EMPIRICAL PROCESS OF RESIDUALS FOR
REGRESSION MODELS WITH LONG MEMORY ERRORS

BY RAFAL KULIK AND Paweł LOREK

University of Ottawa and University of Wrocław

We consider the residual empirical process in random design regression with long memory errors. We establish its limiting behaviour, showing that its rates of convergence are different from the rates of convergence for the empirical process based on (unobserved) errors. Also, we study a residual empirical process with estimated parameters. Its asymptotic distribution can be used to construct Kolmogorov-Smirnov, Cramér-Smirnov-von Mises, or other goodness-of-fit tests. Theoretical results are justified by simulation studies.

1. Introduction. Consider a random design regression model,

\( Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \)

where \( \{X, X_i, i \geq 1\} \) is a stationary sequence of random variables with a density \( f = f_X \), independent of a centered, stationary long memory error sequence \( \{\varepsilon, \varepsilon_i, -\infty < i < \infty\} \), with a distribution \( F_\varepsilon \) and density \( f_\varepsilon \). The goal of this paper is to study the asymptotic properties of the empirical process of residuals,

\[ \hat{K}_n(x) := \sum_{i=1}^{n} \left( 1_{\{\varepsilon_i \leq x\}} - F_\varepsilon(x) \right), \]

where \( \hat{\varepsilon}_i = Y_i - \hat{m}(X_i) \) and \( \hat{m}(\cdot) \) is an estimator of the function \( m(\cdot) \).

Residual-based inference is a standard tool in regression analysis. With this in mind, several authors considered empirical process of residuals in case of weakly dependent stationary time series, starting with [3], followed by [9] or [1], just to mention few. A general message from these papers is that an estimation of parameters (in a parametric regression) does not influence empirical process of residuals. Similar conclusion can be drawn in nonparametric case, see e.g. [5].
As for stationary long memory errors, a picture is less clear. In [4], the authors obtained that in case of a parametric regression, \( m(x) = \beta_0 + \beta_1 x \), with a known intercept, the limiting behaviour of \( \hat{K}_n(\cdot) \) is similar to the limiting behaviour of

\[
K_n(x) := \sum_{i=1}^{n} \left( 1_{\{\varepsilon_i \leq x\}} - F_{\varepsilon}(x) \right),
\]

in the sense that \( \sigma_{n,1}^{-1}K_n(\cdot) \) and \( \sigma_{n,1}^{-1}\hat{K}_n(\cdot) \) converge weakly to, respectively, \( f_{\varepsilon}(x)Z_1 \), where \( Z_1 \) is standard normal and \( \sigma_{n,1} \) is an appropriate scaling factor. However, if one considers a parametric regression when both slope and intercept are unknown, from the latter paper one can only conclude that

\[
\sigma_{n,1}^{-1}\sup_{x \in \mathbb{R}}|\hat{K}_n(x)| = o_P(1).
\]

The main goal of this paper is to establish an asymptotic behaviour for \( \hat{K}_n(\cdot) \) in case of parametric regression with unknown intercept. Also, we obtain results for nonparametric regression. We will show in this paper, that convergence properties of \( \hat{K}_n(\cdot) \) may be completely different from the asymptotics of \( K_n(\cdot) \). To do this, we will establish a second order expansion for \( \hat{K}_n(\cdot) \) (see Theorems 3.1 and 3.2), which will be applied to both parametric and nonparametric regression (see Sections 3.3 and 3.4, respectively).

The established results can be used, in principle, to test whether the errors \( \varepsilon_1, \ldots, \varepsilon_n \) are consistent with a given distribution \( F_{\varepsilon} \). If \( F_{\varepsilon} \) belongs to a one-parameter family \( \{ F_{\varepsilon}(\cdot, \theta), \theta \in \mathbb{R} \} \), then one needs to know the value of the parameter \( \theta \). Therefore, we discuss asymptotic properties of an empirical process of residuals, when a parameter \( \theta \) is estimated. The appropriate limit theorems are established in Section 3.5. Our theoretical results are confirmed by small simulation studies in Section 4.

The results for empirical processes in Sections 3.3 and 3.4 can be applied directly to establish limiting behaviour of quantiles. Furthermore, in a spirit of [7, Section 3], it is tempting to use empirical process to establish limiting behaviour of nonparametric error density estimator. However, in view of results in [12], such approach does not seem to be appropriate. Finally, it would be interesting to establish corresponding results in case of fixed-design regression.

2. Preliminaries: LRD error sequence. In the sequel, \( F_U(\cdot), f_U(\cdot) \) denote a distribution and a density, respectively, of a given random variable
Also, if $U$ has finite mean, we denote $U^* = U - E[U]$.

We shall consider the following assumption on the error sequence:

\[(E)\] \(\epsilon_i, i \geq 1\), is an infinite order moving average

\[\epsilon_i = \sum_{k=0}^{\infty} c_k \eta_{i-k}, \quad \text{with } c_0 = 1,\]

where \(\eta_i, -\infty < i < \infty\), is a sequence of centered i.i.d. random variables, independent of \(X_i\), \(i \geq 1\). We assume that \(E[\epsilon^4] < \infty, E[\epsilon^2] = 1\), and for some \(\alpha_\epsilon \in (0, 1)\), \(c_k \sim k^{-(\alpha_\epsilon+1)/2}L_0(k)\) as \(k \to \infty\), where \(L_0(\cdot)\) is slowly varying at infinity.

Let

\[
(2.1) \quad \epsilon_{n,r} = \sum_{i=1}^{n} \sum_{1 \leq j_1 < \cdots < j_r} \prod_{s=1}^{r} c_{j_s} \eta_{i-j_s}.
\]

In particular, \(\epsilon_{n,1} = \sum_{i=1}^{n} \epsilon_i\) and if \(r \alpha_\epsilon < 1\),

\[
(2.2) \quad \sigma_{n,r}^2 := \text{Var}(\epsilon_{n,r}) \sim n^{2-r\alpha_\epsilon}L_0^2(n).
\]

From [7] we know that for \(r < \alpha_\epsilon^{-1}\), as \(n \to \infty\),

\[
(2.3) \quad \sigma_{n,r}^{-1} \epsilon_{n,r} \xrightarrow{d} Z_r,
\]

where \(Z_r\) is a random variable which can be represented by appropriate multiple Wiener-Itô integrals. In particular, \(Z_1\) is standard normal. Moreover, the random variables \(Z_1, \ldots, Z_p\) are independent, see e.g. [8, Eq. (1.22)].

Furthermore, let

\[
S_{n,p}(x) = \sum_{i=1}^{n} \left(1_{\{\epsilon_i \leq x\}} - F_\epsilon(x)\right) + \sum_{r=1}^{p} (-1)^{r-1} F_\epsilon^{(r)}(x) \epsilon_{n,r}.
\]

Assume that \(F_\eta(\cdot)\) is 5 times differentiable with bounded, continuous and integrable derivatives. We note in passing that these properties are transferable to \(F_\epsilon(\cdot)\). Following [14, Theorem 3] and [7, Theorem 2.2] we conclude, in particular, that for \(\alpha_\epsilon < 1/2\),

\[
(2.4) \quad \sigma_{n,2}^{-1} S_{n,1}(x) \Rightarrow f_\epsilon^{(1)}(x) Z_2,
\]

where \(Z_2\) is the same random variable as in (2.3). Otherwise, if \(\alpha_\epsilon > 1/2\), then

\[
(2.5) \quad n^{-1/2} S_{n,1}(x) \Rightarrow W_1(x),
\]
where \( \{W_1(x), x \in \mathbb{R}\} \) is a Gaussian process. Furthermore, for \( \alpha > 1/3 \)
\[
\sigma_n^{-1} \sup_{x \in \mathbb{R}} |S_{n,2}(x)| \xrightarrow{a.s.} 0.
\]

The structure of these Gaussian processes is rather complicated and their covariance is given in rather complicated form; see [14] for more details.

3. Results. Let
\[
\Delta_i := \varepsilon_i - \hat{\varepsilon}_i = \varepsilon_i - (Y_i - \hat{m}(X_i)) = \hat{m}(X_i) - m(X_i), \quad \Delta = (\Delta_1, \ldots, \Delta_n).
\]

3.1. Empirical process of residuals: \( \alpha < 1/2 \). The following result provides an uniform expansion of the process \( \hat{K}_n(\cdot) \) and forms a basis for further analysis.

**Theorem 3.1.** Assume (E) with \( \alpha < 1/2 \). Assume that \( F_\eta(\cdot) \) is 3 times differentiable with bounded, continuous and integrable derivatives. Suppose that \( \Delta \) can be written as \( \Delta = \Delta_0 + (\Delta_0, \ldots, \Delta_0) \), where
\[
\begin{align*}
\delta_n + \frac{n^2 \delta_n}{\sigma_{n,2}^4} + \frac{n^2 \delta_n^2}{\sigma_{n,2}^3} + \frac{n \delta_n}{\sigma_{n,2}^3} & \to 0; \\
|\Delta_0| & = o_P(\delta_n), \text{ uniformly in } i.
\end{align*}
\]

Then
\[
\sup_{x \in \mathbb{R}} \left| \hat{K}_n(x) - K_n(x) - f_\varepsilon(x) \sum_{i=1}^n \Delta_i - \frac{1}{2} f_\varepsilon^{(1)}(x) \sum_{i=1}^n \Delta_i^2 + f_\varepsilon^{(1)}(x) \Delta_0 \varepsilon_n,1 \right| = O_P(\delta_n^{1-\nu} \sigma_{n,2}) + o_P(\delta_n \sigma_{n,1}) + O_P \left( \sum_{i=1}^n \Delta_i^2 \right).
\]

In principle, this result is very similar to [4, Theorem 2.1]. However, first, our result is established in the general context of either nonparametric or parametric regression, second, we provide \( o_P(\cdot) \) rates of the approximation. This is crucial to establish limit theorems for the process \( \hat{K}_n(\cdot) \).
To have some intuition, let us write
\[
\hat{K}_n(x) - K_n(x) = \sum_{i=1}^{n} \mathbf{1}_{\{\varepsilon_i \leq x + \Delta_i\}} - \sum_{i=1}^{n} \mathbf{1}_{\{\varepsilon_i \leq x\}}
\]
\[
= \sum_{i=1}^{n} \left( \mathbf{1}_{\{\varepsilon_i \leq x + \Delta_i\}} - F_{\varepsilon}(x + \Delta_i) \right) - \sum_{i=1}^{n} \left( \mathbf{1}_{\{\varepsilon_i \leq x\}} - F_{\varepsilon}(x) \right)
\]
(3.2)
\[
+ \sum_{i=1}^{n} \left( F_{\varepsilon}(x + \Delta_i) - F_{\varepsilon}(x) \right).
\]

From Theorem 3.1 and (2.6) we conclude for \( \alpha_{\varepsilon} < 1/2 \) that, uniformly in \( x \),
\[
\hat{K}_n(x) = K_n(x) + f_{\varepsilon}(x) \sum_{i=1}^{n} \Delta_i + \frac{1}{2} f_{\varepsilon}'(x) \sum_{i=1}^{n} \Delta_i^2 - f_{\varepsilon}^{(1)}(x) \Delta_0 \varepsilon_{n,1}
\]
\[
+ o_P(\sigma_{n,2} + \delta_n \sigma_{n,1}) + O_P \left( \sum_{i=1}^{n} \Delta_i^3 \right).
\]
(3.3)
\[
= -f_{\varepsilon}(x) \sum_{i=1}^{n} \varepsilon_i + f_{\varepsilon}(x) \sum_{i=1}^{n} \Delta_i + f_{\varepsilon}^{(1)}(x) \varepsilon_{n,2} + \frac{1}{2} f_{\varepsilon}'(x) \sum_{i=1}^{n} \Delta_i^2 - f_{\varepsilon}^{(1)}(x) \Delta_0 \varepsilon_{n,1}
\]
\[
+ o_P(\sigma_{n,2} + \delta_n \sigma_{n,1}) + O_P \left( \sum_{i=1}^{n} \Delta_i^3 \right).
\]

We note in passing that in order to obtain the above expansion via (2.6) one has to assume that \( F_{\eta}(\cdot) \) is 5 times differentiable.

As we will see below (Sections 3.3 and 3.4), it may happen that the first order contribution
\[
-f_{\varepsilon}(x) \sum_{i=1}^{n} \varepsilon_i + f_{\varepsilon}(x) \sum_{i=1}^{n} \Delta_i
\]
is negligible. In other words, the rates of convergence of \( \hat{K}_n(\cdot) \) will be different from those for \( K_n(\cdot) \). The rates of convergence will be determined by the second order term
\[
f_{\varepsilon}^{(1)}(x) \varepsilon_{n,2} + \frac{1}{2} f_{\varepsilon}'(x) \sum_{i=1}^{n} \Delta_i^2 - f_{\varepsilon}^{(1)}(x) \Delta_0 \varepsilon_{n,1}.
\]

3.2. Empirical process of residuals: \( \alpha_{\varepsilon} > 1/2 \). Let \( \xi_i = \varepsilon_i - \eta_i \). Define \( \xi_{n,r} \) in the analogous way as \( \varepsilon_{n,r} \); see (2.1).
Theorem 3.2. Assume (E) with \( \alpha_\varepsilon > 1/2 \). Assume that \( F_\eta(\cdot) \) is 3 times differentiable with bounded, continuous and integrable derivatives. Suppose that \( \Delta \) can be written as \( \Delta_0 \mathbf{1} + (\Delta_{01}, \ldots, \Delta_{0n}) \), where

\[
\delta_n + \sqrt{n} \delta_n^2 \rightarrow 0;
\]

- \( \frac{1}{\delta_n} \Delta_0 \overset{d}{\rightarrow} V \), where \( V \) is a nondegenerate random variable;
- \( |\Delta_{0i}| = o_P(\delta_n) \), uniformly in \( i \).

Then

\[
\hat{K}_n(x) = K_n(x) + f_\varepsilon(x) \sum_{i=1}^n \varepsilon_i + \left( f_\varepsilon(x) \sum_{i=1}^n \Delta_i - f_\varepsilon(x) \sum_{i=1}^n \varepsilon_i - f_\varepsilon^{(1)}(x) \Delta_0 \xi_{n,1} \right)
\]

\[+ O\left( \sum_{i=1}^n \Delta_i^2 \right) + o_P(\sqrt{n}) + O_P(\delta_n \sigma_{n,1})\]

where \( n^{-1/2} (K_n(x) + f_\varepsilon(x) \sum_{i=1}^n \varepsilon_i) \) converges weakly to \( W_1(x) \) from (2.5).

3.3. Parametric regression. The results of Theorems 3.1 and 3.2 are the tools to establish a limit theorem for \( \hat{K}_n(\cdot) \) in case of parametric model

\[
m(x) = \beta_0 + \beta_1 x.
\]

We assume that the regression parameters are estimated using standard least squares. We make the following assumptions on the predictors \( X_i, i \geq 1 \):

(P1) \( X_i, i \geq 1 \), is a sequence of i.i.d. random variables with \( \mathbb{E}X_1 = \mu < \infty \) and \( \text{Var}(X_1) < \infty \).

(P2) \( X_i, i \geq 1 \), is a random sequence such that \( \sup_i \mathbb{E}|X_i| + |\bar{X}| < \infty \).

Corollary 3.3. Assume (P2) and (E) and that

\[
\hat{\beta}_1 - \beta_1 = o_P(\sigma_{n,1}/n).
\]

Assume that \( F_\eta(\cdot) \) is 5 times differentiable with bounded, continuous and integrable derivatives.

(a) If \( \alpha_\varepsilon < 1/2 \), then

\[
\frac{1}{\sigma_{n,2}} \hat{K}_n(x) \Rightarrow f_\varepsilon^{(1)}(x)(Z_2 - \frac{1}{2} Z_1^2),
\]

where \( Z_1, Z_2 \) are defined in (2.3).

(b) If \( \alpha_\varepsilon > 1/2 \), then \( n^{-1/2} \hat{K}_n(x) \Rightarrow W_1(x) \).
Remark 3.4. Note that the rate of convergence \( \sigma_{n,1} \) for the original process \( K_n(\cdot) \) changes to \( \sigma_{n,2} \) or \( \sqrt{n} \) for \( \hat{K}_n(\cdot) \). The similar phenomena was observed in a context of empirical processes with estimated parameters in [10] (see also [2]). Note further that a possible LRD of predictors does not play any role.

Furthermore, from the proof of Corollary 3.3 below, we may conclude that in case \( \beta_0 = 0 \) the limiting behaviour of \( K_n(x) \) and \( \hat{K}_n(x) \) is the same. In other words, for the model (3.5) with \( \beta_0 = 0 \), we have (see also [4])

\[
\sigma_{n,1}^{-1} \hat{K}_n(x) \Rightarrow f_\epsilon(x)Z_1.
\]

Remark 3.5. The condition (3.6) can be verified for many stationary sequences. In particular, if \( X_i, i \geq 1, \) is LRD linear sequence with parameter \( \alpha_X \), then the rate of convergence of \( \hat{\beta}_1 - \beta_1 \) is either \( \sqrt{n} \) or \( n(\alpha_X + \alpha_\epsilon)/2 \), for \( \alpha_X + \alpha_\epsilon > 1 \) or max(\( \alpha_X, \alpha_\epsilon \)) < 1/2, respectively; see [13] and [6].

Proof of Corollary 3.3. Least squares estimation leads to the following expressions:

\[
\hat{\beta}_1 - \beta_1 = \frac{1}{s_n} \left( \frac{1}{n} \sum_{j=1}^{n} X_j \varepsilon_j - \bar{X} \bar{\varepsilon} \right), \quad \hat{\beta}_0 - \beta_0 = \bar{\varepsilon} - \bar{X}(\hat{\beta}_1 - \beta_1),
\]

where \( \bar{X} \) and \( \bar{\varepsilon} \) are sample means based on \( X_1, \ldots, X_n \) and \( \varepsilon_1, \ldots, \varepsilon_n \), respectively, and \( s_n = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X})^2 \). We have

\[
\Delta_i = \hat{m}(X_i) - m(X_i) = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)X_i = \bar{\varepsilon} + (\hat{\beta}_1 - \beta_1)(X_i - \bar{X}).
\]

From (2.2) we conclude that

\[
\bar{\varepsilon} = O_P(\sigma_{n,1}/n), \quad \sigma_{n,1}^2/n \sim \sigma_{n,2}, \quad \text{as } n \to \infty.
\]

From (3.6) and Assumption (P2) we conclude \( \Delta_i = \bar{\varepsilon} + o_P(\sigma_{n,1}/n)O_P(1) \). Let now \( \delta_n = \sigma_{n,1}/n \). It is straightforward to check that such \( \delta_n \) fulfills (3.1). Therefore, the conditions of Theorem 3.1 are fulfilled with \( \Delta_0 = \bar{\varepsilon} \) and \( V = Z_1 \).

Furthermore, from (3.8), \( \sum_{i=1}^{n} \Delta_i = n\bar{\varepsilon} = \varepsilon_{n,1} = \sum_{i=1}^{n} \varepsilon_i \) and via (3.9),

\[
\sum_{i=1}^{n} \Delta_i^2 = n\bar{\varepsilon}^2 + n\bar{\varepsilon} o_P(\sigma_{n,1}/n) + o_P(n\sigma_{n,1}^2/n^2) = n\bar{\varepsilon}^2 + o_P(\sigma_{n,2}).
\]
Consequently, noting that $\delta_n \sigma_{n,1} \sim \sigma_{n,2}$ and $n\bar{\varepsilon}^2 = \bar{\varepsilon}_n \varepsilon_{n,1}$, the expansion (3.3) reads

$$\hat{K}_n(x) = f^{(1)}(x)\varepsilon_{n,2} - \frac{1}{2}f^{(1)}(x)n\bar{\varepsilon}^2 + o_P(\sigma_{n,2}).$$

The result of part (a) follows now from (2.3).

As for part (b), we recall that $\sum_{i=1}^{n} \Delta_i - \sum_{i=1}^{n} \varepsilon_i = 0$. Also, since $\alpha_\varepsilon > 1/2$, $\Delta_0 \xi_{n,1} = O_P(\sigma_{n,1}^2/n) = O_P(\sigma_{n,2}) = o_P(\sqrt{n})$ and via (3.10), $\sum_{i=1}^{n} \Delta_i^2 = O_P(\sigma_{n,2}) = o_P(\sqrt{n})$. Finally, the choice of $\delta_n$ yields $\delta_n \sigma_{n,1} = o_P(\sqrt{n})$. Therefore, part (b) follows from Theorem 3.2. □

3.4. Nonparametric regression. Now, we establish the result for nonparametric regression case. It is assumed that $m(\cdot)$ is estimated by the usual Nadaraya-Watson estimator, i.e.

$$\hat{m}(x) = \hat{m}_b(x) = \frac{1}{nb} \sum_{j=1}^{n} Y_j K_b(x - X_j),$$

with

$$\hat{f}_b(x) = \frac{1}{nb} \sum_{j=1}^{n} K_b(x - X_j),$$

where $K_b(\cdot) = K(\cdot/b)$ and $K(\cdot)$ is a positive kernel, which fulfills standard conditions: $\int K(u) du = 1$, $\int uK(u) du = 0$ and $\int u^2K(u) du < \infty$.

Here we shall assume for simplicity that the predictors are i.i.d. It can be extended to e.g. LRD stationary predictors using estimates from [12].

**Corollary 3.6.** Assume (P1) and (E). Assume that $F_\eta(\cdot)$ is 5 times differentiable with bounded, continuous and integrable derivatives. Also, suppose that the bandwidth fulfills

$$b + (nb)^{-1} \rightarrow 0,$$

and

$$b\sigma_{n,1}^2/n \rightarrow \infty.$$

(a) If $\alpha_\varepsilon < 1/2$, and

$$b^2 n/\sigma_{n,1} + nb^4/\sigma_{n,2} + b^2 \sigma_{n,1}/\sigma_{n,2} + \frac{\sigma_{n,1}^2}{\sigma_{n,2}^2 nb} \rightarrow 0.$$
then
\[
\frac{1}{\sigma_{n,2}} \hat{K}_n(x) \Rightarrow f_{\varepsilon}^{(1)}(x)(Z_2 - \frac{1}{2} Z_1^2),
\]
where \( Z_1, Z_2 \) are defined in (2.3).

(b) If \( \alpha_{\varepsilon} > 1/2 \), and
\[
(3.17) \quad \frac{b^2 n}{\sigma_{n,1}} + nb^4/\sqrt{n} + b^2 \sigma_{n,1}/\sqrt{n} + \sigma^2_{n,1}/(n^2 b) \to 0.
\]
then \( n^{-1/2} \hat{K}_n(x) \Rightarrow W_1(x) \).

Remark 3.7. The condition (3.14) is standard in nonparametric estimation. With the standard bandwidth choice \( b = C n^{-1/5} \) (see e.g. [11]) condition (3.15) is valid for \( \alpha_{\varepsilon} < 4/5 \). Likewise, one can easily verify that (3.16) holds for \( \alpha_{\varepsilon} < 4/5 \) as well and so for all \( \alpha_{\varepsilon} < 1/2 \). Finally, (3.17) holds for \( 1/5 < \alpha_{\varepsilon} < 4/5 \) and so for all \( 1/2 < \alpha_{\varepsilon} < 4/5 \).

Proof of Corollary 3.6. In the nonparametric regression model we have
\[
(3.18) \quad \Delta_i = \hat{m}_b(X_i) - m(X_i) = R_b(X_i) + \frac{1}{nb \hat{f}_b(X_i)} \sum_{j=1}^n K_b(X_i - X_j) \varepsilon_j,
\]
where
\[
(3.19) \quad R_b(y) = \frac{1}{nb \hat{f}_b(y)} \sum_{j=1}^n (m(X_j) - m(y)) K_b(y - X_j).
\]
Denote \( \rho(y) = (mf)'(y) - m(y) f''(y) \). Uniformly over \( \{y : f(y) > 0\} \),
\[
(3.20) \quad R_b(y) - \frac{b^2 \kappa_2 \rho(y)}{2} = O(b^4(1 + o_P(1)))).
\]
Now, in the second part of (3.18), we may replace \( \hat{f}_b(X_i) \) with \( f(X_i) \). This is allowed since, first, \( \hat{f}_b(\cdot) \) is the consistent estimator of \( f(\cdot) \); second, since \( K(\cdot) \) has bounded support \( \mathcal{I} \) and \( f(x) > 0, x \in \mathcal{I} \). Define for \( j \neq i \),
\[
L_b(X_i, X_j) = \frac{1}{bf(X_i)} K_b(X_i - X_j).
\]
We may write (recall that \( L_b^*(X_i, X_j) \) is the centered version of \( L_b(X_i, X_j) \))
\[
\Delta_i = R_b(X_i) + E[L_b(X_1, X_2)] \varepsilon + \frac{1}{n} \sum_{j=1}^n L_b^*(X_i, X_j) \varepsilon_j.
\]
Using (3.20) and (5.21) below we argue that
\[ \Delta_i = O_P(b^2) + \bar{\varepsilon} + o_P(\sigma_{n,1}/n), \]
uniformly in \( i \), provided that (3.15) holds. Therefore, the conditions of Theorem 3.1 are fulfilled with \( \Delta_0 = \bar{\varepsilon}, \delta_n = \sigma_{n,1}/n \) and \( V = Z_1 \), as long as (3.15) and the first part of (3.16) hold.

From (3.21),
\[ \sum_{i=1}^{n} \Delta_i^2 = n\bar{\varepsilon}^2 + O_P(b^2\sigma_{n,1}) + O_P(nb^4) + o_P(\sigma_{n,1}/n) = n\bar{\varepsilon}^2 + o_P(\sigma_{n,2}), \]
if \( \alpha < 1/2 \) and (3.16) holds. Likewise, if (3.17) holds and \( \alpha > 1/2 \),
\[ \sum_{i=1}^{n} \Delta_i^2 = n\bar{\varepsilon}^2 + o_P(\sqrt{n}) = O_P(\sigma_{n,1}^2/n) + o_P(\sqrt{n}) = o_P(\sqrt{n}). \]

Also, from Section 5.5 we obtain
\[ \sum_{i=1}^{n} \Delta_i = \sum_{i=1}^{n} \varepsilon_i + o_P(\sigma_{n,2} \sqrt{n}). \]
This finishes the proof.

3.5. **Residual empirical process with estimated parameters.** Let us focus on the parametric regression model of Section 3.3. Application of Corollary 3.3 yields that
\[ \frac{1}{\sigma_{n,2}} \sup_{x \in \mathbb{R}} |\hat{K}_n(x)| \overset{d}{\to} \sup_{x \in \mathbb{R}} |f_\varepsilon^{(1)}(x)|(Z_2 - \frac{1}{2}Z_1^2), \]
and
\[ n^{-1/2} \sup_{x \in \mathbb{R}} |\hat{K}_n(x)| \overset{d}{\to} \sup_{x \in \mathbb{R}} |W_1(x)|, \]
respectively for \( \alpha \varepsilon < 1/2 \) and \( \alpha \varepsilon > 1/2 \). The above results can be used, in principle, to test whether the errors \( \varepsilon_1, \ldots, \varepsilon_n \) are consistent with a given distribution \( F_\varepsilon \). If however \( F_\varepsilon \) belongs to, say, a one-parameter family \( \{F_\varepsilon(\cdot, \theta), \theta \in \mathbb{R}\} \), then one needs to know the value of the parameter \( \theta \). A straightforward procedure would be to estimate it and use the statistic
\[ \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^{n} 1_{\{\varepsilon_i < x\}} - F_\varepsilon(x; \hat{\theta}_n) \right|, \]
where \( F_\varepsilon(x; \hat{\theta}_n) \) is the distribution function \( F_\varepsilon(x) = F_\varepsilon(x; \theta) \) in which the parameter \( \theta \) has been replaced with its estimator \( \hat{\theta}_n \).

Therefore, this section is devoted to study the limiting behaviour of

\[
\hat{L}_n(x) := \sum_{i=1}^{n} \left( 1_{\{\hat{\varepsilon}_i \leq x \}} - F_\varepsilon(x; \hat{\theta}_n) \right).
\]

The results below may be seen as counterpart to the asymptotic results for

\[
L_n(x) := \sum_{i=1}^{n} \left( 1_{\{\varepsilon_i \leq x \}} - F_\varepsilon(x; \hat{\theta}_n) \right),
\]

see [10] for results and references therein for more discussion on this approach.

Many estimators \( \hat{\theta}_n \) of \( \theta \) can be obtained with help of partial sums \( \sum_{i=1}^{n} H(\hat{\varepsilon}_i) \), where \( H \) is a function that does not depend on \( n \). Let us note that from Theorem 3.1 we may have two scenarios for \( \alpha_\varepsilon < 1/2 \):

(A) \( \sigma_{n,2}^{-1} (\sum_{i=1}^{n} H(\hat{\varepsilon}_i) - \text{E}[H(\varepsilon_i)]) \) converges in distribution to a nondegenerate random variable;

(B) \( \sigma_{n,2}^{-1} (\sum_{i=1}^{n} H(\hat{\varepsilon}_i) - \text{E}[H(\varepsilon_i)]) = o_P(1) \).

**Example 3.8.** Consider \( H(u) = u^2 \) which yields the estimator of \( \text{Var}(\varepsilon) \).

We obtain for \( \alpha_\varepsilon < 1/2 \):

\[
\sigma_{n,2}^{-1} \left( \sum_{i=1}^{n} H(\hat{\varepsilon}_i) - \text{E}[H(\varepsilon_i)] \right) \overset{d}{\rightarrow} \int f_\varepsilon^{(1)}(v) dH(v) \left( Z_1^2 - \frac{1}{2} Z_2^2 \right) = 2 \left( Z_1^2 - \frac{1}{2} Z_2^2 \right).
\]

Consider now \( H(u) = u^3 \). We have for \( \alpha_\varepsilon < 1/2 \):

\[
\sigma_{n,2}^{-1} \left( \sum_{i=1}^{n} H(\hat{\varepsilon}_i) - \text{E}[H(\varepsilon_i)] \right) \overset{d}{\rightarrow} 6 \int v f_\varepsilon(v) dv \left( Z_1^2 - \frac{1}{2} Z_2^2 \right).
\]

Consequently, if \( f_\varepsilon \) is symmetric, then the right hand side is simply 0 and thus we are in scenario (B).

In what follows, we will write \( f_\varepsilon(\cdot; \theta) \) to indicate the density with the true parameter \( \theta \).

**Corollary 3.9.** Assume that \( \hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} H(\hat{\varepsilon}_i) \) and \( \theta = \text{E}[H(\varepsilon)] \).

Under the conditions of Corollary 3.3, we have

\[
\frac{1}{\sigma_{n,2}} \tilde{L}_n(x) \Rightarrow \left( f_\varepsilon^{(1)}(x; \theta) + f_\varepsilon(x; \theta) \int f_\varepsilon^{(1)}(u; \theta) dH(u) \right) \left( Z_2 - \frac{1}{2} Z_1^2 \right),
\]
and
\[ n^{-1/2} \hat{L}_n(x) \Rightarrow W_1(x) + f_\varepsilon(x; \theta) \int W_1(u) \, dH(u), \]
respectively for \( \alpha_\varepsilon < 1/2 \) and \( \alpha_\varepsilon > 1/2 \), provided that the integrals at the right hand sides are finite.

**Remark 3.10.** We note that rates of convergence for \( \hat{L}_n(\cdot) \), residual empirical process with estimated parameters, are the same as for \( \hat{K}_n(\cdot) \), the ordinary residual empirical process. This is different as compared to \( K_n(\cdot) \) and its ”estimated” version; see [10].

**Proof of Corollary 3.9.** We conduct the proof for \( \alpha_\varepsilon < 1/2 \). For a function \( g(x; \theta) \) denote by \( \nabla^r \theta g(x; \theta) \) its \( r \)th order derivative with respect to \( \theta \), evaluated at \( \theta = \theta \). In particular, \( \nabla = \nabla^1 \). Then
\[
\hat{L}_n(x) = \hat{K}_n(x) + n(\theta - \hat{\theta}_n) \nabla_{\theta} F_\varepsilon(x; \theta) + \frac{1}{2} n(\theta - \hat{\theta}_n)^2 \nabla_{\theta}^2 F_\varepsilon(x; \theta^*_n)
\]
with some \( \hat{\theta}_n^* \) such that \( |\theta^*_n - \hat{\theta}_n| \leq |\theta - \hat{\theta}_n^*| \). Therefore
\[
\hat{L}_n(x) = \hat{K}_n(x) + f_\varepsilon(x; \theta) \left( \sum_{i=1}^{n} (E[H(\varepsilon)] - H(\hat{\varepsilon}_i)) \right) + o_P(\sigma_{n,2})
\]
\[
= \hat{K}_n(x) - f_\varepsilon(x; \theta) \left( \int H(u) \, d\hat{K}_n(u) \right) + o_P(\sigma_{n,2})
\]
\[
= \hat{K}_n(x) + f_\varepsilon(x; \theta) \left( \int \hat{K}_n(u) \, dH(u) \right) + o_P(\sigma_{n,2})
\]
and the result follows from Corollary 3.3.

**4. Simulation studies.** We conducted simulations justifying our results on asymptotic behaviour of supremum of the empirical process of residuals \( \hat{K}_n(\cdot) \). First, we simulated \( n = 100 \) i.i.d. random variables \( \varepsilon_i, i = 1, \ldots, n \) from \( N(0,1) \) distribution. Then, supremum \( \sup_{x \in \mathbb{R}} K_n(x) \) was calculated. This procedure was repeated 100 times. Quartiles and standard deviation of the empirical distribution of the supremum was calculated. Next, for the same errors, model \( Y_i = 1 + 4X_i + \varepsilon_i \) was considered, and residuals were calculated using estimators of \( \beta_0, \beta_1 \) given in (3.7). Also, for the same errors, we assumed that \( \beta_0 = 1 \) is known. The same procedure was repeated with errors following LRD Gaussian process with \( \alpha_\varepsilon \in \{0.2, 0.4, 0.6, 0.8\} \). The results are given in Table 1.

- Column 3: For the empirical process \( K_n \) based on errors, the variability of the supremum increases with the dependence, which is in agreement with the asymptotic theory for the LRD-based empirical processes.
• Column 4: We consider the empirical process $L_n$, where $F_\varepsilon(\cdot; \hat{\theta}_n)$, $\hat{\theta}_n$ being sample standard deviation based on errors $\varepsilon_1, \ldots, \varepsilon_n$. The results are similar to column 3. In other words, estimation of variance does not influence asymptotic behaviour of the empirical process. This agrees with theoretical results; see [10, Remark 1.6]. This happens since variance can be estimated with rate $\sigma_n^2 \sqrt{n}$, whereas the rate of convergence for $K_n(\cdot)$ is $\sigma_{n,1}$.

• Column 5: We consider the residual-based empirical process $\hat{K}_n$ in the linear regression model. Both slope and intercept are estimated. We note that the variability for $\alpha_\varepsilon = 0.8$ or $\alpha_\varepsilon = 0.6$ is almost the same as for i.i.d. case. In other words, LRD does not play any role, which is in agreement with Corollary 3.3.

• Column 6: Results for the residual-based empirical process $\hat{L}_n$ with estimated variance are similar as for $\hat{K}_n$. Recall that Corollary 3.9 indicates that rates of convergence for $\hat{L}_n$ is the same as for $\hat{K}_n$.

• Column 7: We consider $\hat{K}_n$, but the intercept is assumed to be known. Results are similar to Column 3. In other words, in case of known intercept the asymptotic behaviour of $\hat{K}_n$ is similar to $K_n$; see Remark 3.4.

|                | $K_n$ | $L_n$ | $\hat{K}_n$ | $\hat{L}_n$ | $K_n; \beta_0 = 1$ | $L_n; \beta_0 = 1$
|----------------|-------|-------|-------------|-------------|---------------------|---------------------
| i.i.d.         | $Q_1$ | 0.0416| 0.0392      | 0.0467      | 0.0413              | 0.0419              | 0.0376              |
|                | $Q_3$ | 0.0880| 0.0859      | 0.0656      | 0.0592              | 0.0873              | 0.0789              |
|                | $s$   | 0.0314| 0.0315      | 0.0169      | 0.0146              | 0.0312              | 0.0313              |
| $\alpha_\varepsilon = 0.8$ | $Q_1$ | 0.0307| 0.0274      | 0.0473      | 0.0448              | 0.0346              | 0.0278              |
|                | $Q_3$ | 0.0994| 0.0940      | 0.0686      | 0.0637              | 0.0963              | 0.0908              |
|                | $s$   | 0.0484| 0.0494      | 0.0149      | 0.013               | 0.0494              | 0.0504              |
| $\alpha_\varepsilon = 0.6$ | $Q_1$ | 0.0303| 0.0150      | 0.0488      | 0.0447              | 0.0274              | 0.0147              |
|                | $Q_3$ | 0.1285| 0.1237      | 0.0718      | 0.0646              | 0.1303              | 0.1192              |
|                | $s$   | 0.0758| 0.0786      | 0.0151      | 0.0139              | 0.0772              | 0.0797              |
| $\alpha_\varepsilon = 0.4$ | $Q_1$ | 0.0062| 0.0038      | 0.0504      | 0.0471              | 0.0072              | 0.0041              |
|                | $Q_3$ | 0.1471| 0.1353      | 0.0784      | 0.0662              | 0.1479              | 0.1353              |
|                | $s$   | 0.0858| 0.0852      | 0.0194      | 0.0147              | 0.0850              | 0.0845              |
| $\alpha_\varepsilon = 0.2$ | $Q_1$ | 0.0015| 0.0023      | 0.0535      | 0.0418              | 0.0021              | 0.0017              |
|                | $Q_3$ | 0.2870| 0.2714      | 0.0826      | 0.0645              | 0.2978              | 0.2770              |
|                | $s$   | 0.1911| 0.1851      | 0.0218      | 0.0178              | 0.1906              | 0.1850              |

Table 1
Simulated values of different dispersion measures.
5. Technical details. Let $H_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$. Let $u = (u_1, \ldots, u_n)$ be a vector of scalars. Define

$$Z_n(x; u) = \sum_{i=1}^{n} \left( 1\{\xi_i \leq x + u_i\} - F_{\xi}(x + u_i) \right) - \sum_{i=1}^{n} \left( 1\{\xi_i \leq x\} - F_{\xi}(x) \right).$$

The process $Z_n(x; u)$ is written as

$$(5.1) \quad Z_n(x; u) = \sum_{i=1}^{n} \left( 1\{x < \xi_i \leq x + u_i\} - E\left[ 1\{x < \xi_i \leq x + u_i\} | H_{i-1} \right] \right)$$

$$+ \sum_{i=1}^{n} \left( E\left[ 1\{x < \xi_i \leq x + u_i\} | H_{i-1} \right] - E\left[ 1\{x < \xi_i \leq x + u_i\} \right] \right) =: M_n(x; u) + N_n(x; u).$$

Recall now that $\Delta = (\Delta_1, \ldots, \Delta_n)$. Recalling (3.2), we decompose

$$(5.2) \quad \hat{K}_n(x) - K_n(x) =$$

$$= M_n(x; \Delta) + N_n(x; \Delta) + f_{\xi}(x) \sum_{i=1}^{n} \Delta_i + \frac{1}{2} f_{\xi}^{(1)}(x) \sum_{i=1}^{n} \Delta_i^2 + O\left( \sum_{i=1}^{n} \Delta_i^3 \right).$$

First, in Corollary 5.2 we will establish an asymptotic expansion for the LRD part $N_n(x; \Delta)$. This will be done by considering a special structure of $N_n(x; u)$ (see Lemma 5.1 and (5.4) below) and then ”replacing” $u$ with $\Delta$ under proper assumptions for the latter.

Furthermore, we have to bound $M_n(x; \Delta)$. This will be done by obtaining an uniform bound on $M_n(x; u)$. In this way, we may utilize the martingale structure of the latter. Clearly, $M_n(x; \Delta)$ is not a martingale. The bounds are given in Lemma 5.3 and Lemma 5.5.

5.1. LRD part. Denote $u_0 = u_0 \mathbf{1}$, where $\mathbf{1}$ is the vector of dimension $n$, consisting of '1'. Recall that $\xi_i = \xi_i - \eta_i$ and $\xi_{n,r}$ is defined in the analogous way as $\varepsilon_{n,r}$. In the first lemma we deal with $N_n(x; u_0)$. The proof is included in Section 5.1.1.

**Lemma 5.1.** Assume that $F_{\eta}(\cdot)$ is 3 times differentiable with bounded, continuous and integrable derivatives. Then with some $0 < \nu < 1/2$ and $\delta_n \to 0$,

$$(5.3) \quad \sup_{|u_0| \leq \delta_n^{1-\nu}} \sup_{x \in \mathbb{R}} \left| N_n(x; u_0) + f_{\xi}^{(1)}(x) u_0 \xi_{n,1} \right| = O_P \left( \delta_n^{1-\nu} (\sigma_{n,2} \vee \sqrt{n}) + \delta_n^{2(1-\nu)} \sigma_{n,1} \right).$$
Note now that the part $N_n(x, u)$ in (5.1) can be written as

$$N_n(x; u) = \sum_{i=1}^{n} (F_\eta(x + u_i - \xi_i) - F_\eta(x - \xi_i) - EF_\eta(x + u_i - \xi_i) + EF_\eta(x - \xi_i)).$$

Let us choose $u = u_0 + (u_0, \ldots, u_{0n})$. If $\max_i(|u_0|) = o(\delta_n)$, then applying first order Taylor expansion, and noting that $\xi_i, i \geq 1$, is LRD moving average with the same properties as $\varepsilon_i, i \geq 1$,

$$N_n(x; u) - N_n(x; u_0) = o(\delta_n) \sum_{i=1}^{n} (f_\eta(x + u_0 - \xi_i) - Ef_\eta(x + u_0 - \xi_i)) = o_P(\delta_n \sigma_n1),$$

uniformly in $u, u_0$ and $x$, since $f^{(1)}_\eta$ is bounded and integrable. Combining this with (5.3), we have (recall $\nu < 1/2$)

$$\sup_{u} \sup_{x} |N_n(x; u) + f^{(1)}_\eta(x)u_0 \xi_n,1| = O_P(\delta_n^{1-\nu} (\sigma_{n,2} \vee \sqrt{n})) + o_P(\delta_n \sigma_n1),$$

where $\sup_u$ is taken over all $u$ such that

$$u = u_0 + (u_0, \ldots, u_{0n}), \quad \max_i(|u_0|) = o(\delta_n), \quad |u_0| = O(\delta_n^{1-\nu}).$$

In this way we end up with the following corollary.

**Corollary 5.2.** Assume that $F_\eta(\cdot)$ is 3 times differentiable with bounded, continuous and integrable derivatives. Assume that $\Delta$ can be written as $\Delta_01 + (\Delta_{01}, \ldots, \Delta_{0n})$, where

$$\Delta_0 = o_P(\delta_n^{1-\nu}), \quad \max_i \Delta_{0i} = o_P(\delta_n).$$

Then

$$\sup_{x \in \mathbb{R}} |N_n(x; \Delta) + f^{(1)}_\xi(x)\Delta_0 \xi_n,1| = O_P(\delta_n^{1-\nu} (\sigma_{n,2} \vee \sqrt{n})) + o_P(\delta_n \sigma_n1).$$

Noting that for $\alpha_\varepsilon < 1/2$ we have $\xi_n,1 - \varepsilon_n,1 = o_P(\sigma_{n,2})$, we may replace $\xi_n,1$ with $\varepsilon_n,1$ in the statement of Theorem 3.1.

5.1.1. **Proof of Lemma 5.1.** Let $F_{n,\xi}(\cdot)$ be an empirical distribution function, associated with $\xi_1, \ldots, \xi_n$ and let $F_{\xi}(\cdot)$ be, respectively, distribution and density function of any of $\xi_i$. Note that $\xi_i$ and $\eta_i$ are independent for each fixed $i$, and $f_{\xi} * f_\eta = f_\xi$. Recall that $\xi_{n,x}$ is defined in the analogous way as $\varepsilon_{n,y}$; see (2.1). From (2.2) we obtain that $\xi_n,1 = O_P(\sigma_{n,1})$. 

imsart-aos ver. 2007/12/10 file: Residuals-AoS.tex date: June 11, 2010
Furthermore, let

\[ \tilde{S}_{n,p}(x) = \sum_{i=1}^{n} \left( 1_{\{\xi_i \leq x\}} - F_{\xi}(x) \right) + \sum_{r=1}^{p} (-1)^{r-1} F^{(r)}_{\xi}(x) \xi_{n,r}. \]

Note that \( \tilde{S}_{n,p} \) is defined in the same way as \( S_{n,p} \), but we use \( \xi_i \)'s in the former instead of \( \varepsilon_i \)'s in the latter. Nevertheless, we conclude from (2.4) and (2.5) that for \( \alpha_\varepsilon < 1/2 \),

\[ \sigma_{n,2}^{-1} \tilde{S}_{n,1}(x) \Rightarrow f^{(1)}_{\xi}(x) Z_2, \]

where \( Z_2 \) is the same random variable as in (2.3). Otherwise, if \( \alpha_\varepsilon > 1/2 \), then

\[ n^{-1/2} \tilde{S}_{n,1}(x) \Rightarrow \Psi(x), \]

where \( \Psi \) is a Gaussian process and the convergence is in the Skorokhod topology.

We compute

\[ N_n(x; u_0) = n \int (F_\eta(x + u_0 - v) - F_\eta(x - v)) d(F_{n, \xi}(v) - F_\xi(v)) \]

\[ = n \int (F_{n, \xi}(v) - F_\xi(v)) (f_\eta(x + u_0 - v) - f_\eta(x - v)) \, dv \]

\[ = n \int (F_{n, \xi}(v) - F_\xi(v) + f_\xi(v) \xi_{n,1}/n) (f_\eta(x + u_0 - v) - f_\eta(x - v)) \, dv \]

\[ - (f_\varepsilon(x + u_0) - f_\varepsilon(x)) \xi_{n,1} \]

\[ = \int \tilde{S}_{n,1}(v) (f_\eta(x + u_0 - v) - f_\eta(x - v)) \, dv - f^{(1)}_\varepsilon(x) u_0 \xi_{n,1} + O(u_0^2) \xi_{n,1} \]

\[ = \int \tilde{S}_{n,1}(v) f^{(1)}_\eta(x - v) u_0(v) \, dv - f^{(1)}_\varepsilon(x) u_0 \xi_{n,1} + O(u_0^2) \xi_{n,1}, \]

where \( u_0(v) \) lies between \( x - v \) and \( x + u_0 - v \). From (2.4) and (2.5) we conclude that \( \sup_v |\tilde{S}_{n,1}(v)| = O_P(\sigma_{n,2} \vee \sqrt{n}) \). Therefore, with a \( 1 > \nu > 0 \),

\[ \sup_{|u_0| \leq \delta_n^{-\nu}} \sup_x N_n(x; u_0) + f^{(1)}_\varepsilon(x) u_0 \xi_{n,1} = O_P \left( \delta_n^{1-\nu} (\sigma_{n,2} \vee \sqrt{n}) + \delta_n^{2(1-\nu)} \sigma_{n,1} \right). \]

\[ \square \]

5.2. Martingale part. The proofs for martingale part are standard, in particular, they are similar as in [4]. However, some details are different, since the main theorems involve non-standard scalings \( n^{-1/2} \) and \( \sigma_{n,2}^{-1} \), rather than \( \sigma_{n,1}^{-1} \).
Lemma 5.3. Assume that $\|f_\eta\|_\infty < \infty$.

(a) Let $x_r = r \frac{1}{\sigma_n^2}$. If $\alpha_\varepsilon < 1/2$ and (3.1) holds, then

$$\sup_u \max_{r \in \mathbb{Z}} |M_n(x_r; u)| = o_P(\sigma_{n,2}).$$

(b) Let $x_r = r \frac{\sqrt{n}}{\sigma_n}$ with $\varepsilon > 0$. If $\alpha_\varepsilon > 1/2$ and (3.4) holds, then

$$\sup_u \max_{r \in \mathbb{Z}} |M_n(x_r; u)| = o_P(\sqrt{n}).$$

In both cases $\sup_u$ is taken over all $u$ such that

$$u = u_0 + (u_{01}, \ldots, u_{0n}), \quad \max_i |u_{0i}| = o(\delta_n), \quad |u_0| = O(\delta_n^{1-\nu}).$$

Let

$$A_n(x; y) = \sum_{i=1}^n \left(1_{\{\varepsilon_i \leq y\}} - F_\varepsilon(y) - (1_{\{\varepsilon_i \leq x\}} - F_\varepsilon(x))\right).$$

The next lemma establishes tightness-like property of the empirical process based on $\varepsilon_i$, $i \geq 1$. Note, however, that it cannot be concluded directly from the tightness of $\sigma_n^{-1/2}K_n(\cdot)$, since the different scaling is involved.

Lemma 5.4. Assume that $\|f_\eta\|_\infty < \infty$.

- If $\alpha_\varepsilon < 1/2$, then $\sup_{|y-x| \leq \sigma_n^{-1}} |A_n(x; y)| = o_P(\sigma_{n,2})$.
- If $\alpha_\varepsilon > 1/2$, then $\sup_{|y-x| \leq \sigma_n^{-1}} |A_n(x; y)| = O_P(\epsilon n^{-1/2})$.

Combining Lemmas 5.3 and 5.4 we obtain the following uniform behaviour of the martingale part.

Lemma 5.5. Under the conditions of Lemma 5.3 we have

$$\sup_u \sup_{x \in \mathbb{R}} |M_n(x; u)| = o_P(\sigma_{n,2}) + O_P(\epsilon \sqrt{n}).$$

As in case of Corollary 5.2 we conclude the following corollary.

Corollary 5.6. Assume that $\|f_\eta\|_\infty < \infty$. Assume that $\Delta$ can be written as $\Delta_01 + (\Delta_{01}, \ldots, \Delta_{0n})$, where

$$\Delta_0 = o_P(\delta_n^{1-\nu}), \quad \max_i \Delta_{0i} = o_P(\delta_n)$$

and that (3.1) or (3.4) holds respectively for $\alpha_\varepsilon < 1/2$ or $\alpha_\varepsilon > 1/2$. Then

$$\sup_{x \in \mathbb{R}} |M_n(x; \Delta)| = o_P(\sigma_{n,2}) + O_P(\epsilon \sqrt{n}).$$
Proof of Lemma 5.3. We prove part (a) only. The proof of the other part is analogous. Let
\[ a_{n,i}(x) = a_i(x) := 1_{\{x \leq \varepsilon_i \leq x + u_i\}} - E[1_{\{x \leq \varepsilon_i \leq x + u_i\}} | \mathcal{H}_{t-1}] , \]
so that \( M_n(x, u) = \sum_{i=1}^n a_i(x) \). We note that \( \{M_n(x, u), \mathcal{H}_n\} \) is a martingale array. Thus, by the Rosenthal’s inequality
\[
E[M_n(x, u)]^4 \leq C \sum_{i=1}^n \left( \frac{E(a_i(x))^2}{|H_i - 1} \right)^2 + C \sum_{i=1}^n Ea_i^4(x).
\]
Furthermore, \( |a_i(x)| \leq 1 \), so that
\[
E[M_n(x, u)]^4 \leq C n \sum_{i=1}^n \left( \frac{E(a_i(x))^2}{|H_i - 1} \right)^2 + C \sum_{i=1}^n Ea_i^2(x).
\]
Note that
\[
E[a_i^2(x)|\mathcal{H}_{i-1}] \leq E[1_{\{\varepsilon_i \leq x + |u_i|\}} | \mathcal{H}_{i-1}] - E[1_{\{\varepsilon_i \leq x - |u_i|\}} | \mathcal{H}_{i-1}] =: H_i^+(x) - H_i^-(x)
\]
and that for each \( i \), \( H_i^+(x) \) and \( H_i^-(x) \) are nondecreasing.

Introduce a partition \( \mathbb{R} = \cup_{r \in \mathbb{Z}} [x_r, x_{r+1}) \). Then
\[
EH_i^+(x_r) = EH_i^+(x_r) : \sigma_n,2 \int_{x_r}^{x_{r+1}} 1 \, dx \leq \sigma_n,2 \int_{x_r}^{x_{r+1}} H_i^+(x) \, dx,
\]
\[
EH_i^-(x_r) = EH_i^-(x_r) : \sigma_n,2 \int_{x_{r-1}}^{x_r} 1 \, dx \geq \sigma_n,2 \int_{x_{r-1}}^{x_r} H_i^-(x) \, dx.
\]
Thus, for arbitrary \( M \),
\[
\sum_{r=-M}^{M} E\left[H_i^+(x_r) - H_i^-(x_r)\right] \leq \sigma_n,2 \sum_{r=-M}^{M} \int_{x_r}^{x_{r+1}} H_i^+(x) \, dx - \int_{x_{r-1}}^{x_r} H_i^-(x) \, dx
\]
\[
\begin{align*}
&= \sigma_n,2 \int_{x_{-M}}^{x_M} (H_i^+(x) - H_i^-(x)) \, dx + \int_{x_M}^{x_{M+1}} H_i^+(x) \, dx - \int_{x_{-M-1}}^{x_{-M}} H_i^-(x) \, dx \\
&\leq \sigma_n,2 \int_{x_{-M}}^{x_M} (H_i^+(x) - H_i^-(x)) \, dx + 2.
\end{align*}
\]
Note that (recall that \( \xi_i = \varepsilon_i - \eta_i \))
\[
H_i^+(x) - H_i^-(x) = F_{\eta}(x - \xi_i + |u_i|) - F_{\eta}(x - \xi_i - |u_i|) = \int_{-|u_i|}^{u_i} f_{\eta}(x - \xi_i + y) \, dy,
\]
\[
(5.12)
\]
and

\begin{equation}
|H_i^+(x) - H_i^-(x)| \leq 2|u_i| \sup_x f_\eta(x).
\end{equation}

Using (5.12) we obtain

\begin{equation}
\sum_{r=-M}^{M} \mathbb{E} \left[ H_i^+(x_r) - H_i^-(x_r) \right] \\
\leq 1 + \sigma_{n,2} \mathbb{E} \left[ \int_{x-M}^{x-M} \int_{-|u_i|}^{|u_i|} f_\eta(x - \xi_i + y) \, dy \, dx \right] \\
\leq 1 + \sigma_{n,2} \mathbb{E} \left[ \int_{-|u_i|}^{|u_i|} \int_{-\infty}^{\infty} f_\eta(x + \xi_i + y) \, dx \, dy \right] \\
= 2 + \sigma_{n,2} \mathbb{E} \left[ \int_{-|u_i|}^{|u_i|} 1 \, dy \right] = 2 + 2\sigma_{n,2}|u_i|.
\end{equation}

Combining (5.12), (5.13) and (5.14),

\begin{equation}
\sum_{r=-M}^{M} \mathbb{E} \left[ \left( H_i^+(x_r) - H_i^-(x_r) \right)^2 \right] \leq C|u_i| + C\sigma_{n,2}u_i^2.
\end{equation}

Also, $\mathbb{E}a_i^2(x) \leq \mathbb{E}[H_i^+(x_r) - H_i^-(x_r)]$. By Markov inequality and (5.11),

\begin{equation}
P\left( \max_{r} \frac{1}{\sigma_{n,2}} |M_n(x_r, u)| > 1 \right) \leq \frac{1}{\sigma_{n,2}^4} \sum_{r} \mathbb{E}M_n^4(x_r, u) = \frac{1}{\sigma_{n,2}^4} \sum_{r} \mathbb{E}\left( \sum_{i=1}^{n} a_i(x_r) \right)^4 \\
\leq \frac{1}{\sigma_{n,2}^4} \left\{ Cn \sum_{r} \sum_{i=1}^{n} \mathbb{E} \left[ (\mathbb{E}a_i^2(x_r)|H_{i-1}) \right]^2 \right\} + C \sum_{r} \mathbb{E}a_i^2(x_r) \\
\leq C \frac{1}{\sigma_{n,2}^4} \left\{ n \sum_{i=1}^{n} |u_i| + n\sigma_{n,2} \sum_{i=1}^{n} u_i^2 + n + \sigma_{n,2} \sum_{i=1}^{n} |u_i| \right\}.
\end{equation}

The bound converges to 0 under the conditions (3.1) and (5.8). \hfill \Box

**Proof of Lemma 5.4.** Similarly to (5.1), $A_n(x; y)$ is decomposed as $\tilde{M}_n(x; y) + \tilde{N}_n(x; y)$, where $\tilde{M}_n(x; y)$ is the martingale part and $\tilde{N}_n(x; y)$ is the LRD part. We have

\begin{equation}
\tilde{N}_n(x; y) = \sum_{i=1}^{n} \left\{ \mathbb{E}[1_{x < \xi_i < y}|H_{i-1}] - (F_\varepsilon(y) - F_\varepsilon(x)) \right\} \leq n\|f_\eta + f_\varepsilon\|_\infty|y - x|.
\end{equation}
From [14, Lemma 14], \( \sup_{|y-x| \leq \epsilon n^{-1/2}} |\tilde{M}_n(x; y)| = O_P(\epsilon n^{-1/2}) \). Therefore, the case \( \alpha < 1/2 \) is proven.

Furthermore, for \( \alpha > 1/2 \),

\[
\sup_{|y-x| \leq \sigma_{n,2}} |\tilde{M}_n(x; y)| \leq 2 \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^{n} \left( 1_{\{\varepsilon_i \leq x\}} - E \left[ 1_{\{\varepsilon_i \leq x\} \mid \mathcal{H}_{i-1}\} \right] \right) \right| = o_P(\sigma_{n,2}).
\]

Proof of Lemma 5.5. We start with \( \alpha < 1/2 \). We can rewrite \( a_i(x) \) as follows:

\[
a_i(x) = 1_{\{\varepsilon_i \leq x + u_i\}} - 1_{\{\varepsilon_i \leq x\}} - (F_\eta(x - \xi_i + u_i) - F_\eta(x - \xi_i)).
\]

Let \( x \in [x_r, x_{r+1}) \), since \( 1_{\{\varepsilon_i \leq x\}} \) and \( F_\eta(x) \) are nondecreasing functions with respect to \( x \) we have

\[
a_i(x) \leq 1_{\{\varepsilon_i \leq x + u_i\}} - 1_{\{\varepsilon_i \leq x\}} - (F_\eta(x - \xi_i + u_i) - F_\eta(x_{r+1} - \xi_i))
= a_i(x_{r+1}) + 1_{\{\varepsilon_i \leq x_{r+1}\}} - 1_{\{\varepsilon_i \leq x\}} + F_\eta(x_{r+1} - \xi_i + u_i) - F_\eta(x - \xi_i + u_i).
\]

Thus, recalling the definition of \( A_n(x; y) \) given in (5.9),

\[
M_n(x, u) = M_n(x_r; u) + \sum_{i=1}^{n} \left( 1_{\{\varepsilon_i \leq x_{r+1}\}} - F_\varepsilon(x_{r+1}) - (1_{\{\varepsilon_i \leq x\}} - F_\varepsilon(x)) \right)
+ \sum_{i=1}^{n} (F_\eta(x_{r+1} - \xi_i + u_i) - F_\eta(x - \xi_i + u_i))
=: M_n(x_r; u) + A_n(x; x_{r+1}) + B_n(x; x_{r+1}; u).
\]

Now,

\[
\sup_{\mathbf{u}} \sup_{x \in \mathbb{R}} |M_n(x, \mathbf{u})| = \sup_{\mathbf{u}} \max_{x \in [x_r, x_{r+1})} \sup_{r \in \mathbb{Z}} |M_n(x; \mathbf{u})| \leq \sup_{\mathbf{u}} \max_{r} |M_n(x_r; \mathbf{u})| + \sup_{x \in [x_r, x_{r+1})} A_n(x; x_{r+1}) + \sup_{\mathbf{u}} \max_{r} \max_{x \in [x_r, x_{r+1})} B_n(x; x_{r+1}; \mathbf{u}).
\]

On account on Lemma 5.3, the first term in (5.17) is \( o_P(\sigma_{n,2}) \). The same holds for the second part by Lemma 5.4. For last term we consider Taylor expansion for \( F_\eta \):

\[
F_\eta(x_{r+1} - \xi_i + u_i) = F_\eta(x - \xi_i + u_i) + f_\eta(s)(x_{r+1} - x),
\]

where

\[
s = x - \xi_i + u_i + \frac{x_{r+1} - x}{2}.
\]
where \( s \in [x - \xi_i + u_i, x_{r+1} - \xi_i + u_i] \). Thus, the bound on \( B_n(x; x_{r+1}; u) \) is independent of \( u \)

\[
B_n(x; x_{r+1}; u) = \sum_{i=1}^{n} f_\eta(s)(x_{r+1} - x) \leq n f_\eta(s) \frac{1}{\sigma_{n,2}} = o(\sigma_{n,2})
\]
since \( n/\sigma_{n,2}^2 \to 0 \) for \( \alpha_\varepsilon < 1/2 \). Thus, the proof for \( \alpha_\varepsilon < 1/2 \) is finished.

If \( \alpha_\varepsilon > 1/2 \), then with the choice \( x_r = r \frac{\varepsilon}{\sqrt{n}} \) the first part in (5.17) is \( o_P(\sqrt{n}) \) and the same holds for the second part by applying Lemma 5.4. The term \( B_n(x; x_{r+1}; u) \) is bounded by

\[
B_n(x; x_{r+1}; u) = \sum_{i=1}^{n} f_\eta(s)(x_{r+1} - x) \leq n f_\eta(s) \frac{\varepsilon}{\sqrt{n}} = O(\varepsilon \sqrt{n}).
\]

\[\square\]

5.3. Proofs of Theorems 3.1 and 3.2. The result of Theorem 3.1 follows from Corollary 5.2 and uniform \( o_P(\sigma_{n,2}) \) negligibility of the martingale part in Lemma 5.5.

Now, let \( \alpha_\varepsilon > 1/2 \). Corollary 5.6 implies that for each \( \eta, \theta > 0 \) we may choose \( \varepsilon > 0 \) small enough so that

\[
P \left( \sup_{x \in \mathbb{R}} |n^{-1/2} M_n(x, \Delta)| > \theta \right) < 1 - \eta.
\]

Recall (5.2). This combined with (5.5) of Corollary 5.2 and (5.18) yields

\[\hat{K}_n(x) = K_n(x) + M_n(x; \Delta) + f_\varepsilon(x) \sum_{i=1}^{n} \Delta_i - f_\varepsilon^{(1)}(x) \Delta_0 \xi_{n,1} + O \left( \sum_{i=1}^{n} \Delta_i^2 \right) + O_P(\delta_n \varepsilon \sqrt{n}) + O_P(\delta_n \sigma_{n,1}), \]

\[
= K_n(x) + f_\varepsilon(x) \sum_{i=1}^{n} \varepsilon_i + \left( f_\varepsilon(x) \sum_{i=1}^{n} \Delta_i - f_\varepsilon(x) \sum_{i=1}^{n} \varepsilon_i - f_\varepsilon^{(1)}(x) \Delta_0 \xi_{n,1} \right) + O \left( \sum_{i=1}^{n} \Delta_i^2 \right) + o_P(\sqrt{n}) + O_P(\delta_n \sigma_{n,1}).
\]

Application of (2.5) yields

\[
n^{-1/2} \left( K_n(x) + f_\varepsilon(x) \sum_{i=1}^{n} \varepsilon_i \right) \Rightarrow W_1(x).
\]

The result of Theorem 3.2 follows.
5.4. Proof of (3.21). We have

\[ \mathbb{E}[L_b(X_1, X_2)] \sim 1 + \frac{O(b^2)}{2} \int s^2 K(s) \, ds \int f^{(2)}(v) \, dv = 1 + O(b^2). \]

Consequently, \( \mathbb{E}[L_b(X_1, X_2)] = \bar{\varepsilon} + O_P(b^2 \sigma_{n,1}/n) = \bar{\varepsilon} + o_P(\sigma_{n,1}/n). \)

Furthermore, since central moments are bounded by ordinary moments,

\[ \text{Var} \left( \frac{1}{n} \sum_{j=1}^{n} L_b^*(X_i, X_j) \varepsilon_j \right) = \frac{O(1)}{n^2} \sum_{j=1}^{n} \mathbb{E}[L_b^2(X_i, X_j)] + \frac{1}{n^2} \sum_{j,j'=1}^{n} \mathbb{E}[L_b^*(X_i, X_j) L_b^*(X_i, X_{j'})] \mathbb{E}[\varepsilon_j \varepsilon_{j'}] \]

It is straightforward to verify that for different indices \( i, j, j', \)

\[ \mathbb{E}[L_b(X_i, X_j) L_b(X_i, X_{j'})] = 1 + O(b). \]

Combining this with (5.20) yields

\[ \mathbb{E}[L_b^*(X_i, X_j) L_b^*(X_i, X_{j'})] = o(b). \]

Consequently, if (3.15) holds, then uniformly in \( i, \)

\[ \text{Var} \left( \frac{1}{n} \sum_{j=1}^{n} L_b^*(X_i, X_j) \varepsilon_j \right) = O((nb)^{-1}) + o(b \sigma_{n,1}^2/n^2) = o(\sigma_{n,1}^2/n^2). \]

5.5. Proof of (3.22). Recall (3.18) and (3.20). Also, recall that once (3.20) is evaluated, we may replace \( \hat{f}_b(X_i) \) with \( f(X_i). \) Therefore, we have

\[ \sum_{i=1}^{n} \Delta_i = O_P(n b^2) + \mathbb{E}[L_b(X_1, X_2)] \varepsilon_{n,1} + \frac{1}{n} \sum_{j=1}^{n} \bar{L}_b^*(X_j) \varepsilon_j, \]
where $\tilde{L}_b(X_j) = \sum_{i=1}^{n} \frac{1}{b} K_b(X_i - X_j)$ and $\tilde{L}_b^*(X_j)$ is its centered version.

Now, the variance of the third term in (5.22) is

$$\frac{1}{n^2} \sum_{j,j'=1}^{n} E[\varepsilon_j \varepsilon_{j'}] E \left[ \tilde{L}_b^*(X_j) \tilde{L}_b^*(X_{j'}) \right] = I_1 + I_2 + I_3 + I_4$$

$$= \frac{1}{n^2 b^2} \sum_{j,j'=1}^{n} E[\varepsilon_j \varepsilon_{j'}] \sum_{i=1}^{n} \text{Cov} \left[ \frac{1}{f(X_i)} K_b(X_i - X_j), \frac{1}{f(X_i)} K_b(X_i - X_{j'}) \right]$$

$$+ \frac{1}{n^2 b^2} \sum_{j,j'=1}^{n} E[\varepsilon_j \varepsilon_{j'}] \sum_{i,i'=1}^{n} \text{Cov} \left[ \frac{1}{f(X_i)} K_b(X_i - X_j), \frac{1}{f(X_{i'})} K_b(X_{i'} - X_{j'}) \right]$$

$$+ \frac{1}{n^2 b^2} \sum_{j=1}^{n} E[\varepsilon_j^2] \sum_{i=1}^{n} \text{Cov} \left[ \frac{1}{f(X_i)} K_b(X_i - X_j), \frac{1}{f(X_j)} K_b(X_j - X_j) \right]$$

$$+ \frac{1}{n^2 b^2} \sum_{j=1}^{n} E[\varepsilon_j^2] \sum_{i,i'=1}^{n} \text{Cov} \left[ \frac{1}{f(X_i)} K_b(X_i - X_j), \frac{1}{f(X_{i'})} K_b(X_{i'} - X_j) \right].$$

We start with $I_1$. We claim that

$$I_1 = O\left( \frac{1}{n^2 b^2} \sigma_{n,1}^2 \left( \frac{nb^2}{j,j',j'\neq j, j'=j} + \frac{b}{j,j',j'\neq j, i=j'=j} \right) \right) = O(\sigma_{n,1}^2/n).$$

Indeed, let us verify the case when $i, j, j'$ are different. We have (recall (P1))

$$\text{Cov} \left[ \frac{1}{f(X_i)} K_b(X_i - X_j), \frac{1}{f(X_i)} K_b(X_i - X_{j'}) \right]$$

$$\leq E \left[ \frac{1}{f(X_i)} K_b(X_i - X_j) \frac{1}{f(X_i)} K_b(X_i - X_{j'}) \right]$$

$$= \iiint \frac{1}{f(u)} K_b(u - v) \frac{1}{f(u)} K_b(u - v') f(v) f(v') du dv dv'$$

$$= b^2 \iiint \frac{1}{f(u)} \frac{1}{f(u)} K(s) f(u - sb) f(u - sb') du ds ds' = O(b^2).$$

In $I_2$, the term with all indices $i, i', j, j'$ different vanishes (recall that we work under (P1)). The other terms are verified in the similar way as for $I_1$, by computing expected values of products instead of covariances. We obtain:

$$I_2 = O\left( \frac{1}{n^2 b^2} \sigma_{n,1}^2 \left( \frac{nb}{i\neq i', j\neq j', i'=j, j'\neq j} + \frac{1}{i\neq i', j\neq j', i=j, j'\neq j} \right) \right) = O(\sigma_{n,1}^2/(nb)).$$
Similarly,
\[
I_3 = \frac{1}{n^2 b^2} \sum_{j=1}^{n} E[\varepsilon_j^2] \left( \sum_{i=1}^{n} O(b) + 1 \right) = O(b^{-1}) = o(n).
\]

Finally, for \(I_4\) let us note that with \(i, i', j\) different we obtain
\[
\text{Cov} \left( \frac{1}{f(X_i)} K_b(X_i - X_j), \frac{1}{f(X_{i'})} K_b(X_{i'} - X_j) \right) = 0.
\]

Thus,
\[
I_4 = \frac{O(1)}{n^2 b^2 n} \left( \sum_{i\neq i', i=j \text{ or } i'=j}^{n b} \right) = O(b^{-1}) = o(n).
\]

From \((5.22)\), \((5.20)\) and the above estimates we obtain
\[
\sum_{i=1}^{n} \Delta_i = O_P(n b^2) + \varepsilon_{n,1} + O_P(b^2 \sigma_{n,1}) + O_P\left( \frac{\sigma_{n,1}}{\sqrt{nb}} \right) + o_P(\sqrt{n}).
\]

If \(\alpha_\varepsilon < 1/2\) and \((3.16)\) holds, then the above estimate is \(o_P(\sigma_{n,2})\). Likewise, if \(\alpha_\varepsilon > 1/2\) and \((3.17)\) holds, then the bound is \(o_P(\sqrt{n})\). Thus, \((3.22)\) is proven. \(\Box\)

**Acknowledgement.** We would like to thank Professor Shiqing Ling for sending us the correction note. The work of the first author was supported by a NSERC (Natural Sciences and Engineering Research Council of Canada) grant. The work of the second author was conducted while being a Postdoctoral Fellow at the University of Ottawa.

**REFERENCES**


