

Robinson, Peter M.

**Working Paper**

## Asymptotic theory for nonparametric regression with spatial data

cemmap working paper, No. CWP11/11

**Provided in Cooperation with:**

The Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Robinson, Peter M. (2011) : Asymptotic theory for nonparametric regression with spatial data, cemmap working paper, No. CWP11/11, Centre for Microdata Methods and Practice (cemmap), London,  
<https://doi.org/10.1920/wp.cem.2011.1111>

This Version is available at:

<https://hdl.handle.net/10419/64675>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Asymptotic theory for nonparametric regression with spatial data

---

**Peter M. Robinson**

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP11/11

# Asymptotic Theory for Nonparametric Regression with Spatial Data

P. M. Robinson \*  
London School of Economics

January 26, 2009

## Abstract

Nonparametric regression with spatial, or spatio-temporal, data is considered. The conditional mean of a dependent variable, given explanatory ones, is a nonparametric function, while the conditional covariance reflects spatial correlation. Conditional heteroscedasticity is also allowed, as well as non-identically distributed observations. Instead of mixing conditions, a (possibly non-stationary) linear process is assumed for disturbances, allowing for long range, as well as short-range, dependence, while decay in dependence in explanatory variables is described using a measure based on the departure of the joint density from the product of marginal densities. A basic triangular array setting is employed, with the aim of covering various patterns of spatial observation. Sufficient conditions are established for consistency and asymptotic normality of kernel regression estimates. When the cross-sectional dependence is sufficiently mild, the asymptotic variance in the central limit theorem is the same as when observations are independent; otherwise, the rate of convergence is slower. We discuss application of our conditions to spatial autoregressive models, and models defined on a regular lattice.

*JEL Classifications:* C13; C14; C21

*Keywords:* Nonparametric regression; Spatial data; Weak dependence; Long range dependence; Heterogeneity; Consistency; Central limit theorem.

---

\*Tel. +44-20-7955-7516; fax: +44-20-7955-6592. *E-mail address:* p.m.robinson@lse.ac.uk

# 1 Introduction

A distinctive challenge facing analysts of spatial econometric data is the possibility of spatial dependence. Typically, dependence is modelled as a function of spatial distance, whether the distance be geographic or economic, say, analogous to the modelling of dependence in time series data. However, unlike with time series, there is usually no natural ordering to spatial data. Moreover, forms of irregular spacing of data are more common with spatial than time series data, and this considerably complicates modelling and developing rules of statistical inference.

Often, as with cross-sectional and time series data, some (parametric or non-parametric) regression relation or conditional moment restriction is of interest in the modelling of spatial data. If the spatial dependence in the left-hand-side variable is entirely explained by the regressors, such that the disturbances are independent, matters are considerably simplified, and the development of rules of large sample statistical inference is, generally speaking, not very much harder than if the actual observations were independent. In parametric regression models, ordinary least squares can then typically deliver efficient inference (in an asymptotic Gauss-Markov sense, at least). Andrews (2005) has developed the theory to allow for arbitrarily strong forms of dependence in the disturbances, but with the data then generated by random sampling, an assumption that is not necessarily plausible in practice.

Substantial activity has taken place in the modelling of spatial dependence, and consequent statistical inference, and this is relevant to handling dependence in disturbances. In the statistical literature, lattice data have frequently been discussed. Here, there is equally-spaced sampling in each of  $d \geq 2$  dimensions, to extend the equally-spaced time series setting ( $d = 1$ ). Familiar time series models, such as autoregressive-moving-averages, have been extended to lattices (see e.g. Whittle, 1954). In parametric modelling there are greater problems of identifiability than in time series, and the "edge effect" complicates statistical inference (see Guyon, 1982, Dahlhaus and Künsch, 1987, Robinson and Vidal-Sanz, 2006, Yao and Brockwell, 2006). Nevertheless there is a strong sense in which results from time series can be extended.

Unfortunately economic data typically are not recorded on a lattice. If the observation locations are irregularly-spaced points in geographic space, it is possible to consider, say, Gaussian maximum likelihood estimation based on a parametric model for dependence defined continuously over the space, though a satisfactory asymptotic statistical theory has not yet been developed. However, even if we feel able to assign a (relative) value to the distance between each pair of data points, we may not have the information to plot the data in, say, 2-dimensional space. Partly as a result, "spatial autoregressive" (SAR) models of Cliff and Ord (1981) have become popular. Here,  $n$  spatial observations (or disturbances) are modelled as a linear transformation of  $n$  independent and identically distributed (iid) unobservable random variables, the  $n \times n$  transformation matrix being usually known apart from finitely many unknown parameters (often only a single such parameter). While we use the description

"autoregressive", forms of the model can be analogous to time series moving average, or autoregressive-moving-average, models, not just autoregressive ones, see (2.9) below. While a relatively *ad hoc* form of model, the flexibility of SAR has led to considerable applications (see e.g. Arbia, 2006).

SAR, and other structures, have been used to model disturbances, principally in parametric, in particular linear, regression models (see e.g. Kelejian and Prucha, 1999, Lee, 2002). On the other hand, nonparametric regression has become a standard tool of econometrics, at least in large cross-sectional data sets, due to a recognition that there can be little confidence that the functional form is linear, or of a specific nonlinear type. Estimates of the nonparametric regression function are typically obtained at several fixed points by some method of smoothing. In a spatial context, nonparametric regression has been discussed by, for example, Tran and Yakowitz (1993), Hallin, Lu and Tran (2004a). The most commonly used kind of smoothed nonparametric regression estimate in econometrics is still the Nadaraya-Watson kernel estimate. While originally motivated by iid observations, its asymptotic statistical behaviour has long been studied in the presence of stationary time series dependence. Under forms of weak dependence, it has been found that not only does the Nadaraya-Watson estimate retain its basic consistency property, but more surprisingly it has the same limit distribution as under independence (see, e.g. Roussas, 1969, Rosenblatt, 1971, Robinson, 1983). The latter finding is due to the "local" character of the estimate, and contrasts with experience with parametric regression models, where dependence in disturbances generally changes the limit distribution, and entails efficiency loss.

The present paper establishes consistency and asymptotic distribution theory for the Nadaraya-Watson estimate in a framework designed to apply to various kinds of spatial data. It would be possible to describe a theory that mimics fairly closely that for the time series case. In particular, strong mixing time series were assumed by Robinson (1983) in asymptotic theory for the Nadaraya-Watson estimate, and various mixing concepts have been generalised to  $d \geq 2$  dimensions in the random fields literature, where they have been employed in asymptotic theory for various parametric, nonparametric and semi-parametric estimates computed from spatial data; a less global condition, in a similar spirit, was employed by Pinkse, Shen and Slade (2007). We prefer to assume, in the case of the disturbances in our nonparametric regression, a linear (in independent random variables) structure, that explicitly covers both lattice linear autoregressive-moving-average and SAR models (with a scale factor permitting conditional or unconditional heteroscedasticity). Our framework also allows for a form of strong dependence (analogous to that found in long memory time series), a property ruled out by the mixing conditions usually assumed in asymptotic distribution theory. In this respect, it seems we also fill some gap in the time series literature because we allow our regressors to be stochastic, unlike in the fixed-design nonparametric regressions with long memory disturbances covered by Hall and Hart (1990), Robinson (1997). As a further, if secondary, innovation, while we have to assume some (mild) falling off of dependence in the regressors as their distance increases, we do not require these to be identically

distributed across observations (as in Andrews, 1995). It should be added that our asymptotic theory is of the “increasing domain” variety. “Infill asymptotics”, on a bounded domain, is popular in some research on spatial statistics and could be employed here, but is likely to yield less useful results. For example it has been found in some settings that under infill asymptotics estimates are inconsistent, by virtue of converging to a nondegenerate probability limit.

The following section describes our basic model and setting. Section 3 introduces the Nadaraya-Watson kernel estimate. Detailed regularity conditions are presented in Sections 4 and 5 for consistency and asymptotic distribution theory, respectively, the proofs resting heavily on a sequence of lemmas, which are stated and proved in appendices. Section 6 discusses implications of our conditions and of our results in particular spatial settings.

## 2 Nonparametric regression in a spatial setting

We consider the conditional expectation of a scalar observable  $Y$  given a  $d$ -dimensional vector observable  $X$ . We have  $n$  observations on  $(Y, X)$ . It is convenient to treat these as triangular arrays, that is, we observe the scalar  $Y_{in}$  and the  $d \times 1$  vector  $X_{in}$ , for  $1 \leq i \leq n$ , where statistical theory will be developed with  $n$  increasing without bound.

The triangular array structure of  $Y$  is partly a consequence of allowing a triangular array structure for the disturbances (the difference between  $Y$  and its conditional expectation) in the model, to cover in particular a common specification of the SAR model. But there is a more fundamental reason for it, and for treating the  $X$  observations as a triangular array also. We can identify each of the indices  $i = 1, \dots, n$  with a location in space. In regularly-observed time series settings, these indices correspond to equi-distant points on the real line, and it is evident what we usually mean by letting  $n$  increase. However there is ambiguity when these are points in space. For example, consider  $n$  points on a 2-dimensional regularly-spaced lattice, where both the number ( $n_1$ ) of rows and the number ( $n_2$ ) of columns increases with  $n = n_1 \cdot n_2$ . If we choose to list these points in lexicographic order (say first row left  $\rightarrow$  right, then second row etc.) then as  $n$  increases there would have to be some re-labelling, as the triangular array permits. Another consequence of this listing is that dependence between locations  $i$  and  $j$  is not always naturally expressed as a function of the difference  $i - j$ , even if the process is stationary (unlike in a time series). For example, this is so if the dependence is isotropic. Of course in this lattice case we can naturally label the locations by a bivariate index, and model dependence relative to this. However, there is still ambiguity in how  $n_1$  and/or  $n_2$  increase as  $n$  increases, and in any case we do not wish to restrict to 2-dimensional lattice data; we could have a higher-dimensional lattice (as with spatio-temporal data, for example) or irregularly-spaced data, or else data modelled using a SAR model, in which only some measures of distance between each pair of observations are employed. As a result our conditions tend to be of a "global" nature, in the sense that all  $n$  locations are involved, with  $n$  increasing, and thus are also relatively unprim-

itive, sometimes requiring a good deal of work to check in individual cases, but this seems inevitable in order to potentially cover many kinds of spatial data.

As a consequence of the triangular array structure many quantities in the paper deserve an  $n$  subscript. To avoid burdening the reader with excessive notational detail we will however tend to suppress the  $n$  subscript, while reminding the reader from time to time of the underlying  $n$ -dependence. Thus, for example, we write  $X_i$  for  $X_{in}$  and  $Y_i$  for  $Y_{in}$ .

We consider a basic conditional moment restriction of the form

$$E(Y_i | X_i) = g(X_i), \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2.1)$$

where  $g(x) : \mathbb{R}^d \rightarrow \mathbb{R}$  is a smooth, nonparametric function. We wish to estimate  $g(x)$  at fixed points  $x$ . Note that  $g$  is constant over  $i$  (and  $n$ ). However (anticipating Nadaraya-Watson estimation, which entails density cancellation asymptotically) we will assume that the  $X_i$  have probability densities,  $f_i(x) = f_{in}(x)$ , that are unusually allowed to vary across  $i$ , though unsurprisingly, given the need to obtain a useful asymptotic theory, they do have to satisfy some homogeneity restrictions, and the familiar identically-distributed case affords simplification. The  $X_i$  are also not assumed independent across  $i$ , but to satisfy "global" assumptions requiring some falling-off in dependence (see e.g. Assumption A6 below).

A key role is played by an assumption on the disturbances

$$U_i = U_{in} = Y_i - g(X_i), \quad 1 \leq i \leq n, \quad n = 1, 2, \dots \quad (2.2)$$

We assume

$$U_i = \sigma_i(X_i) V_i, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2.3)$$

where  $\sigma_i(x) = \sigma_{in}(x)$  and  $V_i = V_{in}$  are both scalars; as desired the first, and also second, moment of  $\sigma_i(X_i)$  exists; and, for all  $n = 1, 2, \dots$ ,  $\{V_i, 1 \leq i \leq n\}$  is independent of  $\{X_i, 1 \leq i \leq n\}$ . We assume that

$$E(V_i) = 0, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2.4)$$

implying immediately the conditional moment restriction  $E\{U_i | X_i\} = 0, 1 \leq i \leq n, n = 1, 2, \dots$ . As the  $\sigma_i^2(x)$  are unknown functions, if  $V_i$  has finite variance, with no loss of generality we fix

$$Var\{V_i\} \equiv 1, \quad (2.5)$$

whence

$$Var\{Y_i | X_i\} = Var\{U_i | X_i\} = \sigma_i^2(X_i), \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2.6)$$

so conditional heteroscedasticity is permitted. We do not assume the  $\sigma_i^2(x)$  are constant across  $i$ , thus allowing unconditional heteroscedasticity also, though again homogeneity restrictions will be imposed.

Dependence across  $i$  is principally modelled via the  $V_i$ . For many, though not all, of our results we assume

$$V_i = \sum_{j=1}^{\infty} a_{ij} \varepsilon_j, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2.7)$$

where for each  $n$ , the  $\varepsilon_j$ ,  $j \geq 1$ , are independent random variables with zero mean, and the nonstochastic weights  $a_{ij} = a_{ijn}$  are at least square-summable over  $j$ , whence with no loss of generality we fix

$$\sum_{j=1}^{\infty} a_{ij}^2 \equiv 1, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots \quad (2.8)$$

When the  $\varepsilon_j$  have finite variance we fix  $Var\{\varepsilon_i\} = 1$ , implying (2.5). An alternative to the linear dependence structure (2.7) is some form of mixing condition, which indeed could cover some heterogeneity as well as dependence. In fact mixing could be applied directly to the  $X_i$  and  $U_i$ , avoiding the requirement of independence between  $\{V_i\}$  and  $\{X_i\}$ , or simply to the observable  $\{Y_i, X_i\}$ . Mixing conditions, when applied to our triangular array, would require a notion of falling off of dependence as  $|i - j|$  increases, which, as previously indicated, is not relevant to all spatial situations of interest. Moreover, we allow for a stronger form of dependence than mixing; we usually do not require, for example, that the  $a_{ij}$  are summable with respect to  $j$ , and thence cover forms of long-range dependence analogous to those familiar in time series analysis.

The linear structure (2.7) obviously covers equally-spaced stationary time series, where  $a_{ij}$  is of the form  $a_{i-j}$ , and lattice extensions, where the infinite series is required not only to model long range dependence but also finite-degree autoregressive structure in  $V_i$ . Condition (2.7) also provides an extension of SAR models. These typically imply

$$V_i = \sum_{j=1}^n a_{ij} \varepsilon_j, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2.9)$$

so there is a mapping from  $n$  independent innovations  $\varepsilon_i$  to the  $n$  possibly dependent  $V_i$ . In particular, we may commence from a parametric structure

$$(I_n - \omega_1 W_1 - \dots - \omega_{m_1} W_{m_1}) U = (I_n - \omega_{m_1+1} W_{m_1+1} - \dots - \omega_{m_1+m_2} W_{m_1+m_2}) \sigma \varepsilon, \quad (2.10)$$

where the integers  $m_1$ ,  $m_2$  are given,  $I_n$  is the  $n \times n$  identity matrix,  $U = (U_1, \dots, U_n)'$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ , the  $\omega_i$  are unknown scalars,  $\sigma$  is an unknown scale factor, and the  $W_i = W_{in}$  are given  $n \times n$  "weight" matrices (satisfying further conditions in order to guarantee identifiability of the  $\omega_i$ ), and such that matrix on the left hand side multiplying  $U$  is nonsingular. Of course (2.10) is similar to autoregressive-moving-average structure for stationary time series, but that is generated from an infinite sequence of innovations, not only  $n$  such (though of course  $n$  will increase in the asymptotic theory). There seems no compelling



reason to limit the number of innovations to  $n$  in spatial modelling, and (2.9) cannot cover forms of long-range dependence, unless somehow the sums  $\sum_{j=1}^n a_{ij}$  are permitted to increase in  $n$  without bound, which is typically ruled out in the SAR literature.

### 3 Kernel regression estimate

We introduce a kernel function  $K(u) : \mathbb{R}^d \Rightarrow \mathbb{R}$ , satisfying at least

$$\int_{\mathbb{R}^d} K(u) du = 1. \quad (3.1)$$

The Nadaraya-Watson kernel estimate of  $g(x)$ , for a given  $x \in \mathbb{R}^d$ , is

$$\hat{g}(x) = \hat{g}_n(x) = \frac{\hat{v}(x)}{\hat{f}(x)}, \quad (3.2)$$

where

$$\hat{f}(x) = \hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^n K_i(x), \quad \hat{v}(x) = \hat{v}_n(x) = \frac{1}{nh^d} \sum_{i=1}^n Y_i K_i(x), \quad (3.3)$$

with

$$K_i(x) = K_{in}(x) = K\left(\frac{x - X_i}{h}\right), \quad (3.4)$$

and  $h = h_n$  is a scalar, positive bandwidth sequence, such that  $h \rightarrow 0$  as  $n \rightarrow \infty$ .

Classically, the literature is concerned with a sequence  $X_i$  of identically distributed variables, having probability density  $f(x)$ , with  $X_i$  observed at  $i = 1, \dots, n$ , so  $f_i(x) \equiv f(x)$ . In this case  $\hat{f}(x)$  estimates  $f(x)$ , and  $\hat{v}(x)$  estimates  $g(x)f(x)$ , so that  $\hat{g}(x)$  estimates  $g(x)$ . The last conclusion results also in our possibly non-identically distributed, triangular array setting, because under suitable additional conditions,

$$\hat{f}(x) - \bar{f}(x) \rightarrow_p 0, \quad (3.5)$$

$$\hat{v}(x) - g(x)\bar{f}(x) \rightarrow_p 0, \quad (3.6)$$

where

$$\bar{f}(x) = \bar{f}_n(x) = \frac{1}{n} \sum_{i=1}^n f_i(x). \quad (3.7)$$

It follows from (3.5) and (3.6) and Slutsky's theorem, that

$$\hat{g}(x) \rightarrow_p g(x), \quad (3.8)$$

so long as  $\lim_{n \rightarrow \infty} \bar{f}(x) > 0$ . In fact though we establish (3.5), we do not employ this result in establishing (3.8), but instead a more subtle argument (that avoids continuity of  $\bar{f}(x)$ ). The consistency results are presented in the following section, with Section 5 then establishing a central limit theorem for  $\hat{g}(x)$ .

## 4 Consistency of kernel regression estimate

We introduce first some conditions of a standard nature on the kernel function  $K(u)$ .

**Assumption A1:**  $K(u)$  is an even function, and

$$\sup_{u \in \mathbb{R}^d} |K(u)| + \int_{\mathbb{R}^d} |K(u)| du < \infty. \quad (4.1)$$

**Assumption A2( $\xi$ ):** As  $\|u\| \rightarrow \infty$ ,

$$K(u) = O(\|u\|^{-\xi}). \quad (4.2)$$

For  $\xi > d$ , A2( $\xi$ ) plus the first part of A1 implies the second part of A1. Note for future use that, for  $\varepsilon > 0$ ,

$$\sup_{\|u\| \geq \varepsilon/h} |K(u)| = O(h^\xi). \quad (4.3)$$

**Assumption A3:**

$$K(u) \geq 0, \quad u \in \mathbb{R}^d. \quad (4.4)$$

Assumption A3 excludes higher-order kernels, but can be avoided if conditions on the  $X_{in}$  are slightly strengthened.

The following condition on the bandwidth  $h$  is also standard.

**Assumption A4:** As  $n \rightarrow \infty$ ,

$$h + (nh^d)^{-1} \rightarrow 0. \quad (4.5)$$

For  $\varepsilon > 0$  define

$$\alpha(x; \varepsilon) = \alpha_n(x; \varepsilon) = \sup_{\|w\| < \varepsilon} \bar{f}(x - w). \quad (4.6)$$

**Assumption A5( $x$ ):** For some  $\varepsilon > 0$ ,

$$\overline{\lim}_{n \rightarrow \infty} \alpha(x; \varepsilon) < \infty. \quad (4.7)$$

**Assumption A6( $x, y$ ):** The joint density of  $X_i$  and  $X_j$ ,  $f_{ij}(x, y) = f_{ijn}(x, y)$ , exists for all  $i, j$ , and for some  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \rho(x, y; \varepsilon) = 0, \quad (4.8)$$

where

$$\rho(u, v; \varepsilon) = \rho_n(u, v; \varepsilon) = \sup_{\|w\| < \varepsilon} |m(u - w, v - w)|, \quad (4.9)$$

$$m(u, v) = m_n(u, v) = \frac{1}{n^2} \sum_{\substack{i, j=1 \\ i \neq j}}^n \{f_{ij}(u, v) - f_i(u)f_j(v)\}. \quad (4.10)$$

**Assumption A7(x):** For some  $\varepsilon > 0$ ,

$$\varliminf_{n \rightarrow \infty} \inf_{\|u\| < \varepsilon} \bar{f}(x - u) > 0 \quad (4.11)$$

**Assumption A8:**  $g$  is continuous at  $x$ , and

$$\varlimsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E |g(X_i)| < \infty. \quad (4.12)$$

Define  $\mu_i(u) = \mu_{in}(u) = \sigma_i^2(u)f_i(u)$ ,  $\nu_{ij}(u, v) = \nu_{ijn}(u, v) = \sigma_i(u)\sigma_j(v)f_{ij}(u, v)$ .

**Assumption A9(x, y):** For some  $\varepsilon > 0$ ,

$$\varlimsup_{n \rightarrow \infty} \max_{1 \leq i \leq n} \sup_{\|u\| < \varepsilon} \mu_i(x - u) < \infty \quad (4.13)$$

and

$$\varlimsup_{n \rightarrow \infty} \max_{1 \leq i, j \leq n} \sup_{\|u\| < \varepsilon, \|v\| < \varepsilon} |\nu_{ij}(x - u, y - v)| < \infty; \quad (4.14)$$

also

$$\varlimsup_{n \rightarrow \infty} \max_{1 \leq i \leq n} E \{\sigma_i^2(X_i)\} < \infty. \quad (4.15)$$

**Assumption A10:** For  $n = 1, 2, \dots$ ,  $\{X_i, 1 \leq i \leq n\}$  is independent of  $\{V_i, 1 \leq i \leq n\}$ , (2.4) and (2.5) hold, and the covariances

$$\gamma_{ij} = \gamma_{ijn} = \text{Cov} \{V_i, V_j\}, \quad 1 \leq i, j \leq n, \quad n = 1, 2, \dots, \quad (4.16)$$

satisfy

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i, j=1, i \neq j}^n |\gamma_{ij}| = 0. \quad (4.17)$$

**Theorem 1:** Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A3, A4, A5(x), A6(x, x), A7, A8, A9(x, x) and A10 hold. Then

$$\hat{g}(x) \rightarrow_p g(x), \text{ as } n \rightarrow \infty. \quad (4.18)$$

**Proof** For any  $\eta > 0$

$$\begin{aligned} P(|\hat{g}(x) - g(x)| > \eta) &\leq P\left(\left|(nh^d)^{-1} \sum_{i=1}^n \{Y_i - g(x)\} K_i(x)\right| > \eta^2\right) \\ &+ P(\hat{f}(x) < \eta) \end{aligned} \quad (4.19)$$

Using Lemmas 1 and 2, for  $\zeta > \eta$ ,

$$P(\widehat{f}(x) < \eta) \leq P(|\widehat{f}(x) - E\{\widehat{f}(x)\}| > \zeta - \eta) \leq \text{Var}\{\widehat{f}(x)\} / (\zeta - \eta)^2 \rightarrow 0. \quad (4.20)$$

It remains to show that the first probability on the right of (4.19) is negligible. But

$$\frac{1}{nh^d} \sum_{i=1}^n \{Y_n - g(x)\} K_i(x) = \widehat{r}_1(x) + \widehat{r}_2(x), \quad (4.21)$$

where

$$\widehat{r}_1(x) = \widehat{r}_{1n}(x) = \frac{1}{nh^d} \sum_{i=1}^n U_i K_i(x), \quad (4.22)$$

$$\widehat{r}_2(x) = \widehat{r}_{2n}(x) = \frac{1}{nh_n^d} \sum_{i=1}^n \{g(X_i) - g(x)\} K_i(x), \quad (4.23)$$

whence it remains only to apply Lemmas 3 and 4.  $\square$

The linear representation (2.7) (or (2.9)) is not imposed in Theorem 1. To provide also a consistency result that avoids finite variance of  $V_i$ , while on the other hand strengthening dependence, we employ (2.7). In this setting,  $\sigma_i(X_i)$  no longer represents a standard deviation, but simply a scale factor. For  $D > 0$ , define  $\varepsilon'_i = \varepsilon'_{in} = \varepsilon_i 1(|\varepsilon_i| \leq D)$ ,  $\varepsilon''_i = \varepsilon''_{in} = \varepsilon_i - \varepsilon'_{in}$ .

**Assumption A11:** (2.7) holds, where, for all  $n \geq 1$ ,  $\{X_i, 1 \leq i \leq n\}$  is independent of  $\{\varepsilon_i, i \geq 1\}$ , the  $\varepsilon_i$  are independent random variables with zero mean and

$$\limsup_{D \rightarrow \infty} \sup_{n \geq 1} \sup_{i \geq 1} E|\varepsilon''_i| = 0, \quad (4.24)$$

$$\sup_{n \geq 1} \sup_{j \geq 1} \sum_{i=1}^{\infty} |a_{ij}| + \sup_{n \geq 1} \max_{1 \leq i \leq n} \sum_{j=1}^{\infty} |a_{ij}| < \infty. \quad (4.25)$$

Notice that if the  $\varepsilon_i$  do in fact have finite variance (taken to be 1),

$$\gamma_{ij} = \sum_{k=1}^{\infty} a_{ik} a_{jk} \quad (4.26)$$

so (cf (4.17)) under (4.25)

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i,j=1}^n |\gamma_{ij}| < \infty. \quad (4.27)$$

**Theorem 2:** Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A3, A4, A5( $x$ ), A6( $x, x$ ), A7, A8, A9( $x, x$ ) and A11 hold. Then

$$\widehat{g}(x) \rightarrow_p g(x), \text{ as } n \rightarrow \infty. \quad (4.28)$$

**Proof** Identical to that of Theorem 2, but Lemma 5 is used in place of Lemma 4.  $\square$

## 5 Asymptotic normality of kernel regression estimate

Define

$$s = s_n = \frac{h^{-d}}{n^2} \sum_{i=1}^n \gamma_{ii} = (nh^d)^{-1}, \quad (5.1)$$

$$t = t_n = \left| \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i \neq j}}^n \gamma_{ij} \right|, \quad (5.2)$$

the third equality in (5.1) using (2.8). Note that as  $n \rightarrow \infty$ ,  $s \rightarrow 0$  under Assumption A4 and  $t \rightarrow 0$  under Assumption A10. In the present section we establish a central limit theorem for  $w^{-1/2} \{\widehat{g}(x) - g(x)\}$ , where  $w = w_n = s$  if  $t = O(s)$  and  $w = t$  if  $s = o(t)$ , as  $n \rightarrow \infty$ . Under our conditions,  $s$  decays at the standard rate  $(nh^d)^{-1}$ , whereas  $t$  is zero when the  $V_i$  are uncorrelated, and decays faster than  $s$  when  $V_i$  is short-range dependent, but slower when  $V_i$  is long-range dependent. The modulus operator in  $t$  is to ensure non-negativity; however, because

$$\sum_{i=1}^n \gamma_{ii} + \sum_{i,j=1; i \neq j}^n \gamma_{ij} = \text{Var} \left\{ \sum_{i=1}^n V_i \right\} \geq 0, \quad (5.3)$$

the second term on the left cannot be negative when  $s = o(t)$ , so when the  $t$  normalization applies the modulus operator is unnecessary.

We have

$$\widehat{g}(x) - g(x) = \frac{\widehat{v}(x) - g(x)\widehat{f}(x)}{\widehat{f}(x)} = \frac{\widehat{r}_1(x)}{\widehat{f}(x)} + \frac{\widehat{r}_2(x)}{\widehat{f}(x)}, \quad (5.4)$$

where  $\widehat{r}_1(x)$ ,  $\widehat{r}_2(x)$  are given in (4.22), (4.23). There is a basic difference in our method of proof from that in, say, Robinson (1983). There (where the assumptions imply that  $\widehat{g}(x)$  is  $(nh^d)^{1/2}$ -consistent) the delta method was used, after establishing asymptotic joint normality of  $(nh^d)^{1/2} \{\widehat{f}(x) - \bar{f}(x)\}$  and  $(nh^d)^{1/2} \{\widehat{v}(x) - g(x)\bar{f}(x)\}$ , which in turn follows from asymptotic joint normality of  $(nh^d)^{1/2} [\widehat{f}(x) - E\{\widehat{f}(x)\}]$  and  $(nh^d)^{1/2} [\widehat{v}(x) - E\{\widehat{v}(x)\}]$ , and the properties  $E\{\widehat{f}(x)\} - \bar{f}(x) = o((nh^d)^{-1/2})$ ,  $E\{\widehat{v}(x)\} - g(x)\bar{f}(x) = o((nh^d)^{-1/2})$ . We, however, proceed from the rightmost expression in (5.4) by establishing asymptotic normality of  $w^{-1/2}\widehat{r}_1(x)$ , along with  $E\{|\widehat{r}_2(x)|\} = o(w^{1/2})$  and  $\widehat{f}(x) - \bar{f}(x) \rightarrow_p 0$ .

We begin by showing the latter result, or rather  $\widehat{f}(x_i) - \bar{f}(x_i) \rightarrow_p 0$  for any finitely many distinct points  $x_1, \dots, x_p$  in  $\mathbb{R}^d$ , because our goal in fact is to

establish asymptotic joint normality of the appropriately normed vector  $\widehat{G}_n - G$ , where

$$G = \{g(x_1), \dots, g(x_p)\}', \quad \widehat{G} = \widehat{G}_n = \{\widehat{g}_n(x_1), \dots, \widehat{g}_n(x_p)\}'. \quad (5.5)$$

We introduce a condition on

$$\beta(x; \varepsilon) = \beta_n(x; \varepsilon) = \sup_{\|w\| < \varepsilon} |\bar{f}(x-w) - \bar{f}(x)|. \quad (5.6)$$

**Assumption B1(x):** For all  $\delta > 0$  there exists  $\varepsilon > 0$  such that for all sufficiently large  $n$

$$\beta(x; \varepsilon) < \delta. \quad (5.7)$$

**Theorem 3:** Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A4, A6( $x_i, x_i$ ) and B1( $x_i$ ) hold,  $i = 1, \dots, p$ . Then

$$\widehat{f}(x_i) - \bar{f}(x_i) \rightarrow_p 0, \quad i = 1, \dots, p. \quad (5.8)$$

**Proof:** Follows from Lemmas 1 and 6.  $\square$

**Assumption B2(x):**  $g$  satisfies a Lipschitz condition of degree  $\theta \in (0, 1]$  in a neighbourhood of  $x$ .

**Assumption B3:** For the same  $\theta$  as in Assumption B2(u),  $h^{2\theta}/w \rightarrow 0$  as  $n \rightarrow \infty$ .

**Assumption B4:** (2.7) and (2.8) hold, where for all  $n$ ,  $\{X_i, 1 \leq i \leq n\}$  is independent of  $\{\varepsilon_i, i \geq 1\}$ , and the  $\varepsilon_i$  are independent random variables with zero mean, unit variance, and

$$\lim_{D \rightarrow \infty} \sup_{n \geq 1} \sup_{i \geq 1} E \left\{ \varepsilon_i''^2 \right\} = 0, \quad (5.9)$$

where  $\varepsilon_i''$  is as defined before Assumption A11.

Define

$$b = b_n = \sup_{j \geq 1} \left( \sum_{i=1}^n |a_{ij}| \right)^2 / \sum_{j=1}^{\infty} \left( \sum_{i=1}^n |a_{ij}| \right)^2. \quad (5.10)$$

**Assumption B5:**

$$\overline{\lim}_{n \rightarrow \infty} \sup_{j \geq 1} \sum_{i=1}^n a_{ij}^2 < \infty, \quad (5.11)$$

and

$$\lim_{n \rightarrow \infty} b \rightarrow 0. \quad (5.12)$$

**Assumption B6:** When  $s = o(t)$ ,

$$\sum_{i,j,k=1}^n \gamma_{ij}\gamma_{ik} = o(n^3t), \text{ as } n \rightarrow \infty. \quad (5.13)$$

**Assumption B7:** The densities  $f_i$  of  $X_i$  ( $1 \leq i \leq n$ ),  $f_{i_1 i_2}$  of  $X_{i_1}, X_{i_2}$  ( $1 \leq i_1 < i_2 \leq n$ ),  $f_{i_1 i_2 i_3}$  of  $X_{i_1}, X_{i_2}, X_{i_3}$  ( $1 \leq i_1 < i_2 < i_3 \leq n$ ), and  $f_{i_1 i_2 i_3 i_4}$  of  $X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}$  ( $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$ ) are bounded uniformly in large  $n$  in neighbourhoods of all combinations of arguments  $x_1, \dots, x_p$ .

**Assumption B8( $x, y$ ):** For all  $\delta > 0$  there exists  $\varepsilon > 0$  such that for all sufficiently large  $n$

$$\max_{1 \leq i \leq n} \sup_{\|w\| < \varepsilon} |f_i(x-w) - f_i(x)| < \delta. \quad (5.14)$$

$$\max_{1 \leq i \leq n} \sup_{\|w\| < \varepsilon} |\sigma_i^2(x-w) - \sigma_i^2(x)| < \delta \quad (5.15)$$

$$\max_{1 \leq i < j \leq n} \sup_{\|w\| < \varepsilon, \|z\| < \varepsilon} |f_{ij}(x-w, y-z) - f_{ij}(x, y)| < \delta. \quad (5.16)$$

**Assumption B9( $x$ ):** For some  $\varepsilon > 0$ ,

$$\overline{\lim}_{n \rightarrow \infty} \max_{1 \leq i \leq n} \left[ \sup_{|u_s| < \varepsilon} |\sigma_{i_s}(x - u_s)| + E \{ \sigma_i^4(X_i) \} \right] < \infty. \quad (5.17)$$

Write

$$\begin{aligned} \rho_{i_1 i_2}(u_1, u_2, \varepsilon) &= \rho_{i_1 i_2 n}(u_1, u_2, \varepsilon) = \sup_{\substack{|v_s| < \varepsilon \\ s=1,2}} |f_{i_1 i_2}(u_1 - v_1, u_2 - v_2) \\ &\quad - f_{i_1}(u_1 - v_1) f_{i_2}(u_2 - v_2)|, \end{aligned} \quad (5.18)$$

$$\begin{aligned} \rho_{i_1 i_2 i_3}(u_1, u_2, u_3, \varepsilon) &= \rho_{i_1 i_2 i_3 n}(u_1, u_2, u_3, \varepsilon) = \sup_{\substack{|v_s| < \varepsilon \\ s=1,2,3}} |f_{i_1 i_2 i_3}(u_1 - v_1, u_2 - v_2, u_3 - v_3) \\ &\quad - f_{i_1}(u_1 - v_1) f_{i_2 i_3}(u_2 - v_2, u_3 - v_3)|, \end{aligned} \quad (5.19)$$

$$\begin{aligned} \rho_{i_1 i_2 i_3 i_4}(u_1, u_2, u_3, u_4, \varepsilon) &= \rho_{i_1 i_2 i_3 i_4 n}(u_1, u_2, u_3, u_4, \varepsilon) \\ &= \sup_{\substack{|v_s| < \varepsilon \\ s=1, \dots, 4}} |f_{i_1 i_2 i_3 i_4}(u_1 - v_1, u_2 - v_2, u_3 - v_3, u_4 - v_4) \\ &\quad - f_{i_1 i_2}(u_1 - v_1, u_2 - v_2) f_{i_3 i_4}(u_3 - v_3, u_4 - v_4)|. \end{aligned} \quad (5.20)$$

**Assumption B10** For  $(u_1, u_2, u_3, u_4) = x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}$ ,  $i_s = 1, \dots, p$ ,  $s = 1, \dots, 4$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{\substack{i_s=1 \\ s=1,2}}^n \rho_{i_1 i_2}(u_1, u_2, \varepsilon) = 0, \quad (5.21)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^2 t} \sum_{\substack{i_s=1 \\ s=1,2,3}}^n \left| \gamma_{i_2 i_3} \right| \rho_{i_1 i_2 i_3}(u_1, u_2, u_3, \varepsilon) = 0, \quad (5.22)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^4 t^2} \sum_{\substack{i_s=1 \\ s=1,\dots,4}}^n \left| \gamma_{i_1 i_2} \gamma_{i_3 i_4} \right| \rho_{i_1 i_2 i_3 i_4}(u_1, u_2, u_3, u_4, \varepsilon) = 0, \quad (5.23)$$

where each primed sum omits terms with any equalities between the relevant  $i_s$ .

**Assumption B11:**

$$\overline{\lim}_{n \rightarrow \infty} \sum_{k=1}^{\infty} \sum_{i,j=1;i \neq j}^n |a_{ik} a_{jk}| / \sum_{k=1}^{\infty} \sum_{i,j=1;i \neq j}^n a_{ik} a_{jk} < \infty. \quad (5.24)$$

**Assumption B12:** *There exists a function  $\phi(x)$ ,  $x \in \mathbb{R}^d$  such that*

$$\bar{f}(x) \sim \phi(x), \text{ as } n \rightarrow \infty. \quad (5.25)$$

When  $t = O(s)$  there exists a function  $\lambda(x)$ ,  $x \in \mathbb{R}^d$ , such that

$$\frac{1}{n} \sum_{i=1}^n \mu_i(x) \sim \lambda(x), \text{ as } n \rightarrow \infty. \quad (5.26)$$

When  $s = O(t)$  there exists a function  $\psi(x, y)$ ,  $x, y \in \mathbb{R}^d$ , such that

$$\frac{\sum_{i,j=1;i \neq j}^n \gamma_{ij} \nu_{ij}(x, y)}{\sum_{i,j=1;i \neq j}^n \gamma_{ij}} \sim \psi(x, y), \text{ as } n \rightarrow \infty. \quad (5.27)$$

Denote  $\Phi = \text{diag} \{ \phi(x_1), \dots, \phi(x_p) \}$ ,  $\Lambda = \text{diag} \{ \lambda(x_1), \dots, \lambda(x_p) \}$ , and  $\Psi$  the  $p \times p$  matrix with  $(i, j)$ th element  $\psi(x_i, y_j)$ .

**Assumption B13( $q$ ):**  $\text{rank} \{ \Psi \} = q \in [1, p]$ .

The reason for allowing  $q < p$  is discussed in the following section. Define by  $J^{(q)}$  any  $q \times p$  matrix formed from  $q$  rows of  $I_p$ ; for given  $q$  there are  $\binom{p}{q}$  of these, but we do not distinguish between them in our notation. Define also

$$\kappa = \int_{\mathbb{R}^d} K^2(u) du, \quad (5.28)$$



which is finite under Assumption A1.

**Theorem 4:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 4d$ , A4, A7( $x_i$ ), A8, A9( $x_i, x_j$ ), A10, B1( $x_i$ ), B2( $x_i$ ), B3, B4, B5, B7, B8( $x_i, x_j$ ), B9( $x_i$ ), B10-B12 and B13( $q$ ) hold,  $i, j = 1, \dots, p$ .*

(i) *If also  $t = o(s)$ ,*

$$s^{-1/2} \left( \widehat{G} - G \right) \rightarrow_d \mathcal{N}(0, \kappa \Phi^{-1} \Lambda \Phi^{-1}), \text{ as } n \rightarrow \infty; \quad (5.29)$$

(ii) *if also  $t \sim \chi s$ ,*

$$s^{-1/2} \left( \widehat{G} - G \right) \rightarrow_d \mathcal{N}(0, \Phi^{-1} (\kappa \Lambda + \chi \Psi) \Phi^{-1}), \text{ as } n \rightarrow \infty; \quad (5.30)$$

(iii) *if also  $s = o(t)$ , for any  $J^{(q)}$*

$$t^{-1/2} J^{(q)} \left( \widehat{G} - G \right) \rightarrow_d \mathcal{N}(0, J^{(q)} \Phi^{-1} \Psi \Phi^{-1} J^{(q)'}), \text{ as } n \rightarrow \infty. \quad (5.31)$$

**Proof:** From (5.4),

$$w^{-1/2} (\widehat{g} - g) = \widehat{f}^{-1} w^{-1/2} (\widehat{r}_1 + \widehat{r}_2), \quad (5.32)$$

where

$$\widehat{f} = \widehat{f}_n = \text{diag} \left\{ \widehat{f}(x_1), \dots, \widehat{f}(x_p) \right\}, \quad \widehat{r}_i = \widehat{r}_{in} = \left\{ \widehat{r}_i(x_1), \dots, \widehat{r}_i(x_p) \right\}', \quad i = 1, 2. \quad (5.33)$$

We deduce  $\widehat{q} \rightarrow_p \Phi$  from Lemmas 1 and 6 and  $\widehat{r}_2 = o_p(w^{1/2})$  from Lemma 7 and Assumptions B2( $x_i$ ),  $i = 1, \dots, p$  and B3. Define

$$\begin{aligned} \Sigma &= \kappa \Lambda, \text{ if } t = o(s), \\ &= \Psi, \text{ if } s = o(t), \\ &= (\kappa \Lambda + \chi \Psi), \text{ if } t \sim \chi s. \end{aligned} \quad (5.34)$$

We have now to allow for the possibility that  $\Psi$  is singular, i.e.  $q < p$ . This only affects part (iii) of the Theorem but no generality is lost by giving a single proof for all three cases for  $J^{(q)} \left( \widehat{G} - G \right)$ . Thus define  $\Sigma^{(q)} = J^{(q)} \Sigma J^{(q)'}$ ,  $\widehat{r}_1^{(q)} = J^{(q)} \widehat{r}_1$ .

It remains to prove

$$w^{-1/2} \widehat{r}_1^{(q)} \rightarrow_d \mathcal{N}(0, \Sigma^{(q)}). \quad (5.35)$$

We can write

$$\widehat{r}_1^{(q)} = \sum_{j=1}^{\infty} Z_j^{(q)} \varepsilon_j, \quad Z_j^{(q)} = Z_{jn}^{(q)} = \frac{1}{nh^d} \sum_{i=1}^n J^{(q)} K_i \sigma_i(X_i) a_{ij}, \quad (5.36)$$

where

$$K_i = \{K_i(x_1), \dots, K_i(x_p)\}' \quad (5.37)$$

For positive integer  $N = N_n$ , increasing with  $n$ , define

$$\widehat{r}_1^{(q)*} = \widehat{r}_{1n}^{(q)*} = \sum_{j=1}^N Z_j^{(q)} \varepsilon_j, \quad \widehat{r}_1^{(q)\#} = \widehat{r}_{1n}^{(q)\#} = \widehat{r}_1^{(q)} - \widehat{r}_1^{(q)*}. \quad (5.38)$$

By Lemma 9, there exists an  $N$  sequence such that  $\widehat{r}_1^{(q)\#} = o_p(w^{-1/2})$ .

For such  $N$ , consider

$$T = T_n = E \left\{ \widehat{r}_1^{(q)*} r_1^{(q)*'} \mid X \right\} = \sum_{j=1}^N Z_j^{(q)} Z_j^{(q)'} \quad (5.39)$$

and introduce a  $q \times q$  matrix  $P = P_n$  such that  $T = PP'$ . For  $n$  large enough  $T$  is positive definite under our conditions. For a  $q \times 1$  vector  $\tau$ , such that  $\tau' \tau = 1$ , write

$$c^* = c_n^* = \tau' P^{-1} \widehat{r}_1^{(q)*}, \quad (5.40)$$

so  $E \{c^{*2}\} = 1$ . We show that, conditionally on  $\{X; \text{all } n \geq 1\}$ ,

$$c^* \rightarrow_d \mathcal{N}(0, 1), \quad (5.41)$$

whence by the Cramer-Wold device,

$$P^{-1} \widehat{r}_1^{(q)*} \rightarrow_d \mathcal{N}(0, I_q), \quad (5.42)$$

which implies unconditional convergence, Then for a  $q \times q$  matrix  $\Pi$  such that  $\Sigma^{(q)} = \Pi \Pi'$  it follows that

$$w^{-1/2} \Pi^{-1} \widehat{r}_1^{(q)*} \rightarrow_d \mathcal{N}(0, I_q) \quad (5.43)$$

if  $w^{-1/2} \Pi^{-1} P$  converges in probability to an orthogonal matrix, which is implied if  $w^{-1} \Pi^{-1} P P' \Pi^{-1'} \rightarrow_p I_q$ , i.e. if

$$w^{-1} T \rightarrow_p \Sigma^{(q)}. \quad (5.44)$$

But

$$E \{T\} = E \left\{ \widehat{r}_1^{(q)*} r_1^{(q)*'} \right\} + E \left\{ \widehat{r}_1^{(q)\#} \widehat{r}_1^{(q)\#'} - \widehat{r}_1^{(q)} \widehat{r}_1^{(q)\#'} - \widehat{r}_1^{(q)\#} \widehat{r}_1^{(q)'} \right\}, \quad (5.45)$$

and the norm of the final expectation is  $o(w)$  by the Schwarz inequality and Lemmas 8 and 9, while  $w^{-1} E \left\{ \widehat{r}_1^{(q)} \widehat{r}_1^{(q)'} \right\} \rightarrow \Sigma^{(q)}$  from Lemma 8. Lemma 10 completes the proof of (5.44).

To prove (5.41), write

$$c^* = \sum_{j=1}^N z_j^{(q)} \varepsilon_j, \quad z_j^{(q)} = z_{jn}^{(q)} = \tau' P^{-1} Z_j^{(q)}. \quad (5.46)$$

Since  $z_j^{(q)} \varepsilon_j$  is a martingale difference sequence, and

$$\sum_{j=1}^N z_j^{(q)2} = 1, \quad (5.47)$$

(5.41) follows, from e.g. Scott (1973), if, for any  $\eta > 0$

$$\sum_{j=1}^N E \left\{ z_j^{(q)2} \varepsilon_j^2 \mathbf{1}(|z_j^{(q)} \varepsilon_j| > \eta) \mid X_1, \dots, X_n \right\} \rightarrow_p 0, \quad (5.48)$$

as  $n \rightarrow \infty$ . Now for any  $r > 0$   $\left\{ |z_j^{(q)} \varepsilon_j| > \eta \right\} \subset \left\{ \varepsilon_j^2 > \eta r \right\} \cup \left\{ z_j^{(q)2} > \eta/r \right\}$ , so by independence of the  $\varepsilon_j$  and  $X_1, \dots, X_n$  the left side is bounded by

$$\sum_{j=1}^N z_j^{(q)2} E \left\{ \varepsilon_j^2 \mathbf{1}(\varepsilon_j^2 > \eta/r) \right\} + \sum_{j=1}^N z_j^{(q)2} \mathbf{1}(z_j^{(q)2} > \eta/r) \quad (5.49)$$

which, from (5.47), is bounded by

$$\max_{j \geq 1} E \left\{ \varepsilon_j^2 \mathbf{1}(\varepsilon_j^2 > \eta/r) \right\} + \frac{1}{\eta r} \sum_{j=1}^N z_j^{(q)4}. \quad (5.50)$$

The first term can be made arbitrarily small for small enough  $r$ , while the second is  $o_p(1)$  by Lemma 11.  $\square$

## 6 Discussion

Under conditions motivated by spatial or spatio-temporal settings, we have established consistency of the Nadaraya-Watson estimate under relatively broad conditions, and asymptotic normality under stronger conditions. Our discussion focusses on the relevance of some of the conditions, and some implications of these results.

1. Assumption A5( $x$ ) is implied by  $\overline{\lim}_{n \rightarrow \infty} \max_{1 \leq i \leq n} \sup_{\|w\| < \varepsilon} f_i(x-w) < \infty$ , and with identically distributed  $X_i$ , both requirements are equivalent to boundedness of  $f$  in a neighbourhood of  $x$ . Other conditions on  $X_i$  likewise simplify. In case of a regularly-spaced stationary time series  $\{X_i\}$ , we have  $f_{ij}(x, y) = f_{|i-j|}(x, y)$ , so

$$m(x, y) = \frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \{f_j(x, y) - f(x)f(y)\}. \quad (6.1)$$

In this setting, with  $d = 1$ , Castellana and Leadbetter (1986) established pointwise consistency of kernel probability density estimates for regularly-spaced time

series under the assumption

$$\lim_{n \rightarrow \infty} \sup_{u, v \in \mathbb{R}^d} \frac{1}{n} \sum_{j=1}^{n-1} |f_j(x, y) - f(x)f(y)| = 0. \quad (6.2)$$

Even after adding stationarity to our conditions, Assumption A6( $x, y$ ), used for consistency of density (Theorem 3) as well regression (Theorems 1 and 2) estimates, is milder than (6.2), both because in (6.2) the modulus is inside the summation and because the supremum is over all  $x, y \in \mathbb{R}^d$ . Castellana and Leadbetter (1986) showed that (6.2) holds in case of a scalar Gaussian process  $\{X_i\}$  with lag- $j$  autocovariance tending to zero as  $j \rightarrow \infty$ , and thus covers arbitrarily strong long range dependence. These observations extend to the vector case  $d > 1$ , whence Assumption A6( $x, y$ ) is satisfied also. Moreover, (5.21)-(5.23) of Assumption B10 are of a similar character, because all involve convergence to zero, with no rate, of weighted averages of density-based measures of dependence (though suprema are now inside the summations). If we employed instead an asymptotic normality proof for the  $\hat{f}(x_i)$  in proving Theorem 4, a possibility mentioned in Section 5, density-based conditions on  $X_i$  would, however, have to entail rates, as indeed would mixing conditions. (Castellana and Leadbetter (1986), like other authors, used mixing conditions in their central limit theorem for density estimates from stationary time series.)

2. With respect to the conditions on  $V_i$ , the requirement A10 for consistency in Theorem 1 implies an arbitrarily slow decay in contributions from  $\gamma_{ij}$  as, say,  $|i - j|$  diverges, and under stationarity is satisfied by arbitrarily strong long range dependence. On the other hand avoiding a second moment in Theorem 2 rules out long range dependence, while easily covering conventional forms of weak dependence. Asymptotic normality in Theorem 4 also permits general long range dependence, though the actual strength of this in part determines the precise outcome, including convergence rate, see parts (i)-(iii). Assumption B11 means that changing the sign of any negative  $a_{ij}$  would not make  $t$  decay slower, and could be satisfied more generally if the  $a_{ij}$  are eventually positive as  $j$  increases. The Lindeberg condition (5.12) was checked for linear time series with arbitrarily strong dependence in the central limit theorem for the sample mean by Ibragimov and Linnik (1971), and for fixed-design nonparametric time series regression by Robinson (1997) (where, incidentally, the kind of trichotomy observed in parts (i)-(iii) of Theorem 4 in our stochastic design setting does not occur). Assumption B6 is a very mild additional decay restriction on the  $\gamma_{ij}$ .

3. While overall our conditions on  $\{V_i\}$  and  $\{X_i\}$  are neither stronger nor weaker than those employed under a mixing framework, in some respects ours provide a more precise tool. Theorem 4 indicates exactly when the usual kind of result with  $(nh^d)^{1/2}$  convergence ceases to apply, whereas a mixing rate can usually only be interpreted as sufficient. On the other hand the linearity of  $V_i$  plays an important role in the achievement of asymptotic normality throughout Theorem 4, despite possible long range dependence. Extending

the nonlinear-functions-of-Gaussian-processes conditions employed in the long range dependent time series literature would sometimes yield non-normal limits, especially given our allowance for strong dependence in  $X_i$ .

4. It is important to stress that the question of which of parts (i)-(iii) of Theorem 4 is relevant depends on  $h$ , as well as the strength of dependence in  $V_i$ . More precisely, under the weak dependence condition  $\sum_{i,j=1,i \neq j}^n \gamma_{ij} = O(n)$ , part (i) is relevant however we choose  $h$ , subject to Assumptions A4 and B3, but not otherwise. Putting the conditions together, (5.29) occurs when

$$\frac{1}{nh^d} + \frac{h^d \sum_{i,j=1,i \neq j}^n \gamma_{ij}}{n} + nh^{d+2\theta} \rightarrow 0, \text{ as } n \rightarrow \infty; \quad (6.3)$$

(5.30) occurs when

$$h \sim \left( \frac{\frac{\chi n}{\sum_{i,j=1,i \neq j}^n \gamma_{ij}}}{\sum_{i,j=1,i \neq j}^n \gamma_{ij}} \right)^{1/d}, \quad \chi > 0; \quad \frac{\sum_{i,j=1,i \neq j}^n \gamma_{ij}}{n^2} + \frac{n^{2(1+\theta/d)}}{\sum_{i,j=1,i \neq j}^n \gamma_{ij}} \rightarrow 0, \text{ as } n \rightarrow \infty; \quad (6.4)$$

and (5.31) occurs when

$$\frac{\sum_{i,j=1,i \neq j}^n \gamma_{ij}}{n^2} + \frac{n}{h^d \sum_{i,j=1,i \neq j}^n \gamma_{ij}} + \frac{n^2 h^{2\theta}}{\sum_{i,j=1,i \neq j}^n \gamma_{ij}} \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (6.5)$$

These conditions also reflect the dimension  $d$  of  $X_i$ , and the curse of dimensionality is always of concern with smoothed nonparametric estimation; Gao, Lu and Tjøstheim (2006), Lu, Lundervold, Tjøstheim and Yao (2007) consider semiparametric and additive models, respectively, in different spatial settings from ours.

5. Our results do not directly address the issue of bandwidth choice, which is always of practical concern in smoothed nonparametric estimation, though they have some implications for it. By adding a bias calculation under twice differentiability of  $g$  to the variance implications of part (i) of Theorem 4 we can reproduce the usual minimum-mean-squared error optimal rate for  $h$  of  $n^{4/5}$ . As our results stand, we do not exploit this degree of smoothness (see Assumptions B2 and B3), and to do so would require a stronger restriction on the dependence in  $X_i$ , similar to that mentioned in the penultimate sentence of point 1 above.

However the condition  $t \sim \chi s$  for part (ii) of Theorem 4 prescribes a rate for  $h$  which ignores bias, while the part (iii) convergence rate does not directly depend on  $h$ , so cannot contribute to an optimal bandwidth calculation.

6. Part (i) of Theorem 4 reproduces the classical asymptotic independence across distinct fixed points familiar from the settings of independent observations and mixing time series. Since  $t = o(s)$  entails  $\sum_{i,j=1}^n \gamma_{ij} = o(n/h)$ , the result nevertheless also holds under some degree of long range dependence in  $V_i$ , while, at least in the Gaussian case, strong versions of long range dependence in  $X_i$  are permitted. The allowance for non-stationarity in both processes leads to a more complicated form of asymptotic variance, the  $i$ th diagonal element of  $\kappa \Phi^{-1} \Lambda \Phi^{-1}$  reducing to the familiar  $\kappa \sigma^2(x_i)/f(x_i)$  under identity of distribution. To carry out inference, such as set confidence regions, we need to consistently estimate the diagonal elements; the extent to which this possible in our more general setting would require discussion that goes beyond the scope of the present paper.

7. Consistent variance estimation is an even more challenging proposition in (ii) and (iii) of Theorem 4, in part due to the difficulty with estimating  $t$ . This in turn stems in part from the possible non-stationarity of  $V_i$ , and estimating  $t$  would require imposing some additional structure to limit this. It also stems from the implied long range dependence in  $V_i$  in both parts (ii) and (iii); note that  $t$  is analogous to quantities arising in the "HAC" literature of econometrics, which extends earlier statistical literature (see e.g. Hanman, 1956, Eicker, 1963), but there weak dependence is typically assumed, in which case we are back to part (i). However, at least for stationary  $V_i$ , proposals of Robinson (2005) in the long range dependent time series case may be extendable. The results in parts (ii) and (iii) are much less attractive for practical use due to the non-diagonality of  $\Psi$ , and even less so than immediately meets the eye. Notice that if the  $X_i$  are iid,  $\Psi$  has unit rank for all  $p$ , so estimates are undesirably perfectly correlated. The same kind of outcome was observed by Robinson (1991) in kernel probability density estimates from long range dependent time series data. Unit rank could result more generally: under similar conditions on  $X_i$  to those in Assumption B10 the numerator of the left side of (5.26) differs from

$$\sum_{i,j=1, i \neq j}^n \gamma_{ij} \sigma_i(x) \sigma_j(y) f_i(x) f_j(y) \quad (6.6)$$

by  $o\left(\sum_{i,j=1}^n \gamma_{ij}\right)$ . Then  $\Psi$  has unit rank for all  $p$  if the  $X_i$  are identically distributed, and is possibly not well conditioned more generally. Nevertheless it is of interest that here non-identity of distribution has the potential to desirably increase rank.

8. Consider some implications of our setting for data observed on a rectangular lattice. For simplicity we focus on a 2-dimensional lattice, where observations are recorded at points  $(t_1, t_2)$ , for  $t_1 = 1, \dots, n_1$ ,  $t_2 = 1, \dots, n_2$ , so  $n = n_1 n_2$

(though interval of observation can differ across dimensions). We can regard either or both of  $n_1, n_2$  as increasing with  $n$ . The  $(t_1, t_2)$ th observation can be indexed by  $i = n_2(t_1 - 1) + t_2$ , say, in our triangular array setting. With this correspondence, consider a process, such that  $V_{in} = v(t_1, t_2)$ . We might define  $v(t_1, t_2)$  for  $t_1, t_2 = 0, \pm 1, \dots$ , and take it to be stationary. Then under broad conditions,  $v(t_1, t_2)$  has a "half-plane" linear representation in terms of orthogonal, homoscedastic innovations, analogous to the Wold representation for time series (see e.g. Whittle, 1954). However, for Theorems 2 and 4, we require a linear representation for  $V_{in}$  in independent innovations, so for non-Gaussian  $v(t_1, t_2)$  a general, multilateral representation, would be preferred, namely

$$v(t_1, t_2) = \sum_{j_1, j_2 = -\infty}^{\infty} b(t_1 - j_1, t_2 - j_2) e(j_1, j_2) \quad (6.7)$$

with independent  $e(j_1, j_2)$  and square-summable  $b(j_1, j_2)$ . To produce a correspondence with (2.7) we might read off the  $j_1, j_2$  in a kind of spiral fashion: taking  $j = 1$  when  $(j_1, j_2) = (0, 0)$ , then  $j = 2, \dots, 9$ , correspond to the points  $(-1, -1), (-1, 0), \dots$ , going clockwise around the square with vertices  $(\pm 1, \pm 1)$ , with  $j = 10, \dots, 25$  generated analogously from the square with vertices  $(\pm 2, \pm 2)$ , and so on. If a "half-plane" representation is desired we simply omit the relevant points on each square. A correspondence between the  $a_{ijn}$  and moving average weights  $b(t_1 - j_1, t_2 - j_2)$  then follows. Now the  $b(t_1 - j_1, t_2 - j_2)$  might be chosen to be the moving average weights in unilateral or multilateral spatial analogues of autoregressive moving averages (see e.g. Whittle, 1954, Guyon, 1982, Robinson and Vidal-Sanz, 2006). These models have the weak dependence to place them firmly in the setting of part (i) of Theorem 4. But (6.7) can also describe long range dependence, in either or both dimensions, so parts (ii) and (iii) of Theorem 4 can also be relevant. Notice that  $\sum_{i, j=1, i \neq j}^n \gamma_{ij}$   
 $= \sum_{j_i=1-n_i, i=1,2}^{n_i-1} Cov \{v(0, 0), v(j_1, j_2)\}$ . Tran (1990), Hallin, Lu, and Tran (2001, 2004b), for example, established consistency results for kernel density estimates with lattice spatial data, under different conditions from those in Theorem 3.

9. Data that are irregularly-observed in space and/or time pose far greater problems in both computation and deriving asymptotic theory for many statistical procedures. Describing a model for irregularly-spaced observations from an underlying continuous model can be difficult even in the time series case; from a first-order stochastic differential equation, Robinson (1977) derived a (time-varying autoregressive) model but this kind of outcome is difficult to extend to higher-order models, or spatial processes. Nonparametric regression, on the other hand, is readily applied, though detailed checking of our conditions would be far more difficult than in the regularly-spaced circumstances just described, especially as observation locations might be regarded as stochastically rather than deterministically generated. However, at least some formal correspondence with our triangular array setting can be constructed. Suppose we

have an underlying Gaussian process, where again for simplicity we consider 2 dimensions only, and denote it  $v(t_1, t_2)$ , though now  $t_1$  and  $t_2$  might be real-valued. Then with some ordering the  $n$   $v(t_1, t_2)$  become the  $n$   $V_i$ . Moreover, (conditionally on the locations when these are stochastic) the  $V_i$  are Gaussian. Denote by  $\Xi$  the covariance matrix of the vector  $V = (V_1, \dots, V_n)'$ . In a general irregularly-spaced framework  $\Xi$  has little structure to exploit. However, we can write  $\Xi = AA'$ , where, for positive definite  $\Xi$  the  $n \times n$  matrix  $A$  is uniquely defined only up to premultiplication by an orthogonal matrix. Due to Gaussianity we can write  $V = A\eta$ , where  $\eta \sim N(0, I_n)$ . We deduce (2.7), indeed (2.9), by taking  $a_{ij}$  to be the  $(i, j)$ th element of  $A$  and  $\varepsilon = \eta$ .

10. Processes in the SAR class (2.10) are more directly placed in our framework. Consider the special case  $m_1 = 1$ ,  $m_2 = 0$  of (2.10), i.e.

$$(I_n - \omega W)U = \sigma\varepsilon, \quad 0 < |\omega| < 1, \quad (6.8)$$

with nonstochastic  $n \times n$   $W = W_n$  having row sums normalized to 1. Note that (6.8) generally implies unconditional heteroscedasticity in the  $U_i$ , as can be covered by our  $\sigma_i(X_i) = \sigma_i$ . As noted by Lee (2002), it follows that  $S = S_n = I_n - \omega W$  is non-singular, and thus we have a solution to (6.8) of form  $U_i = \sum_{j=1}^n b_{ij}\varepsilon_j$ , with  $Var\{U_i\} = \sum_{j=1}^n b_{ij}^2 = \sigma_i^2$ . Thus on taking  $a_{ij} = b_{ij}/\sigma_i$  we have (2.9) with (2.8). Moreover, because  $\sum_{i,j=1}^n \gamma_{ij} = \sigma^2 \mathbf{1}' S^{-1} S^{-1} \mathbf{1}$ , where  $\mathbf{1} = \mathbf{1}_n$  is the  $n \times 1$  vector of 1's, it follows that if  $S^{-1}$  has uniformly bounded column sums (for which a condition is given in Lemma 1 of Lee, 2002), then  $\sum_{i,j=1}^n \gamma_{ij} = O(n)$ , so "weak dependence" is implied, and part (i) of Theorem 4 applies.

## 7 Appendix A: Technical Lemmas for Section 4

**Lemma 1:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A4, A5( $x$ ), and A6( $x, x$ ) hold. Then as  $n \rightarrow \infty$ , for some  $\varepsilon > 0$*

$$Var\{\hat{q}(x)\} = O((nh^d)^{-1}\alpha(x; \varepsilon) + n^{-1}h^{2(\xi-d)}\rho(x, x; \varepsilon) + h^{\xi-2d}) \rightarrow 0. \quad (A.1)$$

**Proof.** We have

$$Var\{\hat{q}(x)\} = \frac{1}{(nh^d)^2} \left[ \sum_{i=1}^n Var\{K_i(x)\} + \sum_{i,j=1, i \neq j}^n Cov\{K_i(x), K_j(x)\} \right]. \quad (A.2)$$

The first term in the square brackets is bounded by



$$\begin{aligned}
n \int_{\mathbb{R}^d} K^2 \left( \frac{x-w}{h} \right) \bar{f}(w) dw &= nh^d \int_{\mathbb{R}^d} K^2(u) \bar{f}(x-hu) du \\
&= nh^d \left\{ \int_{\|hu\| < \varepsilon} K^2(u) \bar{f}(x-hu) du \right. \\
&\quad \left. + \int_{\|hu\| \geq \varepsilon} K^2(u) \bar{f}(x-hu) du \right\}, \quad (\text{A.3})
\end{aligned}$$

for any  $\varepsilon > 0$ . The first term in the braces is bounded by  $\kappa\alpha(x; \varepsilon)$ , where, throughout,  $C$  denotes an arbitrarily large generic constant. The second term in the braces is bounded by

$$\sup_{\|u\| \geq \varepsilon/h} K(u)^2 \int_{\mathbb{R}^d} \bar{f}(x-hu) du = O(h^{2\xi-d}), \quad (\text{A.4})$$

so

$$\frac{1}{(nh^d)^2} \sum_{i=1}^n \text{Var}\{K_i(x)\} = O((nh^d)^{-1}\alpha(x; \varepsilon) + n^{-1}h^{2(\xi-d)}). \quad (\text{A.5})$$

The second term in the square brackets in (A.2) is

$$\begin{aligned}
&(nh^d)^2 \left\{ \int_{J_1(\varepsilon)} K(u) K(v) m(x-hu, x-hv) dudv + 2 \int_{J_2(\varepsilon)} K(u) K(v) m(x-hu, x-hv) dudv \right. \\
&\quad \left. + \int_{J_3(\varepsilon)} K(u) K(v) m(x-hu, x-hv) dudv \right\}, \quad (\text{A.6})
\end{aligned}$$

where  $J_1(\varepsilon) = J_{1n}(\varepsilon) = \{u, v : \|hu\| < \varepsilon, \|hv\| < \varepsilon\}$ ,  $J_2(\varepsilon) = J_{2n}(\varepsilon) = \{u, v : \|hu\| < \varepsilon, \|hv\| \geq \varepsilon\}$ ,  $J_3(\varepsilon) = J_{3n}(\varepsilon) = \{u, v : \|hu\| \geq \varepsilon, \|hv\| \geq \varepsilon\}$ . The first integral in the braces is bounded by

$$\rho(x, x; \varepsilon) \left\{ \int_{\mathbb{R}^d} |K(u)| du \right\}^2. \quad (\text{A.7})$$

The second integral is bounded by

$$\begin{aligned}
&2 \int_{u \in \mathbb{R}^d, \|hv\| \geq \varepsilon} |K(u) K(v)| |m(x-hu, x-hv)| dudv \\
&\leq 2 \sup_{u \in \mathbb{R}^d} |K(u)| \sup_{\|v\| \geq \varepsilon/h} |K(v)| \\
&\quad \times \int_{\mathbb{R}^{2d}} \frac{1}{n^2} \sum_{i,j=1, i \neq j}^n \{f_{ij}(x-hu, x-hv) + f_i(x-hu)f_j(x-hv)\} dudv \\
&= O(h^{\xi-2d}). \quad (\text{A.8})
\end{aligned}$$

The third integral is bounded by

$$\begin{aligned}
& \int_{\|hu\| \geq \varepsilon, \|hv\| \geq \varepsilon} |K(u)K(v)| |m(x-hu, x-hv)| dudv \\
\leq & \sup_{\|u\| \geq \varepsilon/h} K(u)^2 \\
& \times \int_{\mathbb{R}^{2d}} \frac{1}{n^2} \sum_{i,j=1, i \neq j}^n \{f_{ij}(x-hu, x-hv) + f_i(x-hu)f_j(x-hv)\} dudv \\
= & O(h^{2\xi-2d}). \tag{A.9}
\end{aligned}$$

Thus

$$\frac{1}{(nh^d)^2} \sum_{i,j=1, i \neq j}^n \text{Cov}\{K_i(x), K_j(x)\} = O(\rho(x, x; \varepsilon) + h^{\xi-2d}). \tag{A.10}$$

□

**Lemma 2:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > d$ , A3, A4 and A7( $x$ ) hold. Then*

$$\varliminf_{n \rightarrow \infty} E\{\widehat{f}(x)\} > 0. \tag{A.11}$$

**Proof:** We have

$$\begin{aligned}
E\{\widehat{f}(x)\} & \geq \int_{\|u\| < \varepsilon/h} K(u) \bar{f}(x-hu) du - \left| \int_{\|u\| \geq \varepsilon/h} K(u) \bar{f}(x-hu) du \right|, \\
& \geq \inf_{\|u\| < \varepsilon} \bar{f}(x-u) \int_{\|u\| < \varepsilon/h} K(u) du - \sup_{\|u\| \geq \varepsilon/h} |K(u)|/h^d \\
& \geq \frac{1}{2} \inf_{\|u\| < \varepsilon} \bar{f}(x-u) - O(h^{\xi-d}) > \zeta \tag{A.12}
\end{aligned}$$

for  $n$  sufficiently large and some  $\zeta > 0$ , by Assumption A7( $x$ ). □

**Lemma 3:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > d$ , A4, A5( $x$ ), and A8 hold. Then*

$$E|\widehat{r}_2(x)| \rightarrow 0, \text{ as } n \rightarrow \infty. \tag{A.13}$$

**Proof:** We have

$$\begin{aligned}
E \left| \sum_{i=1}^n \{g(X_i) - g(x)\} K_i(x) \right| &\leq nh^d \int_{\mathbb{R}^d} |K(u)| |g(x-hu) - g(x)| \bar{f}(x-hu) du \\
&\leq nh^d \sup_{\|u\| \leq \varepsilon} |g(x-u) - g(x)| \sup_{\|u\| \leq \varepsilon} \bar{f}(x-u) \int_{\mathbb{R}^d} |K(u)| du \\
&\quad + nh^d \sup_{\|u\| \geq \varepsilon/h} |K(u)| \left\{ \int_{\mathbb{R}^d} |g(x-hu)| \bar{f}(x-hu) du \right. \\
&\quad \left. + |g(x)| \int_{\mathbb{R}^d} \bar{f}(x-hu) du \right\} \\
&\leq \delta nh^d + Cnh^\xi \left\{ \frac{1}{n} \sum_{i=1}^n E |g(X_i)| + |g(x)| \right\} \tag{A.14}
\end{aligned}$$

for any  $\delta > 0$ , to complete the proof.  $\square$

**Lemma 4:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A4, A9, and A10 hold. Then*

$$E \{ \hat{r}_1(x)^2 \} \rightarrow 0, \text{ as } n \rightarrow \infty. \tag{A.15}$$

**Proof:** The left side of (A.15) is

$$\frac{1}{(nh^d)^2} \left[ \sum_{i=1}^n E \{ \sigma_i^2(X_i) K_i^2(x) \} + \sum_{i,j=1, i \neq j}^n \gamma_{ij} E \{ \sigma_i(X_i) \sigma_j(X_j) K_i(x) K_j(x) \} \right]. \tag{A.16}$$

recalling that  $\gamma_{ii} = \text{Var} \{ V_i \} = 1$ . The first expectation is

$$h^d \left\{ \int_{\|hu\| < \varepsilon} K^2(u) \mu_i(x-hu) du + \int_{\|hu\| \geq \varepsilon} K^2(u) \mu_i(x-hu) du \right\}. \tag{A.17}$$

The first term in braces is bounded by

$$C \sup_{\|u\| < \varepsilon} \mu_i(x-u). \tag{A.18}$$

The second term is bounded by

$$\sup_{\|u\| \geq \varepsilon/h} K(u)^2 \int_{\mathbb{R}^d} \mu_i(x-hu) du \leq Ch^{2\xi-d} E \sigma_i^2(X_i). \tag{A.19}$$

The second expectation in (A.16) is

$$\begin{aligned}
&\int_{\mathbb{R}^{2d}} K\left(\frac{x-w}{h}\right) K\left(\frac{x-z}{h}\right) \nu_{ij}(w,z) dw dz \\
&= h^{2d} \left\{ \int_{J_1(\varepsilon)} + 2 \int_{J_2(\varepsilon)} + \int_{J_3(\varepsilon)} \right\} K(u) K(v) \nu_{ij}(x-hu, x-hv) dudv. \tag{A.20}
\end{aligned}$$

Proceeding as in the proof of Lemma 1, this is bounded by

$$\begin{aligned}
& Ch^{2d} \sup_{\|u\| < \varepsilon, \|v\| < \varepsilon} |\nu_{ij}(x-u, x-v)| \\
& + h^{2d} \left\{ 2 \sup_{\|u\| \geq \varepsilon/h} |K(u)| \sup_{v \in \mathbb{R}^d} |K(v)| + \left( \sup_{\|u\| \geq \varepsilon/h} |K(u)| \right)^2 \right\} \\
& \times \int_{\mathbb{R}^{2d}} |\nu_{ij}(x-h_n u, x-h_n v)| dudv, \tag{A.21}
\end{aligned}$$

whose last term is bounded by  $Ch^\xi E|\sigma_i(X_i)\sigma_j(X_j)| \leq Ch^\xi E\{\sigma_i^2(X_i)\}^{1/2} E\{\sigma_j^2(X_j)\}^{1/2} = O(h^\xi)$  uniformly in  $i, j$ . Thus (A.16) is

$$O\left(\frac{1}{(nh^d)^2} \left\{ n(h^d + h^{2\xi-d}) + (h^{2d} + h^\xi) \sum_{i,j=1, i \neq j}^n |\gamma_{ij}| \right\}\right) = O\left(\frac{1}{nh^d} + \frac{1}{n^2} \sum_{i,j=1, i \neq j}^n |\gamma_{ij}| \right) \rightarrow 0. \tag{A.22}$$

□

Introduce

$$\begin{aligned}
V'_i &= V'_{in} = \sum_{j=1}^{\infty} a_{ij} \varepsilon'_j, \quad U'_i = U'_{in} = \sigma_i(X_i) \{V'_i - E(V'_i)\}, \\
V''_i &= V''_{in} = \sum_{j=1}^{\infty} a_{ij} \varepsilon''_j, \quad U''_i = U''_{in} = \sigma_i(X_i) \{V''_i - E(V''_i)\}. \tag{A.23}
\end{aligned}$$

**Lemma 5:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A4, A9( $x, x$ ) and A11 hold. Then*

$$E|\hat{r}_1(x)| \rightarrow 0, \text{ as } n \rightarrow \infty. \tag{A.24}$$

**Proof:**

$$\frac{1}{nh^d} \sum_{i=1}^n U'_i K_i(x) = \frac{1}{nh^d} \sum_{i=1}^n \sigma_i(X_i) K_i(x) \sum_{j=1}^{\infty} a_{ij} (\varepsilon'_j - E\{\varepsilon'_j\}) \tag{A.25}$$

has mean zero and variance

$$\begin{aligned}
& \frac{1}{(nh^d)^2} \sum_{i=1}^n E\{\sigma_i^2(X_i) K_i^2(x)\} \sum_{k=1}^{\infty} a_{ik}^2 \text{Var}\{\varepsilon'_k\} \\
& + \frac{1}{(nh^d)^2} \sum_{i,j=1, i \neq j}^n E\{\sigma_i(X_i) \sigma_j(X_j) K_i(x) K_j(x)\} \\
& \times \sum_{k=1}^{\infty} a_{ik} a_{jk} \text{Var}\{\varepsilon'_k\}. \tag{A.26}
\end{aligned}$$

From the proof of Lemma 4, this is bounded by

$$\frac{C}{nh^d} \sum_{k=1}^{\infty} a_{ik}^2 + \frac{C}{n^2 h^d} \sum_{i,j=1}^n \sum_{k=1}^{\infty} |a_{ik}| |a_{jk}| \leq \frac{C}{nh^d} (1 + \max_{1 \leq i \leq n} \sum_{k=1}^{\infty} |a_{ik}| \max_{k \geq 1} \sum_{j=1}^n |a_{jk}|) \rightarrow 0 \quad (\text{A.27})$$

as  $n \rightarrow \infty$ . On the other hand

$$E \left| \frac{1}{nh^d} \sum_{i=1}^n U_i'' K_i(x) \right| \leq \frac{2}{nh^d} \sum_{i=1}^n E |\sigma_i(X_i) K_i(x)| \sum_{j=1}^{\infty} |a_{ij}| E |\varepsilon_j''|. \quad (\text{A.28})$$

In a similar way as before,  $E |\sigma_i(X_i) K_i(x)| \leq Ch^d$ , whence (A.28) is bounded by

$$C \max_{n \geq 1} \max_{1 \leq j \leq n} E |\varepsilon_j''| \max_{i \geq 1} \sum_{j=1}^{\infty} |a_{ij}| \rightarrow 0 \quad (\text{A.29})$$

as  $D \rightarrow \infty$ .  $\square$

## 8 Appendix B: Technical Lemmas for Section 5

**Lemma 6:** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > d$ , A4 and B1( $x$ ) hold. Then for all  $\delta > 0$  there exists  $\varepsilon > 0$  such that*

$$\left| E\{\widehat{f}(x)\} - \bar{f}(x) \right| \leq C(\beta(x; \varepsilon) + h^{\xi-d}) < \delta, \quad (\text{B.1})$$

for all sufficiently large  $n$ .

**Proof:** We have

$$\begin{aligned} E\{\widehat{f}(x)\} - \bar{f}(x) &= \int_{\mathbb{R}^d} K(u) \{\bar{f}(x - hu) - \bar{f}(x)\} du \\ &= \int_{\|hu\| < \varepsilon} K(u) \{\bar{f}(x - hu) - \bar{f}(x)\} du \\ &\quad + \int_{\|hu\| \geq \varepsilon} K(u) \{\bar{f}(x - hu) - \bar{f}(x)\} du, \end{aligned} \quad (\text{B.2})$$

for  $\varepsilon > 0$ . The first term is bounded by

$$\beta(x; \varepsilon) \int_{\mathbb{R}^d} |K(u)| du. \quad (\text{B.3})$$

The second term is bounded by

$$\sup_{\|u\| \geq \varepsilon/h} |K(u)| \int_{\mathbb{R}^d} \bar{f}(x - hu) du + \bar{f}(x) \int_{\|u\| \geq \varepsilon/h} |K(u)| du = O(h^{\xi-d}). \quad (\text{B.4})$$

□

**Lemma 7:** Let Assumptions A1, A2( $\xi$ ) for  $\xi > d + \theta$ , A4, A5( $x$ ), A8 and B2( $x$ ) hold. Then

$$E \|\widehat{r}_2(x)\| = O(h_n^\theta), \text{ as } n \rightarrow \infty. \quad (\text{B.5})$$

**Proof:** This is very similar to the proof of Lemma 3, and thus omitted. □

**Lemma 8** Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A4, A9( $x, y$ ), B4, B8( $x, y$ ), B10 and B11 hold. Then

$$\begin{aligned} \text{Cov}\{\widehat{r}_1(x), \widehat{r}_1(y)\} &\sim \kappa\lambda(x)s, \text{ if } x = y \text{ and } t = o(s), \\ &= o(s), \text{ if } x \neq y \text{ and } t_n = o(s), \\ &\sim \psi(x, y)t, \text{ if } s = o(t), \\ &\sim \{\kappa\lambda(x) + \chi\psi(x, y)\}s, \text{ if } t \sim \chi s, \chi \in (0, \infty). \end{aligned} \quad (\text{B.6})$$

**Proof:** Proceeding as in the proof of Lemma 4,  $\text{Cov}\{\widehat{r}_1(x), \widehat{r}_1(y)\}$  is

$$\frac{1}{(nh^d)^2} \sum_{i=1}^n E\{\sigma_i^2(X_i)K_i(x)K_i(y)\} \quad (\text{B.7})$$

$$+ \frac{1}{(nh^d)^2} \sum_{i,j=1, i \neq j}^n \gamma_{ij} E\{\sigma_i(X_i)\sigma_j(X_j)K_i(x)K_j(y)\}. \quad (\text{B.8})$$

When  $x = y$ , from (A.17) (B.7) equals

$$\frac{h^{-d}}{n^2} \sum_{i=1}^n \int_{\mathbb{R}^d} K^2(u) \mu_i(x - hu) du. \quad (\text{B.9})$$

The difference between the integral and  $\kappa\sigma_i^2(x)f_i(x)$  is

$$\begin{aligned} &\int_{\mathbb{R}^d} K^2(u) \{\sigma_i^2(x - hu) - \sigma_i^2(x)\} f_i(x - hu) du \\ &+ \sigma_i^2(x) \int_{\mathbb{R}^d} K^2(u) \{f_i(x - hu) - f_i(x)\} du. \end{aligned} \quad (\text{B.10})$$

The first term is bounded by

$$\begin{aligned} &\kappa \max_{1 \leq i \leq n} \sup_{\|u\| < \varepsilon} |\sigma_i^2(x - u) - \sigma_i^2(x)| \sup_{\|u\| < \varepsilon} f_i(x - hu) \\ &+ \max_{1 \leq i \leq n} \sup_{\|hu\| \geq \varepsilon} K^2(u) \int_{\mathbb{R}^d} \{\sigma_i^2(x - hu) + \sigma_i^2(x)\} f_i(x - hu) du \\ &= o(1) + O(h^{2\xi - d}), \end{aligned} \quad (\text{B.11})$$

uniformly in  $i$ , proceeding as in the proof of Lemma 4. The second term is bounded by

$$\begin{aligned}
& \kappa \max_{1 \leq i \leq n} \sigma_i^2(x) \sup_{\|hu\| < \varepsilon} |f_i(x - hu) - f_i(x)| \\
& + \max_{1 \leq i \leq n} \sigma_i^2(x) \sup_{\|u\| \geq \varepsilon/h} K^2(u) \int_{\mathbb{R}^d} f_i(x - hu) du \\
& + \max_{1 \leq i \leq n} f_i(x) \int_{\|u\| \geq \varepsilon/h} K^2(u) du \\
& = o(1) + O(h^{2\xi-d}). \tag{B.12}
\end{aligned}$$

uniformly in  $i$ . Thus, when  $x = y$ , there is a sequence  $\Delta = \Delta_n \rightarrow 0$  such that

$$(B.7) \sim \kappa \frac{h^{-d}}{n^2} \sum_{i=1}^n \{\mu_i(x) + \Delta\} \sim \kappa \lambda(x) s. \tag{B.13}$$

When  $x \neq y$ , (B.7) equals

$$\frac{h^{-d}}{n^2} \sum_{i=1}^n \int_{\mathbb{R}^d} K(u) K\left(u + \frac{y-x}{h}\right) \mu_i(x - hu) du. \tag{B.14}$$

The difference between the integral and

$$\mu_i(x) \int_{\mathbb{R}^d} K(u) K\left(u + \frac{y-x}{h}\right) du \tag{B.15}$$

is, by essentially the same proof,  $o(1) + O(h^{2\xi-d})$  uniformly in  $i$ . Splitting the range of integration into the sets  $\|u\| > \|x-y\|/2h$  and  $\|u\| \leq \|x-y\|/2h$ , and noting that the latter inequality implies that  $\|u + (y-x)/h\| \geq \|x-y\|/2h$ , the integral in (B.15) is bounded by

$$2 \left\{ \sup_{\|u\| > \|x-y\|/2h} |K(u)| \right\} \int_{\mathbb{R}^d} |K(u)| du \leq Ch^{\xi-d}. \tag{B.16}$$

Thus when  $x \neq y$ , for  $\Delta$  as before,

$$(B.7) = \frac{h^{-d}}{n^2} \sum_{i=1}^n \mu_i(x) \Delta = o(s), \tag{B.17}$$

As in (A.20), (B.8) is

$$\frac{1}{n^2} \sum_{i,j=1, i \neq j}^n \gamma_{ij} \int_{\mathbb{R}^{2d}} K(u) K(v) \nu_{ij}(x - hu, y - hv) dudv. \tag{B.18}$$

Now  $\nu_{ij}(x - u, y - v) - \nu_{ij}(x, y)$  can be written

$$\begin{aligned}
& \{\sigma_i(x - u) - \sigma_i(x)\} \sigma_j(y - v) f_{ij}(x - u, y - v) + \sigma_i(x) \{\sigma_j(y - v) - \sigma_j(y)\} f_{ij}(x - u, y - v) \\
& + \sigma_i(x) \sigma_j(y) \{f_{ij}(x - u, y - v) - f_{ij}(x, y)\}. \tag{B.19}
\end{aligned}$$

By proceeding much as before with each of these three terms, it may thus be seen that

$$(B.18) \sim \frac{1}{n^2} \sum_{i,j=1, i \neq j}^n \gamma_{ij} \{\nu_{ij}(x, y) + \Delta\} \sim \psi(x, y)t. \quad (B.20)$$

□

**Lemma 9** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 2d$ , A4, A9( $x_i, x_j$ ), B4, B8( $x_i, x_j$ ),  $i, j = 1, \dots, p$ , B10 and B11 hold. Then there exists a sequence  $N = N_n$ , increasing with  $n$ , such that*

$$E \left\| \widehat{r}_1^\# \right\|^2 = o(w), \text{ as } n \rightarrow \infty. \quad (B.21)$$

**Proof:** We have

$$E \left\| \widehat{r}_1^\# \right\|^2 = \frac{1}{(nh^d)^2} \sum_{i,j=1}^n E \{K_i' K_j \sigma_i(X_i) \sigma_j(X_j)\} \sum_{k=N+1}^{\infty} a_{ik} a_{jk} \quad (B.22)$$

From the proof of Lemma 8 the expectation is

$$\kappa h^d \left\{ \sum_{k=1}^p \mu_i(x_k) \right\} 1(i=j)(1+o(1)) + \text{tr}(D_{ij}) 1(i \neq j)(1+o(1)) \quad (B.23)$$

uniformly in  $i, j$ , where  $1(\cdot)$  denotes the indicator function and  $D_{ij}$  is the  $p \times p$  matrix with  $(k, l)$ th element  $h^{2d} \nu_{ij}(x_k, x_l)$ . Thus for large enough  $n$  (B.22) is bounded by

$$\frac{Ch^{2d}}{(nh^d)^2} \sum_{i,j=1}^n \left| \sum_{k=N+1}^{\infty} a_{ik} a_{jk} \right|. \quad (B.24)$$

First suppose  $t = O(s)$ . By the Cauchy inequality (B.24) is bounded by

$$\frac{C}{n^2} \left\{ \sum_{i=1}^n \left( \sum_{k=N+1}^{\infty} a_{ik}^2 \right)^{1/2} \right\}^2. \quad (B.24)$$

In view of (2.8),

$$\lim_{N \rightarrow \infty} \max_{1 \leq i \leq n} \sum_{k=N+1}^{\infty} a_{ik}^2 \rightarrow 0. \quad (B.25)$$

Thus  $E \left\| \widehat{r}_1^\# \right\|^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

Now suppose  $s = o(t)$ . We have

$$\sum_{i,j=1, i \neq j}^n \left| \sum_{k=N+1}^{\infty} a_{ik} a_{jk} \right| \leq \sum_{k=N+1}^{\infty} \sum_{i,j=1, i \neq j}^n |a_{ik} a_{jk}|. \quad (B.26)$$



Now

$$\sum_{k=1}^{\infty} b_k \equiv 1, \quad (\text{B.27})$$

where

$$b_k = b_{kn} = \sum_{i,j=1, i \neq j}^n |a_{ik}a_{jk}| / \sum_{k=1}^{\infty} \sum_{i,j=1, i \neq j}^n |a_{ik}a_{jk}|. \quad (\text{B.28})$$

Thus

$$\sum_{k=N+1}^{\infty} b_k \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (\text{B.29})$$

and from (5.26),

$$E \left\| \hat{r}_1^{\#} \right\|^2 = \frac{h^{2d}}{(nh^d)^2} \sum_{i,j=1}^n \sum_{k=N+1}^{\infty} |a_{ik}a_{jk}| \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (\text{B.30})$$

□

**Lemma 10** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 4d$ , A4, A9( $x_i, x_j$ ), B4, B6, B7, B9( $x_i$ ) and B10 hold,  $i, j = 1, \dots, p$ . Then*

$$E \|T - E\{T\}\|^2 = o(w^2), \text{ as } n \rightarrow \infty. \quad (\text{B.31})$$

**Proof:** It suffices to check (B.31) in case  $p = 1$ , so we put  $x_1 = x$ . We have

$$E \{T - E\{T\}\}^2 = \sum_{j,k=1}^N [E \{Z_j^2 Z_k^2\} - E \{Z_j^2\} E \{Z_k^2\}]. \quad (\text{B.32})$$

The summand is

$$\begin{aligned} & \frac{1}{(nh^d)^4} E \left\{ \left( \sum_{i=1}^n K_i(x) \sigma_i(X_i) a_{ij} \sum_{i=1}^n K_i(x) \sigma_i(X_i) a_{ik} \right)^2 \right\} \\ & - E \left\{ \left( \sum_{i=1}^n K_i(x) \sigma_i(X_i) a_{ij} \right)^2 \right\} E \left\{ \left( \sum_{i=1}^n K_i(x) \sigma_i(X_i) a_{ik} \right)^2 \right\} \\ & = \frac{1}{(nh^d)^4} \sum_{\substack{i_s=1 \\ s=1, \dots, 4}}^n a_{i_1 j} a_{i_2 j} a_{i_3 k} a_{i_4 k} \left[ E \left\{ \prod_{s=1}^4 K_{i_s}(x) \sigma_{i_s}(X_{i_s}) \right\} \right. \\ & \quad \left. - E \left\{ \prod_{s=1}^2 K_{i_s}(x) \sigma_{i_s}(X_{i_s}) \right\} E \left\{ \prod_{s=3}^4 K_{i_s}(x) \sigma_{i_s}(X_{i_s}) \right\} \right]. \quad (\text{B.33}) \end{aligned}$$

The quadruple sum yields terms of seven kinds, depending on the nature of equalities, if any, between the  $i_s$ , and bearing in mind the fact that  $i_1, i_2$  are linked with  $j$ , and  $i_3, i_4$  are linked with  $k$ . Symbolically, denote such a term

$\langle A, B, C, D \rangle - \langle A, B \rangle \langle C, D \rangle$ , when all  $i_s$  are unequal, and repeat the corresponding letters in case of any equalities. The other six kinds of term are thus

$$\begin{aligned} & \langle A, A, A, B \rangle - \langle A, A \rangle \langle A, B \rangle, \quad \langle A, A, A, A \rangle - \langle A, A \rangle \langle A, A \rangle. \\ & \langle A, A, B, C \rangle - \langle A, A \rangle \langle B, C \rangle, \quad \langle A, B, A, C \rangle - \langle A, B \rangle \langle A, C \rangle, \\ & \langle A, B, A, B \rangle - \langle A, B \rangle \langle A, B \rangle, \quad \langle A, A, B, B \rangle - \langle A, A \rangle \langle B, B \rangle, \end{aligned} \quad (\text{B.34})$$

For an  $\langle A, B, C, D \rangle - \langle A, B \rangle \langle C, D \rangle$  term, the quantity in square brackets in (B.33) is

$$\begin{aligned} h_n^{4d} & \int_{\mathbb{R}^{4d}} \{ f_{i_1 i_2 i_3 i_4}(x - hu_1, x - hu_2, x - hu_3, x - hu_4) \\ & - f_{i_1 i_2}(x - hu_1, x - hu_2) f_{i_3 i_4}(x - hu_3, x - hu_4) \} \\ & \times \prod_{s=1}^4 \{ K(u_s) \sigma_{i_s}(x - hu_s) du_s \}. \end{aligned} \quad (\text{B.35})$$

By arguments similar to those in Lemma 4 the contribution to (B.32) is thus bounded by

$$\frac{C}{n^4} \sum_{\substack{i_s=1 \\ s=1, \dots, 4}}^n \left| \gamma_{i_1 i_2} \gamma_{i_3 i_4} \right| \rho_{i_1 i_2 i_3 i_4}(x, x, x, \varepsilon) \quad (\text{B.36})$$

for some  $\varepsilon > 0$ . This is  $o(t^2)$ , and thus  $o(s^2)$  if  $t = O(s)$ .

In a similar way, the contribution of an  $\langle A, A, B, C \rangle - \langle A, A \rangle \langle B, C \rangle$  term is bounded by

$$\frac{Ch^{-d}}{n^3} \sum_{\substack{i_s=1 \\ s=1, 2, 3}}^n \left| \gamma_{i_2 i_3 n} \right| \rho_{i_1 i_2 i_3}(x, x, x, \varepsilon). \quad (\text{B.37})$$

This is  $o(st)$ , which is  $o(s^2)$  if  $t = O(s)$ , and  $o(t^2)$  if  $s = o(t)$ . Likewise the contribution of an  $\langle A, A, B, B \rangle - \langle A, A \rangle \langle B, B \rangle$  term is bounded by

$$\frac{Ch^{-2d}}{n^2} \rho_{i_1 i_2}(x, x, \varepsilon). \quad (\text{B.38})$$

This is  $o(s^2)$ , and thus  $o(t^2)$  if  $s = o(t)$ .

The remaining terms in (B.34) are handled by showing that the individual components of each difference are  $o(w^2)$ . The  $\langle A, B, A, C \rangle$  contribution is (using Assumption B6) bounded by

$$\frac{Ch^{-d}}{n^4} \sum_{\substack{i_s=1 \\ s=1, 2, 3}}^n \left| \gamma_{i_1 i_2} \right| \left| \gamma_{i_1 i_3} \right| \sup_{\substack{|u_s| < \varepsilon \\ s=1, 2, 3}} f_{i_1 i_2 i_3}(x - u_1, x - u_2, x - u_3) = o(st); \quad (\text{B.39})$$

the  $\langle A, B \rangle \langle A, C \rangle$  one by

$$\frac{C}{n^4} \sum_{\substack{i_s=1 \\ s=1,2,3}}^n \left| \gamma_{i_1 i_2} \right| \left| \gamma_{i_1 i_3} \right| \sup_{\substack{|u_s| < \varepsilon \\ s=1,2}} f_{i_1 i_2}(x-u_1, x-u_2) \sup_{\substack{|u_s| < \varepsilon \\ s=1,3}} f_{i_1 i_3}(x-u_1, x-u_3) = o(st); \quad (\text{B.40})$$

the  $\langle A, B, A, B \rangle$  one by

$$\frac{Ch^{-2d}}{n^4} \sum_{\substack{i_s=1 \\ s=1,2}}^n \gamma_{i_1 i_2}^2 \sup_{\substack{|u_s| < \varepsilon \\ s=1,2}} f_{i_1 i_2}(x-u_1, x-u_2) = O(s^2 t); \quad (\text{B.41})$$

the  $\langle A, B \rangle \langle A, B \rangle$  one by

$$\frac{C}{n^4} \sum_{\substack{i_s=1 \\ s=1,2}}^n \gamma_{i_1 i_2}^2 \sup_{\substack{|u_s| < \varepsilon \\ s=1,2}} f_{i_1 i_2}^2(x-u_1, x-u_2) = o(s^2 t); \quad (\text{B.42})$$

the  $\langle A, A, A, B \rangle$  one by

$$\frac{Ch^{-2d}}{n^4} \sum_{\substack{i_s=1 \\ s=1,2}}^n \left| \gamma_{i_1 i_2} \right| \sup_{\substack{|u_s| < \varepsilon \\ s=1,2}} f_{i_1 i_2}(x-u_1, x-u_2) = O(s^2 t); \quad (\text{B.43})$$

the  $\langle A, A \rangle \langle A, B \rangle$  one by

$$\frac{Ch^{-d}}{n^4} \sum_{\substack{i_s=1 \\ s=1,2}}^n \left| \gamma_{i_1 i_2} \right| \sup_{|u| < \varepsilon} f_{i_1}(x-u) \sup_{\substack{|u_s| < \varepsilon \\ s=1,2}} f_{i_1 i_2}(x-u_1, x-u_2) = o(s^2 t); \quad (\text{B.44})$$

the  $\langle A, A, A, A \rangle$  one by

$$\frac{Ch^{-3d}}{n^4} \sum_{i=1}^n \sup_{|u| < \varepsilon} f_i(x-u) = O(s^3); \quad (\text{B.45})$$

and the  $\langle A, A \rangle \langle A, A \rangle$  one by

$$\frac{Ch^{-2d}}{n^4} \sum_{i=1}^n \sup_{|u| < \varepsilon} f_i^2(x-u) = o(s^3). \quad (\text{B.46})$$

□

**Lemma 11** *Let Assumptions A1, A2( $\xi$ ) for  $\xi > 4d$ , A4, B4, B5, B7, B9( $x_i$ ) and B10 hold,  $i = 1, \dots, p$ . Then*

$$E \left\{ \sum_{j=1}^N z_j^{(q)4} \right\} = O(b + n^{-1} + s + t), \text{ as } n \rightarrow \infty. \quad (\text{B.47})$$

**Proof:** Arguing as before and applying (5.44), the left side is bounded by

$$\begin{aligned}
& \frac{Cw^{-2}}{(nh^d)^4} \sum_{j=1}^N \sum_{\substack{i_s=1 \\ s=1,\dots,4}}^n |a_{i_1j}a_{i_2j}a_{i_3j}a_{i_4j}| \{h^{4d}1(i_s \neq i_t, s, t = 1, \dots, 4, s \neq t) \\
& + h^{3d}1(i_1 = i_2 \neq i_3, i_4; i_3 \neq i_4) + h^{2d}1(i_1 = i_2 \neq i_3 = i_4) + h^d1(i_1 = i_2 = i_3 = i_4)\} \\
\leq & \frac{Cw^{-2}}{n^4} \sum_{j=1}^N \left\{ \left( \sum_{i=1}^n |a_{ij}| \right)^4 + h^{-d} \left( \sum_{i=1}^n |a_{ij}| \right)^2 \sum_{i=1}^n a_{ij}^2 \right. \\
& \left. + h^{-2d} \left( \sum_{i=1}^n a_{ij}^2 \right)^2 + h^{-3d} \left( \sum_{i=1}^n a_{ij}^4 \right) \right\}. \tag{B.48}
\end{aligned}$$

We have first

$$\begin{aligned}
\sum_{j=1}^N \left( \sum_{i=1}^n |a_{ij}| \right)^4 & \leq \sup_{j \geq 1} \left( \sum_{i=1}^n |a_{ij}| \right)^2 \sum_{j=1}^N \left( \sum_{i=1}^n |a_{ij}| \right)^2 \\
& \leq b \left\{ \sum_{j=1}^{\infty} \left( \sum_{i=1}^n |a_{ij}| \right)^2 \right\}^2 \tag{B.49}
\end{aligned}$$

Now

$$\begin{aligned}
\sum_{j=1}^{\infty} \left( \sum_{i=1}^n |a_{ij}| \right)^2 & \leq \sum_{k=1}^{\infty} \left( \sum_{i,j=1; i \neq j}^n |a_{ik}a_{jk}| + \sum_{i=1}^n a_{ik}^2 \right) \\
& = O \left( \sum_{k=1}^{\infty} \sum_{i,j=1; i \neq j}^n a_{ik}a_{jk} + \sum_{i=1}^n a_{ik}^2 \right) \\
& = O \left( \sum_{i,j=1}^n \gamma_{ij} \right), \tag{B.50}
\end{aligned}$$

as  $n \rightarrow \infty$ , the penultimate bound using Assumption B11. We have

$$\begin{aligned}
\frac{w^{-2}}{n^4} \left( \sum_{i,j=1}^n \gamma_{ij} \right)^2 & \leq t^2 s^{-2} \leq C \text{ if } t = O(s), \\
& \leq t^{-2} t^2 \leq C \text{ if } s = o(t). \tag{B.51}
\end{aligned}$$

Thus the contribution to (B.48) from the first expression in its braces is  $O(b)$ .

Next

$$\begin{aligned}
\sum_{j=1}^N \left( \sum_{i=1}^n |a_{ij}| \right)^2 \sum_{i=1}^n a_{ij}^2 &\leq \sup_{j \geq 1} \sum_{i=1}^n a_{ij}^2 \left\{ \sum_{j=1}^N \left( \sum_{i=1}^n |a_{ij}| \right)^2 \right\} \\
&= O \left( \sum_{j=1}^{\infty} \left( \sum_{i=1}^n |a_{ij}| \right)^2 \right) \\
&= O \left( \sum_{i,j=1}^n \gamma_{ij} \right), \tag{B.52}
\end{aligned}$$

as  $n \rightarrow \infty$ , the second and last bounds using Assumption B5 and (B.50). But

$$\begin{aligned}
h^{-d} \frac{w^{-2}}{n^4} \sum_{i,j=1}^n \gamma_{ij} &\leq h^{-d} t s^{-2} / n^2 \leq C/n \text{ if } t = O(s) \\
&\leq h^{-d} t^{-1} / n^2 \leq C/n \text{ if } s = O(t) \tag{B.53}
\end{aligned}$$

Thus the contribution to (B.48) from the second expression in its braces is  $O(n^{-1})$ . Next

$$\begin{aligned}
\frac{Cw^{-2}h^{-2d}}{n^4} \sum_{j=1}^N \left( \sum_{i=1}^n a_{ij}^2 \right)^2 &\leq \frac{Cw^{-2}}{n^4} n h^{-2d} \left( \max_{1 \leq j \leq n} \sum_{i=1}^n a_{ij}^2 \right) \sum_{i=1}^n \sum_{j=1}^N a_{ij}^2 \\
&\leq Cw^{-2} s^2 / n \\
&\leq C/n, \tag{B.54}
\end{aligned}$$

for both  $t = O(s)$  and  $s = o(t)$ . Finally

$$\begin{aligned}
\frac{Cw^{-2}}{n^4} \sum_{j=1}^N h^{-3d} \left( \sum_{i=1}^n a_{ij}^4 \right) &\leq \frac{Cw^{-2}}{n^4} h^{-3d} \left( \sum_{i=1}^n \sum_{j=1}^N a_{ij}^4 \right) \\
&\leq \frac{Cw^{-2}}{n^3} h^{-3d} \leq C s^3 s^{-2} \leq C s \text{ if } t = O(s) \\
&\leq C s^3 t^{-2} = o(t) \text{ if } s = o(t), \tag{B.55}
\end{aligned}$$

and both tend to zero in view of Assumptions A4 and A10.  $\square$

## Acknowledgement

This research was supported by ESRC Grant RES-062-23-0036. The paper has been improved as a result of helpful comments of two referees, Supachoke Thawornkaiwong and Marcia Schafgans.

## References

- Andrews, D. W. K., 1995, Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560-596.
- Andrews, D. W. K., 2005. Cross-section regression with common shocks. *Econometrica* 73, 551-585.
- Arbia, G., 2006. *Spatial Econometrics: Statistical Foundations and Applications to Regional Analysis*. Springer-Verlag, Berlin.
- Castellana, J. V., Leadbetter, M.R., 1986. On smoothed probability density estimation for stationary processes. *Stochastic Processes and Their Applications* 21, 179-193.
- Cliff, A., Ord, J.K., 1981. *Spatial Processes: Models & Applications*, Pion, London.
- Dahlhaus, R., Künsch, H., 1987. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* 74, 877-882.
- Eicker, F., 1963. Asymptotic normality and consistency of the least squares estimator for families of linear regressions. *Annals of Mathematical Statistics* 34, 447-456.
- Gao, J., Lu, Z. and Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. *Annals of Statistics* 34, 1395-1435.
- Guyon, X., 1982. Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* 69, 95-106.
- Hall, P. and Hart, J.D., 1990. Nonparametric regression with long-range dependence. *Stochastic Processes and their Applications* 36, 339-351.
- Hallin, M., Lu, Z., Tran, L.T., 2001. Density estimation for spatial linear processes. *Bernoulli* 7, 657-688.
- Hallin, M., Lu, Z., Tran, L.T., 2004a. Local linear spatial regression. *Annals of Statistics* 32, 2469-2500.
- Hallin, M., Lu, Z., Tran, L.T., 2004b. Kernel density estimation for spatial processes: the  $L_1$  theory. *Journal of Multivariate Analysis* 88, 61-75.
- Hannan, E.J., 1957. The variance of the mean of a stationary process. *Journal of the Royal Statistical Society Series B* 19, 22-285.
- Ibragimov I.A., Linnik Yu.V., 1971. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
- Kelejian, H.H. and Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.
- Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory* 18, 252-277.
- Lu, Z., Lundervold, A., Tjøstheim, D., Yao, Q., 2007. Exploring spatial non-linearity using additive approximation. *Bernoulli* 13, 447-472.
- Pinkse, J., Shen, L. and Slade, M.E., 2007. A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics* 140, 215-225.
- Robinson, P.M., 1977. Estimation of a time series model from unequally spaced data. *Stochastic Processes and Their Applications* 6, 9-24.

- Robinson, P.M., 1983. Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185-207.
- Robinson, P.M., 1991. Nonparametric function estimation for long-memory time series. In Barnett, W., Powell, J., and Tauchen, G. ( eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press, Cambridge, 437-457.
- Robinson, P.M., 1997. Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics* 28, 2054-2083.
- Robinson, P.M., 2005. Robust covariance matrix estimation: ‘HAC’ estimates with long memory/antipersistence correction. *Econometric Theory* 21, 171-180.
- Robinson, P.M., Vidal Sanz, J., 2006. Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis* 97, 1090-1120.
- Rosenblatt, M., 1971. Curve estimates. *Annals of Mathematical Statistics* 42, 1815-1842.
- Roussas, G., 1969. Nonparametric estimation of the transition distribution function of a Markov process. *Annals of Mathematical Statistics* 40, 1386-1400.
- Scott, D.J., 1973. Central limit theorems for martingales using a Skorokhod representation approach. *Advances in Applied Probability* 5, 119-137.
- Tran, L.T., 1990. Kernel density estimation on random fields. *Journal of Multivariate Analysis* 34, 37-53.
- Tran, L.T., Yakowitz, S., 1993. Nearest neighbour estimators on random fields. *Journal of Multivariate Analysis* 44, 23-46.
- Whittle, P., 1954. On stationary processes in the plane. *Biometrika* 41, 434-449.
- Yao, Q., Brockwell, P.J. 2006. Gaussian maximum likelihood estimation for ARMA models II: spatial processes. *Bernoulli* 12, 403-429.