Ann. Inst. Statist. Math.
Vol. 56, No. 2, 225–250 (2004)

ASYMPTOTIC DISTRIBUTIONS OF M-ESTIMATORS IN A SPATIAL REGRESSION MODEL UNDER SOME FIXED AND STOCHASTIC SPATIAL SAMPLING DESIGNS*

S. N. LAHIRI[1] AND KANCHAN MUKHERJEE[2]

[1]*Department of Statistics, Iowa State University, 102 Snedecor Hall, Ames, IA 50011-1210, U.S.A.*
[2]*Department of Statistics and Applied Probability, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore*

(Received October 21, 2002; revised April 22, 2003)

Abstract. In this paper, we consider M-estimators of the regression parameter in a spatial multiple linear regression model. We establish consistency and asymptotic normality of the M-estimators when the data-sites are generated by a class of deterministic as well as a class of stochastic spatial sampling schemes. Under the deterministic sampling schemes, the data-sites are located on a regular grid but may have an *infill* component. On the other hand, under the stochastic sampling schemes, locations of the data-sites are given by the realizations of a collection of independent random vectors and thus, are irregularly spaced. It is shown that scaling constants of different orders are needed for asymptotic normality under different spatial sampling schemes considered here. Further, in the stochastic case, the asymptotic covariance matrix is shown to depend on the spatial sampling density associated with the stochastic design. Results are established for M-estimators corresponding to certain non-smooth score functions including Huber's ψ-function and the sign functions (corresponding to the sample quantiles).

Key words and phrases: Central limit theorem, infill sampling, increasing-domain asymptotics, long range dependence, random field, strong mixing, stochastic design, spatial sampling design.

1. Introduction

Consider the multiple linear regression model

$$Y(\boldsymbol{s}_i) = \boldsymbol{\beta}'\boldsymbol{w}_n(\boldsymbol{s}_i) + Z(\boldsymbol{s}_i); \tag{1.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown parameter, $\boldsymbol{w}_n(\cdot)$ is a known non-random $\mathbb{R}^p$-valued function on $\mathbb{R}^d$, $Z(\cdot)$ is a stationary strong mixing random field (r.f) with marginal cdf and pdf G_0 and g_0, respectively, and $\{Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \dots\}$ are observed at the sampling sites $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \dots\}$. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a nondecreasing, right continuous function satisfying

$$E[\psi\{Z(\boldsymbol{s})\}] = 0. \tag{1.2}$$

*Research of Lahiri is partially supported by NSF grant no. DMS-0072571. Research of Mukherjee is partially supported by the Academic Research Grant R-155-000-003-112 from the National University of Singapore.

Define $\hat{\boldsymbol{\beta}}_n$ as a solution (in $\boldsymbol{t} \in I\!R^p$) of the equation $\boldsymbol{M}_n(\boldsymbol{t}) = 0$, where

$$\boldsymbol{M}_n(\boldsymbol{t}) = \sum \boldsymbol{w}_n(\boldsymbol{s}_i)\psi\{Y(\boldsymbol{s}_i) - \boldsymbol{w}_n'(\boldsymbol{s}_i)\boldsymbol{t}\}, \tag{1.3}$$

and the sum is over the total number of sampling sites $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots\}$. (The sample size depends on the spatial sampling design generating the sampling sites $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots\}$ and will be denoted by different symbols in the sequel. As a result, the range of summation in (1.3) is not indicated explicitly). In this paper, we derive the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$ corresponding to both smooth (absolutely continuous) and nonsmooth (bounded) score function ψ under certain fixed and stochastic spatial sampling designs. In particular, in the smooth ψ case, the results of the paper yield asymptotic distributions of robust-estimators of $\boldsymbol{\beta}$, including the ones obtained through Huber's ψ_k-function

$$\psi_k(x) = \begin{cases} k & \text{if } \ x > k \\ x & \text{if } \ |x| \leq k \\ -k & \text{if } \ x < -k, \end{cases} \tag{1.4}$$

$k \in (0, \infty)$ under both random and nonrandom sampling designs. In the nonsmooth case, it covers the score functions that yield estimates of quantiles of $Z(\cdot)$.

Asymptotic properties of estimators based on spatial data can be carried out under more than one asymptotic framework, primarily due to the fact that points in space do not admit a natural ordering as does the (time-) points in one dimension. Indeed, there are two basic types of asymptotic frameworks that are commonly used in the context of a continuous parameter random field. When the minimum distance between the neighboring data-sites remain bounded away from zero as the sample size increases, one obtains the *pure increasing domain* asymptotic structure (cf. Cressie (1993)). In this case, the sampling region at the n-th stage R_n (say) necessarily becomes unbounded as $n \to \infty$. On the other end, if the sampling region R_n remains confined in a bounded subset of $I\!R^d$ and an increasing number of data-sites are selected from it, then one obtains what is known as the *infill* asymptotic structure. In this case, the minimum distance among the data-sites goes down to zero as $n \to \infty$. In most studies on large sample properties of estimators based on spatial data, the pure increasing domain asymptotic structure with a grid based sampling design is used. However, a combination of the two asymptotic structures is also relevant for some applications (cf. Lahiri *et al.* (1999)). This is called a mixed asymptotic structure which has both an increasing domain component and an infill component. The increasing domain component is specified by the structure of the sampling region R_n which becomes unbounded as $n \to \infty$. The infill component (defined in Section 2) allows one to fill in any given subregion of R_n with an increasing number of data-sites using a scaled down integer grid on $I\!R^d$. See Section 2 for more details.

For the nonrandom or fixed design case, we derive asymptotic distribution of the M-estimator $\hat{\boldsymbol{\beta}}_n$ of (1.3) under the mixed asymptotic structure which has both an increasing domain component and an infill component. For the sake of completeness, here we also obtain the limit distribution of $(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ under the pure increasing domain case. For both asymptotic structures, we denote the sample size by N_n (cf. Subsection 2.1). It is observed that for a nondegenerate limit distribution of $(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ under the fixed design with either of the two asymptotic structures, the right choice of the scaling constant is given by D_n (defined in Section 3) which is comparable to the square root of the volume of R_n. For the pure increasing domain asymptotics, this volume is of the order of the total number of sampling sites N_n and the rate of convergence of $\hat{\beta}_n$ to β is the usual rate for strongly mixing observations. However, for the mixed asymptotic structure, the order of the volume of R_n is strictly less than the common choice N_n, reflecting the fact that neighbouring observations are strongly dependent and hence the rate of convergence of the estimators is slower than the usual $N_n^{-1/2}$ rate.

For the stochastic design case, we denote the sample size by n and suppose that the sampling sites $s_1, \ldots, s_n$ are given by the realization of a set of independent random vectors taking values in R_n. This provides a more flexible framework for modelling irregularly spaced data-locations than the standard approach that uses a homogeneous Poisson process. As in the fixed design case, here also we assume that the sampling region R_n becomes unbounded as $n \to \infty$. As explained in Section 2 below, in the stochastic design case, the distinction between a pure increasing domain structure and a mixed asymptotic structure lies in the relative growth rates of the sample size n *vis-a-*vis the volume of the sampling region. Infill sampling occurs when the sample size n grows at a faster rate than the volume of R_n. For a nondegenerate limit distribution of $(\hat{\beta}_n - \beta)$ under the stochastic design with both pure increasing domain and the mixed increasing domain sampling paradigm, the scaling factor is $\check{D}_n$ (defined in Section 3) which depends on both the volume inflating factor λ_n (defined in Section 2) and the density generating the stochastic design. For the pure increasing domain asymptotics, this volume is of the order of the total number of sampling sites n; however, under strict infilling, the order of volume is strictly less than n. Under appropriate regularity conditions, $(\hat{\beta}_n - \beta)$ is asymptotically p-variate normal under both asymptotic paradigms. The limiting normal distributions have zero mean but distinct convariance matrices that, among other things, depend on the spatial sampling density generating the points $s_1, \ldots, s_n$. Further, the results also show that the mixed increasing domain structure, with its partial infill component, leads to a "smaller" asymptotic covariance matrix than the pure increasing domain case under the stochastic design considered here. The results also allow us to determine the optimal spatial sampling design for estimating the parameter β. See Section 4 for more details.

The rest of the paper is organized as follows. In Section 2, we describe the spatial asymptotic framework and the spatial sampling designs. We state the main results of the paper under the nonrandom designs in Section 3 and those under the stochastic designs in Section 4. The proofs of the results are separately presented in Sections 5 and 6, respectively for the nonrandom and the stochastic design cases.

2. The spatial asymptotic framework

Let R_0 be a subset of $(-1/2, 1/2]^d$ containing the origin and let R_0^* be an open set satisfying $R_0^* \subset R_0 \subset \bar{R}_0^*$, where for any set $A \subset \mathbb{R}^d$, $\bar{A}$ denotes its closure. We regard R_0 as a 'prototype' of the sampling region R_n. Let $\{\lambda_n\}$ be a sequence of positive real numbers such that $\lambda_n \to \infty$ as $n \to \infty$. We assume that at the n-th stage, the sampling region R_n is obtained by 'inflating' the set R_0 by a scaling factor λ_n, i.e.

$$R_n = \lambda_n R_0. \tag{2.1}$$

Since the origin is assumed to lie in R_0, the shape of R_n remains the same for different values of n. Furthermore, this formulation allows the sampling region to have a wide range of shapes, encompassing common convex subsets of $\mathbb{R}^d$, such as spheres, ellipsoids, polyhedrons, as well as certain nonconvex regions (in $\mathbb{R}^d$), such as star-shaped sets. (Recall that a set $A \subset \mathbb{R}^d$ is called star-shaped if for any $x \in A$, the line segment joining x to the origin lies in A.) The latter class of sets may have fairly irregular boundaries. To avoid pathological cases, we shall always assume that for any sequence of real numbers $\{a_n\}$ with $a_n \to 0$ as $n \to \infty$, the number of cubes of the lattice $a_n\mathbb{Z}^d$ that intersect both R_0 and R_0^c is $O((a_n)^{-(d-1)})$ as $n \to \infty$. This assumption would ensure that the number of "boundary" sampling sites in the sampling region R_n is negligible compared to the total number of sampling sites. This condition holds for all regions R_n of practical interest. For example, if the boundary of R_0 is delineated by a smooth function $\tilde{g} : [0,1]^k \to \mathbb{R}^d$ for some $k < d$, then this condition holds. Sherman and Carlstein (1994) consider a simiar class of regions in the plane (i.e. $d = 2$), with boundaries given by closed rectifiable curves of finite lengths.

2.1 *Fixed designs*

Let $\delta_1, \ldots, \delta_d$ be a given set of d positive real numbers and let Δ be the $d \times d$ diagonal matrix with diagonal elements $\delta_1, \ldots, \delta_d$. The fixed designs we consider in this paper are based on the transformed integer lattice $\mathbb{Z}^d$, given by

$$\mathcal{Z}^d = \{\Delta \boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^d\}.$$

Thus, the lattice $\mathcal{Z}^d$ has an increment δ_i in the i-th direction, $1 \leq i \leq d$. For the *fixed sampling design* under the pure-increasing-domain structure, we assume that the random process $Z(\boldsymbol{s})$ is observed at N_n number of sampling sites $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_n}\}$ defined by

$$\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_n}\} = \{\boldsymbol{s} \in \mathcal{Z}^d : \boldsymbol{s} \in R_n\}. \tag{2.2}$$

In this case, the points $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_n}$ are separated by a distance $\delta_0 \equiv \min\{\delta_1, \ldots, \delta_d\}$ for all n and the sampling region R_n grows to $\mathbb{R}^d$ eventually. Note that the sample size N_n under this case satisfies the relation

$$N_n \sim \text{vol.}(\Delta^{-1}R_0)\lambda_n^d \tag{2.3}$$

where vol.(A) denotes the volume (i.e. the Lebesgue measure) of a set A in $\mathbb{R}^d$ and for any two sequences $\{r_n\}$ and $\{t_n\}$ of positive real numbers, we write $r_n \sim t_n$ if $r_n/t_n \to 1$ as $n \to \infty$.

For the fixed sampling design under the 'mixed' paradigm, any given subregion of the sampling region R_n is filled in with an increasing density of sampling sites. To that end, let $\{\eta_n\}$ be a sequence of positive real numbers such that $\eta_n \downarrow 0$ as $n \to \infty$. We assume that the sampling sites $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_n}\}$ under the *fixed sampling design* in the mixed-increasing-domain-structure case are given by the points on the scaled lattice $\eta_n\mathcal{Z}^d$ that lie in the sampling region R_n, i.e.

$$\begin{aligned} \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_n}\} &= \{\boldsymbol{s} \in \eta_n\mathcal{Z}^d : \boldsymbol{s} \in R_n\} \\ &= R_n \cap (\eta_n\mathcal{Z}^d). \end{aligned} \tag{2.4}$$

The scaled lattice $\eta_n\mathcal{Z}^d$ becomes finer for larger values of n and thus fills in any given region of $\mathbb{R}^d$ (and hence, of R_n) with an increasing density. Indeed, the maximum distance between any two adjacent sampling sites is bounded by $\max\{\delta_1, \ldots, \delta_d\}\eta_n$, which goes to zero as $n \to \infty$. The sample size N_n in this case satisfies the growth condition

$$N_n \sim \text{vol.}(\Delta^{-1} R_0)\lambda_n^d \eta_n^{-d}, \tag{2.5}$$

which is of a *larger* order of magnitude than the volume of R_n, given by vol.$(\Delta^{-1}R_0)\lambda_n^d$. This type of sampling designs may be useful for analysis of high resolution image data, e.g. remotely sensed satellite data on a fine scale.

2.2 *Stochastic designs*

For the stochastic design, let $f(x)$ be a continuous, everywhere positive probability density function on R_0, and let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be iid random vectors with probability density $f(x)$, that are independent of the r.f. $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. We assume that the sampling sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ lying in the sampling region R_n are obtained from a realization $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of the random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, by the relation

$$\boldsymbol{s}_i = \lambda_n \boldsymbol{x}_i, \qquad 1 \le i \le n.$$

Since $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ take values in R_0, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ are distributed over the entire sampling region $R_n \equiv \lambda_n R_0$. In analogy to nonparametric regression under stochastic designs, the results proved in this paper under stochastic designs are to be interpreted as valid with probability one under the joint probability distribution, say, $P_{\boldsymbol{X}}$ of the $\boldsymbol{X}_i$s (i.e. almost surely, $P_{\boldsymbol{X}}$).

In the stochastic design case, the distinction between the pure-increasing domain structure and the mixed-increasing domain structures is determined by the relative growth rates of the sample size n and the volume of the sampling region R_n. When $n \sim K\lambda_n^d$ for some $0 < K < \infty$, the sample size is of the same order as in the pure-increasing-domain-fixed design case (cf. (2.2)), and it should be regarded as the pure-increasing-domain analog under the stochastic design. On the other hand, if $n/\lambda_n^d \to \infty$ as $n \to \infty$ (as in (2.5)), it corresponds to the 'mixed-increasing domain' case under the stochastic design.

3. Results under fixed designs

3.1 *Notation and conditions*

We begin by introducing some notations. Denote the transpose of a matrix B by B'. For $\boldsymbol{x} = (x_1, \ldots, x_k)' \in \mathbb{R}^k$, $k \ge 1$, let $\|\boldsymbol{x}\|_1 \equiv |x_1| + \cdots + |x_k|$ and $\|\boldsymbol{x}\| = (x_1^2 + \cdots + x_k^2)^{1/2}$ respectively denote the ℓ^1 and ℓ^2 norms on $\mathbb{R}^k$. For any set $A \subset \mathbb{R}^k$, let vol.(A) denote the volume (i.e. the Lebesgue measure) of A and let $|A|$ denote the total number of elements in A. Let $\mathbb{1}(\cdot)$ denote the indicator function. For $y \in \mathbb{R}$, write $y_+ = \max\{y, 0\}$. Let $\mathcal{F}_Z(T) = \sigma\langle Z(\boldsymbol{s}) : \boldsymbol{s} \in T\rangle$ be the σ-field generated by the variables $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in T\}$, $T \subset \mathbb{R}^d$. For any two subsets T_1 and T_2 of $\mathbb{R}^d$, let

$$d(T_1, T_2) = \inf\{\|\boldsymbol{x} - \boldsymbol{s}\|_1 : \boldsymbol{x} \in T_1, \boldsymbol{s} \in T_2\}.$$

Also, let $\mathcal{R}(b) \equiv \{\cup_{i=1}^k D_i : \sum_{i=1}^k \text{vol.}(D_i) \le b, k \ge 1\}$ be the collection of all finite *disjoint unions* of *cubes* $D_1, \ldots, D_k$ in $\mathbb{R}^d$, $b > 0$. Then, the strong-mixing coefficient for the r.f. $Z(\cdot)$ is defined as

$$\alpha(a; b) = \sup\{\tilde{\alpha}(T_1, T_2) : d(T_1, T_2) \ge a, T_1, T_2 \in \mathcal{R}(b)\}, \tag{3.1}$$

$a > 0$, $b > 0$, where $\tilde{\alpha}(T_1, T_2) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_Z(T_1), B \in \mathcal{F}_Z(T_2)\}$, $T_1, T_2 \subset I\!R^d$. Note that the supremum in the definition of $\alpha(a; b)$ is taken over sets T_1, T_2 that are *bounded*. For $d > 1$, this is very important; some interesting results of Bradley (1989, 1993) (cf. Doukhan (1994)) show that a r.f. on $I\!R^d$, $d \geq 2$ that satisfies a mixing condition of the form

$$\lim_{a \to \infty} \sup\{\tilde{\alpha}(T_1, T_2) : T_1, T_2 \subset I\!R^d, d(T_1, T_2) \geq a\} = 0, \tag{3.2}$$

with the supremum taken over possibly *unbounded* sets, necessarily belongs to the more restricted class of ρ-mixing r.f.s. As a result, to ensure that the main results of the paper are applicable to a larger class of r.f.s, we do not allow unbounded sets in the definition of the strong mixing condition here.

For clarity of exposition, as in Lahiri (2003), we shall make the assumption that there exist a nonincreasing left continuous function $\alpha_1(\cdot)$ and a nondecreasing function $g_1(\cdot)$ such that $\lim_{a\to\infty} \alpha_1(a) = 0$ and $\lim_{b\to\infty} g_1(b) = \infty$, and the strong-mixing coefficient $\alpha(a, b)$ in (3.1) satisfies the inequality

$$\alpha(a, b) \leq \alpha_1(a) g_1(b) \qquad a > 0, \ b > 0. \tag{3.3}$$

Next, denote the autocovariance function of the transformed process $\psi(Z(\cdot))$ by $\sigma(\cdot)$, i.e., $\sigma(\boldsymbol{s}) = E\psi(Z(\boldsymbol{x}))\psi(Z(\boldsymbol{x} + \boldsymbol{s}))$ for all $\boldsymbol{s}, \boldsymbol{x} \in I\!R^d$. Let $\chi_0 = E\psi'(Z(\mathbf{0}))$ (whenever ψ' is defined) and $\chi_1 = \int g_0(x)\psi(dx)$, where recall that g_0 denotes the density function (with respect to the Lebesgue measure) of $Z(\mathbf{0})$. Let D_{1n} be a $p \times p$ matrix satisfying

$$D_{1n} D_{1n}' = \sum_{i=1}^{N_n} \boldsymbol{\omega}_n(\boldsymbol{s}_i) \boldsymbol{\omega}_n(\boldsymbol{s}_i)'.$$

Next define $D_n = \eta_n^{d/2} D_{1n}$, $C_n = \eta_n^{-d} D_n$ and $\boldsymbol{v}_i = D_n^{-1}\boldsymbol{\omega}_n(\boldsymbol{s}_i)$. Here, D_n would serve as the scaling matrix for $(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and C_n^{-1} as the scaling matrix for $\boldsymbol{M}_n(\boldsymbol{\beta})$. Also, let $m_n = \sup\{\|D_n^{-1}\boldsymbol{\omega}_n(\boldsymbol{s})\| : \boldsymbol{s} \in R_n\}$, $n \geq 1$. In addition to the assumption that ψ is monotone nondecreasing, right continuous with $E[\psi\{Z(\boldsymbol{s})\}] = 0$, we shall make the following assumptions for the 'smooth' and the 'nonsmooth' cases.

CONDITIONS.

(S.1) Let ψ be absolutely continuous with its almost everywhere derivative ψ' satisfying

(a) $\lim_{v\to 0} E|\psi'(Z(\mathbf{0}) - v) - \psi'(Z(\mathbf{0}))| = 0$.

(b) $E|\psi'(Z(\mathbf{0}))| < \infty$ and $\chi_0 = E\psi'(Z(\mathbf{0})) \neq 0$.

(S.2) For some $\delta \in (0, \infty)$ and $\tau > d(2 + \delta)/\delta$,

(a) $E|\psi(Z(\mathbf{0}))|^{2+\delta} < \infty$;

(b) $\alpha_1(y) \leq K y^{-\tau}$ for all $y \geq 1$, for some constant $K \in (0, \infty)$;

(c) $g_1(y) = o(y^{(\tau - d)/4d})$ as $y \to \infty$;

(d) $m_n^2 \equiv \sup\{\|D_n^{-1}\boldsymbol{w}_n(\boldsymbol{s})\|^2 : \boldsymbol{s} \in R_n\} = o(\lambda_n^{(\tau-d)/4d - d})$ as $n \to \infty$.

(S.3) (a) The $p \times p$ matrix $\sum_{i=1}^{N_n} \boldsymbol{w}_n(\boldsymbol{s}_i)\boldsymbol{w}_n(\boldsymbol{s}_i)'$ is nonsingular for all large n.

(b) There exists a $p \times p$ matrix valued function Q such that for any $\boldsymbol{h}_n \to \boldsymbol{h}$ in $I\!R^d$,

$$\lim_{n\to\infty} D_{1n}^{-1} \left[\sum_{i: \boldsymbol{s}_i, \boldsymbol{s}_i + \boldsymbol{h}_n \in R_n} \boldsymbol{w}_n(\boldsymbol{s}_i)\boldsymbol{w}_n(\boldsymbol{s}_i + \boldsymbol{h}_n)' \right] (D_{1n}^{-1})' = Q(\boldsymbol{h}).$$

In the nonsmooth case, we shall assume the following:

(N.1) ψ is bounded and the error distribution $G_0(\cdot)$ has a continuous density g_0 such that

$$\chi_1 = \int g_0(x)\psi(dx) \neq 0.$$

(N.2) For some $\tau > 3d(4d-1)/(4d-3)$ and $K \in (0,\infty)$,

$$\alpha_1(y) \leq K y^{-\tau} \quad \text{for all} \quad y \geq 1.$$

Some discussion about the conditions are in order. Note that the smoothness condition on ψ in (S.1) is stronger than assuming continuity but weaker than assuming that ψ is differentiabile everywhere on $\mathbb{R}$. This weaker condition allows us to apply the main results of the paper to M-estimators corresponding to Huber's ψ function. Part (a) of (S.1) holds if the almost everywhere derivative ψ' is bounded and continuous on a set $\mathcal{Z}$ for which $G_0\{\mathcal{Z}\} = 1$, where G_0 is the error distribution. In particular, this holds automatically for Huber's function ψ_k function for which $\psi_k'(x) = \mathbb{1}(-k \leq x \leq k)$. Part (b) of (S.1) is a necessary condition, as the reciprocal of the quantity $E\psi'(Z(\mathbf{0}))$ appears in the asymptotic covariance matrix of the M-estimator $\hat{\boldsymbol{\beta}}_n$. Condition (S.2) specifies a set of moment and mixing conditions. For deriving the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$, we make use of a Central Limit Theorem (CLT) result of Lahiri (2003), applied to a weighted sum of the variables $\psi(Z(\boldsymbol{s}_i))$, $i = 1,\ldots,n$. As a result, the moment condition is assumed on the transformed variables $\psi(Z(\boldsymbol{s}))$'s rather than on the $Z(\boldsymbol{s})$'s directly. Asymptotic normality of the M-estimator may hold even when $Z(\boldsymbol{s})$ does not have enough moments but $\psi(Z(\boldsymbol{s}))$ does. The mixing condition in part (b) is almost minimal as follows from the discussion in Lahiri (2003). Indeed, for $d = 1$, the requirement $\alpha_1(y) = O(y^{-\tau})$ as $y \to \infty$ for some $\tau > (2+\delta)/\delta$ is only slightly stronger than the more familiar condition

$$\sum_{n=1}^{\infty} \alpha_1(n)^{\delta/(2+\delta)} < \infty$$

used for proving the Central Limit Theorem for sums of stationary random variables in the time series case. Part (c) of (S.2) allows the function $g_1(\cdot)$ to be unbounded. As explained earlier, this is important for spatial processes in dimensions $d \geq 2$. And part (d) of (S.2) specifies the rate of decay for the scaled weight function $\boldsymbol{w}_n(\cdot)$. Note that when $\boldsymbol{w}_n(\cdot) \equiv \boldsymbol{w}(\cdot)$ for all $n \geq 1$ and the function $\|\boldsymbol{w}(\cdot)\|^2$ is Lebesgue integrable over $\mathbb{R}^d$ with

$$W \equiv \int \boldsymbol{w}(\boldsymbol{s})\boldsymbol{w}(\boldsymbol{s})'d\boldsymbol{s} \quad \text{nonsingular} \tag{3.4}$$

then it is easy to show that $\lambda_n^{-d} D_n D_n'$ converges to a scaler multiple of W, and hence, condition (S.2)(d) holds if $\sup\{\|\boldsymbol{w}(\boldsymbol{s})\| : \boldsymbol{s} \in R_n\} = o(\lambda_n^{(\tau-d)/4d})$ as $n \to \infty$. Condition (S.3) is a version of the well-known Grenader condition (cf. Grenander (1954)) for the spatial stochastic design case. We refer the reader to Grenander (1954), Anderson (1971) and Lahiri (2003) for examples and further discussion of Condition (S.3).

In the nonsmooth case, a different line of argument is needed to establish asymptotic normality of $\hat{\boldsymbol{\beta}}_n$. Here the score function ψ need not even be continuous and may have jump discontinuity. Note that Condition (N.1) is satisfied by the 'sign'-score function

$$\psi(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x > 0 \\ -1 & \text{if} \quad x < 0, \end{cases}$$

if $\chi_1 \neq 0$. Further, to establish the CLT in the nonsmooth case, we need a different condition on the rate of decay of the function $\alpha_1(\cdot)$, which is specified in Condition (N.2).

3.2 *Main results*

Now we are ready to state the main result of this section for the smooth score function case.

THEOREM 3.1. *Under Conditions* (S.1)–(S.3),

$$D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \chi_0^{-1} C_n^{-1} M_n(\boldsymbol{\beta}) + o_p(1),$$

and hence

$$D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to^d N\left(0, \left(\prod_{i=1}^d \delta_i\right)^{-1} \int \sigma(s)Q(s)ds/\chi_0^2\right). \tag{3.5}$$

Thus, under the mixed increasing domain asymptotic structure, $\hat{\boldsymbol{\beta}}_n$ is asymptotically normal for sampling regions of a wide variety of shapes and for rectangular grids with different spacings along different directions. When $\omega_n(\cdot)$, $n \geqq 1$ are given by a single Lebesgue integrable function $\omega(\cdot)$ satisfying the nonsingularity condition (S.3), $\|D_n\| = O(\lambda_n^{d/2})$ as $n \to \infty$, and thus, the M-estimator $\hat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}$ at the rate $O_p(\lambda_n^{-d/2})$ as $n \to \infty$.

Note that under the mixed increasing domain asymptotic structure, the infilling factor $\eta_n \to 0$ as $n \to \infty$ and therefore, by (3.5), the sample size N_n grows at a faster rate than the volume of the sampling region R_n, given by vol.$(R_0)\lambda_n^d$. As a consequence, the convergence rate $O_p(\lambda_n^{-d/2})$ of $\hat{\boldsymbol{\beta}}$ under the mixed asymptotic structure is much slower than the usual rate $O_p(N_n^{-1/2})$.

A similar conclusion holds for the case of nonsmooth score functions.

THEOREM 3.2. *Suppose that Conditions* (N.1), (N.2), (S.2) *and* (S.3) *hold. Then*

$$D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = -\chi_1^{-1} C_n^{-1} M_n(\boldsymbol{\beta}) + o_p(1),$$

and hence (3.5) *holds with* χ_0^2 *replaced by* χ_1^2.

For the sake of completeness, we also note down the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$ in the *pure* increasing domain case. In this case $\eta_n \equiv 1$ for all $n \geq 1$ and the sample size N_n grows at the *same* rate as the volume of the sampling region R_n. In this case, $\hat{\boldsymbol{\beta}}_n$ is a $N_n^{1/2}$-consistent estimator of $\boldsymbol{\beta}$, i.e. $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = O_p(N_n^{-1/2})$, as follows from the following result.

THEOREM 3.3. (a) *Suppose that conditions* (S.1)–(S.3) *hold with* $\eta_n \equiv 1$ *for all* $n \geq 1$. *Then,*

$$D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to^d N\left(\mathbf{0}, \chi_0^{-2} \sum_{\boldsymbol{i} \in \mathbb{Z}^d} \sigma(\Delta \boldsymbol{i}) Q(\Delta \boldsymbol{i})\right). \tag{3.6}$$

(b) *Suppose that* (N.1), (N.2), (S.2) *and* (S.3) *hold. Then* (3.6) *holds with* χ_0 *replaced by* χ_1.

In the time series context, Koul (1977) obtained the asymptotic behavior of M-estimators in a regression model with stationary strongly mixing errors (weak dependence) whereas Yajima (1991), Beran (1991) and Koul (1992) discussed least squares and M-estimation in linear models with long range dependent errors (strong dependence). Under weak dependence, M-estimators are asymptotically normal with normalizing constant of the order of square-root of the sample size which can be viewed as a spacial case of our results in Theorem 3.3 under the pure increasing domain case on $\mathbb{R}^1$. However, under strong dependence, the rate of convergence of the estimators may typically be slower than square-root of the sample size. This shows that the mixed asymptotic structure leads to strong dependence in the observations and our result gives asymptotic distribution of the M-estimator $\hat{\boldsymbol{\beta}}_n$ for a class of strongly dependent (time series as well as) spatial data. In the context of a spatial regression model with Gaussian errors, asymptotics of the maximum likelihood estimators of the regression and covariance parameters was discussed by Mardia and Marshall (1984) under the pure increasing domain framework. In contrast, here we are concerned with the estimation of the regression parameters only, but under a wide variety of asymptotic frameworks and for a large class of (possibly nonsmooth) score functions.

4. Results under stochastic designs

4.1 *Notation and conditions*

We introduce some notation and conditions, in addition to those of Section 3. Recall that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ are iid random vectors with common density f (with respect to the Lebesgue measure on $\mathbb{R}^d$). As in the nonstochastic case, define the $p \times p$ scaling matrices $\check{D}_{1n}, \check{D}_n, \check{C}_n$ by

$$\begin{aligned} \check{D}_{1n}\check{D}_{1n}' &= \int \boldsymbol{w}_n(\lambda_n \boldsymbol{x}) \boldsymbol{w}_n(\lambda_n \boldsymbol{s})' f(\boldsymbol{s}) d\boldsymbol{s}, \\ \check{D}_n &= \lambda_n^{d/2} \check{D}_{1n}, \qquad \text{and} \\ \check{C}_n &= (n\lambda_n^{-d}) \check{D}_n, \qquad n \geq 1. \end{aligned}$$

Also, let $\check{M}_{1n} = \sup\{\|\check{D}_{1n}^{-1} \boldsymbol{\omega}_n(\boldsymbol{s})\| : \boldsymbol{s} \in R_n\}$ and $\check{m}_n \equiv \lambda_n^{-d/2} \check{M}_{1n} = \sup\{\|\check{D}_n^{-1} \boldsymbol{\omega}_n(\boldsymbol{s})\| : \boldsymbol{s} \in R_n\}$.

CONDITIONS.

(C.1) There exists a $p \times p$ matrix valued function Q_1 on $\mathbb{R}^d$ such that for all $\boldsymbol{h} \in \mathbb{R}^d$,

$$\check{D}_{1n}^{-1} \left[\int \boldsymbol{\omega}_n(\lambda_n \boldsymbol{x} + \boldsymbol{h}) \boldsymbol{\omega}_n(\lambda_n \boldsymbol{x})' f(\boldsymbol{x})^2 d\boldsymbol{x}\right] (\check{D}_{1n}^{-1})' \to Q_1(\boldsymbol{h}) \qquad \text{as} \quad n \to \infty.$$

(C.2) There exist $\delta \in (0, \infty)$ and $\tau > d(2+\delta)/\delta$ such that

(a) $E|\psi(Z(\mathbf{0}))|^{2+\delta} < \infty$
(b) $\alpha_1(t) = O(t^{-\tau})$ as $t \to \infty$
(c) $g_1(t) = o(t^{(\tau-d)/4d})$ as $t \to \infty$.

(C.3) $\check{M}_{1n}^2 = o(\min\{(\log n)^{-2}\lambda_n^{(\tau-d)/4\tau}, n^a\})$ for some $a \in [0, 1/8)$, where τ is as in (C.2).

(C.4) The pdf $f(\cdot)$ of $\boldsymbol{X}_1$ is continuous and everywhere positive on $\bar{R}_0$.

Condition (C.1) is the analog of Condition (S.3), used in the nonrandom design case. Condition (C.2) is identical to conditional (S.2), except for the last part. Condition (C.3) specifies the growth rate of the "scaled" weight function $\boldsymbol{\omega}_n(\boldsymbol{s})$, as in (S.2)(d). Condition (C.4) says that every part of the sampling region R_n is "sampled" with positive probability under the given stochastic sampling scheme. This condition can be somewhat weakened at the expense of some additional regularity conditions on the weight function $\boldsymbol{\omega}_n(\cdot)$. See the discussion in Section 3 of Lahiri (2003) for more details. For $0 < c_0 \le \infty$, let $\chi_0 = E\psi'(Z(\mathbf{0}))$ and with $\mathbb{I}_p$ denoting the identity matrix of order p, let

$$\Sigma_{\infty,c_0} = \begin{cases} c_0^{-1}\sigma(\mathbf{0})\mathbb{I}_p + \int \sigma(\boldsymbol{x})Q_1(\boldsymbol{x})d\boldsymbol{x}, & \text{if } \ c_0 \in (0,\infty) \\ \int \sigma(\boldsymbol{x})Q_1(\boldsymbol{x})d\boldsymbol{x}, & \text{if } \ c_0 = \infty. \end{cases} \tag{4.1}$$

Then, we have the following result.

THEOREM 4.1. *Suppose that conditions* (C.1)–(C.4) *and* (S.1) *hold.*
(i) *If* $n\lambda_n^{-d} \to c_0 \in (0,\infty)$, *then*

$$\check{D}_n^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to^d N(0, \chi_0^{-2}\Sigma_{\infty,c_0}).$$

(ii) *If* $n\lambda_n^{-d} \to \infty$ *as* $n \to \infty$, *then*

$$\check{D}_n^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to^d N(0, \chi_0^{-2}\Sigma_{\infty,\infty}).$$

Thus, the M-estimators are asymptotically normal for both the pure increasing domain case ($c_0 < \infty$) and the mixed increasing domain case ($c_0 = \infty$). Note that the limiting covariance matrix differ in the two cases depending on whether $c_0 = \infty$ or not. Thus, if the sample size n grows at the rate $n \sim \text{vol.}(R_n)$, then the asymptotic covariance has an additional variance term $\text{vol.}(R_0)^{-1}\sigma(\mathbf{0})\mathbb{I}_p$ compared to the $c_0 = \infty$ case. Since this additional term is a nonnegative definite matrix, it follows that the $c_0 = \infty$ case allows for the most accurate estimation of $\boldsymbol{\beta}$ and its linear functions by $\hat{\boldsymbol{\beta}}_n$ and the corresponding linear combinations of $\hat{\boldsymbol{\beta}}_n$, respectively. Thus, the mixed asymptotic structure leads to a more accurate estimation of the parameter $\boldsymbol{\beta}$ compared to the pure increasing domain case. Intuitively this may be explained by noting that in the mixed case, one has many more observations ($n \gg \lambda_n^d$) than the pure increasing domain case where $n = O(\lambda_n^d)$. Although the additional data values do not improve the *rate* of convergence of $\hat{\boldsymbol{\beta}}_n$ to $\boldsymbol{\beta}$, it does reduce the asymptotic variability.

A second notable feature of the results is that the scaling matrix $\check{D}_n$ grows at a different rate than the square root of the sample size n under the mixed increasing domain asymptotics, due to the effect of infill sampling! Thus, under the present framework, the *volume* of the sampling region, not the *sample size* n serves as a common scaling constant that produces a nondegenerate limit distribution under both types of asymptotic structures.

The following result shows that a similar conclusion holds for the *nonsmooth* case.

THEOREM 4.2. *Suppose that Conditions* (C.1)–(C.4) *and* (N.1)–(N.2) *hold. Let* Σ_{∞,c_0} *be as in* (4.1) *and let* $\chi_1 = \int g_0(x)\psi(dx)$.

(i) *If* $n\lambda_n^{-d} \to c_0 \in (0,\infty)$, *then*

$$\check{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to^d N(\mathbf{0}, \chi_1^{-2}\Sigma_{\infty,c_0}).$$

(ii) *If* $n\lambda_n^{-d} \to \infty$ *as* $n \to \infty$, *then*

$$\check{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to^d N(\mathbf{0}, \chi_1^{-2}\Sigma_{\infty,\infty}).$$

5. Proofs for the fixed design case

Let $C = (0,1]^d$. For a σ-field $\mathcal{A}$ on a nonempty set Ω and a function $T : \Omega \to \mathbb{R}^k$ $(1 \le k < \infty)$, we write $T \in \mathcal{A}$ if T is $\langle \mathcal{A}, \mathcal{B}(\mathbb{R}^k)\rangle$-measureable, where $\mathcal{B}(\mathbb{R}^k)$ denotes the Borel σ-field on $\mathbb{R}^k$. For a function $h : A \to \mathbb{R}$, defined on a nonempty set A, let $\|h\|_\infty = \sup\{|h(a)| : a \in A\}$. In particular, $\|x\|_\infty = \max\{|x_i| : 1 \le i \le k\}$ for $x = (x_1, \ldots, x_k)' \in \mathbb{R}^k$, $1 \le k < \infty$. Let $A(t;a,b) = 1 + \sum_{k=1}^{[t]} k^{a-1}\alpha(k\delta_0; 3\delta^{*d})^{1/b}$, $a, b \ge 1$, $t \in \mathbb{R}^1$, where $[t]$ is the largest integer less than or equal to t and recall that $\delta_0 = \min\{\delta_1, \ldots, \delta_d\}$ and $\delta^* = \max\{\delta_1, \ldots, \delta_d\}$. Let $K, K_i, K(\cdot)$ denote generic constants in $(0,\infty)$ that may depend on their arguments (if any) but not on the variable n and on the probability element ω of the underlying probability space $(\Omega, \mathcal{F}, P)$. Also, unless otherwise specified, limits in order symbols will be taken by letting $n \to \infty$, where the qualification 'as $n \to \infty$' will be dropped for brevity.

LEMMA 5.1. *Let* $e_n(\boldsymbol{t}) \in \sigma\langle Z(\boldsymbol{t})\rangle$, $\boldsymbol{t} \in R_n$ *be a collection of centered Bernoulli variables with* $P[e_n(\boldsymbol{s}) = -\pi_n(\boldsymbol{s})] = 1 - P[e_n(\boldsymbol{s}) = 1 - \pi_n(\boldsymbol{s})] = 1 - \pi_n(\boldsymbol{s})$ *for some real number* $\pi_n(\boldsymbol{s}) \in [0, \pi_{0n}]$ *for all* $\boldsymbol{s} \in R_n$, *where* $0 < \pi_{0n} \le 1$. *Let* $w_{0n}(\boldsymbol{s}) : R_n \to \mathbb{R}$ *be a nonrandom real-valued weight function and let* $M_{0n} \equiv \sup\{|w_{0n}(\boldsymbol{s})| : \boldsymbol{s} \in R_n\} < \infty$, $n \ge 1$. *Also, let* $\{\eta_n\}_{n\ge1} \subset (0,1]$. *Then, for any real numbers* $r > 1$, $s > 2$ *with* $\frac{1}{r} + \frac{2}{s} = 1$,

$$E\left[\sum_{\boldsymbol{i}:\boldsymbol{i}\eta_n \in R_n} w_{0n}(\Delta\boldsymbol{i}\eta_n)e_n(\Delta\boldsymbol{i}\eta_n)\right]^4 \tag{5.1}$$
$$\le K(d,\Delta,r)\lambda_n^d\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}A(t_{0n};3d,r) + K(d,\Delta,r)\lambda_n^{2d}\eta_n^{-4d}M_{0n}^4\pi_{0n}^{4/s}A(t_{0n};d,r)^2$$

for all $n \ge 1$, *where* $t_{0n} = d\lambda_n/\delta_0$.

PROOF. We proceed as in the proof of Lemma 6.1 of Lahiri (2003). Note that $\Delta\boldsymbol{i}\eta_n \in R_n$ if and only if $\boldsymbol{i}\eta_n \in \lambda_n\Delta^{-1}R_0$. Set $\Delta^{-1}R_0 = R^1$ and $\lambda_n R^1 = R_n^1$. Note that the boundary of the set R^1 satisfy the same regularity property as that of R_0. Hence, instead of working with the data-sites $\Delta \boldsymbol{i}\eta_n \in R_n$, we work with the indices $\boldsymbol{i}\eta_n \in R_n^1$ which lie on the square grid $\mathbb{Z}^d$. Recall that $\mathcal{C}$ denotes the unit cube in $\mathbb{R}^d$. Define

$$\begin{aligned} I_n &\equiv \{\boldsymbol{i} \in \mathbb{Z}^d : [\boldsymbol{i} + \mathcal{C}] \cap R_n^1 \neq \phi\} \\ J_n(\boldsymbol{i}) &\equiv \{\boldsymbol{j} \in \mathbb{Z}^d : \eta_n \boldsymbol{j} \in [\boldsymbol{i} + \mathcal{C}] \cap R_n^1\}, \qquad \text{and} \\ Y(\boldsymbol{i}) &= \sum_{\boldsymbol{j} \in J_n(\boldsymbol{i})} w_{0n}(\Delta \boldsymbol{i}\eta_n) e_n(\Delta \boldsymbol{i}\eta_n), \qquad \boldsymbol{i} \in I_n. \end{aligned}$$

Then, $\{\boldsymbol{i} + \mathcal{C} : \boldsymbol{i} \in I_n\}$ is a covering of the set R_n^1 by unit cubes in $\mathbb{R}^d$ and $Y(\boldsymbol{i})$ denotes the summation of $w_{0n}(\boldsymbol{s})e_n(\boldsymbol{s})$ over all sampling sites $\boldsymbol{s}$ (lying on the grid $\mathcal{Z}^d$) that lie in the rectangle $\Delta[\boldsymbol{i} + \mathcal{C}] \cap R_n$. Hence, the left side of (5.1) is equal to $E(\sum_{\boldsymbol{i} \in I_n} Y(\boldsymbol{i}))^4$. Expanding the fourth power, we get

$$\begin{aligned} (5.2) \quad & E\left(\sum_{\boldsymbol{i} \in I_n} Y(\boldsymbol{i})\right)^4 \\ & \leq K(d) \Bigg[\sum_{\boldsymbol{i} \in I_n} E(Y(\boldsymbol{i}))^4 + \sum_{\boldsymbol{i} \neq \boldsymbol{j}} E(Y(\boldsymbol{i})^3 Y(\boldsymbol{j}) + \sum_{\boldsymbol{i} \neq \boldsymbol{j}} E(Y(\boldsymbol{i})^2 Y(\boldsymbol{j})^2 \\ & \qquad + \sum_{\boldsymbol{i} \neq \boldsymbol{j} \neq \boldsymbol{k}} EY(\boldsymbol{i})^2 Y(\boldsymbol{j}) Y(\boldsymbol{k}) + \sum_{\boldsymbol{i} \neq \boldsymbol{j} \neq \boldsymbol{k} \neq \boldsymbol{m}} EY(\boldsymbol{i}) Y(\boldsymbol{j}) Y(\boldsymbol{k}) Y(\boldsymbol{m})\Bigg] \\ & \equiv Q_{1n} + Q_{2n} + Q_{3n} + Q_{4n} + Q_{5n}, \qquad \text{say.} \end{aligned}$$

Note that for any $\boldsymbol{i}, \boldsymbol{j} \in I_n$, the maximum value of $|\boldsymbol{i} - \boldsymbol{j}|$ is bounded above by $\lambda_n(\delta_1^{-1} + \cdots + \delta_d^{-1}) \leq d\lambda_n/\delta_0 \equiv t_{0n}$, and for any $\boldsymbol{i}, \boldsymbol{j} \in I_n$ with $\boldsymbol{i} \neq \boldsymbol{j}$, the minimum distance (in the ℓ^1-norm) between the points of the rectangles $\Delta[\boldsymbol{i} + \mathcal{C}]$ and $\Delta[\boldsymbol{j} + \mathcal{C}]$ is $\delta_0(\|\boldsymbol{i} - \boldsymbol{j}\|_1 - d)_+$. Also, note that for any real number $a \geq 1$, $E|e_n(\boldsymbol{s})|^a = \pi_n(\boldsymbol{s})^a(1 - \pi_n(\boldsymbol{s})) + (1 - \pi_n(\boldsymbol{s}))^a \pi_n(\boldsymbol{s})$, so that by Jensen's inequality,

$$\begin{aligned} (5.3) \qquad E|Y(\boldsymbol{i})|^a &\leq |J_n(\boldsymbol{i})|^a E\left|\frac{1}{|J_n(\boldsymbol{i})|} \sum_{\boldsymbol{j} \in J_n(\boldsymbol{i})} w_{1n}(\Delta \eta_n \boldsymbol{j}) e_n(\Delta \eta_n \boldsymbol{j})\right|^a \\ &\leq |J_n(\boldsymbol{i})|^a M_{0n}^a \left\{\frac{1}{|J_n(\boldsymbol{i})|} \sum_{\boldsymbol{j} \in J_n(\boldsymbol{i})} E|e_n(\Delta \eta_n \boldsymbol{j})|^a\right\} \\ &\leq 2\pi_{0n} |J_n(\boldsymbol{i})|^a M_{0n}^a \end{aligned}$$

for all $\boldsymbol{i} \in I_n$. Hence, by (5.3) and the strong mixing property of $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$, we have

$$\begin{aligned} (5.4) \quad & Q_{1n} + Q_{2n} + Q_{3n} \\ & \leq K(d, r) \Bigg[2\pi_{0n} \sum_{\boldsymbol{i} \in I_n} |J_n(\boldsymbol{i})| M_{0n}^4 + \left(\sum_{\boldsymbol{i} \in I_n} EY(\boldsymbol{i})^2\right)^2 \\ & \qquad + \sum_{k=1}^{t_{0n}} |\{(\boldsymbol{i}, \boldsymbol{j}) \in I_n \times I_n : \|\boldsymbol{i} - \boldsymbol{j}\|_1 = k\}| \alpha([(k - d)_+]\delta_0; \delta^{*d})^{1/r} \end{aligned}$$

$$\left. \times \left\{(E|Y(\boldsymbol{i})|^{2s})^{2/s} + (E|Y(\boldsymbol{i})|^{3s})^{1/s}(E|Y(\boldsymbol{i})|^{s})^{1/s}\right\}\right]$$

$$\leq K(d,r)|I_n| \Bigg[\pi_{0n}\eta_n^{-4d}M_{0n}^4 + \pi_{0n}^2\eta_n^{-4d}M_{0n}^4 + \left\{\sum_{k=1}^{t_{0n}} k^{d-1}\alpha((k-d)_+\delta_0;\delta^{*d})^{1/r}\right\}\{\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}\}\Bigg]$$

$$\leq K(d,r,\delta_0)\lambda_n^d\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}A(t_{0n};d,r).$$

Now, using the combinatorial arguments in the proof of Lemma 4.1 of Lahiri (1999) and the arguments leading to (6.4) and (6.5) of Lahiri (2003), we get

$$\begin{aligned}
Q_{4n} \leq\ & K(d)\sum_{k=1}^{t_{0n}} |\{(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) \in I_n^3 : \boldsymbol{i} \neq \boldsymbol{j} \neq \boldsymbol{k}, d_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) = k\}| \\
& \times \alpha((k-d)_+\delta_0;3\delta^{*d})^{1/r} \\
& \times \{(E|Y(\boldsymbol{i})^2Y(\boldsymbol{j})|^s)^{1/s}(E(Y(\boldsymbol{k})|^s)^{1/s} \\
& \quad + (E|Y(\boldsymbol{i})^2Y(\boldsymbol{k})|^s)^{1/2}(E|Y(\boldsymbol{j})|^s)^{1/s}\} \\
\leq\ & K(d,p)|I_n| \left[\sum_{k=1}^{t_{0n}} k^{2d-1}\alpha((k-d)_+\delta_0;3\delta^{*d})^{1/r}\right]\{\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}\},
\end{aligned} \tag{5.5}$$

and writing $\sum^*$ and $\sum^{**}$ for summations over $I \subset I_n^4$ with $d_2(I) \geq d_3(I)$ and $d_2(I) < d_3(I)$ respectively, we get

$$\begin{aligned}
Q_{5n} =\ & \sum^{*} + \sum^{**}\{EY(\boldsymbol{i})Y(\boldsymbol{j})Y(\boldsymbol{k})Y(\boldsymbol{m})\} \\
\leq\ & K(d)\sum_{k=1}^{t_{0n}}\sum_{s=1}^{k} |\{I \subset I_n^4 : d_2(I) = k, d_3(I) = s\}| \\
& \times \alpha((k-d)_+\delta_0;3\delta^{*d})^{1/r}\{\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}\} \\
& + K(d)\Bigg[\left\{\sum_{\boldsymbol{i}\neq\boldsymbol{j}} |EY(\boldsymbol{i})Y(\boldsymbol{j})|\right\}^2 \\
& \qquad + \sum_{k=1}^{t_{0n}}\sum_{s=1}^{k-1} |\{I \subset I_n^4 : d_2(I) = s, d_3(I) = k\}| \\
& \qquad \times \alpha((k-d)_+\delta_0;3\delta^{*d})^{1/r}\{\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}\}\Bigg] \\
\leq\ & K(d)\eta_n^{-4d}M_{0n}^4\pi_{0n}^{2/s}\left[\sum_{k=1}^{t_{0n}}(2k)^{3d-1}|I_n|\alpha((k-d)_+\delta_0;3\delta^{*d})^{1/r}\right] \\
& + K(d)\left[|I_n|\sum_{k=1}^{t_{0n}} k^{d-1}\alpha((k-d)_+\delta_0;\delta^{*d})^{1/r}\eta_n^{-2d}M_{0n}^2\pi_{0n}^{2/s}\right]^2.
\end{aligned} \tag{5.6}$$

Hence, the lemma follows from (5.2), (5.4), (5.5) and (5.6).

LEMMA 5.2. *Let $g : I\!R \to I\!R$ be a Borel measurable function satisfying $E|g(Z(\mathbf{0}))| < \infty$ and $Eg(Z(\mathbf{0})) = 0$. Let $a_{1n}, \ldots, a_{N_n n} \in I\!R$ be constants satisfying $\sum_{i=1}^{N_n} |a_{in}| = O(1)$ and $\eta_n^{-d} \sum_{i=1}^{N_n} a_{in}^2 = o(1)$ as $n \to \infty$. If $\int t^{d-1}\alpha_1(t)dt < \infty$, then $\sum_{i=1}^{N_n} a_{in} g(Z(\boldsymbol{s}_i)) \to 0$ in probability.*

PROOF. For $1 \le i \le N_n$, define

$$W_i = g(Z(\boldsymbol{s}_i)), \quad W_{1i} = W_i \mathbb{1}(|W_i| \le t_{1n}) \quad \text{and} \quad W_{2i} = W_i - W_{1i},$$

where $t_{1n} \to \infty$ as $n \to \infty$ at a suitable rate, to be specified later. Since $E|\sum_{i=1}^{N_n} a_{in} W_{2i}| \le \sum_{i=1}^{N_n} |a_{in}| E|W_1| \mathbb{1}(|W_1| > t_{1n}) = o(1)$, it remains to show that $\sum_{i=1}^{N_n} a_{in} W_{1i} = o_p(1)$. Using $E(W_{1i}) = -E(W_{2i})$,

$$\left| E\left[\sum_{i=1}^{N_n} a_{in} W_{1i}\right]\right| \le \sum_{i=1}^{N_n} |a_{in}| E|W_1| \mathbb{1}(|W_1| > t_{1n}) = o(1). \tag{5.7}$$

To calculate the variance, let $J_n = \{\boldsymbol{s}_i - \boldsymbol{s}_j : 1 \le i, j \le N_n\}$ and for $\boldsymbol{h} \in J_n$, let $J_n(\boldsymbol{h}) = \{(i,j) : 1 \le i,j \le N_n, \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$. Then,

$$\begin{aligned} \operatorname{Var}\left(\sum_{i=1}^{N_n} a_{in} W_{1i}\right) &= \sum_{\boldsymbol{h} \in J_n} \left[\sum_{(i,j) \in J_n(\boldsymbol{h})} a_{in} a_{jn} \operatorname{Cov}(W_{1i}, W_{1j})\right] \\ &\le \sum_{\boldsymbol{h} \in J_n} \left[\left(\sum_{i=1}^{N_n} a_{in}^2\right) 4\alpha(\|\boldsymbol{h}\|_1; 1)(2t_{1n})^2\right] \\ &\le K\left[\eta_n^{-d} \sum_{i=1}^{N_n} a_{in}^2\right] (t_{1n}^2) \left[\int_0^\infty y^{d-1}\alpha_1(y)dy \cdot g_1(1)\right] \\ &= o(1), \end{aligned} \tag{5.8}$$

if we set $t_{1n} = [\eta_n^{-d} \sum_{i=1}^{N_n} a_{in}^2]^{-1/4}$. Now, Lemma 5.2 follows from (5.7) and (5.8).

PROOF OF THEOREM 3.1. For clarity of exposition, we explicitly indicate and separate out the main steps in the proof of Theorem 3.1.

Step (I): *Uniform one-step asymptotic approximation for $\boldsymbol{M}_n(\boldsymbol{u})$.* First in (5.12) below we obtain a one-step Taylor-type expansion of the M-score $\boldsymbol{M}_n(\boldsymbol{u})$ uniformly over every compact intervals $\{\|u\| \le b\}$, $b > 0$. Towards that, for $\boldsymbol{u} \in I\!R^p$ let $I = I_n(\boldsymbol{u}) = \{i : 1 \le i \le N_n, \boldsymbol{v}_i'\boldsymbol{u} \ge 0\}$ and $J = J_n(\boldsymbol{u}) = \{1, \ldots, N_n\} \setminus I$. Then, using the smoothness of ψ, we have

$$\begin{aligned} &C_n^{-1}\{\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\boldsymbol{u}) - \boldsymbol{M}_n(\boldsymbol{\beta})\} \\ &= C_n^{-1} \sum_{i=1}^{N_n} \boldsymbol{w}_n(\boldsymbol{s}_i)[\psi(Z(\boldsymbol{s}_i) - \boldsymbol{v}_i'\boldsymbol{u}) - \psi(Z(\boldsymbol{s}_i))] \\ &= -C_n^{-1} \sum_{i \in I} \boldsymbol{w}_n(\boldsymbol{s}_i) \int_{Z(\boldsymbol{s}_i) - \boldsymbol{v}_i'\boldsymbol{u}}^{Z(\boldsymbol{s}_i)} \psi'(t)dt + C_n^{-1} \sum_{i \in J} \boldsymbol{w}_n(\boldsymbol{s}_i) \int_{Z(\boldsymbol{s}_i)}^{Z(\boldsymbol{s}_i) - \boldsymbol{v}_i'\boldsymbol{u}} \psi'(t)dt \end{aligned}$$

$$= -C_n^{-1}\sum_{i\in I} \boldsymbol{w}_n(\boldsymbol{s}_i)\int_{-\boldsymbol{v}_i'\boldsymbol{u}}^{0} \psi'(Z(\boldsymbol{s}_i)+t)dt$$
$$+ C_n^{-1}\sum_{i\in J} \boldsymbol{w}_n(\boldsymbol{s}_i)\int_{0}^{-\boldsymbol{v}_i'\boldsymbol{u}} \psi'(Z(\boldsymbol{s}_i)+t)dt.$$

Fix any $b \in (0,\infty)$ arbitrarily. Writing $t_i \equiv \sup\{|\boldsymbol{v}_i'\boldsymbol{u}| : \|\boldsymbol{u}\| \le b\}$, bounding the limits of the integral by $-t_i$ and t_i and then taking expectation, we get

$$\text{(5.9)} \quad E\sup_{\|\boldsymbol{u}\|\le b}\left\|C_n^{-1}\{\boldsymbol{M}_n(\boldsymbol{\beta}+D_n^{-1}\boldsymbol{u})-\boldsymbol{M}_n(\boldsymbol{\beta})\}+C_n^{-1}\sum_{i=1}^{N_n}\boldsymbol{w}_n(\boldsymbol{s}_i)(\boldsymbol{v}_i'\boldsymbol{u})\psi'(Z(\boldsymbol{s}_i))\right\|$$
$$\le \sum_{i\in I}\|C_n^{-1}\boldsymbol{w}_n(\boldsymbol{s}_i)\|\int_{-t_i}^{0} E|\psi'(Z(\boldsymbol{0})+t)-\psi'(Z(\boldsymbol{0}))|dt$$
$$+\sum_{i\in J}\|C_n^{-1}\boldsymbol{w}_n(\boldsymbol{s}_i)\|\int_{0}^{t_i} E|\psi'(Z(\boldsymbol{0})+t)-\psi'(Z(\boldsymbol{0}))|dt$$
$$= o\left(\sum_{i=1}^{N_n}\|C_n^{-1}\boldsymbol{w}_n(\boldsymbol{s}_i)\|\|\boldsymbol{v}_i\|\right).$$

However,

$$\text{(5.10)} \quad \sum_{i=1}^{N_n}\|C_n^{-1}\boldsymbol{w}_n(\boldsymbol{s}_i)\boldsymbol{v}_i'\| \le \sum_{i=1}^{N_n}\|C_n^{-1}\boldsymbol{w}_n(\boldsymbol{s}_i)\|\|\boldsymbol{v}_i\|$$
$$= \eta_n^d\cdot\operatorname{tr}\left(\sum_{i=1}^{n}\boldsymbol{v}_i\boldsymbol{v}_i'\right) = \operatorname{tr}(\mathbb{I}_p) = p.$$

Hence right hand side of (5.9) is $o(1)$. Next we use Lemma 5.2 to show that

$$C_n^{-1}\sum_{i=1}^{n}\boldsymbol{w}_n(\boldsymbol{s}_i)\boldsymbol{v}_i'\{\psi'(Z(\boldsymbol{s}_i))-E\psi'(Z(\boldsymbol{0}))\} = o_p(1).$$

Note that

$$\text{(5.11)} \quad \sum_{i=1}^{n}\|C_n^{-1}\boldsymbol{w}_n(\boldsymbol{s}_i)\boldsymbol{v}_i'\|^2 \le \eta_n^{2d}\sum_{i=1}^{n}\|\boldsymbol{v}_i\|^4 \le p\eta_n^d m_n^2.$$

Hence, using (5.10) and (5.11) and applying Lemma 5.2 componentwise, we have

$$\sup_{\|\boldsymbol{u}\|\le b}\left\|C_n^{-1}\sum_{i=1}^{n}\boldsymbol{w}_n(\boldsymbol{s}_i)(\boldsymbol{v}_i'\boldsymbol{u})\{\psi'(Z(\boldsymbol{s}_i))-E\psi'(Z(\boldsymbol{0}))\}\right\|$$
$$\le b\left\|C_n^{-1}\sum_{i=1}^{n}\boldsymbol{w}_n(\boldsymbol{s}_i)\boldsymbol{v}_i'\{\psi'(Z(\boldsymbol{s}_i))-E\psi'(Z(\boldsymbol{0}))\}\right\| = o_p(1),$$

and consequently for any given $b \in (0,\infty)$,

$$\text{(5.12)} \quad \sup_{\|\boldsymbol{u}\|\le b}\|C_n^{-1}\{\boldsymbol{M}_n(\boldsymbol{\beta}+D_n^{-1}\boldsymbol{u})-\boldsymbol{M}_n(\boldsymbol{\beta})\}+E\psi'(Z(\boldsymbol{0}))\boldsymbol{u}\| = o_p(1).$$

Step (II): *D_n^{-1}-consistency of $\hat{\boldsymbol{\beta}}_n$.* Now we show the existence of a *tight* sequence of solutions of (1.3). Towards that, for any $M, \eta > 0$,

$$
\begin{aligned}
(5.13)\quad & P[\|D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\| \geq M] \\
& \leq P[\|D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\| \geq M, \|C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}))\| \leq \eta] \\
& \quad + P[\|C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}))\| \geq \eta] \\
& \leq P\left[\inf_{\|\boldsymbol{v}\| \geq M} \|C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\boldsymbol{v})\| \leq \eta\right] + P[\|C_n^{-1}\boldsymbol{M}_n(\hat{\boldsymbol{\beta}}_n)\| \geq \eta] \\
& = (I) + (II), \qquad \text{say.}
\end{aligned}
$$

Using the polar representation of $\boldsymbol{v} = r\theta$ and the Cauchy-Schwarz inequality we have,

$$
\begin{aligned}
& \inf_{\|\boldsymbol{v}\| \geq M} \|C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\boldsymbol{v})\| \\
& = \inf_{\|\theta\|=1, r \geq M} \|C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\theta r)\| \\
& \geq \inf_{\|\theta\|=1, r \geq M} -\theta' C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\theta r) \\
& = \inf_{\|\theta\|=1} -\theta' C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\theta M) \\
& = \inf_{\|\theta\|=1} \{-\theta'[C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\theta M) - C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta}) + E\{\psi'(Z(\mathbf{0}))\}\theta M] \\
& \qquad\qquad - \theta' C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta})\} \\
& \quad + E\{\psi'(Z(\mathbf{0}))\}M,
\end{aligned}
$$

where the second equality follows since using (1.3),

$$
-\theta' C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}r\theta) = \eta_n^d \sum(-\theta'\boldsymbol{v}_i)\psi\{Y(\boldsymbol{s}_i) - (\theta'\boldsymbol{v}_i)r\}
$$

is monotonically nondecreasing in r. The first term in the last equality is $o_p(1)$ by (5.12) and the second term is $O_p(1)$ by Theorem 4.2 of Lahiri (2003). Therefore (I) can be made very small by choosing sufficiently large M. Also, (II) can be made small by the definition of the M-estimator in (1.3) and thus $D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1)$.

Step (III): *Asymptotic distribution of $D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$.* Now substituting $u = D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ in (5.12) (which is possible because of uniform convergence on compact sets and $D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1)$), we obtain

$$
D_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \chi_0^{-1} C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta}) + o_p(1).
$$

Hence using the Cramer-Wold device and Theorem 4.2 and Proposition 4.2(ii) of Lahiri (2003) with $L(\cdot) \equiv K$ and $\alpha_1(y) = y^{-\tau}$, one can show that for any $\boldsymbol{a} \in \mathbb{R}^p$ with $\boldsymbol{a} \neq \mathbf{0}$,

$$
(5.14)\qquad \boldsymbol{a}' C_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta}) \rightarrow^d N\left(\mathbf{0}, \left[\prod_{i=1}^d \delta_i\right]^{-1} \int \sigma(\boldsymbol{s})[\boldsymbol{a}'Q(\boldsymbol{s})\boldsymbol{a}]d\boldsymbol{s}\right).
$$

Hence, Theorem 3.1 now follows from (5.14).

PROOF OF THEOREM 3.2. Note that Steps (II) and (III) in the proof of Theorem 3.1 are valid even for a nonsmooth ψ. Hence, it is enough to establish an analog of Step

(I), i.e., the uniform one-step asymptotic approximation for $\boldsymbol{M}_n(\boldsymbol{\beta})$. For $i = 1, \ldots, N_n$, let $\zeta_i(x; \boldsymbol{v}) = \mathbb{1}(Z(\boldsymbol{s}_i) \leq x + \boldsymbol{w}_n'(\boldsymbol{s}_i) D_n^{-1} \boldsymbol{v}) - \mathbb{1}(Z(\boldsymbol{s}_i) \leq x)$, $x \in \mathbb{R}$, $\boldsymbol{v} \in \mathbb{R}^p$. Then, it is easy to check that for any $\boldsymbol{u} \in \mathbb{R}^p$,

$$\eta_n^d \int \sum_{i=1}^{N_n} \boldsymbol{v}_i \zeta_i(x; \boldsymbol{v}) \psi(dx) = -C_n^{-1}\{\boldsymbol{M}_n(\boldsymbol{\beta} + D_n^{-1}\boldsymbol{v}) - \boldsymbol{M}_n(\boldsymbol{\beta})\}$$

$$\eta_n^d \int \sum_{i=1}^{N_n} \boldsymbol{v}_i E\zeta_i(x; \boldsymbol{v}) \psi(dx) = \boldsymbol{v} \int g_0(x)\psi(dx) + o(1). \tag{5.15}$$

Therefore, in analogy to (5.12), it is enough to show that for any $b \in \mathbb{N}$,

$$\begin{aligned} \Delta_n(b) \equiv & \sup_{\boldsymbol{v} \in [-b,b]^p} \left\| \eta_n^d \int \left[\sum_{i=1}^{N_n} \boldsymbol{v}_i \{\zeta_i(x; \boldsymbol{v}) - E\zeta_i(x; \boldsymbol{v})\} \right] \psi(dx) \right\| \\ = & \ o_p(1) \quad \text{as} \quad n \to \infty. \end{aligned} \tag{5.16}$$

To that end, we proceed in two steps. In the *first* step, we obtain a bound on $\Delta_n(b)$ on a discretized version of the supremum (leading to (5.19) below) and in the *second* step, we use Lemma 5.1 to show that each of the three terms in the resulting bound is negligible. Let $\{\epsilon_{1n}\} \subset (0, 1)$ be a sequence of real numbers such that $\epsilon_{1n} \to 0$ as $n \to \infty$ and $\epsilon_{1n}^{-1} \in \mathbb{N}$ for all $n \geq 1$. The exact rate at which $\{\epsilon_{1n}\}$ goes to zero will be specified later. Next, let $\mathcal{T}_n \equiv \{\boldsymbol{i}\epsilon_{1n} : \boldsymbol{i} \in \mathbb{Z}^p, \boldsymbol{i}\epsilon_{1n} \in [-b, b]^p\}$. Then $|\mathcal{T}_n| \equiv$ the size of $\mathcal{T}_n$ is equal to $\{(2b + 1)/\epsilon_{1n}\}^p$. Write any $\boldsymbol{v} \in [-b, b]^p$ as $\boldsymbol{v} = \boldsymbol{t} + \boldsymbol{u}$ where $\boldsymbol{t}$ is the nearest member of $\boldsymbol{v}$ which is in $\mathcal{T}_n$. The supremum in (5.16) will be taken in two stages; at the first stage, the maximum over the members $\{\boldsymbol{t}\}$ of $\mathcal{T}_n$ is taken and at the second stage the supremum over the members $\{\boldsymbol{u}\}$ in the balls of radius ϵ_{1n} with center at members of $\mathcal{T}_n$ is taken. Write $\epsilon_{2n} = \sup\{\|\boldsymbol{u}\| : \boldsymbol{u} \in [0, \epsilon_{1n}]^p\} = \sqrt{p}\epsilon_{1n}$. At the second stage, we will invoke the monotonicity of the indicator functions involving $\zeta_i(x; \boldsymbol{t} + \boldsymbol{u})$ and their expectations to replace the supremum over $\{\boldsymbol{u}\}$ in the balls of radius ϵ_{1n} by ϵ_{2n}. Then

$$\begin{aligned} \Delta_n(b) \leq & \max_{\boldsymbol{t} \in \mathcal{T}_n} \left\| \eta_n^d \int \left[\sum_{i=1}^{N_n} \boldsymbol{v}_i \{\zeta_i(x; \boldsymbol{t}) - E\zeta_i(x; \boldsymbol{t})\} \right] \psi(dx) \right\| \\ & + \max_{\boldsymbol{t} \in \mathcal{T}_n} \sup_{\boldsymbol{u} \in [0, \epsilon_{1n}]^p} A_{3n}(\boldsymbol{t}; \boldsymbol{u}) \eta_n^d, \end{aligned} \tag{5.17}$$

where

$$\begin{aligned} A_{3n}(\boldsymbol{t}; \boldsymbol{u}) = & \left\| \int \left[\sum_{i=1}^{N_n} \boldsymbol{v}_i \{\zeta_i(x; \boldsymbol{t} + \boldsymbol{u}) - E\zeta_i(x; \boldsymbol{t} + \boldsymbol{u})\} \right. \right. \\ & \left. \left. - \sum_{i=1}^{N_n} \boldsymbol{v}_i \{\zeta_i(x; \boldsymbol{t}) - E\zeta_i(x; \boldsymbol{t})\} \right] \psi(dx) \right\|. \end{aligned}$$

Note that for any $\boldsymbol{u}, \boldsymbol{t} \in \mathbb{R}^p$, with $I = I(\boldsymbol{u}) = \{1 \leq i \leq N_n : \boldsymbol{v}_i'\boldsymbol{u} \geq 0\}$ and $J = I^c$, we have

$$\left\| \int \left[\sum_{i=1}^{N_n} \boldsymbol{v}_i \{\zeta_i(x; \boldsymbol{u} + \boldsymbol{t}) - \zeta_i(x; \boldsymbol{t})\} \right] \psi(dx) \right\|$$

$$
\begin{aligned}
&= \left\| \int \left[\sum_{i \in I} \boldsymbol{v}_i \mathbb{1}(x + \boldsymbol{v}_i' \boldsymbol{t} < Z(\boldsymbol{s}_i) \le x + \boldsymbol{v}_i'(\boldsymbol{t} + \boldsymbol{u})) \right] \psi(dx) \right. \\
&\qquad \left. + \int \left[\sum_{i \in J} \boldsymbol{v}_i \mathbb{1}(x + \boldsymbol{v}_i'(\boldsymbol{t} + \boldsymbol{u}) < Z(\boldsymbol{s}_i) \le x + \boldsymbol{v}_i' \boldsymbol{t}) \right] \psi(dx) \right\| \\
&\le \sum_{i=1}^{N_n} \|\boldsymbol{v}_i\| \int \mathbb{1}(x + \boldsymbol{v}_i' \boldsymbol{t} - \|\boldsymbol{v}_i\| \|\boldsymbol{u}\| < Z(\boldsymbol{s}_i) \le x + \boldsymbol{v}_i' \boldsymbol{t} + \|\boldsymbol{v}_i\| \|\boldsymbol{u}\|) \psi(dx) \\
&\equiv A_{1n}(\boldsymbol{t}; \|\boldsymbol{u}\|), \qquad \text{say.}
\end{aligned}
$$

Similarly, for any $\boldsymbol{t}, \boldsymbol{u} \in \mathbb{R}^p$,

$$
\begin{aligned}
&\left\| \int \left[\sum_{i=1}^{N_n} \boldsymbol{v}_i \{ E\zeta_i(x; \boldsymbol{t} + \boldsymbol{u}) - E\zeta_i(x; \boldsymbol{t}) \} \right] \psi(dx) \right\| \\
&\qquad \le \sum_{i=1}^{N_n} \|\boldsymbol{v}_i\| \int P(x + \boldsymbol{v}_i' \boldsymbol{t} - \|\boldsymbol{v}_i\| \|\boldsymbol{u}\| < Z(\boldsymbol{0}) \le x + \boldsymbol{v}_i' \boldsymbol{t} + \|\boldsymbol{v}_i\| \|\boldsymbol{u}\|) \psi(dx) \\
&\qquad \equiv A_{2n}(\boldsymbol{t}; \|\boldsymbol{u}\|), \qquad \text{say.}
\end{aligned}
$$

Therefore, by the monotonicity of $A_{kn}(\boldsymbol{t}; \cdot)$, $k = 1, 2$ in the second argument,

$$
\begin{aligned}
\text{(5.18)} \qquad \sup_{\boldsymbol{u} \in [0, \epsilon_{1n}]^p} A_{3n}(\boldsymbol{t}, \boldsymbol{u}) &\le A_{1n}(\boldsymbol{t}; \epsilon_{2n}) + A_{2n}(\boldsymbol{t}; \epsilon_{2n}) \\
&\le |A_{1n}(\boldsymbol{t}; \epsilon_{2n}) - A_{2n}(\boldsymbol{t}; \epsilon_{2n})| + 2A_{2n}(\boldsymbol{t}; \epsilon_{2n})
\end{aligned}
$$

for all $\boldsymbol{t} \in \mathbb{R}^p$.

Combining (5.17) and (5.18) we have

$$
\begin{aligned}
\text{(5.19)} \qquad \Delta_n(b) &\le \max_{\boldsymbol{t} \in \mathcal{T}_n} \left\| \eta_n^d \int \left[\sum_{i=1}^{N_n} \boldsymbol{v}_i \{ \zeta_i(x; \boldsymbol{t}) - E\zeta_i(x; \boldsymbol{t}) \} \right] \psi(dx) \right\| \\
&\quad + \max_{\boldsymbol{t} \in \mathcal{T}_n} |A_{1n}(\boldsymbol{t}; \epsilon_{2n}) - A_{2n}(\boldsymbol{t}; \epsilon_{2n})| \eta_n^d \\
&\quad + 2 \max_{\boldsymbol{t} \in \mathcal{T}_n} A_{2n}(\boldsymbol{t}; \epsilon_{2n}) \eta_n^d \\
&\equiv I_{1n} + I_{2n} + I_{3n}, \qquad \text{say}
\end{aligned}
$$

Since G_0 has a bounded continuous density g_0, for any $\epsilon > 0$,

$$
\text{(5.20)} \qquad \sup_{a \in \mathbb{R}} |G_0(a + \epsilon) - G_0(a - \epsilon)| \le 2\epsilon \|g_0\|_\infty.
$$

Hence, by (5.11) and (5.20), from (5.19) we have

$$
\text{(5.21)} \qquad I_{3n} \le \eta_n^d \sum_{i=1}^{n} \|\boldsymbol{v}_i\| \int (2\|\boldsymbol{v}_i\| \epsilon_{2n}) \|g_0\|_\infty \psi(dx) \le K(p, \psi_0) \epsilon_{1n}
$$

where $\psi_0 \equiv \int \psi(dx)$.

Note that by applying Lemma 5.1 with $r = (4d-1)/(4d-3) > 1$, for any collection of Borel sets $\{B_{ix} : 1 \le i \le n, x \in \mathbb{R}\}$ and for any nonrandom weights $\tilde{w}_{1n}, \ldots, \tilde{w}_{nn} \in \mathbb{R}$, we get

$$
\begin{aligned}
(5.22)\qquad & E\left|\psi_0^{-1}\int\sum_{i=1}^{n}\tilde{w}_{in}\{\mathbb{1}(Z(s_i)\in B_{i,x}) - P(Z(\mathbf{0})\in B_{i,x})\}\psi(dx)\right|^4 \\
&\le \psi_0^{-1}\int E\left|\sum_{i=1}^{n}\tilde{w}_{in}\{\mathbb{1}(Z(s_i)\in B_{i,x}) - P(Z(\mathbf{0})\in B_{i,x})\}\right|^4\psi(dx) \\
&\le K_1\psi_0^{-1}\int[\alpha_{1n}\lambda_n^d\eta_n^{-4d}\tilde{w}_{0n}^4\pi_n^{2/s} + \alpha_{2n}^2\lambda_n^{2d}\eta_n^{-4d}\tilde{w}_{0n}^4\pi_n^{4/s}]\psi(dx) \\
&\le K_2\eta_n^{-4d}\tilde{w}_{0n}^4[\lambda_n^d\pi_n^{2/s} + \lambda_n^{2d}\pi_n^{4/s}],
\end{aligned}
$$

where $\tilde{w}_{0n} = \max\{|\tilde{w}_{in}| : 1 \le i \le n\}$, $\pi_n = \max\{P(Z(\mathbf{0}) \in B_{i,x}) : i = 1, \ldots, n, x \in \mathbb{R}\}$, $\alpha_{1n} = \sum_{i=1}^{t_{0n}} k^{3d-1}\alpha(k\delta_0; 3\delta^{*d})^{1/r}$ and $\alpha_{2n} = \sum_{i=1}^{t_{0n}} k^{d-1}\alpha(k\delta_0; \delta^{*d})^{1/r}$. Here, the factor s is defined by the equation $1/r + 2/s = 1$ and, by our choice of r and the condition on τ, $\alpha_{1n} + \alpha_{2n} = O(1)$ as $n \to \infty$.

Hence, by (5.19), (5.21) and (5.22), for any $\epsilon > 0$, there exists $n_\epsilon \in \mathbb{N}$ such that for all $n \ge n_\epsilon$,

$$
\begin{aligned}
A_{4n} &= P(\Delta_n(b) > 3\epsilon) \\
&\le P(I_{1n} > \epsilon) + P(I_{2n} > \epsilon) \\
&\le |\mathcal{T}_n|\cdot K_3\epsilon^{-4}m_n^4[\lambda_n^d\pi_{1n}^{2/s} + \lambda_n^{2d}\pi_{1n}^{4/s}] + |\mathcal{T}_n|\cdot K_4\epsilon^{-4}m_n^4[\lambda_n^d\pi_{2n}^{2/s} + \lambda_n^{2d}\pi_{2n}^{4/s}],
\end{aligned}
$$

where $\pi_{1n} = \sup\{|E\zeta_i(x;\boldsymbol{t})| : i = 1, \ldots, n, x \in \mathbb{R}, \boldsymbol{t} \in \mathcal{T}_n\}$, $\pi_{2n} = \sup\{P(x + \boldsymbol{v}_i'\boldsymbol{t} - \|\boldsymbol{v}_i\|\epsilon_{2n} < Z(\mathbf{0}) \le x + \boldsymbol{v}_i'\boldsymbol{t} + \|\boldsymbol{v}_i\|\epsilon_{2n}) : i = 1, \ldots, n, x \in \mathbb{R}, \boldsymbol{t} \in \mathcal{T}_n\}$. Note that by (5.20), $\pi_{1n} \le 2\max\{\|\boldsymbol{v}_i\|\cdot\|t\| : i = 1, \ldots, n, \boldsymbol{t} \in \mathcal{T}_n\}\|g_0\|_\infty \le 2\|g_0\|_\infty(b\sqrt{p})m_n$ and similarly, $\pi_{2n} \le 2\|g_0\|_\infty\epsilon_{2n}\cdot m_n$. Hence, by the condition on m_n,

$$
\begin{aligned}
A_{4n} &\le K_5\epsilon^{-4}\{(2b+1)/\epsilon_{1n}\}^p[(m_n^2\lambda_n^d)m_n^2m_n^{2/s} + (m_n^2\lambda_n^d)^2m_n^{4/s}] \\
&\le K_6\epsilon^{-4}\cdot\epsilon_{1n}^{-p}[o(\lambda_n^{(\tau-d)/2\tau}\{\lambda_n^{(\tau-d)/4\tau-d}\}^{2/s})] \\
&= o(1),
\end{aligned}
$$

if we choose $\epsilon_{1n} \sim [\lambda_n^{(\tau-d)/4\tau(2+2/s)-2d/s}]^{1/2p}$, say. Note that by our choice of r, the exponent $\frac{(\tau-d)}{4\tau}(2+\frac{2}{s}) - \frac{2d}{s} < \frac{1}{4}(3-\frac{1}{r}) - d(1-\frac{1}{r}) = 0$. This completes the proof of (5.16). The rest follows as in the proof of Theorem 3.1 with the analogous use of (5.13) and (5.14).

Note. A less stringent condition on τ obtains if we choose r and s to ensure $\frac{\tau-d}{4\tau}(2+\frac{2}{s}) - \frac{2d}{s} = 0$. It is easy to check that this leads to the requirement that $\tau > 3dr_1$ for $r_1 = [4\tau d - (\tau-d)]/[4\tau d - 3(\tau-d)]$, in place of $\tau > 3d(4d-1)/(4d-3)$. The new bound is better when τ is close to d, i.e., when δ is large.

Also, condition (N.1) can be slightly weakened by assuming that $t \to \int F(x+t)\psi(dx)$ is Lipschitz continuous.

PROOF OF THEOREM 3.3. Follows by straight forward modification of the steps in the proofs of Theorems 3.1 and 3.2. We omit the details.

6. Proofs for the stochastic design case

Suppose that $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in I\!R^d\}$ and $\{\boldsymbol{X}_n\}_{n\geq 1}$ are defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{X}$ denote the σ-field generated by $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$. Also, let $P_{\cdot|\boldsymbol{X}}$, $E_{\cdot|\boldsymbol{X}}$ and $\mathrm{Var}_{\cdot|\boldsymbol{X}}$ denote the conditional probability, expectation and variance, given $\mathcal{X}$. The (marginal) joint distribution of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ on $(I\!R^d)\infty$ is denoted by $P_{\boldsymbol{X}}$, and the corresponding expectation by $E_{\boldsymbol{X}}$. Let $\mathcal{G}$ be the σ-field generated by $\mathcal{X}$ and $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in I\!R^d\}$. For $\mathcal{G}$-measurable random vectors $\boldsymbol{T}_0$ and $\{\boldsymbol{T}_n\}_{n\geq 1}$, we shall say that

$$\boldsymbol{T}_n \to \boldsymbol{T}_0 \text{ in } P_{\cdot|\boldsymbol{X}} \text{ probability} \quad \text{a.s.} \quad (P_{\boldsymbol{X}})$$

if for every $\epsilon > 0$,

$$P_{\cdot|\boldsymbol{X}}(\|\boldsymbol{T}_n - \boldsymbol{T}_0\| > \epsilon) \to 0 \quad \text{as} \quad n \to \infty \quad \text{a.s.} \quad (P_{\boldsymbol{X}}).$$

Also, write $\eta_{1n} = \max\{\log n, n\lambda_n^{-d}\}^{-1/d}$, $n \geq 2$.

LEMMA 6.1. *Let A be a Borel subsets of R_0 and let $p \equiv \int_A f(\boldsymbol{x})d\boldsymbol{x}$.*

(i) *For every $0 < \epsilon < 1$, there exists a constant $C(\epsilon) > 1$, depending only on ϵ, such that for any $n \geq 1$ and $\beta > 0$ satisfying $np < C(\epsilon)\beta$,*

$$P_{\boldsymbol{X}}\left(\sum_{i=1}^{n} 1\!\!1(\boldsymbol{X}_i \in A) > np + \beta\right) \leq \exp(-(1-\epsilon)\beta \log[\beta/(np)]).$$

(ii) *For any $0 < \epsilon < 1$, $\beta > 0$, $n \geq 1$ with $\beta < np\log(1+\epsilon)$,*

$$P_{\boldsymbol{X}}\left(\sum_{i=1}^{n} 1\!\!1(\boldsymbol{X}_i \in A) > np + \beta\right) \leq \exp(-(1-\epsilon)\beta^2/[2np]).$$

PROOF OF LEMMA 6.1. This is Lemma 5.1 of Lahiri (2003).

LEMMA 6.2. *Let $g : R \to I\!R$ be a Borel measurable function satisfying $E|g(Z(\mathbf{0}))| < \infty$ and $Eg(Z(\mathbf{0})) = 0$. Also, let $a_{in} \equiv a_{in}(\boldsymbol{X})$, $i = 1, \ldots, n$ be $\mathcal{X}$-measurable random variables satisfying*

$$\sum_{i=1}^{n} |a_{in}(\boldsymbol{X})| = O(1) \quad a.s. \quad (P_{\boldsymbol{X}}) \tag{6.1}$$

and

$$\max\{\log n, n\lambda_n^{-d}\} \cdot \sum_{i=1}^{n} a_{in}^2(\boldsymbol{X}) \equiv o(1) \quad a.s. \quad (P_{\boldsymbol{X}}). \tag{6.2}$$

Then, $\sum_{i=1}^{n} a_{in}(\boldsymbol{X})g(Z(\boldsymbol{s}_i)) \to 0$ in $P_{\boldsymbol{X}}$-probability, a.s. $(P_{\boldsymbol{X}})$.

PROOF. Define the variables $W_i, W_{1i}, W_{2i}, 1 \leq i \leq n$ as in the proof of Lemma 5.2, where $t_{1n} \to \infty$ as $n \to \infty$ is to be specified. Then as in (5.7),

$$\begin{aligned} &\left|\sum_{i=1}^{n} a_{in} E_{\boldsymbol{X}}(W_{1i})\right| + E_{\boldsymbol{X}}\left|\sum_{i=1}^{n} a_{1n} W_{2i}\right| \\ &\quad \leq 2\left(\sum_{i=1}^{n} |a_{in}|\right) E|g(Z(\mathbf{0}))| 1\!\!1(|g(Z(\mathbf{0}))| > t_{1n}) = o(1), \qquad \text{a.s.} \quad P_{\boldsymbol{X}}. \end{aligned} \tag{6.3}$$

Next, let $J_{1n} = \{\boldsymbol{i} \in \mathbb{Z}^d : \boldsymbol{i} + \mathcal{C} \cap R_n \neq \emptyset\}$. Then, by Lemma 6.1, with $\beta_n = K_1 \max\{\log n, n\lambda_n^{-d}\}$ for some suitable $K_1 \in (0, \infty)$,

$$\sum_{n=1}^{\infty} P_{\boldsymbol{X}} \left(\sum_{j=1}^{n} \mathbb{1}(\lambda_n \boldsymbol{X}_j \in [i + \mathcal{C}]) > \beta_n \text{ for some } \boldsymbol{i} \in J_n \right)$$
$$\leq K_2 \sum_{n=1}^{\infty} \lambda_n^d \exp(-3 \log n) < \infty.$$

Hence, there exists a set A with $P(A) = 1$ such that for every $\omega \in A$, there is a $n_0 = n_0(\omega) \geq 1$ such that

$$\max_{i \in J_n} \sum_{j=1}^{n} \mathbb{1}(\lambda_n \boldsymbol{X}_j \in [i + \mathcal{C}]) \leq \beta_n \quad \text{for all} \quad n \geq n_0(\omega). \tag{6.4}$$

Next, for $\boldsymbol{j} \in J_{1n}$, let $Y_n(\boldsymbol{j})$ denote the sum of all $a_{in}(\boldsymbol{X}) W_{1i}$ such that $\boldsymbol{s}_i \in [\boldsymbol{j} + \mathcal{C}]$, i.e., $Y_n(\boldsymbol{j}) = \sum_{i=1}^{n} a_{in}(\boldsymbol{X}) W_{1i} \mathbb{1}(\boldsymbol{s}_i \in [\boldsymbol{j} + \mathcal{C}])$. Then for $\omega \in A$ and for all $n \geq n_0(\omega)$, $Y_n(\boldsymbol{j})$ is the sum of atmost β_n variables. W.l.g, suppose that $\sum_{i=1}^{n} |a_{in}(\boldsymbol{X})| = O(1)$ and $\beta_n \sum_{i=1}^{n} |a_{in}^2(\boldsymbol{X})| = o(1)$ on A. Hence, as in the proof of Lemma 5.2, for $\omega \in A$ and for $n \geq n_0(\omega)$,

$$\begin{aligned}
&\operatorname{Var}_{\cdot|\boldsymbol{X}} \left(\sum_{i=1}^{n} a_{in}(\boldsymbol{X}) W_{1i} \right) \\
&= \operatorname{Var}_{\cdot|\boldsymbol{X}} \left(\sum_{\boldsymbol{j} \in J_{1n}} Y_n(\boldsymbol{j}) \right) \\
&= \sum_{\boldsymbol{j} \in J_{1n}} \sum_{\boldsymbol{k} \in J_{1n}} \operatorname{Cov}_{\cdot|\boldsymbol{X}}(Y_n(\boldsymbol{j}), Y_n(\boldsymbol{k})) \\
&\leq \sum_{\boldsymbol{j} \in J_{1n}} \sum_{\boldsymbol{k} \in J_{1n}} 4\alpha(\|\boldsymbol{j} - \boldsymbol{k}\|_1; 1) \left\{ \sum_{i=1}^{n} |a_{in}(\boldsymbol{X})| \mathbb{1}(\boldsymbol{s}_i \in [\boldsymbol{j} + \mathcal{C}]) \right\} \\
&\quad \times \left\{ \sum_{i=1}^{n} |a_{1n}(\boldsymbol{X})| \mathbb{1}(\boldsymbol{s}_i \in [\boldsymbol{k} + \mathcal{C}]) \right\} (2t_{1n})^2 \\
&\leq 16 \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \alpha(\|\boldsymbol{h}\|_1; 1) \left(\sum_{\boldsymbol{j} \in J_{1n}} \left\{ \sum_{i=1}^{n} |a_{in}(\boldsymbol{X})| \mathbb{1}(\boldsymbol{s}_i \in (\boldsymbol{j} + \mathcal{C}]) \right\}^2 \right) t_{1n}^2 \\
&\leq K_3 \left[\int y^{d-1} \alpha_1(y) dy \cdot g_1(1) \right] \left[\sum_{i=1}^{n} a_{in}(\boldsymbol{X})^2 \right] t_{1n}^2 \beta_n,
\end{aligned}$$

by the Cauchy-Schwarz inequality. Hence, by (6.2), setting $t_{1n} = \{\beta_n [\sum_{i=1}^{n} a_{in}(\boldsymbol{X})^2]\}^{-1/4}$, the lemma is proved.

LEMMA 6.3. *For each $n \geq 1$, let $e_n(\boldsymbol{s}) \in \langle Z(\boldsymbol{s}) \rangle$, $\boldsymbol{s} \in R_n$ be a collection of centered Bernoulli variables with $P(e_n(\boldsymbol{s}) = -\pi_n(\boldsymbol{s})) = 1 - P(e_n(\boldsymbol{s}) = 1 - \pi_n(\boldsymbol{s})) = 1 - \pi_n(\boldsymbol{s})$, where $\pi_n(\boldsymbol{s}) \in [0, \pi_{0n}]$ for all $\boldsymbol{s} \in R_n$ for some $\pi_{0n} \in (0, 1]$. Let $\{\omega_{0n}(\boldsymbol{s}) : \boldsymbol{s} \in R_n\}$ be* *nonrandom weights with* $\sup_{s \in R_n} |\omega_{0n}(s)| = M_{0n} < \infty$ *for all* $n \geq 1$. *Then, there exists a random variable* $N \in \mathcal{X}$ *such that for almost all realizations of* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$,

$$
\begin{aligned}
(6.5)\qquad & E_{\cdot|\boldsymbol{X}} \left(\sum_{i=1}^{n} \omega_{0n}(\boldsymbol{s}_i) e_n(\boldsymbol{s}_i) \right)^4 \\
& \leq K(d,s) \lambda_n^d \eta_{1n}^{-4d} M_{0n}^4 \pi_{0n}^{2/s} \left[\sum_{k=0}^{d\lambda_n} (k+1)^{3d-1} \alpha(k;3)^{1/r} \right] \\
& \quad + K(d,s) \lambda_n^{2d} \eta_{1n}^{-4d} M_{0n}^4 \pi_{0n}^{4/s} \left[\sum_{k=0}^{d\lambda_n} (k+1)^{d-1} \alpha(k;1)^{1/r} \right]^2
\end{aligned}
$$

for all $n \geq N$, *where* $\boldsymbol{s}_i = \lambda_n \boldsymbol{X}_i$, $1 \leq i \leq n$, $\eta_{1n}^{-d} = \max\{\log n, n\lambda_n^{-d}\}$, *and where* $r \in (1, \infty)$ *and* $s \in (2, \infty)$ *are real numbers satisfying* $\frac{1}{r} + \frac{2}{s} = 1$.

PROOF. As in the proof of Lemma 5.1, we define the sets

$$\check{I}_n = \{\boldsymbol{i} \in \mathbb{Z}^d : \boldsymbol{i} + \mathcal{C} \cap R_n \neq \emptyset\},$$

and

$$\check{J}_n(\boldsymbol{i}) = \{j : 1 \leq j \leq n, \boldsymbol{s}_j \in (\boldsymbol{i} + \mathcal{C}) \cap R_n\}, \qquad \boldsymbol{i} \in \check{I}_n.$$

Also, let $\check{Y}(\boldsymbol{i}) = \sum_{j \in J_n(\boldsymbol{i})} \omega_{0n}(\boldsymbol{s}_j) e_n(\boldsymbol{s}_j)$ denote the sum over all $\omega_{0n}(\boldsymbol{s}_j) e_n(\boldsymbol{s}_j)$ such that the sampling site $\boldsymbol{s}_j \in (\boldsymbol{i} + \mathcal{C}) \cap R_n$, $\boldsymbol{i} \in \check{I}_n$. Then, it follows that the left side of (6.5) is equal to $E_{\cdot|\boldsymbol{X}} (\sum_{\boldsymbol{i} \in \check{I}_n} \check{Y}(\boldsymbol{i}))^4$. Note that $\check{Y}_n(\boldsymbol{i})$ is measurable with respect to $\sigma\langle \{Z(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{i} + \mathcal{C}\} \rangle$, $\boldsymbol{i} \in \check{I}_n$ and the minimum distance (in the ℓ^1-norm) between the points in $\boldsymbol{i} + \mathcal{C}$ and $\boldsymbol{j} + \mathcal{C}$, $\boldsymbol{i}, \boldsymbol{j} \in \check{I}_n \subset \mathbb{Z}^d$ is $(\|\boldsymbol{i} - \boldsymbol{j}\|_1 - d)_+$. Although the number of $e_n(s)$-variables entering in the sum $\check{Y}_n(\boldsymbol{i})$ may be different, by (6.4), it follows that there exists a random $\mathcal{X}$-measurable variable N such that for almost all realizations of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$,

$$(6.6)\qquad \max_{\boldsymbol{i} \in \check{I}_n} \sum_{j=1}^{n} \mathbb{1}(\lambda_n \boldsymbol{X}_j \in (\boldsymbol{i} + \mathcal{C}) \cap R_n) \leq \beta_n \quad \text{for all} \quad n \geq N,$$

where $\beta_n = K_1 \eta_{1n}^{-d}$ for some nonrandom $K_1 \in (0, \infty)$.

Hence, as in the derivation of (5.3), for almost all realizations of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$,

$$(6.7)\qquad E_{\cdot|\boldsymbol{X}} |\check{Y}(\boldsymbol{i})|^a \leq 2\pi_{0n} |\check{J}_n(\boldsymbol{i})|^a M_{1n}^a \leq k_2 \beta_n^{-a} M_{1n}^a \pi_{0n}$$

uniformly in $\boldsymbol{i} \in \check{I}_n$, whenever $n \geq N$. Now the proof of the lemma may be completed by retracing the arguments in the proof of Lemma 5.1. We omit the routine details.

PROOF OF THEOREM 4.1. The main steps in the proof of Theorem 4.1 is similar to those of Theorem 3.1, and hence we only point out the necessary modifications. Using the arguments in the proof of (5.9), for any given $b \in (0, \infty)$, we have

$$(6.8)\qquad E_{\cdot|\boldsymbol{X}} \left\{ \sup_{\|\boldsymbol{u}\| \leq b} \left\| \check{C}_n^{-1} \{ \boldsymbol{M}_n(\boldsymbol{\beta} + \check{D}_n^{-1} \boldsymbol{u}) - \boldsymbol{M}_n(\boldsymbol{\beta}) \} \right. \right.$$

$$- \check{C}_n^{-1} \sum_{i=1}^{n} \boldsymbol{\omega}_n(\boldsymbol{s}_i)(\check{\boldsymbol{v}}_i'\boldsymbol{u})\psi(Z(\boldsymbol{s}_i))\Bigg\|\Bigg\}$$

$$\leq \sum_{i=1}^{n} \|\check{C}_n^{-1}\boldsymbol{\omega}_n(\boldsymbol{s}_i)\| \int_{-t_{0i}}^{t_{0i}} E_{\cdot|\boldsymbol{X}}|\psi'(Z(\mathbf{0})+y)-\psi'(Z(\mathbf{0}))|dy$$

$$\leq K_1 b \sum_{i=1}^{n} \|\check{C}_n^{-1}\boldsymbol{\omega}_n(\boldsymbol{s}_i)\|\|\check{\boldsymbol{v}}_i\|\gamma_n,$$

where $t_{0i} = \sup\{|\check{\boldsymbol{v}}_i'\boldsymbol{u}| : \|\boldsymbol{u}\| \leq b\}$ and $\gamma_n = \sup\{E|\psi'(Z(\mathbf{0})+y)-\psi'(Z(\mathbf{0}))| : |y| \leq b\cdot\check{m}_{0n}\}$ and $\check{m}_{0n} = \max\{\|\check{\boldsymbol{v}}_i\| : 1 \leq i \leq n\}$.

Note that with probability 1,

$$\|\check{\boldsymbol{v}}_1\| \leq \lambda_n^{-d/2} \sup_{\boldsymbol{s}\in R_n} \|\check{D}_{1n}^{-1}\boldsymbol{\omega}_n(\boldsymbol{s})\| \equiv \lambda_n^{-d/2}\check{M}_{1n}. \tag{6.9}$$

Hence, by (C.3), there exists a (nonrandom) $n_0 \in I\!N$ such that for all $n \geq n_0$,

$$P_{\boldsymbol{X}}((n^{-1}\lambda_n^d)\|\check{\boldsymbol{v}}_1\|^2 \leq n^{-3/4}) = 1. \tag{6.10}$$

By (6.10) and Hoeffding's inequality (cf. Theorem 2, Hoeffding (1963)), for any $\epsilon > 0$,

$$\begin{aligned} P_{\boldsymbol{X}}\left(\left|n^{-1}\lambda_n^d \sum_{i=1}^{n}\{\|\check{\boldsymbol{v}}_i\|^2 - E_{\boldsymbol{X}}\|\check{\boldsymbol{v}}_i\|^2\}\right| > \epsilon\right) & \\ \leq 2\exp(-2\epsilon^2/[n(n^{-3/4})^2]) & \\ = 2\exp(-2\epsilon^2 n^{1/2}) \quad \text{for all} \quad n \geq n_0. & \end{aligned}$$

Hence, by the Borel-Cantelli Lemma,

$$n^{-1}\lambda_n^d \sum_{i=1}^{n}(\|\check{\boldsymbol{v}}_i\|^2 - E_{\boldsymbol{X}}\|\check{\boldsymbol{v}}_i\|^2) = o(1) \quad \text{a.s.}$$

Note that

$$\begin{aligned} n^{-1}\lambda_n^d \sum_{i=1}^{n} E_{\boldsymbol{X}}\|\check{\boldsymbol{v}}_i\|^2 &= \lambda_n^d \operatorname{tr}(E_{\boldsymbol{X}}\check{\boldsymbol{v}}_1\check{\boldsymbol{v}}_1') \\ &= \operatorname{tr}(\check{D}_{1n}^{-1}[E_{\boldsymbol{X}}\boldsymbol{\omega}_n(\lambda_n\boldsymbol{X}_1)\boldsymbol{\omega}_n(\lambda_n\boldsymbol{X}_1)'](\check{D}_{1n}^{-1})') \\ &= \operatorname{tr}(I\!I_p) = p. \end{aligned}$$

Hence, it follows that

$$n^{-1}\lambda_n^d \sum_{i=1}^{n} \|\check{\boldsymbol{v}}_i\|^2 \to p \quad \text{a.s.} \quad (P_{\boldsymbol{X}}). \tag{6.11}$$

Next, using (C.3) and (6.9)–(6.11), we have

$$(\log n \vee n\lambda_n^{-d}) \sum_{i=1}^{n} \|\check{\boldsymbol{v}}_i\|^4 (n^{-1}\lambda_n^d)^2 \tag{6.12}$$

$$\leq \begin{cases} O(\log n)\cdot n^{-3/4}\sum_{i=1}^{n}\|\check{\boldsymbol{v}}_i\|^2(n^{-1}\lambda_n^d) & \text{if } n\lambda_n^{-d}=O(\log n) \\ O(1)\cdot\lambda_n^{-d}\check{M}_{1n}^2\sum_{i=1}^{n}\|\check{\boldsymbol{v}}_i\|^2(n^{-1}\lambda_n^d) & \text{if } \log n=O(n\lambda_n^{-d}) \end{cases}$$
$$= o(1) \quad \text{a.s.} \quad (P_{\boldsymbol{X}}).$$

Hence using (6.11), (6.12) and Lemma 6.2 as in the proof of (5.11), one can show that for any $b\in(0,\infty)$,

$$\sup_{\|\boldsymbol{u}\|\leq b}\left\|\check{C}_n^{-1}\sum_{i=1}^{n}\boldsymbol{\omega}_n(\boldsymbol{s}_i)\check{\boldsymbol{v}}_i\boldsymbol{u}\{\psi'(Z(\boldsymbol{s}_i))-E\psi'(Z(\boldsymbol{s}))\}\right\| \to 0 \quad \text{in } P_{\cdot|\boldsymbol{X}}\text{-probability}, \quad \text{a.s.} \quad (P_{\boldsymbol{X}}). \tag{6.13}$$

Now using steps analogous to (5.12)–(5.14) as in the proof of Theorem 3.1 above and using Theorem 3.2 and Proposition 3.1 of Lahiri (2003), one can show that for almost all realizations of $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,$

$$\boldsymbol{a}'\check{C}_n^{-1}\boldsymbol{M}_n(\boldsymbol{\beta})\to^d N(\boldsymbol{0},\boldsymbol{a}'\Sigma_{\infty,c_0}\boldsymbol{a})$$

for every $\boldsymbol{a}\in\mathbb{R}^p$ with $\|\boldsymbol{a}\|=1$ and hence the result follows.

PROOF OF THEOREM 4.2. Like the proof of Theorem 4.1, here we outline the necessary modifications to the proof of Theorem 3.2 for the stochastic design case. For $i=1,\ldots,n$, $x\in\mathbb{R}$, $a\in(0,\infty)$ and $\boldsymbol{t},\boldsymbol{u}\in\mathbb{R}^p$, define

$$\check{\zeta}_i(x;\boldsymbol{u}) = \mathbb{1}(Z(\boldsymbol{s}_i)\leq x+\check{\boldsymbol{v}}_i'\boldsymbol{u})-\mathbb{1}(Z(\boldsymbol{s}_i)\leq x);$$
$$\check{A}_{1n}(\boldsymbol{t};a) = \sum_{i=1}^{n}\|\check{\boldsymbol{v}}_i\|\int\mathbb{1}(x+\check{\boldsymbol{v}}_i'\boldsymbol{t}-\|\check{\boldsymbol{v}}_i\|a<Z(\boldsymbol{s}_i)\leq x+\check{\boldsymbol{v}}_i'\boldsymbol{t}+\|\check{\boldsymbol{v}}_i\|a)\psi(dx);$$
$$\check{A}_{2n}(\boldsymbol{t};a) = E_{\cdot|\boldsymbol{X}}A_{1n}(\boldsymbol{t};a).$$

Then, as in (5.15), it is easy to check that

$$\check{C}_n^{-1}[\boldsymbol{M}_n(\boldsymbol{\beta}+\check{D}_n^{-1}\boldsymbol{u})-\boldsymbol{M}_n(\boldsymbol{\beta})] = -(n^{-1}\lambda_n^d)\sum_{i=1}^{n}\check{\boldsymbol{v}}_i\int\check{\zeta}_i(x;\boldsymbol{u})\psi(dx), \qquad \boldsymbol{u}\in\mathbb{R}^p.$$

Let $\{\epsilon_{1n}\}_{n\geq1}$ be a sequence of positive real numbers such that $\epsilon_{1n}\downarrow0$ as $n\to\infty$ and $\epsilon_{1n}^{-1}\in\mathbb{N}$. The exact rate at which ϵ_{1n} decays will be specified later. Next, set $\epsilon_{2n}=\sqrt{p}\epsilon_{1n}$ and define $\mathcal{T}_n$ as in the proof of Theorem 3.2. Then, repeating the arguments used in (5.16)–(5.19), one can show that for any given $b\in(0,\infty)$,

$$\check{\Delta}_n(b) \equiv \sup_{\boldsymbol{u}\in[-b,b]^p}\left\|\check{C}_n^{-1}\{\boldsymbol{M}_n(\boldsymbol{\beta}+\check{D}_n^{-1}\boldsymbol{u})-\boldsymbol{M}_n(\boldsymbol{\beta})\} - (n^{-1}\lambda_n^d)\int\left[\sum_{i=1}^{n}\boldsymbol{v}_iE_{\cdot|\boldsymbol{X}}\check{\zeta}_i(x;\boldsymbol{u})\right]\psi(dx)\right\| \tag{6.14}$$

$$
\begin{aligned}
&\le \max_{t\in\mathcal{T}_n}\left\| (n^{-1}\lambda_n^d)\int\left[\sum_{i=1}^n \check{v}_i\{\check{\zeta}_i(x;t)-E_{\cdot|X}\check{\zeta}_i(x;t)\}\right]\psi(dx)\right\| \\
&\quad + (n^{-1}\lambda_n^d)\max_{t\in\mathcal{T}_n}|\check{A}_{1n}(t;\epsilon_{2n})-\check{A}_{2n}(t;\epsilon_{2n})| \\
&\quad + 2(n^{-1}\lambda_n^d)\max_{t\in\mathcal{T}_n}\check{A}_{2n}(t;\epsilon_{2n}) \\
&\equiv \check{I}_{1n}+\check{I}_{2n}+\check{I}_{3n}, \qquad \text{say.}
\end{aligned}
$$

By (5.20) and (6.11), it follows that

$$
\begin{aligned}
\check{I}_{3n} &\le 2(n^{-1}\lambda_n^d)\sum_{i=1}^n\|\check{v}_i\|\int(2\|\check{v}_i\|\epsilon_{2n})\|g_0\|_\infty\psi(dx) \\
&= O(\epsilon_{1n}) \quad \text{a.s.} \quad (P_X).
\end{aligned} \tag{6.15}
$$

Next, for any collection of Borel sets $\{B_{i,x}:1\le i\le n, x\in\mathbb{R}\}$ and any nonrandom weight function $\omega_{0n}(\cdot):R_n\to\mathbb{R}$, by Lemma 6.3 and arguments in the proof of (5.22), there exist a constant $K_2\in(0,\infty)$ and a $\mathcal{X}$-measurable random variable N such that a.s. (P_X), for all $n\ge N$,

$$
\begin{aligned}
&E_{\cdot|X}\left(\psi_0^{-1}\int\left[\sum_{i=1}^n\omega_{0n}(s_i)\{\mathbb{1}(Z(s_i)\in B_{i,x})-P_{\cdot|X}(Z(\mathbf{0})\in B_{i,x})\}\right]\psi(dx)\right)^4 \\
&\le K_2\eta_{1n}^{-4d}M_{0n}^4[\lambda_n^d\pi_{0n}^{2/s}+\lambda_n^{2d}\pi_{0n}^{4/s}],
\end{aligned} \tag{6.16}
$$

where $\pi_{0n}=\sup\{P_{\cdot|X}(Z(\mathbf{0})\in B_{i,x}):1\le i\le n, x\in\mathbb{R}\}$ and s is determined by the relation $\frac{1}{r}+\frac{2}{s}=1$ with $r=(4d-1)/(4d-3)$.

Next note that by (5.20), $\sup\{P(Z(\mathbf{0})\in[x-\|\check{v}_i\|\|t\|, x+\|\check{v}_i\|\|t\|]):x\in\mathbb{R}, t\in\mathcal{T}_n, i=1,\dots,n\}\le K(b,p\|g_0\|_\infty)\max\{\|\check{v}_i\|:1\le i\le n\}\equiv\check{\pi}_{0n}$, say. Now setting $\check{M}_{1n}^2\equiv\lambda_n^{-d}M_{1n}^2$ and using (C.3), (6.16) and arguments leading to (5.23), one can show that a.s. (P_X),

$$
\begin{aligned}
P_{\cdot|X}(\check{\Delta}_n(b)>3\epsilon) &\le K_3|J_n|\epsilon^{-4}\max\{(n^{-1}\lambda_n^d)\|\check{v}_i\|:1\le i\le n\}^4[\lambda_n^d\check{\pi}_{0n}^{2/s}+\lambda_n^{2d}\check{\pi}_{0n}^{4/s}] \\
&\le \frac{K_4}{\epsilon^4}\epsilon_{1n}^{-p}(\log n)^4\check{m}_{1n}^4[\lambda_n^d\check{m}_{1n}^{2/s}+\lambda_n^{2d}\check{m}_{1n}^{4/s}] \\
&\le \frac{K_5}{\epsilon^4}\epsilon_{1n}^{-p}o(\lambda_n^{(\tau-d)/2\tau}[\lambda_n^{(\tau-d)/4\tau-d}]^{2/s}) \\
&= o(1),
\end{aligned}
$$

if we set $\epsilon_{1n}\sim(\lambda_n^{(\tau-d)/2\tau}[\lambda_n^{(\tau-d)/4\tau-d}]^{2/s})^{1/(2p)}$, say.

The rest follows as in the proof of Theorem 3.1 with the anologous use of (5.13) and (5.14).

References

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.

Beran, J. (1991). M-estimators of location for Gaussian and related processes with slowly decaying serial correlations, *Journal of the American Statistical Association*, **86**, 704–708.

Bradley, R. C. (1989). A caution on mixing conditions for random fields, *Statistics & Probability Letters*, **8**, 489–491.

Bradley, R. C. (1993). Equivalent mixing conditions for random fields, *The Annals of Probability*, **21**, 1921–1926.

Cressie, N. (1993). *Statistics for Spatial Data*, Wiley, New York.

Doukhan, P. (1994). *Mixing*, Lecture Notes in Statistics, No. 85. Springer, New York.

Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance, *Annals of Mathematical Statistics*, **25**, 252–272.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, **58**, 13–30.

Koul, H. L. (1977). Behavior of robust estimators in regression model with dependent errors, *The Annals of Statistics*, **5**, 681–699.

Koul, H. L. (1992). M-estimators in linear models with long range dependent errors, *Statistics & Probability Letters*, **14**, 153–164.

Lahiri, S. N. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method, *Probability Theory and Related Fields*, **114**, 55–84.

Lahiri, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs, *Sankhya Ser. A* (to appear).

Lahiri, S. N., Kaiser, M. S., Cressie, N. and Hsu, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling (with discussion), *Journal of the American Statistical Association*, **94**, 86–97.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **71**, 135–146.

Sherman, M. and Carlstein, E. (1994). Nonparametric estimation of the moments of a general statistic computed from spatial data, *Journal of American Statistical Association*, **89**, 496–500.

Yajima, Y. (1991). Asymptotic properties of the LSE in a linear regression model with long memory stationary errors, *The Annals of Statistics*, **19**, 158–177.