USING RANDOMIZATION TO IMPROVE PERFORMANCE OF REGRESSION ESTIMATORS UNDER DEPENDENCE

By Artur Bryk and Jan Mielniczuk

Dedicated to Sándor Csörgő on his 60th birthday

Abstract We consider a fixed-design regression model with long-range dependent errors which form a moving average process. Taking into account different behavior of regression estimators in such a model and in a random-design regression model discussed in Csörgő and Mielniczuk [5], we introduce an artificial randomization of grid points at which observations are taken in order to diminish the impact of strong dependence of errors. The resulting estimator is shown to exhibit smoothing dichotomy with the variance in both cases tending to 0 more quickly than in the fixed design case. Moreover, we establish a uniform convergence rate of the regression function estimators which also reflects the dichotomous behaviour of the regression estimator. Simulation results indicate significant improvement for moderate sample sizes when randomization is employed.

AMS Subject Classification: 62g07, 60f17

Introduction

Consider a fixed-design regression (FDR) model

\[ Y_{i,n} = g(i/n) + \varepsilon_{i,n}, \quad i = 1, 2, \ldots, n, \]

where \( g : [0, 1] \rightarrow \mathbb{R} \) 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, \( \mathbb{E}\varepsilon_{i,n} = 0, \mathbb{E}\varepsilon_{i,n}^{2} = \sigma_{\varepsilon}^{2} > 0, \)

\[ r(k) := \text{Cov}(\varepsilon_{i,n}, \varepsilon_{i+k,n}) = \hat{L}(k)k^{-\alpha} , \quad k = 1, 2, \ldots, \]

where \( 0 < \alpha < 1 \) is a fixed constant and \( \hat{L}(\cdot) \) is a function defined on \([0, +\infty)\), slowly varying at infinity and positive in some neighborhood of infinity. 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 a nonparametric setting the regression function \( g \) at a given point \( x \) is usually estimated by one of many methods including local polynomials, smoothing splines or kernel estimators. Any of these methods assigns weights to concomitants of grid points around
In such a way that those being closer to $x$ contribute more to a value of the estimator. As the concomitants corresponding to a small neighbourhood of $x$ form a block of consecutive observations which are strongly dependent, the resulting estimator is more variable than in a weakly dependent case. In order to alleviate the effect of dependence on variability of regression estimator we consider a randomly chosen permutation $\sigma = \sigma_n$ of the set $\{1, \ldots, n\}$ from a set $\Sigma_n$ 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 $1/n, 2/n, \ldots, 1$. Assuming dependence of observations reflects solely the temporal order in which they are taken, the appropriate model of this observational scheme is

$$Y_{i,n} = g\left(\frac{\sigma_n(i)}{n}\right) + \varepsilon_{i,n}, \quad i = 1, \ldots, n,$$

(2)

when $(\varepsilon_{i,n})$ form a triangular array defined above. The random permutation $\sigma_n$ is chosen independently of $(\varepsilon_{i,n})$. We will refer to (2) as to the Randomized Fixed-Design Regression (RFDR) model. The idea of considering (2) is based on observation made in Csörgő and Mielniczuk [4] that the regression estimators in random design regression model with LRD errors are less variable than when they are used for a deterministic design and is in line with a general discussion in Künsch et al. [11] on benefits of randomization in ANOVA model with strongly dependent errors when estimating contrasts. For a thorough discussion of effect of design type on regression estimation with LRD errors see Csörgő [3]. We stress that plausibility of model (2) is based on the assumption that the dependence between 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 grid points at which an experimenter takes the observations.

In order to motivate usage of randomization in case of LRD errors assume momentarily that $g$ in (1) is an affine function i.e. we deal with an ordinary univariate linear regression model. Consider estimating a slope coefficient $\beta_1$ when $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, x_i = i/n, i = 1, 2, \ldots, n, \varepsilon_i$ are LRD and such that $r(i) > 0$ for each $i$. Let $\hat{\beta}_1$ be the ordinary least squares estimator (OLSE) of $\beta_1$. Imagine that in order to guard against the 'curse of LRD errors' (a phrase coined by S. Efromovich [6]) we randomize grid points $x_i$ at which observations are taken, obtaining the model $Y_i = \beta_0 + \beta_1 x_{\sigma(i)} + \varepsilon_i$, where $\sigma(\cdot)$ is some random permutation. Denote the OLSE of $\beta_1$ in this model by $\tilde{\beta}_1$. Observe that both
estimators are unbiased. As \( \hat{\beta}_1 = \sum_{i=1}^{n} Y_i (x_i - \bar{x}) / S \) with \( S = \sum_{i=1}^{n} (x_i - \bar{x})^2 \) one has
\[
\text{Var} \hat{\beta}_1 = \frac{\sigma^2 \varepsilon}{S} + \frac{\sum_{i \neq j} r(|i - j|)(x_i - \bar{x})(x_j - \bar{x})}{S^2},
\] (3)
where \( \sigma^2 = r(0) \). At the same time (cf Proposition 2)
\[
\text{Var} \tilde{\beta}_1 = \frac{\sigma^2 \varepsilon}{S} + \frac{\sum_{i \neq j} \bar{r}(|i - j|)(x_i - \bar{x})(x_j - \bar{x})}{S^2},
\] (4)
with \( \bar{r}(|i - j|) = \bar{r} = (n(n - 1))^{-1} \sum_{s \neq t} r(|s - t|) \). As \( \sum_{i \neq j} (x_i - \bar{x})(x_j - \bar{x}) = -S < 0 \) and in case of positively correlated errors we have \( \bar{r} > 0 \), \( \text{Var} \hat{\beta}_1 \) is smaller than the variance \( \sigma^2 \varepsilon / S = O(n^{-1}) \) of the OLSE \( \hat{\beta}_1 \) under independence. On the other hand, it follows from Yajima [19] that under some regularity conditions and for constant slowly varying function \( n^\alpha \text{Var}(\hat{\beta}_1) \to c \), where \( c \) is a finite positive constant. Thus \( \hat{\beta}_1 \) is much more variable than \( \tilde{\beta}_1 \). The same property holds for the best linear unbiased estimator (BLUE) of \( \beta_1 \).

The Figure 1 illustrates the gain in efficiency for the randomized design when \( \varepsilon_i \) are sampled from the fractional Gaussian noise with \( n = 100 \) for various values of \( \alpha \). For this sample size the ratio \( \text{Var} (\hat{\beta}_1) / \text{Var} (\tilde{\beta}_1) \) is more than tenfold for \( \alpha = 0.2 \) and approximately twofold for \( \alpha = 0.8 \).
Figure 1. Efficiency of OLSE with prior randomization of grid points wrt OLSE for LRD errors (ratio1) and independent errors (ratio2). Errors are fractional Gaussian noise with $H = 1 - \alpha/2$ and $n = 100$.

In the paper we assume that $(\varepsilon_i)$ is a one-sided moving average (linear) process given by

$$\varepsilon_i = \sum_{k=0}^{\infty} c_k \eta_{i-k}, \quad i = 1, 2, \ldots, \tag{5}$$

where $(\eta_k)_{k=-\infty}^{\infty}$ is a sequence of independent, identically distributed innovations such that $E\eta_1 = 0$, $E(\eta_1^2) = 1$ and $c_k$ satisfy $\sum_{k=0}^{\infty} c_k^2 < \infty$. Moreover, assume that $c_0 = 1$ and $c_k = L(k)k^{-\beta}$ for $k \geq 1$, where $1/2 < \beta < 1$ and $L(\cdot)$ is a function defined on $[0, +\infty)$, slowly varying at infinity and positive in some neighborhood of infinity. Routine calculation based on the Karamata theorem and properties of slowly varying functions (see e.g. Theorem 0.6(a) and Proposition 0.5 in Resnick [15]) implies that the covariance of $(\varepsilon_i)$ decays approximately hiperbolically and thus assumptions on errors stated below (1) are indeed satisfied, namely, $r(k) \sim C(\beta) L^2(k) k^{-\alpha}$, where $C(\beta) := \int_{0}^{\infty} (x + x^2)^{-\beta} dx$ and $\alpha = 2\beta - 1$. Thus in this case sum of absolute values of covariances diverge. Note that if $\sum_{k=0}^{\infty} |c_k| < \infty$, or $\beta > 1$ in the hyperbolic decay condition given above, $(\varepsilon_i)$ is short-range dependent. Indeed, since $r(k) = \sum_{t=0}^{\infty} c_t c_{t+k}$, we have

$$\sum_{k=1}^{\infty} |r(k)| \leq \sum_{k=1}^{\infty} \sum_{t=0}^{\infty} |c_t| |c_{t+k}| \leq \left( \sum_{t=0}^{\infty} |c_t| \right)^2.$$ 

Put $c_k = 0$ for $k < 0$. Then (5) can be written as

$$\varepsilon_i = \sum_{j=-\infty}^{\infty} c_{i-j} \eta_j, \quad i = 1, 2, \ldots, n, \tag{6}$$

Let $\gamma_n^2 = \text{Var}(\varepsilon_1 + \ldots + \varepsilon_n)$. It is easily seen that $\gamma_n^2 = \sum_{k=-\infty}^{n} (\sum_{t=1}^{n} c_{t-k})^2 \sim D(\beta)n^{2-\alpha}L^2(n)$, where $D(\beta) := [(2 - 2\beta)(3/2 - \beta)]^{-1}C(\beta)$. Then noting that $\gamma_n^2 \to \infty$ when $n \to \infty$ it follows from Theorem 18.6.5 in Ibragimov and Linnik [10] that

$$\frac{1}{\gamma_n} \sum_{i=1}^{n} \varepsilon_i \overset{D}{\to} Z, \tag{7}$$

where $Z$ is a standard normal random variable. Note that $\gamma_n^{-1} = o(n^{-1/2})$. 

4
We return now to the models (1) and (2). We will use Priestley-Chao [14] kernel estimator to estimate $g$ in both cases. In the FDR model it is defined as follows

$$\hat{g}_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{x - i/n}{b_n}\right)Y_i, \quad 0 \leq x \leq 1,$$

where the kernel $K$ is some density function chosen by a statistician and bandwidths (smoothing parameters) satisfy natural conditions $b_n \to 0$ and $nb_n \to \infty$. However, the results described below extend to the Gasser-Müller estimator as well as to the local linear regression smoothers (see e.g. Fan and Gijbe ls [7] for definitions of these estimators). A modified Priestley-Chao estimator in the RFDR model is

$$\hat{g}_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{x - \sigma(i)/n}{b_n}\right)Y_i, \quad 0 \leq x \leq 1$$

We estimate $g$ at fixed distinct points $x_1, \ldots, x_k \in (0, 1)$ and show that depending on the size of the bandwidth, two different norming factors are required to get a nondegenerate asymptotic distribution. A small bandwidth case reflects ‘whitening by windowing’ principle i.e. for such bandwidths asymptotic behaviour of regression estimator is the same as for independent errors. A borderline of the dichotomy is established and it turns out to be the same as in the random-design regression (RDR) model with LRD linear errors determined in Csörgő and Mielniczuk [5]. More importantly, for both parts of the dichotomy, asymptotic variances are of higher order, i.e. converge to 0 more quickly, 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 the last mentioned paper and Mielniczuk [12] as artificial randomized variables $\sigma_n(\cdot)/n$ mimic the behaviour of independent explanatory random variables uniformly distributed on $[0,1]$.

In the paper we also investigate the uniform convergence rates of $\hat{g}_n$. We prove that under certain conditions imposed on $g(\cdot), K(\cdot), (\eta_k)_{k=1}^{\infty}$ and bandwidths this rate is $O(\log n \max(L(n) n^{-\alpha/2}, (nb_n)^{-1/2}))$ in probability, reflecting the dichotomous behaviour of $\hat{g}_n(\cdot)$.

The paper concludes with a simulation example showing effect of randomization in practice. It indicates that randomization has a non-negligible impact on integrated square error for sample size around $n = 1000$. Numerical evidence in Efroymovitch [6] suggests that the random design is more robust than fixed design against departures from normality. Our
simulation results indicate that similar robustness property is exhibited by the randomized fixed design.

As an interesting byproduct of the main considerations of the paper we prove in Proposition 1 below that the variance of Priestley-Chao estimator for uniform $K$ does not increase when any fixed permutation is used to change the order of the sampling points provided the covariance of the errors is nonincreasing. In the last section we construct a special fixed permutation such that the pertaining estimator performs on par or better than randomized Priestley-Chao estimator in our limited simulation study. Further research is needed into asymptotic and small sample properties of such estimators.

Results

Before we focus on the case of linear LRD errors consider the following general result for uniform $K(s) = 0.5I_{[-1,1]}(s)$ when Priestley-Chao estimator is known as a running mean. In this case for FDR model we have $\hat{g}_n^f(x) = 0.5(n b_n)^{-1} \sum_{i: |x-i/n| \leq b_n} Y_i$ whereas in RFDR model $\hat{g}_n^r(x) = 0.5(n b_n)^{-1} \sum_{i: |x-\sigma(i)/n| \leq b_n} Y_i$. Then for any fixed permutation $\sigma$ we have $\text{Var}(\hat{g}_n^f(x)) \geq \text{Var}(\hat{g}_n^r(x))$ provided only that the covariance function $r(\cdot)$ is nonincreasing. This follows easily from the following proposition which is proved in the next section.

**Proposition 1.** Suppose that $\pi : \{1, ..., n\} \rightarrow \mathbb{N}$ is 1-1 function and $r(\cdot)$ is nonincreasing function. Then

$$\sum_{1 \leq t \neq s \leq n} r(|t-s|) \geq \sum_{1 \leq t \neq s \leq n} r(|\pi(t) - \pi(s)|).$$

Note that since the above relation between variances holds for any $\sigma$ its analogue is valid for randomly chosen $\sigma$. Moreover, it can be extended to the analogous property of mean integrated square errors. As the property for a random permutation is a counterpart of the result for a fixed distribution obtained by averaging one can expect a greater gain in efficiency for a suitably chosen permutation. The idea to choose a single appropriate permutation to decrease variability of Priestley-Chao estimator is exploited in the last section.

In the following we assume that the RFDR model (2) holds. Let $b = b_n$ and $K_b(x) = b^{-1}K(x/b)$. We use $b_n$ in lieu of $b$ at places when dependence on $n$ needs to be stressed.
The Priestley-Chao estimator given by (9) has the following representation:

\[
\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_b(x - i/n) g\left(\frac{i}{n}\right) + \frac{1}{n} \sum_{i=1}^{n} K_b(x - i/n) \varepsilon_{\sigma^{-1}(i)} =: \tilde{g}_n(x) + A_n(x). \tag{10}
\]

Using (6) the second term in (10) can be written as weighted sum of i.i.d. summands with random coefficients

\[
A_n(x) = \sum_{j=-\infty}^{\infty} v_{j,n} \eta_j, \tag{11}
\]

where \( v_{j,n} := v_{j,n}(x) = (n)^{-1} \sum_{i=1}^{n} K_b(x - i/n) c_{\sigma^{-1}(i)-j} \).

We consider first how the introduced randomization affects properties of long-range dependent errors and two first moments of \( \hat{g}_n(x) \). Let \( a_n^2 = 2((1-\alpha)(2-\alpha))^{-1}L^2(n)n^{-\alpha} \sim \operatorname{Var}(n^{-1} \sum_{i=1}^{n} \varepsilon_i) \).

**Proposition 2.** Let \( \bar{\varepsilon}_{i,n} = \varepsilon_{\sigma^{-1}(i)}, \ i = 1, \ldots, n \). Then

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

(ii) \( \operatorname{Cov}(\bar{\varepsilon}_{i,n}, \varepsilon_{j,n}) \sim a_n^2, \) for \( i \neq j \) and \( \operatorname{Var}(\bar{\varepsilon}_{i,n}) = \operatorname{Var}(\varepsilon_{i,n}) \).

\[
\tag{12}
\]

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

\[
\frac{1}{n(n-1)} \sum_{1 \leq k \neq l \leq n} \operatorname{Cov}(\bar{\varepsilon}_{i,n}, \varepsilon_{j,n}|\sigma(k) = i, \sigma(l) = j) = \frac{1}{n(n-1)} \sum_{1 \leq k \neq l \leq n} \operatorname{Cov}(\varepsilon_{k,n}, \varepsilon_{l,n}),
\]

where the last equality is implied by independence of \( \sigma_n \) and \( (\varepsilon_{i,n}) \). Routine application of the Karamata theorem yields (ii).

**Proposition 3.** Assume that \( K \) is compactly supported and Lipschitz continuous. Then we have

(i) \( \mathbb{E}\hat{g}_n(x) = \tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_b\left(x - \frac{i}{n}\right) g\left(\frac{i}{n}\right); \)

(ii) \( \operatorname{Var}\hat{g}_n(x) = (nb_n)^{-1} \sigma_{\varepsilon}^2 \int K^2(s)ds + a_n^2 + o((nb_n)^{-1} + a_n^2). \tag{13} \)
Property (ii) follows from Proposition 2 and equality

\[
\Var \hat{g}_n(x) = \frac{1}{n^2} \left( \sum_{1 \leq i \neq j \leq n} K_b(x - i/n)K_b(x - j/n)\Cov(\bar{\varepsilon}_{i,n}, \bar{\varepsilon}_{j,n}) + \sum_{i=1}^n K_b^2(x - i/n)\Var(\bar{\varepsilon}_{i,n}) \right).
\]

The main terms in (13) correspond to the limits of the second and first terms in the decomposition above, respectively.

Note that the variance of \(\hat{g}_n(x)\) does not depend on \(g(\cdot)\). Proposition 3 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 dichotomous behavior depending on the size of the bandwidth. Namely, \(\Var \hat{g}_n(x) \sim a_n^{-2} = o(nb_n)\), and is equivalent to \((nb_n)^{-1}\sigma^2 \int K^2(s)\,ds\) when the opposite condition \(nb_n = o(a_n^{-2})\) holds. The results below show that the analogy between behavior of the Priestley-Chao estimator in the RFDR and RDR models extends to asymptotic laws.

Consider distinct points \(x_1, \ldots, x_k \in (0,1)\). Let \(C^1(\mathbb{R})\) denote a family of continuously differentiable real functions.

One part of the smoothing dichotomy, for large bandwidths satisfying \(a_n^{-2} = o(nb_n)\) that allow the long memory of the errors to prevail, 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 \(a_n^{-1} = o(n^{1/2})\) i.e. \(n = o(\Var(\sum_{i=1}^n \varepsilon_i))\).

**Theorem 1.** Assume that \(K\) is Lipschitz continuous supported on \((-1,1)\) and \(a_n^{-2} = o(nb_n)\). Then

\[
a_n^{-1}(\hat{g}_n(x_1) - \bar{g}_n(x_1), \ldots, \hat{g}_n(x_k) - \bar{g}_n(x_k)) \xrightarrow{D} (Z, \ldots, Z),
\]

where \(Z\) is a standard normal random variable and \(\bar{g}_n(x)\) is defined in Proposition 3(i).

If \(g \in C^2(U_x)\) for some neighborhood \(U_x\) of \(x\), \(g\) is Lipschitz and \(K\) is symmetric, it is easily seen that \(\hat{g}_n(x) - g(x) = O(b_n^2 + (nb_n)^{-1})\). Then \(a_n^{-1}(\hat{g}_n(x) - g(x)) \to 0\) provided \(nb_n^5 \to 0\) and in such case \(\hat{g}_n(x_i)\) may be replaced in (14) by \(g(x_i)\), \(i = 1, 2, \ldots, k\).

Set \(\tilde{\sigma}^2 = \sigma^2 \int K^2(s)\,ds\), where \(\sigma^2\) is the variance of the errors. 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 $K \in \mathcal{C}^1(\mathbb{R})$ is supported on $(-1, 1)$ and $nb_n = o(a_n^{-2})$. Then

$$(nb_n)^{1/2} (\hat{g}_n(x_1) - \tilde{g}_n(x_1), \ldots, \hat{g}_n(x_k) - \tilde{g}_n(x_k)) \overset{D}{\to} \tilde{\sigma} (Z_1, \ldots, Z_k),$$

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

The observation before the statement of Theorem 2 suggests that the conclusion should hold for all bandwidths when the errors are short-range dependent.

Theorem 3. Suppose that $\sum_{t=0}^{\infty} |c_t| < \infty$ and $K$ satisfies the assumptions of Theorem 2. Then we have (15).

Next we establish the uniform convergence rate of $\hat{g}_n$. First we state the Bernstein inequality ( cf. Bosq [2], p. 26).

Lemma 1. Let $X_{i,n}, i = 1, \ldots, n$, be mean zero, finite variance independent random variables. Assume, additionally, that they satisfy Cramér’s condition: for some $B = B(n) < \infty$

$$E|X_{i,n}|^k \leq B^{k-2}k! EX_{i,n}^2 \quad k = 2, 3, \ldots, \quad i = 1, \ldots, n.$$  

Let $S_n = \sum_{i=1}^{n} X_{i,n}$, $s_n^2 = \sum_{i=1}^{n} Var(X_{i,n})$. Then, for any $\xi > 0$,

$$P\left(|S_n| > \xi \right) \leq 2 \exp \left\{ \frac{-\xi^2}{4s_n^2 + 2B\xi} \right\}.$$  

Let $r_n^2 = a_n^2 + \sigma^2 \int K^2(s) \, ds/nb_n$. In order to establish a rate of uniform convergence of $\hat{g}_n(\cdot)$ we additionally assume that innovations in (5) satisfy Cramér’s condition and $L(t) \geq 0$ for all $t \geq 0$.

Theorem 4. Assume that $g \in \mathcal{C}^2[0,1]$, $K$ is symmetric and satisfies assumptions of Theorem 1. Suppose that $nb_n^5 = O(1)$, $L(\cdot) \geq 0$ and $\{\eta_t\}$ satisfy Cramér’s condition. Then, for any $0 < \delta < 1/2$ and $D > 1$, we have

$$P\left( \sup_{x \in [\delta, 1-\delta]} |\hat{g}_n(x) - g(x)| > Dr_n \log n \right) \to 0.$$  

(17)
Observe that (17) is a slightly stronger statement than \( \sup_{x \in [\delta, 1 - \delta]} |\hat{g}_n(x) - g(x)| = O_P(r_n \log n) \).
Moreover, the order \( r_n \log n \) reflects the dichotomous behaviour of \( \hat{g}_n(x) \).

**Proofs**

In all proofs \( C \) and \( \check{C} \) denote generic constants, values of which may change.

**Proof of Proposition 1.** Let \( A \) be any set of integers of size \( n \), \( r(A) := \sum_{t \neq s, t \notin A} r(t - s) \) and \( \Delta(A) = \max_{t \in A} \min_{s \in A \setminus t} |t - s| \). It is enough to prove that for \( A \) with \( \Delta(A) > 1 \) there exists \( \hat{A} \) such that \( \Delta(\hat{A}) < \Delta(A) \) and \( r(A) \leq r(\hat{A}) \). Let \( t_0 \in A \) be such that \( \Delta(A) = \min_{s \in A, s \neq t_0} |t_0 - s| > 1 \). We move each element towards \( t_0 \) by one i.e. \( f(t) = t + 1 \) when \( t < t_0 \), \( f(t) = t - 1 \) when \( t > t_0 \) and \( f(t_0) = t_0 \). Then \( \hat{A} = f(A) \) enjoys the listed properties in view of the fact of \( r(\cdot) \) being a nonincreasing function. The proof is concluded by noting that when \( \Delta(A) = 1 \) then \( A \) is a block of \( n \) consecutive integers.

**Proof of Theorem 1.** Let \( k = 1 \) and \( x = x_1 \). Let \( K_b(x) := b^{-1}K(x/b) \). Note that the left-hand side of (14) for \( k = 1 \) can be written as

\[
a_n^{-1}\left(\hat{g}_n(x) - \check{g}_n(x) - EK_b\left(x - \frac{\sigma(1)}{n}\right)\sum_{i=1}^{n} \varepsilon_i\right) + a_n^{-1}EK_b\left(x - \frac{\sigma(1)}{n}\right)\sum_{i=1}^{n} \varepsilon_i =: T_{1,n}(x) + T_{2,n}(x).
\]

It is easy to check that \( EK_b(x - \sigma(1)/n) - 1 \to 0 \) as \( K \) is a Lipschitz continuous function with a compact support integrating to 1. It follows from (7) that \( (1/n\varepsilon_n) \sum_{i=1}^{n} \varepsilon_i \xrightarrow{D} Z \). Thus it is enough to show that \( T_{1,n}(x) \xrightarrow{p} 0 \). Note that \( a_n T_{n,1}(x) = A_n(x) - E(A_n(x)|\varepsilon_1, \ldots, \varepsilon_n) \), where \( A_n(x) \) is defined in (10).

Using the fact that \( \sigma_n \) and \( (\varepsilon_i) \) are independent we have

\[
a_n^2 \mathbb{E}(T_{1,n}^2(x)) = \frac{1}{n^2} \sum_{i,j=1}^{n} \text{Cov}\left(K_b\left(x - \frac{\sigma(i)}{n}\right), K_b\left(x - \frac{\sigma(j)}{n}\right)\right) \mathbb{E}\varepsilon_i \varepsilon_j.
\]

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

\[
\gamma_{ij} = \frac{1}{n(n-1)} \sum_{1 \leq s \neq t \leq n} K_b\left(x - \frac{s}{n}\right) K_b\left(x - \frac{t}{n}\right) - \frac{1}{n^2} \sum_{1 \leq s, t \leq n} K_b\left(x - \frac{s}{n}\right) K_b\left(x - \frac{t}{n}\right)
\]

\[
= \left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) \sum_{1 \leq s, t \leq n} K_b\left(x - \frac{s}{n}\right) K_b\left(x - \frac{t}{n}\right) - \frac{1}{n(n-1)} \sum_{s=1}^{n} K_b^2\left(x - \frac{s}{n}\right)
\]

\[
= O(n^{-1} + (nb_n)^{-1})
\]
Choosing $\xi$ in view of a
In this case (1)
For any
\[ K \]
Using the fact that $\xi > 0$ twice, we obtain that for any
\[ L \]
Let
\[ \text{Proof of Theorem 2.} \]
Lemma
\[ 2 \]
Remark
\[ 1 \]
Theorem 1 can be extended to the case when $(\eta_k)$ are martingale differences. In this case $(1/na_n) \sum_{i=1}^{n} \varepsilon_i \overset{D}{\to} Z$ follows e.g. from Peligrad and Utev [13] The rest of the proof is exactly the same as that of Theorem 1.

In order to prove Theorem 2 we will need some properties of a slowly varying function. Seneta [18] proved the following

**Lemma 2.** Let $x^{-\gamma} \bar{L}(x) = \sup_{t \leq t < \infty} \{t^{-\gamma} L(t)\}, \text{ where } \gamma > 0.$ Then for $x > 0$ (i) $x^{-\gamma} \bar{L}(x)$ is a nonincreasing function; (ii) $x^{-\gamma} \bar{L}(x) > 0$; (iii) $x^{-\gamma} L(x) \leq x^{-\gamma} \bar{L}(x)$; (iv) $L(x) \sim 1/\bar{L}(x)$ as $x \to \infty$ and thus $\bar{L}(x)$ is slowly varying.

**Proof of Theorem 2.** Let $k_i(x) = K((x - i/n)/b_n).$ We prove first the result for $k = 1.$ Using the fact that $K$ is bounded, Cauchy-Schwarz inequality and Karamata theorem twice, we obtain that for any $\xi > 0$ and uniformly in $0 \leq x \leq 1$,
\[
|v_{j,n}| = \frac{1}{nb} \left| \left( \sum_{|\sigma^{-1}(i)-j| > \xi} + \sum_{|\sigma^{-1}(i)-j| \leq \xi} \right) k_i(x) c_{\sigma^{-1}(i)-j} \right|
\leq (nb)^{-1/2} \left( \frac{1}{nb} \sum_{i=1}^{n} k_i^2(x) \right)^{1/2} \left( \sum_{j > \xi} |c_j| \right)^{1/2} + \frac{1}{nb} \max_{1 \leq i \leq n} |k_i(x)| \sum_{|j| \leq \xi} |c_j|
\leq \left[ (nb)^{-1/2} \left( \xi^{-1-\beta} L^2(\xi) \right)^{1/2} + (nb)^{-1} \xi^{1-\beta} |L(\xi)| \right].
\]
Choosing $\xi = nb$ in this bound yields that for all sufficiently large $n$,
\[
\max_{1 \leq j \leq n} \sup_{0 \leq x \leq 1} |v_{j,n}| \leq C(nb)^{-\beta} L(nb).
\]
(18)
For any $N = N_n > 0$ and all $0 < x < 1$ using (11) we have
\[
A_n(x) = S_{-\infty}^{-N-1} + S_{-N}^{-N} + S_{N+1}^{-\infty},
\]
(19)
where \( S_p^q = S_p^q(x) = \sum_{j=p}^q v_{j,n} \eta_j \). Let \( n = o(N_n) \). Then \( S_{N+1}^N \equiv 0 \) for sufficiently large \( n \). Thus (15) will follow for \( k = 1 \) by Slutsky’s theorem if we show that

\[
(nb_n)^{1/2} S_{-\infty}^{N-1} \overset{p}{\to} 0 \quad \text{and} \quad (nb_n)^{1/2} S_{-N}^{N} \overset{D}{\to} N(0, \sigma^2).
\]

(20)

We observe that

\[
Var(S_{-\infty}^{N-1}) = \frac{1}{(nb)^2} \sum_{j<-N} \left[ \sum_{i=1}^n k_i^2(x) \mathbb{E}(c_{\sigma^{-1}(i-j)}) + \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \mathbb{E}(c_{\sigma^{-1}(i_1-i_2)j}) \right]
=: \Theta_{1,n} + \Theta_{2,n}.
\]

(21)

It is easy to see using square summability of \( c_j \) and \( \sum k_i^2(x) = O(nb_n) \) that

\[
\Theta_{1,n} \leq (nb)^{-2} \sum_{i=1}^n k_i^2(x) \sum_{j>N_n} c_j^2 = o((nb)^{-1}).
\]

Moreover,

\[
\Theta_{2,n} = (nb)^{-2} \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \frac{2}{n(n-1)} \sum_{p=0}^n \sum_{j>N_n} c_j c_{j+(q-p)}.
\]

Let \( \tilde{c}_j = j^{-\beta} L(j) = \sup_{t \leq \tilde{t}<\infty} \{t^{-\beta} L(t)\} \). By Lemma 2 and the fact that \( c_j > 0 \) for large \( j \) we have

\[
\Theta_{2,n} \leq \frac{1}{(nb)^2} \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \frac{2}{n-1} \sum_{k=1}^n \sum_{j>N} \tilde{c}_j \tilde{c}_{j+k}
\leq \frac{1}{(nb)^2} \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \frac{2}{n-1} \sum_{k=1}^n \sum_{j>N} \tilde{c}_j^2 \leq C \bar{L}^2(N) N^{1-2\beta} \sim C L^2(N) N^{-\alpha} = o(a_n^2)
\]

as \( n = o(N) \), where the last equality follows from the Karamata representation of slowly varying functions. Thus \( \text{Var}(nb_n^{1/2} S_{-\infty}^{N-1}) = o(1) \) in view of \( nb_n = o(a_n^{-2}) \).

Let \( V_n(x) := (nb_n)^{1/2} S_{-N}^{N} = \sum_{t=1}^{2N+1} x_{t,n} \), where \( x_{t,n} = (nb_n)^{1/2} v_{t-N-1,n} \eta_{t-N-1} \). Reasoning as in the proof of Theorem 2 in Csörgő and Mielniczuk [4] one can check that

\[
\mathbb{E} \left[ \exp \left( it V_n(x) | \sigma_n \right) \right] \overset{p}{\to} \exp(-t^2 \dot{\sigma}^2/2),
\]

(22)

with \( \dot{\sigma}^2 = \int K^2(s) ds \), implies convergence in distribution of \( V_n(x) \) to \( N(0, \dot{\sigma}^2) \). Note that given \( \sigma_n \), \( V_n(x) \) is a sum of independent random variables such that \( \mathbb{E}(V_n(x) | \sigma_n) = 0 \) and \( \text{Var}(V_n(x) | \sigma_n) \to \dot{\sigma}^2 \) in probability. The second fact follows from (13) and \( \text{Var}((nb_n)^{1/2} S_{-\infty}^{N-1}) \)
→ 0. Thus in order to prove (22) it is enough to check the conditional Lindeberg condition. It amounts to showing
\[ nb_n \sum_{t=1}^{2N+1} v_{t-N-1,n}^2 \mathbb{E}(\eta_{t-N-1}^2 I\{ (nb_n)^{1/2} |v_t - N - 1, n| \geq \delta \} | \sigma_n) \rightarrow 0 \] (23)
in probability for any \( \delta > 0 \). However, in view of (18) expression in (23) is bounded by
\[ nb_n \sum_{t=1}^{2N+1} v_{t-N-1,n}^2 D_n, \] where \( D_n = \mathbb{E}(\eta_{t-N-1}^2 I\{ |\eta_t - N - 1| \geq C\delta (nb_n)^{\beta-1/2} \} | \sigma_n) \). Note that \( D_n \rightarrow 0 \) as \( nb_n \rightarrow \infty \). Thus (23) follows as
\[ nb_n \sum_{t=1}^{2N+1} \mathbb{E}v_{t-N-1,n}^2 \leq nb_n \text{Var}(A_n(x)) = nb_n \text{Var}(\hat{g}_n(x)) = O(1) \]
in view of Proposition 3(ii).

The general case is proved analogously using Cramér-Wald device. Namely, compactness of support of \( K \) implies that \( \text{Var}(d_1 V_n(x_1) + \ldots + d_k V_n(x_k)) \rightarrow (d_1^2 + \ldots + d_k^2) \sigma^2 \) and checking the conditional Lindeberg condition proceeds in the same way as above.

**Remark 2.** By appealing to martingale CLT instead of Lindeberg theorem one can extend Theorem 2 to martingale differences \( \{\eta_t\} \). The proof differs from that of Theorem 2 at one place, namely when establishing the second part in (20). In this case \( \{x_{t,n}\} \) is a sequence of martingale differences with respect to \( \mathcal{F}_{t,n} = \sigma(\ldots, \eta_{t-1}, \eta_t, \sigma_n) \). Then it is enough to check the usual sufficient condition for an analogue of Lindeberg condition, namely
\[ \sum_{t=1}^{2N+1} \mathbb{E}(x_{t,n}^2 I\{|x_{t,n}| \geq \delta \}) \leq nb_n \sum_{t=1}^{2N+1} \mathbb{E}v_{t-N-1,n}^2 D_n \rightarrow 0 \]
as \( n \rightarrow \infty \) for any \( \delta > 0 \).

**Proof of Theorem 3.** The proof is similar to that of Theorem 2. Using the fact that \( K \) is bounded we have
\[ |v_{j,n}| \leq (nb_n)^{-1} \max_{1 \leq i \leq n} |k_i(x)| \sum_{i=1}^n |c_{\sigma^{-1}(i)-j}| \leq C (nb_n)^{-1}. \] (24)

The variance of the left-hand side of (15) for \( k = 1 \) equals \( nb \sum_{j=-\infty}^{\infty} \mathbb{E}v_{j,n}^2 \) and can be written as
\[ (nb_n)^{-1} \sum_{i=1}^n k_i^2(x) \sum_{j=-\infty}^{\infty} \mathbb{E}(c_{\sigma^{-1}(i)-j}) + (nb_n)^{-1} \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \sum_{j=-\infty}^{\infty} \mathbb{E}(c_{\sigma^{-1}(i_1)-j}c_{\sigma^{-1}(i_2)-j}). \]
The first term is \( r(0) \int K^2(s)ds + o(1) \) and the second term equals 
\[
\frac{1}{n(n-1)} \sum_{t \neq s} r(|t - s|) \leq Cn b_2^4 = o(1).
\]

The reasoning is slightly different when showing that \((nb)^{1/2} S_{-\infty}^{-N-1}\), where \(S_{-\infty}^{-N-1}\) is defined in (19), is negligible. Namely
\[
\Theta_{2,n} \leq \frac{1}{(nb)^2} \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \frac{2}{n(n-1)} \sum_{p<q,j>N} |c_j c_{j+(q-p)}| 
\leq \frac{1}{(nb)^2} \sum_{i_1 \neq i_2} k_{i_1}(x)k_{i_2}(x) \frac{2}{n(n-1)} \sum_{p=1}^n \sum_{j>N} |c_j| \sum_{s=1}^n |c_{j+s}| = o\left(\frac{1}{n}\right).
\]

Checking the conditional Lindeberg condition (23) proceeds analogously with \(D_n = \mathbb{E}(\eta_{-N-1} I\{(nb_2)^{1/2} |\eta_{-N-1} v_{-N-1,n}| \geq \delta\}) \leq \mathbb{E}(\eta_2^2 I\{|\eta_{-N-1}| \geq C\delta (nb)^{1/2}\})\) in view of (24).

The following two lemmas are needed to prove Theorem 4. In the first one we assume more generally that \(\pi_n : \{1,2,\ldots,n\} \rightarrow \mathbb{N}\).

**Lemma 3.** Let \(\varepsilon_i\) be defined in (6) and \(\pi_n\) be a random 1-1 transform of \(\{1,2,\ldots,n\} \rightarrow \mathbb{N}\) independent of \(\varepsilon_1,\ldots,\varepsilon_n\). Suppose that \(\eta_i\) satisfy the Cramér condition. Then
\[
\max_{1 \leq i \leq n} |\varepsilon_{\pi(i)}| = \mathcal{O}_P(\log n). \quad (25)
\]

Recall that \(r_n^2 = a_n^2 + \sigma_\varepsilon^2 \int K^2(s)ds/(nb_n)\).

**Lemma 4.** Let \(s_{n,1}^2 = \text{Var}(S_{-N}^N(x)|\sigma_n)\). Then \(\text{Var}(s_{n,1}^2) = o(r_n^4)\) provided \(n = o(N)\) and \(L(\cdot)\) is nonnegative.

**Proof of Lemma 3.** The proof uses the Bernstein inequality stated in Lemma 1 and the truncation method used in the proof of Theorem 2. For any \(N = N_n > 0\) rewrite the moving average process
\[
\varepsilon_{\pi_n(i)} = T_{i,-\infty}^{-N-1} + T_{i,-N}^N + T_{i,N+1}^\infty,
\]
where \(T_{i,p}^q = \sum_{j=p}^q c_{\pi_n(i)-j} \eta_j\). Let \(n = o(N_n)\). Then \(T_{i,N+1}^\infty \equiv 0\) for sufficiently large \(n\).

It is easy to see that as \(c_j\) are bounded, the random variables \(X_j := c_{\pi_n(i)-j} \eta_j\) satisfy the Cramér condition with some constant \(\tilde{B}\) which does not depend on \(\pi_n\).
Note that \( s_n^2 = \text{Var}(T_i^{N,N}|\pi_n) = \sum_{j=-N}^{N} c_{\pi_n(i)-j}^2 \leq \sum_{j=0}^{\infty} c_j^2 = C \) for any \( i \). Applying the Bernstein inequality with \( \xi_n = A \log n \), where \( A > 2B \) we obtain that

\[
P\left( |T_i^{N,N}| > \xi_n \right) \leq 2 \exp \left\{ \frac{-A^2 \log^2 n}{4C + 2AB \log n} \right\}.
\]

The right hand side of above inequality does not depend on \( \pi_n \) and note that in calculation of \( \text{Var}(\sum_{i=1}^{N}) \) indices are of the stated order. Consider \( \text{Var}(\sum_{i=1}^{N}) \) first. Observe that

Next, we observe that \( \text{Var}(T_i^{-N,-\infty}) = \sum_{j<0} \text{E}(c_{\pi_n(i)-j}^2) \leq \sum_{j>N} c_j^2 = \mathcal{O}(L^2(N) N^{-\alpha}) \).

Thus, by choosing \( N \) such that \( N \sim n^{1/\alpha+\varepsilon} \) for some \( \varepsilon > 0 \) we have

\[
P\left( \max_{1 \leq i \leq n} |T_i^{N,N}| > \xi_n \right) \leq C n \frac{L^2(N) N^{-\alpha}}{\log^2 n} = o(1).
\]

Proof of Lemma 4. Observe that we have \( s_n^2 = s_{n,1}^2 + s_{n,2}^2 \), where \( s_n^2 = \text{Var}(A_n(x)|\sigma_n) \) and \( s_{n,2}^2 = \text{Var}(S_{-N,-\infty}(x)|\sigma_n) \). Moreover, \( \text{Var}(s_{n,1}^2) \leq \text{Var}(s_n^2) + \text{Var}(s_{n,2}^2) + |2\text{Cov}(s_{n,1}^2, s_{n,2}^2)| \leq 2(\text{Var}(s_n^2) + \text{Var}(s_{n,2}^2)) \). Thus it is enough to show that both terms on the rhs of the last equality are of the stated order. Consider \( \text{Var}(s_n^2) \) first. Observe that

\[
s_n^2 = (nb_n)^{-2} \sum_{1 \leq i_1, i_2 \leq n} k_{i_1}(x)k_{i_2}(x)r(||\sigma^{-1}(i_1) - \sigma^{-1}(i_2)||)
\]

and note that in calculation of \( \text{Var}(s_n^2) \) the last sum can be restricted to pairs of different indices \( i_1, i_2 \). Thus \( \text{Var}(s_n^2) \) equals

\[
(nb_n)^{-4} \sum_{i_1, i_2, j_1, j_2}^{} k_{i_1}(x)k_{i_2}(x)k_{j_1}(x)k_{j_2}(x)\text{Cov}(r(||\sigma^{-1}(i_1) - \sigma^{-1}(i_2)||), r(||\sigma^{-1}(j_1) - \sigma^{-1}(j_2)||)),
\]

where \( \sum^* \) is over all \( 1 \leq i_1, i_2, j_1, j_2 \leq n \) such that \( i_1 \neq i_2 \) and \( j_1 \neq j_2 \). Write the above sum as \( T_0 + T_1 + T_2 \), where definition of \( T_i \) pertains to the cardinality \( i \) of \( \{i_1, i_2\} \cap \{j_1, j_2\} \). Consider first \( T_0 \). Then as \( L(\cdot) \) is nonnegative

\[
\text{E}(r(||\sigma^{-1}(i_1) - \sigma^{-1}(i_2)||)r(||\sigma^{-1}(j_1) - \sigma^{-1}(j_2)||) =
\]

\[
\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k_1, k_2, k_3, k_4} r(|k_1 - k_2|)r(|k_3 - k_4|) \leq \frac{1}{n(n-1)(n-2)(n-3)} \left( \sum_{k_1, k_2} r(|k_1 - k_2|) \right)^2 \sim \frac{n^4}{n(n-1)(n-2)(n-3)} a_n^4 \sim a_n^4.
\]
where the sum in the second line is over different \(k_1, k_2, k_3, k_4\). As \(\mathbb{E}^2(r(|\sigma^{-1}(i_1) - \sigma^{-1}(i_2)|)) \sim a_n^4\), it follows that \(T_0 = o(a_n^4)\). Further we have that \(T_1 = O(a_n^4/(nb_n))\). Namely, the order follows from the inequality

\[
\frac{1}{n(n-1)(n-2)} \sum_{k_1,k_2,k_3} r(|k_1-k_2|)r(|k_1-k_3|) \leq \frac{n}{n(n-1)(n-2)} \left( \sum_{k=1}^{n} r(k) \right)^2 = O(a_n^4).
\]

In the same way we get \(T_2 = o((nb_n)^{-2})\). Indeed,

\[
\mathbb{E}(r^2(|\sigma^{-1}(i_1) - \sigma^{-1}(i_2)|)) = \frac{1}{n(n-1)} \sum_{k_1,k_2} r^2(|k_1-k_2|) \leq \frac{n}{n(n-1)} \sum_{k=1}^{n} r^2(k).
\]

The last expression is \(O(n^{-1})\) for \(\alpha > 1/2\) and \(O(a_n^4)\) for \(\alpha < 1/2\). For \(\alpha = 1/2\) we have \(\sum_{k=1}^{n} r^2(k) = O(n^\rho \log n)\) for any \(\rho > 0\) as \(L(i) = O(i^\rho)\). Thus we have \(T_2 = O(max(n^\rho, a_n^4)/(nb_n)^2)\) for any \(0 < \varepsilon < 1\). The reasoning for \(s_{n2}^2\) is similar to that for \(\Theta_{2,n}\) in Theorem 2 and consists in showing that \(\mathbb{E}((s_{n2}^2)^2) \leq CL^4(N)N^{-2\alpha} = o(a_n^4)\) for \(n = o(N)\). We omit the details.

**Proof of Theorem 4.** By (10) and the triangle inequality,

\[
\sup_{x \in [\delta, 1-\delta]} |\hat{g}_n(x) - g(x)| \leq \sup_{x \in [\delta, 1-\delta]} |A_n(x)| + \sup_{x \in [\delta, 1-\delta]} |\hat{g}_n(x) - g(x)|.
\]

If \(g \in C^2[0,1]\), \(g\) is Lipschitz and \(K\) is symmetric it is easily seen that for any \(\delta > 0\), \(\sup_{x \in [\delta, 1-\delta]} |\hat{g}_n(x) - g(x)| = O(b^2 + (nb)^{-1})\). As \(nb_5^2 = O(1)\), we obtain that the second term in the above bound is of order \(o(\log n r_n)\).

To deal with the first term, let \(l = l_n\) be a sequence of positive integers such that \(l_n \sim n\). Let \(\delta = x_0 < x_1 < \ldots < x_{l_n} = 1 - \delta\) denote an equipartition of the interval \([\delta, 1-\delta]\). Write \(A_n(x) = A_n(x_k) + A_n(x) - A_n(x_k)\), so that

\[
\sup_{x \in [\delta, 1-\delta]} |A_n(x)| \leq \max_{1 \leq k \leq l_n} |A_n(x_k)| + \max_{1 \leq k \leq l_n} \sup_{x_{k-1} \leq x \leq x_k} |A_n(x) - A_n(x_k)|. \tag{26}
\]

By decomposition (19), choosing \(N = N_n > n\), we have

\[
\max_{1 \leq k \leq l_n} |A_n(x_k)| \leq \max_{1 \leq k \leq l_n} |S_{-\infty}^{-N-1}(x_k)| + \max_{1 \leq k \leq l_n} |S_N^{-N}(x_k)|. \tag{27}
\]

Proof of Theorem 2 indicates that \(\text{Var}(S_{-\infty}^{-N-1}) = O\left(L^2(N)N^{-\alpha}\right)\). Thus

\[
P\left( \max_{1 \leq k \leq l_n} |S_{-\infty}^{-N-1}(x_k)| > \log n r_n \right) \leq CN \frac{L^2(N)}{(r_n^2 N^\alpha \log^2 n)} = o(1) \tag{28}
\]
for $N$ such that $(n^{-2})^{1/\alpha} = o(NL^{-2/\alpha}(N))$ which is satisfied when $N \sim n^{2/\alpha+\varepsilon}$, for any $\varepsilon > 0$.

Next, consider the second term on the rhs of (27). Note that given $\sigma_n$, $S_{N}^{N}(x)$ is a sum of independent r.v.’s. Let $X_j := v_{j,n} \eta_j$. Since $\eta_j$ satisfy the Cramér’s condition, we obtain

$$
\mathbb{E}\left(|X_j|^k | \sigma_n \right) \leq |v_{j,n}|^k B^{k-2}k! \mathbb{E}|\eta_j|^2 \leq |v_{j,n}|^{k-2}B^{k-2}k! \mathbb{E}\left(X_j^2 | \sigma_n \right) \leq \tilde{B}^{k-2}k! \mathbb{E}\left(X_j^2 | \sigma_n \right),
$$

where $\tilde{B} = O(L(nb_n)/(nb_n)^{\beta})$ uniformly in $x$ in view of (18). Thus conditionally on $\sigma_n$, $X_j$ satisfy the Cramér’s condition with a constant $\tilde{B}$.

In order to apply the Bernstein inequality given $\sigma_n$, we need to analyze the variance of these r.v.’s. Recall that

$$
S_{n,1}^2 = Var(S_{N}^{-N}(x)|\sigma_n) = \Theta_{1,n} + \Theta_{2,n},
$$

where $\Theta_{1,n} = (nb)^{-2} \sum_{i=1}^{n} k_i^2(x) \sum_{j=-N}^{N} c_{\sigma^{-1}(i)-j} = (nb)^{-2} \sum_{i=1}^{n} k_i^2(x) \sum_{j=\sigma^{-1}(i)+N}^{\sigma^{-1}(i)-N} c_j$

and

$$
\Theta_{2,n} = \frac{1}{(nb)^2} \sum_{i \neq i_2} k_i(x)k_{i_2}(x) \sum_{j=-N}^{N} c_{\sigma^{-1}(i_1)-j}c_{\sigma^{-1}(i_2)-j} \geq 0
$$

by assumption that $L(\cdot) \geq 0$. Thus taking $N \geq 2n$ we see that $s_{n,1}^2 \geq C/nb_n$, where $C$ is positive and does not depend on $\sigma$. Applying the Bernstein inequality with $\xi_n = \log ns_{n,1}$ for given $\sigma_n$ we obtain

$$
P\left(|S_{N}^{-N}(x)| > \xi_n | \sigma_n \right) \leq 2 \exp \left\{ \frac{-\log^2 n}{4C + 2B \log n/s_{n,1}} \right\}.
$$

Note that in view of the bound on $s_{n,1}$ and (18) we have for some $D$

$$
\tilde{B} \log n/s_{n,1} \leq DL(nb_n) \log n\ (nb_n)^{-\beta+1/2} = o(\log n).
$$

Thus the rhs of (29) is bounded by a sequence independent of $\sigma$ of the form $2 \exp\{-\log^2 n/(4C + A \log n)\}$ for arbitrary small $A$ when $n$ is sufficiently large. Integrating (29) wrt $\sigma$ and applying ensuing inequality for every grid point, we obtain

$$
P\left( \max_{1 \leq k \leq l_n} |S_{N}^{-N}(x_k)| > \xi_n \right) \leq 2l_n \exp \left\{ \frac{-\log^2 n}{4C + A \log n} \right\} = o(1)
$$

for $A < 1$. Note that in view of Chebyshev inequality and Lemma 4 we have for any $d > 0$

$$
P(|s_{n,1}^2 - \mathbb{E}s_{n,1}^2| \geq d\mathbb{E}s_{n,1}^2) \leq \text{Var}(s_{n,1}^2)/(d\mathbb{E}s_{n,1}^2)^2 \rightarrow 0 \text{ and thus } P(s_{n,1} > Dr_n) \leq P(s_{n,1}^2 > D^2\mathbb{E}s_{n,1}^2) \rightarrow 0 \text{ for } D > 1.
$$

This together with (27), (28) and (30) implies that

$$
P\left( \max_{1 \leq k \leq l_n} |A_n(x_k)| > D \log n r_n \right) = o(1).
$$
Now, it is enough to prove that second term in the bound of (26) is $O_P(r_n \log n)$. Using the Lipschitz condition for the kernel function $K$ and Lemma 3, we have

$$
\max_{1 \leq k \leq l_n} \sup_{x_{k-1} \leq x \leq x_k} \left| \frac{1}{nb} \sum_{i=1}^{n} \left[ K\left( \frac{x - \sigma(i)/n}{b} \right) - K\left( \frac{x_k - \sigma(i)/n}{b} \right) \right] \cdot \varepsilon_i \right|
$$

\leq \max_{1 \leq k \leq l_n} \sup_{x_{k-1} \leq x \leq x_k} \frac{C}{nb} \frac{\max_{1 \leq k \leq l_n} \sup_{x_{k-1} \leq x \leq x_k} \left| x_k - x \right|}{b} \sum_{i=-nx-nb}^{nx+nb} |\varepsilon_{\sigma^{-1}(i)}| \\
= O_P( (nb)^{-1} \log n ) = o_P(r_n \log n).

\[ \blacksquare \]

**Simulation Study**

A limited simulation study has been conducted to investigate the effect of randomization of the fixed design regression in practice. Series $(Y_i)$ of length $n = 1025$ has been generated with trend functions $g_1(x) = 2 \sin(4\pi x)$ and $g_2(x) = 2 - 5x + 5 \exp\{-100(x - 0.5)^2\}$. The considered errors follow a fractionally differenced FD($d$) (cf. Granger and Joyeux [8] and Hosking [9]) with $d = 0, 0.1, 0.2, 0.3, 0.4$. It is known that $L(n) \sim C$ and one-sided moving average representation exists in this case. For FD($d$) process $\varepsilon_t = (1 - B)^{-d}\eta_t$, where $(\eta_t)$ is a Gaussian white noise with marginal variance $\sigma^2_\eta$ and $B\eta_t = \eta_{t-1}$, we have $C = \sigma^2_\eta \Gamma(1-2d)/\Gamma(d)\Gamma(1-d)$. We refer to Beran [1] for more information on this process. The sample paths of the process have been obtained with aid of `fracdiff.sim` procedure in R package `fracdiff` which uses Fast Fourier Transform to generate a Gaussian process with a given covariance. Besides RFDR and FDR models investigated in the paper we also considered Random Design Regression (RDR) model in which independent explanatory random variables are uniformly distributed on $[0,1]$ and independent of errors.

Moreover, we also consider a modification of the RFDR model, called RFDR-p, in which a permutation $\sigma$ is not random but chosen in advance. The permutation is constructed by induction in such a way that adjacent and close values are mapped to values far apart. We give here the construction for $n = 2^N + 1$ only. Define $\sigma(1) = n$, $\sigma(2) = 1$ and $\sigma(3) = (\sigma(1) + \sigma(2))/2 = 2^{N-1} + 1$. In the first step $\sigma(4)$ and $\sigma(5)$ are defined as the following averages: $\sigma(4) = (\sigma(2) + \sigma(3))/2 = 2^{N-2} + \sigma(2)$ and $\sigma(5) = (\sigma(1) + \sigma(3))/2 = 2^{N-2} + \sigma(3)$. In the $k^{th}$ step ($k = 2, \ldots, N-1$) assume that $\sigma(1), \ldots, \sigma(2^k + 1)$ are already defined. Then $\sigma(2^k + 2), \ldots, \sigma(2^{k+1} + 1)$ are determined as specific averages of already defined values, namely $\sigma(2^k + i) = 2^{N-k-1} + \sigma(i)$ for $i = 2, 3, \ldots, 2^k + 1$. 

18
We compared the performance of estimators (8) and (9) and that of their obvious counterparts in the RDR and the RFDR-p models. The number of replications of each experiment was 1000. The employed kernel was either 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 discuss the choice of bandwidth in both cases.

**RFDR model** As it can be conjectured from Proposition 3 that the asymptotic form of 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 [17] data-based bandwidth commonly used for independent data. R procedure `dpill` from KernSmooth package has been employed for its calculation. The same bandwidth choice method was used for RDR and RFDR-p models.

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

$$b_n^f = \left(\frac{\alpha D_1}{D_2}\right)^{1/(4+\alpha)} n^{-\alpha/(4+\alpha)}, \quad (31)$$

where $D_1 = C \iint (x-y)^{-\alpha} K(x) K(y) dx dy$ and $D_2 = (\int s^2 K(s) ds)^2 \int g''^2(s) ds$. In order to develop a version of $b_n^f$ a local Whittle method proposed by Robinson [16] was used to estimate $\alpha$. More precisely, a pair $(d, G)$ is estimated, where $\alpha = 1 - 2d$ and a spectral density $f$ of $(\epsilon_i)$ satisfies $f(\lambda) \sim G\lambda^{-2d}$ for $\lambda \to 0$. As it can be checked that for FD$(d)$ process $\sigma_n^2 = 2\pi G$ we have that $C = 2\pi G \Gamma(1 - 2d)/\Gamma(d)\Gamma(1 - d)$. An employed bandwidth 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 $(\epsilon_i)$ (generated from FARIMA model) are known as well as the value of $\int g''^2(s) ds$ and the previous relation linking $C$ with $\alpha$ and $G$. Thus $\hat{\alpha}$ and $\hat{G}$ are local Whittle estimators based on $(\epsilon_i)$, $\hat{C}$ is obtained from them under FARIMA model, and then $\hat{\alpha}$ and $\hat{C}$ are plugged in (31). In this way we give this method an advantage over bandwidth choice method for RFDR model which is completely data-dependent. Fully data-dependent bandwidth for FDR model performed significantly worse. The results of the simulation study are summarized in Tables 1 and 2. Medians of a distribution of Integrated Squared Error $ISE = n^{-1} \sum_{i=1}^n (\hat{g}(x_i) - g(x_i))^2$ are used as a measure of performance as the distribution is pronouncedly skewed. Despite the fact that some quantities were assumed known for the FDR model, the performance of
Priestley-Chao estimator in the presence of LRD is inferior to its performance when prior randomization is used. The ratio of medians of ISE for FDR to ISE for RFDR is more than 2 for \( d \geq 0.2 \) in the case of normal kernel. The same qualitative conclusion holds when theoretically optimal bandwidths are used for both models instead of their empirical counterparts (results not shown). What is also remarkable is much better performance of regression estimator under randomized discrete uniform grid than for the case when explanatory variables were uniformly distributed on \([0,1]\). This is likely to be caused by the fact that explanatory variables are more evenly distributed across the interval \([0,1]\) when the equispaced grid is used. It has been noticed by Efromovitch [6] that RDR model is much more robust against departures from independence of errors than FDR. Tables 1 and 2 indicate that RFDR method also enjoys this property. Note also that the results for a fixed permutation are consistently on par or better than for RFDR model. This is in concordance with discussion at the beginning of Section 2 for the running mean. An interesting open problem is to choose a permutation in a data-dependent way which would asymptotically minimize the MISE of the Priestley-Chao estimator in the RFDR-p model for LRD errors.

Table 1: Medians of Integrated Squared Error for \( g_1(x) \)

<table>
<thead>
<tr>
<th>kernel</th>
<th>normal</th>
<th>Epanechnikov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FDR</td>
<td>RDR</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0191</td>
<td>0.0347</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0310</td>
<td>0.0365</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0649</td>
<td>0.0440</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1612</td>
<td>0.0725</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5025</td>
<td>0.2547</td>
</tr>
</tbody>
</table>

Table 2: Medians of Integrated Squared Error for \( g_2(x) \)
<table>
<thead>
<tr>
<th>kernel</th>
<th>normal</th>
<th>Epanechnikov</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>FDR</td>
<td>RDR</td>
</tr>
<tr>
<td>0</td>
<td>0.0482</td>
<td>0.0966</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0670</td>
<td>0.0987</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1072</td>
<td>0.1110</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2161</td>
<td>0.1417</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5770</td>
<td>0.3233</td>
</tr>
</tbody>
</table>

**Acknowledgment** Detailed comments of an anonymous reviewer which helped to improve the presentation of the paper are acknowledged.

**References**


Artur Bryk, Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, e-mail: abryk@mini.pw.edu.pl

J. Mielniczuk, Institute of Computer Science PAS, Ordona 21, 01-237, Warsaw; Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, e-mail: miel@mini.pw.edu.pl