Statistica Sinica **10**(2000), 771-787

THE SMOOTHING DICHOTOMY IN RANDOM-DESIGN REGRESSION WITH LONG-MEMORY ERRORS BASED ON MOVING AVERAGES

Sándor Csörgő and Jan Mielniczuk

Universities of Michigan and Szeged and Polish Academy of Sciences

Abstract: We consider the random-design nonparametric regression model with errors an unknown function of long-range dependent moving averages and of the independent and identically distributed explanatory random vectors. We show that the Nadaraya–Watson kernel estimator of the regression function may exhibit a dichotomous asymptotic behavior depending on the amount of smoothing employed: its finite-dimensional distributions converge either to those of a not necessarily Gaussian degenerate process with completely dependent marginals or to those of a Gaussian white noise process. The borderline situation results in a limiting convolution of the two cases. Convergence to Gaussian white noise is also established when the resulting errors lack a long memory. The main results here are general analogues of those in Csörgő and Mielniczuk (1999), where the smoothing dichotomy was disclosed in the case when the long-range dependence of the errors is produced by a Gaussian sequence.

Key words and phrases: Finite-dimensional distributions, kernel estimators, long-memory errors, moving averages, random-design nonparametric regression, smoothing dichotomy.

1. Introduction

Let $\{Y_i\}_{i=1}^{\infty}$ and $\{\mathbf{X}_i\}_{i=1}^{\infty}$ be jointly stationary processes with respective values in $\mathbb{R}$ and $\mathbb{R}^d$, and suppose that $E(|Y|) < \infty$, where $(Y, \mathbf{X}) = (Y_1, \mathbf{X}_1)$ and $d \in \mathbb{N}$. The literature is extensive on the estimation of the multivariate regression function $g(\mathbf{x}) = E(Y \,|\, \mathbf{X} = \mathbf{x})$ of Y, given $\mathbf{X} = \mathbf{x}$, from the observations $\{(Y_i, \mathbf{X}_i)\}_{i=1}^{n}$ when $(Y_1, \mathbf{X}_1), (Y_2, \mathbf{X}_2), \ldots$ are independent or weakly dependent. There has been an increasing recent interest in the problem for observations which exhibit long-range dependence (see Csörgő and Mielniczuk (1999) for references). Various modeling assumptions have been entertained to describe long-range dependence. Here we consider the model

$$Y_i = g(\mathbf{X}_i) + \eta_i\,, \quad \text{where } \eta_i = G(Z_i, \mathbf{X}_i)\,, \;\; i = 1, 2, \ldots, \tag{1.1}$$

where $G : \mathbb{R}^{1+d} \mapsto \mathbb{R}$ is a Borel measurable function. We assume condition that $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent and identically distributed, $\mathbf{X}$ 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, assumed to be independent of the sequence $\{\mathbf{X}_i\}_{i=1}^{\infty}$, is a moving-average process given by $Z_i = \sum_{j=1}^{\infty} a_j \varepsilon_{i-j}$. Here $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ is a sequence of independent, identically distributed innovations with $E(\varepsilon) = 0$ and $E(\varepsilon^2) = 1$, where $\varepsilon = \varepsilon_0$ throughout, and the nonzero constants $\{a_j\}_{j=1}^{\infty}$ are such that $a_j = L_0(j)/j^{\beta}$ for some $\beta \in (1/2, 1)$ and a function $L_0 : [1, \infty) \mapsto \mathbb{R}$ slowly varying at infinity. This implies

$$r(i) := E(Z_1 Z_{i+1}) = \frac{L(i)}{i^{\alpha}} \tag{1.2}$$

for the exponent $\alpha = 2\beta - 1 \in (0, 1)$ and the slowly varying function $L(\cdot) = C_\beta L_0^2(\cdot)$, with $C_\beta = \int_0^\infty (x + x^2)^{-\beta} dx$. (The statement in (1.2) appears to be folklore. It is Lemma 2 in Mielniczuk (1997) under an unnecessary condition on L_0, which can be removed by applying Theorem 1.9.7 in Bingham, Goldie and Teugels (1987) to the sequence $\{L_0(j)\}_{j=1}^{\infty}$.)

Equation (1.2) implies that the sequence $\{Z_i\}_{i=1}^{\infty}$ exhibits long-range dependence, or long memory, in the sense that the lagged autocovariances $r(\cdot)$ are not summable. Csörgő and Mielniczuk (1999) considered the estimation of g for that version of the model (1.1) where, in place of the present linear process, $\{Z_i\}_{i=1}^{\infty}$ was a long-memory stationary Gaussian sequence. In this paper we allow for a much more general form of $\{Z_i\}_{i=1}^{\infty}$ (see Brockwell and Davis (1987, Theorem 5.7.1)). We estimate g at the points $\mathbf{x}_1, \ldots, \mathbf{x}_l \in \mathbb{R}^d$ for some $l \in \mathbb{N}$ when the sample $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$ is available, considering the time-honored Nadaraya–Watson kernel estimator $(\widehat{g}_n(\mathbf{x}_1), \ldots, \widehat{g}_n(\mathbf{x}_l))$ of the vector $(g(\mathbf{x}_1), \ldots, g(\mathbf{x}_l))$, given by (2.1) below. We show that if the amount of smoothing is large in a specified sense, meaning that the weighted average in (2.1) 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 $Z_1, Z_2, \ldots$ 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 is shown in Theorem 3 to result in a convolution of the limiting distributions in the two main cases.

The same dichotomous phenomenon was disclosed by Csörgő and Mielniczuk (1999) for a Gaussian $\{Z_i\}$. It was conceivable that the smoothing dichotomy should extend in some form or other to the present situation; indeed the structure of the proofs is retained here, where Sections 2, 3 and 4 respectively contain preliminaries, the main results and all the proofs. However, the basic ingredients of the proofs are necessarily very different and the proofs themselves depend, in part, upon understanding Ho and Hsing's (1996, 1997) recent breakthrough. We also note that the heuristic method of bandwidth choice based on the dichotomy may be imported from Csörgő and Mielniczuk (1999) for the general case here. The problem of the optimal bandwidth in a model analogous to (1.1), when $d = 1$ and the design is deterministic, is discussed by Hall, Lahiri and Polzehl (1995).

2. Preliminaries

Let K_0 be a univariate kernel, a Borel measurable function on $\mathbb{R}$, with properties specified later. For $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, define $K(\mathbf{x}) = K_0(x_1) \cdots K_0(x_d)$. Putting $\mathbf{x}/b = (x_1/b, \ldots, x_d/b)$ for $b > 0$, consider the Nadaraya–Watson estimate of $g(\mathbf{x})$:

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

where $b_n > 0$ is a sequence of deterministic bandwidths tending to zero. Setting $K_{b_n}(\mathbf{x}) = b_n^{-d} K(\mathbf{x}/b_n)$, let $\widehat{f}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^n K_{b_n}(\mathbf{x} - \mathbf{X}_i)$ be the corresponding kernel estimate of the density f of $\mathbf{X}$ at $\mathbf{x}$, and introduce also $f_n(\mathbf{x}) = E(\widehat{f}_n(\mathbf{x}))$, $\mathbf{x} \in \mathbb{R}^d$.

We write $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$, and let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution. All asymptotic relations are meant as $n \to \infty$. Using the modeling assumptions in the introduction throughout, consider

$$Y_{n,m} = \sum_{k=1}^n \sum_{1 \le j_1 < j_2 < \cdots < j_m} \prod_{i=1}^m a_{j_i} \varepsilon_{k-j_i} \quad \text{for some } m \in \mathbb{N},$$

the building blocks of Ho and Hsing's (1996, 1997) theory. Avram and Taqqu (1987) proved that if $m\alpha < 1$ and $E(\varepsilon^{2m}) < \infty$, then $\{Y_{n,m}\}_{n=1}^\infty$ is long-range dependent and

$$\frac{a_{n,m}}{n} Y_{n,m} \xrightarrow{\mathcal{D}} Y_m^*, \tag{2.2}$$

where $a_{n,m} \sim n/[E(Y_{n,m}^2)]^{1/2}$. Then by Lemma 6.1 in Ho and Hsing (1996) (in the statement of which the constant must be squared) $a_{n,m}$ may be chosen as

$$a_{n,m} = C_m(\beta) \frac{n^{m\alpha/2}}{L^{m/2}(n)}, \quad \text{where} \quad C_m(\beta) = \sqrt{\frac{m!\,(1-m\alpha)(2-m\alpha)}{2}}\,.$$

The random variable Y_m^* in (2.2) is given by the multiple Wiener–Itô integral

$$Y_m^* = K_m(\beta) \int \cdots \int_{\{-\infty < u_1 < \cdots < u_m < 1\}} \Big[\int_0^1 \prod_{j=1}^m [(v-u_j)^+]^{-\beta} dv \Big] dB(u_1) \cdots dB(u_m)\,,$$

where $K_m(\beta) = C_m(\beta)/C_\beta^{m/2}$ for the C_β in (1.2), and B is a standard Brownian motion on $\mathbb{R}$. Thus Y_m^* is the value $Y_m^*(1)$ of a Hermite process $Y_m^*(t)$, $t \geq 0$, of rank m, given by the same formula with t replacing 1 (see e.g. Avram and Taqqu (1987)). We have $E(Y_m^*) = 0$ and $E([Y_m^*]^2) = 1$. Here Y_1^* is normally distributed, but $Y_2^*, Y_3^*, \ldots$ are not.

We say that the Borel measurable function K_0 on $\mathbb{R}$ is a kernel of order $\kappa \in \mathbb{N}$, $\kappa \geq 2$, if $\int K_0(s)ds = 1$, $\int K_0(s)s^i ds = 0$ for $i = 1, \ldots, \kappa - 1$ and $\int K_0(s)s^\kappa ds \neq 0$. For $h : \mathbb{R} \mapsto \mathbb{R}$ and $k = 0, 1, \ldots$, let $h^{(k)}(\cdot)$ denote the kth derivative of the function $h(\cdot)$, with $h^{(0)}(\cdot) = h(\cdot)$. Let $H_j(\cdot)$ and $\widetilde{H}_j(\cdot)$ be the distribution functions of $\sum_{i=1}^{j} a_i \varepsilon_{1-i}$ and $\sum_{i=j+1}^{\infty} a_i \varepsilon_{1-i}$, respectively, and introduce the functions $G_j(z, \mathbf{x}) = \int_{\mathbb{R}} G(z + v, \mathbf{x})\, dH_j(v)$, $j = 0, 1, \ldots$, so that, an empty sum being understood as zero, $H_0(v) = 0$ or 1 according as $v < 0$ or $v \geq 0$, and thus $G_0(z, \mathbf{x}) = G(z, \mathbf{x})$, $z \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^d$. Furthermore, $H(\cdot) := \widetilde{H}_0(\cdot)$ is the distribution function of $Z = Z_1$. For a fixed $\mathbf{x} \in \mathbb{R}^d$ and an integer $m \in \mathbb{N}$, consider the following conditions.

$\mathbb{C}_0(\mathbf{x}, m)$: There exist an integer $\tau \in \{0\} \cup \mathbb{N}$ and a number $\lambda > 0$ such that for every $r = 0, \ldots, m + 2$, the derivative $G_\tau^{(r)}(u, \mathbf{x}) = \partial^r G_\tau(u, \mathbf{x})/\partial u^r$ exists and is continuous on $\mathbb{R}$. If $G_{\tau,\lambda}^{(r)}(u, \mathbf{x}) := \sup_{-\lambda \leq v \leq \lambda} |G_\tau^{(r)}(u + v, \mathbf{x})|$, $u \in \mathbb{R}$,

$$C_{\tau,\lambda}^m(G(\cdot, \mathbf{x}); z) := \max_{0 \leq r \leq m+2} \sup_{I \subset \mathbb{N}} E\Big(\Big[\Big| G_{\tau,\lambda}^{(r)}\Big(z + \sum_{i \in I} a_i \varepsilon_i, \mathbf{x} \Big) \Big| \Big]^4 \Big) < \infty$$

for all $z \in \mathbb{R}$, where the supremum is taken over all subsets I of the natural numbers;

$\mathbb{C}_0^*(\mathbf{x}, m)$: There exist a neighborhood $U_\mathbf{x} \subset \mathbb{R}^d$ of $\mathbf{x}$, an integer $\tau \in \{0\} \cup \mathbb{N}$ and a number $\lambda > 0$ such that for every $\mathbf{y} \in U_\mathbf{x}$ the derivative $G_\tau^{(r)}(z, \mathbf{y}) = \partial^r G_\tau(z, \mathbf{y})/\partial z^r$ exists and is continuous for each $r = 0, \ldots, m + 2$, and $\sup_{\mathbf{y} \in U_\mathbf{x}} C_{\tau,\lambda}^m(G(\cdot, \mathbf{y}); z) < \infty$ for every $z \in \mathbb{R}$;

$\mathbb{C}_1(0)$: K_0 is bounded and $K_0(y) = 0$ for $y \notin [-1, 1]$;

$\mathbb{C}_1(\kappa)$: K_0 is a bounded kernel of order κ such that $K_0(y) = 0$ for $y \notin [-1, 1]$;

$\mathbb{C}_2(\mathbf{x})$: g is twice continuously differentiable in a neighborhood of $\mathbf{x}$;

$\mathbb{C}_3(\mathbf{x})$: $f(\mathbf{x}) > 0$ and f is continuously differentiable in a neighborhood of $\mathbf{x}$;

$\mathbb{C}_4(\mathbf{x})$: $E(G(Z, \mathbf{x})) = 0$, $E(G^2(Z, \mathbf{x})) > 0$ and the function $E(G^2(Z, \cdot))$ is bounded in a neighborhood of $\mathbf{x}$;

$\mathbb{C}_5(\mathbf{x})$: $E(G(Z, \mathbf{x})) = 0$, $E(G^2(Z, \mathbf{x})) > 0$ and $E([\sup_{\mathbf{y} \in V_\mathbf{x}} |G(Z, \mathbf{y})|]^2) < \infty$ for a neighborhood $V_\mathbf{x} \subset \mathbb{R}^d$ of $\mathbf{x}$.

Conditions $\mathbb{C}_2(\mathbf{x}) - \mathbb{C}_4(\mathbf{x})$ were used in Theorem 1 of Csörgő and Mielniczuk (1999), $\mathbb{C}_5(\mathbf{x})$ is somewhat stronger than $\mathbb{C}_4(\mathbf{x})$, while $\mathbb{C}_0(\mathbf{x}, m)$ is the basic condition of Ho and Hsing (1997) on the function $G(\cdot, \mathbf{x})$. As they point out, $\mathbb{C}_0(\mathbf{x}, m)$ holds with $\tau = 0$ if the continuous derivatives $G^{(r)}(\cdot, \mathbf{x})$ are all bounded on $\mathbb{R}$, $r = 0, \ldots, m+2$. Similarly, the version $\mathbb{C}_0^*(\mathbf{x}, m)$, local uniformity in the vector variable, is satisfied with $\tau = 0$ if there exists a neighborhood $U_{\mathbf{x}}$ of $\mathbf{x}$ such that the derivatives $G^{(r)}(\cdot, \mathbf{y})$ exist and are continuous for each $\mathbf{y} \in U_{\mathbf{x}}$, $r = 0, \ldots, m+2$, and $\max_{0 \le r \le m+2} \sup_{\mathbf{y} \in U_{\mathbf{x}}} \sup_{z \in \mathbb{R}} |G^{(r)}(z, \mathbf{y})| < \infty$.

However, the generality of condition $\mathbb{C}_0^*(\mathbf{x}, m)$ is greatly restricted by such sufficient conditions expressed in terms of G only. Note that conditions $\mathbb{C}_0^*(\mathbf{x}, m)$ and $\mathbb{C}_5(\mathbf{x})$ may hold for various unbounded functions G when some weak assumptions on the distribution of innovations are permissible. For example, consider the situation when $G(z, \mathbf{y}) = P_k(z) G_*(\mathbf{y})$, where P_k is a polynomial of order k and G_* is bounded with $\mathbf{y}$ in a neighborhood of $\mathbf{x}$. Then conditions $\mathbb{C}_5(\mathbf{x})$ and $\mathbb{C}_0^*(\mathbf{x}, m)$ hold (with $\tau = 0$) provided the innovations have finite moments up to the order $4k$. This follows easily from the Hölder inequality. Moreover, both conditions are also satisfied if P_k is replaced by the indicator function of an arbitrary interval, provided the distribution function of ε has a continuous and integrable second derivative (see Ho and Hsing (1997, Remark 1)).

As another nontrivial example consider the Tobin model, a kind of a censored regression model widely used in econometrics, as expressed in the context of possibly long-range dependent observations by Cheng and Robinson (1994). In this model,

$$Y_i = \begin{cases} \langle \mathbf{b}\,, \mathbf{X}_i \rangle + W_i\,, & \text{if } \langle \mathbf{b}\,, \mathbf{X}_i \rangle + W_i > 0, \\ 0\,, & \text{otherwise}, \end{cases} \qquad i = 1, 2, \ldots,$$

where $\langle \cdot\,, \cdot \rangle$ stands for the inner product in $\mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^d$ is a fixed unknown vector and $W_i = R(Z_i)$ for some Borel measurable function $R : \mathbb{R} \mapsto \mathbb{R}$, and where we now assume that $\{Z_i\}_{i=1}^{\infty}$ is a moving-average process as defined between (1.1) and (1.2). Here, if $S(\cdot)$ denotes the distribution function of $W = R(Z)$ and $I\{\,\cdot\,\}$ is the indicator function, then, as Cheng and Robinson (1994) point out, 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 $G(z, \mathbf{x}) = [\langle \mathbf{b}\,, \mathbf{x} \rangle + R(z)] I\{R(z) > -\langle \mathbf{b}\,, \mathbf{x} \rangle\} - g(\mathbf{x})$, $(z, \mathbf{x}) \in \mathbb{R}^{1+d}$, whenever $E(|R(Z)|) < \infty$. Assuming that $R(\cdot)$ is bounded and that the distribution function $H_\varepsilon(t) = P\{\varepsilon \le t\}$, $t \in \mathbb{R}$, has a continuous and integrable second derivative $H_\varepsilon''(\cdot)$, we see that for each $\mathbf{x} \in \mathbb{R}$ the function $G_\tau^{(r)}(\cdot, \mathbf{x})$ is bounded and continuous whenever $\tau \ge r$. Moreover, condition $\mathbb{C}_0^*(\mathbf{x}, m)$ holds for any $\mathbf{x} \in \mathbb{R}$ and $m \in \mathbb{N}$ since each integer $\tau \ge m+2$ is a good choice. Indeed, changing variables, differentiating under the integral sign and changing the variable back, and then repeating all this $r-1$ times, for any bounded Borel measurable function $h : \mathbb{R} \mapsto \mathbb{R}$ we get

$$\frac{d^r}{z^r} E\Big(h\Big(z + \sum_{i=1}^{\tau} a_i \varepsilon_{1-i}\Big)\Big)$$
$$= \frac{(-1)^r}{a_1 \cdots a_r} \int_{\mathbb{R}^r} h\Big(z + \sum_{i=1}^{\tau} a_i t_i\Big) \Big[\prod_{i=1}^{r} H_\varepsilon''(t_i) \prod_{j=r+1}^{\tau} H_\varepsilon'(t_j)\Big] dt_1 \dots dt_r$$

for every $r = 0, \dots, m+2$. The statement then is justified by two applications of this equation with the bounded functions $h(z) = I\{R(z) > -\langle \mathbf{b}, \mathbf{x}\rangle\}$ and $h(z) = R(z)I\{R(z) > -\langle \mathbf{b}, \mathbf{x}\rangle\}$, $z \in \mathbb{R}$. Since the function $G(\cdot, \mathbf{x})$ is discontinuous for each $\mathbf{x} \in \mathbb{R}^d$, it is indeed the most general form of condition $\mathbb{C}_0^*(\mathbf{x}, m)$ that is needed for the Tobin model, with some $\tau \geq m+2$, as stressed above in general terms. Of course, the required boundedness in condition $\mathbb{C}_4$ is trivially satisfied if $R(\cdot)$ is bounded, and $\mathbb{C}_2$ holds whenever $S(\cdot)$ has a continuously differentiable density.

Conditions $\mathbb{C}_2(\mathbf{x}), \mathbb{C}_3(\mathbf{x})$ and $\mathbb{C}_5(\mathbf{x})$ are used in Theorem 1 here, and while $\mathbb{C}_2(\mathbf{x})$ is also required in Theorem 2, a stronger form of the smoothness condition $\mathbb{C}_3(\mathbf{x})$ and more smoothness in the vector variable of G is needed besides $\mathbb{C}_4(\mathbf{x})$, exactly as in Csörgő and Mielniczuk (1999). Letting $|\mathbf{y}|$ be the Euclidean norm of $\mathbf{y} \in \mathbb{R}^d$ and putting

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

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

$\mathbb{C}_6(\mathbf{x})$: $f(\mathbf{x}) > 0$ and f is twice differentiable in a neighborhood of $\mathbf{x}$;

$\mathbb{C}_7(\mathbf{x})$: For each $z \in \mathbb{R}$ outside a set of Lebesgue measure zero, the function $G(z, \cdot)$ is twice differentiable in a neighborhood of $\mathbf{x}$ such that $E(\sup\{|G'(Z, \mathbf{y})|^2 : |\mathbf{y} - \mathbf{x}| \leq \delta\}) < \infty$ and $E(\sup\{|G''(Z, \mathbf{y})| : |\mathbf{y} - \mathbf{x}| \leq \delta\}) < \infty$ for some $\delta = \delta(\mathbf{x}) > 0$.

We note that condition $\mathbb{C}_7(\mathbf{x})$ is satisfied for the Tobin model at every $\mathbf{x}$ if the distribution function $S(\cdot)$ of $W = R(Z)$ has a bounded second derivative and the level sets $\{z \in \mathbb{R} : R(z) = a\}$ have Lebesgue measure zero for all $a \in \mathbb{R}$.

For a Borel measurable function $h : \mathbb{R} \mapsto \mathbb{R}$, set $h_\infty(z) = E(h(z+Z))$, $z \in \mathbb{R}$, and

$$m_h = \min_{k \in \mathbb{N}} \Big\{k : h_\infty^{(k)}(0) \text{ exists, } h_\infty^{(j)}(0) = 0 \text{ for } j = 1, \dots, k-1, \text{ and } h_\infty^{(k)}(0) \neq 0\Big\},$$

when meaningful. Ho and Hsing (1997) call m_h the power rank of h with respect to the distribution of Z. As they point out, if Z is standard normal and $E(h^2(Z)) < \infty$, then m_h is the Hermite rank of h defined as $\min_{k\in\mathbb{N}}\{k : E(H_k(Z)h(Z)) \neq 0\}$, with H_k being the k^{th} Hermite polynomial. Hermite ranks are used in the random-design regression context by Csörgő and Mielniczuk (1999). Now for the G in (1.1) and $\mathbf{x} \in \mathbb{R}^d$, let $m(\mathbf{x})$ denote the power rank of $G(\cdot, \mathbf{x})$ with respect to the distribution of Z, and consider $G_\infty(z, \mathbf{x}) = E(G(z + Z, \mathbf{x}))$ with derivatives $G_\infty^{(r)}(z, \mathbf{x}) = \partial^r G_\infty(z, \mathbf{x})/\partial z^r$, $z \in \mathbb{R}$.

To motivate part of the development, we note that, as applied to the present function $G(\cdot, \mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^d$, Theorem 3.1 of Ho and Hsing (1997) states that if $E(\varepsilon^8) < \infty$, $E(G(Z, \mathbf{x})) = 0$, $E(G^2(Z, \mathbf{x})) < \infty$ and condition $\mathbb{C}_0(\mathbf{x}, p)$ is satisfied for some $p \in \mathbb{N}$ for which $p\alpha < 1$, then for every $\gamma > 0$,

$$E\Big(\Big[\sum_{k=1}^{n} G(Z_k, \mathbf{x}) - \sum_{r=1}^{p} G_\infty^{(r)}(0, \mathbf{x}) Y_{n,r}\Big]^2\Big) \leq C_{\mathbf{x}}(\gamma)\, \max(n, n^{2-(p+1)\alpha+\gamma}) \tag{2.3}$$

for a finite constant $C_{\mathbf{x}}(\gamma) > 0$.

For a fixed $\mathbf{x} \in \mathbb{R}^d$, consider now the functions

$$G_n(z, \mathbf{x}) = E(G(z, \mathbf{X})K_{b_n}(\mathbf{x} - \mathbf{X})) \text{ and } G_{n,\infty}(z, \mathbf{x}) = E(G_n(z + Z, \mathbf{x}))\,, \quad z \in \mathbb{R}\,,$$

with derivatives $G_n^{(r)}(z, \mathbf{x}) = \partial^r G_n(z, \mathbf{x})/\partial z^r$ and $G_{n,\infty}^{(r)}(z, \mathbf{x}) = \partial^r G_{n,\infty}(z, \mathbf{x})/\partial z^r$. If we assume condition $\mathbb{C}_0^*(\mathbf{x}, m)$ with some $\tau \geq 0$ and $U_{\mathbf{x}}$, then the function $G_\infty(z, \mathbf{y}) := E(G(z+Z, \mathbf{y})) = \int_{\mathbb{R}} G(z+v, \mathbf{y})\, dH(v)$ and its derivatives $G_\infty^{(r)}(z, \mathbf{y}) = \partial^r G_\infty(z, \mathbf{y})/\partial z^r$, $z \in \mathbb{R}$, are well defined for each $\mathbf{y} \in U_{\mathbf{x}}$; in fact, $G_\infty^{(r)}(z, \mathbf{y}) = \int_{\mathbb{R}} G_\tau^{(r)}(z + v, \mathbf{y})\, d\widetilde{H}_\tau(v)$ for every $z \in \mathbb{R}$ by an application of Lemma 1.1 in Ho and Hsing (1997), $r = 0, \ldots, m + 2$. Introducing the rectangle $R_n(\mathbf{x}) = [x_1 - b_n, x_1 + b_n] \times \cdots \times [x_d - b_n, x_d + b_n]$ with center $\mathbf{x} = (x_1, \ldots, x_d)$ and assuming also $\mathbb{C}_1(0)$,

$$\begin{aligned} G_{n,\infty}^{(r)}(z, \mathbf{x}) &= \int_{R_n(\mathbf{x})} G_\infty^{(r)}(z, \mathbf{y}) K_{b_n}(\mathbf{x} - \mathbf{y}) f(\mathbf{y})\, d\mathbf{y} \\ &= \int_{U_{\mathbf{x}}} G_\infty^{(r)}(z, \mathbf{y}) K_{b_n}(\mathbf{x} - \mathbf{y}) f(\mathbf{y})\, d\mathbf{y}\,, \quad z \in \mathbb{R}\,, \end{aligned} \tag{2.4}$$

for every $r = 0, \ldots, m + 2$ and $n \geq n_0(U_{\mathbf{x}}) := \min\{k \in \mathbb{N} : R_n(\mathbf{x}) \subset U_{\mathbf{x}}, n \geq k\}$. Here $n_0(U_{\mathbf{x}}) \in \mathbb{N}$ since $b_n \to 0$, by Fubini's theorem and the fact that differentiation and integration can be interchanged under $\mathbb{C}_0^*(\mathbf{x}, m)$. By (1.1), $E(G(Z, \mathbf{X}) \,|\, \mathbf{X}) = E(\eta \,|\, \mathbf{X}) = 0$ almost surely. Then we see from (2.4) that $G_{n,\infty}^{(0)}(0, \mathbf{x}) = G_{n,\infty}(0, \mathbf{x}) = 0$ for every $\mathbf{x} \in \mathbb{R}^d$ and $n \in \mathbb{N}$ if, for example, $\mathbb{C}_1(0)$ holds.

The following result is crucial for Theorem 1, the case of large bandwidths. It is based on an extension of Ho and Hsing's (1997) Theorem 3.1 in (2.3), which we state at the beginning of the proof of the lemma. The extension itself is obtained by simply monitoring an issue of uniformity in the original proof.

Lemma 1. *Suppose* $E(\varepsilon^8) < \infty$, *conditions* $\mathbb{C}_1(0)$, $\mathbb{C}_5(\mathbf{x})$ *and* $\mathbb{C}_0^*(\mathbf{x}, p)$ *hold for some* $\mathbf{x} \in \mathbb{R}^d$ *and some* $p \in \mathbb{N}$, *and the density function* $f(\cdot)$ *is bounded in a neighborhood of this* $\mathbf{x}$. *If* $p\alpha < 1$, *then for every* $\gamma > 0$,

$$E\Big(\Big[\sum_{k=1}^{n} G_n(Z_k, \mathbf{x}) - \sum_{r=1}^{p} G_{n,\infty}^{(r)}(0, \mathbf{x}) Y_{n,r}\Big]^2\Big) \le C_{\mathbf{x}}^*(\gamma) \max(n, n^{2-(p+1)\alpha+\gamma})$$

for a finite constant $C_{\mathbf{x}}^*(\gamma) > 0$.

The next lemma is also a consequence of (2.3). It gives an upper bound for the stochastic order of $\sum_{i=1}^{n} G(Z_i, \mathbf{x})$ when $m(\mathbf{x})\alpha \ge 1$, needed in the proof of Theorem 2*.

Lemma 2. *If* $E(\varepsilon^8) < \infty$, $E(G(Z, \mathbf{x})) = 0$, $E(G^2(Z, \mathbf{x})) < \infty$ *and condition* $\mathbb{C}_0(\mathbf{x}, m(\mathbf{x}))$ *is satisfied for the power rank* $m(\mathbf{x}) \ge 2$ *of* $G(\cdot, \mathbf{x})$, *and if* $m(\mathbf{x})\,\alpha \ge 1$ *for the given* $\mathbf{x} \in \mathbb{R}^d$, *then for every* $\gamma > 0$,

$$M_n^2(\mathbf{x}) := E\Big(\Big[\sum_{k=1}^{n} G(Z_k, \mathbf{x})\Big]^2\Big) \le C_{\mathbf{x}}(\gamma)\, n^{1+\gamma}$$

for the constant $C_{\mathbf{x}}(\gamma)$ *in* (2.3).

3. Results

Let the different points $\mathbf{x}_1, \ldots, \mathbf{x}_l \in \mathbb{R}^d$ be given. In all four theorems below (and also in Theorem 1*, not stated in detail) we assume that the respective power ranks $m(\mathbf{x}_1), \ldots, m(\mathbf{x}_l)$ of the l functions $G(\cdot, \mathbf{x}_1), \ldots, G(\cdot, \mathbf{x}_l)$ are well defined with respect to the distribution of Z, and from now on we use the symbol $m := \min(m(\mathbf{x}_1), \ldots, m(\mathbf{x}_l))$ for the smallest power rank pertaining to the points $\mathbf{x}_1, \ldots, \mathbf{x}_l$. Recall (2.1) and that $f_n(\cdot) = E(\widehat{f}_n(\cdot))$, where $f_n(\mathbf{x}) \to f(\mathbf{x})$ under $\mathbb{C}_3(\mathbf{x})$ and $\mathbb{C}_1(\kappa)$ for any $\kappa \ge 2$.

One part of the smoothing dichotomy, for "large" bandwidths that allow the long memory of the errors in (1.1) to prevail, is expressed by the first result.

Theorem 1. *Suppose that condition* $\mathbb{C}_1(\kappa)$ *holds for some* $\kappa \in \{2, 3, \ldots\}$, *conditions* $\mathbb{C}_2(\mathbf{x}_j)$, $\mathbb{C}_3(\mathbf{x}_j)$, $\mathbb{C}_5(\mathbf{x}_j)$ *and* $\mathbb{C}_0^*(\mathbf{x}_j, m)$ *hold for all* $j = 1, \ldots, l$ *and the smallest power rank* $m \in \mathbb{N}$, *and that* $E(\varepsilon^{\max(8,2m)}) < \infty$. *If* $m\alpha < 1$, *the sequence* $\{nb_n^{d+4}\}$ *is bounded,* $n^{m\alpha} L^{-m}(n) = o(nb_n^d)$ *and*

$$\frac{n^{(m-r)\alpha/2}}{L^{(m-r)/2}(n)}\, E\Big([G_\infty^{(r)}(0, \mathbf{X}) - G_\infty^{(r)}(0, \mathbf{x}_j)] K_{b_n}(\mathbf{x}_j - \mathbf{X})\Big) \to 0\,, \quad r = 1, \ldots, m\,, \tag{3.1}$$

for all $j = 1, \ldots, l$, *then, with* $a_{n,m}$ *as in* (2.2),

$$a_{n,m}\Big(\widehat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \ldots, \widehat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)\Big) \xrightarrow{\mathcal{D}} Y_m^*\Big(G_\infty^{(m)}(0, \mathbf{x}_1), \ldots, G_\infty^{(m)}(0, \mathbf{x}_l)\Big).$$

Concerning the size requirements on the bandwidth sequence $\{b_n\}$, we refer to the discussion in Csörgő and Mielniczuk (1999), from which the analysis of the other conditions can also be extended to corresponding versions in this paper.

Observe that if $\mathbf{x}_j$ for some $j \in \{1, \ldots, l\}$ is such that $m(\mathbf{x}_j) > m$ then it follows by Theorem 1 that $a_{n,m}[\hat{g}_n(\mathbf{x}_j) - g(\mathbf{x}_j)] \xrightarrow{P} 0$. However, if the conditions of the theorem are satisfied with m replaced by $m(\mathbf{x})$ for a single $\mathbf{x} \in \mathbb{R}^d$, the result applied with $l = 1$ yields $a_{n,m(\mathbf{x})}[\widehat{g}_n(\mathbf{x}_j) - g(\mathbf{x}_j)] \xrightarrow{\mathcal{D}} G_\infty^{(m(\mathbf{x}))}(0, \mathbf{x}) Y_{m(\mathbf{x})}^*$, where the limit is nondegenerate since $G_\infty^{(m(\mathbf{x}))}(0, \mathbf{x}) \neq 0$. If we proved only the latter conclusion, then of course its l-fold application could not give the joint asymptotic distribution in the theorem.

Special attention must be paid to condition (3.1). Notice first that Lemma 1 yields $a_{n,m}[\sum_{i=1}^n G_n(Z_k, \mathbf{x}_j) - \sum_{r=1}^m G_{n,\infty}^{(r)}(0, \mathbf{x}_j) Y_{n,r}]/n \xrightarrow{\mathcal{P}} 0$ for all $j = 1, \ldots, l$. However, it may happen that $G_{n,\infty}^{(r)}(0, \mathbf{x}_j) \neq 0$ for some n and $r < m$ even though $G_\infty^{(r)}(0, \mathbf{x}_j) = 0$, and in this case $a_{n,m} G_{n,\infty}^{(r)}(0, \mathbf{x}_j) Y_{n,r}/n$ is not necessarily $o_P(1)$. Thus the role of condition (3.1) is to make $a_{n,m} \sum_{r=1}^{m-1} G_{n,\infty}^{(r)}(0, \mathbf{x}_j) Y_{n,r}/n$ asymptotically negligible. On the other hand, to understand the condition, observe also that since by (2.4),

$$\begin{aligned} E\Big([G_\infty^{(r)}(0, \mathbf{X}) - G_\infty^{(r)}(0, \mathbf{x})] K_{b_n}(\mathbf{x} - \mathbf{X})\Big) &= G_{n,\infty}^{(r)}(0, \mathbf{x}) - G_\infty^{(r)}(0, \mathbf{x}) f_n(\mathbf{x}) \\ &= \int_{U_\mathbf{x}} \Big[G_\infty^{(r)}(0, \mathbf{y}) - G_\infty^{(r)}(0, \mathbf{x})\Big] f(\mathbf{y}) K_{b_n}(\mathbf{x} - \mathbf{y})\, d\mathbf{y} \end{aligned} \tag{3.2}$$

for all $n \geq n_0(U_\mathbf{x})$ under $\mathbb{C}_0^*(\mathbf{x}, m)$, condition (3.1) is in fact a problem for deterministic kernel estimation, which depends only upon the smoothness of the $m+1$ functions $G_\infty^{(1)}(0, \mathbf{y}), \ldots, G_\infty^{(m)}(0, \mathbf{y})$ and $f(\mathbf{y})$ with $\mathbf{y}$ near the points $\mathbf{x}_1, \ldots, \mathbf{x}_l$. It is for the sake of this problem that we entertain kernels with some order κ, i.e. use condition $\mathbb{C}_1(\kappa)$, rather than assuming simply as in Csörgő and Mielniczuk (1999) that K_0 is a symmetric density. It is for the sake of easy comparison between the two sets of conditions for Theorem 1, there and here, that we delineate condition (3.1) for separate discussion; with a normal sequence $\{Z_i\}$ in (1.1) no such extra condition enters the picture.

Notice first of all that for the case of $r = m$, condition (3.1) does not demand any rate of convergence. Hence, under $\mathbb{C}_0^*(\mathbf{x}_j, m)$, assuming only that the function $G_\infty^{(m)}(0, \cdot)$ is continuously differentiable in a neighborhood of $\mathbf{x}_j$ and condition $\mathbb{C}_3(\mathbf{x}_j)$ is also satisfied for $f(\cdot)$, $j = 1, \ldots, l$, we have (3.1) for $r = m$ whenever $\mathbb{C}_1(\kappa)$ holds for any order $\kappa \geq 2$. In particular, if $m = 1$ in Theorem 1, then with the only extra condition that $G_\infty^{(1)}(0,\cdot)$ is continuously differentiable in a neighborhood of $\mathbf{x}_j$, $j=1,\ldots,l$, the conclusion holds with the standard normal Y_1^* in the limit. To handle the general case, introduce

$$\nu_m(r)=\nu_m^{d,\alpha}(r)=\min\Big\{n\in\mathbb{N}:\ n>\alpha\,\frac{d+4}{2d}\,(m-r)\Big\},\quad r=1,\ldots,m,$$

and let $\kappa_m=\max(\nu_m(1),2)\in\mathbb{N}$. Clearly, $1=\nu_m(m)\le\nu_m(m-1)\le\cdots\le\nu_m(1)$ and $\kappa_1=2$. Of course, the full force of $\mathbb{C}_0^*(\mathbf{x}_j,m)$ is not needed to ensure (3.2) at $\mathbf{x}_j$, $j=1,\ldots,l$, but we assume it to avoid further conditions irrelevant to the present context. The condition on $f(\cdot)$ below reduces to $\mathbb{C}_3(\mathbf{x}_j)$, $j=1,\ldots,l$, whenever $\nu_m(1)=1$.

Proposition. *Suppose that condition $\mathbb{C}_0^*(\mathbf{x}_j,m)$ is satisfied, $f(\cdot)$ is $\nu_m(1)$ times continuously differentiable in a neighborhood of $\mathbf{x}_j$, $j=1,\ldots,l$, the kernel satisfies condition $\mathbb{C}_1(\kappa)$ with an order $\kappa\ge\kappa_m$ and the sequence $\{nb_n^{d+4}\}$ is bounded. If for each $r=1,\ldots,m$, the function $G_\infty^{(r)}(0,\cdot)$ is $\nu_m(r)$ times continuously differentiable in a neighborhood of $\mathbf{x}_j$, $j=1,\ldots,l$, then* (3.1) *holds.*

Direct smoothness conditions on $\partial^r G(z,\cdot)/\partial x^r$, $z\in\mathbb{R}$, ensure these conditions on $G_\infty^{(r)}(0,\cdot)$. However, we emphasize that under $\mathbb{C}_0^*(\mathbf{x},m)$ the function $G(\cdot,\mathbf{x})$ does not even have to be continuous to talk about the functions $G_\infty^{(r)}(0,\cdot)$. This point is important as we saw when discussing Tobin models in the previous section; this is what makes Theorem 1 applicable to these models.

We also note that in the case that $\{Z_i\}$ in (1.1) is Gaussian satisfying (1.2) for some $\alpha\in(0,1)$, Csörgő and Mielniczuk's (1999) Theorem 1* extends their Theorem 1 to a situation in which the sequence $\{\mathbf{X}_i\}$ of explanatory vectors is weakly dependent, under the assumption that the errors do not depend on the explanatory vectors, so that for a Borel function $G:\mathbb{R}\mapsto\mathbb{R}$, the modeling assumption in (1.1) simplifies to $Y_i=g(\mathbf{X}_i)+\eta_i=g(\mathbf{X}_i)+G(Z_i)$, $i=1,\ldots,n$. In this model, a corresponding result remains valid for the present case when $\{Z_i\}$ is a linear process: assuming the partially simplified conditions of the present Theorem 1 and condition (2.10) of Csörgő and Mielniczuk (1999) on the weak dependence of $\{\mathbf{X}_i\}$, the conclusion above holds with the limiting vector $Y_m^*(G_\infty^{(m)}(0),\ldots,G_\infty^{(m)}(0))$, provided $r_G(n):=E(G(Z_1)G(Z_{1+n}))=|r^m(n)|(1+o(1))$ and $\sum_{1\le i\ne j\le n}|r_G(j-i)|=\mathcal{O}(n^{2-m\alpha}L^m(n))$. Under further conditions on G, depending on the value of m, the latter proviso can be established by techniques of Ho and Hsing (1997), and hence Theorem 1* can be formulated for the present moving averages $\{Z_i\}$.

Setting $\sigma^2(\mathbf{x})=[\int K^2(\mathbf{y})\,d\mathbf{y}]\,E(G^2(Z,\mathbf{x}))/f(\mathbf{x})$, $\mathbf{x}\in\mathbb{R}^d$, where $E(G^2(Z,\mathbf{x}))=E(\eta^2\,|\,\mathbf{X}=\mathbf{x})$ and $\int K^2(\mathbf{y})\,d\mathbf{y}=[\int_{-1}^1 K_0^2(y)\,dy]^d$, we now come to the opposite part of the dichotomy for "small" bandwidths, when the Nadaraya–Watson estimator $\widehat{g}_n$ behaves as if the errors were independent, even though they are long-range dependent.

Theorem 2. *Suppose* $E(\varepsilon^{\max(8,2m)}) < \infty$, *condition* $\mathbb{C}_1(\kappa)$ *holds for some* $\kappa \in \{2, 3, \ldots\}$ *and conditions* $\mathbb{C}_2(\mathbf{x}_j)$, $\mathbb{C}_4(\mathbf{x}_j)$, $\mathbb{C}_6(\mathbf{x}_j)$, $\mathbb{C}_7(\mathbf{x}_j)$ *and* $\mathbb{C}_0(\mathbf{x}_j, m)$ *hold for all* $j = 1, \ldots, l$ *and the smallest power rank* $m \in \mathbb{N}$. *If* $m\alpha < 1$, $nb_n^{d+4} \to 0$, $nb_n^d \to \infty$, *but* $nb_n^d = o(n^{m\alpha} L^{-m}(n))$, *then*

$$\sqrt{nb_n^d}\left(\widehat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \ldots, \widehat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)\right) \xrightarrow{\mathcal{D}} \left(\sigma(\mathbf{x}_1)N_1, \ldots, \sigma(\mathbf{x}_l)N_l\right), \quad (3.3)$$

where $N_1, \ldots, N_l$ *are independent standard normal random variables.*

Note that Theorem 2 establishes the asymptotic independence of $\widehat{g}_n(\mathbf{x}_i)$ and $\widehat{g}_n(\mathbf{x}_j)$ for $i \neq j$ and hence is much stronger than its univariate special case. Of course, the small-bandwidth condition $nb_n^d = o(n^{m\alpha} L^{-m}(n))$ for the minimal rank m implies the same for all the individual ranks $m(\mathbf{x}_1), \ldots, m(\mathbf{x}_l)$, and, except for those that are equal to m, we can have $m(\mathbf{x}_j)\alpha \geq 1$, $j = 1, \ldots, l$, This is not surprising in view of Theorem 2* below. The next theorem describes the borderline case between "large" and "small" bandwidths, resulting in an asymptotic convolution of the two parts of the dichotomy.

Theorem 3. *Suppose* $E(\varepsilon^{\max(8,2m)}) < \infty$, *condition* $\mathbb{C}_1(\kappa)$ *holds for some* $\kappa \in \{2, 3, \ldots\}$ *and conditions* $\mathbb{C}_2(\mathbf{x}_j)$, $\mathbb{C}_4(\mathbf{x}_j)$, $\mathbb{C}_6(\mathbf{x}_j)$, $\mathbb{C}_7(\mathbf{x}_j)$ *and* $\mathbb{C}_0(\mathbf{x}_j, m)$ *hold for all* $j = 1, \ldots, l$ *and the smallest power rank* $m \in \mathbb{N}$. *If* $m\alpha < 1$, $nb_n^{d+4} \to 0$ *and* $nb_n^d / a_{n,m}^2 \to C_b^2$ *for some constant* $C_b \in (0, \infty)$, *then*

$$\sqrt{nb_n^d}\left(\widehat{g}_n(\mathbf{x}_1) - g(\mathbf{x}_1), \ldots, \widehat{g}_n(\mathbf{x}_l) - g(\mathbf{x}_l)\right)$$
$$\xrightarrow{\mathcal{D}} \left(C_b\, G_\infty^{(m)}(0, \mathbf{x}_1)\, Y_m^* + \sigma(\mathbf{x}_1)N_1, \ldots, C_b\, G_\infty^{(m)}(0, \mathbf{x}_l)\, Y_m^* + \sigma(\mathbf{x}_l)N_l\right),$$

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

Although the sequence $\{Z_i\}_{i=1}^\infty$ is always long-range dependent if $\beta \in (1/2, 1)$, so that $\alpha = 2\beta - 1 \in (0, 1)$ in (1.2), the transformed sequences $\{G(Z_i, \mathbf{x}_j)\}_{i=1}^\infty$, $j = 1, \ldots, l$, may lack long memory if $m\alpha \geq 1$ for their smallest power rank m. It is natural to expect in this case that the conclusion of Theorem 2 holds for *all* bandwidth sequences $\{b_n\}$ such that $nb_n^d \to \infty$ and $nb_n^{d+4} \to 0$. Indeed, this is the content of the next result.

Theorem 2*. *Suppose* $E(\varepsilon^8) < \infty$, *condition* $\mathbb{C}_1(\kappa)$ *holds for some* $\kappa \in \{2, 3, \ldots\}$ *and conditions* $\mathbb{C}_2(\mathbf{x}_j)$, $\mathbb{C}_4(\mathbf{x}_j)$, $\mathbb{C}_6(\mathbf{x}_j)$, $\mathbb{C}_7(\mathbf{x}_j)$ *and* $\mathbb{C}_0(\mathbf{x}_j, m(\mathbf{x}_j))$ *hold for the power ranks* $m(\mathbf{x}_j)$ *for all* $j = 1, \ldots, l$. *If* $m\alpha \geq 1$ *for the smallest power rank* $m \in \mathbb{N}$, $nb_n^d \to \infty$ *and* $nb_n^{d+4} \to 0$, *then* (3.3) *holds.*

Again, this result is a counterpart of Theorem 2* in Csörgő and Mielniczuk (1999), in which $\{Z_i\}_{i=1}^{\infty}$ is Gaussian and the sequences $\{G(Z_i,\mathbf{x}_j)\}_{i=1}^{\infty}$, $j = 1,\ldots,l$, all have short memory in an arbitrary fashion.

4. Proofs

Proof of Lemma 1. First we state an extension of Ho and Hsing's (1997) Theorem 3.1, for sums of a *sequence* of functions of $\{Z_k\}_{k=1}^{n}$ instead of sums of a single function. Let $n_0 \in \mathbb{N}$ and consider a sequence $\{\overline{G}_n(\cdot)\}_{n=n_0}^{\infty}$ of Borel measurable real functions on $\mathbb{R}$ with the associated functions $\overline{G}_{n,j}(z) = E(\overline{G}_n(z+\sum_{i=1}^{j} a_i\varepsilon_{1-i}))$, $j \in \{0\}\cup\mathbb{N}$, and $\overline{G}_{n,\infty}(z) = E(\overline{G}_n(z+Z))$, and their derivatives $\overline{G}_{n,j}^{(r)}(z)$ and $\overline{G}_{n,\infty}^{(r)}(z)$, $z\in\mathbb{R}$, $r=0,1,\ldots.$ Setting $Y_{n,0}=n$, a careful analysis of Ho and Hsing's (1997) Theorem 3.1 proof, including their Lemmas 6.1 and 6.2, reveals the following extension. *If* $E(\varepsilon^8)<\infty$, $\sup_{n\ge n_0} E(\overline{G}_n^2(Z))<\infty$, $\sup_{n\ge n_0}\max_{1\le r\le p}|\overline{G}_{n,\infty}^{(r)}(0)|<\infty$ *and* $\sup_{n\ge n_0} C_{\tau,\lambda}^{p}(\overline{G}_n;z)<\infty$ *at every* $z\in\mathbb{R}$ *for some* $\tau\in\{0\}\cup\mathbb{N}$, $\lambda>0$ *and some* $p\in\mathbb{N}$ *for which* $p\alpha<1$, *then for every* $\gamma>0$,

$$E\Big(\Big[\sum_{k=1}^{n}\overline{G}_n(Z_k)-\sum_{r=0}^{p}\overline{G}_{n,\infty}^{(r)}(0)Y_{n,r}\Big]^2\Big)\le C(\gamma)\max(n,n^{2-(p+1)\alpha+\gamma})$$

for a finite constant $C(\gamma)>0$.

The lemma will follow if we show that under the stated conditions at $\mathbf{x}\in\mathbb{R}^d$, the three assumptions above are satisfied for the choice $\overline{G}_n(\cdot)=G_n(\cdot,\mathbf{x})$, $n\ge n_0(\mathbf{x}) := \max(n_0(U_{\mathbf{x}}),n_0(V_{\mathbf{x}}))$, Here $n_0(U_{\mathbf{x}})$ is the threshold number defined at (2.4), $n_0(V_{\mathbf{x}})$ is the analogous threshold pertaining to $V_{\mathbf{x}}$ in condition $\mathbb{C}_5(\mathbf{x})$ and

$$G_n(z,\mathbf{x})=\int_{R_n(\mathbf{x})}G(z,\mathbf{y})K_{b_n}(\mathbf{x}-\mathbf{y})f(\mathbf{y})d\mathbf{y}=\int_{U_{\mathbf{x}}\cap V_{\mathbf{x}}}G(z,\mathbf{y})K_{b_n}(\mathbf{x}-\mathbf{y})f(\mathbf{y})d\mathbf{y},\ z\in\mathbb{R},$$

for every $n\ge n_0(\mathbf{x})$, as an analogue of (2.4), and where we also assume without loss of generality that both $U_{\mathbf{x}}$ and $V_{\mathbf{x}}$ are contained in that neighborhood of $\mathbf{x}$ where $f(\cdot)$ is bounded. First,

$$\begin{aligned}\sup_{n\ge n_0(V_{\mathbf{x}})} E(G_n^2(Z,\mathbf{x})) &= \sup_{n\ge n_0(V_{\mathbf{x}})}\int_{\mathbb{R}}\Big[\int_{V_{\mathbf{x}}}G(z,\mathbf{y})K_{b_n}(\mathbf{x}-\mathbf{y})f(\mathbf{y})\,d\mathbf{y}\Big]^2 dH(z)\\ &\le\Big\{\int_{\mathbb{R}}\Big[\sup_{\mathbf{y}\in V_{\mathbf{x}}}|G(z,\mathbf{y})|\Big]^2 dH(z)\Big\}\Big\{\int_{V_{\mathbf{x}}}|K_{b_n}(\mathbf{x}-\mathbf{y})|f(\mathbf{y})\,d\mathbf{y}\Big\}^2\\ &\le f_{V_{\mathbf{x}}}^2 C_K^2 E\Big(\Big[\sup_{\mathbf{y}\in V_{\mathbf{x}}}|G(Z,\mathbf{y})|\Big]^2\Big)<\infty\end{aligned}$$

by condition $\mathbb{C}_5(\mathbf{x})$, where $f_U = \sup_{\mathbf{y}\in U} f(\mathbf{y})$ for $U \subset \mathbb{R}^d$ and $C_K = \int_{[-1,1]^d} |K(\mathbf{w})|d\mathbf{w}$. Similarly, from (2.4),

$$\sup_{n\geq n_0(U_{\mathbf{x}})} \max_{1\leq r\leq m} |G_{n,\infty}^{(r)}(0,\mathbf{x})| \leq f_{U_{\mathbf{x}}} C_K \max_{1\leq r\leq m} \sup_{\mathbf{y}\in U_{\mathbf{x}}} \int_{\mathbb{R}} |G_{\tau}^{(r)}(v,\mathbf{y})|\, d\widetilde{H}_\tau(v) < \infty$$

for the τ in condition $\mathbb{C}_0^*(\mathbf{x}, m)$.

Finally, introducing $G_{n,j}(z,\mathbf{x}) = E(G_n(z + \sum_{i=1}^{j} a_i\varepsilon_{1-i}, \mathbf{x}))$, $j \in \{0\} \cup \mathbb{N}$, so that $G_{n,0}(z,\mathbf{x}) = G_n(z,\mathbf{x})$, we see by Fubini's theorem that for the τ in condition $\mathbb{C}_0^*(\mathbf{x}, m)$ and for all $n \geq n_0(U_{\mathbf{x}})$ and $z \in \mathbb{R}$, we have

$$\begin{aligned} G_{n,\tau}(z,\mathbf{x}) &= \int_{\mathbb{R}} \Big[\int_{U_{\mathbf{x}}} G(z+v,\mathbf{y}) K_{b_n}(\mathbf{x}-\mathbf{y}) f(\mathbf{y})\, d\mathbf{y} \Big] dH_\tau(v) \\ &= \int_{U_{\mathbf{x}}} G_\tau(z,\mathbf{y}) K_{b_n}(\mathbf{x}-\mathbf{y}) f(\mathbf{y})\, d\mathbf{y}\,. \end{aligned}$$

Thus, since differentiation and integration can again be interchanged under $\mathbb{C}_0^*(\mathbf{x}, m)$,

$$\begin{aligned} \sup_{n\geq n_0(U_{\mathbf{x}})} C_{\tau,\lambda}^m(G_n(\cdot,\mathbf{x}); z) &\leq \sup_{\mathbf{y}\in U_{\mathbf{x}}} C_{\tau,\lambda}^m(G(\cdot,\mathbf{y}); z) \int_{U_{\mathbf{x}}} |K_{b_n}(\mathbf{x}-\mathbf{y})| f(\mathbf{y})\, d\mathbf{y} \\ &\leq f_{U_{\mathbf{x}}} C_K \sup_{\mathbf{y}\in U_{\mathbf{x}}} C_{\tau,\lambda}^m(G(\cdot,\mathbf{y}); z) \end{aligned}$$

for every $z \in \mathbb{R}$. Thus if $\mathbb{C}_1(0)$ holds and $b_n \to 0$, the locally uniform condition $\mathbb{C}_0^*(\mathbf{x}, m)$ for G implies the condition $\mathbb{C}_0(\mathbf{x}, m)$ for G_n uniformly in n.

Proof of Theorem 1. The distributional convergence in (2.2) and inspection of the proof of Theorem 1 in Csörgő and Mielniczuk (1999) shows that it suffices to prove that

$$\Delta_{n,m}(\mathbf{x}) := \Big| \frac{a_{n,m}}{n} \sum_{k=1}^{n} G_n(Z_k,\mathbf{x}) - G_\infty^{(m)}(0,\mathbf{x}) f_n(\mathbf{x}) \frac{a_{n,m}}{n} Y_{n,m} \Big| \xrightarrow{P} 0\,, \qquad (4.1)$$

where $\xrightarrow{P}$ denotes convergence in probability and $\mathbf{x}$ is any one of $\mathbf{x}_1, \ldots, \mathbf{x}_l$; the desired convergence in (4.1) is the counterpart of equation (4.5) in that paper. Indeed, the present condition $\mathbb{C}_5(\mathbf{x})$ is somewhat stronger than condition $\mathbb{C}_4(\mathbf{x})$ there. The only other difference in the stipulated regularities, used in the proof to yield the sufficiency of (4.1) in the present context, is that condition $\mathbb{C}_1$ there requires that K_0 be a symmetric *density*, and hence the present condition $\mathbb{C}_1(\kappa)$ is weaker. However, $\mathbb{C}_1(\kappa)$ suffices both there and here since what is needed is the consistency of $\widehat{f}_n(\mathbf{x})$, which under $\mathbb{C}_3(\mathbf{x})$ follows for any kernel of order κ satisfying the rest of condition $\mathbb{C}_1(\kappa)$, by a simple adaptation of the proof of Devroye and Wagner's (1979) theorem.

To show (4.1), we write $\Delta_{n,m}(\mathbf{x}) \le \Delta_{n,m}^{(1)}(\mathbf{x}) + \Delta_{n,m}^{(2)}(\mathbf{x}) + \Delta_{n,m}^{(3)}(\mathbf{x})$, where, using $G_\infty^{(r)}(0,\mathbf{x}) = 0$ for all $r = 1,\dots,m-1$,

$$\Delta_{n,m}^{(1)}(\mathbf{x}) = \frac{a_{n,m}}{n}\Big|\sum_{k=1}^{n} G_n(Z_k,\mathbf{x}) - \sum_{r=1}^{m} G_{n,\infty}^{(r)}(0,\mathbf{x})Y_{n,r}\Big|,$$

$$\Delta_{n,m}^{(2)}(\mathbf{x}) = \sum_{r=1}^{m-1}\Big|G_{n,\infty}^{(r)}(0,\mathbf{x})\Big|\,\frac{a_{n,m}}{a_{n,r}}\Big|\frac{a_{n,r}}{n}Y_{n,r}\Big|$$
$$= \sum_{r=1}^{m-1}\frac{C_m(\beta)}{C_r(\beta)}\Big|G_{n,\infty}^{(r)}(0,\mathbf{x}) - G_\infty^{(r)}(0,\mathbf{x})f_n(\mathbf{x})\Big|\frac{n^{(m-r)\alpha/2}}{L^{(m-r)/2}(n)}\Big|\frac{a_{n,r}}{n}Y_{n,r}\Big|\xrightarrow{P}0$$

by (2.2) with $r = 1,\dots,m-1$ replacing m and by condition (3.1), and

$$\Delta_{n,m}^{(3)}(\mathbf{x}) = \Big|G_{n,\infty}^{(m)}(0,\mathbf{x}) - G_\infty^{(m)}(0,\mathbf{x})f_n(\mathbf{x})\Big|\,\Big|\frac{a_{n,m}}{n}\,Y_{n,m}\Big| \xrightarrow{P} 0,$$

also by (2.2) and the case $r = m$ of condition (3.1). Finally, choosing $\gamma \in (0,\alpha)$ in Lemma 1 and using Chebyshev's inequality, we have

$$P\Big\{\Delta_{n,m}^{(1)}(\mathbf{x}) \ge \theta\Big\} \le \frac{C_{\mathbf{x}}^*(\gamma)C_m^2(\beta)}{\theta^2 L^m(n)}\,\max\Big(\frac{1}{n^{1-m\alpha}}\,,\frac{1}{n^{\alpha-\gamma}}\Big) \to 0$$

for every $\theta > 0$, proving the theorem.

Proof of the Proposition. Let $\mathbf{x}$ be any one of $\mathbf{x}_1,\dots,\mathbf{x}_l$. By (3.2) we write

$$G_{n,\infty}^{(r)}(0,\mathbf{x}) - G_\infty^{(r)}(0,\mathbf{x})f_n(\mathbf{x})$$
$$= \int_{[0,1]^d}\Big[G_\infty^{(r)}(0,\mathbf{x}-b_n\mathbf{w}) - G_\infty^{(r)}(0,\mathbf{x})\Big]f(\mathbf{x}-b_n\mathbf{w})K(\mathbf{w})\,d\mathbf{w}$$

for all n large enough. Expanding $G_\infty^{(r)}(0,\mathbf{x}-b_n\mathbf{w})$ and $f(\mathbf{x}-b_n\mathbf{w})$ about $\mathbf{x}$ to $\nu_m(r)$ and $\nu_m(1)$ terms, respectively, and using that the order of K_0 is at least κ_m, it is routine to see that

$$\Big|G_{n,\infty}^{(r)}(0,\mathbf{x}) - G_\infty^{(r)}(0,\mathbf{x})f_n(\mathbf{x})\Big| = \mathcal{O}\Big(b_n^{d\nu_m(r)}\Big), \quad r = 1,\dots,m\,.$$

Thus (3.1) will follow if we show that

$$\frac{n^{(m-r)\alpha/2}}{L^{(m-r)/2}(n)}\,b_n^{d\nu_m(r)} \to 0\,, \quad r = 1,\dots,m\,.$$

This is true for $r = m$. Let $m > 1$ and $r = 1,\dots,m-1$. Since $nb_n^{d+4} \le C^{d+4}$ (and so $b_n \le C/n^{1/(d+4)}$ for all $n \in \mathbb{N}$, for some constant $C > 0$) and

$$\nu_m(r) > \alpha\,\frac{d+4}{2d}\,(m-r) \quad \text{implies} \quad \mu_m(r) := \frac{2d\nu_m(r)}{\alpha(m-r)} - (d+4) > 0\,,$$

it follows that

$$\frac{nb_n^{2d\nu_m(r)/[\alpha(m-r)]}}{L^{1/\alpha}(n)} \le \frac{C^{\mu_m(r)+d+4}}{L^{1/\alpha}(n)} \frac{1}{n^{\mu_m(r)/(d+4)}} \to 0\,.$$

Raising this to the power of $(m-r)\alpha/2$, the desired convergence follows.

Proof of Theorem 2. Since $\mathbb{C}_6(\mathbf{x}_j)$ is stronger than $\mathbb{C}_3(\mathbf{x}_j)$, inspection of the proof of Theorem 2 in Csörgő and Mielniczuk (1999) reveals that if $nb_n^d \to \infty$ and $nb_n^{d+4} \to 0$ for the bandwidths, then under the conditions $\mathbb{C}_2(\mathbf{x}_j)$, $\mathbb{C}_4(\mathbf{x}_j)$, $\mathbb{C}_6(\mathbf{x}_j)$, $\mathbb{C}_7(\mathbf{x}_j)$ and $\mathbb{C}_1(\kappa)$ for some integer $\kappa \ge 2$ (the last of which and $\mathbb{C}_6(\mathbf{x}_j)$ also make $\widehat{f}_n(\mathbf{x}_j)$ consistent for $f(\mathbf{x}_j)$ as noted above, $j = 1,\ldots,l$), the conclusion of Theorem 2 there and here holds true for *all* stationary sequences $\{Z_k\}_{k=1}^\infty$ which satisfy only two additional conditions. One is that the sequence $\{h(Z_k)\}_{k=1}^\infty$ is ergodic for ($5l$ choices of) a non-negative Borel measurable $h : \mathbb{R} \mapsto \mathbb{R}$ such that $E(h(Z)) < \infty$, while the other is that for each $j = 1,\ldots,l$, the sequence $\{n^{-1}A_n \sum_{k=1}^n G(Z_k,\mathbf{x}_j)\}_{n=1}^\infty$ is stochastically bounded for a numerical sequence $A_n \to \infty$ for which $nb_n^d/A_n^2 \to 0$.

In our present situation, it is well known that the linear process $\{Z_k\}_{k=1}^\infty$ is ergodic as a "moving function" of the ergodic stationary sequence $\{\varepsilon_i\}_{i=-\infty}^\infty$ of independent and identically distributed variables, and hence the sequence $\{h(Z_k)\}_{k=1}^\infty$ is also ergodic.

Also, assuming $E(G(Z,\mathbf{x}_1)) = \cdots = E(G(Z,\mathbf{x}_l)) = 0$ and that the l conditions $\mathbb{C}_0(\mathbf{x}_1,m),\ldots,\mathbb{C}_0(\mathbf{x}_l,m)$ hold, $m\alpha < 1$ and $E(\varepsilon^{\max(8,2m)}) < \infty$, the first statement of Corollary 3.3 of Ho and Hsing (1997) implies that

$$\frac{a_{n,m}}{n}\Big(\sum_{k=1}^n G(Z_k,\mathbf{x}_1),\ldots,\sum_{k=1}^n G(Z_k,\mathbf{x}_l)\Big) \xrightarrow{\mathcal{D}} Y_m^*\Big(G_\infty^{(m)}(0,\mathbf{x}_1),\ldots,G_\infty^{(m)}(0,\mathbf{x}_l)\Big), \tag{4.2}$$

with the $a_{n,m}$ as before. Indeed, for any $j = 1,\ldots,l$, convergence in the jth component follows directly from that statement if $m(\mathbf{x}_j) = m$. If $m(\mathbf{x}_j) > m$, then, applying (2.3) with $p = m$ and $\gamma \in (0,\alpha)$,

$$E\Big(\Big[\frac{a_{n,m}}{n}\sum_{k=1}^n G(Z_k,\mathbf{x}_j)\Big]^2\Big) \le \frac{C_{\mathbf{x}_j}(\gamma)C_m^2(\beta)}{L^m(n)} \max\Big(\frac{1}{n^{1-m\alpha}}\,,\frac{1}{n^{\alpha-\gamma}}\Big) \to 0\,,$$

but because $G_\infty^{(m)}(0,\mathbf{x}_j) = 0$, the claimed convergence in the jth component holds again.

Since we assumed $nb_n^d/a_{n,m}^2 \to 0$, the second additional condition holds by (4.2) with the choice of $A_n \equiv a_{n,m}$.

Proof of Theorem 3. The proof of Theorem 3 in Csörgő and Mielniczuk (1999) remains valid if (4.2) is substituted for the present sequence $\{Z_k\}_{k=1}^\infty$.

Proof of Lemma 2. Consider $m_\alpha(\mathbf{x}) = \max\{k \in \{1,\dots,m(\mathbf{x})-1\} : k\alpha < 1\}$. Distinguishing the two cases $m(\mathbf{x}) - 1 < \frac{1}{\alpha} \le m(\mathbf{x})$ and $1 < \frac{1}{\alpha} \le m(\mathbf{x}) - 1$, where the second case can occur only if $m(\mathbf{x}) \ge 3$, we see that $m_\alpha(\mathbf{x}) = m(\mathbf{x}) - 1$ in the first case, while in the second case $m_\alpha(\mathbf{x}) = \frac{1}{\alpha} - 1$ if $\frac{1}{\alpha} \in \mathbb{N}$ and, with $\lfloor\cdot\rfloor$ standing for integer part, $m_\alpha(\mathbf{x}) = \lfloor\frac{1}{\alpha}\rfloor > \frac{1}{\alpha} - 1$ if $\frac{1}{\alpha} \notin \mathbb{N}$. Hence in both cases $m_\alpha(\mathbf{x}) \le m(\mathbf{x}) - 1$ and $(m_\alpha(\mathbf{x}) + 1)\,\alpha \ge 1$. Since $G_\infty^{(r)}(0,\mathbf{x}) = 0$ for all $r = 1,\dots,m(\mathbf{x})-1$ by the definition of the rank $m(\mathbf{x})$, denoting by $D_n^2(p,\mathbf{x})$ the left-hand side of (2.3), we have

$$M_n^2(\mathbf{x}) = D_n^2(m_\alpha(\mathbf{x}),\mathbf{x}) \le C_{\mathbf{x}}(\gamma)\,\max(n, n^{2-(m_\alpha(\mathbf{x})+1)\alpha+\gamma}) \le C_{\mathbf{x}}(\gamma)\,n^{1+\gamma}$$

for every $\gamma > 0$.

Proof of Theorem 2*. Uniting the opening discussion in the proof of Theorem 2 above and that of Theorem 2* in Csörgő and Mielniczuk (1999), it suffices to show that

$$W_n(\mathbf{x}_j) := \frac{\sqrt{nb_n^d}}{n}\sum_{k=1}^n G(Z_k,\mathbf{x}_j) \xrightarrow{P} 0 \quad \text{for each } j = 1,\dots,l.$$

Since $m(\mathbf{x}_j)\,\alpha \ge 1$ and $nb_n^{d+4} \to 0$, by Lemma 2 we obtain

$$E(W_n^2(\mathbf{x}_j)) = \frac{b_n^d}{n}\,M_n^2(\mathbf{x}_j) \le C_{\mathbf{x}_j}(d/(d+4))\,n^{d/(d+4)}\,b_n^d \to 0\,,$$

so the desired convergence follows for each $j = 1,\dots,l$.

Acknowledgements

S. Csörgő was supported in part by the National Science Foundation, USA, Grant DMS–96-25732 and by the National Scientific Program for Basic Research, Hungary, Grant T 032025, while J. Mielniczuk was supported in part by the Committee for Scientific Research, Poland, Grant 2 P301 001 06.

References

Avram, F. and Taqqu, M. S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.* **15**, 767-775.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods.* Springer, Berlin. [Second Edition: Springer, New York, 1991.]

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. (1999). Random-design regression under long-range dependent errors. *Bernoulli* **5**, 209-224.

Devroye, L. P. and Wagner, T. J. (1979). The L^1-consistency of kernel density estimates. *Ann. Statist.* **7**, 1136-1139.

Hall, P., Lahiri, S. N. and Polzehl, J. (1995) On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.* **23**, 1921-1936.

Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24**, 992-1024.

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25**, 1636-1669.

Mielniczuk, J. (1997). Long and short-range dependent sums of infinite-order moving averages and regression estimation. *Acta Sci. Math.* (*Szeged*) **63**, 301-316.

Department of Statistics, University of Michigan, 1440 Mason Hall, Ann Arbor, MI 48109–1027, U.S.A.

Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, Szeged, Hungary–6720.

Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01–237 Warsaw, Poland.

E-mail: miel@ipipan.waw.pl

(Received June 1998; accepted October 1999)