APPLICATIONES MATHEMATICAE
41,1 (2014), pp. 67–80

Artur Bryk (Warszawa)

SMOOTHING DICHOTOMY IN RANDOMIZED FIXED-DESIGN REGRESSION WITH STRONGLY DEPENDENT ERRORS BASED ON A MOVING AVERAGE

Abstract. We consider a fixed-design regression model with errors which form a Borel measurable function of a long-range dependent moving average process. We introduce an artificial randomization of grid points at which observations are taken in order to diminish the impact of strong dependence. We show that the Priestley–Chao kernel estimator of the regression fuction exhibits a dichotomous asymptotic behaviour depending on the amount of smoothing employed. Moreover, the resulting estimator is shown to exhibit weak consistency (i.e. in probability). Simulation results indicate significant improvement when randomization is employed.

1. Introduction. Consider a fixed-design regression model (FDR)

$$Y_{i,n} = g(i/n) + \varepsilon_{i,n}, \quad i = 1, \dots, n, \tag{1.1}$$

where $g : [0,1] \to \mathbb{R}$ is some function with smoothness properties to be described. For each n, we observe the random variables $Y_{1,n}, Y_{2,n}, \dots, Y_{n,n}$ and the aim is to estimate the unknown function g based on this information. Here $(\varepsilon_{i,n})$ is a triangular array such that for each n, the finite sequence $\{\varepsilon_{i,n}\}_{i=1}^n$ is stationary, $\mathbb{E}\varepsilon_{i,n} = 0$, $\mathbb{E}\varepsilon_{i,n}^2 < \infty$, $\operatorname{Cov}(\varepsilon_{i,n}, \varepsilon_{j,n}) = r_\varepsilon(|i-j|)$, where $r_\varepsilon(\cdot)$ is the covariance function which does not depend on n.

In a nonparametric setting, the regression function g at a given point x is usually estimated by one of many methods involving local polynomials, smoothing splines or kernel estimators. Any of these methods weighs concomitants of grid points around x in such a way that those closer to x contribute more to the value of the estimator. As the concomitants corresponding to a small neighbourhood of x form a block of consecutive observa-

---

2010 *Mathematics Subject Classification*: 62G08, 62G20, 62M10.

Key words and phrases: kernel estimators, linear process, long-range dependence, randomization, fixed-design regression, smoothing dichotomy.

DOI: 10.4064/am41-1-6

tions which are strongly dependent, the resulting estimator is more variable than in the weakly dependent case. In order to alleviate the effect of dependence on variability of the regression estimator we consider a randomly chosen permutation $\sigma = \sigma_n$ of the set $\{1, \ldots, n\}$ from the set Σ_n of all such permutations and assume that the observations are taken consecutively at the points $\sigma(1)/n, \sigma(2)/n, \ldots, \sigma(n)/n$ instead of $1/n, 2/n, \ldots, 1$. As dependence of the observations reflects solely the temporal order in which they are taken, the appropriate model of this observational scheme is

$$Y_{i,n} = g(\sigma_n(i)/n) + \varepsilon_{i,n}, \quad i = 1, \ldots, n. \tag{1.2}$$

The random permutation σ_n is chosen independently of $(\varepsilon_{i,n})$. We will refer to (1.2) as the *Randomized Fixed-Design Regression* model (RFDR). The idea of considering (1.2) is based on the observation made in [9] that the regression estimators in a random-design regression model with LRD errors are less variable than in the fixed-design case and is in line with a general discussion in [14]. For a thorough discussion of the effect of design type on regression estimation with LRD errors see [7]. We stress that plausibility of model (1.2) is based on the assumption that dependence between the observations is due to their temporal and not spatial proximity. Thus, for example, dependence of two consecutive observations $(t = i, i+1)$ will be the same regardless of the grid points at which the experimenter takes the observations. Another insight into advantages of a randomization scheme can be found in [3].

The condition we impose on the process $(\varepsilon_{i,n})$ is

$$\varepsilon_{i,n} = G(Z_{i,n}), \quad i = 1, \ldots, n, \tag{1.3}$$

where $G : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function such that $\int |G(z)|\, dz < \infty$ and $\mathbb{E}G^2(Z_{i,n}) < \infty$. The last restriction is satisfied if we assume, for instance, that G is square integrable or bounded. Here $(Z_{i,n})$ is a one-sided moving average process given by

$$Z_{i,n} = \sum_{t=0}^{\infty} c_t \eta_{i-t}^{[n]}, \quad i = 1, \ldots, n, \tag{1.4}$$

where $(\eta_t^{[n]})_{t=-\infty}^{\infty}$ is a sequence of independent, identically distributed innovations such that $\mathbb{E}\eta = 0$, $\mathbb{E}(\eta^2) = \sigma_\eta^2 < \infty$ and the c_t satisfy $\sum_{t=0}^{\infty} c_t^2 < \infty$. Let

$$c_t = L(t) t^{-\beta} \quad (c_0 = 1), \tag{1.5}$$

where $1/2 < \beta < 1$ and $L(\cdot)$ is a function defined on $[0, \infty)$, slowly varying at infinity and positive in some neighbourhood of infinity. A routine calculation based on the Karamata theorem (see [11, p. 281]) implies that $r(k) = \mathrm{Cov}(Z_{1,n}, Z_{1+k,n}) \sim \sigma_\eta^2 C(\beta) L^2(k) k^{-\alpha}$, $k = 1, \ldots, n-1$, where

$C(\beta) = \int_0^\infty (x + x^2)^{-\beta}\,dx$ and $\alpha = 2\beta - 1$. Thus, in this case, the array $(Z_{i,n})$ is long-range dependent (LRD) in the sense that $\sum_{k=1}^\infty |r(k)| = \infty$. Note that if $\sum_{t=0}^\infty |c_t| < \infty$, or $\beta > 1$ in the hyperbolic decay condition given above, then $(Z_{i,n})$ is short-range dependent.

Set $c_t = 0$ for $t < 0$. Then (1.4) can be written as

$$Z_{i,n} = \sum_{j=-\infty}^{\infty} c_{i-j}\,\eta_j^{[n]}, \qquad i = 1, \ldots, n.$$

It is easily seen that

$$\gamma_n^2 = \mathrm{Var}\Big(\sum_{i=1}^n Z_{i,n}\Big) = \sigma_\eta^2 \sum_{k=-\infty}^{n} \Big(\sum_{t=1}^{n} c_{t-k}\Big)^2 \sim \sigma_\eta^2 D(\beta) n^{2-\alpha} L^2(n),$$

where $D(\beta) = [(2 - 2\beta)(3/2 - \beta)]^{-1} C(\beta)$. Then noting that $\gamma_n^2 \to \infty$ as $n \to \infty$, it follows from [13, Theorem 18.6.5] that

$$\frac{1}{\gamma_n} \sum_{i=1}^n Z_{i,n} \xrightarrow{\mathcal{D}} \mathcal{N}, \tag{1.6}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $\mathcal{N}$ is a standard normal random variable. Note that $\gamma_n^{-1} = o(n^{-1/2})$.

In the following we suppress the dependence of $Y_{i,n}$, $Z_{i,n}$, $(\eta_t^{[n]})$ and $\varepsilon_{i,n}$ on n.

We return now to the models (1.1) and (1.2). We will use the Priestley–Chao kernel estimator (see [16]) to estimate g in both models. In the FDR model it is defined as follows:

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

where the kernel K is a not necessarily positive function such that $\int K(s)\,ds = 1$ and the bandwidths (smoothing parameters) satisfy natural conditions: $b_n \to 0$ and $nb_n \to \infty$. The modified Priestley–Chao estimator in the RFDR model is

$$\hat{g}_n(x) = \frac{1}{nb_n} \sum_{i=1}^n K\Big(\frac{x - \sigma(i)/n}{b_n}\Big) Y_i, \qquad 0 \le x \le 1. \tag{1.7}$$

We estimate g at fixed distinct points $x_1, \ldots, x_k \in (0, 1)$ for some $k \in \mathbb{N}$ and show that depending on the size of the bandwidths, two different norming factors are required to get a nondegenerate asymptotic distribution. A small bandwidth case reflects the 'whitening by windowing' principle, i.e. for such bandwidths the asymptotic behaviour of the regression estimator is the same as for independent errors. The same dichotomous asymptotic behaviour of the Priestley–Chao estimator was shown in [3] and [4], but to prove it in the case of model (1.2) with errors (1.3) we will need a new technique. In [3], LRD errors form a moving average process with coefficients given in (1.5). In [4], the case of positively correlated LRD Gaussian errors is considered. A borderline of the dichotomy is established and turns out to be the same as in the random-design regression model with LRD errors (cf. [10]). More importantly, for both parts of the dichotomy, asymptotic variances are of a lower order than in the fixed-design case, indicating superiority of this design. Actually, the different behaviour of $\hat{g}_n$ in the FDR and RFDR models can be conjectured from [10] and [8], as the artificial randomized variables $\sigma_n(\cdot)/n$ mimic the behaviour of independent explanatory random variables uniformy distributed on $[0,1]$.

The paper concludes with a simulation example showing the effect of randomization in practice. It indicates that randomization has a nonnegligible impact on the integrated square error even if we consider sample size $n = 1000$.

2. Results. The following notation will be used throughout the paper. Let $x \in (0,1)$,

$$K_{b_n}(x) := \frac{1}{b_n} K\left(\frac{x}{b_n}\right), \qquad J_{n,i}(x) := K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) G(Z_i),$$
$$Z_{i;s} := \sum_{j=i-s}^{\infty} c_j \eta_{i-j}.$$

Let $W_i = \sigma(\ldots, \eta_{i-1}, \eta_i)$ be the σ-field generated by all innovations up to time i, and let h_i be the density of $Z_i - Z_{i;0} = \sum_{j=0}^{i-1} c_j \eta_{i-j}$. In particular, h_1 stands for the density of η_1, and $h = h_\infty$ is the marginal density of the linear process (Z_i). Let $\|\xi\| = (\mathbb{E}(\xi^2))^{1/2}$ be the $\mathcal{L}^2$-norm of the random variable ξ and let

$$\mathcal{P}_k \xi = \mathbb{E}(\xi \,|\, W_k) - \mathbb{E}(\xi \,|\, W_{k-1}), \qquad k \in \mathbb{N},$$

be the projection differences.

The Priestley–Chao estimator given by (1.7) has the following representation in the RFDR model:

$$\begin{aligned}\hat{g}_n(x) &= \frac{1}{n}\sum_{i=1}^{n} K_{b_n}\left(x - \frac{i}{n}\right) g\left(\frac{i}{n}\right) + \frac{1}{n}\sum_{i=1}^{n} [J_{n,i}(x) - \mathbb{E}(J_{n,i}(x) \,|\, W_{i-1})] \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(J_{n,i}(x) \,|\, W_{i-1}) \\ &=: \widetilde{g}_n(x) + M_n(x) + N_n(x).\end{aligned}$$

Note that $M_n(x)$ has a martingale structure and $\mathbb{E}\hat{g}_n(x) = \widetilde{g}_n(x)$.

Consider distinct points $x_1, \dots, x_k \in (0,1)$. First we state crucial auxiliary lemmas which provide the asymptotic law for $M_n(x)$ and an asymptotic representation for $N_n(x)$.

LEMMA 1. *Assume that*

$$\mathbb{E}K_{b_n}(x - \sigma(i)/n) = \mathcal{O}(1), \qquad b_n \mathbb{E}K_{b_n}^2(x - \sigma(i)/n) = \mathcal{O}(1)$$

for any $x = x_l$, $l = 1, \dots, k$. Let the density h_1 and the kernel K be bounded. Then

$$\sqrt{nb_n}\big(M_n(x_1), \dots, M_n(x_k)\big) \xrightarrow{\mathcal{D}} \big(\rho(x_1)\mathcal{N}_1, \dots, \rho(x_k)\mathcal{N}_k\big)$$

where

$$\rho^2(x) = b_n \mathbb{E}K_{b_n}^2(x - \sigma(i)/n) \cdot \mathbb{E}G^2(Z_i)$$

and $\mathcal{N}_1, \dots, \mathcal{N}_k$ are independent standard normal random variables.

REMARK 1. If K is compactly supported, satisfies the Lipschitz condition and $\int K^2(s)\,ds < \infty$ then the first two conditions in Lemma 1 are satisfied in view of

$$\mathbb{E}K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) \to \int K(s)\,ds = 1 \quad \text{and} \quad b_n\mathbb{E}K_{b_n}^2\left(x - \frac{\sigma(i)}{n}\right) \to \int K^2(s)\,ds.$$

Then $\rho^2(x) = \int K^2(s)\,ds \cdot \mathbb{E}G^2(Z_i)$.

Note that

$$\begin{aligned} N_n(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) \cdot \int G(z + Z_{i;i-1})h_1(z)\,dz \\ &= \frac{1}{n}\mathbb{E}K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) \cdot \sum_{i=1}^{n} \int G(z)h_1(z - Z_{i;i-1})\,dz. \end{aligned}$$

Let

$$\widetilde{N}_n(x) := -\frac{1}{n}\mathbb{E}K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) \cdot \int G(z)h'(z)\,dz \cdot \sum_{i=1}^{n} Z_{i;i-1}.$$

LEMMA 2. *Assume that $\mathbb{E}K_{b_n}(x - \sigma(i)/n) = \mathcal{O}(1)$ and h_1 is twice continuously differentiable with bounded derivatives. Suppose that there exists $C > 0$ such that*

$$\left\| \int G(z)[h_{i-1}'(z - \xi) - h_{i-1}'(z)]\,dz \right\| \le C\|\xi\| \tag{2.1}$$

for $\xi = Z_{i;0}$ and $\xi = Z_{i;1}$, and

$$\left\| \int G(z) R_{2,i}(z)\,dz \right\| \le C|c_{i-1}|^2 \tag{2.2}$$

where $R_{2,i}(z) = h_{i-1}(z - Z_{i;1}) - h_{i-1}(z - Z_{i;0}) + h_{i-1}'(z - Z_{i;0})c_{i-1}\eta_1$. Then

$$\|N_n(x) - \widetilde{N}_n(x)\| = \mathcal{O}\left(\frac{\Xi_n}{n}\right) \tag{2.3}$$

where

$$\Xi_n^2 := n\Theta_n^2 + \sum_{i=1}^{\infty} (\Theta_{n+i} - \Theta_i)^2, \qquad \Theta_n = \sum_{i=1}^{n} \theta_i,$$

$$\theta_i = |c_{i-1}|\sqrt{A_{i-1}}, \qquad A_i = \sum_{j=i}^{\infty} c_j^2.$$

Let

$$a_n^2 := \mathrm{Var}\Big(n^{-1} \sum_{i=1}^{n} Z_i\Big) \sim 2((1-\alpha)(2-\alpha))^{-1} L^2(n) n^{-\alpha}.$$

One part of the smoothing dichotomy, for large bandwidths satisfying $a_n^{-2} = o(nb_n)$ that allow the long memory of the errors to prevail, is expressed by our 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 the conditions of Lemmas* 1 *and* 2 *hold and* $a_n^{-2} = o(nb_n)$. *Then in the RFDR model*

$$a_n^{-1}\big(\hat{g}_n(x_1) - \widetilde{g}_n(x_1), \ldots, \hat{g}_n(x_k) - \widetilde{g}_n(x_k)\big) \xrightarrow{\mathcal{D}} \bar{\rho}(\mathcal{N}, \ldots, \mathcal{N}), \tag{2.4}$$

where

$$\bar{\rho} = -\mathbb{E}K_{b_n}(x - \sigma(i)/n) \cdot \int G(z) h'(z)\, dz$$

and $\mathcal{N}$ *is a standard normal random variable.*

REMARK 2. (a) If $g \in \mathcal{C}^2(U_x)$ for some neigbourhood U_x of $x \in (0,1)$ and K is compactly supported, satisfies the Lipschitz condition and is symmetric, then it is easily seen that

$$\widetilde{g}_n(x) - g(x) = \mathcal{O}(b_n^2 + (nb_n)^{-1}).$$

(b) Assuming additionally $nb_n^5 \to 0$, we have $a_n^{-1}(\widetilde{g}_n(x) - g(x)) \to 0$ and in that case $\widetilde{g}_n(x)$ may be replaced by $g(x)$ in (2.4).

The opposite part of the dichotomy for small bandwidths in the given sense is stated below. Both the norming sequence and the limiting process are the same if the errors were independent.

THEOREM 2. *Assume that the conditions of Lemmas* 1 *and* 2 *hold and* $nb_n = o(a_n^{-2})$. *Then in the RFDR model*

$$\sqrt{nb_n}\big(\hat{g}_n(x_1) - \widetilde{g}_n(x_1), \ldots, \hat{g}_n(x_k) - \widetilde{g}_n(x_k)\big) \xrightarrow{\mathcal{D}} \big(\rho(x_1)\mathcal{N}_1, \ldots, \rho(x_k)\mathcal{N}_k\big), \tag{2.5}$$

where $\rho^2(x) = b_n \mathbb{E}K_{b_n}^2(x - \sigma(i)/n) \cdot \mathbb{E}G^2(Z_i)$ *and* $\mathcal{N}_1, \ldots, \mathcal{N}_k$ *are independent standard normal random variables.*

In Theorem 3 below we investigate conditions under which weak consistency of $\hat{g}_n$ holds. In this case no decay rate of the coefficients c_i is needed. The proof of Theorem 3 proceeds through the following lemma.

LEMMA 3. *Assume that* $\mathbb{E}K_{b_n}(x-\sigma(i)/n) = \mathcal{O}(1)$, $b_n\mathbb{E}K_{b_n}^2(x-\sigma(i)/n) = \mathcal{O}(1)$ *for any* $x \in (0,1)$ *and the density* h_1 *of* η_1 *is Lipschitz continuous with Lipschitz constant* A. *Then*

$$\operatorname{Var}\hat{g}_n(x) \le \frac{8C_1^2\|\eta_1\|^2}{n^2}\sum_{k=-\infty}^{n}\Big(\sum_{t=1}^{n}|c_{t-k}|\Big)^2 + \frac{2}{nb_n}\mathbb{E}G^2(Z_i), \tag{2.6}$$

where $C_1 = A\int|G(z)|\,dz$.

THEOREM 3. *Assume that the conditions of Lemma* 3 *and Remark* 2(a) *hold. Then* $\hat{g}_n(x) \to g(x)$ *in probability for every* $x \in (0,1)$.

A sequence $(\xi_n)_{n=1}^{\infty}$ of random variables converges to 0 *completely* if $\sum_{n=1}^{\infty} P(|\xi_n| > \varepsilon) < \infty$ for any $\varepsilon > 0$.

PROPOSITION 1. *Let* K *and* G *be bounded functions,* $b_n\mathbb{E}K_{b_n}^2(x-\sigma(i)/n) = \mathcal{O}(1)$ *for any* $x \in (0,1)$ *and* $nb_n/\log n \to \infty$. *Then* $M_n(x) \to 0$ *completely.*

3. Proofs. In all proofs, C denotes a generic constant whose value may change, and $x \in (0,1)$.

Proof of Lemma 1. We prove the lemma for $k=1$; the extension to the case $k>1$ is obtained using a similar reasoning based on the Cramér–Wald device.

Recall that $J_{n,i}(x) := K_{b_n}(x-\sigma(i)/n)G(Z_i)$. Let $T_{n,i}(x) := J_{n,i}(x) - \mathbb{E}(J_{n,i}(x)\,|\,W_{i-1})$. By the martingale central limit theorem, it suffices to show that

$$\mathbb{E}\bigg|\frac{b_n}{n}\sum_{i=1}^{n}\mathbb{E}(T_{n,i}^2(x)\,|\,W_{i-1}) - \rho^2(x)\bigg| \to 0 \tag{3.1}$$

and that the Lindeberg condition

$$b_n\mathbb{E}[T_{n,i}^2(x)\cdot\mathbb{I}_{\sqrt{b_n/n}\,|T_{n,i}(x)|>\epsilon}] = o(1)$$

holds for any $\epsilon > 0$. In order to prove (3.1), observe that since h_1 is bounded,

$$\begin{aligned}\bigg|\mathbb{E}\bigg(\sqrt{\frac{b_n}{n}}J_{n,i}(x)\,\bigg|\,W_{i-1}\bigg)\bigg| &= \sqrt{\frac{b_n}{n}}\bigg|\mathbb{E}K_{b_n}\Big(x-\frac{\sigma(i)}{n}\Big)\bigg|\cdot\bigg|\int G(z)h_1(z-Z_{i;i-1})\,dz\bigg|\\ &\le C\sqrt{\frac{b_n}{n}},\end{aligned}$$

where $C = \sup h_1(z) \cdot \int |G(z)|\,dz \cdot |\mathbb{E}K_{b_n}(x - \sigma(i)/n)| < \infty$. Thus we have

$$\Big|\sum_{i=1}^n \mathbb{E}\Big(\frac{b_n}{n} J_{n,i}^2(x) \,\Big|\, W_{i-1}\Big) - \sum_{i=1}^n \mathbb{E}\Big(\frac{b_n}{n} T_{n,i}^2(x) \,\Big|\, W_{i-1}\Big)\Big| \le \sum_{i=1}^n \Big(\mathbb{E}\Big[\sqrt{\frac{b_n}{n}}\, J_{n,i}(x) \,\Big|\, W_{i-1}\Big]\Big)^2 = \mathcal{O}(b_n).$$

Hence we only need to check (3.1) with $T_{n,i}(x)$ replaced by $J_{n,i}(x)$. Thus we show that $\mathbb{E}|n^{-1}\sum_{i=1}^n p_i(x) - \rho^2(x)| \to 0$, where

$$\begin{aligned} p_i(x) &:= b_n \mathbb{E}(J_{n,i}^2(x) \mid W_{i-1}) \\ &= b_n \mathbb{E}K_{b_n}^2(x - \sigma(i)/n) \cdot \int G^2(z) h_1(z - Z_{i;i-1})\,dz. \end{aligned}$$

Let $S_i := \int G^2(z) h_1(z - Z_{i;i-1})\,dz$. Note that $\mathbb{E}S_i = \mathbb{E}G^2(Z_i)$ in view of the fact that $\mathbb{E}h_1(z - Z_{i;i-1}) = h(z)$ and (S_i) is ergodic as instantaneous transformation of a linear process which is ergodic (cf. [20, Theorem 1.3.3]). By the Ergodic Theorem, $\mathbb{E}|n^{-1}\sum_{i=1}^n p_i(x) - \rho^2(x)|$ tends to 0.

The Lindeberg condition results from [5, Corollary 9.5.2], which implies that

$$\begin{aligned} b_n \mathbb{E}\big[T_{n,i}^2(x) \cdot \mathbb{I}_{\{\sqrt{b_n/n}\,|T_{n,i}(x)| > \epsilon\}}\big] &\le 4 b_n \mathbb{E}\big[J_{n,i}^2(x) \cdot \mathbb{I}_{\{|J_{n,i}(x)| > \epsilon/(2\sqrt{n/b_n})\}}\big] \\ &\le C\mathbb{E}\big[G^2(Z_i) \cdot \mathbb{I}_{\{|G(Z_i)| > \epsilon/(C\sqrt{n b_n})\}}\big]. \end{aligned}$$

The right-hand side is $o(1)$ since $n b_n \to \infty$. ∎

Proof of Lemma 2. Let

$$V_i := \int G(z) h_1(z - Z_{i;i-1})\,dz + \int G(z) h'(z)\,dz \cdot Z_{i;i-1}.$$

Then

$$\|N_n(x) - \widetilde{N}_n(x)\| = \frac{1}{n}\Big|\mathbb{E}K_{b_n}\Big(x - \frac{\sigma(i)}{n}\Big)\Big| \cdot \Big\|\sum_{i=1}^n V_i\Big\|.$$

Thus (2.3) follows from application of the projection method (cf. [2]), more precisely from

$$\Big\|\sum_{i=1}^n V_i\Big\|^2 \le \sum_{k=-\infty}^n \Big(\sum_{i=1}^n \|\mathcal{P}_1 V_{i-k+1}\|\Big)^2 \le C \sum_{k=-\infty}^n \Big(\sum_{i=1}^n \theta_{i-k+1}\Big)^2 = \mathcal{O}(\Xi_n^2)$$

provided

$$\|\mathcal{P}_1 V_i\| \le C\theta_i \tag{3.2}$$

for $i \ge 1$. Note that

$$\mathcal{P}_1 V_i = \int G(z)[\mathcal{P}_1 h_1(z - Z_{i;i-1}) + c_{i-1}\eta_1 h'(z)]\,dz. \tag{3.3}$$

Let $R_{1,i}(z) = h'_{i-1}(z - Z_{i;1}) - h'_{i-1}(z)$. By [15, Lemma 1], we have $h'(z) = \mathbb{E}h'_{i-1}(z - Z_{i;1})$. Hence $\mathbb{E}R_{1,i}(z) = h'(z) - h'_{i-1}(z)$. Using $|\mathbb{E}\xi| \le \|\xi\|$ and

(2.1) we get

$$\Big|\int G(z)[h'(z) - h'_{i-1}(z)]\,dz\Big| \le C\|Z_{i;1}\| = C\sqrt{A_{i-1}}.$$

Using the last equality and (2.1) again we deduce via the triangle inequality that

$$\Big\|\int G(z)[h'_{i-1}(z - Z_{i;0}) - h'(z)]\,dz\Big\| \le C\|Z_{i;0}\| + C\|Z_{i;1}\| \le 2C\sqrt{A_{i-1}}.$$

The last inequality implies

$$\Big\|\int G(z)[h'_{i-1}(z - Z_{i;0}) - h'(z)]c_{i-1}\eta_1\,dz\Big\| \le 2C|c_{i-1}|\sqrt{A_{i-1}} = 2C\theta_i. \tag{3.4}$$

Let $(\eta_i^*)_{i=-\infty}^{\infty}$ be an iid copy of $(\eta_i)_{i=-\infty}^{\infty}$. Let $R_{2,i}^*(z)$ be $R_{2,i}(z)$ with η_1 replaced by η_1^*. Hence condition (2.2) entails $\|\int G(z)R_{2,i}^*(z)\,dz\| \le C|c_{i-1}|^2$ and

$$\Big\|\int G(z)[R_{2,i}(z) - R_{2,i}^*(z)]\,dz\Big\| \le 2C|c_{i-1}|^2. \tag{3.5}$$

Observe that

$$\mathcal{P}_1 h_1(z - Z_{i;i-1}) + h'_{i-1}(z - Z_{i;0})c_{i-1}\eta_1 = \mathbb{E}[R_{2,i}(z) - R_{2,i}^*(z) \mid W_1],$$

which implies (3.2) by (3.3)–(3.5) and $|c_{i-1}|^2 = \mathcal{O}(\theta_i)$. ∎

Proof of Theorem 1. Let $k = 1$ and $x = x_1$. Note that the left-hand side of (2.4) for $k = 1$ can be written as

$$a_n^{-1}\big(M_n(x) + N_n(x) - \widetilde{N}_n(x)\big) + a_n^{-1}\widetilde{N}_n(x) =: T_{1,n}(x) + T_{2,n}(x).$$

It follows from (1.6) that

$$T_{2,n}(x) = \bar{\rho}(na_n)^{-1}\sum_{i=1}^{n} Z_{i;i-1} \xrightarrow{\mathcal{D}} \bar{\rho}\mathcal{N}.$$

From [2, Theorem 2] we know that $\Xi_n/n = o(a_n)$. Thus $T_{1,n}(x) = o_p(1)$ in view of $a_n^{-2} = o(nb_n)$ and Lemmas 1 and 2. For the general case $k \in \mathbb{N}$ note that it easily follows that $a_n^{-1}\big(\hat{g}_n(x_1) - \widetilde{g}_n(x_1), \ldots, \hat{g}_n(x_k) - \widetilde{g}_n(x_k)\big)$ is equivalent to $(T_{2,n}(x_1), \ldots, T_{2,n}(x_k))$, and thus the proof proceeds along the same lines. ∎

Proof of Theorem 2. Note that (2.5) is a direct consequence of $\sqrt{nb_n} = o(a_n^{-1})$, Lemmas 1 and 2, convergence (1.6) and $\Xi_n/n = o(a_n)$. ∎

Proof of Lemma 3. Let

$$U_i = K_{b_n}\Big(x - \frac{\sigma(i)}{n}\Big)g\Big(\frac{\sigma(i)}{n}\Big) - \mathbb{E}\Big[K_{b_n}\Big(x - \frac{\sigma(i)}{n}\Big)g\Big(\frac{\sigma(i)}{n}\Big)\Big] + K_{b_n}\Big(x - \frac{\sigma(i)}{n}\Big)G(Z_i).$$

Note that $n^2 \operatorname{Var} \hat{g}_n(x) = \|\sum_{i=1}^n U_i\|^2$ and $\mathcal{P}_k U_t = 0$ for $t < k$. Observe that for $i > 1$ we have

$$\begin{aligned}\mathcal{P}_1 U_i &= \mathbb{E} K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) \cdot \left[\int G(z) h_{i-1}(z - Z_{i;1})\, dz - \int G(z) h_i(z - Z_{i;0})\, dz\right] \\ &= \mathbb{E} K_{b_n}\left(x - \frac{\sigma(i)}{n}\right) \cdot \left[\int G(z)\{h_{i-1}(z - Z_{i;1}) - h_i(z - Z_{i;1})\}\, dz\right. \\ &\qquad \left. + \int G(z)\{h_i(z - Z_{i;1}) - h_i(z - Z_{i;0})\}\, dz\right] =: I + II.\end{aligned}$$

Thus, reasoning analogously to [2, proof of Lemma 1], we get $|I| \le C_1 \mathbb{E}|c_{i-1}\eta_1|$, $\|\mathcal{P}_1 U_i\| \le 2C_1 |c_{i-1}|\, \|\eta_1\|$, $\|\mathcal{P}_1 U_1\|^2 \le \mathbb{E} K_{b_n}^2(x - \sigma(i)/n) \mathbb{E} G^2(Z_i)$ and finally (2.6). ∎

Proof of Theorem 3. The second term on the right-hand side of (2.6) tends to 0 as $nb_n \to \infty$. Moreover,

$$\begin{aligned}\frac{1}{n^2} \sum_{k=-\infty}^{n} \Big(\sum_{t=1}^{n} |c_{t-k}|\Big)^2 &= \frac{1}{n^2} \sum_{k=-\infty}^{n} \sum_{1 \le t,t' \le n} |c_{t-k} c_{t'-k}| \\ &\le \frac{2}{n} \sum_{i=0}^{n} \sum_{j=0}^{\infty} |c_j c_{j+i}| \le \Big(\sum_{j=0}^{\infty} c_j^2\Big)^{1/2} \frac{2}{n} \sum_{i=0}^{n} \Big(\sum_{j=0}^{\infty} c_{j+i}^2\Big)^{1/2} \to 0\end{aligned}$$

as $\sum_{j=0}^\infty c_j^2 < \infty$. Thus in view of (2.6), $\operatorname{Var} \hat{g}_n(x) \to 0$ and weak consistency of $\hat{g}_n(x)$ follows from Remark 2(a). ∎

Proof of Proposition 1. Recall that $J_{n,i}(x) := K_{b_n}(x - \sigma(i)/n) G(Z_i)$. Let $T_i := T_{n,i}(x) := J_{n,i}(x) - \mathbb{E}(J_{n,i}(x) \mid W_{i-1})$. We use a special case of Freedman's exponential inequality (see [12]) stating that for the sum $S_n = nM_n(x)$ of bounded martingale differences T_i with $|T_i| \le B$, we have, for $\lambda > 0$,

$$\mathbb{E} \exp(\lambda S_n) \le \exp(\beta B^{-2} e(B\lambda)),$$

where $e(\lambda) = e^\lambda - \lambda - 1$ and β is a bound of the conditional variances of T_i such that $P(\sum_{i=1}^n \mathbb{E}(T_i^2 \mid W_{i-1}) \le \beta) = 1$. In our case

$$\begin{aligned}\mathbb{E}(T_i^2 \mid W_{i-1}) \le \mathbb{E}(J_{n,i}^2(x) \mid W_{i-1}) &= \mathbb{E} K_{b_n}^2\left(x - \frac{\sigma(i)}{n}\right) \int G^2(z + Z_{i,i-1}) h_1(z)\, dz \\ &\le \frac{C \sup G^2}{b_n}\end{aligned}$$

and $B = 2 \sup K \sup G / b_n$. Thus $\sum_{i=1}^n \mathbb{E}(T_i^2 \mid \mathcal{F}_{i-1}) \le nC/b_n$ for some positive C, and for $\lambda > 0$ we have

$$P(M_n \ge \epsilon) \le \mathbb{E} \exp(\lambda S_n) / \exp(n\lambda\epsilon) \le \exp\{CB^{-2} n e(B\lambda)/b_n - n\lambda\epsilon\}.$$

Let $\lambda_n = \epsilon_n b_n / C$, where $\epsilon_n = (8C \log n/(nb_n))^{1/2}$. Note that as $e(\lambda) \sim 2^{-1}\lambda^2$ for $\lambda \to 0$ we have $e(B\lambda_n) < (3/4)(B\lambda_n)^2$ for sufficiently large n. Thus for such n the last bound is smaller than

$$\exp(6\log n - 8\log n) = \exp(-2\log n).$$

Now, the Borel–Cantelli lemma implies that $M_n(x) \geq (8C\log n/(nb_n))^{1/2}$ finitely often almost surely. Analogously we show that

$$\sum_n P(M_n(x) \leq -\epsilon_n) < \infty$$

and hence $M_n(x) \to 0$ completely. ∎

REMARK 3. The proof of Proposition 1 indicates that

$$M_n(x) = \mathcal{O}\bigg(\bigg(\frac{\log n}{nb_n}\bigg)^{1/2}\bigg) \quad \text{a.s.}$$

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

$$g_1(x) = 2\sin(4\pi x), \qquad g_2(x) = 2 - 5x + 5\exp\{-100(x-0.5)^2\}.$$

These are the two regression functions used in [17]. The corresponding errors follow the functions

$$G_1(x) = \frac{1}{1+x^2}, \qquad G_2(x) = \exp\{-x^2\}$$

of a FARIMA$(0,d,0)$ (fractional autoregressive integrated moving average process) with $d = 0, 0.1, 0.2, 0.3, 0.4$. It is known that for this process a one-sided moving average representation exists and the slowly varying function $L(\cdot)$ is equivalent to a constant τ, $L(n) \sim \tau$. For the FARIMA$(0,d,0)$ process $Z_t = (1-B)^{-d}\eta_t$, where (η_t) is a Gaussian white noise with marginal variance σ_η^2 and $B\eta_t = \eta_{t-1}$, we have $\tau = \sigma_\eta^2\Gamma(1-2d)/(\Gamma(d)\Gamma(1-d))$. We refer to [1] for more information on this process.

Besides RFDR and FDR models investigated in this paper we also considered a random-design regression (RDR) model in which the explanatory random variables form an independent sequence which is uniformly distributed on $[0,1]$ and independent of the errors.

The number of replications of each experiment was 1000. The kernel employed was either the normal kernel $K(x) = (2\pi)^{-1/2}\exp\{-x^2/2\}$, $x \in \mathbb{R}$, or the Epanechnikov kernel $K(x) = 0.75(1-x^2)$, $|x| \leq 1$. We now discus the choice of bandwidths in both cases.

RFDR model. As it can be conjectured from the results of [3, Section 3] that the asymptotic form AMISE of MISE in the RFDR model differs from AMISE for the FDR model with *independent* errors by a term which does not depend on b_n, asymptotically optimal theoretical bandwidths in both cases should coincide. In view of this we used Ruppert, Sheather and Wand's [19] data-based bandwidth commonly used for independent data,

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

where $D_2 = (\int s^2 K(s)\, ds)^2 \int g''^2(s)\, ds$ and σ_ε^2 is the variance of the errors. The procedure *dpill* from the **KernSmooth** package was employed for calculating it. The same bandwidths choice method was employed for the RDR model.

FDR model. The bandwidth minimizing AMISE of $\hat{g}_n$ is of the form

$$b_n^f = \left(\frac{\alpha D_1}{D_2} \right)^{1/(4+\alpha)} n^{-\alpha/(4+\alpha)}, \tag{4.1}$$

where $D_1 = \tau \int\int |x-y|^{-\alpha} K(x)K(y)\, dx\, dy$ and $\alpha \in (0,1]$. In order to develop a version of b_n^f, a local Whittle method proposed by Robinson [18] was used to estimate α. Specifically, a pair (d, Q) is estimated, where $\alpha = 1 - 2d$ and the spectral density f of the FARIMA$(0, d, 0)$ process satisfies $f(\lambda) \sim Q\lambda^{-2d}$ for $\lambda \to 0$. It can be checked that for the FARIMA process, $\sigma_\eta^2 = 2\pi Q$. Thus we have $\tau = 2\pi Q \Gamma(1-2d)/(\Gamma(d)\Gamma(1-d))$. The bandwidth sequence employed is quasi data-dependent in the sense that it assumes that certain quantities in the model are known. Namely, we assume that the true errors (ε_i) are known together with the value of $\int g''^2(s)\, ds$ and the previous relation linking τ with α and Q. Thus $\hat{\alpha}$ and $\hat{Q}$ are local Whittle estimators based on (ε_i), $\hat{\tau}$ is obtained from them under the FARIMA model, and $\hat{\alpha}$ and $\hat{\tau}$ are plugged in (4.1). In this way we give this method an advantage over the bandwidth choice method for the RFDR model.

The results of the simulation study are summarized in Tables 1–4.

The medians of the distribution of the Integrated Square Error ISE $= n^{-1} \sum_{i=1}^n (\hat{g}_n(x_i) - g(x_i))^2$ are used as the measure of performance. It follows that despite the fact that some quantities were assumed known for the FDR model, the performance of Priestley–Chao estimator is inferior to its performance when prior randomization is used. What is also remarkable is a much better performance of the regression estimator under a randomized discrete uniform grid than for the case when the explanatory variables were uniformly distributed on $[0,1]$. This may be caused by the fact that the explanatory variables are more evenly distributed across the interval $[0,1]$ when the equispaced grid is used. For all designs the accuracy of estimation is better when the G_2 function is used, and estimation of g_2 is more difficult than that of g_1. Moreover, there is no significant difference in the medians of ISE for the FDR and RFDR models when the normal or Epanechnikov kernel is used.

Table 1. Medians of ISE for $g_1(x)$ and normal kernel

	$G_1(x)$			$G_2(x)$		
d	**FDR**	**RDR**	**RFDR**	**FDR**	**RDR**	**RFDR**
0	0.4256	0.4910	0.4239	0.3321	0.3837	0.3309
0.1	0.4208	0.4830	0.4189	0.3275	0.3780	0.3259
0.2	0.4060	0.4730	0.4047	0.3109	0.3634	0.3095
0.3	0.3729	0.4306	0.3708	0.2758	0.3223	0.2735
0.4	0.3059	0.3489	0.3012	0.2120	0.2420	0.2052

Table 2. Medians of ISE for $g_2(x)$ and normal kernel

	$G_1(x)$			$G_2(x)$		
d	**FDR**	**RDR**	**RFDR**	**FDR**	**RDR**	**RFDR**
0	0.4458	0.5932	0.4484	0.3538	0.4790	0.3565
0.1	0.4424	0.5953	0.4444	0.3491	0.4737	0.3526
0.2	0.4280	0.5754	0.4296	0.3336	0.4500	0.3357
0.3	0.3964	0.5270	0.3955	0.2997	0.4154	0.2979
0.4	0.3386	0.4493	0.3296	0.2426	0.3317	0.2319

Table 3. Medians of ISE for $g_1(x)$ and Epanechnikov kernel

	$G_1(x)$			$G_2(x)$		
d	**FDR**	**RDR**	**RFDR**	**FDR**	**RDR**	**RFDR**
0	0.4297	0.5811	0.4243	0.3345	0.4589	0.3300
0.1	0.4263	0.5784	0.4208	0.3313	0.4537	0.3260
0.2	0.4102	0.5554	0.4052	0.3133	0.4309	0.3095
0.3	0.3766	0.5172	0.3718	0.2770	0.3874	0.2742
0.4	0.3100	0.4238	0.3079	0.2130	0.3052	0.2109

Table 4. Medians of ISE for $g_2(x)$ and Epanechnikov kernel

	$G_1(x)$			$G_2(x)$		
d	**FDR**	**RDR**	**RFDR**	**FDR**	**RDR**	**RFDR**
0	0.4489	0.7721	0.4409	0.3574	0.6316	0.3485
0.1	0.4457	0.7675	0.4377	0.3528	0.6203	0.3440
0.2	0.4312	0.7551	0.4223	0.3383	0.5972	0.3278
0.3	0.3998	0.7051	0.3878	0.3048	0.5540	0.2910
0.4	0.3397	0.6122	0.3209	0.2477	0.4675	0.2243

References

[1] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, London, 1994.
[2] A. Bryk and J. Mielniczuk, *Asymptotic properties of density estimates for linear processes: application of projection method*, J. Nonparametr. Stat. 17 (2005), 121–133.
[3] A. Bryk and J. Mielniczuk, *Using randomization to improve performance of regression estimators under dependence*, Acta Sci. Math. (Szeged) 73 (2007), 817–838.
[4] A. Bryk and J. Mielniczuk, *Randomized fixed design regression under long-range dependent errors*, Comm. Statist. Theory Methods 37 (2008), 520–531.
[5] Y. S. Chow and H. Teicher, *Probability Theory*, 2nd ed., Springer, New York, 1988.
[6] I. Cornfeld, S. Fomin and Y. Sinai, *Ergodic Theory*, Springer, 1982.
[7] S. Csörgő, *Smoothing dichotomy in nonparametric regression under long-range dependent errors*, Statistica Neerlandica 56 (2002), 132–142.
[8] S. Csörgő and J. Mielniczuk, *Distant long-range dependent sums and regression estimation*, Stoch. Process. Appl. 59 (1995), 143–155.
[9] S. Csörgő and J. Mielniczuk, *Random design regression under long-range dependent errors*, Bernoulli 5 (1999), 209–224.
[10] S. Csörgő and J. Mielniczuk, *The smoothing dichotomy in random-design regression with long-memory errors based on moving averages*, Statistica Sinica 10 (2000), 771–787.
[11] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York, 1971.
[12] D. Freedman, *On tail probabilities for martingales*, Ann. Probab. 3 (1975), 100–118.
[13] I. A. Ibragimov and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, 1971.
[14] H. Künsch, J. Beran and F. Hampel, *Contrasts under long-range dependent correlations*, Ann. Statist. 21 (1993), 943–964.
[15] J. Mielniczuk and W. B. Wu, *On random-design model with dependent errors*, Statistica Sinica 14 (2004), 1105–1126.
[16] M. B. Priestley and M. T. Chao, *Nonparametric function fitting*, J. Roy. Statist. Soc. Ser. B 34 (1972), 385–392.
[17] B. K. Ray and R. S. Tsay, *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.), Lecture Notes in Statistics 115, Springer, 1996, 339–351.
[18] P. Robinson, *Gaussian semiparametric estimation of time series with long-range dependence*, Ann. Statist. 23 (1995), 1630–1661.
[19] D. Ruppert, S. J. Sheather and M. D. Wand, *An effective bandwidth for local least squares regression*, J. Amer. Statist. Assoc. 90 (1995), 1257–1270.
[20] M. Taniguchi and Y. Kakizawa, *Asymptotic Theory of Statistical Inference for Time Series*, Springer, 2000.

Artur Bryk
Department of Mathematics and Mathematical Economics
Warsaw School of Economics
Al. Niepodległości 162
02-554 Warszawa, Poland
E-mail: abryk@sgh.waw.pl

Received on 19.6.2013;
revised version on 30.4.2014 (2187)