and* (B2) *hold. Suppose that there exists a sequence of positive integers* $q_{\mathbf{n}}$ *such that* $q_{\mathbf{n}} \to \infty$, $q_{\mathbf{n}} = \mathrm{o}((\widehat{\mathbf{n}}h_{\mathbf{n}}^d)^{1/2N})$, *and* $\widehat{\mathbf{n}}q_{\mathbf{n}}^{-\mu} \to 0$ *as* $\mathbf{n} \Longrightarrow \infty$. *Moreover, let the bandwidth* $h_{\mathbf{n}}$ *tend to zero in such a manner that*

$$\liminf_{\mathbf{n}\Longrightarrow\infty} q_{\mathbf{n}}h_{\mathbf{n}}^{d/a} > 1 \qquad \text{for some } N < a < \mu - N. \tag{2.4}$$

Then, for any $\mathbf{x}$ *and* $0 < p < 1$, *as* $\mathbf{n} \Longrightarrow \infty$,

$$\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}\left[\begin{pmatrix} \widehat{q}_p(\mathbf{x}) - q_p(\mathbf{x}) \\ h_{\mathbf{n}}\big(\widehat{\dot{q}}_p(\mathbf{x}) - \dot{q}_p(\mathbf{x})\big)\end{pmatrix} - \frac{1}{2}\begin{pmatrix} B_0(\mathbf{x}) \\ \mathbf{B}_1(\mathbf{x})\end{pmatrix}h_{\mathbf{n}}^2\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_0^2(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}_1^2(\mathbf{x})\end{pmatrix}\right),$$

so that $\widehat{q}_p(\mathbf{x})$ *and* $\widehat{\dot{q}}_p(\mathbf{x})$ *are asymptotically independent.*

The asymptotic normality results in Theorem 2.2 are stated for $\widehat{q}_p(\mathbf{x})$ and $\widehat{\dot{q}}_p(\mathbf{x})$ at a given $\mathbf{x}$. They are easily extended, via the traditional Cramér–Wold device, into a joint asymptotic normality result for any couple $(\mathbf{x}_1, \mathbf{x}_2)$ (or any finite collection of $\mathbf{x}$ values); the asymptotic covariance terms (between $\widehat{q}_p(\mathbf{x}_1)$ and $\widehat{q}_p(\mathbf{x}_2)$, $\widehat{q}_p(\mathbf{x}_1)$ and $\widehat{\dot{q}}_p(\mathbf{x}_2)$, etc.) all are equal to zero (cf. [24], page 2478). The same remark also holds for Theorems 2.3–2.6 below.

An important advantage of local polynomial (and linear) fitting over the Nadaraya–Watson approach is its much better boundary behavior. This advantage often has been emphasized in the usual regression and time series settings when the regressors take values on a compact subset of $\mathbb{R}^d$. For example, considering a univariate ($d = 1$) regressor X with bounded support ([0, 1], say), it can be proved, using an argument similar to the one developed in the proof of Theorem 3.1 of [24], that asymptotic normality still holds at boundary points of the form $ch_{\mathbf{n}}$, $c \in \mathbb{R}^+$, but with asymptotic bias and variances

$$B_0 = \{f_X(0^+)\}^{-1}\left[\ddot{q}_p(0^+)\int_{-c}^{\infty} u^2 K(u)\,\mathrm{d}u\right] \quad \text{and} \quad \sigma_0^2 = \eta^*(0^+)\int_{-c}^{\infty} K^2(u)\,\mathrm{d}u,$$

and

$$B_1 = \{f_X(0^+)\}^{-1}\left[\ddot{q}_p(0^+)\int_{-c}^{\infty} u^3 K(u)\,\mathrm{d}u\right] \quad \text{and} \quad \sigma_1^2 = \eta^*(0^+)\int_{-c}^{\infty} u^2 K^2(u)\,\mathrm{d}u,$$

respectively, where $\eta^*(0^+) = \eta_p^2(0^+)p(1-p)f_X(0^+) = p(1-p)/f_X(0^+)f_{Y|X}^2(q_p(0^+)|0^+)$; similar results can be found in [16,18] for mean regression. As pointed out in [24], this advantage is likely to be more substantial as N grows.

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

Theorem 2.3. *Let Assumptions* (A1), (A2), (A3′) *and* (A4) *hold, with* $\varphi(x) = \mathrm{O}(\mathrm{e}^{-\xi x})$ *as* $x \to \infty$ *for some* $\xi > 0$. *Then, if* $h_{\mathbf{n}} \to 0$ *as* $\mathbf{n} \Longrightarrow \infty$ *in such a manner that* $(\widehat{\mathbf{n}} h_{\mathbf{n}}^{d(1+2N/a)})^{1/2N} \times (\log \widehat{\mathbf{n}})^{-1} \to \infty$ *for some* $a > N$, *the conclusions of Theorem* 2.2 *still hold.*

Note that, for $N = 1$ and "large" values of a, this condition is "close" to the classical one (for independent observations) that $nh_n^d \to \infty$. Next, we consider the situation under which $\mathbf{n}$ tends to ∞ in the "weak" sense ($\mathbf{n} \to \infty$ instead of $\mathbf{n} \Longrightarrow \infty$).

Theorem 2.4. *Let Assumptions* (A1), (A2), (A3′) *and* (A4) *hold, with* $\varphi(x) = \mathrm{O}(x^{-\mu})$ *as* $x \to \infty$ *for some* $\mu > 2N$. *Let* $q_{\mathbf{n}}$ *be a sequence of positive integers such that* $q_{\mathbf{n}} \to \infty$ *as* $\mathbf{n} \to \infty$, *and assume that the bandwidth* $h_{\mathbf{n}}$ *factorizes into* $h_{\mathbf{n}} := \prod_{i=1}^N h_{n_i}$, *with* $\widehat{\mathbf{n}} q_{\mathbf{n}}^{-\mu} \to 0$, $q_{\mathbf{n}} = \mathrm{o}((\min_{1\le k\le N}(n_k h_{n_k}^d))^{1/2})$, *and* $\liminf_{\mathbf{n}\to\infty} q_{\mathbf{n}} h_{\mathbf{n}}^{d/a} > 1$ *for some* $N < a < \mu - N$. *Then the conclusions of Theorem* 2.2 *hold as* $\mathbf{n} \to \infty$.

For $\varphi(x) \to 0$ at an exponential rate, parallel to Theorem 2.3, we have the following:

Theorem 2.5. *Let Assumptions* (A1), (A2), (A3′) *and* (A4) *hold, with* $\varphi(x) = \mathrm{O}(\mathrm{e}^{-\xi x})$ *as* $x \to \infty$ *for some* $\xi > 0$. *Let the bandwidth* $h_{\mathbf{n}}$ *factorize into* $h_{\mathbf{n}} := \prod_{i=1}^N h_{n_i}$ *in such a way that*

$\min_{1\le k\le N}\{(n_k h_{n_k}^d)^{1/2}\} h_{\mathbf{n}}^{d/a}(\log\widehat{\mathbf{n}})^{-1}\to\infty$ *for some* $a>N$ *as* $\mathbf{n}\to\infty$. *Then the conclusions of Theorem* 2.2 *hold as* $\mathbf{n}\to\infty$.

Under (A3″), we then have the following counterpart of Theorem 2.2:

Theorem 2.6. *Let Assumptions* (A1), (A2), (A3″) *and* (A4) *hold, with* $\varphi(x)=\mathrm{O}(x^{-\mu})$ *as* $x\to\infty$ *for some* $\mu>2N$. *Denote by* $q_{\mathbf{n}}$ *a sequence of positive integers such that* $q_{\mathbf{n}}\to\infty$, $q_{\mathbf{n}}=\mathrm{o}((\widehat{\mathbf{n}}b_{\mathbf{n}}^d)^{1/2N})$ *and* $\widehat{\mathbf{n}}^{\kappa+1}q_{\mathbf{n}}^{-\mu-N}\to 0$ *as* $\mathbf{n}\Longrightarrow\infty$. *Assume that the bandwidth* $h_{\mathbf{n}}$ *tends to zero in such a manner that* (2.4) *is satisfied as* $\mathbf{n}\Longrightarrow\infty$. *Then the conclusions of Theorem* 2.2 *hold as* $\mathbf{n}\Longrightarrow\infty$.

See Section A.3 for the proofs. Analogues of Theorems 2.3, 2.4 and 2.5 can also be obtained under Assumption (A3″); details are left to the reader.

3. Random fields with a spatial trend

In Section 2, the observed process $\{Y_{\mathbf{i}},\mathbf{X}_{\mathbf{i}}\}$ was assumed to be stationary – an assumption that is often violated in practice. As a reasonable alternative, we can assume that non-stationarity is due to the presence of a spatial trend and that, instead of the stationary process $\{Y_{\mathbf{i}},\mathbf{X}_{\mathbf{i}}\}$, we actually observe $\{\tilde{Y}_{\mathbf{i}},\tilde{\mathbf{X}}_{\mathbf{i}}\}$, with

$$\tilde{Y}_{\mathbf{i}}=\mu_Y(\mathbf{s}_{\mathbf{i}})+Y_{\mathbf{i}},\qquad \tilde{\mathbf{X}}_{\mathbf{i}}=\boldsymbol{\mu}_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}})+\mathbf{X}_{\mathbf{i}},\qquad \mathbf{i}\in\mathcal{I}_{\mathbf{n}},\tag{3.1}$$

where $\mathbf{s}_{\mathbf{i}}=(s_{i_1},\dots,s_{i_N}):=(i_1/n_1,\dots,i_N/n_n)$ and $\mathbf{s}\in[0,1]^N\mapsto(\mu_Y(\mathbf{s}),\boldsymbol{\mu}_{\mathbf{X}}(\mathbf{s}))$ is some deterministic but unknown trend function. The procedure described in Section 2 then is applied to the residuals $\{(\hat{Y}_{\mathbf{i}},\hat{\mathbf{X}}_{\mathbf{i}}):=(\tilde{Y}_{\mathbf{i}}-\hat{\mu}_Y(\mathbf{s}_{\mathbf{i}}),\tilde{\mathbf{X}}_{\mathbf{i}}-\hat{\boldsymbol{\mu}}_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}))\}$ of some preliminary spatial smoothing $(\hat{\mu}_Y(\mathbf{s}),\hat{\boldsymbol{\mu}}_{\mathbf{X}}(\mathbf{s}))$ of the original $\{\tilde{Y}_{\mathbf{i}},\tilde{\mathbf{X}}_{\mathbf{i}}\}$.

For the sake of simplicity, we assume throughout this section that $N=2$, which is also the most frequent case in practice. Letting

$$w(\mathbf{s}_{\mathbf{i}},\mathbf{s}):=W\big((\mathbf{s}_{\mathbf{i}}-\mathbf{s})/g\big)\Big/\Big[\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}W\big((\mathbf{s}_{\mathbf{j}}-\mathbf{s})/g\big)\Big],$$

where $g=g_{\mathbf{n}}$ is some bandwidth tending to 0, a simple smoothing is obtained as

$$\hat{\mu}_Y(\mathbf{s})=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\tilde{Y}_{\mathbf{i}}w(\mathbf{s}_{\mathbf{i}},\mathbf{s}),\qquad \hat{\boldsymbol{\mu}}_{\mathbf{X}}(\mathbf{s})=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\tilde{\mathbf{X}}_{\mathbf{i}}w(\mathbf{s}_{\mathbf{i}},\mathbf{s}).\tag{3.2}$$

A sufficient technical requirement for the convergence of this kernel smoothing is the following set of assumptions (inspired by Theorem 2 in [25]).

(C0) (i) $\{\mathbf{X}_{\mathbf{i}},Y_{\mathbf{i}}\}$ is a stationary spatial process with the spatial mixing coefficient specified in Assumption (A3) with $\varphi(m)\le Cm^{-\beta}$, where $0<C<\infty$ and $\beta>\{1+(s-1)(1+N)\}/(s-2)$;

(ii) For some $s > 2$, $E|Y_{\mathbf{i}}|^s < \infty$ and $E\|\mathbf{X}_{\mathbf{i}}\|^s < \infty$;
(iii) For $\theta = \{\beta - 1 - N - (1+\beta)/(s-1)\}/\{\beta + 3 - N - (1+\beta)/(s-1)\}$, $(\ln \hat{\mathbf{n}}/\hat{\mathbf{n}}^{\theta} g_{\mathbf{n}}^N) = \mathrm{o}(1)$.

The local linear estimators of the quantile regression are defined as in (2.3),

$$(\check{a}_0, \check{\mathbf{a}}_1) := \arg\min_{(a_0,\mathbf{a}_1)} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \rho_p\big(\hat{Y}_{\mathbf{i}} - a_0 - \mathbf{a}_1'(\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x})\big) K_h(\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x}), \tag{3.3}$$

yielding the estimators $\check{q}_p(\mathbf{x}) := \check{a}_0$ and $\check{\dot{q}}_p(\mathbf{x}) := \check{\mathbf{a}}_1$ of $q_p(\mathbf{x})$ and $\dot{q}_p(\mathbf{x})$, respectively.

In order to study the asymptotic behavior of these estimators, we need some additional regularity conditions.

(C1) $\mathbf{s} \mapsto \mu_Y(\mathbf{s})$ and $\mathbf{s} \mapsto \boldsymbol{\mu}_{\mathbf{X}}(\mathbf{s})$ are r times differentiable with bounded derivatives on $\mathcal{S} := [0,1]^2$, where r is some positive integer.
(C2) There exists a continuous *sampling intensity (density) function* f defined on $\mathcal{S}$ such that $0 < c_0 \le f(\mathbf{s}) \le c_1 < \infty$ for any $\mathbf{s} \in \mathcal{S}$ and $\hat{\mathbf{n}}^{-1}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} I(\mathbf{s}_{\mathbf{i}} \in A) \to \int_A f(\mathbf{s})\,\mathrm{d}\mathbf{s}$ for any measurable set $A \subset \mathcal{S}$, as $\mathbf{n} \to \infty$.

Assumption (C1) is a classical smoothness assumption on spatial trend functions; Assumption (C2) is mentioned for the sake of generality, and is trivially satisfied in the case of a regular grid. Depending on r in Assumption (C1), we require the following conditions on the kernel W and the bandwidth $g_{\mathbf{n}}$.

(C3) The kernel $W(\cdot)$, defined on $\mathbb{R}^2$, has bounded support with Lipschitz property, that is $|W(\mathbf{u}) - W(\mathbf{u}')| \le C\|\mathbf{u} - \mathbf{u}'\|$ for all $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^2$, where $C > 0$ is a generic constant, and satisfies ($\mathbf{u}^{\otimes k}$ stands for the kth Kronecker power of $\mathbf{u}$)

$$\int W(\mathbf{u})\,\mathrm{d}\mathbf{u} = 1, \qquad \int \mathbf{u}^{\otimes k} W(\mathbf{u})\,\mathrm{d}\mathbf{u} = 0, \qquad k = 1, 2, \ldots, r-1,$$

$$\int \mathbf{u}^{\otimes r} W(\mathbf{u})\,\mathrm{d}\mathbf{u} \neq 0.$$

(C4) As $\mathbf{n} \to \infty$, $g_{\mathbf{n}} \to 0$ and $\hat{\mathbf{n}} g_{\mathbf{n}}^2 h_{\mathbf{n}} \to \infty$; moreover, $g_{\mathbf{n}}^r/h_{\mathbf{n}} \to 0$, $h_{\mathbf{n}}^{d+2}/(\hat{\mathbf{n}} g_{\mathbf{n}}^4) \to 0$, $h_{\mathbf{n}}^d \ln \hat{\mathbf{n}}/g_{\mathbf{n}}^2 \to 0$, and $\hat{\mathbf{n}} h_{\mathbf{n}}^d g_{\mathbf{n}}^{2r} \to 0$.

Assumption (C3) requires a higher order kernel function $W(\cdot)$ of order r, which ensures that the bias term of the smoothing estimators of the spatial trends is of order $\mathrm{O}(g_{\mathbf{n}}^r)$ (the same objective could be achieved through a local polynomial fitting of order $(r-1)$). Assumption (C4) on the bandwidths $g_{\mathbf{n}}$ and $h_{\mathbf{n}}$ is satisfied, for instance, if we let $h_{\mathbf{n}} = \hat{\mathbf{n}}^{-1/(d+4)}$, which is optimal under Assumption (A2), with $\hat{\mathbf{n}} g_{\mathbf{n}}^{2(d+4)} \to \infty$ and $\hat{\mathbf{n}} g_{\mathbf{n}}^{r(d+4)/2} \to 0$, which holds for $r > 4$.

Theorem 3.1 (Bahadur representation). *Let Assumptions* (A1), (A3), (A4), (B1) *and* (C0)–(C4) *hold. Assume that* $\mathbf{x} \mapsto q_p(\mathbf{x})$ *is continuously differentiable at* $\mathbf{x}$, *with gradient* $\dot{q}_p(\mathbf{x})$. *Then,* *as* $\mathbf{n} \to \infty$,

$$(\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{1/2} \begin{pmatrix} \check{q}_p(\mathbf{x}) - q_p(\mathbf{x}) \\ h_{\mathbf{n}}(\check{\dot{q}}_p(\mathbf{x}) - \dot{q}_p(\mathbf{x})) \end{pmatrix} = \frac{\eta_p(\mathbf{x})}{\sqrt{\widehat{\mathbf{n}} h_{\mathbf{n}}^d}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \psi_p(Y_{\mathbf{i}}^*) \begin{pmatrix} 1 \\ \frac{\mathbf{X_i} - \mathbf{x}}{h_{\mathbf{n}}} \end{pmatrix} K\left(\frac{\mathbf{X_i} - \mathbf{x}}{h_{\mathbf{n}}}\right) + o_{\mathrm{P}}(1).$$

The proof of Theorem 3.1 is postponed to Section A.4.

It readily follows from this theorem that the asymptotic normality results of Section 2.4 still hold true for $\check{q}_p(\mathbf{x})$ and $\check{\dot{q}}_p(\mathbf{x})$. Details are left to the reader.

4. An application to environmental data

The data set we are analyzing here was collected as part of a project entitled "Geostatistical Analysis of Plant Community Transitions in the Outer Hebrides", led by Martin Kent (University of Plymouth), and was kindly provided by his colleague and coauthor Rana Moyeed. This project aims at a better understanding of the endangered coastal ecosystems in the Outer Hebrides of Scotland known as *machairs*. Of particular interest in that context are the rates of spatial change in plant species composition, and the environmental and biotic factors across landscape boundaries as well as within landscape patches.

Machair is a Gaelic word that describes a distinctive type of coastal grassland found in the north and west of Scotland, and in western Ireland. It is associated with calcareous sand blown inland by very strong prevailing winds from beaches and mobile dunes. Machair grassland plains are a complex mosaic of wet and dry grassland communities and ecosystems. Machair systems have high conservation value related to their rarity on a global scale, their species composition and botanical significance, in addition to their geomorphological, archaeological and landscape importance.

One of the major threats on the fragile equilibrium of the machair ecosystem is the increase of soil acidity induced, mainly, by an excess of organic matter, possibly related with intensive use of fertilizers containing ammonium or urea. One way of balancing the observed increase of soil acidity consists in replacing the lost cation nutrients, particularly calcium. A better understanding of the interaction between organic matter and Ca concentrations on one hand, soil acidity on the other hand, is thus crucial, and spatial quantile regression is particularly well adapted for an in-depth analysis of said interaction. The instance we are treating here is a good example of what would go unnoticed in a traditional regression/correlation approach but can be detected by our method.

The analysis we are conducting is deliberately simple, with a minimal number of two covariates ($N = d = 2$). Data were collected as explained, over a grid of 217 sites. The covariates X_1 and X_2 are densities of Ca (in mg/kg) and organic matter (in %), respectively. The response Y is a measure of soil acidity (pH) – a pH less than (resp., greater than) seven is considered acidic (resp., basic or alkaline), seven being the pH of pure water at 25°C. Figure 1 presents a spatial plot of raw data.

As a preliminary step, the data were "detrended" via the standard R function 'sm.regression' in the R library 'sm' (see [6] for details). A direct regression of Y against X_1 and X_2 indeed could lead to spurious relations induced by spatial trend (hence, spatial non-stationarity). We then apply

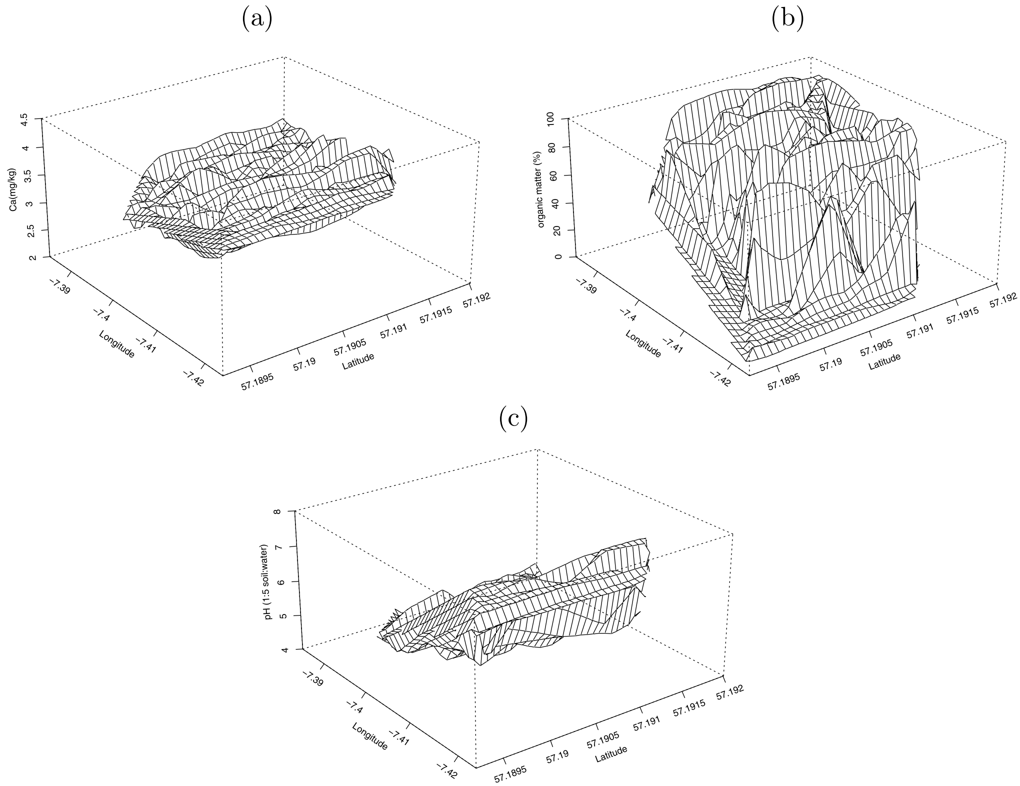

Figure 1. Componentwise spatial perspectives of the machairs data set: (a) Ca density (mg/kg), (b) organic matter concentration (in %) and (c) soil acidity (pH).

our methods to the resulting residuals, for $q = 0.12$ (bandwidth 1.7), 0.50 (bandwidth 1.3) and 0.88 (bandwidth 1.5). This yields the conditional regression quantile surfaces shown in Figure 2.

Due to the impossibility of plotting three q-values in one figure, these figures, however, are not easily readable, and we therefore also provide, in Figure 3, simultaneous plots of the same quantiles, (a) against (detrended) Ca density for three chosen values of organic matter concentration, and (b) against (detrended) organic matter concentration for three chosen values of Ca density – along with the corresponding estimated conditional mean. These graphs clearly show that the (positive) dependence of soil pH on Ca density and its (negative) dependence on organic matter concentrations are not linear. For low Ca densities (irrespective of organic matter concentrations) and high organic matter concentrations (irrespective of Ca densities), pH is uniformly low (i.e., the soil is rather acid), with pretty limited impact of the covariates. In particular, a minimal Ca density apparently is required, whatever the organic matter concentration, for inducing any noticeable acidity reduction effect (see Figure 3(a) for this threshold effect); on the other hand, median pH values are pretty stable (low pH values, hence high acidity) for high organic matter concentrations (see Figure 3(b)). Conditional pH distributions moreover look highly asymmetric

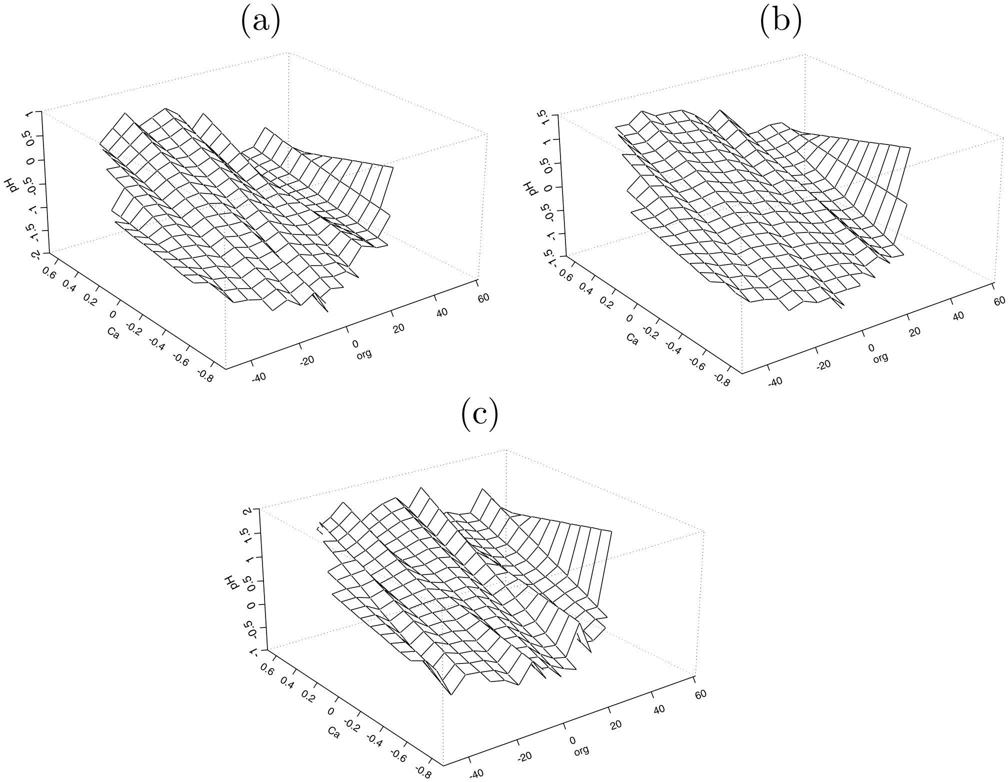

Figure 2. Estimated conditional quantiles of order q for soil acidity (pH), conditional on Ca density (mg/kg) and organic matter concentration (in %) for (a) $q = 0.12$, (b) $q = 0.50$ and (c) $q = 0.88$ (all variables detrended).

and highly "heteroskedastic", with much higher spread in right-hand tails (higher uncertainty on alkalinity) than in the left-hand ones (less uncertainty on acidity). Such facts could not be revealed by a traditional study of conditional means; neither would they be revealed by a simpler LAD estimation of conditional location. For a localized interpretation at given site $\mathbf{i}$, however, the estimated trend also should be taken into account, as explained at the end of Section 1.

Asymmetry of the densities involved is confirmed by Figure 4, where kernel estimates of marginal densities (after preliminary detrending) are provided. Those estimates indicate that pH measurements exhibit a strongly bimodal profile, meaning that a simple study of conditional means or conditional medians, contrary to our method, is bound to miss some of the essential features of the data set.

In this analysis, we restricted ourselves to conditioning on the two covariates, treating the spatial dependence as a nuisance. Further analysis of the data set might be carried out by introducing neighbouring observations into the set of covariates, which is made possible by our theoretical results. This is likely to improve the results, but also would increase the dimension of the covari-

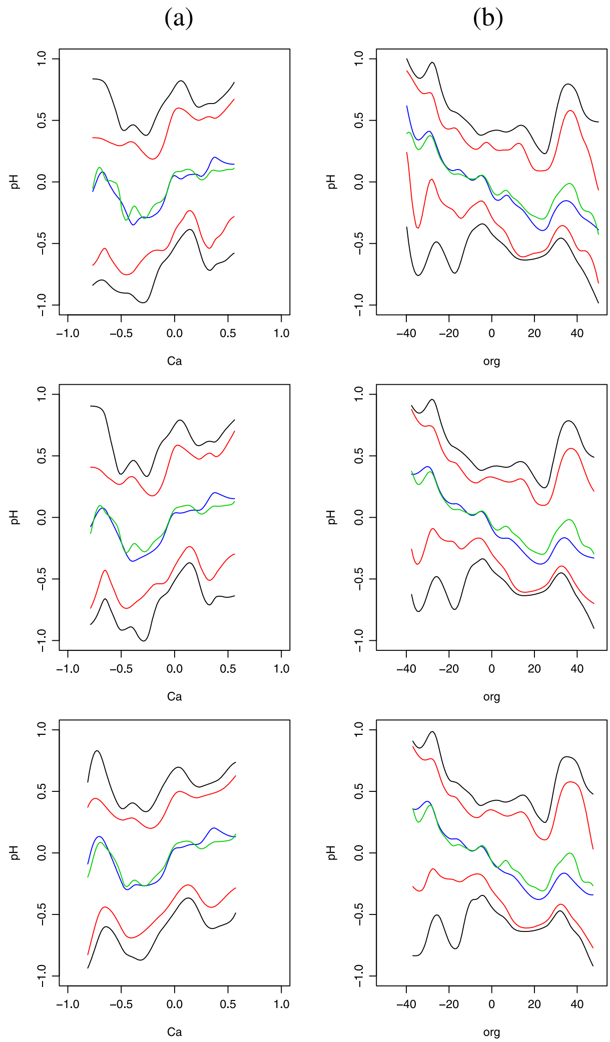

Figure 3. Estimated conditional quantiles of order $q = 0.03$ and 0.97 (black), $q = 0.12$ and 0.88 (red) and $q = 0.50$ (blue) and estimated conditional mean (solid green) for soil acidity (pH), (a) conditional on Ca density (mg/kg) and organic matter concentrations −28.5, 0.0 and 39.6, and (b) conditional on organic matter concentration and Ca densities −0.82, 0.00 and 0.67 (all variables detrended).

ate space. Semi-parametric dimension-reduction techniques then should be considered, as in [19] and [38]. Our purpose here was voluntarily limited to a simple illustration of the spatial regression quantile methods described in Section 2. Limited as it is, we hope this short study provides a good picture of how our method may provide a better understanding of complex spatial processes.

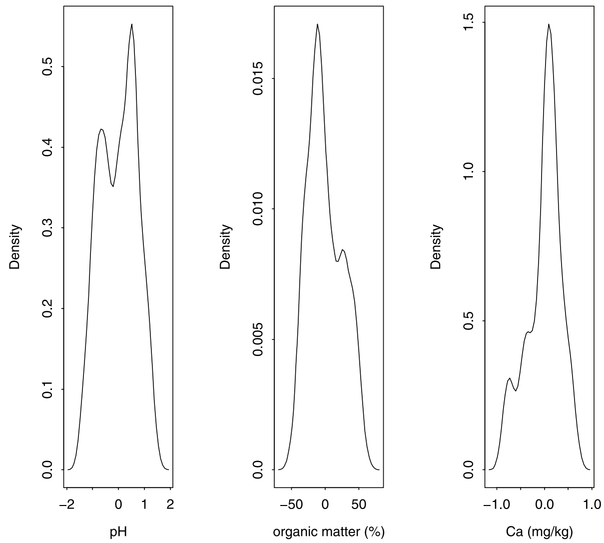

Figure 4. Kernel estimates of the marginal densities of soil acidity (pH), organic matter concentration (in %) and Ca density (mg/kg) (after preliminary detrending).

Appendix: Proofs

A.1. A preliminary lemma

The following lemma is an improved version of the *cross-term inequality* of Lemma 5.2 of [24], adapted to the quantile regression context, and plays a crucial role in the subsequent sections. For the sake of generality, and in order for this lemma to apply beyond the specific framework of this paper, we do not necessarily assume that the mixing coefficient take the form imposed in Assumption (A3). Let $\{(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}}); \mathbf{j} \in \mathbb{Z}^N\}$ denote a stationary spatial process with general mixing coefficient

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

Let $(y, \mathbf{x}) \mapsto \tilde{b}(y, \mathbf{x})$ be a bounded measurable function defined on $\mathbb{R}^1 \times \mathbb{R}^d$. Set

$$\eta_{\mathbf{j}}(\mathbf{x}) := \tilde{b}(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}}) K\bigl((\mathbf{x} - \mathbf{X}_{\mathbf{j}})/h_{\mathbf{n}}\bigr), \qquad \Delta_{\mathbf{j}}(\mathbf{x}) := \eta_{\mathbf{j}}(\mathbf{x}) - \mathrm{E}\eta_{\mathbf{j}}(\mathbf{x}),$$

and $\tilde{R}(\mathbf{x}) := (\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1} \sum_{\mathbf{i} \neq \mathbf{j} \in \mathcal{I}_{\mathbf{n}}} \mathrm{E}[\Delta_{\mathbf{i}}(\mathbf{x}) \Delta_{\mathbf{j}}(\mathbf{x})]$. For any $\mathbf{c}_{\mathbf{n}} := (c_{\mathbf{n}1}, \ldots, c_{\mathbf{n}N}) \in \mathbb{Z}^N$ with $1 < c_{\mathbf{n}k} < n_k$ for all $k = 1, \ldots, N$, define $\tilde{J}_1(\mathbf{x}) := h_{\mathbf{n}}^{2d} \prod_{k=1}^N (n_k c_{\mathbf{n}k})$ and

$$\tilde{J}_2(\mathbf{x}) := \widehat{\mathbf{n}} \sum_{k=1}^{N} \sum_{\substack{|j_s|=1 \\ s=1,\ldots,k-1}}^{n_s} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{\substack{|j_s|=1 \\ s=k+1,\ldots,N}}^{n_s} \varphi(j_1, \ldots, j_N).$$

Lemma A.1 (Cross-term lemma). *Under Assumptions* (A1), (A2) *and* (A4), *there exists a constant* $C > 0$ *such that* $|\tilde{R}(\mathbf{x})| \le C(\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1}[\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})]$. *If, furthermore,* $\varphi(j_1, \dots, j_N)$ *takes the form* $\varphi(\|\mathbf{j}\|)$, *then* $\tilde{J}_2(\mathbf{x}) \le C\widehat{\mathbf{n}} \sum_{k=1}^N \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1}\varphi(t)$.

Proof. The main idea of the proof is similar to that of Lemma 5.2 of [24], though details are different. We only briefly sketch it here. Writing $Z_{\mathbf{j}}$ for $\tilde{b}(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}})$, we have $\eta_{\mathbf{j}}(\mathbf{x}) = Z_{\mathbf{j}} K((\mathbf{x} - \mathbf{X}_{\mathbf{j}})/h_{\mathbf{n}})$, where $|Z_{\mathbf{j}}|$ is bounded by some $L > 0$. For $\mathbf{i} \neq \mathbf{j}$, letting $K_{\mathbf{n}}(\mathbf{x}) := (1/h_{\mathbf{n}}^d)K(\mathbf{x}/h_{\mathbf{n}})$,

$$h_{\mathbf{n}}^{-d}[\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})] = h_{\mathbf{n}}^d \iint K_{\mathbf{n}}(\mathbf{x}-\mathbf{u})K_{\mathbf{n}}(\mathbf{x}-\mathbf{v}) \times \left\{g_{1\mathbf{ij}}(\mathbf{u},\mathbf{v}) f_{\mathbf{i},\mathbf{j}}(\mathbf{u},\mathbf{v}) - g_1^{(1)}(\mathbf{u}) g_1^{(1)}(\mathbf{v}) f(\mathbf{u}) f(\mathbf{v})\right\} \mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v},$$

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

$$\left|g_{1\mathbf{ij}}(\mathbf{u},\mathbf{v}) f_{\mathbf{i},\mathbf{j}}(\mathbf{u},\mathbf{v}) - g_1^{(1)}(\mathbf{u}) g_1^{(1)}(\mathbf{v}) f(\mathbf{u}) f(\mathbf{v})\right| \le L^2 |f_{\mathbf{i},\mathbf{j}}(\mathbf{u},\mathbf{v}) - f(\mathbf{u})f(\mathbf{v})| + 2L^2 f(\mathbf{u})f(\mathbf{v}).$$

It then follows from Assumption (A1) and the Lebesgue density Theorem (see Chapter 2 of [13]) that

$$\begin{aligned} h_{\mathbf{n}}^{-d}|\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})| &\le h_{\mathbf{n}}^d \iint K_{\mathbf{n}}(\mathbf{x}-\mathbf{u})K_{\mathbf{n}}(\mathbf{x}-\mathbf{v}) L^2 |f_{\mathbf{i},\mathbf{j}}(\mathbf{u},\mathbf{v}) - f(\mathbf{u})f(\mathbf{v})| \,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v} \\ &\quad + 2h_{\mathbf{n}}^d \iint K_{\mathbf{n}}(\mathbf{x}-\mathbf{u})K_{\mathbf{n}}(\mathbf{x}-\mathbf{v}) L^2 f(\mathbf{u})f(\mathbf{v})\} \,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{v} \\ &\le C h_{\mathbf{n}}^d L^2 = C h_{\mathbf{n}}^d. \end{aligned} \tag{A.1}$$

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

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

and

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

Clearly, $\mathrm{Card}(\mathcal{S}_1) \le 2^N \widehat{\mathbf{n}} \prod_{k=1}^N c_{\mathbf{n}k}$. Splitting $\tilde{R}(\mathbf{x})$ into $(\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1}(J_1 + J_2)$, with $J_\ell := \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}_\ell} \mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})$, $\ell = 1, 2$, it follows from (A.1) that

$$|J_1| \le C h_{\mathbf{n}}^{2d}\,\mathrm{Card}(\mathcal{S}_1) \le 2^N C h_{\mathbf{n}}^{2d} \widehat{\mathbf{n}} \prod_{k=1}^N c_{\mathbf{n}k}. \tag{A.2}$$

Turning to J_2, we have $|J_2| \le \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}_2} |\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})|$. Davydov's inequality (cf. Lemma 2.1 of [48]) and the boundedness of $\Delta_{\mathbf{i}}(\mathbf{x})$ yield $|\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})| \le C\varphi(\mathbf{j} - \mathbf{i})$. Hence,

$$|J_2| \le C \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}_2} \varphi(\mathbf{j} - \mathbf{i}) =: C\Sigma_2, \qquad \text{say.}$$

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

$$\begin{aligned}\mathcal{S}(\ell_1, \dots, \ell_N) := \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_\mathbf{n} : |j_k - i_k| > c_{\mathbf{n}k} \text{ if } \ell_k = 1 \\ \text{and } |j_k - i_k| \le c_{\mathbf{n}k} \text{ if } \ell_k = 0, k = 1, \dots, N\}\end{aligned}$$

and

$$V(\ell_1, \dots, \ell_N) := \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}(\ell_1,\dots,\ell_N)} \varphi(\mathbf{j} - \mathbf{i}).$$

Then,

$$\Sigma_2 = \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}_2} \varphi(\mathbf{j} - \mathbf{i}) = \sum_{\mathbf{0} \neq \boldsymbol{\ell} \in \{0,1\}^N} V(\ell_1, \dots, \ell_N)$$

where, as in equation (5.11) of [24],

$$V(\ell_1, \ell_2, \dots, \ell_N) \le \widehat{\mathbf{n}} \sum_{|j_1|} \cdots \sum_{|j_k|} \cdots \sum_{|j_N|} \varphi(j_1, \dots, j_N),$$

with the sums $\sum_{|j_k|}$ running over all j_k's such that $1 \le |j_k| \le n_k$ when $\ell_k = 0$, such that $c_{\mathbf{n}1} \le |j_k| \le n_k$ when $\ell_k = 1$. Since all terms are non-negative, for $1 \le c_{\mathbf{n}k} \le n_k$, sums of the form $\sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \cdots$ are smaller than those of the form $\sum_{|j_k|=1}^{n_k} \cdots$, and

$$|J_2| \le C\widehat{\mathbf{n}} \sum_{k=1}^{N} \sum_{|j_1|=1}^{n_1} \cdots \sum_{|j_{k-1}|=1}^{n_{k-1}} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{|j_{k+1}|=1}^{n_{k+1}} \cdots \sum_{|j_N|=1}^{n_N} \varphi(j_1, \dots, j_N). \tag{A.3}$$

The first part of the lemma is a consequence of (A.2) and (A.3). The second part follows from the fact that, if $\varphi(j_1, \dots, j_N)$ depends on $\|\mathbf{j}\|$ only,

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

□

A.2. Proof of the Bahadur representation result

We first introduce some notation. Throughout, let $C > 0$ denote a generic constant. Let $\mathbf{X}_{h\mathbf{i}} := (\mathbf{X}_\mathbf{i} - \mathbf{x})/h_\mathbf{n}$, $\boldsymbol{\mathcal{X}}_{h\mathbf{i}} := (1, \mathbf{X}'_{h\mathbf{i}})'$, $K_\mathbf{i} := K(\mathbf{X}_{h\mathbf{i}})$, $H_\mathbf{n} = (\widehat{\mathbf{n}} h_\mathbf{n}^d)^{1/2}$,

$$\boldsymbol{\theta} := H_\mathbf{n}\big(a_0 - q_p(\mathbf{x}), h_\mathbf{n}\big(\mathbf{a}_1 - \dot{q}(\mathbf{x})\big)'\big)', \qquad \overline{\boldsymbol{\theta}}_\mathbf{n} := H_\mathbf{n}\big(\widehat{a}_0 - q_p(\mathbf{x}), h_\mathbf{n}\big(\widehat{\mathbf{a}}_1 - \dot{q}(\mathbf{x})\big)'\big)'$$

and $\tilde{\boldsymbol{\theta}} := H_\mathbf{n}(\tilde{a}_0 - q_p(\mathbf{x}), h_\mathbf{n}(\tilde{\mathbf{a}}_1 - \dot{q}(\mathbf{x}))')'$, where $(a_0, \mathbf{a}'_1)', (\tilde{a}_0, \tilde{\mathbf{a}}'_1)' \in \mathbb{R}^{1+d}$. With $Y^*_\mathbf{i}$ defined in Theorem 2.1, put $Y^*_{\mathbf{ni}}(\boldsymbol{\theta}) := Y^*_\mathbf{i} - \boldsymbol{\theta}'\boldsymbol{\mathcal{X}}_{h\mathbf{i}}/H_\mathbf{n}$, $T_{\mathbf{ni}} := (\dot{q}_p(\mathbf{x}))'\mathbf{X}_{h\mathbf{i}} h_\mathbf{n}$ and $U_{\mathbf{ni}} := U_{\mathbf{ni}}(\boldsymbol{\theta}) = T_{\mathbf{ni}} +$

$\boldsymbol{\theta}'\boldsymbol{\mathcal{X}}_{h\mathbf{i}}/H_{\mathbf{n}}$. With this notation, $Y^*_{\mathbf{i}} = Y_{\mathbf{i}} - q_p(\mathbf{x}) - T_{\mathbf{ni}}$ and $Y^*_{\mathbf{ni}}(\boldsymbol{\theta}) = Y_{\mathbf{i}} - q_p(\mathbf{x}) - U_{\mathbf{ni}}(\boldsymbol{\theta}) = Y_{\mathbf{i}} - a_0 - \mathbf{a}_1'(\mathbf{X}_{\mathbf{i}} - \mathbf{x})$. Since K is a bounded function with bounded support,

$$\|\mathbf{X}_{h\mathbf{i}}\| \le C \quad \text{and} \quad \|\boldsymbol{\mathcal{X}}_{h\mathbf{i}}\| \le C \qquad \text{when } K_{\mathbf{i}} > 0. \tag{A.4}$$

When $\|\boldsymbol{\theta}\| \le M$ and $K_{\mathbf{i}} > 0$, $|T_{\mathbf{ni}}| \le Ch_{\mathbf{n}}$ and $|U_{\mathbf{ni}}| \le Ch_{\mathbf{n}} + CH_{\mathbf{n}}^{-1} \to 0$ as $\mathbf{n} \to \infty$. It follows from (2.3) that

$$\overline{\boldsymbol{\theta}}_{\mathbf{n}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{1+d}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \rho_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta}))K_{\mathbf{i}}. \tag{A.5}$$

Finally, define $\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) := H_{\mathbf{n}}^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta}))\boldsymbol{\mathcal{X}}_{h\mathbf{i}}K_{\mathbf{i}}$. The following lemma provides an asymptotic representation result for sequences $\boldsymbol{\theta}_{\mathbf{n}}$ of solutions of $\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) = \mathbf{0}$ or, more generally, for any sequence $\boldsymbol{\theta}_{\mathbf{n}}$ such that $\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) = o_{\mathrm{P}}(1)$ as $\mathbf{n} \to \infty$. This spatial version of Lemma A.4 of [32] plays a key role in the proof of Theorem 2.1.

Lemma A.2. *Let* $\boldsymbol{\delta} \mapsto \mathbf{V}_{\mathbf{n}}(\boldsymbol{\delta})$ *satisfy* (i) $\boldsymbol{\delta}'\mathbf{V}_{\mathbf{n}}(\lambda\boldsymbol{\delta}) \le \boldsymbol{\delta}'\mathbf{V}_{\mathbf{n}}(\boldsymbol{\delta})$ *for all* $\lambda \ge 1$ *and* (ii) $\sup_{\|\boldsymbol{\delta}\| \le M} \|\mathbf{V}_{\mathbf{n}}(\boldsymbol{\delta}) + f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x})\mathbf{D}\boldsymbol{\delta} - \mathbf{A}_{\mathbf{n}}\| = o_{\mathrm{P}}(1)$ *as* $\mathbf{n} \to \infty$, *where* $\|\mathbf{A}_{\mathbf{n}}\|$ *is* $O_{\mathrm{P}}(1)$, $0 < M < \infty$, $f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x}) > 0$ *and* $\mathbf{D}$ *is a positive definite matrix. Suppose that* $\boldsymbol{\delta}_{\mathbf{n}}$ *is such that* $\|\mathbf{V}_{\mathbf{n}}(\boldsymbol{\delta}_{\mathbf{n}})\| = o_{\mathrm{P}}(1)$. *Then,* $\|\boldsymbol{\delta}_{\mathbf{n}}\| = O_{\mathrm{P}}(1)$ *and* $\boldsymbol{\delta}_{\mathbf{n}} = [f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x})]^{-1}\mathbf{D}^{-1}\mathbf{A}_{\mathbf{n}} + o_{\mathrm{P}}(1)$ *as* $\mathbf{n} \to \infty$.

Proof. The proof follows along the same lines as in [32], page 809; details are left to the reader. □

In order to establish the Bahadur representation result of Theorem 2.1, it is now sufficient to check that the assumptions of Lemma A.2 are satisfied. To do this, we repeatedly use the next lemma, the proof of which is essentially the same as in the time series case (cf. [37]), and hence is omitted.

Lemma A.3. *Let Assumptions* (A1)(ii)–(iii) *and* (A2) *hold. Then, for* $\mathbf{n}$ *large enough*,

$$\mathrm{E}[|\psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta})) - \psi_p(Y^*_{\mathbf{ni}}(\tilde{\boldsymbol{\theta}}))|K_{\mathbf{i}}] \le C\mathrm{E}\big[I_{(|Y^*_{\mathbf{ni}}(\tilde{\boldsymbol{\theta}})| < C\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|/H_{\mathbf{n}})}K_{\mathbf{i}}\big] \le C\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|h_{\mathbf{n}}^d/H_{\mathbf{n}},$$

$$\mathrm{E}[|\psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta})) - \psi_p(Y^*_{\mathbf{ni}}(\tilde{\boldsymbol{\theta}}))|^2K_{\mathbf{i}}^2] \le C\mathrm{E}\big[I_{(|Y^*_{\mathbf{ni}}(\tilde{\boldsymbol{\theta}})| < C\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|/H_{\mathbf{n}})}K_{\mathbf{i}}^2\big] \le C\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|h_{\mathbf{n}}^d/H_{\mathbf{n}}$$

for any $\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \le M\}$.

Lemma A.4. *Under the conditions of Theorem* 2.1,

$$\sup_{\|\boldsymbol{\theta}\| \le M} \big\|\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) - \mathbf{V}_{\mathbf{n}}(\mathbf{0}) - \mathrm{E}\big(\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) - \mathbf{V}_{\mathbf{n}}(\mathbf{0})\big)\big\| = o_{\mathrm{P}}(1).$$

Proof. The proof is divided into two steps. The first step consists in proving that

$$\big\|\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) - \mathbf{V}_{\mathbf{n}}(\mathbf{0}) - \mathrm{E}\big(\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) - \mathbf{V}_{\mathbf{n}}(\mathbf{0})\big)\big\| = o_{\mathrm{P}}(1). \tag{A.6}$$

for any fixed $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta}\| \le M$. Note that

$$\mathbf{V_n}(\boldsymbol{\theta}) - \mathbf{V_n}(\mathbf{0}) = H_{\mathbf{n}}^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} [\psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta})) - \psi_p(Y_{\mathbf{i}}^*)] \boldsymbol{\mathcal{X}}_{h\mathbf{i}} K_{\mathbf{i}} =: H_{\mathbf{n}}^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathbf{V_{ni}}(\boldsymbol{\theta}), \tag{A.7}$$

where $\mathbf{V_{ni}}(\boldsymbol{\theta}) := (V_{\mathbf{ni}}^0(\boldsymbol{\theta}), (\mathbf{V}_{\mathbf{ni}}^1(\boldsymbol{\theta}))')'$, with

$$V_{\mathbf{ni}}^0(\boldsymbol{\theta}) := [\psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta})) - \psi_p(Y_{\mathbf{i}}^*)] K_{\mathbf{i}} \quad \text{and} \quad \mathbf{V}_{\mathbf{ni}}^1(\boldsymbol{\theta}) = [\psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta})) - \psi_p(Y_{\mathbf{i}}^*)] \mathbf{X}_{h\mathbf{i}} K_{\mathbf{i}}.$$

Then, from (A.7), the left-hand side of (A.6) is bounded by

$$H_{\mathbf{n}}^{-1} \left| \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \big(V_{\mathbf{ni}}^0(\boldsymbol{\theta}) - \mathrm{E} V_{\mathbf{ni}}^0(\boldsymbol{\theta})\big) \right| + H_{\mathbf{n}}^{-1} \left\| \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \big(\mathbf{V}_{\mathbf{ni}}^1(\boldsymbol{\theta}) - \mathrm{E}\mathbf{V}_{\mathbf{ni}}^1(\boldsymbol{\theta})\big) \right\| =: V_{\mathbf{n}}^0 + V_{\mathbf{n}}^1. \tag{A.8}$$

It follows from stationarity together with Lemma A.1 that

$$\begin{aligned} \mathrm{E}(V_{\mathbf{n}}^0)^2 &= (\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1} \left[\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \operatorname{var}(V_{\mathbf{ni}}^0(\boldsymbol{\theta})) + \sum_{\mathbf{i}\ne\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \operatorname{cov}(V_{\mathbf{ni}}^0(\boldsymbol{\theta}), V_{\mathbf{nj}}^0(\boldsymbol{\theta})) \right] \\ &\le h_{\mathbf{n}}^{-d} \operatorname{var}(V_{\mathbf{n}1}^0(\boldsymbol{\theta})) + (\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1} [\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})], \end{aligned} \tag{A.9}$$

where $\tilde{J}_1(\mathbf{x}) \le C\widehat{\mathbf{n}} h_{\mathbf{n}}^{2d} \prod_{k=1}^N c_{\mathbf{n}k}$ and $\tilde{J}_2(\mathbf{x}) \le C\widehat{\mathbf{n}} \sum_{k=1}^N \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1}\varphi(t)$, as implied by Lemma A.1. Here $c_{\mathbf{n}k}$, $k = 1, \dots, N$, are positive integers depending on $\mathbf{n}$, to be specified later on. In order to bound (A.9), we apply Lemma A.3 with $\tilde{\boldsymbol{\theta}} = \mathbf{0}$; for $\|\boldsymbol{\theta}\| \le M$, $\operatorname{var}(V_{\mathbf{n}1}^0(\boldsymbol{\theta})) \le \mathrm{E}[(V_{\mathbf{n}1}^0)^2] = \mathrm{E}[|\psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta})) - \psi_p(Y_{\mathbf{i}}^*)|^2 K_{\mathbf{i}}^2] \le C h_{\mathbf{n}}^d / H_{\mathbf{n}}$. Then it follows from (A.9) with $c_{\mathbf{n}k} = h_{\mathbf{n}}^{-d/a}$ for $k = 1, \dots, N$, that

$$\mathrm{E}[(V_{\mathbf{n}}^0)^2] \le C H_{\mathbf{n}}^{-1} + C h_{\mathbf{n}}^{(1-N/a)d} + C \sum_{k=1}^N c_{\mathbf{n}k}^a \sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1}\varphi(t) = \mathrm{o}(1), \tag{A.10}$$

in view of Assumption (A3) and the fact that $h_{\mathbf{n}} \to 0$, $\widehat{\mathbf{n}} h_{\mathbf{n}}^d \to \infty$ and $a > N$. Similar to (A.10), we have $\mathrm{E}(V_{\mathbf{n}}^1)^2 = \mathrm{o}(1)$ which, with (A.8) and (A.10), implies (A.6).

The second step consists in establishing the uniform consistency with respect to $\|\boldsymbol{\theta}\| \le M$ by a chaining argument. Decompose $\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \le M\}$ into cubes based on the grid $(j_1\gamma M, \dots, j_{d+1}\gamma M)$, $j_i = 0, \pm 1, \dots, \pm[1/\gamma] + 1$, where $[1/\gamma]$ denotes the integer part of $1/\gamma$ and γ is a small positive number that does not depend on $\mathbf{n}$. Let $\mathbf{R}(\boldsymbol{\theta})$ be the lower vertex of the cube that contains $\boldsymbol{\theta}$. Clearly, $\|\mathbf{R}(\boldsymbol{\theta}) - \boldsymbol{\theta}\| \le C\gamma$ and the number of elements of $\{\mathbf{R}(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| \le M\}$ is finite. Then

$$\sup_{\|\boldsymbol{\theta}\| \le M} \big\| \mathbf{V_n}(\boldsymbol{\theta}) - \mathbf{V_n}(\mathbf{0}) - \mathrm{E}\big(\mathbf{V_n}(\boldsymbol{\theta}) - \mathbf{V_n}(\mathbf{0})\big) \big\| \le V_{\mathbf{n}1}^* + V_{\mathbf{n}2}^* + V_{\mathbf{n}3}^* \tag{A.11}$$

where, following (A.6), $V_{\mathbf{n}1}^* := \sup_{\|\boldsymbol{\theta}\| \le M} \|\mathbf{V_n}(\mathbf{R}(\boldsymbol{\theta})) - \mathbf{V_n}(\mathbf{0}) - \mathrm{E}(\mathbf{V_n}(\mathbf{R}(\boldsymbol{\theta})) - \mathbf{V_n}(\mathbf{0}))\|$ is $o_{\mathrm{P}}(1)$, $V_{\mathbf{n}2}^* := \sup_{\|\boldsymbol{\theta}\| \le M} \|\mathbf{V_n}(\boldsymbol{\theta}) - \mathbf{V_n}(\mathbf{R}(\boldsymbol{\theta}))\|$, and $V_{\mathbf{n}3}^* := \sup_{\|\boldsymbol{\theta}\| \le M} \|\mathrm{E}(\mathbf{V_n}(\boldsymbol{\theta}) - \mathbf{V_n}(\mathbf{R}(\boldsymbol{\theta})))\|$. Using (A.4) and, for $\|\boldsymbol{\theta}\| \le M$, applying Lemma A.3 with $\tilde{\boldsymbol{\theta}} = \mathbf{R}(\boldsymbol{\theta})$ for $\mathbf{n}$ large enough,

$$V^*_{\mathbf{n}3} \le C H_{\mathbf{n}}^{-1} \widehat{\mathbf{n}} \sup_{\|\boldsymbol{\theta}\| \le M} \mathrm{E}[|\psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta})) - \psi_p(Y^*_{\mathbf{ni}}(\mathbf{R}(\boldsymbol{\theta})))|K_{\mathbf{i}}]$$
$$\le C \sup_{\|\boldsymbol{\theta}\| \le M} \|\boldsymbol{\theta} - \mathbf{R}(\boldsymbol{\theta})\| \le C\gamma.$$

Therefore, letting $\mathbf{n} \to \infty$ and $\gamma \to 0$, we have $V^*_{\mathbf{n}3} = \mathrm{o}(1)$.

Let $B_{\mathbf{i}}(\boldsymbol{\theta}) := I_{(|Y^*_{\mathbf{ni}}(\boldsymbol{\theta})| < C\gamma/H_{\mathbf{n}})} \|\boldsymbol{\mathcal{X}}_{h\mathbf{i}}\| K_{\mathbf{i}}$. Since $|I_{(y<a)} - I_{(y<0)}| \le I_{(|y| \le |a|)}$,

$$V^*_{\mathbf{n}2} \le \sup_{\|\boldsymbol{\theta}\| \le M} \|\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) - \mathbf{V}_{\mathbf{n}}(\mathbf{R}(\boldsymbol{\theta}))\| \le C \sup_{\|\boldsymbol{\theta}\| \le M} H_{\mathbf{n}}^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} B_{\mathbf{i}}(\mathbf{R}(\boldsymbol{\theta})) \le B_{\mathbf{n}1} + B_{\mathbf{n}2},$$

where, by the same argument as above, $B_{n1} := C \sup_{\|\boldsymbol{\theta}\| \le M} H_{\mathbf{n}}^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \mathrm{E} B_{\mathbf{i}}(\mathbf{R}(\boldsymbol{\theta})) = \mathrm{o}(1)$, and, similar to (A.10), $B_{n2} := C \sup_{\|\boldsymbol{\theta}\| \le M} |H_{\mathbf{n}}^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} (B_{\mathbf{i}}(\mathbf{R}(\boldsymbol{\theta})) - \mathrm{E} B_{\mathbf{i}}(\mathbf{R}(\boldsymbol{\theta})))| = \mathrm{o}_{\mathrm{P}}(1)$. Thus, $V^*_{\mathbf{n}2} = \mathrm{o}_{\mathrm{P}}(1)$, and Lemma 4.4 follows from (A.11). □

Lemma A.5. *Let* $\mathbf{D} := f_{\mathbf{X}}(\mathbf{x}) \,\mathrm{diag}(1, \int \mathbf{u}\mathbf{u}' K(\mathbf{u})\, \mathrm{d}\mathbf{u})$. *Under Assumptions* (A1)(iii) *and* (A2), $\sup_{\|\boldsymbol{\theta}\| \le M} \|\mathrm{E}(\mathbf{V}_{\mathbf{n}}(\boldsymbol{\theta}) - \mathbf{V}_{\mathbf{n}}(\mathbf{0})) + f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x}) \mathbf{D}\boldsymbol{\theta}\| = \mathrm{o}(1)$.

Proof. The proof again is similar to that in the time series case (see [37]). □

Lemma A.6. *Denote by* $\overline{\boldsymbol{\theta}}_{\mathbf{n}}$ *the minimizer in* (A.5). *Then*, $\|\mathbf{V}_{\mathbf{n}}(\overline{\boldsymbol{\theta}}_{\mathbf{n}})\|$ *is* $\mathrm{o}_{\mathrm{P}}(H_{\mathbf{n}}^{-1})$.

Proof. The proof is similar to that of Lemma A.2 of [44]. □

Lemma A.7. *Under Assumptions* (A1) *and* (A2), *if* $a \ge N$ *and* $h_{\mathbf{n}} \to 0$,

$$\mathrm{E}\big[\big(\mathbf{c}'\mathbf{V}_{\mathbf{n}}(\mathbf{0}) - \mathbf{c}'\mathrm{E}\mathbf{V}_{\mathbf{n}}(\mathbf{0})\big)^2\big] \to p(1-p) f_{\mathbf{X}}(\mathbf{x}) \int (c_0 + \mathbf{c}_1'\mathbf{u})^2 K^2(\mathbf{u})\, \mathrm{d}\mathbf{u}$$

as $\mathbf{n} \to \infty$, *where* $\mathbf{c} = (c_0, \mathbf{c}_1')' \in \mathbb{R}^{1+d}$.

Proof. Let $v_{\mathbf{i}} := \psi_p(Y^*_{\mathbf{i}})(c_0 + c_1' \mathbf{X}_{h\mathbf{i}}) K_{\mathbf{i}}$. Lemma A.1 with $c_{\mathbf{n}k} = h_{\mathbf{n}}^{-d/a}$ for $k = 1, \dots, N$ yields

$$\mathrm{E}\big[\big(\mathbf{c}'\mathbf{V}_{\mathbf{n}}(\mathbf{0}) - \mathbf{c}'\mathrm{E}\mathbf{V}_{\mathbf{n}}(\mathbf{0})\big)^2\big] = (\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1} \Bigg[\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \mathrm{var}(v_{\mathbf{i}}) + \sum_{\mathbf{i} \ne \mathbf{j} \in \mathcal{I}_{\mathbf{n}}} \mathrm{cov}(v_{\mathbf{i}}, v_{\mathbf{j}})\Bigg] \tag{A.12}$$
$$= h_{\mathbf{n}}^{-d} \mathrm{var}(v_{\mathbf{1}}) + \mathrm{O}(1) h_{\mathbf{n}}^{(1-N/a)d} + \mathrm{O}(1) \sum_{k=1}^{N} c_{\mathbf{n}k}^a \sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1} \varphi(t)$$
$$=: v_{n1} + v_{n2} + v_{n3}. \tag{A.13}$$

Theorem 3 of [13] (page 8) entails

$$\begin{aligned}\mathrm{E}\big[I_{(Y_1^*<0)}(c_0+\mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})^2K_{\mathbf{1}}^2\big] &= \mathrm{E}\big[F_{Y|\mathbf{X}}\big(q_p(\mathbf{x})+\dot{q}_p(\mathbf{X_1}-\mathbf{x})|\mathbf{X_1}\big)(c_0+\mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})^2K_{\mathbf{1}}^2\big]\\ &\longrightarrow pf_{\mathbf{X}}(\mathbf{x})\int(c_0+\mathbf{c}_1'\mathbf{u})^2K^2(\mathbf{u})\,\mathrm{d}\mathbf{u}\end{aligned}$$

and $\mathrm{E}[I_{(Y_1^*<0)}(c_0+\mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})K_{\mathbf{1}}] \longrightarrow pf_{\mathbf{X}}(\mathbf{x})\int(c_0+\mathbf{c}_1'\mathbf{u})K(\mathbf{u})\,\mathrm{d}\mathbf{u}$. This in turn implies

$$\begin{aligned}h_{\mathbf{n}}^{-d}\mathrm{E}[v_{\mathbf{1}}^2] &= \mathrm{E}\big[\big(p^2-2pI_{(Y_{\mathbf{1}}^*<0)}+I_{(Y_{\mathbf{1}}^*<0)}\big)(c_0+\mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})^2K_{\mathbf{1}}^2\big]\\ &\longrightarrow p(1-p)f_{\mathbf{X}}(\mathbf{x})\int(c_0+\mathbf{c}_1'\mathbf{u})^2K^2(\mathbf{u})\,\mathrm{d}\mathbf{u},\end{aligned}$$

and

$$\begin{aligned}h_{\mathbf{n}}^{-d}\mathrm{E}[v_{\mathbf{1}}] &= \mathrm{E}\big[\big(p-I_{(Y_{\mathbf{1}}^*<0)}\big)(c_0+\mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})K_{\mathbf{1}}\big]\\ &\longrightarrow (p-p)f_{\mathbf{X}}(\mathbf{x})\int(c_0+\mathbf{c}_1'\mathbf{u})K(\mathbf{u})\,\mathrm{d}\mathbf{u}=0.\end{aligned}$$

Hence, $v_{n1}=h_{\mathbf{n}}^{-d}\mathrm{E}[v_1^2]-h_{\mathbf{n}}^{-d}(\mathrm{E}v_1)^2 \longrightarrow p(1-p)f_X(\mathbf{x})\int(c_0+\mathbf{c}_1'\mathbf{u})^2K^2(\mathbf{u})\,\mathrm{d}\mathbf{u}$. On the other hand, it clearly follows from the fact that $h_{\mathbf{n}}\to 0$ and Assumption (A3) with $a>N$, that $|v_{n2}+v_{n3}|=\mathrm{O}(1)h_{\mathbf{n}}^{(1-N/a)d}+\mathrm{O}(1)\sum_{k=1}^N c_{\mathbf{n}k}^a\sum_{t=c_{\mathbf{n}k}}^{\infty}t^{N-1}\varphi(t)\longrightarrow 0$. The result follows. □

Proof of Theorem 2.1. As already mentioned, it is sufficient to check that the conditions of Lemma A.2 are fulfilled. First we note that Lemmas A.4 and A.5 lead to (ii) of Lemma A.2. Also, it follows from Lemma A.6 together with Assumptions (A2) and (A3) that $\|\mathbf{V_n}(\overline{\boldsymbol{\theta}}_{\mathbf{n}})\|=o_{\mathrm{P}}(1)$. Take $A_{\mathbf{n}}=\mathbf{V_n}(\mathbf{0})$. Then it is clear from Lemma A.7 that $A_{\mathbf{n}}=\mathrm{O}_{\mathrm{P}}(1)$. Since $y\mapsto\psi_p(y)$ is monotone increasing, the function $\lambda\mapsto-\boldsymbol{\theta}'\mathbf{V_n}(\lambda\boldsymbol{\theta})=H_{\mathbf{n}}^{-1}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\psi_p(y_{\mathbf{i}}^*-\lambda\boldsymbol{\theta}'\mathcal{X}_{h\mathbf{i}}/H_{\mathbf{n}})(-\boldsymbol{\theta}'\boldsymbol{\mathcal{X}}_{h\mathbf{i}})K_{\mathbf{i}}$ also is. Therefore, condition (i) of Lemma A.2 holds. The theorem follows. □

A.3. Proof of asymptotic normality

On the basis of the Bahadur representation of Theorem 2.1, the asymptotic normality of our estimators in Theorems 2.2–2.6 follows exactly as in the corresponding proofs for mean regression in [24], with the "cross-term" Lemma A.1 replacing the corresponding Lemma A.2 in that paper, yielding the asymptotic normality with the bias (i.e., the expectation) of the first term on the right-hand side of (2.5) as

$$\begin{aligned}&\mathrm{E}\left[\frac{\eta_p(\mathbf{x})}{\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\psi_p(Y_{\mathbf{i}}^*)\begin{pmatrix}1\\ \dfrac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\end{pmatrix}K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right)\right]\\ &\quad=\frac{\eta_p(\mathbf{x})}{\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}}\widehat{\mathbf{n}}\mathrm{E}\left[\psi_p(Y_{\mathbf{i}}^*)\begin{pmatrix}1\\ \dfrac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\end{pmatrix}K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right)\right]\end{aligned}$$

$$= \eta_p(\mathbf{x})\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}h_{\mathbf{n}}^{-d}\mathrm{E}\Bigg[\big(F_{Y|\mathbf{X}}(q_p(\mathbf{X_i})|\mathbf{X_i}) - F_{Y|\mathbf{X}}\big(q_p(\mathbf{x}) + (\dot{q}_p(\mathbf{x}))'(\mathbf{X_i}-\mathbf{x})|\mathbf{X_i}\big)\big)$$

$$\times \begin{pmatrix} 1 \\ \frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}} \end{pmatrix} K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right)\Bigg]$$

$$= \sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}\left[\big(1+\mathrm{o}(1)\big)\frac{1}{2}\begin{pmatrix} B_0(\mathbf{x}) \\ \mathbf{B}_1(\mathbf{x}) \end{pmatrix} h_{\mathbf{n}}^2\right],$$

where the last equality is derived via a first-order Taylor expansion of $y \mapsto F_{Y|\mathbf{X}}(y|\cdot)$ and a second-order Taylor expansion of $\mathbf{x} \mapsto q_p(\mathbf{x})$ (these expansions exist in view of Assumptions (A1)(ii) and (A2)). The $(1+\mathrm{o}(1))$ factor is eliminated in Theorems 2.2–2.6 by using Assumption (B2). Details are omitted.

A.4. Proof Theorem 3.1

Recall that $N = 2$ has been assumed throughout this section. Following Hansen [25], $\hat{\mu}_Y(\mathbf{s})$ and $\hat{\boldsymbol{\mu}}_{\mathbf{X}}(\mathbf{s})$ are such that

$$\sup_{\mathbf{s}\in[0,1]^2} |\hat{\mu}_Y(\mathbf{s}) - \mu_Y(\mathbf{s})| = \mathrm{O_P}(\epsilon_{\mathbf{n}}) \quad \text{and} \quad \sup_{\mathbf{s}\in[0,1]^2} \|\hat{\boldsymbol{\mu}}_{\mathbf{X}}(\mathbf{s}) - \mu_{\mathbf{X}}(\mathbf{s})\| = \mathrm{O_P}(\epsilon_{\mathbf{n}}),$$

with $\epsilon_{\mathbf{n}} = (\ln\hat{\mathbf{n}}/(\hat{\mathbf{n}}g^2))^{1/2} + g^r =: \epsilon_{\mathbf{n}}^1 + \epsilon_{\mathbf{n}}^2$, where $\epsilon_{\mathbf{n}}^1$ is obtained as in the proof of Theorem 2 of Hansen [25] under Assumptions (C0), (C1), (C3) and (A3), while $\epsilon_{\mathbf{n}}^2$ readily follows from Assumptions (C1) and (C3). Therefore, we have

$$\max_{\mathbf{i}} |\hat{Y}_{\mathbf{i}} - Y_{\mathbf{i}}| = \mathrm{O_P}(\epsilon_{\mathbf{n}}) \quad \text{and} \quad \max_{\mathbf{i}} \|\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{X}_{\mathbf{i}}\| = \mathrm{O_P}(\epsilon_{\mathbf{n}}). \tag{A.14}$$

Hence $\Upsilon_{\mathbf{n}} := \max\{\max_{\mathbf{i}}|\hat{Y}_{\mathbf{i}} - Y_{\mathbf{i}}|, \max_{\mathbf{i}}\|\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{X}_{\mathbf{i}}\|\} = \mathrm{O_P}(\epsilon_{\mathbf{n}})$, that is, $\mathrm{P}[\Upsilon_{\mathbf{n}} > C\epsilon_{\mathbf{n}}] \to 0$ as $\hat{\mathbf{n}} \to \infty$ and $C \to \infty$. Thus we can assume that, with probability arbitrarily close to one, $\Upsilon_{\mathbf{n}} \le C\epsilon_{\mathbf{n}}$ for some C and $\mathbf{n}$ sufficiently large. Let

$$\hat{V}_{\mathbf{n}}(\boldsymbol{\theta}) := H_{\mathbf{n}}^{-1}\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\psi_p(\hat{Y}_{\mathbf{ni}}^*(\boldsymbol{\theta}))\hat{\mathcal{X}}_{h\mathbf{i}}\hat{K}_{\mathbf{i}}$$

with

$$\hat{Y}_{\mathbf{ni}}^*(\boldsymbol{\theta}) = \hat{Y}_{\mathbf{i}} - a_0 - \mathbf{a}_1'(\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x}), \qquad \hat{\mathcal{X}}_{h\mathbf{i}} = (1, \hat{\mathbf{X}}_{h\mathbf{i}}')',$$

$$\hat{\mathbf{X}}_{h\mathbf{i}} = \frac{\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x}}{h}, \qquad \hat{K}_{\mathbf{i}} = K\left(\frac{\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{x}}{h}\right).$$

We only need to show that $\sup_{|\boldsymbol{\theta}|\le M} |\hat{V}_{\mathbf{n}}(\boldsymbol{\theta}) - V_{\mathbf{n}}(\boldsymbol{\theta})| = o_{\mathrm{P}}(1)$, since Lemma A.3 with $\hat{V}_{\mathbf{n}}(\boldsymbol{\theta})$ instead of $V_{\mathbf{n}}(\boldsymbol{\theta})$ still applies. Now,

$$\hat{V}_{\mathbf{n}}(\boldsymbol{\theta}) - V_{\mathbf{n}}(\boldsymbol{\theta}) = H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} [\psi_p(\hat{Y}_{\mathbf{ni}}^*(\boldsymbol{\theta}))\hat{\mathcal{X}}_{h\mathbf{i}}\hat{K}_{\mathbf{i}} - \psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta}))\mathcal{X}_{h\mathbf{i}}K_{\mathbf{i}}] = B_{\mathbf{n}3} + B_{\mathbf{n},4}, \tag{A.15}$$

where

$$B_{\mathbf{n}3} := H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} [\psi_p(\hat{Y}_{\mathbf{ni}}^*(\boldsymbol{\theta})) - \psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta}))]\hat{\mathcal{X}}_{h\mathbf{i}}\hat{K}_{\mathbf{i}}$$

and

$$B_{\mathbf{n}4} := H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta}))[\hat{\mathcal{X}}_{h\mathbf{i}}\hat{K}_{\mathbf{i}} - \mathcal{X}_{h\mathbf{i}}K_{\mathbf{i}}].$$

Proceeding as in the proof of Lemma A.4 (from equation (A.11) on), by noting that $|\psi_p(\hat{Y}_{\mathbf{ni}}^*(\boldsymbol{\theta})) - \psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta}))| \le I_{\{|Y_{\mathbf{ni}}^*(\boldsymbol{\theta})|\le \Upsilon_{\mathbf{n}}\}}$, we obtain, since $\Upsilon_{\mathbf{n}} \le C\epsilon_{\mathbf{n}}$,

$$\begin{aligned}
|\mathbf{c}'B_{\mathbf{n}3}| &\le CH_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} I_{\{|Y_{\mathbf{ni}}^*(\theta)|\le \Upsilon_{\mathbf{n}}\}} |K_{\mathbf{c}}(\hat{\mathbf{X}}_{h\mathbf{i}})| \\
&\le C\big(1 + o_{\mathrm{P}}(1)\big) H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} I_{\{|Y_{\mathbf{ni}}^*(\theta)|\le C\epsilon_{\mathbf{n}}\}} |K_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})|.
\end{aligned}$$

In view of (A.14), the $o_{\mathrm{P}}(1)$ quantity here is uniform in $\mathbf{i}$. Since $\mathrm{E}[I_{\{|Y_{\mathbf{ni}}^*(\theta)|\le C\epsilon_{\mathbf{n}}\}}|K_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})|]$ is $\mathrm{O}(\epsilon_{\mathbf{n}} h_{\mathbf{n}}^d)$ uniformly with respect to $\boldsymbol{\theta}$, it follows that

$$\begin{aligned}
H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \mathrm{E}\big[I_{\{|Y_{\mathbf{ni}}^*(\theta)|\le C\epsilon_{\mathbf{n}}\}} |K_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})|\big] &= H_{\mathbf{n}}^{-1}\hat{\mathbf{n}}\mathrm{O}(\epsilon_{\mathbf{n}} h_{\mathbf{n}}^d) \\
&= \mathrm{O}(H_{\mathbf{n}}\epsilon_{\mathbf{n}}) = \mathrm{O}\big((h_{\mathbf{n}}^d \ln \hat{\mathbf{n}}/g^2)^{1/2} + (\hat{\mathbf{n}} h_{\mathbf{n}}^d g^{2r})^{1/2}\big).
\end{aligned}$$

Therefore $|\mathbf{c}'B_{\mathbf{n}3}| = o_{\mathrm{P}}(1)$, and $B_{\mathbf{n}3} = o_{\mathrm{P}}(1)$ uniformly with respect to $\boldsymbol{\theta}$.

On the other hand, $K_{\mathbf{c}}(\mathbf{x})$ is continuously differentiable since $K(\mathbf{x})$ is, so that

$$\mathbf{c}'[\hat{\mathcal{X}}_{h\mathbf{i}}\hat{K}_{\mathbf{i}} - \mathcal{X}_{h\mathbf{i}}K_{\mathbf{i}}] = K_{\mathbf{c}}(\hat{\mathbf{X}}_{h\mathbf{i}}) - K_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}}) = \big(1 + o_{\mathrm{P}}(1)\big) h^{-1} (\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}}))'(\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{X}_{\mathbf{i}}),$$

where $\dot{K}_{\mathbf{c}}(\mathbf{x})$ denotes the gradient of $K_{\mathbf{c}}(\mathbf{x})$ with respect to $\mathbf{x}$; from (A.14), the $o_{\mathrm{P}}(1)$ quantity here is uniform with respect to $\mathbf{i}$. Then,

$$B_{\mathbf{n}4} = \big(1 + o_{\mathrm{P}}(1)\big) H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \psi_p(Y_{\mathbf{ni}}^*(\boldsymbol{\theta})) h^{-1} \dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})(\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{X}_{\mathbf{i}}).$$

Note that, by (3.2) and (3.1),

$$\hat{\mathbf{X}}_{\mathbf{i}} - \mathbf{X}_{\mathbf{i}} = \mu_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}) - \hat{\mu}_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}) = \mu_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}) - \sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2} \tilde{\mathbf{X}}_{\mathbf{j}} w(\mathbf{s}_{\mathbf{j}}, \mathbf{s}_{\mathbf{i}}) = \tilde{w}(\mathbf{s}_{\mathbf{i}}) - \sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2} \mathbf{X}_{\mathbf{j}} w(\mathbf{s}_{\mathbf{j}}, \mathbf{s}_{\mathbf{i}}),$$

where $\tilde{w}(\mathbf{s}_{\mathbf{i}}) = \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}(\mu_{\mathbf{X}}(\mathbf{s}_{\mathbf{i}}) - \mu_{\mathbf{X}}(\mathbf{s}_{\mathbf{j}}))w(\mathbf{s}_{\mathbf{j}}, \mathbf{s}_{\mathbf{i}}) = \mathrm{O}(g^r)$ uniformly, under Assumptions (C1)–(C3), with respect to $\mathbf{i}$. It easily follows that

$$B_{\mathbf{n}4} = \bigl(1 + o_{\mathrm{P}}(1)\bigr)\biggl[B_{\mathbf{n}41} - \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \mathbf{X}(\mathbf{s}_{\mathbf{j}}) B_{\mathbf{n}42\mathbf{j}}\biggr], \tag{A.16}$$

where

$$B_{\mathbf{n}41} := B_{\mathbf{n}41}(\boldsymbol{\theta}) := H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2} \psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta}))h^{-1}\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})\tilde{w}(\mathbf{s}_{\mathbf{i}}),$$

and

$$B_{\mathbf{n}42\mathbf{j}} := B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta}) := H_{\mathbf{n}}^{-1} \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2} \psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta}))h^{-1}\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})w(\mathbf{s}_{\mathbf{j}}, \mathbf{s}_{\mathbf{i}}).$$

Along the same lines as in the proof of Lemma A.4, it can be shown that, since $\tilde{w}(\mathbf{s}_{\mathbf{i}}) = \mathrm{O}(g^r)$ uniformly in $\mathbf{s}_{\mathbf{i}}$,

$$\begin{aligned}
&\mathrm{E}\Bigl[\bigl|B_{\mathbf{n}41}(\boldsymbol{\theta}) - B_{\mathbf{n}41}(\mathbf{0}) - \bigl(\mathrm{E}B_{\mathbf{n}41}(\boldsymbol{\theta}) - \mathrm{E}B_{\mathbf{n}41}(\mathbf{0})\bigr)\bigr|^2\Bigr] \\
&\quad = H_{\mathbf{n}}^{-2}\Biggl[\hat{\mathbf{n}}(h_{\mathbf{n}}^d/H_{\mathbf{n}})(g^r/h)^2 + \hat{\mathbf{n}}\prod_{k=1}^{2} c_{\mathbf{n}k}h_{\mathbf{n}}^{2d}(g^r/h)^2 + \hat{\mathbf{n}}(g^r/h)^2 \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t\varphi(t)\Biggr] \\
&\quad = H_{\mathbf{n}}^{-1}(g^r/h)^2 + h_{\mathbf{n}}^{(1-2/a)d}(g^r/h)^2 + (g^r/h)^2 c_{\mathbf{n}k}^a \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t\varphi(t) = \mathrm{o}(1).
\end{aligned}$$

Similarly, taking $c_{\mathbf{n}k} = [h_{\mathbf{n}}^{-d}]^{1/a} \to \infty$, and since $w(\mathbf{s}_{\mathbf{j}}, \mathbf{s}_{\mathbf{i}}) = \mathrm{O}((\hat{\mathbf{n}}g^2)^{-1})$ uniformly in $\mathbf{s}_{\mathbf{i}}$ and $\mathbf{s}_{\mathbf{j}}$,

$$\begin{aligned}
&\mathrm{E}\bigl|B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta}) - B_{\mathbf{n}42\mathbf{j}}(\mathbf{0}) - \bigl(\mathrm{E}B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta}) - \mathrm{E}B_{\mathbf{n}42\mathbf{j}}(\mathbf{0})\bigr)\bigr|^2 \\
&\quad = H_{\mathbf{n}}{}^{-1}\bigl(1/(\hat{\mathbf{n}}g^2)h\bigr)^2 + \prod_{k=1}^{2} c_{\mathbf{n}k}h_{\mathbf{n}}^{d}\bigl(1/(\hat{\mathbf{n}}g^2)h\bigr)^2 + h_{\mathbf{n}}^{-d}\bigl(1/(\hat{\mathbf{n}}g^2)h\bigr)^2 \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t\varphi(t) \\
&\quad = H_{\mathbf{n}}{}^{-1}\bigl(1/(\hat{\mathbf{n}}g^2)h\bigr)^2 + h_{\mathbf{n}}^{(1-2/a)d}\bigl(1/(\hat{\mathbf{n}}g^2)h\bigr)^2 + \bigl(1/(\hat{\mathbf{n}}g^2)h\bigr)^2 c_{\mathbf{n}k}^a \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t\varphi(t) = \mathrm{o}(1),
\end{aligned}$$

uniformly with respect to $\mathbf{j}$. Thus, applying the chaining argument as in the proof of Lemma A.4, it follows that

$$\sup_{\|\boldsymbol{\theta}\|\le M}\left\|B_{\mathbf{n}41}(\boldsymbol{\theta})-B_{\mathbf{n}41}(\mathbf{0})-\big(\mathrm{E}B_{\mathbf{n}41}(\boldsymbol{\theta})-\mathrm{E}B_{\mathbf{n}41}(\mathbf{0})\big)\right\|=o_{\mathrm{P}}(1), \tag{A.17}$$

$$\sup_{\|\boldsymbol{\theta}\|\le M}\left\|B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta})-B_{\mathbf{n}42\mathbf{j}}(\mathbf{0})-\big(\mathrm{E}B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta})-\mathrm{E}B_{\mathbf{n}42\mathbf{j}}(\mathbf{0})\big)\right\|=o_{\mathrm{P}}(1), \tag{A.18}$$

uniformly with respect to $\mathbf{j}$. Further, it is easily shown that

$$B_{\mathbf{n}41}(\mathbf{0})-\mathrm{E}B_{\mathbf{n}41}(\mathbf{0})=o_{\mathrm{P}}(1),\qquad B_{\mathbf{n}42\mathbf{j}}(\mathbf{0})-\mathrm{E}B_{\mathbf{n}42\mathbf{j}}(\mathbf{0})=o_{\mathrm{P}}(1) \tag{A.19}$$

uniformly with respect to $\mathbf{j}$, while (the last step follows from integration by parts)

$$\begin{aligned}
\mathrm{E}_{\mathbf{n}41}(\boldsymbol{\theta})&=H_{\mathbf{n}}^{-1}\hat{\mathbf{n}}E\psi_p(Y^*_{\mathbf{ni}}(\boldsymbol{\theta}))h^{-1}\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})\tilde{w}(\mathbf{s_i})\\
&=H_{\mathbf{n}}^{-1}\hat{\mathbf{n}}\mathrm{E}\big[F(q_p(\mathbf{X_i})|\mathbf{X_i})-F\big(q_p(\mathbf{x})+U_{\mathbf{ni}}(\boldsymbol{\theta})|\mathbf{X_i}\big)\big]h^{-1}\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})\tilde{w}(\mathbf{s_i})\\
&=H_{\mathbf{n}}^{-1}\hat{\mathbf{n}}\big(1+o(1)\big)f_{Y|\mathbf{X}}(q_p(x)|x)\\
&\quad\times\mathrm{E}\big[\tfrac{1}{2}(\mathbf{X}_i-\mathbf{x})'\ddot{q}_p(\mathbf{x})(\mathbf{X}_i-\mathbf{x})+\boldsymbol{\theta}'\mathcal{X}_{h\mathbf{i}}H_{\mathbf{n}}^{-1}\big]\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})(g^r/h)\\
&=\mathrm{O}\big((\hat{\mathbf{n}}h_{\mathbf{n}}^{d+2}g^4)^{1/2}+g^r/h\big)
\end{aligned} \tag{A.20}$$

uniformly with respect to $\boldsymbol{\theta}$. Similarly,

$$\begin{aligned}
\mathrm{E}B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta})&=H_{\mathbf{n}}^{-1}\hat{\mathbf{n}}\mathrm{E}\big[F(q_p(\mathbf{X_i})|\mathbf{X_i})-F\big(q_p(\mathbf{x})+U_{\mathbf{ni}}(\boldsymbol{\theta})|\mathbf{X_i}\big)\big]h^{-1}\dot{K}_{\mathbf{c}}(\mathbf{X}_{h\mathbf{i}})w(\mathbf{s_i},\mathbf{s_j})\\
&=\big(1+o(1)\big)w(\mathbf{s_i},\mathbf{s_j})[(\hat{\mathbf{n}}h_{\mathbf{n}}^{d+2})^{1/2}A(\mathbf{x})+h^{-1}\boldsymbol{\theta}'\mathbf{B}(\mathbf{x})]
\end{aligned} \tag{A.21}$$

uniformly with respect to $\|\boldsymbol{\theta}\|\le M$, $\mathbf{i}$ and $\mathbf{j}$, where

$$A(\mathbf{x}):=f_{Y|\mathbf{X}}(q_p(x)|x)f_{\mathbf{X}}(\mathbf{x})\,\mathrm{tr}\bigg(\ddot{q}_p(\mathbf{x})\int\mathbf{uu}'\dot{K}_{\mathbf{c}}(\mathbf{u})\,\mathrm{d}\mathbf{u}\bigg)$$

and $\mathbf{B}(\mathbf{x}):=f_{Y|\mathbf{X}}(q_p(x)|x)f_{\mathbf{X}}(\mathbf{x})\int(1,\mathbf{u}')'\dot{K}_{\mathbf{c}}(\mathbf{u})\,\mathrm{d}\mathbf{u}$. Then, (A.17)–(A.21) imply that

$$B_{\mathbf{n}41}(\boldsymbol{\theta})=\mathrm{O}\big((\hat{\mathbf{n}}h_{\mathbf{n}}^{d+2}g^4)^{1/2}+g^r/h\big) \tag{A.22}$$

uniformly with respect to $\boldsymbol{\theta}$, and that

$$B_{\mathbf{n}42\mathbf{j}}(\boldsymbol{\theta})=\big(1+o(1)\big)w(\mathbf{s_i},\mathbf{s_j})[(\hat{\mathbf{n}}h_{\mathbf{n}}^{d+2})^{1/2}A(\mathbf{x})+h^{-1}\boldsymbol{\theta}'\mathbf{B}(\mathbf{x})] \tag{A.23}$$

uniformly with respect to $\boldsymbol{\theta}$, $\mathbf{i}$ and $\mathbf{j}$. Since $\mathbf{X_i}$ is α-mixing and stationary, with $\mathrm{E}\mathbf{X_i}=\mathbf{0}$ and $w(\mathbf{s_i},\mathbf{s_j})=\mathrm{O}(1/\hat{\mathbf{n}}g^2)$ uniformly in $\mathbf{i}$ and $\mathbf{j}$, we easily show that

$$\begin{aligned}
\mathrm{E}\bigg\|\sum_{\mathbf{j}\in\mathcal{I}_\mathbf{n}}\mathbf{X_j}w(\mathbf{s_i},\mathbf{s_j})\bigg\|^2&=\mathrm{E}[\mathbf{X}_\mathbf{j}'\mathbf{X_j}]\frac{\sum_{\mathbf{j}\in\mathcal{I}_\mathbf{n}}W^2((\mathbf{s_i}-\mathbf{s_j})/g)}{(\sum_{\mathbf{j}\in\mathcal{I}_\mathbf{n}}W((\mathbf{s_i}-\mathbf{s_j})/g))^2}+\sum_{\mathbf{j}_1\ne\mathbf{j}_2}\mathrm{E}[\mathbf{X}_{\mathbf{j}_1}'\mathbf{X}_{\mathbf{j}_2}]w(\mathbf{s_i},\mathbf{s}_{\mathbf{j}_1})w(\mathbf{s_i},\mathbf{s}_{\mathbf{j}_2})\\
&=\mathrm{O}(1)\big((1/\hat{\mathbf{n}}g^2)+(1/\hat{\mathbf{n}}g^2)^2\big)=\mathrm{O}(1/\hat{\mathbf{n}}g^2)
\end{aligned}$$

uniformly with respect to $\mathbf{i}$. Hence, in view of (A.23), we obtain

$$\begin{aligned}\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{X}_{\mathbf{j}}B_{\mathbf{n}42\mathbf{j}} &= \bigl(1+\mathrm{o}(1)\bigr)\bigl((\hat{\mathbf{n}}h_{\mathbf{n}}^{d+2})^{1/2}A(\mathbf{x})+h^{-1}\boldsymbol{\theta}'\mathbf{B}(\mathbf{x})\bigr)\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}}\mathbf{X}_{\mathbf{j}}w(\mathbf{s}_{\mathbf{i}},\mathbf{s}_{\mathbf{j}})\\ &= \bigl(1+\mathrm{o}(1)\bigr)\bigl((\hat{\mathbf{n}}h_{\mathbf{n}}^{d+2})^{1/2}A(\mathbf{x})+h^{-1}\boldsymbol{\theta}'\mathbf{B}(\mathbf{x})\bigr)\mathrm{O}(1/\hat{\mathbf{n}}g^2) \qquad (\mathrm{A}.24)\\ &= \mathrm{O}_{\mathrm{P}}\bigl((h_{\mathbf{n}}^{d+2}/\hat{\mathbf{n}}g^4)^{1/2}+(1/\hat{\mathbf{n}}hg^2)\bigr)=\mathrm{o}_{\mathrm{P}}(1)\end{aligned}$$

uniformly with respect to $\boldsymbol{\theta}$, $\|\boldsymbol{\theta}\|\le M$; (A.16), (A.22) and (A.24) thus imply that $B_{\mathbf{n}4}$ is $\mathrm{o}_{\mathrm{P}}(1)$, uniformly over $\|\boldsymbol{\theta}\|\le M$ and, in view of (A.15), we have that $\hat{V}_{\mathbf{n}}(\boldsymbol{\theta})-V_{\mathbf{n}}(\boldsymbol{\theta})=B_{\mathbf{n}3}+B_{\mathbf{n}4}=\mathrm{o}_{\mathrm{P}}(1)$ uniformly over $|\boldsymbol{\theta}|\le M$, which completes the proof.

Acknowledgement

Zudi Lu's work was partially supported by Discovery Grant from the Australian Research Council.

References

[1] Altman, N. (2000). Krige, smooth, both or neither? *Aust. N. Z. J. Stat.* **42** 441–461. MR1802968
[2] Anselin, L. and Florax, R.J.G.M. (1995). *New Directions in Spatial Econometrics*. Berlin: Springer.
[3] Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–225. MR0373208
[4] Bhattacharya, P.K. and Gangopadhyay, A. (1990). Kernel and nearets neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415. MR1062716
[5] Bolthausen, E. (1982). On the central limit theorem for stationary random fields. *Ann. Probab.* **10** 1047–1050. MR0672305
[6] Bowman, A.W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations.* Oxford: Oxford Univ. Press.
[7] Cai, Z.W. (2002). Regression quantiles for time series. *Econometric Theory* **18** 169–192. MR1885356
[8] Carbon, M., Hallin, M. and Tran, L.T. (1996). Kernel density estimation for random fields: The L_1 theory. *J. Nonparamet. Stat.* **6** 157–170. MR1383049
[9] Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777. MR1105843
[10] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836. MR0556476
[11] Cressie, N.A.C. (1993). *Statistics for Spatial Data*. New York: Wiley. MR1239641
[12] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916
[13] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*. New York: Wiley. MR0780746
[14] Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statist. Sinica* **1** 93–125. MR1101317
[15] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004. MR1209561

[16] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman & Hall. MR1383587

[17] Fan, J., Hu, T.C. and Truong, Y.K. (1994), Robust nonparametric function estimation. *Scand. J. Statist.* **21** 433–446. MR1310087

[18] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer. MR1964455

[19] Gao, J., Lu, Z. and Tjøstheim, D. (2006). Estimation in semi-parametric spatial regression. *Ann. Statist.* **34** 1395–1435. MR2278362

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

[21] Guyon, X. (1995). *Random Fields on a Network.* New York: Springer Verlag. MR1344683

[22] Hallin, M., Lu, Z. and Tran, L.T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7** 657–668. MR1849373

[23] Hallin, M., Lu, Z. and Tran, L.T. (2004). Density estimation for spatial processes: The L_1 theory. *J. Multivariate Anal.* **88** 61–75. MR2021860

[24] Hallin, M., Lu, Z. and Tran, L.T. (2004). Local linear spatial regression. *Ann. Statist.* **32** 2469–2500. MR2153992

[25] Hansen, B. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748. MR2409261

[26] He, X. and Shao, Q.M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochatic designs. *Ann. Statist.* **24** 2608–2630. MR1425971

[27] Honda T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52** 459–470. MR1794246

[28] Koenker, R. (2005). *Quantile Regression*. Cambridge, U.K.: Cambridge Univ. Press. MR2268657

[29] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[30] Koenker, R. and Mizera, I. (2004). Penalized triograms: Total variation regularization for bivariate smoothing. *J. Roy. Statist. Soc. Ser. B* **66** 145–164. MR2035764

[31] Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857. MR0909992

[32] Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12** 793–813. MR1421404

[33] Koul, H.L. and Mukherjee, K. (1994). Regression quantiles and related processes under long-range dependence. *J. Multivariate Anal.* **51** 318–337. MR1321301

[34] Loader, C. (1999). *Local Regression and Likelihood*. New York: Springer. MR1704236

[35] Lu, Z. and Chen, X. (2002). Spatial nonparametric regression estimation: non-isotropic case. *Acta Math. Appl. Sin.* **18** 641–656. MR2012328

[36] Lu, Z. and Chen, X. (2004). Spatial kernel regression estimation: weak consistency. *Statist. Probab. Lett.* **68** 125–136. MR2066167

[37] Lu, Z., Hui, Y.V. and Zhao, Q. (1998). Local linear quantile regression under dependence: Bahadur representation and application. Discussion paper, Dept. Management Sciences, City Univ. Hong Kong.

[38] Lu, Z., Lundervold, A., Tjøstheim, D. and Yao, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli* **13** 447–472. MR2331259

[39] Nakhapetyan, B.S. (1980). The central limit theorem for random fields with mixing conditions. In *Advances in Probability* **6**, *Multicomponent Systems* (R.L. Dobrushin and Ya.G. Sinai, eds.) 531–548. New York: Dekker. MR0599547

[40] Neaderhouser, C.C. (1980). Convergence of blocks spins defined on random fields. *J. Stat. Phys.* **22** 673–684. MR0579097

[41] Portnoy, S.L. (1991). Asymptotic behavior of regression quantiles in non-stationary dependent cases. *J. Multivariate Anal.* **38** 100–113. MR1128939

[42] Possolo, A. (1991). *Spatial Statistics and Imaging*. Hayward: Institute of Mathematical Statistics. MR1195556

[43] Ripley, B. (1981). *Spatial Statistics*. New York: Wiley. MR0624436

[44] Ruppert, D. and Carroll, R.J. (1980). Trimmed least square estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838. MR0600964

[45] Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. MR1311979

[46] Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620. MR0443204

[47] Takahata, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. *Z. Wahrsch. Verw. Gebiete* **64** 445–456. MR0717753

[48] Tran, L.T. (1990). Kernel density estimation on random fields. *J. Multivariate Anal.* **34** 37–53. MR1062546

[49] Tran, L.T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *J. Multivariate Anal.* **44** 23–46. MR1208468

[50] Welsh, A.H. (1996). Robust estimation of smooth regression and spread functions and their derivatives. *Statist. Sinica* **6** 347–366. MR1399307

[51] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450

[52] Whittle, P. (1963). Stochastic process in several dimensions. *Bulletin of the International Statistical Institute* **40** 974–985. MR0173287

[53] Yu, K. and Jones, M.C. (1997). A comparison of local constant and local linear regression quantile estimation. *Comput. Statist. Data Anal.* **25** 159–166.

[54] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628

[55] Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scand. J. Statist.* **31** 333–346. MR2087829

[56] Zhang, H. and Zimmerman, D.L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195

Received June 2007 and revised September 2008