Empirical process of residuals for regression models with long memory errors

Paweł Lorek[a,*], Rafał Kulik[b]

[a] *Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland*

[b] *Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa, ON, Canada K1N 6N5*

Abstract

We consider the residual empirical process in random design regression with long memory. We establish its limiting behaviour, showing that its rates of convergence are different from the rates of convergence for to the empirical process based on (unobserved) errors.

Keywords: long memory, empirical process of residuals, regression analysis

1. Introduction

Consider a random design regression model,

$$Y_i = m(X_{i1}, \dots, X_{ip}) + \varepsilon_i, \qquad i = 1, \dots, n,$$

where $m : \mathbb{R}^p \to \mathbb{R}$ is a deterministic function, $\{X_i = (X_{i1}, \dots, X_{ip})^T, i \geq 1\}$ is a p-dimensional time series, independent of a centered, stationary long range dependent (LRD) error sequence $\{\varepsilon, \varepsilon_i, -\infty < i < \infty\}$, with a distribution F_ε and density f_ε. The goal of this paper is to study the asymptotic properties of the empirical process of residuals,

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

*Work supported by NCN Research Grant UMO-2011/01/B/ST1/01305

Email addresses: `Pawel.Lorek@math.uni.wroc.pl` (Paweł Lorek), `rkulik@uottawa.ca` (Rafał Kulik)

where

$$\hat{\varepsilon}_i = Y_i - \hat{m}(X_{i1}, \ldots, X_{ip}) = \varepsilon_i - \Delta_i,$$

$$\Delta_i := \hat{m}(X_{i1}, \ldots, X_{ip}) - m(X_{i1}, \ldots, X_{ip})$$

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 independent random variables or weakly dependent stationary time series, see e.g. Koul and Ossiander (1994), Bai (1994), Müller et al. (2009), just to mention few.

As for regression models with long memory errors, in Chan and Ling (2008), 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

$$K_n(x) := \sum_{i=1}^{n} \left(\mathbf{1}_{\{\varepsilon_i \le x\}} - F_\varepsilon(x) \right),$$

in the sense that both $\sigma_{n,1}^{-1} K_n(\cdot)$ and $\sigma_{n,1}^{-1} \hat{K}_n(\cdot)$ converge weakly to $f_\varepsilon(x) Z_1$, 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

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

This conclusion was explicitly stated and proven in Proposition 3.1 in Koul and Surgailis (2010). In the latter article the authors obtained, among other results, a correct second-order expansion for an empirical process in the following framework. Let $Y_i = \mu + \varepsilon_i$, $\mu \in \mathbb{R}$. We want to test that

$$H_0 : F_\varepsilon(x) = F_0(x - \mu), \qquad \text{for all } x \in \mathbb{R},$$

against the alternative "H_0 is not true", where F_0 is a known distribution function. A possible test statistics can be based on an empirical process

$$\tilde{K}_n(x) := \sum_{i=1}^{n} \left(\mathbf{1}_{\{\varepsilon_i \le x + \bar{Y} - \mu\}} - F_\varepsilon(x) \right),$$

where $\bar{Y} = (Y_1 + \cdots + Y_n)/n$. We can see that the above empirical process has a similar form as $\hat{K}_n(x)$, that is

$$\tilde{K}_n(x) = \sum_{i=1}^{n} \left(\mathbf{1}_{\{\varepsilon_i \le x + \Delta_i\}} - F_\varepsilon(x) \right),$$

where $\Delta_i = \bar{Y} - \mu$.

The main goal of this paper is to extend results from Koul and Surgailis (2010) and establish a general theory on asymptotic behaviour for $\hat{K}_n(\cdot)$. In particular, this theory is applied to the parametric regression and a non-parametric regression; the latter in Kulik and Lorek (2011a) (a longer, `arXiv` version of the paper). 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).

The results for empirical processes in Sections 3.3 can be applied directly to establish limiting behaviour of quantiles (see (Ho and Hsing, 1996, Section 5)), empirical processes with estimated parameters (see Kulik (2008)) or in regression analysis (see Kulik and Lorek (2011b)). Furthermore, in a spirit of (Ho and Hsing, 1996, Section 3), our results should be applicable to the error density estimation. However, a precise proof requires at least third order expansion of the residual-based empirical process. In the extended version of the paper Kulik and Lorek (2011a) our theoretical results are confirmed by small simulation studies.

We note also that our considerations have a direct connection with a theory of weighted empirical processes based on long range dependent random variables. We refer to Koul and Surgailis (1997), Koul and Surgailis (2000) for original results in such framework, as well as to a recent monograph Giraitis et al. (2012). Introduction to general theory of weighted empirical processes can be found in Koul (2002).

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 U. Also, if U has finite mean, we denote $U^* = U - \mathrm{E}[U]$.

We shall consider the following assumption on the error sequence:

(E) ε_i, $i \geq 1$, is an infinite order moving average

$$\varepsilon_i = \sum_{k=0}^{\infty} c_k \eta_{i-k}, \qquad \text{with } c_0 = 1,$$

where η_i, $-\infty < i < \infty$, is a sequence of centered i.i.d. random variables, independent of X_i, $i \geq 1$. We assume that $\mathrm{E}[\eta^4] < \infty$, $\mathrm{E}[\eta_1^2] = 1$, and for some $\alpha_\varepsilon \in (0,1)$, $c_k \sim k^{-(\alpha_\varepsilon+1)/2} L_0(k)$ as $k \to \infty$, where $L_0(\cdot)$ is slowly varying at infinity.

Let

$$\varepsilon_{n,r} = \sum_{i=1}^{n} \sum_{0 \leq j_1 < \cdots < j_r < \infty} \prod_{s=1}^{r} c_{j_s} \eta_{i-j_s}. \tag{1}$$

In particular, $\varepsilon_{n,1} = \sum_{i=1}^{n} \varepsilon_i$ and if $r\alpha_\varepsilon < 1$,

$$\sigma_{n,r}^2 := \mathrm{Var}(\varepsilon_{n,r}) \sim \kappa(\alpha_\varepsilon, r) n^{2-r\alpha_\varepsilon} L_0^{2r}(n), \tag{2}$$

where $\kappa(\alpha_\varepsilon, r)$ is a multiplicative constant that depends on α_ε and r; see Lemma 6.1 in Ho and Hsing (1996).

From Ho and Hsing (1996) we know that for $r < \alpha_\varepsilon^{-1}$, as $n \to \infty$,

$$\sigma_{n,r}^{-1} \varepsilon_{n,r} \xrightarrow{d} Z_r, \qquad r = 1, 2, \tag{3}$$

where Z_r is a random variable which can be represented by a multiple Wiener-Itô integral. In particular, Z_1 is standard normal. Moreover, the random variables Z_1, Z_2 are uncorrelated, see e.g. (Koul and Surgailis, 1997, Eq. (1.22)). We also note that the convergence in (3) holds jointly.

Furthermore, let

$$S_{n,p}(x) = \sum_{i=1}^{n} \left(\mathbf{1}_{\{\varepsilon_i \leq x\}} - F_\varepsilon(x) \right) + \sum_{r=1}^{p} (-1)^{r-1} F_\varepsilon^{(r)}(x) \varepsilon_{n,r}.$$

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 $F_\varepsilon(\cdot)$. Following (Wu, 2003, Theorem 3) and (Ho and Hsing, 1996, Theorem 2.2) we conclude, in particular, that for $\alpha_\varepsilon < 1/2$,

$$\sigma_{n,2}^{-1} S_{n,1}(x) \Rightarrow f_\varepsilon^{(1)}(x) Z_2, \tag{4}$$

where Z_2 is the same random variable as in (3) and $\Rightarrow$ denotes weak convergence in $D(\mathbb{R})$. Otherwise, if $\alpha_\varepsilon > 1/2$, then

$$n^{-1/2} S_{n,1}(x) \Rightarrow W_1(x), \tag{5}$$

where $\{W_1(x), x \in \mathbb{R}\}$ is a Gaussian process. Furthermore, for $\alpha_\varepsilon > 1/3$,

$$\sigma_{n,2}^{-1} \sup_{x\in\mathbb{R}} |S_{n,2}(x)| \stackrel{a.s.}{\to} 0.$$

The structure of this Gaussian process and its covariance is given in a rather complicated form; see Wu (2003) for more details.

3. Results

Denote $X_i = (X_{i1}, \dots, X_{ip})$. Let $\mathbf{\Delta} = (\Delta_1, \dots, \Delta_n)$, where recall that

$\Delta_i = \varepsilon_i - \hat{\varepsilon}_i = \varepsilon_i - (Y_i - \hat{m}(X_{i1}, \dots, X_{ip})) = \hat{m}(X_{i1}, \dots, X_{ip}) - m(X_{i1}, \dots, X_{ip})$.

3.1. Empirical process of residuals: $\alpha_\varepsilon < 1/2$

The following result provides a uniform expansion of the process $\hat{K}_n(\cdot)$ and forms a basis for further analysis. A proof is given in Section 4.

Theorem 3.1. *Assume* (E) *with $\alpha_\varepsilon < 1/2$. Assume that $F_\eta(\cdot)$ is 3 times differentiable with bounded, continuous and integrable derivatives. Suppose that $\mathbf{\Delta}$ can be written as $\Delta_0 \mathbf{1} + (\Delta_{01}, \dots, \Delta_{0n})$, where*

- *$\Delta_0 = O_P(\delta_n)$ as $n \to \infty$;*
- *$|\Delta_{0i}| = o_P(\delta_n)$, uniformly in i,*

with a deterministic sequence δ_n such that

$$\delta_n + \frac{n^2 \delta_n}{\sigma_{n,2}^4} + \frac{n^2 \delta_n^2}{\sigma_{n,2}^3} + \frac{n\delta_n}{\sigma_{n,2}^3} \to 0. \tag{6}$$

Then with some $\nu \in (0, 1/2)$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| \hat{K}_n(x) - K_n(x) - f_\varepsilon(x) \sum_{i=1}^n \Delta_i - \frac{1}{2} f_\varepsilon^{(1)}(x) \sum_{i=1}^n \Delta_i^2 + f_\varepsilon^{(1)}(x) \Delta_0 \varepsilon_{n,1} \right| \\
&\quad = \quad O_P(\delta_n^{1-\nu} \sigma_{n,2}) + o_P\left(\delta_n \sigma_{n,1}\right) + O_P\left(\sum_{i=1}^n \Delta_i^3 \right).
\end{aligned}$$

In principle, this result is very similar to (Chan and Ling, 2008, Theorem 2.1). However, we provide $o_P(\cdot)$ rates of the approximation. This is crucial to establish limit theorems for the process $\hat{K}_n(\cdot)$.

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 (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 $\mathbf{\Delta}$ can be written as $\Delta_0\mathbf{1} + (\Delta_{01}, \ldots, \Delta_{0n})$, where*

- $\Delta_0 = O_P(\delta_n)$ *as $n \to \infty$;*
- $|\Delta_{0i}| = o_P(\delta_n)$, *uniformly in i,*

with a deterministic sequence δ_n such that

$$\delta_n + \sqrt{n}\delta_n^2 \to 0. \tag{7}$$

Then

$$\begin{aligned}\hat{K}_n(x) &= K_n(x) + f_\varepsilon(x)\sum_{i=1}^n \varepsilon_i + \left(f_\varepsilon(x)\sum_{i=1}^n \Delta_i - f_\varepsilon(x)\sum_{i=1}^n \varepsilon_i - f_\varepsilon^{(1)}(x)\Delta_0\xi_{n,1}\right) \\ &\quad + O_P\left(\sum_{i=1}^n \Delta_i^2\right) + o_P(\sqrt{n}) + O_P\left(\delta_n\sigma_{n,1}\right)\end{aligned}$$

where $n^{-1/2}\left(K_n(x) + f_\varepsilon(x)\sum_{i=1}^n \varepsilon_i\right)$ converges weakly to $W_1(x)$ from (5).

3.3. Application to 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

$$m(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

Hence, our model becomes

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i = \beta_0 + \beta^T X_i + \varepsilon_i,$$

where $\beta = (\beta_1, \ldots, \beta_p)^T$ and $X_i = (X_{i1}, \ldots, X_{ip})^T$. Let $\hat{\beta}$ be an estimator of β. The intercept β_0 is estimated using

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}^T\bar{X},$$

where

$$\bar{X} = (\bar{X}_1, \ldots, \bar{X}_p) = \left(n^{-1}\sum_{i=1}^n X_{i1}, \ldots, n^{-1}\sum_{i=1}^n X_{ip}\right)^T. \tag{8}$$

We make the following assumption on the predictors $X_{i1}, \ldots, X_{ip}$, $i \geq 1$:

(P) $X_i = (X_{i1}, \ldots, X_{ip})^T$ are such that for $r = 1, \ldots, p$, $\sup_i \mathrm{E}[|X_{ir}|] < \infty$.

Corollary 3.3. *Assume* (P) *and* (E) *and that for each* $i = 0, \ldots, p$

$$\hat{\beta}_i - \beta_i = o_P(\sigma_{n,1}/n). \tag{9}$$

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

(a) *Consider* Z_1, Z_2 *defined in* (3). *If* $\alpha_\varepsilon < 1/2$, *then*

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

(b) *If* $\alpha_\varepsilon > 1/2$, *then* $n^{-1/2}\hat{K}_n(x) \Rightarrow W_1(x)$, *where* W_1 *is the Gaussian process defined in (5).*

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 Kulik (2008). 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.

Remark 3.5. The condition (9) can be verified for many stationary sequences. For example, assume that $p = 1$ and $X_i = X_{i1}$, $i \geq 1$, is LRD linear sequence with the long memory parameter α_X, then the rate of convergence of the least squares estimator of $(\hat{\beta} - \beta) = (\hat{\beta}_1 - \beta_1)$ is either $\sqrt{n}$ or $n^{(\alpha_X + \alpha_\varepsilon)/2}$, for $\alpha_X + \alpha_\varepsilon > 1$ or $\max(\alpha_X, \alpha_\varepsilon) < 1/2$, respectively; see Robinson and Hidalgo (1997) and Guo and Koul (2008). The rate of convergence of a weighted LSE is $\sqrt{n}$; see Robinson and Hidalgo (1997). .

Proof of Corollary 3.3. We have

$$\hat{\beta}_0 - \beta_0 = \bar{\varepsilon} - (\hat{\beta}^T - \beta^T)\bar{X}$$

where $\bar{\varepsilon}$ is the sample means based on $\varepsilon_1, \ldots, \varepsilon_n$ and the sample mean vector is defined in (8). We have

$$\Delta_i \;=\; \hat{m}(X_i) - m(X_i) = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}^T - \beta^T)X_i = \bar{\varepsilon} + (\hat{\beta}^T - \beta^T)(X_i - \bar{X}). \tag{10}$$

From (2) we conclude that

$$\bar{\varepsilon} = O_P(\sigma_{n,1}/n), \qquad \sigma_{n,1}^2/n \sim \sigma_{n,2}, \quad \text{as } n \to \infty. \tag{11}$$

From (9) and Assumption (P) 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 δ_n fulfills (6). Therefore, the conditions of Theorem 3.1 are fulfilled with $\Delta_0 = \bar{\varepsilon}$ and $V = Z_1$.

Furthermore, from (10), $\sum_{i=1}^n \Delta_i = n\bar{\varepsilon} = \varepsilon_{n,1} = \sum_{i=1}^n \varepsilon_i$ and via (11),

$$\sum_{i=1}^n \Delta_i^2 = n\bar{\varepsilon}^2 + n\bar{\varepsilon} o_P(\sigma_{n,1}/n) + o_P(n\sigma_{n,1}^2/n^2) = n\bar{\varepsilon}^2 + o_P(\sigma_{n,2}). \tag{12}$$

Consequently, noting that $\delta_n \sigma_{n,1} \sim \sigma_{n,2}$ and $n\bar{\varepsilon}^2 = \bar{\varepsilon}\varepsilon_{n,1}$, we have

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

uniformly in x. The result of part (a) follows now from (3).

As for part (b), we recall that $\sum_{i=1}^n \Delta_i - \sum_{i=1}^n \varepsilon_i = 0$. Also, since $\alpha_\varepsilon > 1/2$, $\Delta_0 \varepsilon_{n,1} = O_P(\sigma_{n,1}^2/n) = O_P(\sigma_{n,2}) = o_P(\sqrt{n})$ and via (12), $\sum_{i=1}^n \Delta_i^2 = O_P(\sigma_{n,2}) = o_P(\sqrt{n})$. Finally, the choice of δ_n yields $\delta_n \sigma_{n,1} = o_P(\sqrt{n})$. Therefore, part (b) follows from Theorem 3.2. $\square$

4. Technical details

Let $\mathcal{H}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$. Let $\mathbf{u} = (u_1, \ldots, u_n)$ be a vector of scalars. Define

$$Z_n(x; \mathbf{u}) = \sum_{i=1}^n \left(\mathbf{1}_{\{\varepsilon_i \le x + u_i\}} - F_\varepsilon(x + u_i) \right) - \sum_{i=1}^n \left(\mathbf{1}_{\{\varepsilon_i \le x\}} - F_\varepsilon(x) \right).$$

The process $Z_n(x; \mathbf{u})$ is written as

$$Z_n(x; \mathbf{u}) = \sum_{i=1}^n \left(\mathbf{1}_{\{x < \varepsilon_i \le x + u_i\}} - \mathrm{E}\left[\mathbf{1}_{\{x < \varepsilon_i \le x + u_i\}} | \mathcal{H}_{i-1} \right] \right) \tag{13}$$

$$+ \sum_{i=1}^n \left(\mathrm{E}\left[\mathbf{1}_{\{x < \varepsilon_i \le x + u_i\}} | \mathcal{H}_{i-1} \right] - \mathrm{E}\left[\mathbf{1}_{\{x < \varepsilon_i \le x + u_i\}} \right] \right) =: M_n(x; \mathbf{u}) + N_n(x; \mathbf{u}).$$

Recall now that $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_n)$. We decompose

$$\hat{K}_n(x) - K_n(x) = \tag{14}$$
$$= M_n(x;\mathbf{\Delta}) + N_n(x;\mathbf{\Delta}) + f_\varepsilon(x)\sum_{i=1}^n \Delta_i + \frac{1}{2} f_\varepsilon^{(1)}(x)\sum_{i=1}^n \Delta_i^2 + O_P\left(\sum_{i=1}^n \Delta_i^3\right).$$

First, in Corollary 4.2 we will establish an asymptotic expansion for the LRD part $N_n(x;\mathbf{\Delta})$. This will be done by considering a special structure of $N_n(x;\mathbf{u})$ (see Lemma 4.1 and (16) below) and then "replacing" $\mathbf{u}$ with $\mathbf{\Delta}$ under proper assumptions for the latter.

Furthermore, we have to bound $M_n(x;\mathbf{\Delta})$. This will be done by obtaining a uniform bound on $M_n(x;\mathbf{u})$. In this way, we may utilize the martingale structure of the latter. Clearly, $M_n(x;\mathbf{\Delta})$ is not a martingale. The bounds are given in Lemma 4.3 and Lemma 4.5.

4.1. LRD part

Denote $\mathbf{u}_0 = u_0\mathbf{1}$, where $\mathbf{1}$ is the vector of dimension n, consisting of '1'. Recall that $\xi_i = \varepsilon_i - \eta_i$ and $\xi_{n,r}$ is defined in the analogous way as $\varepsilon_{n,r}$. In the first lemma we deal with $N_n(x;\mathbf{u}_0)$. The proof is included in Section 4.1.1.

Lemma 4.1. *Assume that $F_\eta(\cdot)$ is 3 times differentiable with bounded, continuous and integrable derivatives. Then with some $0<\nu<1/2$ and $\delta_n \to 0$,*

$$\sup_{|u_0|\le \delta_n^{1-\nu}} \sup_{x\in\mathbb{R}} \left| N_n(x;\mathbf{u}_0) + f_\varepsilon^{(1)}(x) u_0 \xi_{n,1}\right| = O_P\left(\delta_n^{1-\nu}\left(\sigma_{n,2}\vee\sqrt{n}\right) + \delta_n^{2(1-\nu)}\sigma_{n,1}\right). \tag{15}$$

Note now that the part $N_n(x,\mathbf{u})$ in (13) can be written as

$$N_n(x;\mathbf{u}) = \sum_{i=1}^n \left(F_\eta(x+u_i-\xi_i) - F_\eta(x-\xi_i) - \mathrm{E}F_\eta(x+u_i-\xi_i) + \mathrm{E}F_\eta(x-\xi_i)\right).$$

Let us choose $\mathbf{u} = \mathbf{u}_0 + (u_{01}, \ldots, u_{0n})$. If $\max_i(|u_{0i}|) = o(\delta_n)$, then applying first order Taylor expansion, and noting that ξ_i, $i\ge 1$, is LRD moving average with the same properties as ε_i, $i \ge 1$,

$$N_n(x;\mathbf{u}) - N_n(x;\mathbf{u}_0) = o(\delta_n)\sum_{i=1}^n \left(f_\eta(x+u_0-\xi_i) - \mathrm{E}f_\eta(x+u_0-\xi_i)\right) = o_P(\delta_n\sigma_{n,1}),$$

uniformly in u, u_0 and x, since $f_\eta^{(1)}$ is bounded and integrable. Combining this with (15), we have (recall $\nu < 1/2$)

$$\sup_{\mathbf{u}} \sup_{x\in\mathbb{R}} \left| N_n(x;\mathbf{u}) + f_\varepsilon^{(1)}(x) u_0 \xi_{n,1} \right| = O_P(\delta_n^{1-\nu} \left(\sigma_{n,2} \vee \sqrt{n}\right)) + o_P\left(\delta_n \sigma_{n,1}\right), \tag{16}$$

where $\sup_{\mathbf{u}}$ is taken over all $\mathbf{u}$ such that

$$\mathbf{u} = \mathbf{u}_0 + (u_{01}, \ldots, u_{0n}), \qquad \max_i(|u_{0i}|) = o(\delta_n), \qquad |u_0| = O(\delta_n^{1-\nu}).$$

In this way we end up with the following corollary.

Corollary 4.2. *Assume that $F_\eta(\cdot)$ is 3 times differentiable with bounded, continuous and integrable derivatives. Assume that $\mathbf{\Delta}$ can be written as $\Delta_0 \mathbf{1} + (\Delta_{01}, \ldots, \Delta_{0n})$, where*

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

Then

$$\sup_{x\in\mathbb{R}} \left| N_n(x;\mathbf{\Delta}) + f_\varepsilon^{(1)}(x) \Delta_0 \xi_{n,1} \right| = O_P(\delta_n^{1-\nu} \left(\sigma_{n,2} \vee \sqrt{n}\right)) + o_P\left(\delta_n \sigma_{n,1}\right).$$

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.

4.1.1. Proof of Lemma 4.1

Let $F_{n,\xi}(\cdot)$ be an empirical distribution function, associated with $\xi_1, \ldots, \xi_n$ and let $F_\xi(\cdot)$, $f_\xi(\cdot)$ be, respectively, distribution and density function of any of ξ_i. Note that ξ_i and η_i are independent for each fixed i, and $f_\xi * f_\eta = f_\varepsilon$. Recall that $\xi_{n,r}$ is defined in the analogous way as $\varepsilon_{n,r}$; see (1). From (2) we obtain that $\xi_{n,1} = O_P(\sigma_{n,1})$. Furthermore, let

$$\tilde{S}_{n,p}(x) = \sum_{i=1}^{n} \left(\mathbf{1}_{\{\xi_i \le x\}} - F_\xi(x)\right) + \sum_{r=1}^{p} (-1)^{r-1} F_\xi^{(r)}(x) \xi_{n,r}.$$

Note that $\tilde{S}_{n,p}$ is defined in the same way as $S_{n,p}$, but we use ξ_i's in the former instead of ε_i's. Nevertheless, we conclude from (4) and (5) that for $\alpha_\varepsilon < 1/2$,

$$\sigma_{n,2}^{-1} \tilde{S}_{n,1}(x) \Rightarrow f_\xi^{(1)}(x) Z_2,$$

where Z_2 is the same random variable as in (3). Otherwise, if $\alpha_\varepsilon > 1/2$, then

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

where Ψ is a Gaussian process and the convergence is in the Skorokhod topology.

We compute

$$\begin{aligned}
N_n(x;\mathbf{u}_0) &= n\int \left(F_\eta(x+u_0-v) - F_\eta(x-v)\right) d(F_{n,\xi}(v) - F_\xi(v)) \\
&= n\int \left(F_{n,\xi}(v) - F_\xi(v)\right)\left(f_\eta(x+u_0-v) - f_\eta(x-v)\right)\, dv \\
&= n\int \left(F_{n,\xi}(v) - F_\xi(v) + f_\xi(v)\xi_{n,1}/n\right)\left(f_\eta(x+u_0-v) - f_\eta(x-v)\right)\, dv \\
&\quad -\left(f_\varepsilon(x+u_0) - f_\varepsilon(x)\right)\xi_{n,1} \\
&= \int \tilde{S}_{n,1}(v)\left(f_\eta(x+u_0-v) - f_\eta(x-v)\right)\, dv - f_\varepsilon^{(1)}(x)u_0\xi_{n,1} + O(u_0^2)\xi_{n,1} \\
&= \int \tilde{S}_{n,1}(v) f_\eta^{(1)}(x-v)u_0(v)\, dv - f_\varepsilon^{(1)}(x)u_0\xi_{n,1} + O(u_0^2)\xi_{n,1},
\end{aligned}$$

where $u_0(v)$ lies between $x-v$ and $x+u_0-v$. From (4) and (5) we conclude that $\sup_v |\tilde{S}_{n,1}(v)| = O_P(\sigma_{n,2} \vee \sqrt{n})$. Therefore, with a $1 > \nu > 0$,

$$\sup_{|u_0| \le \delta_n^{1-\nu}} \sup_x \left| N_n(x;\mathbf{u}_0) + f_\varepsilon^{(1)}(x)u_0\xi_{n,1}\right| = O_P\left(\delta_n^{1-\nu}(\sigma_{n,2} \vee \sqrt{n}) + \delta_n^{2(1-\nu)}\sigma_{n,1}\right).$$

□

4.2. Martingale part

The proofs for martingale part are standard, in particular, they are similar as in Chan and Ling (2008). 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 4.3. *Assume that $\|f_\eta\|_\infty < \infty$.*

(a) *Let $x_r = r\frac{1}{\sigma_{n,2}}$. If $\alpha_\varepsilon < 1/2$ and (6) holds, then*

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

(b) *Let* $x_r = r\frac{\epsilon}{\sqrt{n}}$ *with* $\epsilon > 0$. *If* $\alpha_\varepsilon > 1/2$ *and* (7) *holds, then*

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

In both cases $\sup_{\mathbf{u}}$ *is taken over all* $\mathbf{u}$ *such that*

$$\mathbf{u} = \mathbf{u}_0 + (u_{01},\ldots,u_{0n}), \qquad \max_i(|u_{0i}|) = o(\delta_n), \qquad |u_0| = O(\delta_n^{1-\nu}). \quad (17)$$

Let

$$A_n(x;y) = \sum_{i=1}^n \left(\mathbf{1}_{\{\varepsilon_i\le y\}} - F_\varepsilon(y) - (\mathbf{1}_{\{\varepsilon_i\le x\}} - F_\varepsilon(x))\right). \quad (18)$$

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

Lemma 4.4. *Assume that* $\|f_\eta\|_\infty < \infty$.

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

Combining Lemmas 4.3 and 4.4 we obtain the following uniform behaviour of the martingale part.

Lemma 4.5. *Under the conditions of Lemma 4.3 we have*

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

As in case of Corollary 4.2 we conclude the following corollary.

Corollary 4.6. *Assume that* $\|f_\eta\|_\infty < \infty$. *Assume that* $\mathbf{\Delta}$ *can be written as* $\Delta_0\mathbf{1} + (\Delta_{01},\ldots,\Delta_{0n})$, *where*

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

and that (6) *or* (7) *holds respectively for* $\alpha_\varepsilon < 1/2$ *or* $\alpha_\varepsilon > 1/2$. *Then*

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

Proof of Lemma 4.3. We prove part (a) only, the other one is analogous. Let

$$a_{n,i}(x) = a_i(x) := \mathbf{1}_{\{x \le \varepsilon_t \le x+u_i\}} - \mathrm{E}[\mathbf{1}_{\{x \le \varepsilon_i \le x+u_i\}} | \mathcal{H}_{i-1}] \,,$$

so that $M_n(x, \mathbf{u}) = \sum_{i=1}^n a_i(x)$. We note that $\{M_n(x, \mathbf{u}), \mathcal{H}_n\}$ is a martingale array. Thus, by the Rosenthal's inequality

$$\mathrm{E}|M_n(x, \mathbf{u})|^4 \le C\mathrm{E}\left[\left(\sum_{i=1}^n \mathrm{E}(a_i(x)^2 | \mathcal{H}_{i-1})\right)^2\right] + C\sum_{i=1}^n \mathrm{E}a_i^4(x).$$

Furthermore, $|a_i(x)| \le 1$, so that

$$\mathrm{E}|M_n(x, \mathbf{u})|^4 \le Cn\sum_{i=1}^n \mathrm{E}\left[\left(\mathrm{E}(a_i^2(x)|\mathcal{H}_{i-1})\right)^2\right] + C\sum_{i=1}^n \mathrm{E}a_i^2(x). \quad (19)$$

Note that

$$\mathrm{E}[a_i^2(x)|\mathcal{H}_{i-1}] \le \mathrm{E}[\mathbf{1}_{\{\varepsilon_i \le x+|u_i|\}}|\mathcal{H}_{i-1}] - \mathrm{E}[\mathbf{1}_{\{\varepsilon_i \le x-|u_i|\}}|\mathcal{H}_{i-1}] =: H_i^+(x) - H_i^-(x)$$

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

$$\mathrm{E}H_i^+(x_r) = \mathrm{E}H_i^+(x_r) \cdot \sigma_{n,2}\int_{x_r}^{x_{r+1}} 1\, dx \le \sigma_{n,2}\mathrm{E}\left[\int_{x_r}^{x_{r+1}} H_i^+(x)\, dx\right],$$

$$\mathrm{E}H_i^-(x_r) = \mathrm{E}H_i^-(x_r) \cdot \sigma_{n,2}\int_{x_r-1}^{x_r} 1\, dx \ge \sigma_{n,2}\mathrm{E}\left[\int_{x_{r-1}}^{x_r} H_i^-(x)dx\right].$$

Thus, for arbitrary M,

$$\begin{aligned}
\sum_{r=-M}^{M} \mathrm{E}\left[H_i^+(x_r) - H_i^-(x_r)\right] &\le \sigma_{n,2}\sum_{r=-M}^{M} \mathrm{E}\left[\int_{x_r}^{x_{r+1}} H_i^+(x)\, dx - \int_{x_{r-1}}^{x_r} H_i^-(x)\, dx\right] \\
&= \sigma_{n,2}\mathrm{E}\left[\int_{x_{-M}}^{x_M} (H_i^+(x) - H_i^-(x))dx + \int_{x_M}^{x_{M+1}} H_i^+(x)dx - \int_{x_{-M-1}}^{x_{-M}} H_i^-(x)dx\right] \\
&\le \sigma_{n,2}\mathrm{E}\left[\int_{x_{-M}}^{x_M} (H_i^+(x) - H_i^-(x))dx\right] + 2\sigma_{n,2}.
\end{aligned}$$

Note that (recall that $\xi_i = \varepsilon_i - \eta_i$)

$$H_i^+(x) - H_i^-(x) = F_\eta(x - \xi_i + |u_i|) - F_\eta(x - \xi_i - |u_i|) = \int_{-|u_i|}^{|u_i|} f_\eta(x - \xi_i + y)\, dy, \tag{20}$$

and

$$|H_i^+(x) - H_i^-(x)| \le 2|u_i| \sup_x f_\eta(x). \tag{21}$$

Using (20) we obtain

$$\begin{aligned} \sum_{r=-M}^{M} \mathrm{E}\left[H_i^+(x_r) - H_i^-(x_r)\right] &\le 2 + \sigma_{n,2}\mathrm{E}\left[\int_{x_{-M}}^{x_M} \int_{-|u_i|}^{|u_i|} f_\eta(x - \xi_i + y)\, dy\, dx\right] \\ &\le 2 + \sigma_{n,2}\mathrm{E}\left[\int_{-|u_i|}^{|u_i|} \int_{-\infty}^{\infty} f_\eta(x + \xi_i + y)\, dx\, dy\right] \le 2 + 2\sigma_{n,2}|u_i|. \end{aligned} \tag{22}$$

Combining (20), (21) and (22),

$$\sum_{r=-M}^{M} \mathrm{E}\left[\left(H_i^+(x_r) - H_i^-(x_r)\right)^2\right] \le C|u_i| + C\sigma_{n,2}u_i^2.$$

Also, $\mathrm{E}a_i^2(x) \le \mathrm{E}[H_i^+(x_r) - H_i^-(x_r)]$. By Markov inequality and (19),

$$\begin{aligned} P\left(\max_r \frac{1}{\sigma_{n,2}}|M_n(x_r, \mathbf{u})| > 1\right) &\le \frac{1}{\sigma_{n,2}^4}\sum_r \mathrm{E}M_n^4(x_r, \mathbf{u}) = \frac{1}{\sigma_{n,2}^4}\sum_r \mathrm{E}\left(\sum_{i=1}^n a_i(x_r)\right)^4 \\ \le\ & \frac{1}{\sigma_{n,2}^4}\left\{Cn\sum_r\sum_{i=1}^n \mathrm{E}\left[\left(\mathrm{E}(a_i^2(x_r)|\mathcal{H}_{i-1})\right)^2\right] + C\sum_r\sum_{i=1}^n \mathrm{E}a_i^2(x_r)\right\} \\ \le\ & \frac{C}{\sigma_{n,2}^4}\left\{n\sum_{i=1}^n |u_i| + n\sigma_{n,2}\sum_{i=1}^n u_i^2 + n + \sigma_{n,2}\sum_{i=1}^n |u_i|\right\}. \end{aligned}$$

The bound converges to 0 under the conditions (6) and (17). □

Proof of Lemma 4.4. Similarly to (13), $A_n(x;y)$ 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

$$\tilde{N}_n(x;y) = \sum_{i=1}^n \left(\mathrm{E}[1_{\{x<\varepsilon_i<y\}}|\mathcal{H}_{i-1}] - (F_\varepsilon(y) - F_\varepsilon(x))\right) \le n\|f_\eta + f_\varepsilon\|_\infty |y - x|.$$

From (Wu, 2003, Lemma 14), $\sup_{|y-x|\le \epsilon n^{-1/2}} |\tilde{M}_n(x;y)| = O_P(\epsilon n^{-1/2})$. Therefore, the case $\alpha_\varepsilon > 1/2$ is proven. Furthermore, for $\alpha_\varepsilon < 1/2$,

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

□

Proof of Lemma 4.5. We start with $\alpha_\varepsilon < 1/2$. We can rewrite $a_i(x)$ as follows:

$$a_i(x) = \mathbf{1}_{\{\varepsilon_i \le x+u_t\}} - \mathbf{1}_{\{\varepsilon_i\le x\}} - (F_\eta(x-\xi_i+u_i) - F_\eta(x-\xi_i)).$$

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

$$\begin{aligned} a_i(x) &\le \mathbf{1}_{\{\varepsilon_i\le x_{r+1}+u_i\}} - \mathbf{1}_{\{\varepsilon_i\le x\}} - (F_\eta(x-\xi_i+u_i) - F_\eta(x_{r+1}-\xi_i)) \\ &= a_i(x_{r+1}) + \mathbf{1}_{\{\varepsilon_i\le x_{r+1}\}} - \mathbf{1}_{\{\varepsilon_i\le x\}} + F_\eta(x_{r+1}-\xi_i+u_i) - F_\eta(x-\xi_i+u_i). \end{aligned}$$

Thus, recalling the definition of $A_n(x;y)$ given in (18),

$$\begin{aligned} M_n(x,\mathbf{u}) &= M_n(x_r;\mathbf{u}) + \sum_{i=1}^{n} \left(\mathbf{1}_{\{\varepsilon_i\le x_{r+1}\}} - F_\varepsilon(x_{r+1}) - (\mathbf{1}_{\{\varepsilon_i\le x\}} - F_\varepsilon(x))\right) \\ &+ \sum_{i=1}^{n} (F_\eta(x_{r+1}-\xi_i+u_i) - F_\eta(x-\xi_i+u_i)) =: M_n(x_r;\mathbf{u}) + A_n(x;x_{r+1}) + B_n(x;x_{r+1};\mathbf{u}). \end{aligned}$$

Now,

$$\begin{aligned} \sup_{\mathbf{u}} \sup_{x\in\mathbb{R}} |M_n(x;\mathbf{u})| &= \sup_{\mathbf{u}} \max_{r\in\mathbb{Z}} \sup_{x\in[x_r,x_{r+1})} |M_n(x;\mathbf{u})| \le \sup_{\mathbf{u}} \max_{r} |M_n(x_r;\mathbf{u})| \\ &+ \sup_{|x_1-x_2|\le\sigma_{n,2}^{-1}} A_n(x;x_{r+1}) + \sup_{\mathbf{u}} \max_{r} \max_{x\in[x_r,x_{r+1})} B_n(x;x_{r+1};\mathbf{u}). \quad (23) \end{aligned}$$

By Lemma 4.3, the first term in (23) is $o_P(\sigma_{n,2})$. The same holds for the second part by Lemma 4.4. For last term we consider Taylor expansion for F_η:

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

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

$$B_n(x; x_{r+1}; \mathbf{u}) = \sum_{i=1}^{n} f_\eta(s)(x_{r+1} - x) \leq n f_\eta(s) \frac{1}{\sigma_{n,2}} = o(\sigma_{n,2})$$

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 $x_r = r\frac{\epsilon}{\sqrt{n}}$ the first part in (23) is $o_P(\sqrt{n})$ and the same holds for the second part by applying Lemma 4.4. The term $B_n(x; x_{r+1}; \mathbf{u})$ is bounded by

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

□

References

Bai, J., 1994. Weak convergence of the sequential empirical processes of residuals in ARMA models. Ann. Statist. 22, 2051–2061.

Chan, N.H., Ling, S., 2008. Residual empirical processes for long and short memory time series. Ann. Statist. 36, 2453–2470. Correction note 2010. Ann. Statist. 38, 3839.

Giraitis, L., Koul, H.L., Surgailis, D., 2012. Large sample inference for long memory processes. Imperial College Press, London.

Guo, H., Koul, H.L., 2008. Asymptotic inference in some heteroscedastic regression models with long memory design and errors. Ann. Statist. 36, 458–487.

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

Koul, H.L., 2002. Weighted empirical processes in dynamic nonlinear models. Second ed., Lecture Notes in Statistics, 166. Springer-Verlag, New York.

Koul, H.L., Ossiander, M., 1994. Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. Ann. Statist. 22, 540–562.

Koul, H.L., Surgailis, D., 1997. Asymptotic expansion of m-estimators with long-memory errors. Ann. Statist. 25, 818–850.

Koul, H.L., Surgailis, D., 2000. Second-order behavior of M-estimators in linear regression with long-memory errors. J. Statist. Plann. Inf. 91, 399–412.

Koul, H.L., Surgailis, D., 2010. Goodness-of-fit testing under long-memory errors. J. Statist. Plann. Inf. 140, 3742–3753.

Kulik, R., 2008. Empirical process of long-range dependent sequences when parameters are estimated. J. Statist. Plann. Inf. 139, 287–294.

Kulik, R., Lorek, P., 2011a. Empirical process of residuals for regression models with long memory errors (longer version). arXiv:1102.4368 [math.PR] .

Kulik, R., Lorek, P., 2011b. Some results on random design regression with long memory errors and predictors. J. Statist. Plann. Inf. 141, 508–523.

Müller, U.U., Schick, A., Wefelmeyer, W., 2009. Estimating the innovation distribution in nonparametric autoregression. Probab. Theory Related Fields 144.

Robinson, P.M., Hidalgo, F.J., 1997. Time series regression with long-range dependence. Ann. Statist. 25, 77–104.

Wu, W.B., 2003. Empirical processes of long-memory sequences. Bernoulli 9, 809–831.