RANDOMIZED FIXED DESIGN REGRESSION UNDER LONG-RANGE DEPENDENT ERRORS

Artur Bryk[1] and Jan Mielniczuk[1,2]

[1]Warsaw University of Technology and [2]Polish Academy of Sciences, Warsaw, Poland

ABSTRACT

We consider a fixed-design regression model with long-range dependent errors and introduce an artificial randomization of grid points at which observations are taken in order to diminish the impact of strong dependence. The resulting estimator is shown to exhibit smoothing dichotomy with the variance in both cases diminishing more quickly than in the fixed design case.

Keywords: Long- and short-range dependence; randomization; kernel estimator; linear process; random- and fixed-design regression; convergence of moments, smoothing dichotomy

1. INTRODUCTION

Consider a fixed-design regression model (FDR)

$$Y_{i,n} = g(i/n) + \varepsilon_{i,n}, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where $g : [0,1] \to \mathbb{R}$ is some function with smoothness properties described later. For each n, we observe the random variables $Y_{1,n}, Y_{2,n}, \ldots, Y_{n,n}$ and the aim is to estimate the unknown function g based on this information. Here $(\varepsilon_{i,n})$ is the triangular array such that for each n, the finite sequence $\{\varepsilon_{i,n}\}_{i=1}^n$ is stationary, $\mathbb{E}\varepsilon_{i,n} = 0$, $\mathbb{E}\varepsilon_{i,n}^2 = \sigma_\varepsilon^2 > 0$, $r(k) := \mathrm{Cov}(\varepsilon_{i,n}, \varepsilon_{i+k,n}) = L(k)k^{-\alpha}$, $k = 1, 2, \ldots$, where $0 < \alpha < 1$ is a fixed constant and $L(\cdot)$ is a function defined on $[\,0, +\infty)$, slowly varying at infinity and positive in some neighborhood of infinity. We also assume that $r(k) < r(0)$ for $k \neq 0$. The array $(\varepsilon_{i,n})$ is long-range dependent (LRD) in the sense that $\sum_{k=1}^\infty |r(k)| = \infty$. In the following we suppress the dependence of $Y_{i,n}$ and $\varepsilon_{i,n}$ on n . In the paper we focus on the kernel regression function estimator $\hat{g}_n$ of g proposed by Priestley and Chao (1972)

$$\hat{g}_n(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K\Big(\frac{x-i/n}{b_n}\Big)Y_i, \qquad 0 \le x \le 1\,, \tag{2}$$

where the kernel K is some, not necessarily positive function such that $\int K(s)\,ds = 1$ and bandwidths (smoothing parameters) satisfy natural conditions $b_n \to 0$ and $nb_n \to \infty$. The properties of $\hat{g}_n(\cdot)$ have been investigated by numerous authors in the case when the errors are independent or weakly dependent, see e.g. Fan and Yao (2003). For the study of long-range dependent case of fixed-design regression we refer to Hall and Hart (1990) and Csörgő and Mielniczuk (1995). Hall and Hart consider the model (1) with general second-order stationary errors for which $r(k) \sim Ck^{-\alpha}$ as $k \to \infty$ for some constant $C > 0$ and $\alpha \in (0,1]$, assuming that g is twice differentiable. They established an asymptotic form of Mean Squared Error (MSE) of $\hat{g}_n(x)$, namely

$$\begin{aligned} MSE(\hat{g}_n(x)) &= \frac{b_n^4}{4}\Big(\int s^2K(s)ds\Big)^2 g''^2(x) + \frac{C}{(nb_n)^\alpha}\iint |x-y|^{-\alpha}K(x)K(y)dxdy + \\ &+ o\Big(b_n^4 + (nb_n)^{-\alpha}\Big) \end{aligned} \tag{3}$$

uniformly in $\delta < x < 1-\delta$ for each $\delta > 0$ from which it follows that the smallest asymptotic mean square error, i.e. the sum of the two main terms in (3), is achieved for an optimal bandwidth proportional to $n^{-\alpha/(4+\alpha)}$. Csörgő and Mielniczuk (1995) show in their Theorem 2 that under certain conditions imposed on $g(\cdot)$, $K(\cdot)$, $L(\cdot)$ and b_n the correct norming factor for $(\hat{g}_n(x) - E\hat{g}_n(x))$ to get a non-degenerate asymptotic distribution is $a_n^\star = (nb_n)^{\alpha/2}/L^{1/2}(nb_n)$.

In order to illustrate the importance of the way the explanatory variable is sampled, consider a random-design regression model (RDR)

$$Y_i = g(X_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{4}$$

where X_i are independent and have the uniform density on $[0,1]$. It is additionally assumed that the two sequences (X_i) and (ε_i) are independent. An estimator $\hat{g}_n(x)$ in this model is defined as in (2) with i/n replaced by X_i. It is known that, unlike in the FDR case, in the RDR model with long-range dependent errors the Priestley-Chao estimator exhibits a dichotomous asymptotic behaviour depending on the amount of smoothing employed. Namely, it is proved in Csörgő and Mielniczuk (2000) that when (ε_i) is a one-sided moving average LRD process described later, then

$$\min\Big((nb_n)^{1/2}, \frac{n^{\alpha/2}}{L^{1/2}(n)}\Big)\ \Big(\hat{g}_n(x) - g(x)\Big) \xrightarrow{\mathcal{D}} \gamma Z\,, \tag{5}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and Z is a standard normal random variable.

Observe that comparison of the norming sequence in (5) and the second term in the expansion of $\mathrm{MSE}(\hat{g}_n(x))$ in (3) corresponding to the asymptotic variance suggests that the asymptotic variance of $\hat{g}_n(x)$ is of a higher order in the random than in the fixed-design case whereas the bias is of the same order in both models. For a heuristic justification of this phenomenon we refer to Csörgő and Mielniczuk (1999). Efromovitch (1999) and Yang (2001) arrived at a similar conclusions by comparing minimax rates of convergence of regression estimators in the RDR and the FDR models.

In order to take advantage of a smaller variance of $\hat{g}_n(x)$ in the random-design case we introduce randomization of explanatory variable in the fixed-design regression model (1). Our motivation is to decrease the dependence between the observations which are effectively used for constructing $\hat{g}_n(x)$. To this end consider a permutation $\sigma = \sigma_n$ of the set $\{1,\ldots,n\}$ randomly chosen from a set Σ_n of all such permutations and assume that observations are taken consecutively at points $\sigma(1)/n, \sigma(2)/n, \ldots, \sigma(n)/n$ instead of points $1/n, 2/n, \ldots, 1$. As dependence of observations reflects solely the temporal order in which they are taken, the appropriate model of this observational scheme is

$$Y_{i,n} = g\Big(\frac{\sigma_n(i)}{n}\Big) + \varepsilon_{i,n}, \qquad i = 1,\ldots,n. \tag{6}$$

The random permutation σ_n is chosen independently of $(\varepsilon_{i,n})$. We will refer to (6) as to the Randomized Fixed-Design Regression model (RFDR). A modified Priestley-Chao estimator is thus

$$\hat{g}_n(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K\Big(\frac{x - \sigma(i)/n}{b_n}\Big)Y_i, \qquad 0 \le x \le 1\,. \tag{7}$$

We also consider a slight modification of (6), called the RFDR-b, in which a bootstrap sample of uniform grid points is considered i.e.

$$Y_i = g\Big(\frac{U_i}{n}\Big) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{8}$$

where $U_1, U_2, \dots, U_n$ are independent variables uniformly distributed on $\{1, \dots, n\}$. The sequence (U_i) is chosen independently from (ε_i). In this case the Priestley-Chao estimator is defined as in (7) with $\sigma(i)$ defined by U_i.

The main condition imposed in the paper on the process (ε_i) is that

$$\frac{1}{na_n} \sum_{i=1}^{n} \varepsilon_i \xrightarrow{\mathcal{D}} Z, \tag{9}$$

where (a_n) is a norming sequence defined before Proposition 2 such that $a_n^{-1} = o(n^{1/2})$ and Z is some nondegenerate univariate random variable. In the rest of the paper we assume that Z has the standard normal distribution. Note that $(na_n)^{-1} = o(n^{-1/2})$.

In particular, consider the case when (ε_i) can be represented as one-sided moving average process $\varepsilon_i = \sum_{t=0}^{\infty} c_t \eta_{i-t}, i = 1, 2, \dots$. Here $(\eta_t)_{t=-\infty}^{\infty}$ is a sequence of independent, identically distributed innovations such that $E\eta_1 = 0$, $E\eta_1^2 = 1$ and c_t satisfy $\sum_{t=0}^{\infty} c_t^2 < \infty$. If $c_t = L^{1/2}(t)t^{-\beta}$ where $1/2 < \beta < 1$, routine calculation based on the Karamata theorem implies that $r(k) \sim C(\beta)L(k)k^{-\alpha}$, where $C(\beta) := \int_0^{\infty}(x + x^2)^{-\beta}\, dx$ and $\alpha = 2\beta - 1$. In this case (9) was proved by Ibragimov and Linnik (1971) in Theorem 18.6.5 with Z being the standard normal random variable. Another model is a subordinated Gaussian process $\varepsilon_i = G(\eta_i)$, (η_i) is a a Gaussian LRD sequence and G is of Hermite order 1 (cf Beran (1994)).

We estimate g at fixed distinct points $x_1, \dots, x_k \in (0, 1)$ for some $k \in \mathbb{N}$ and show that asymptotic behavior of $\hat{g}_n$ in the RFDR model is analogous to its behavior in the case of random explanatory variables. Namely, depending on the size of a bandwidth, different norming factors are required to get a nondegenerate asymptotic distribution. A borderline of the dichotomy is the same as in the RDR model. More importantly, for both parts of the dichotomy, asymptotic variances are of a higher order than in the fixed-design case indicating superiority of this design (compare section 3 in Csörgő and Mielniczuk (1999)). Yang (2001, p. 641) also conjectured this type of result.

2. RESULTS

Let $K_b(x) := b^{-1}K(x/b)$. The Priestley-Chao estimator given by (7) has the following representation in the RFDR model:

$$\begin{aligned}\hat{g}_n(x) &= \frac{1}{n}\sum_{i=1}^{n} K_b\Big(x-\frac{\sigma(i)}{n}\Big)g\Big(\frac{\sigma(i)}{n}\Big)+\frac{1}{n}\sum_{i=1}^{n} K_b\Big(x-\frac{\sigma(i)}{n}\Big)\varepsilon_i \\ &= \frac{1}{n}\sum_{i=1}^{n} K_b\Big(x-\frac{i}{n}\Big)g\Big(\frac{i}{n}\Big)+\frac{1}{n}\sum_{i=1}^{n} K_b\Big(x-\frac{i}{n}\Big)\varepsilon_{\sigma^{-1}(i)} =: \tilde{g}(x)+A_n(x). \qquad (10)\end{aligned}$$

We consider first how the introduced randomization affects properties of long-range dependent errors and two first moments of $\hat{g}_n(x)$.

Proposition **1**. *Let* $\bar{\varepsilon}_{i,n} = \varepsilon_{\sigma_n^{-1}(i)}$, $i = 1,\ldots,n$, $n \in \mathbb{N}$. *Then for the RFDR model*

(i) $(\bar{\varepsilon}_{i,n}), i \le n, n \in \mathbb{N}$, *is a rowwise exchangeable array of random variables*;

(ii) $Cov(\bar{\varepsilon}_{i,n}, \bar{\varepsilon}_{j,n}) \sim \dfrac{2L(n)n^{-\alpha}}{(1-\alpha)(2-\alpha)}$, *for* $i \neq j$ *and* $Var(\bar{\varepsilon}_{i,n}) = Var(\varepsilon_{i,n})$.

(11)

Property (ii) follows from noting that $\mathrm{Cov}(\bar{\varepsilon}_{i,n}, \bar{\varepsilon}_{j,n})$ equals for $i \neq j$

$$\frac{1}{n(n-1)}\sum_{1\le k\neq l\le n}\mathrm{Cov}(\bar{\varepsilon}_{i,n}, \bar{\varepsilon}_{j,n}|\sigma(k)=i,\sigma(l)=j) = \frac{1}{n(n-1)}\sum_{1\le k\neq l\le n}\mathrm{Cov}(\varepsilon_{k,n}, \varepsilon_{l,n}),$$

where the last equality is implied by independence of σ_n and $(\varepsilon_{i,n})$. Routine application of Karamata theorem yields (ii). Let $a_n^2 = 2((1-\alpha)(2-\alpha))^{-1}L(n)n^{-\alpha} \sim Var(n^{-1}\sum_{i=1}^n \varepsilon_i)$.

Proposition **2**. *Assume that K is compactly supported and satisfies Lipschitz condition. Then for the RFDR and RFDR-b models we have*

(i) $E\hat{g}_n(x) = \tilde{g}(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} K_b\Big(x-\frac{i}{n}\Big)g\Big(\frac{i}{n}\Big)$;

(ii) $Var\,\hat{g}_n(x) = (nb_n)^{-1}\sigma_\varepsilon^2\displaystyle\int K^2(s)ds + a_n^2 + o((nb_n)^{-1}+a_n^2)$. (12)

Assumptions on K are used in part (ii) only. Property (ii) for the RFDR model follows from Proposition 1 and equality

$$\mathrm{Var}\,\hat{g}_n(x) = \frac{1}{n^2}\Big(\sum_{1\le i\neq j\le n} K_b(x-i/n)K_b(x-j/n)\mathrm{Cov}(\bar{\varepsilon}_{i,n}, \bar{\varepsilon}_{j,n}) + \sum_{i=1}^{n} K_b^2(x-i/n)\mathrm{Var}(\bar{\varepsilon}_{i,n})\Big),$$

after noting that under assumed conditions $b^{l-1}n^{-1}\sum_{i=1}^{n}K_b^l(x-i/n)\to\int K^l$ for $l=1,2$. Proposition 2 implies that the asymptotic variance of $\hat{g}_n(x)$ in the RFDR model coincides with the asymptotic variance of its counterpart in the RDR model and exhibits dichotmous behavior depending on the size of the bandwidth. Namely, $\operatorname{Var}\hat{g}_n(x)\sim a_n^2$ provided $a_n^{-2}=o(nb_n)$, and is equivalent to $(nb_n)^{-1}\sigma_\varepsilon^2\int K^2(s)\,ds$ when the opposite condition $nb_n=o(a_n^{-2})$ holds. The results below show that the analogy between behavior of the Pristley-Chao estimator in the RFDR and RDR models extends to asymptotic laws.

Consider distinct points $x_1,\ldots,x_k\in(0,1)$. Let $\mathcal{C}^1(\mathbb{R})$ denote a family of continuously differentiable real functions. One part of the smoothing dichotomy, for large bandwidths satisfying $a_n^{-2}=o(nb_n)$ is expressed by the first result. Note that as $b_n=o(1)$ the last condition can be satisfied only in the LRD case when $a_n^{-1}=o(n^{1/2})$.

Theorem **1**. *Assume that (9) holds, K is compactly supported on $(-1,1)$, satisfies Lipschitz condition and moreover $a_n^{-2}=o(nb_n)$. Then in the RFDR and RFDR-b models*

$$a_n^{-1}(\hat{g}_n(x_1)-\tilde{g}(x_1),\ldots,\hat{g}_n(x_k)-\tilde{g}(x_k))\xrightarrow{\mathcal{D}}(Z,\ldots,Z). \tag{13}$$

where Z is a standard normal random variable and $\tilde{g}(x)=E\hat{g}_n(x)$.

If $g\in\mathcal{C}^2(U_x)$ for some neigborhood U_x of x and K satisfies assumptions of Proposition 2 and is symmetric it is easily seen that $\tilde{g}(x)-g(x)=\mathcal{O}(b_n^2+(nb_n)^{-1})$. Then $a_n^{-1}(\tilde{g}(x)-g(x))\to 0$ provided $nb_n^5\to 0$ and in such case $\tilde{g}(x)$ may be replaced by $g(x)$ in (13).

Set $\widetilde{\sigma}^2=\sigma_\varepsilon^2\int K^2(s)ds$, where σ_ε^2 is the variance of the errors. The opposite part of the dichotomy for small bandwidths satisfying $nb_n=o(a_n^{-2})$, will be proved for the RFDR model in the special case of positively correlated Gaussian errors (ε_i). Gaussianity of (ε_i) is exploited by use of diagram formula (cf e.g. Arcones (1994)), however, Proposition 2 (ii) and Theorem 3 suggest that the result holds under weaker assumptions.

Theorem **2**. *Let (ε_i) be Gaussian random variables such that $0\le r(i)<1$ for $i\neq 0$. Assume that $K\in\mathcal{C}^1(\mathbb{R})$ is supported on $(-1,1)$ and $nb_n=o(a_n^{-2})$. Then in the RFDR model*

$$(nb_n)^{1/2}(\hat{g}_n(x_1)-\tilde{g}(x_1),\ldots,\hat{g}_n(x_k)-\tilde{g}(x_k))\xrightarrow{\mathcal{D}}\ \widetilde{\sigma}\ (Z_1,\ldots,Z_k) \tag{14}$$

where $Z_1, \ldots, Z_k$ are independent standard normal random variables.

The last result concerns the RFDR-b model. Here, the second part of the dichotomy can be proved under weaker conditions taking adventage of the fact that explanatory random variables U_i are independent.

THEOREM **3**. *Assume conditions of Theorem 2 on K and b_n and moreover, let $(\varepsilon_{i,n})$ be an ergodic array in a sense that $n^{-1}\sum_{i=1}^n G(\varepsilon_{i,n}) \xrightarrow{\mathcal{P}} \mathbb{E}G(\varepsilon_{1,1})$ for any G such that $\mathbb{E}|G(\varepsilon_{1,1})| < \infty$. Then the convergence (14) holds in the RFDR-b model.*

As for a linear one-sided linear process we have $\varepsilon_t = T(\ldots, \eta_{t-1}, \eta_t)$, it is easy to see that an array of linear processes satisfies the conditions of the above theorem.

3. PROOFS

Proof of Theorem 1. Let $k = 1$ and $x = x_1$. We will prove the result for the RFDR model, the reasoning for the RFDR-b model is similar but simpler. Note that the left-hand side of (13) for $k = 1$ can be written as

$$a_n^{-1}\Big(\hat{g}_n(x)-\tilde{g}(x)-EK_b\Big(x-\frac{\sigma(1)}{n}\Big)\frac{1}{n}\sum_{i=1}^n \varepsilon_i\Big)+a_n^{-1}\, EK_b\Big(x-\frac{\sigma(1)}{n}\Big)\frac{1}{n}\sum_{i=1}^n \varepsilon_i =: T_{1,n}(x)+T_{2,n}(x).$$

It is easy to check that $EK_b(x-\sigma(1)/n)-1 \to 0$ as K is Lipschitz continuous with a compact support integrating to 1. By assumption $(1/na_n)\sum_{i=1}^n \varepsilon_i \xrightarrow{\mathcal{D}} Z$. Thus it is enough to show that $T_{1,n}(x) \xrightarrow{\mathcal{P}} 0$. Using the fact that σ and ε are independent we have

$$a_n^2 E(T_{1,n}^2(x)) = \frac{1}{n^2}\sum_{i,j=1}^n \mathrm{Cov}\Big(K_b\Big(x-\frac{\sigma(i)}{n}\Big), K_b\Big(x-\frac{\sigma(j)}{n}\Big)\Big)\, E\varepsilon_i\varepsilon_j\ .$$

Let $\gamma_{i,j} := \mathrm{Cov}(K_b(x-\sigma(i)/n), K_b(x-\sigma(j)/n))$ and observe that for $i \neq j$

$$\begin{aligned}
\gamma_{ij} &= \frac{1}{n(n-1)}\sum_{1\le k\neq l\le n} K_b\Big(x-\frac{k}{n}\Big)K_b\Big(x-\frac{l}{n}\Big) - \frac{1}{n^2}\sum_{1\le k,l\le n} K_b\Big(x-\frac{k}{n}\Big)K_b\Big(x-\frac{l}{n}\Big)\\
&= \Big(\frac{1}{n(n-1)}-\frac{1}{n^2}\Big)\sum_{1\le k,l\le n} K_b\Big(x-\frac{k}{n}\Big)K_b\Big(x-\frac{l}{n}\Big) - \frac{1}{n(n-1)}\sum_{k=1}^n K_b^2\Big(x-\frac{k}{n}\Big)\\
&= \mathcal{O}(n^{-1}+(nb_n)^{-1})
\end{aligned}$$

and analogously $\gamma_{ii} = \mathcal{O}(b_n^{-1})$. Thus as γ_{ij} does not depend on i and j

$$a_n^2 E(T_{1,n}^2(x)) = \mathcal{O}(\frac{1}{n^3 b_n} \sum_{1 \le i,j \le n} \mathbb{E}\varepsilon_i \varepsilon_j + \frac{1}{nb_n}) = o(a_n^2)$$

in view of $a_n^{-2} = o(nb_n)$. For the general case $k \in \mathbb{N}$ note that it easily follows that $a_n^{-1}(\hat{g}_n(x_1) - \tilde{g}(x_1), \ldots, \hat{g}_n(x_k) - \tilde{g}(x_k))$ is equivalent to $(T_{2,n}(x_1), \ldots, T_{2,n}(x_k))$ and thus the proof proceeds by the same token. ■

In order to prove Theorem 2 we will need diagram formula (cf e.g. Arcones (1994)). A diagram G of order $(k_1, k_2, \ldots, k_l)$ is a set of vertices $\{(i,h) : 1 \le i \le l, 1 \le h \le k_i\}$ and a set of edges $\{((i,h),(j,m)) : 1 \le i < j \le l, 1 \le h \le k_i, 1 \le m \le k_j\}$, such that each vertex is of degree one. The set of edges will be denoted by $E(G)$. We denote by $\Gamma(k_1, k_2, \ldots, k_l)$ the set of diagrams of order $(k_1, k_2, \ldots, k_l)$. Given an edge $w = ((i,h),(j,m))$ let $d_1(w) = i$ and $d_2(w) = j$. Then the diagram formula is

LEMMA **1**. *Let (ε_i) be a Gaussian stationary sequence such that $\mathbb{E}\,\varepsilon_i = 0$ and $\mathrm{Var}\,\varepsilon_i = 1$. Then*

$$\mathbb{E}\left(\prod_{s=1}^{l} H_{k_s}(\varepsilon_s)\right) = \sum_{G \in \Gamma(k_1,k_2,\cdots,k_l)} \prod_{w \in E(G)} r(d_2(w) - d_1(w)). \tag{15}$$

LEMMA **2**. *Let (ε_i) be as in Lemma 1 and assume additionally that its covariance $r(i) = L(i)i^{-\alpha}$ for $0 < \alpha < 1$ and $r(i) \ge 0$ for $i \in \mathbb{N}$. Moreover, $i_1, i_2, \ldots, i_l \in \mathbb{N}$ are different indices and $k_1, k_2, \ldots, k_l \in \mathbb{N}$. Then*
(i) $\mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \ldots \bar{\varepsilon}_{i_l}^{k_l}) = 0$ when $k_1 + k_2 + \ldots + k_l = 2k+1$ with $k \in \mathbb{N}$.
(ii) $\mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \ldots \bar{\varepsilon}_{i_l}^{k_l}) = O(a_n^{2\lceil s/2 \rceil})$ when $k_1 + k_2 + \ldots + k_l = 2k$ and $s = \#\{k_j, 1 \le j \le l : k_j = 1\}$.

Proof of Lemma 2. In order to prove part (i) observe that

$$\mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \bar{\varepsilon}_{i_2}^{k_2} \cdots \bar{\varepsilon}_{i_l}^{k_l}) = \frac{1}{n(n-1)\ldots(n-l+1)} \sum_{j_1,j_2,\ldots,j_l}^{*} \mathbb{E}(\varepsilon_{j_1}^{k_1} \varepsilon_{j_2}^{k_2} \ldots \varepsilon_{j_l}^{k_l}), \tag{16}$$

where $*$ denotes summation over all sequences $j_1, j_2, \ldots, j_l$ of different indices belonging to $\{1, 2, \ldots, n\}$. By the diagram formula it follows that $\mathbb{E}(H_{k_1}(\varepsilon_{j_1}) \cdots H_{k_l}(\varepsilon_{j_l})) = 0$ as the set $\Gamma(k_1, k_2, \ldots, k_l)$ is empty. Then the proof of (i) is easily obtained by induction with respect to k by noting that $\mathbb{E}(\varepsilon_{j_1}^{k_1} \cdots \varepsilon_{j_l}^{k_l})$ differs from $\mathbb{E}(H_{k_1}(\varepsilon_{j_1}) \cdots H_{k_l}(\varepsilon_{j_l}))$ by a linear combination of terms of the form $\mathbb{E}(\varepsilon_{j_1}^{s_1}\cdots\varepsilon_{j_l}^{s_l})$, where $k_i \geq s_i \geq 0$, $\sum s_i < \sum k_i$ and $s_i \equiv k_i \pmod 2$. This in turn follows by observing that $H_{2k}(x)$ (respectively, $H_{2k+1}(x)$) is a linear combination of even (respectively, odd) powers of x. Note also that the number of ones among $s_1,\ldots,s_l$ is greater of equal the number of ones among $k_1,k_2,\ldots,k_l$. This will be used in the proof of part (ii).

Proof of part (ii). Observe that the conclusion automatically holds for $s=0$ as by Hölder inequality $|E(\varepsilon_{j_1}^{k_1}\varepsilon_{j_2}^{k_2}\cdots\varepsilon_{j_l}^{k_l})| \leq \left(\prod_{i=1}^{l}\mathbb{E}\left(|\varepsilon_{j_i}|^{k_i l}\right)\right)^{1/l}$ and thus the left hand side of (16) is bounded by a constant $C(k_1,\ldots,k_l)$ independent of n. Consider now the case $s>0$ and let without loss of generality $k_1 = 1$. Consider $\Gamma_i \subset \Gamma(k_1,\ldots,k_l)$ consisting of diagrams containing an edge which joins the first and i^{th} level. Observe that by removing this edge and the first level, Γ_i is mapped onto $\Gamma(k_2,\ldots,k_i',\ldots,k_l)$ where $k_i' = k_i - 1$ in such a way that groups of k_i graphs from Γ_i differing only by the edge emanating from the first level are mapped onto the same graph. Thus using the fact that $r(i)\geq 0$ and diagram formula we have by summing over j_1

$$\sum_{j_1,j_2,\ldots,j_l}^{*}\mathbb{E}(\varepsilon_{j_1}H_{k_2}(\varepsilon_{j_2})\ldots H_{k_l}(\varepsilon_{j_l})) \leq Cn^{1-\alpha}L(n)\max(k_1,\ldots,k_l)S(k_2,k_3,\ldots,k_l), \tag{17}$$

where $S(k_2,k_3,\ldots,k_l)$ equals

$$\sum_{j_2,j_3,\ldots,j_l}^{*}\mathbb{E}(H_{k_2-1}(\varepsilon_{j_2})H_{k_3}(\varepsilon_{j_3})\ldots H_{k_l}(\varepsilon_{j_l})) + \cdots + \mathbb{E}(H_{k_2}(\varepsilon_{j_2})H_{k_3}(\varepsilon_{j_3})\cdots H_{k_l-1}(\varepsilon_{j_l})).$$

The same reasoning can be applied to every term on the right hand side of the above equation $\lceil s/2\rceil$ times by noting that the number of levels of order 1 in any diagram belonging to $\Gamma(k_2,\ldots,k_i',\ldots,k_l)$ is at least $s-2$. It follows that

$$\frac{1}{n(n-1)(n-2)\cdots(n-l+1)}\sum_{j_1,j_2,\ldots,j_l}^{*}\mathbb{E}(H_{k_1}(\varepsilon_{j_1})\ldots H_{k_l}(\varepsilon_{j_l})) = O(a_n^{2\lceil s/2\rceil}). \tag{18}$$

The proof of (ii) is now obtained by induction on k using (18) and exploiting the relation between $E(H_{k_1}(\varepsilon_{j_1})\ldots H_{k_l}(\varepsilon_{j_l}))$ and $\mathbb{E}(\varepsilon_{j_1}^{k_1}\cdots\varepsilon_{j_l}^{k_l})$ used in the proof of part (i). ■

Proof of Theorem 2. We prove the result for $k=1$, the general case is obtained using similar reasoning based on Cramér-Wald device. Without loss of generality we assume that $\sigma_\varepsilon^2 = 1$.

Let $T_n(x) = (nb_n)^{1/2} A_n(x)$. We will use the method of moments and show that $\mathbb{E}T_n(x)^q$, $q = 0, 1, \ldots$ converge to moments of $N(0, \int K^2(s)\,ds)$ equal to $(\int K^2(s)\,ds)^m (2m-1)!!$ for $q = 2m$ and 0 for $q = 2m+1$. From Lemma 2 it follows that $\mathbb{E}T_n(x)^{2m+1} = 0$. Thus it is enough to consider the convergence of even moments of $T_n(x)$. We have

$$\mathbb{E}T_n(x)^{2m} = \sum_{l=1}^{2m} W_l =:$$

$$(nb_n)^m \frac{1}{n^{2m}} \sum_{l=1}^{2m} \sum_{\substack{k_1+k_2\ldots+k_l=2m \\ k_1 \geq k_2 \ldots \geq k_l > 0}} C_{\mathbf{k},l} \sum_{i_1,i_2,\ldots,i_l}^{*} K_b^{k_1}\Big(x - \frac{i_1}{n}\Big) \ldots K_b^{k_l}\Big(x - \frac{i_l}{n}\Big) \mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \cdots \bar{\varepsilon}_{i_l}^{k_l}),$$

where $C_{\mathbf{k},l} = C_{k_1,k_2,\ldots,k_l} = (2m)!/(k_1! \cdots k_l! p_1! \cdots p_{2m}!)$ and $p_j = \sum_{i=1}^{l} I(k_i = j)$ for $j = 1, 2, \ldots, 2m$. Observe that

$$\begin{aligned} W_l &= \sum_{k_1+k_2\ldots+k_l=2m} C_{\mathbf{k},l} \frac{(nb_n)^l}{(nb_n)^m} \sum_{i_1,i_2,\ldots,i_l}^{*} \Big(\frac{b_n^{k_1-1}}{n}\Big) K_b^{k_1}\Big(x - \frac{i_1}{n}\Big) \ldots \Big(\frac{b_n^{k_l-1}}{n}\Big) K_b^{k_l}\Big(x - \frac{i_l}{n}\Big) \mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \cdots \bar{\varepsilon}_{i_l}^{k_l}) \\ &= \mathcal{O}((nb_n)^{l-m} a_n^{2\lceil s/2 \rceil}) \end{aligned}$$

in view of Lemma 2 (ii). Note that $k_1 + \ldots + k_l = 2m$ implies $s + 2(l-s) \leq 2m$ and thus $l - m \leq [s/2]$. Therefore the imposed condition on bandwidth implies that $W_l \to 0$ for $l > m$ when $n \to \infty$ and obviously $W_l \to 0$ when $l < m$ as $\mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \cdots \bar{\varepsilon}_{i_l}^{k_l}) \leq C(k_1, k_2, \ldots, k_l)$. Consider now the remaining case $l = m$ and note that if there is a power $k_i > 2$ then $s > 0$ and thus in view of the Lemma 2(ii) $\mathbb{E}(\bar{\varepsilon}_{i_1}^{k_1} \ldots \bar{\varepsilon}_{i_l}^{k_l}) \to 0$. Thus it is enough to show that

$$\frac{1}{n(n-1)\ldots(n-l+1)} \sum_{j_1,j_2,\ldots,j_l}^{*} \mathbb{E}(\varepsilon_{j_1}^2 \cdots \varepsilon_{j_l}^2) \to 1 \tag{19}$$

as it is easy to see that $C_{\mathbf{k},m} = (2m-1)!!$ for $\mathbf{k} = (2, 2, \ldots, 2)$. In order to prove (19) observe that as $H_2(x) = x^2 - 1$

$$\sum_{j_1,j_2,\ldots,j_l}^{*} \mathbb{E}(H_2(\varepsilon_{j_1}) \cdots H_2(\varepsilon_{j_l})) - \sum_{t=0}^{l-1} (-1)^{l-t} \sum_{j_1,j_2,\ldots,j_l}^{*} \mathbb{E}(\varepsilon_{j_1}^2 \cdots \varepsilon_{j_t}^2) = \sum_{j_1,j_2,\ldots,j_l}^{*} \mathbb{E}(\varepsilon_{j_1}^2 \cdots \varepsilon_{j_l}^2)\,.$$

Convergence (19) follows from above equality by an easy induction by noting that

$$\sum_{j_1,j_2,\ldots,j_l}^{*} \mathbb{E}(\varepsilon_{j_1}^2 \cdots \varepsilon_{j_t}^2) = \binom{l}{t} (n-t) \ldots (n-l) \sum_{j_1,j_2,\ldots,j_t}^{*} \mathbb{E}(\varepsilon_{j_1}^2 \cdots \varepsilon_{j_t}^2)$$

provided we prove $1/n(n-1)\cdots(n-l+1)\sum^*_{j_1,j_2,\ldots,j_l}\mathbb{E}(H_2(\varepsilon_{j_1})\cdots H_2(\varepsilon_{j_l}))\to 0$. Using diagram formula and $1\geq r(i)\geq 0$ we have

$$\sum^*_{j_1,j_2,\ldots,j_l} E(H_2(\varepsilon_{j_1})\cdots H_2(\varepsilon_{j_l}))\leq \sum_{G\in\Gamma(2,2,\ldots,2)}\sum^*_{j_1,j_2,\ldots,j_l}\sum_{w\in E(G)} r(d_2(w)-d_1(w)).$$

Moreover, we have for an edge w joining levels s and t

$$\sum^*_{j_1,j_2,\ldots,j_l} r(d_2(w)-d_1(w))\leq n^{l-2}\sum_{j_t\neq j_s} r(j_t-j_s)=\mathcal{O}(n^{l-\alpha}L(n))$$

from which the needed property follows since $\Gamma(2,2,\ldots,2)$ consists of a finite number of diagrams. ■

Proof of Theorem 3. Let $V_n(x)=(b_n/n)^{1/2}\sum_{k=1}^n K_b(x-U_k/n)\varepsilon_k$, $\mu_n(x)=\mathbb{E}(V_n(x)|\varepsilon_1,\ldots,\varepsilon_n)$ and $s_n^2(x)=\mathrm{Var}(V_n(x)|\varepsilon_1,\ldots,\varepsilon_n)$. Reasoning as in proof of Theorem 2 in Csörgő and Mielniczuk (1999) one can check that with $\tilde\sigma^2=\sigma_\varepsilon^2\int K^2(s)\,ds$

$$\mathbb{E}\exp(i(V_n(x)-\mu_n(x))|\varepsilon_1,\ldots,\varepsilon_n)\xrightarrow{\mathcal{P}}\exp(-t^2\tilde\sigma^2/2) \tag{20}$$

implies convergence in distribution of $V_n(x)-\mu_n(x)$ to $N(0,\tilde\sigma^2)$. Thus in order to prove Theorem 3 for $k=1$ it is enough to prove that $\mu_n(x)\to 0$ in probability and (20). But

$$\mu_n(x)=(nb_n)^{1/2}a_n\mathbb{E}(K_b(x-U_1/n))a_n^{-1}\sum_{i=1}^n\varepsilon_i/n=o_P(1),$$

in view of assumptions on (b_n) and (ε_i). Note that given $(\varepsilon_1,\ldots,\varepsilon_n)$ $V_n(x)$ is a sum of i.i.d. random variables such that conditional variance $s_n^2(x)$ tends in probability to $\tilde\sigma^2$ in view of ergodic property. Thus in order to prove (20) it is enough to check the Lindeberg condition. It will follow from

$$\frac{b_n}{n}\sum_{k=1}^n\varepsilon_k^2\mathbb{E}_{U_k}\Big(K_b^2(x-\frac{U_k}{n})I\{(b_n/n)^{1/2}|(K_b(x-U_k/n)-\mathbb{E}_{U_k}K_b(x-U_k/n))\varepsilon_k|\geq\eta\}\Big)\to 0 \tag{21}$$

in probability for any $\eta>0$. However, taking into account that K has compact support in $(-1,1)$ and is bounded we have that left-hand side of (21) is bounded by

$$\frac{b_n}{n}\sum_{k=1}^n\varepsilon_k^2 n^{-1}\sum_{j=1}^n K_b^2(x-\frac{j}{n})I\{(b_n/n)^{1/2}|(K_b(x-j/n)-\mathbb{E}_{U_k}K_b(x-U_k/n))\varepsilon_k|\geq\eta\}$$

$$\leq \frac{b_n}{n} \sum_{k=1}^{n} \varepsilon_k^2 n^{-1} \sum_{j:\, U_j \in (x-nb_n, x+nb_n)} K_b^2(x - \frac{j}{n}) I\{|\varepsilon_k| \geq C_1 \eta (nb_n)^{1/2}\}$$

$$\leq C_2 \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k^2 I\{|\varepsilon_k| \geq C_1 \eta (nb_n)^{1/2}\} \to 0$$

in probability in view of assumed ergodic property of $(\varepsilon_{i,n})$ as $nb_n \to \infty$.

The general case is proved analogously using Cramér-Wald device. Namely, compactness of suppport of K implies that $\mathrm{Var}(c_1 V_n(x_1) + \ldots + c_k V_n(x_k)) \to (c_1^2 + \ldots + c_k^2)\tilde{\sigma}^2$ and checking random Lindeberg condition proceeds in the same way as above. ∎

5. SIMULATION RESULTS

We conducted a simulation study to investigate the effect of randomization of the fixed design regression in practice. We generated series (Y_i) of length $n = 1000$ with trend functions

(i) $g_1(x) = 2\sin(4\pi x)$;

(ii) $g_2(x) = 2 - 5x + 5\exp\{-100(x - 0.5)^2\}$.

These are two regression functions used in Ray and Tsay (1996). The considered errors follow either a fractional autoregressive integrated moving average process FARIMA$(0, d, 0)$ or a fractional Gaussian noise process fGn(H) with $d = 0.1, 0.2, 0.3, 0.4$, where the Hurst exponent H satisfies $H = 1 - \alpha/2$ and $H = d + 0.5$. It is known that $L(n) \sim C$ and one-sided moving average representation exists in both cases. For FARIMA$(0, d, 0)$ process $\varepsilon_t = (1 - B)^{-d}\eta_t$, where (η_t) is a Gaussian white noise with marginal variance σ_η^2 and $B\eta_t = \eta_{t-1}$, we have $C = \sigma_\eta^2 \Gamma(1 - 2d)/\Gamma(d)\Gamma(1 - d)$. Let η_t be now a fractional Brownian motion (fBm) with Hurst exponent H such that $\mathrm{Var}(\eta_t) = \sigma_\eta^2 |t|^{2H}$. Then for fGn process $\varepsilon_t = \eta_{t+1} - \eta_t$, $C = \sigma_\eta^2 (1 - \alpha/2)(1 - \alpha)k^{-\alpha}$. We refer to Beran (1994) for more information on both processes.

The number of replications of each experiment was 500. The employed kernel was the Epanechnikov kernel $K(x) = 0.75\,(1 - x^2),\ |x| \leq 1$. For each series the performance of the following bandwidths was investigated:

1. The asymptotically optimal bandwidth, i.e. the bandwidth minimizing asymptotic $MISE(\hat{g}_n)$ in the respective model.

- For the FDR model we have: $b_n^f = (\alpha D_1/D_2)^{1/(4+\alpha)} n^{-\alpha/(4+\alpha)}$, where

$D_1 = C \iint |x-y|^{-\alpha} K(x)K(y)dxdy$ and $D_2 = (\int s^2 K(s)ds)^2 \int g''^2(s)ds$.

- For the RDR, RFDR and RFDR-b models we have:

$$b_n^{rf} = (\sigma_\varepsilon^2 \int K^2(s)ds \,/\, D_2)^{1/5} n^{-1/5}.$$

2. The empirically optimal bandwidth, i.e. the bandwidth minimizing

$$ISE(b) = \sum_{i=1}^{n} \{\hat{g}((i-0.5)/n) - g((i-0.5)/n)\}^2$$

over a grid of 20 equally spaced points between 0 and 0.5.

Obviously, neither of these bandwidths are known when an unknown regression function is estimated. We stress that the aim of the simulation study is not to construct data-based bandwith selection method for LRD regression but rather to compare the performance of regression estimators in the FDR and RFDR models when optimal parameters are chosen for the respective model. Tables 1 and 2 show the average values of asymptotically and empirically optimal bandwitdhs together with the medians of corresponding distributions of the Integrated Squared Error (ISE). The medians were considered because of significant skewness of underlying distributions of ISE. It follows that the medians of the ISE for the RFDR model are the smallest ones among *all* considered designs.

The same conclusion holds true when MISE is considered instead of the median of the ISE. Randomization of the fixed design yields significant improvement of estimation accuracy (measured by the median of the ISE) with the effect becoming more pronounced for stronger dependence. For weaker dependence (d=0.1) the fixed design yields better results than the random design for $n = 1000$ with the reverse conclusion for stronger dependence. The results for the RDR and RFDR-b models are very similar indicating that there is a negligible difference between sampling from uniform distributions on [0,1] and on $\{1/n, 2/n, \ldots, 1\}$. For all designs accuracy of estimation decreases with increasing d or H and estimation of g_2 is more difficult than that of g_1. The medians of ISE for each considered design are more variable with changing d for FARIMA than fGn errors. Empirically optimal bandwidths are close to the asymptotic values and the same is true for the corresponding medians of the ISE suggesting that $n = 1000$ is sufficiently large sample size for validity of asymptotic analysis.

Table 1: Average bandwidths and medians of Integrated Squared Error for $g_1(x)$

		FARIMA								fGn							
		FDR		RDR		RFDR		RFDR-b		FDR		RDR		RFDR		RFDR-b	
d		b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE
0.1	as.	0.057	0.028	0.050	0.037	0.050	0.017	0.050	0.038	0.057	0.031	0.032	0.054	0.032	0.021	0.032	0.055
	e.	0.055	0.029	0.061	0.033	0.051	0.018	0.064	0.034	0.057	0.029	0.060	0.033	0.050	0.017	0.063	0.033
0.2	as.	0.065	0.065	0.051	0.045	0.051	0.024	0.051	0.044	0.066	0.064	0.038	0.055	0.038	0.023	0.038	0.051
	e.	0.066	0.060	0.061	0.040	0.050	0.023	0.063	0.041	0.066	0.065	0.061	0.040	0.050	0.023	0.063	0.040
0.3	as.	0.077	0.155	0.052	0.075	0.052	0.050	0.052	0.071	0.074	0.145	0.043	0.073	0.043	0.048	0.043	0.073
	e.	0.078	0.148	0.064	0.067	0.052	0.050	0.064	0.068	0.076	0.137	0.063	0.063	0.050	0.047	0.065	0.069
0.4	as.	0.092	0.486	0.057	0.252	0.057	0.232	0.057	0.257	0.079	0.260	0.046	0.135	0.046	0.112	0.046	0.139
	e.	0.098	0.467	0.075	0.226	0.063	0.224	0.076	0.235	0.082	0.256	0.065	0.120	0.053	0.102	0.066	0.110

Table 2: Average bandwidths and medians of Integrated Squared Error for $g_2(x)$

		FARIMA								fGn							
		FDR		RDR		RFDR		RFDR-b		FDR		RDR		RFDR		RFDR-b	
d		b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE	b_n	ISE
0.1	as.	0.050	0.073	0.044	0.103	0.044	0.054	0.044	0.103	0.050	0.074	0.029	0.130	0.029	0.046	0.029	0.129
	e.	0.033	0.065	0.054	0.098	0.025	0.045	0.053	0.100	0.033	0.069	0.053	0.100	0.025	0.046	0.052	0.100
0.2	as.	0.057	0.112	0.045	0.113	0.045	0.062	0.045	0.111	0.057	0.115	0.034	0.123	0.034	0.053	0.034	0.123
	e.	0.045	0.107	0.054	0.105	0.025	0.053	0.054	0.105	0.045	0.112	0.053	0.115	0.025	0.052	0.054	0.112
0.3	as.	0.067	0.219	0.046	0.148	0.046	0.090	0.046	0.146	0.064	0.197	0.038	0.153	0.038	0.081	0.038	0.144
	e.	0.058	0.220	0.055	0.149	0.027	0.086	0.055	0.142	0.053	0.187	0.054	0.126	0.025	0.072	0.054	0.134
0.4	as.	0.079	0.579	0.051	0.339	0.051	0.304	0.051	0.355	0.068	0.349	0.041	0.223	0.041	0.166	0.041	0.223
	e.	0.079	0.584	0.062	0.353	0.036	0.302	0.060	0.345	0.060	0.323	0.054	0.192	0.027	0.138	0.057	0.188

REFERENCES

Arcones, M.A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. Ann. Probab. 22:2242–2274

Beran, J. (1994). *Statistics for Long-memory Processes.* London: Chapman & Hall.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Lecture Notes in Statistics. no. 110, Springer

Csörgő, S. and Mielniczuk, J. (1995). Distant long-range dependent sums and regression estimation. Stochastic Processes Applic. 59:143–155

Csörgő, S. and Mielniczuk, J. (1999). Random design regression under long-range dependent errors. Bernoulli, 5:209–224

Csörgő, S. and Mielniczuk, J. (2000). The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. Statistica Sinica, 10:771–787

Dobrushin, R.L. and Major, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. Z. Wahrsh. Verw. Gebiete, 50:27–52

Efromovitch, S. (1999). How to overcome the curse of long-memory errors. IEEE Trans. Inform. Th. 45:1735–1741

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods*, Springer

Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. Stochastic Process. Appl. 36:339–351

Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters Noordhoff

Priestley, M.B. and Chao, M.T. (1972). Nonparametric function fitting. J. Roy. Statist. Soc. Ser. B. 34:385–392.

Ray, B.K. and Tsay, R.S. (1996). Iterative bandwidth estimation for nonparametric regression with long-range dependent errors, in Athens Conference on Applied Probability and Time Series, Vol. 2 (P.M. Robinson, M. Rosenblatt, eds), 339–351, Lecture Notes in Statistics no. 115, Springer

Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. Z. Wahrsh. Verw. Gebiete. 50:53–83

Yang, Y. (2001). Nonparametric regression with dependent errors. Bernoulli 7:633–655