

Random-design regression under long-range dependent errors

SÁNDOR CSÖRGÖ¹ and JAN MIELNICZUK²

¹*Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1027, USA*

²*Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland.
e-mail: miel@ipipan.waw.pl*

We consider the random-design nonparametric regression model with long-range dependent errors that may also depend on the independent and identically distributed explanatory variables. Disclosing a smoothing dichotomy, we show that the finite-dimensional distributions of the Nadaraya–Watson kernel estimator of the regression function converge either to those of a degenerate process with completely dependent marginals or to those of a Gaussian white-noise process. The first case occurs when the bandwidths are large enough in a specified sense to allow long-range dependence to prevail. The second case is for bandwidths that are small in the given sense, when both the required norming sequence and the limiting process are the same as if the errors were independent. This conclusion is also derived for all bandwidths if the errors are short-range dependent. The borderline situation results in a limiting convolution of the two cases. The main results contrast with previous findings for deterministic-design regression.

Keywords: asymptotic finite-dimensional distributions; kernel estimators; long-range and short-range dependent errors; random-design regression; smoothing dichotomy

1. Introduction

Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be independent and identically distributed d -dimensional random vectors, $d \in \mathbb{N}$, and Z_1, Z_2, \dots be real random variables such that the two sequences $\{\mathbf{X}_i\}_{i=1}^\infty$ and $\{Z_i\}_{i=1}^\infty$ are independent. Set $(Z, \mathbf{X}) = (Z_1, \mathbf{X}_1)$, so that Z and \mathbf{X} are independent, and suppose that for $\epsilon = G(Z, \mathbf{X})$ we have $E(\epsilon|\mathbf{X}) = 0$ almost surely, where $G: \mathbb{R}^{1+d} \mapsto \mathbb{R}$ is a Borel-measurable function. Let Y be a real random variable with a finite mean, jointly distributed with \mathbf{X} . We consider the problem of estimating the regression function $g(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, using the $(1 + d)$ -dimensional observable sequence $\{(Y_i, \mathbf{X}_i)\}_{i=1}^\infty$, given by

$$Y_i = g(\mathbf{X}_i) + \epsilon_i, \quad \text{where } \epsilon_i = G(Z_i, \mathbf{X}_i), \quad i = 1, 2, \dots, \quad (1.1)$$

under the assumption that \mathbf{X} has a density function f with respect to the Lebesgue measure on \mathbb{R}^d and that the sequence $\{Z_i\}_{i=1}^\infty$ of latent variables is a stationary Gaussian process with zero mean and unit variance such that, for some $0 < \alpha < 1$,

$$r(i) = E(Z_1 Z_{i+1}) = \frac{L(i)}{i^\alpha}, \quad i = 1, 2, \dots, \quad (1.2)$$

where $L: [1, \infty) \mapsto \mathbb{R}$ is an eventually positive function, slowly varying at infinity.

Condition (1.2) implies that the sequence $\{Z_i\}_{i=1}^\infty$ exhibits long-range dependence in the sense that the lagged autocovariances $r(\cdot)$ are not summable. Long-range dependence may describe better than customary types of weak dependence the behaviour of many empirical time series encountered, for example, in geophysics and econometric studies (Beran 1992, 1994; Robinson 1994). However, such a dependence structure of the data may have significant effects on the properties of statistical estimators. This is in contrast with the typical situation of weakly dependent observations, in which estimators behave basically the same way as with independent observations. In particular, Dobrushin and Major (1979) and Taqu (1979) have shown that the partial sum of n long-range dependent variables requires a norming different from the usual $n^{1/2}$ to have a non-degenerate asymptotic law and, moreover, this law may not be normal.

We estimate g at the fixed points $\mathbf{x}_1, \dots, \mathbf{x}_l \in \mathbb{R}^d$ for some $l \in \mathbb{N}$ when the sample $(Y_1, X_1), \dots, (Y_n, X_n)$ is available. Specifically, we investigate the asymptotic distribution of kernel estimators of the vector $(g(\mathbf{x}_1), \dots, g(\mathbf{x}_l))$ and show that, if the amount of smoothing is large in a specified sense, meaning that the weighted average in (2.1) below is taken over many observations, then the effect of dependence prevails in determining the form of the asymptotic law. This is described in Theorem 1 in which the required norming sequence does not further depend on the amount of smoothing. In the opposite case of Theorem 2, when the amount of smoothing is small in the given sense, the estimators behave asymptotically as if Z_1, Z_2, \dots were independent. Thus, depending on the size of the smoothing parameter, the marginals of the asymptotic law are either completely dependent or independent. The borderline case of the *smoothing dichotomy* found is shown in Theorem 3 to result in a convolution of the limiting distributions in the two main cases.

The observed dichotomy is explained by a simple probabilistic fact. Namely, let momentarily $\epsilon_i = Z_i$, $i = 1, 2, \dots$, and consider an independent array X_{ni} of row-wise independent and identically distributed random variables with $E(X_{ni}) = \mu_n$ and $\text{var}(X_{ni}) = \sigma_n^2$, $i = 1, \dots, n$. Put $S_n = \sum_{i=1}^n X_{ni} \epsilon_i$. Then $S_n = \mu_n \sum_{i=1}^n Z_i + \sum_{i=1}^n (X_{ni} - \mu_n) Z_i =: S'_n + S''_n$. By Karamata's theorem $\text{var}(S'_n) \sim C \mu_n^2 L(n) n^{2-\alpha}$ for some constant $C > 0$ and $\text{var}(S''_n) = n \sigma_n^2$. So, the asymptotic distribution of S_n is determined either by S'_n or by S''_n , depending on whether $\mu_n^2 L(n) n^{2-\alpha} = o(n \sigma_n^2)$ or $n \sigma_n^2 = o(\mu_n^2 L(n) n^{2-\alpha})$. Theorems 1 and 2 convert this variance dichotomy into the smoothing dichotomy when $X_{ni} = K(x - X_i/b_n)$, $i = 1, \dots, n$, for a suitable kernel function K and a bandwidth sequence b_n . Because of the local character of X_{ni} , the reasoning remains valid for the general case $\epsilon_i = G(Z_i, X_i)$, $i = 1, 2, \dots$, under appropriate natural conditions.

Some analogous phenomena were discussed for smoothing operations performed on long-range dependent observations for density estimation. Hall and Hart (1990a) observed that the mean integrated squared error of kernel estimators of a density from a long-range dependent infinite-order moving-average sequence incorporating a small bandwidth is asymptotically equivalent to the mean integrated squared error of the same kernel estimator

based on the same number of independent observations drawn from the same marginal density. Ho (1996) has shown that the asymptotic law of a kernel estimator of a density from long-range dependent observations depends on the amount of smoothing involved; see also Csörgő and Mielniczuk (1995a) for the effect of long-range dependence with large bandwidths. For the related problem of the bandwidth choice in long- and short-range dependent cases see Hall *et al.* (1995b).

Theorem 1* is an analogue of Theorem 1 for $\{X_i\}_{i=1}^\infty$ satisfying certain weak dependence conditions in the special case when $G(z, \mathbf{x}) = G(z)$ for all $(z, \mathbf{x}) \in \mathbb{R}^{1+d}$. For linear regression this situation received investigation in depth by Koul (1992); see also Robinson and Hidalgo (1997) and many of their references for related work.

Theorem 2* derives the same conclusion as Theorem 2, by essentially the same proof, for all bandwidths (as in the corresponding statement with independent errors) when the errors in (1.1) are short-range dependent. This parallels Theorem 2 of Csörgő and Mielniczuk (1995c) for fixed-design regression with short-range dependent errors.

The present form of the errors in (1.1) was first considered by Cheng and Robinson (1994), who deal with the estimation of certain moment-type functionals. It allows for the dependence between the error ϵ_i and the variable X_i . The nature of this is a compromise between the restrictive assumption that $\{\epsilon_i\}_{i=1}^\infty$ is independent from $\{X_i\}_{i=1}^\infty$ and a general untractable assumption of dependence without an underlying structure. Furthermore, the assumed form of the errors turns out to yield, for specific functions G , models frequently considered in the statistical literature. For example, with $\langle \cdot, \cdot \rangle$ as the inner product in \mathbb{R}^d , the censored regression model, introduced by Tobin (1958),

$$Y_i = \begin{cases} \langle \mathbf{b}, X_i \rangle + W_i, & \text{if } \langle \mathbf{b}, X_i \rangle + W_i > 0, \\ 0, & \text{otherwise,} \end{cases} \quad i = 1, 2, \dots, \tag{1.3}$$

where $\mathbf{b} \in \mathbb{R}^d$ is a fixed unknown vector to be estimated and $W_i = R(Z_i)$ for some function $R: \mathbb{R} \mapsto \mathbb{R}$, is easily seen to be a special case of (1.1). As Cheng and Robinson (1994) pointed out, if S denotes the distribution function of $W = R(Z)$ and $I\{\cdot\}$ is the indicator function, we have $g(\mathbf{x}) = \langle \mathbf{b}, \mathbf{x} \rangle [1 - S(-\langle \mathbf{b}, \mathbf{x} \rangle)] + \int_{-\langle \mathbf{b}, \mathbf{x} \rangle}^\infty w dS(w)$ and $G(z, \mathbf{x}) = \{\langle \mathbf{b}, \mathbf{x} \rangle + R(z)\} I\{R(z) > -\langle \mathbf{b}, \mathbf{x} \rangle\} - g(\mathbf{x})$, $(z, \mathbf{x}) \in \mathbb{R}^{1+d}$ if $E(|R(Z)|) < \infty$.

In the univariate case ($d = 1$), a model parallel to (1.1) is regression with a deterministic design and long-range dependent errors in which the sequence $\{X_i\}_{i=1}^\infty$ is replaced by a predetermined triangular array of partition points of an interval. The linear case has been investigated by Koul and Mukherjee (1993) and Giraitis *et al.* (1994). For a general g such a model was considered by Hall and Hart (1990b) and Csörgő and Mielniczuk (1995b,d) when $\epsilon_i = G(Z_i)$, $i \in \mathbb{N}$. In particular, Csörgő and Mielniczuk (1995d) determined the asymptotic distribution of kernel estimators of g . It turns out that regression estimators behave very differently in the two models; this is discussed briefly in Section 3. The problem of bandwidth choice in this setting is considered by Hall *et al.* (1995a).

It is natural to conjecture that suitable versions of the smoothing dichotomy will hold for other forms of long-range dependent errors, such as (functions of) non-Gaussian moving averages, and for other techniques of smoothing, different from the kernel method.

2. Results and discussion

Let K_0 be a univariate probability density and, for $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, define $K(\mathbf{x}) = K_0(x_1) \dots K_0(x_d)$. Putting $\mathbf{x}/b = (x_1/b, \dots, x_d/b)$ for $b > 0$, we consider the Nadaraya (1964)–Watson (1964) estimate of $g(\mathbf{x})$:

$$\hat{g}_n(\mathbf{x}) = \frac{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{b_n}\right) Y_i}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{b_n}\right)}, \quad \mathbf{x} \in \mathbb{R}^d, \tag{2.1}$$

where $b_n > 0$ is a sequence of deterministic bandwidths tending to zero. Let $H_j(z) = (-1)^j e^{-z^2/2} d^j e^{-z^2/2} / dz^j$, $z \in \mathbb{R}$, denote the j th Hermite polynomial, $j = 0, 1, 2, \dots$, and φ the standard normal density. If $E\{G^2(Z, \mathbf{x})\} < \infty$ and $E\{G(Z, \mathbf{x})\} = 0$ for some $\mathbf{x} \in \mathbb{R}^d$, the function $G(\cdot, \mathbf{x})$ admits an $\mathcal{L}^2(\mathbb{R}, \varphi)$ Fourier–Hermite expansion:

$$G(z, \mathbf{x}) = \sum_{j=m(\mathbf{x})}^{\infty} \frac{c_j(\mathbf{x})}{j!} H_j(z), \quad z \in \mathbb{R}, \tag{2.2}$$

where the integer $m(\mathbf{x}) = \min\{j: c_j(\mathbf{x}) \neq 0\} \in \mathbb{N}$ is called the Hermite rank of $G(\cdot, \mathbf{x})$. Let $\mathbf{x}_1, \dots, \mathbf{x}_l$ be different fixed points in \mathbb{R}^d . Along with the conditions in the first paragraph of the introduction, we assume throughout that for $m = m(\mathbf{x}_{j^*}) = \min\{m(\mathbf{x}_1), \dots, m(\mathbf{x}_l)\} \in \mathbb{N}$ we have $m\alpha < 1$. This is the familiar condition of Taquq (1975) for the long-range dependence of the sequence $\{G(Z_i, \mathbf{x}_{j^*})\}_{i=1}^{\infty}$, under which it follows from results of Dobrushin and Major (1979) and Taquq (1979) that

$$\frac{a_{mn}}{n} \left(\sum_{k=1}^n G(Z_k, \mathbf{x}_1), \dots, \sum_{k=1}^n G(Z_k, \mathbf{x}_l) \right) \xrightarrow{\mathcal{D}} \frac{Y_m^*}{m!} (c_m(\mathbf{x}_1), \dots, c_m(\mathbf{x}_l)), \tag{2.3}$$

provided that the first two conditions in \mathbb{C}_4 below hold, where

$$a_{mn} = \left(\frac{(1 - m\alpha)(2 - m\alpha)}{2m!} \right)^{1/2} \frac{n^{m\alpha/2}}{L^{m/2}(n)}$$

and, understanding all convergence relations as $n \rightarrow \infty$ unless otherwise specified, $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, while Y_m^* denotes the value at $t = 1$ of a Hermite process of rank m , given by Taquq (1979) for each argument $t \in [0, 1]$ as a multiple Wiener–Itô integral, so that $E(Y_m^*) = 0$ and $E\{(Y_m^*)^2\} = 1$ for all $m \in \mathbb{N}$. The random variable Y_1^* is normal, but Y_2^*, Y_3^*, \dots are not normally distributed.

We consider two situations that are opposite to each other and are determined by the relative strength of dependence with respect to the amount of smoothing, in which both the order of the norming constants and the limiting law are different for the estimator \hat{g}_n . The case of Theorem 1 is determined by having $n^{m\alpha} L^{-m}(n) = o(nb_n^d)$ in condition (2.5) below. Setting $r(0) = 1$, by Karamata’s theorem together with Mehler’s classical formula

$$E\{H_n(Z_i)H_m(Z_j)\} = m!r^m(|i - j|)\delta_{nm}, \quad n, m = 0, 1, 2, \dots; i, j \in \mathbb{N}, \quad (2.4)$$

where δ_{nm} stands for Kronecker's symbol, an equivalent form of the condition is $(nb_n^d)^{-1} = o[\text{var}\{\sum_{i=1}^n G(Z_i, \mathbf{x}_j^*)/n\}]$. Here $(nb_n^d)^{-1}$ is asymptotically proportional to $\text{var}\{\hat{g}_n(\mathbf{x}_j^*)\}$ when the errors $\epsilon_1, \epsilon_2, \dots$ in (1.1) are independent and identically distributed. Hence under (2.5) such a variance will be dominated by the variance of the sample mean of the errors in the present model (1.1). Since our L is ultimately positive, when $d = 1$ the first condition in (2.5) is the same as in Ho (1996) for density estimation. The case of Theorem 2 is determined by the reverse condition $nb_n^d = o\{n^{m\alpha}L^{-m}(n)\} = o(a_{mn}^2)$.

Introducing the deviations $\Delta_n^G(x) = E([\int\{G(Z, \mathbf{x}) - G(Z, \mathbf{s})\}K_{b_n}(\mathbf{x} - \mathbf{s})f(\mathbf{s})d\mathbf{s}]^2) = E\{[E\{G(Z, \mathbf{x}) - G(Z, \mathbf{X})\}K_{b_n}(\mathbf{x} - \mathbf{X})|Z]^2\}$ and letting $|\mathbf{x}|$ be the Euclidean length of $\mathbf{x} \in \mathbb{R}^d$, the regularity conditions that are used in Theorem 1 are the following.

- \mathbb{C}_1 : K_0 is a symmetric bounded probability density such that $K_0(x) = 0$ for $x \notin [-1, 1]$.
- \mathbb{C}_2 : g is twice continuously differentiable in a neighbourhood of $\mathbf{x}_j, j = 1, \dots, l$.
- \mathbb{C}_3 : $f(\mathbf{x}_j) > 0$ and f is continuously differentiable in a neighbourhood of $\mathbf{x}_j, j = 1, \dots, l$.
- \mathbb{C}_4 : $E\{G(Z, \mathbf{x}_j)\} = 0, E\{G^2(Z, \mathbf{x}_j)\} > 0$ and the function $E\{G^2(Z, \cdot)\}$ is bounded in a neighbourhood of $\mathbf{x}_j, j = 1, \dots, l$.
- \mathbb{C}_5 : $\min_{1 \leq j \leq l} \min\{m(\mathbf{x}): |\mathbf{x} - \mathbf{x}_j| \leq \epsilon\} = m$ for some $\epsilon > 0$ and $\max_{1 \leq j \leq l} \Delta_n^G(\mathbf{x}_j) \rightarrow 0$.

While Theorem 2 does not require the first part of \mathbb{C}_5 , it does require stronger forms of the smoothness conditions in the second part of \mathbb{C}_5 and in \mathbb{C}_3 . Putting

$$|G'(z, \mathbf{x})| = \sum_{k=1}^d \left| \frac{\partial G(z, \mathbf{x})}{\partial x_k} \right| \quad \text{and} \quad |G''(z, \mathbf{x})| = \sum_{j=1}^d \sum_{k=1}^d \left| \frac{\partial^2 G(z, \mathbf{x})}{\partial x_j \partial x_k} \right|$$

for each $z \in \mathbb{R}$ and $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, these conditions are the following.

- \mathbb{C}_6 : $f(\mathbf{x}_j) > 0$ and f is twice differentiable in a neighbourhood of $\mathbf{x}_j, j = 1, \dots, l$.
- \mathbb{C}_7 : for each $z \in \mathbb{R}$ outside of a set of Lebesgue measure zero, the function $G(z, \cdot)$ is twice differentiable in a neighbourhood of \mathbf{x}_j such that $E[\text{sup}\{|G'(Z, \mathbf{x})|^2: |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\}] < \infty$ and $E[\text{sup}\{|G''(Z, \mathbf{x})|: |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\}] < \infty$ for some $\epsilon_j > 0, j = 1, \dots, l$.

Now we can state the two main results of the paper for “large” and “small” bandwidths.

Theorem 1. *Suppose that $\alpha m < 1$ and conditions $\mathbb{C}_1 - \mathbb{C}_5$ hold. If*

$$n^{m\alpha}L^{-m}(n) = o(nb_n^d) \text{ and the sequence } \{nb_n^{d+4}\}_{n=1}^\infty \text{ is bounded,} \quad (2.5)$$

then

$$a_{mn}(\hat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \dots, \hat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)) \xrightarrow{\mathcal{D}} \frac{Y_m^*}{m!}(c_m(\mathbf{x}_1), \dots, c_m(\mathbf{x}_l)). \quad (2.6)$$

Set $\kappa^2 = \int K^2(\mathbf{u})d\mathbf{u} = (\int_{-1}^1 K_0^2(u)du)^d$ and $\sigma^2(\mathbf{x}_j) = \kappa^2 E(G^2(Z, \mathbf{x}_j))/f(\mathbf{x}_j)$. Noting that $E\{G^2(Z, \mathbf{x}_j)\} = E(\epsilon^2|X = \mathbf{x}_j) = E(\{Y - g(X)\}^2|X = \mathbf{x}_j), j = 1, \dots, l$, we see that the limiting distribution below is the same as if Z_1, Z_2, \dots were independent.

Theorem 2. *Suppose that $\alpha m < 1$ and that $\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_4, \mathbb{C}_6$ and \mathbb{C}_7 hold. If*

$$nb_n^d = o\{n^{m\alpha} L^{-m}(n)\}, \quad nb_n^d \rightarrow \infty \quad \text{and} \quad nb_n^{d+4} \rightarrow 0, \quad (2.7)$$

then

$$(nb_n^d)^{1/2}(\hat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \dots, \hat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)) \xrightarrow{\mathcal{D}} (\sigma(\mathbf{x}_1)N_1, \dots, \sigma(\mathbf{x}_l)N_l), \quad (2.8)$$

where N_1, \dots, N_l are independent standard normal random variables.

The borderline case between “large” and “small” bandwidths is described in the following theorem, resulting in an asymptotic convolution of the two main cases.

Theorem 3. *Suppose that $\alpha m < 1$ and that $\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_4, \mathbb{C}_6$ and \mathbb{C}_7 hold. If*

$$\frac{nb_n^d}{a_{mn}^2} \rightarrow C_b^2 \quad \text{for some constant } 0 < C_b < \infty \quad \text{and} \quad nb_n^{d+4} \rightarrow 0,$$

then

$$(nb_n^d)^{1/2}(\hat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \dots, \hat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)) \xrightarrow{\mathcal{D}} \left(C_b c_m(\mathbf{x}_1) \frac{Y_m^*}{m!} + \sigma(\mathbf{x}_1)N_1, \dots, C_b c_m(\mathbf{x}_l) \frac{Y_m^*}{m!} + \sigma(\mathbf{x}_l)N_l \right),$$

where N_1, \dots, N_l are standard normal and Y_m^* is as in Theorem 1 such that the $l + 1$ random variables Y_m^*, N_1, \dots, N_l are independent.

The condition $nb_n^{d+4} \rightarrow 0$ in (2.7) is the familiar condition for eliminating the bias; this is what also brings in the derivatives in \mathbb{C}_2 and \mathbb{C}_7 . (See Schuster (1972) for $d = 1$, whose result for independent and identically distributed errors is routinely extended for a general d .) The essence of the restrictions on the bandwidth sequence is most easily illustrated if we consider the sequence $b_n \equiv C/n^\delta$ for some constants $C > 0$ and $\delta > 0$. Then condition (2.5) of Theorem 1 is satisfied if $1/(d + 4) \leq \delta < (1 - m\alpha)/d$, so that we must have $m\alpha < 4/(d + 4)$ to give room for δ . Condition (2.7) of Theorem 2 is satisfied whenever $\max\{(1 - m\alpha)/d, 1/(d + 4)\} < \delta < 1/d$.

Conditions $\mathbb{C}_4, \mathbb{C}_5$ and \mathbb{C}_7 are satisfied if, for instance, $G(z, \mathbf{x}) = G_1(z)G_2(\mathbf{x})$, $z \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^d$, for some Borel functions G_1 and G_2 such that $E\{G_1(Z)\} = 0$, $0 < E\{G_1^2(Z)\} < \infty$, $G_2(\mathbf{x}_j) \neq 0$, so that $c_m(\mathbf{x}_j) = c_m G_2(\mathbf{x}_j) \neq 0$, $j = 1, \dots, l$, where the constant c_m is the first non-zero coefficient of the Hermite expansion of G_1 , and G_2 satisfies the corresponding smoothness conditions. Here the first condition in \mathbb{C}_5 is empty and the second holds if G_2 is continuous in a neighbourhood of \mathbf{x}_j , while condition \mathbb{C}_7 holds if G_2 is twice differentiable in a neighbourhood of \mathbf{x}_j , $j = 1, \dots, l$.

Since $\lim_{b \downarrow 0} E[\{\sup\{|G(Z, \mathbf{x}) - G(Z, \mathbf{y})|: |\mathbf{x} - \mathbf{y}| \leq b\}\}^2] = 0$ implies $\Delta_n^G(\mathbf{x}) \rightarrow 0$ in general, the latter condition is satisfied for all $\mathbf{x} \in \mathbb{R}^d$ in the Tobin model (1.3) if $E\{R^2(Z)\} < \infty$ and \mathbb{C}_2 holds, while \mathbb{C}_2 is satisfied if $R(Z)$ has a continuously differentiable density. Since $E\{G(Z, \mathbf{x})\} = 0$ for all $\mathbf{x} \in \mathbb{R}^d$, Theorem 1 becomes applicable for this

model provided that the design variables satisfy \mathbb{C}_3 and that the first part of \mathbb{C}_5 is satisfied. Of course, this part of \mathbb{C}_5 is trivial in general if $m = m(\mathbf{x}_j^*) = 1$. Now, $m(\mathbf{x}) = 1$ for some $\mathbf{x} \in \mathbb{R}^d$ for the Tobin model if and only if $c_1(\mathbf{x}) = E\{G(Z, \mathbf{x})H_1(Z)\} = \int_{\{z: R(z) > -\langle \mathbf{b}, \mathbf{x} \rangle\}} \{\langle \mathbf{b}, \mathbf{x} \rangle + R(z)\} z \varphi(z) dz \neq 0$. It is not too difficult to see that this holds whenever $R(\cdot)$ is a non-constant monotone function such that $R(z) > -\langle \mathbf{b}, \mathbf{x} \rangle$ for at least one $z \in \mathbb{R}$. Condition \mathbb{C}_7 of Theorem 2 is also satisfied if $E\{R^2(Z)\} < \infty$ and the Lebesgue measure of the set $\{z: R(z) = -\langle \mathbf{b}, \mathbf{x}_j \rangle\}$ is zero for all $j = 1, \dots, l$.

The remaining discussion for long-range dependent errors is for the special case when $G(z, \mathbf{x}) = G(z)$, $(z, \mathbf{x}) \in \mathbb{R}^{1+d}$, so that (1.1) reduces to

$$Y_i = g(\mathbf{X}_i) + \epsilon_i = g(\mathbf{X}_i) + G(Z_i), \quad i = 1, 2, \dots, \tag{2.9}$$

for some Borel function $G \in \mathcal{L}^2(\mathbb{R}, \phi)$ such that $E\{G(Z)\} = 0$. Let c_j , $j = 1, \dots$, denote the coefficients in the Fourier–Hermite expansion of G and m denote its Hermite rank. The next result is an analogue of Theorem 1 and shows that (2.6) still holds true in the model (2.9) for some weakly dependent identically distributed explanatory variables $\mathbf{X}_1, \mathbf{X}_2, \dots$ with density f . We suppose that the Lebesgue density $f_{i,j}(\mathbf{x}, \mathbf{y})$ of $(\mathbf{X}_i, \mathbf{X}_j)$ exists for all $i, j \in \mathbb{N}$, $i \neq j$, and consider weak dependence in terms of conditions imposed on the functions $p_{i,j}(\mathbf{x}, \mathbf{y}) = f_{i,j}(\mathbf{x}, \mathbf{y}) - f(\mathbf{x})f(\mathbf{y})$ and their first and second partial derivatives. For $\mathbf{x} = (x_1, \dots, x_d)$, $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ and $b > 0$, introduce

$$p'_{i,j}(\mathbf{x}, \mathbf{y}) = \sum_{1 \leq k, n \leq d} \left(\left| \frac{\partial p_{i,j}(\mathbf{x}, \mathbf{y})}{\partial x_k} \right| + \left| \frac{\partial p_{i,j}(\mathbf{x}, \mathbf{y})}{\partial y_n} \right| \right),$$

$$p''_{i,j}(\mathbf{x}, \mathbf{y}) = \sum_{1 \leq k, n \leq d} \sum_{\substack{u, v=0,1,2 \\ u+v=2}} \left| \frac{\partial^2 p_{i,j}(\mathbf{x}, \mathbf{y})}{\partial x_k^u \partial y_n^v} \right|,$$

$$P_n(\mathbf{x}) = \sum_{1 \leq i \neq j \leq n} p_{i,j}(\mathbf{x}, \mathbf{x}), \quad P'_{n,b}(\mathbf{x}) = \sum_{1 \leq i \neq j \leq n} \sup_{|s|, |t| \leq b} |p'_{i,j}(\mathbf{x} + \mathbf{s}, \mathbf{x} + \mathbf{t})|,$$

$$P''_{n,b}(\mathbf{x}) = \max_{1 \leq i \neq j \leq n} \sup_{|s|, |t| \leq b} |p''_{i,j}(\mathbf{x} + \mathbf{s}, \mathbf{x} + \mathbf{t})|$$

and $Q_n(\mathbf{x}) = \sum_{1 \leq i \neq j \leq n} |p_{i,j}(\mathbf{x}, \mathbf{x})| |r(|j - i|)|^m$. Weak dependence of the type considered here was entertained by Rosenblatt (1970) and Castellana and Leadbetter (1986) for one-dimensional stationary sequences in the context of density estimation.

Theorem 1*. *Suppose that the modelling assumption (2.9) and conditions $\mathbb{C}_1 - \mathbb{C}_4$ and (2.5) hold. If, furthermore,*

$$P'_{n,b_n}(\mathbf{x}_j) = \mathcal{O}(n^2), \quad P''_{n,b_n}(\mathbf{x}_j) = o(b_n^{-2}), \tag{2.10}$$

$$P_n(\mathbf{x}_j) = o(n^2) \quad \text{and} \quad Q_n(\mathbf{x}_j) = o\{n^{2-m\alpha} L^m(n)\}$$

for all $j = 1, \dots, l$, then we have (2.6) with $c_m(\mathbf{x}_j)$ replaced by c_m .

The sum defining Q_n is $\mathcal{O}\{n^{2-m\alpha} L^m(n)\}$ if the factors $|p_{i,j}(\mathbf{x}, \mathbf{x})|$ are all absent from it.

Then it is easy to see, for example, that, if the sequence $\mathbf{X}_1, \mathbf{X}_2, \dots$ is stationary and $p_{1,i+1}(\mathbf{x}_j, \mathbf{x}_j) = \mathcal{O}(i^{-\beta})$ for some $\beta > 0$ as $i \rightarrow \infty$ for all $j = 1, \dots, l$, then the third and the fourth conditions in (2.10) are satisfied.

Finally, we return to the general model (1.1). The observation before the statement of Theorem 2 suggests that the conclusion should hold, for all $\{b_n\}$ such that $nb_n^d \rightarrow \infty$ and $nb_n^{d+4} \rightarrow 0$, when the errors $\epsilon_1, \epsilon_2, \dots$ in (1.1) are short-range dependent. With the Hermite rank m as defined between (2.2) and (2.3), this means that the covariances $r(i) = E(Z_1 Z_{i+1})$ satisfy $\sum_{i=1}^{\infty} |r(i)|^m < \infty$ instead of (1.2). Provided that $r(n) \rightarrow 0$ the last condition is equivalent to $E\{G^2(Z_1, \mathbf{x}_j)\} + 2\sum_{i=1}^{\infty} E\{G(Z_1, \mathbf{x}_j)G(Z_{i+1}, \mathbf{x}_j)\} < \infty$, $j = 1, \dots, l$; see Lemma 5 of Giraitis and Surgailis (1985).

Theorem 2*. *Suppose that $\sum_{i=1}^{\infty} |r(i)|^m < \infty$ and that conditions $\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_4, \mathbb{C}_6$ and \mathbb{C}_7 hold. If $nb_n^d \rightarrow \infty$ and $nb_n^{d+4} \rightarrow 0$, then we have (2.8).*

Although the sources of the arising asymptotic normality are completely different, Theorem 2* is a natural analogue of Theorem 2 in Csörgő and Mielniczuk (1995c) for deterministic-design regression with short-range dependent errors.

We note in passing that the dichotomous conclusions of Theorems 1 and 2, manifesting either asymptotic independence or full linear dependence, suggest a heuristically appealing data-driven choice of a bandwidth \hat{b}_n in situations when the statistician has reason to believe that the d -dimensional random explanatory variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independent or weakly dependent and, at the same time, suspects possible long-range dependence in the errors. Briefly, test successively for the independence of the variables $\hat{g}_n(\mathbf{x}_1), \dots, \hat{g}_n(\mathbf{x}_l)$, for example through some suitable bootstrap replicas, to find the largest $b_n = b_n^\circ$ at which independence is *not* rejected at a given level of significance, and proceed to find by the many existing methods a suitable $\hat{b}_n \leq b_n^\circ$ as if the errors were independent.

3. Deterministic and random design: a comparative discussion

For $d = 1$, the deterministic-design regression model

$$Y_i^{(n)} = g\left(\frac{i}{n}\right) + G(Z_{i,n}), \quad i = 1, \dots, n, \quad (3.1)$$

was considered by Hall and Hart (1990b) and Csörgő and Mielniczuk (1995d), where $\{Z_{i,n}\}_{i=1}^n$ satisfies (1.2) for every n . Here the random explanatory variables are replaced by the grid $1/n, \dots, (n-1)/n$ and $G(z, x) = G(z)$, $(z, x) \in \mathbb{R}^2$. The difference between the asymptotic behaviours of similar estimators for deterministic and random regression becomes particularly striking for long-range dependent errors. For a heuristic reason, note that estimation of a local functional of the conditional distribution $P(Y \leq \cdot | X = x)$ involves concomitants of the X_i or the grid points falling in a neighbourhood of x . If the design is deterministic, these concomitants form a block of consecutive observations. In contrast, if the design is random, they are randomly chosen from the Y sequence. Consequently, the dependence between the observations on which estimation is based is larger in the

deterministic case. Therefore, one expects greater variability of the estimators for deterministic designs in comparison to that with random designs.

Quantitative support of this intuition follows from Theorems 1 and 2 for the special model (2.9) and from Theorem 2 of Csörgö and Mielniczuk (1995d). Namely, the latter result shows that under (3.1) the correct norming factor for the Priestley–Chao regression estimator $\tilde{g}_n(x) = (nb_n)^{-1} \sum_{i=1}^n K([x - i/n]/b_n) Y_i$ to get a non-degenerate asymptotic distribution is $a_{mn}^* = (nb_n)^{ma/2} / L^{m/2} (nb_n)$. (As shown by Csörgö and Mielniczuk (1995c), the norming factor for $\tilde{g}_n(\cdot)$ is the classical $(nb_n)^{1/2}$ when the array of errors in (3.1) is short-range dependent.) The same result holds for the exact analogue of the Nadaraya–Watson estimator \hat{g}_n in (2.1) for the model (3.1), i.e. for the estimator $\hat{g}_n(x) = \tilde{g}_n(x) [(nb_n)^{-1} \sum_{i=1}^n K([x - i/n]/b_n)]^{-1}$. Since $b_n \rightarrow 0$ and $nb_n \rightarrow \infty$, it is easy to see that for the norming factor a_{mn} in Theorem 1 we have $a_{mn}^* = o(a_{mn})$ as well as $a_{mn}^* = o\{(nb_n)^{1/2}\}$ under the conditions of Theorem 2. These facts completely confirm the intuition described above. Moreover, since the bias of the Nadaraya–Watson estimator is of the same order under both deterministic and random designs, this remark suggests that random design with the uniform density f on $[0, 1]$ should be superior to the deterministic design in (3.1) from the point of view of mean squared errors.

4. Proofs

Let \xrightarrow{P} denote convergence in probability, put $K_b(x) = b^{-d} K(x/b)$ for any $b > 0$, and introduce $\hat{h}_n(x) = n^{-1} \sum_{i=1}^n K_{b_n}(x - X_i) Y_i$ and $\hat{f}_n(x) = n^{-1} \sum_{i=1}^n K_{b_n}(x - X_i)$, so that $\hat{g}_n(x) = \hat{h}_n(x) / \hat{f}_n(x)$ for all $x \in \mathbb{R}^d$ in (2.1).

Proof of Theorem 1. Since $b_n \rightarrow 0$ and, by the first of the two conditions assumed in (2.5), $nb_n^d \rightarrow \infty$, note first that $\hat{f}_n(x_j) \xrightarrow{P} f(x_j)$, $j = 1, \dots, l$, under \mathbb{C}_1 and \mathbb{C}_3 (Devroye and Wagner 1979). Hence, by the vector form of Slutsky’s theorem it clearly suffices to prove that $a_{mn}(\hat{h}_n(x_1) - g(x_1)\hat{f}_n(x_1), \dots, \hat{h}_n(x_l) - g(x_l)\hat{f}_n(x_l))$ converges in distribution to $Y_m^*(c_m(x_1)f(x_1), \dots, c_m(x_l)f(x_l))/m!$. This will follow from (2.3) once we show that

$$H_n^*(x_j) = \hat{h}_n(x_j) - g(x_j)\hat{f}_n(x_j) = \frac{f(x_j)}{n} \sum_{i=1}^n G(Z_i, x_j) + o_P\left(\frac{1}{a_{mn}}\right) \tag{4.1}$$

for every $j = 1, \dots, l$. For the rest of the proof, x denotes any one of x_1, \dots, x_l .

Write $H_n^*(x) = \hat{h}_n(x) - g(x)\hat{f}_n(x) = I_n(x) + J_n(x)$, where

$$I_n(x) = \frac{1}{n} \sum_{i=1}^n \{Y_i - g(X_i)\} K_{b_n}(x - X_i) \quad \text{and} \quad J_n(x) = \frac{1}{n} \sum_{i=1}^n \{g(X_i) - g(x)\} K_{b_n}(x - X_i).$$

First we verify that, under conditions \mathbb{C}_2 and \mathbb{C}_3 ,

$$E\{J_n^2(x)\} = \mathcal{O}(b_n^4 + n^{-1}b_n^{2-d}). \tag{4.2}$$

We have $E\{J_n^2(x)\} = A_n(x) + B_n(x)$, where, after changing variables also in $B_n(x)$,

$$\begin{aligned}
 A_n(\mathbf{x}) &= \frac{1}{n} \int f(\mathbf{s}) \{g(\mathbf{s}) - g(\mathbf{x})\}^2 K_{b_n}^2(\mathbf{x} - \mathbf{s}) \, d\mathbf{s} \\
 &= \frac{1}{nb_n^d} \int f(\mathbf{x} - b_n \mathbf{u}) \{g(\mathbf{x} - b_n \mathbf{u}) - g(\mathbf{x})\}^2 K^2(\mathbf{u}) \, d\mathbf{u}
 \end{aligned}$$

and

$$B_n(\mathbf{x}) = \frac{n(n-1)}{n^2} \left(\int f(\mathbf{x} - b_n \mathbf{u}) \{g(\mathbf{x} - b_n \mathbf{u}) - g(\mathbf{x})\} K(\mathbf{u}) \, d\mathbf{u} \right)^2. \tag{4.3}$$

Taking the one- and two-term Taylor expansions of $f(\mathbf{x} - \mathbf{u}b_n)$ and $g(\mathbf{x} - \mathbf{u}b_n) - g(\mathbf{x})$, respectively, about \mathbf{x} , we see by \mathbb{C}_1 - \mathbb{C}_3 that $B_n(\mathbf{x}) = \mathcal{O}(b_n^4)$. Similarly, we get $A_n(\mathbf{x}) = \mathcal{O}(b_n^{2-d}/n)$, establishing (4.2). Hence by the second condition in (2.5) it follows that $J_n(\mathbf{x}) = \mathcal{O}_P\{(nb_n^d)^{-1/2}\}$, and thus $J_n(\mathbf{x}) = o_P(a_{mn}^{-1})$ by the first condition in (2.5).

Next, we define $W_{in}(\mathbf{x}) = E(\epsilon_i K_{b_n}(\mathbf{x} - \mathbf{X}_i) | Z_i) = \int G(Z_i, \mathbf{s}) K_{b_n}(\mathbf{x} - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s}$ and claim that

$$\Delta_n^{(1)}(\mathbf{x}) = E \left\{ \left(I_n(\mathbf{x}) - \frac{1}{n} \sum_{i=1}^n W_{in}(\mathbf{x}) \right)^2 \right\} = \frac{1}{n^2} E \left\{ \left(\sum_{i=1}^n U_{in}(\mathbf{x}) \right)^2 \right\} = \mathcal{O} \left(\frac{1}{nb_n^d} \right), \tag{4.4}$$

where $U_{in}(\mathbf{x}) = \epsilon_i K_{b_n}(\mathbf{x} - \mathbf{X}_i) - W_{in}(\mathbf{x})$. Indeed, putting $W_n(\mathbf{x}) = E(\epsilon K_{b_n}(\mathbf{x} - \mathbf{X}) | Z)$, by the conditional independence of $U_{in}(\mathbf{x})$ and $U_{jn}(\mathbf{x})$ given Z_i and Z_j , we have

$$\begin{aligned}
 \Delta_n^{(1)}(\mathbf{x}) &= \frac{1}{n} E[\{\epsilon K_{b_n}(\mathbf{x} - \mathbf{X}) - W_n(\mathbf{x})\}^2] = \frac{1}{n} E[E(\{\epsilon K_{b_n}(\mathbf{x} - \mathbf{X}) - W_n(\mathbf{x})\}^2 | Z)] \\
 &\leq \frac{1}{n} \int E\{G^2(Z, \mathbf{s})\} K_{b_n}^2(\mathbf{x} - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s}.
 \end{aligned}$$

Hence (4.4) follows by \mathbb{C}_1 , \mathbb{C}_4 and that f is bounded near \mathbf{x} by \mathbb{C}_3 . Thus, by the first condition in (2.5), $\Delta_n^{(1)}(\mathbf{x}) = o(a_{mn}^{-2})$, whence $I_n(\mathbf{x}) - n^{-1} \sum_{i=1}^n W_{in}(\mathbf{x}) = o_P(a_{mn}^{-1})$.

Finally, for $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_l$ and $f_n(\mathbf{x}) = E\{\hat{f}_n(\mathbf{x})\} = \int f(\mathbf{s}) K_{b_n}(\mathbf{x} - \mathbf{s}) \, d\mathbf{s}$ we show that

$$\Delta_n^{(2)}(\mathbf{x}) = E \left\{ \left(\frac{1}{n} \sum_{i=1}^n W_{in}(\mathbf{x}) - \frac{f_n(\mathbf{x})}{n} \sum_{i=1}^n G(Z_i, \mathbf{x}) \right)^2 \right\} = o(a_{mn}^{-2}). \tag{4.5}$$

Since $f_n(\mathbf{x}) \rightarrow f(\mathbf{x})$ by conditions \mathbb{C}_1 and \mathbb{C}_3 and $a_{mn} \sum_{i=1}^n G(Z_i, \mathbf{x})/n = \mathcal{O}_P(1)$ by (2.3), this then implies (4.1) and hence the theorem.

Setting $v_{kn}(\mathbf{x}) = \int \{c_k(\mathbf{x}) - c_k(\mathbf{s})\} K_{b_n}(\mathbf{x} - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s}$, $k \in \mathbb{N}$, and using (2.2), (2.4), $b_n \rightarrow 0$, \mathbb{C}_1 and \mathbb{C}_5 , so that $m(\mathbf{s}) \geq m$ if \mathbf{s} is close enough to \mathbf{x} , for all large n we get

$$\begin{aligned}
 \Delta_n^{(2)}(\mathbf{x}) &= \frac{1}{n^2} \mathbb{E} \left\{ \left(\sum_{i=1}^n \int \{G(Z_i, \mathbf{x}) - G(Z_i, \mathbf{s})\} K_{b_n}(\mathbf{x} - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s} \right)^2 \right\} \\
 &= \frac{1}{n^2} \mathbb{E} \left\{ \left(\sum_{i=1}^n \sum_{k=m}^{\infty} H_k(Z_i) \int \frac{c_k(\mathbf{x}) - c_k(\mathbf{s})}{k!} K_{b_n}(\mathbf{x} - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s} \right)^2 \right\} \\
 &= \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \sum_{k=m}^{\infty} \frac{r^k (|i - j|)}{k!} v_{kn}^2(\mathbf{x}) \\
 &\leq \frac{1}{n^2} \sum_{k=0}^{\infty} \frac{v_{kn}^2(\mathbf{x})}{k!} \sum_{1 \leq i, j \leq n} |r(|i - j|)|^m \\
 &= \mathbb{E} \left\{ \left(\int \{G(Z, \mathbf{x}) - G(Z, \mathbf{s})\} K_{b_n}(\mathbf{x} - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s} \right)^2 \right\} \frac{1}{n^2} \sum_{1 \leq i, j \leq n} |r(|i - j|)|^m
 \end{aligned}$$

by the Parseval equality. Hence the second condition in \mathbb{C}_5 yields $\Delta_n^{(2)}(\mathbf{x}) = \Delta_n^G(\mathbf{x}) \mathcal{O}\{n^{-m\alpha} L^m(n)\} = o\{n^{-m\alpha} L^m(n)\}$, which is (4.5). \square

Proof of Theorem 1*. Since, with $f_n = \mathbb{E}(\hat{f}_n)$ as in (4.5), $f_n(\mathbf{x}_j) \xrightarrow{P} f(\mathbf{x}_j)$ by \mathbb{C}_1 and \mathbb{C}_3 and since, as we point out at the end of the proof, $\hat{f}_n(\mathbf{x}_j) - f_n(\mathbf{x}_j) \xrightarrow{P} 0$, $j = 1, \dots, l$, we may follow the outline of the proof of Theorem 1. Again, \mathbf{x} is any one of $\mathbf{x}_1, \dots, \mathbf{x}_l$.

To establish (4.2), we now have $\mathbb{E}\{J_n^2(\mathbf{x})\} = A_n(\mathbf{x}) + B_n^*(\mathbf{x})$, where $B_n^*(\mathbf{x}) = B_n(\mathbf{x}) + C_n(\mathbf{x})$ with $B_n(\mathbf{x}) = \mathcal{O}(b_n^4)$ as in (4.3) and $C_n(\mathbf{x})$ defined by the formula

$$\frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \int p_{i,j}(\mathbf{x} - b_n \mathbf{u}, \mathbf{x} - b_n \mathbf{v}) \{g(\mathbf{x} - b_n \mathbf{u}) - g(\mathbf{x})\} \{g(\mathbf{x} - b_n \mathbf{v}) - g(\mathbf{x})\} K(\mathbf{u}) K(\mathbf{v}) \, d\mathbf{u} \, d\mathbf{v}.$$

Hence a two-term expansion of g about \mathbf{x} and a one-term expansion of $p_{i,j}$ about (\mathbf{x}, \mathbf{x}) yield by $\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3$ and the first condition in (2.10) that $C_n(\mathbf{x}) = \mathcal{O}(b_n^4)$, giving (4.2). Since now $W_{in}(\mathbf{x}) = \epsilon_i f_n(\mathbf{x}) = G(Z_i) f_n(\mathbf{x})$, $i \in \mathbb{N}$, the last step of the proof of Theorem 1 is vacuous, and it suffices to establish that $\Delta_n^{(1)}(\mathbf{x}) = o(a_{mn}^{-2})$ directly for the present G . Elementary calculation shows that

$$\begin{aligned}
 \Delta_n^{(1)}(\mathbf{x}) &= \frac{\mathbb{E}(\epsilon^2)}{n} \text{var}\{K_{b_n}(\mathbf{x} - \mathbf{X})\} + \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \mathbb{E}(\epsilon_i \epsilon_j) p_{i,j}(\mathbf{x}, \mathbf{x}) \\
 &\quad + \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \mathbb{E}(\epsilon_i \epsilon_j) \int \{p_{i,j}(\mathbf{x} - b_n \mathbf{u}, \mathbf{x} - b_n \mathbf{v}) - p_{i,j}(\mathbf{x}, \mathbf{x})\} K(\mathbf{u}) K(\mathbf{v}) \, d\mathbf{u} \, d\mathbf{v}.
 \end{aligned}$$

Since the first term is $o(a_{mn}^{-2})$ by \mathbb{C}_3 and (2.5) and since $|\mathbb{E}(\epsilon_1 \epsilon_{1+k})| = |r(k)|^m \{1 + o(1)\}$ as $k \rightarrow \infty$ and $\sum_{1 \leq i \neq j \leq n} |\mathbb{E}(\epsilon_i \epsilon_j)| = \mathcal{O}\{n^{2-m\alpha} L^m(n)\}$, we see by a two-term expansion of the $p_{i,j}$ about (\mathbf{x}, \mathbf{x}) that $\Delta_n^{(1)}(\mathbf{x}) = o(a_{mn}^{-2})$, as desired, in view of (2.10).

Finally, note that, if all the factors involving ϵ are replaced by 1, then $\Delta_n^{(1)}(\mathbf{x})$ becomes $\text{var}\{\hat{f}_n(\mathbf{x})\}$ and is $o(1)$ by (2.10). Thus, consistency of $\hat{f}_n(\mathbf{x})$ follows. \square

Proof of Theorem 2. Fix any $u_1, \dots, u_l \in \mathbb{R}$ such that $u_1^2 + \dots + u_l^2 > 0$. Let $v_1 = u_1/\sigma_*, \dots, v_l = u_l/\sigma_*$, where $\sigma_*^2 = \sum_{j=1}^l u_j^2 \sigma^2(j)$ with $\sigma(j) = f(\mathbf{x}_j)\sigma(\mathbf{x}_j)$. The beginning of the proof of Theorem 1 can be followed by the consistency of $\hat{f}_n(\mathbf{x}_j)$, $j = 1, \dots, l$, and, by (4.2), the third condition in (2.7) and the Cramér–Wold device it suffices to prove that

$$S_n = \frac{(nb_n^d)^{1/2}}{n} \sum_{k=1}^n V_{nk}(\mathbf{X}_k) \xrightarrow{\mathcal{D}} Z,$$

where $V_{nk}(\mathbf{x}) = \sum_{j=1}^l v_j G(Z_k, \mathbf{x}) K_{b_n}(\mathbf{x}_j - \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, $k \in \mathbb{N}$.

Introducing the random variables $\mu_n = E(S_n | Z_1, \dots, Z_n)$ and $\sigma_n^2 = b_n^d s_n^2/n = \text{var}(S_n | Z_1, \dots, Z_n)$, this will follow by Slutsky’s theorem if we show that

$$\mu_n \xrightarrow{P} 0 \quad \text{and} \quad \sigma_n^2 \xrightarrow{P} 1, \tag{4.6}$$

together with $(S_n - \mu_n)/\sigma_n \xrightarrow{\mathcal{D}} Z$. By the moment convergence theorem this, in turn, follows if we show that $\phi_n(t) = \prod_{k=1}^n \int \exp\{it s_n^{-1} X_{nk}(\mathbf{x})\} f(\mathbf{x}) d\mathbf{x} \xrightarrow{P} \exp(-t^2/2)$ for each fixed $t \neq 0$, where i is the imaginary unit and $X_{nk}(\mathbf{x}) = V_{nk}(\mathbf{x}) - \int V_{nk}(s)f(s)ds$. Since $\int X_{nk}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0$, $k = 1, \dots, n$, and $\sum_{k=1}^n \int X_{nk}^2(\mathbf{x})f(\mathbf{x})d\mathbf{x} = s_n^2$ for all $n \in \mathbb{N}$, the last convergence relation follows if $s_n^{-2} \sum_{k=1}^n \int_{\{\mathbf{x}: |X_{nk}(\mathbf{x})| \geq \epsilon s_n\}} X_{nk}^2(\mathbf{x})f(\mathbf{x})d\mathbf{x} \xrightarrow{P} 0$ for each $\epsilon > 0$. This random Lindeberg condition is satisfied, once (4.6) is proved, if we show that

$$L_n(\epsilon) = \frac{b_n^d}{n} \sum_{k=1}^n \int_{\{\mathbf{x}: (b_n^d)^{1/2} |X_{nk}(\mathbf{x})| \geq \epsilon n^{1/2}\}} X_{nk}^2(\mathbf{x})f(\mathbf{x})d\mathbf{x} \xrightarrow{P} 0 \quad \text{for each } \epsilon > 0. \tag{4.7}$$

The proof of (4.6) is partially based on the ergodic theorem. Since the sequence $\{Z_k\}_{k=1}^\infty$ is ergodic and hence the transformed sequences $\{h_{p,j}(Z_k)\}_{k=1}^\infty$ are also ergodic, where $h_{1,j}(z) = |G(z, \mathbf{x}_j)|$, $h_{2,j}(z) = G^2(z, \mathbf{x}_j)$, $h_{3,j}(z) = |G'(z, \mathbf{x}_j)|$, $h_{4,j}(z) = \sup\{|G''(z, \mathbf{x})|: |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\}$ and $h_{5,j}(z) = \sup\{|G'(z, \mathbf{x})|^2: |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\}$ for some $\epsilon_j > 0$, $z \in \mathbb{R}$, by \mathbb{C}_4 and \mathbb{C}_7 it implies that

$$\frac{1}{n} \sum_{k=1}^n h_{p,j}(Z_k) \rightarrow E\{h_{p,j}(Z)\} \quad \text{almost surely for } p = 1, \dots, 5 \text{ and } j = 1, \dots, l. \tag{4.8}$$

To prove (4.6), set $\epsilon_{jn}(z) = \int \{G(z, \mathbf{x}_j - b_n \mathbf{s})f(\mathbf{x}_j - b_n \mathbf{s}) - G(z, \mathbf{x}_j)f(\mathbf{x}_j)\} K(\mathbf{s})d\mathbf{s}$ and $\eta_{jn}(z) = \int G^2(z, \mathbf{x}_j - b_n \mathbf{s})f(\mathbf{x}_j - b_n \mathbf{s})K^2(\mathbf{s})d\mathbf{s} - G^2(z, \mathbf{x}_j)f(\mathbf{x}_j)\kappa^2$, and note that by expanding G , in its vector variable, and f about \mathbf{x}_j and using elementary inequalities, conditions \mathbb{C}_1 , \mathbb{C}_6 and (4.8), as applied with $p = 1, 3, 4$ for the first conclusion and $p = 2, 5$ for the second, yield

$$\max_{1 \leq j \leq l} \frac{1}{n} \sum_{k=1}^n |\epsilon_{jn}(Z_k)| = \mathcal{O}(b_n^2) \quad \text{and} \quad \max_{1 \leq j \leq l} \frac{1}{n} \sum_{k=1}^n |\eta_{jn}(Z_k)| = \mathcal{O}(b_n) \tag{4.9}$$

almost surely. First we consider the statement $\mu_n \xrightarrow{P} 0$. We have

$$\begin{aligned} \mu_n &= \frac{(nb_n^d)^{1/2}}{n} \sum_{k=1}^n \sum_{j=1}^l v_j \int G(Z_k, \mathbf{x}) f(\mathbf{x}) K_{b_n}(\mathbf{x}_j - \mathbf{x}) \, d\mathbf{x} \\ &= \sum_{j=1}^l v_j f(\mathbf{x}_j) \frac{(nb_n^d)^{1/2}}{a_{mn}} \frac{a_{mn}}{n} \sum_{k=1}^n G(Z_k, \mathbf{x}_j) + \frac{(nb_n^d)^{1/2}}{n} \sum_{j=1}^l v_j \sum_{k=1}^n \epsilon_{jn}(Z_k) \\ &=: \mu_n^{(1)} + \mu_n^{(2)}. \end{aligned} \tag{4.10}$$

Thus, by (4.9) and the third condition in (2.7), $\mu_n^{(2)} = \mathcal{O}\{(nb_n^{d+4})^{1/2}\} = o(1)$ almost surely, while $\mu_n^{(1)} \rightarrow 0$ by (2.3) and the first condition in (2.7). Similarly, by simple calculation, by the second relation in (4.9) and by (4.8) with $p = 2$,

$$\sigma_n^2 = \frac{\kappa^2}{n} \sum_{k=1}^n \sum_{j=1}^l v_j^2 G^2(Z_k, \mathbf{x}_j) f(\mathbf{x}_j) + \mathcal{O}(b_n + b_n^d) \rightarrow \frac{1}{\sigma_*^2} \sum_{j=1}^l u_j^2 \sigma^2(j) = 1 \text{ almost surely.}$$

Finally, to prove (4.7), put $R_n(\mathbf{x}) = [x_1 - b_n, x_1 + b_n] \times \dots \times [x_d - b_n, x_d + b_n]$ for $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. Using the convexity of the square function twice and setting

$$I_{nk}^{j\epsilon}(\mathbf{x}) = I \left\{ \frac{b_n^d}{(nb_n^d)^{1/2}} \left(|v_j G(Z_k, \mathbf{x}) K_{b_n}(\mathbf{x}_j - \mathbf{x})| + \sum_{j=1}^l \left| v_j \int G(Z_k, \mathbf{s}) K_{b_n}(\mathbf{x}_j - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s} \right| \right) \geq \epsilon \right\}$$

for $\mathbf{x} \in \mathbb{R}^d$, for all n large enough we obtain $L_n(\epsilon) \leq 2l\{L_n^*(\epsilon) + W_n^*\}$, where

$$W_n^* = \frac{b_n^d}{n} \sum_{k=1}^n \sum_{j=1}^l v_j^2 \left(\int_{R_n(\mathbf{x}_j)} G(Z_k, \mathbf{s}) K_{b_n}(\mathbf{x}_j - \mathbf{s}) f(\mathbf{s}) \, d\mathbf{s} \right)^2 = \mathcal{O}(b_n^d + b_n^{d+2} + b_n^{d+4}) = o(1)$$

almost surely and

$$\begin{aligned} L_n^*(\epsilon) &= \frac{b_n^d}{n} \sum_{k=1}^n \sum_{j=1}^l v_j^2 \int_{R_n(\mathbf{x}_j)} I_{nk}^{j\epsilon}(\mathbf{x}) G^2(Z_k, \mathbf{x}) K_{b_n}^2(\mathbf{x}_j - \mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^l v_j^2 \int_{[-1,1]^d} I_{nk}^{j\epsilon}(\mathbf{x}_j - b_n \mathbf{u}) G^2(Z_k, \mathbf{x}_j - b_n \mathbf{u}) f(\mathbf{x}_j - b_n \mathbf{u}) K^2(\mathbf{u}) \, d\mathbf{u}. \end{aligned}$$

Introducing $h_j(z) = |v_j| \sup\{|G(z, \mathbf{x})|: |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\}$ and $g_j(z) = v_j^2 \sup\{G^2(z, \mathbf{x}): |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\}$, $z \in \mathbb{R}$, where the $\epsilon_j > 0$ are those in \mathbb{C}_7 , $j = 1, \dots, l$, and

$$h(z; x) = \sum_{j=1}^l C_j g_j(z) I \left(K^* h_j(z) + \sum_{i=1}^l C_i h_i(z) \geq x \right), \quad z \in \mathbb{R}, x > 0,$$

where $K^* = \sup\{K(\mathbf{u}): \mathbf{u} \in [-1, 1]^d\} < \infty$ and $C_j = \sup\{f(\mathbf{x}): |\mathbf{x} - \mathbf{x}_j| \leq \epsilon_j\} < \infty$, $j = 1, \dots, l$, it is now easy to see that for all n large enough,

$$L_n(\epsilon) \leq 2l\kappa^2 \frac{1}{n} \sum_{k=1}^n h(Z_k; \epsilon(nb_n^d)^{1/2}) + 2lW_n^*.$$

Now fix any $x > 0$ and by the second condition in (2.7) choose n so large that $nb_n^d \geq x^2$. Then $L_n(\epsilon) \leq 2l\kappa^2 n^{-1} \sum_{k=1}^n h(Z_k; \epsilon x) + 2lW_n^*$. Hence, by a final application of the ergodic theorem, $\limsup_{n \rightarrow \infty} L_n(\epsilon) \leq 2l\kappa^2 E\{h(Z; \epsilon x)\}$ almost surely for all $\epsilon > 0$ and $x > 0$. Since $\sum_{j=1}^l C_j E(g_j(Z)) < \infty$ by \mathbb{C}_7 , we have $E\{h(Z; \epsilon x)\} \downarrow 0$ as $x \rightarrow \infty$. Whence $\lim_{n \rightarrow \infty} L_n(\epsilon) = 0$ almost surely for each $\epsilon > 0$. \square

Proof of Theorem 2*. Long-range dependence and the first condition in (2.7) are used in the proof of Theorem 2 only at one single (but crucial) place, namely when showing that the first term $\mu_n^{(1)}$ of the conditional bias μ_n in (4.10) is negligible. The entire proof is otherwise valid for any stationary Gaussian sequence $\{Z_k\}_{k=1}^\infty$ under the rest of the conditions. Setting $\tilde{S}_n(\mathbf{x}_j) = \sum_{k=1}^n G(Z_k, \mathbf{x}_j)/n^{1/2}$, we have

$$E\{(\mu_n^{(1)})^2\} = b_n^d \sum_{1 \leq j, k \leq l} v_j v_k f(\mathbf{x}_j) f(\mathbf{x}_k) E\{\tilde{S}_n(\mathbf{x}_j) \tilde{S}_n(\mathbf{x}_k)\} \leq b_n^d l \sum_{j=1}^l v_j^2 f^2(\mathbf{x}_j) E\{\tilde{S}_n^2(\mathbf{x}_j)\}.$$

Since $E\{\tilde{S}_n^2(\mathbf{x}_j)\} \leq E\{G^2(Z, \mathbf{x}_j)\} + 2 \sum_{k=1}^\infty |E\{G(Z_1, \mathbf{x}_j) G(Z_{1+k}, \mathbf{x}_j)\}|$ and, using (2.4) and the present short-range condition, the second term of this bound is not greater than $2E\{G^2(Z, \mathbf{x}_j)\} \sum_{k=1}^\infty |r(k)|^m < \infty$, we see that $E\{(\mu_n^{(1)})^2\} = \mathcal{O}(b_n^b)$. \square

Proof of Theorem 3. An inspection of the proof of Theorem 2, with a basic point emphasized in the proof of Theorem 2*, shows that, if we set

$$I_n^*(\mathbf{x}) = \frac{(nb_n^d)^{1/2}}{f(\mathbf{x})} I_n(\mathbf{x}) - \mu_n^*(\mathbf{x}) = \frac{(nb_n^d)^{1/2}}{nf(\mathbf{x})} \sum_{k=1}^n G(Z_k, \mathbf{X}_k) K_{b_n}(\mathbf{x} - \mathbf{X}_k) - \mu_n^*(\mathbf{x}),$$

where

$$\mu_n^*(\mathbf{x}) = \frac{\mu_n^{(1)}(\mathbf{x})}{f(\mathbf{x})} = \frac{(nb_n^d)^{1/2}}{n} \sum_{k=1}^n G(Z_k, \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

then the asymptotic distribution is determined by the relation

$$(nb_n^d)^{1/2}(\hat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \dots, \hat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)) = (I_n^*(\mathbf{x}_1) + \mu_n^*(\mathbf{x}_1), \dots, I_n^*(\mathbf{x}_l) + \mu_n^*(\mathbf{x}_l)) + o_P(1).$$

This holds under the conditions of Theorem 3, using only the implication that $nb_n^d \rightarrow \infty$ of the “borderline” condition $(nb_n^d)^{1/2}/a_{mn} \rightarrow C_b$, where, putting $\underline{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_l)$, we have $I_n^*(\underline{\mathbf{x}}) = (I_n^*(\mathbf{x}_1), \dots, I_n^*(\mathbf{x}_l)) \rightarrow (\sigma(\mathbf{x}_1)N_1, \dots, \sigma(\mathbf{x}_l)N_l)$ and, in fact, with $\langle \cdot, \cdot \rangle_l$ as the inner product in \mathbb{R}^l ,

$$\phi_n^*(s; \underline{\mathbf{x}}) = E(e^{i\langle s, I_n^*(\underline{\mathbf{x}}) \rangle_l} | Z_1, \dots, Z_n) \xrightarrow{P} \prod_{j=1}^l e^{-\sigma^2(\mathbf{x}_j) s_j^2 / 2} =: \phi_{\underline{\mathbf{x}}}(s) \tag{4.11}$$

for all $s = (s_1, \dots, s_l) \in \mathbb{R}^l$. The present “borderline” condition implies by (2.3) that

$$\mu_n^*(\underline{\mathbf{x}}) \xrightarrow{\mathcal{D}} C_b \frac{Y_m^*}{m!}(c_m(\mathbf{x}_1), \dots, c_m(\mathbf{x}_l)). \quad (4.12)$$

Denoting the characteristic function of the last limit by $\psi_{\underline{\mathbf{x}}}(\cdot)$, since the random variable $\mu_n^*(\underline{\mathbf{x}})$ is measurable with respect to the σ -algebra generated by Z_1, \dots, Z_n , for the $2l$ -dimensional random vector $(I_n^*(\underline{\mathbf{x}}), \mu_n^*(\underline{\mathbf{x}}))$ we obtain

$$E(e^{i((s,t),(I_n^*(\underline{\mathbf{x}}), \mu_n^*(\underline{\mathbf{x}})))2l}) = E(\phi_n^*(s; \underline{\mathbf{x}}) e^{i(t, \mu_n^*(\underline{\mathbf{x}}))l}) \rightarrow \phi_{\underline{\mathbf{x}}}(s) \psi_{\underline{\mathbf{x}}}(t)$$

for all $s, t \in \mathbb{R}^l$, where the convergence results from (4.10), (4.11) and an application of the moment convergence theorem. This implies the desired conclusion. \square

Acknowledgements

S. Csörgő was supported in part by the National Science Foundation, USA, Grants DMS-9208067 and DMS-9625732. J. Mielniczuk was supported in part by the Committee for Scientific Research, Poland, Grant 2 P301 001 06; he thanks Ola Hössjer for his remarks.

References

- Beran, J. (1992) Statistical methods for data with long-range dependence. *Statist. Sci.*, **7**, 404–427.
- Beran, J. (1994) *Statistics for Long-memory Processes*. London: Chapman & Hall.
- Castellana, J.V. and Leadbetter, M.R. (1986) On smoothed probability density estimation. *Stochastic Processes Applic.*, **21**, 179–193.
- Cheng, B. and Robinson, P.M. (1994) Semiparametric estimation from time series with long-range dependence. *J. Econometrics*, **64**, 335–353.
- Csörgő, S. and Mielniczuk, J. (1995a) Density estimation under long-range dependence. *Ann. Statist.*, **23**, 990–999.
- Csörgő, S. and Mielniczuk, J. (1995b) Nonparametric regression under long-range dependent normal errors. *Ann. Statist.*, **23**, 1000–1014.
- Csörgő, S. and Mielniczuk, J. (1995c) Close short-range dependent sums and regression estimation. *Acta Sci. Math. (Szeged)*, **60**, 177–196.
- Csörgő, S. and Mielniczuk, J. (1995d) Distant long-range dependent sums and regression estimation. *Stochastic Processes Applic.*, **56**, 143–155.
- Devroye, L.P. and Wagner, T.J. (1979) The L^1 -consistency of kernel density estimates. *Ann. Statist.*, **7**, 1136–1139.
- Dobrushin, R.L. and Major, P. (1979) Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **50**, 27–52.
- Giraitis, L., Koul, H.L. and Surgailis, D. (1994) Asymptotic normality of regression estimators with long memory errors. *Statist. Probab. Lett.*, **29**, 317–335.
- Giraitis, L. and Surgailis, D. (1985) CLT and other limit theorems for functionals of Gaussian processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **70**, 191–212.
- Hall, P. and Hart, J.D. (1990a) Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory Related Fields*, **87**, 253–274.

- Hall, P. and Hart, J.D. (1990b) Nonparametric regression with long-range dependence. *Stochastic Processes Applic.*, **36**, 339–351.
- Hall, P., Lahiri, S.N. and Polzehl, J. (1995a) On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.*, **23**, 1921–1936.
- Hall, P., Lahiri, S.N. and Truong, J.K. (1995b) On bandwidth choice for density estimation with dependent data. *Ann. Statist.*, **23**, 2241–2263.
- Ho, H.-C. (1996) On central and non-central limit theorems in density estimation for sequences of long-range dependence. *Stochastic Processes Applic.*, **63**, 153–174.
- Koul, H.L. (1992) M-estimators in linear models with long range dependent errors. *Statist. Probab. Lett.*, **14**, 153–164.
- Koul, H.L. and Mukherjee, K. (1993) Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors. *Probab. Theory Related Fields*, **95**, 535–553.
- Nadaraya, E.A. (1964) Smooth regression analysis. *Theory Probab. Applic.*, **9**, 141–142.
- Robinson, P.M. (1994) Time series with strong dependence. In C.A. Sims (ed.), *Advances in Econometrics. Sixth World Congress*, Vol. 1, pp. 47–95. Cambridge: Cambridge University Press.
- Robinson, P.M. and Hidalgo, F.J. (1997) Time series regression with long range dependence. *Ann. Statist.*, **25**, 77–104.
- Rosenblatt, M. (1970) Density estimates and Markov sequences. In M.L. Puri (ed.), *Nonparametric Techniques in Statistical Inference*, pp. 199–210. Cambridge: Cambridge University Press.
- Schuster, E.F. (1972) Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.*, **43**, 84–88.
- Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **31**, 287–302.
- Taqqu, M.S. (1979) Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **50**, 53–83.
- Tobin, J. (1958) Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24–36.
- Watson, G.S. (1964) On estimating regression. *Sankhyā A*, **26**, 359–372.

Received August 1996 and revised September 1997