Electronic Journal of Statistics
Vol. 5 (2011) 856–898
ISSN: 1935-7524
DOI: 10.1214/11-EJS629

Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors

Rafał Kulik

Department of Mathematics and Statistics
University of Ottawa
585 King Edward Av.
Ottawa ON K1N 6N5, Canada

and

Cornelia Wichelhaus

Institute of Applied Mathematics
University of Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg, Germany

Abstract: This paper considers nonparametric regression models with long memory errors and predictors. Unlike in weak dependence situations, we show that the estimation of the conditional mean has influence on the estimation of both, the conditional variance and the error density. In particular, the estimation of the conditional mean has a negative effect on the asymptotic behaviour of the conditional variance estimator. On the other hand, surprisingly, estimation of the conditional mean may reduce convergence rates of the residual-based Parzen-Rosenblatt density estimator, as compared to the errors-based one. Our asymptotic results reveal small/large bandwidth dichotomous behaviour. In particular, we present a method which guarantees that a chosen bandwidth implies standard weakly dependent-type asymptotics. Our results are confirmed by an extensive simulation study. Furthermore, our theoretical lemmas may be used in different problems related to nonparametric regression with long memory, like cross-validation properties, bootstrap, goodness-of-fit or quadratic forms.

AMS 2000 subject classifications: Primary 62G05; secondary 62G08, 62E20, 62M10.
Keywords and phrases: long memory, nonparametric regression, random design, conditional variance, density estimation.

Received September 2010.

1. Introduction: Random design regression with long memory errors

Consider the random design regression model

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \qquad i = 1, \dots, n. \tag{1.1}$$

We shall assume that the predictors X_i, $i \geq 1$, are random variables with unit variance and density $f = f_X$, independent of ε_i, $i \geq 1$. The error sequence is assumed to be centered with unit variance and density f_ε.

The goal of this paper is two-fold. First, we study the asymptotic properties of a nonparametric kernel estimator of $\sigma^2(\cdot)$ in the model (1.1). We estimate the conditional mean by the usual Nadaraya-Watson estimator

$$\hat{m}(x) = \hat{m}_b(x) = \frac{1}{nb\hat{f}_b(x)} \sum_{i=1}^{n} Y_i K_b(x - X_i), \tag{1.2}$$

where

$$\hat{f}_b(x) = \hat{f}_{b,X}(x) = \frac{1}{nb} \sum_{i=1}^{n} K_b(x - X_i) \tag{1.3}$$

and $K_b(\cdot) = K(\cdot/b)$, $K(\cdot)$ being a kernel function. Having done this we compute the residuals $Y_i - \hat{m}_b(X_i)$ and estimate $\sigma^2(x)$ by

$$\hat{\sigma}^2(x) = \hat{\sigma}_h^2(x) = \frac{1}{nh\hat{f}_h(x)} \sum_{i=1}^{n} (Y_i - \hat{m}(X_i))^2 K_h(x - X_i). \tag{1.4}$$

Second, we are interested in estimating $f_\varepsilon(\cdot)$ in a nonparametric way in the model (1.1). To simplify the exposition, in this case we assume that $\sigma(\cdot) \equiv 1$ and compute the residuals as $\hat{\varepsilon}_i = (Y_i - \hat{m}_b(X_i))$. The error density is estimated by the standard Parzen-Rosenblatt estimator

$$\hat{f}_{h,\Delta}(x) = \frac{1}{nh} \sum_{i=1}^{n} K_h(x - \hat{\varepsilon}_i). \tag{1.5}$$

The problem of nonparametric estimation of the conditional variance received a lot of attention in the past decade. Most of the work focuses on a fixed design regression model $Y_i = m(x_i) + \sigma(x_i)\varepsilon_i$, $i \geq 1$, where $x_i = i/n$ and ε_i, $i \geq 1$, are independent and identically distributed (i.i.d.) random variables. In this set-up it is shown in [3] and [35] that if $m(\cdot)$ is differentiable, then there is no influence of the estimation of $m(\cdot)$ on the minimax rates of convergence for the estimation of $\sigma(\cdot)$. In case of random design regression, a similar lack of influence was proven in [10] for weakly dependent data; see also [37], where the authors considered dependent predictors and i.i.d. errors.

However, very little is known in case of long memory errors and/or predictors. In the statistical literature, long range dependence (LRD) is modeled typically by linear processes with slowly decaying coefficients. To be more specific, the predictors and the errors will be described as

$$X_i = \mu + \sum_{k=0}^{\infty} a_k \zeta_{i-k}, \qquad \varepsilon_i = \sum_{k=0}^{\infty} c_k \eta_{i-k}, \tag{1.6}$$

where ζ_i, η_i, $-\infty < i < \infty$, are mutually independent sequences of centered i.i.d. random variables, and the coefficients a_k, c_k behave like $a_k \sim k^{-(\alpha_X+1)/2}$,

$c_k \sim k^{-(\alpha_\varepsilon+1)/2}$ for some $\alpha_X, \alpha_\varepsilon \in (0,1)$. In particular, the covariances are non-summable. If the predictors are LRD and the errors are i.i.d., it is shown in [37] that there is no influence of estimating the conditional mean. However, this is basically due to their imposed conditions on the bandwidth choice: for a *small* bandwidth the nonparametric kernel estimation behaves as if data were independent. On the other hand, in [15] the authors studied a model with a parametric conditional mean and established a *large bandwidth* behaviour. We are not aware of any further results regarding conditional variance estimation in random design regression with LRD.

As for the error density estimation, let us note first that there is a number of results describing the behaviour of the empirical process of residuals and different estimators of the error distribution under *weak dependence*. The reader is referred to [20, 30, 31, 32] and references therein. As for the empirical processes of residuals in long memory regression models, the reader is referred to [4] (together with a correction note), [21] or [24].

On the other hand, if the error density is considered in the nonparametric case, let us start with a quote from [9]: *Surprisingly, despite the widespread use of residuals as proxies for unobserved errors, to the best of the author's knowledge, no results about optimal (in any sense) estimation of a nonparametric error density is known.* In fact, in the aforementioned paper the author shows that in case of independent errors and predictors, estimation of the conditional mean does not influence the rates of convergence for an estimator of f_ε. One has to point out here, that Efromovich does not consider kernel estimators and a lot of technical problems come from the fact the he considers a data-based adaptive estimator. Moreover, there are very few results describing the asymptotic behaviour of the much simpler Parzen-Rosenblatt or a histogram-type error density estimator, even if errors and predictors are independent or weakly dependent, see e.g. [6] or [28]. Finally, we are not aware of any single result which describes the asymptotic behaviour of any error density estimator in the model (1.1) with LRD.

The goal of this paper is to present the full asymptotic theory for the conditional variance and the error density estimation in the model (1.1), when errors and/or predictors have long memory. Such situations are very often encountered in, especially, financial time series. We will distinguish between the *oracle* and the *non-oracle* case. In the first situation, we assume that $m(\cdot)$ is known, which amounts to estimation of $\sigma(\cdot)$ or $f_\varepsilon(\cdot)$ from the direct observations $\sigma^2(X_i)\varepsilon_i^2$. By applying a log-transformation, we may see that the first problem is similar to nonparametric estimation of the conditional mean. Therefore, one can link our results to the existing literature, e.g. [7, 8] and [29]. One has to mention though that the rates of convergence for the conditional variance are different than those for the conditional mean. This is due to the fact that for long memory the sequences $\sum_{i=1}^n \varepsilon_i$ (the conditional mean case) and $\sum_{i=1}^n (\varepsilon_i^2 - \mathrm{E}[\varepsilon_i^2])$ (the conditional variance case) show a different limiting behaviour.

For the conditional variance estimation the results for the oracle case are given in Proposition 3.1. In the non-oracle situation we have to estimate $m(\cdot)$ first. In Theorem 3.5 we state the results which show the influence of estimating $m(\cdot)$ on the estimation of $\sigma(\cdot)$. If the errors ε_i, $i \geq 1$, are i.i.d., there is no influence, i.e. the oracle and non-oracle case are the same, regardless whether the predictors are LRD or not. This agrees with the findings in [10] and [37]. However, if the errors are LRD, there is an additional term which may contribute to the limiting behavior of the conditional variance estimator. Our results extend the findings in [15], where the case of a parametric mean and Gaussian predictors was considered.

In case of the density estimation in Theorem 3.10 it turned out that the limiting behaviour in the oracle case (for this, the reader is referred to [5, 36]) is also different as compared to the non-oracle case.

Finally, we discuss the bandwidth choice. There are many solutions for this problem in the i.i.d. setting, however, there are very few results in the LRD case. Let us note that from a practical point of view, the *large bandwidth* asymptotics (see (3.5), (3.6) and (3.8) and [15, Theorem 3.1]) are of a very limited use: one has to estimate the long memory parameter of the errors and one has to estimate several other parameters.

As for the first problem, in the model (1.1), we are not aware of any results regarding consistency of the long memory parameter. Such results, with $\log(n)$-rate of convergence, are known in case of a parametric regression; see [15]. The second problem is very difficult, for example LRD-based bootstrap typically fails (see [27]). Therefore, in Section 3.3 we discuss how to use a plug-in method in order to justify that the chosen bandwidth guarantees the *small bandwidth* asymptotics.

From the technical point of view, this paper can be viewed as a generalization of [37, 15] (in the conditional variance case only). In the first paper, the authors considered dependent predictors and i.i.d. errors. The latter assumption greatly simplifies computations and suppresses a possible additional effect coming from the non-oracle parts. Furthermore, the authors considered small bandwidth asymptotics, which suppress a possible LRD effect of the predictors on the convergence rates of the oracle part. In the second paper, the authors considered a parametric mean and Gaussian LRD predictors. Furthermore, they focused on large bandwidth asymptotics. Here, we consider LRD errors and predictors modeled by infinite order moving averages (1.6). Consequently, we have to develop some new results on multivariate density expansions and new limit theorems for weighted quadratic forms (for some results in the latter the reader may also be referred to the unpublished manuscript [13]). Needless to mention, the proofs, as usual in the long memory case, are very technical, so that very often we present additionally some heuristic. Furthermore, our theoretical results may be potentially used to establish asymptotics in such problems like cross-validation ([25]), nonparametric goodness-of-fit ([13]) or asymptotics of weighted quadratic forms.

Our theoretical findings are verified by extensive simulations in Section 4. Finally, we apply the estimation procedures to some real data in Section 5.

We would also like to mention, that there is a parallel paper, [26]. There, the authors deal with the parametric mean case, $m(x) = \beta_0 + \beta_1 x$, but predictors are allowed to form different long memory sequences, including linear processes and stochastic volatility models. The latter are important especially in modeling financial time series, which are typically uncorrelated, but have long memory in squares.

2. Preliminaries

For any random variable Z, we denote by $\bar{Z} = Z - \mathrm{E}[Z]$. Also, $f_Z(\cdot)$ denotes the density of a given random variable Z.

2.1. Predictors and the error sequence

We consider 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 $\mathrm{E}X_1 = \mu < \infty$ and $\mathrm{Var}(X_1) = 1$. In this case we write $X_i = X_i^*$ and denote $\mathcal{X}_i = \sigma(X_1^*, \ldots, X_i^*)$.

(P2) X_i, $i \geq 1$, is the infinite order moving average given by

$$X_i = \mu + \sum_{k=0}^{\infty} a_k \zeta_{i-k}, \qquad \text{with } a_0 = 1 \text{ and } \mu < \infty,$$

where ζ_i, $-\infty < i < \infty$, is a sequence of centered, i.i.d. random variables. Furthermore, as $k \to \infty$, $a_k \sim A_0 k^{-(\alpha_X+1)/2}$ for some $\alpha_X \in (0,1)$ and $0 < A_0 < \infty$. In this case we denote $\mathcal{X}_i = \sigma(\zeta_i, \zeta_{i-1}, \ldots)$ and assume that f_ζ, f'_ζ are bounded and integrable.

Consequently, under (P2), $\sigma_{n,X}^2 = \mathrm{Var}\left(\sum_{i=1}^n X_i\right) \sim A_1^2 n^{2-\alpha_X}$, where (see e.g. [18, Lemma 6.1])

$$A_1^2 = \frac{2A_0^2}{(2-\alpha_X)(1-\alpha_X)} \left[\int_0^\infty (x+x^2)^{-(\alpha_\varepsilon+1)/2} dx\right].$$

Similarly, we shall consider the corresponding assumptions on the error sequence:

(E1) ε_i, $i \geq 1$, is a sequence of centered i.i.d. random variables with finite fourth moment and $\mathrm{E}\varepsilon_1^2 = 1$.

(E2) 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 with finite fourth moment, $\mathrm{E}[\varepsilon_1^2] = 1$, and for some $\alpha_\varepsilon \in (0,1)$ and $0 < C_0 < \infty$ we have $c_k \sim C_0 k^{-(\alpha_\varepsilon+1)/2}$ as $k \to \infty$. Denote $\mathcal{H}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$ and assume that f_η, f'_η are bounded and integrable.

Under (E2), $\sigma_{n,\varepsilon}^2 = \mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i\right) \sim C_1^2 n^{2-\alpha_\varepsilon}$, where

$$C_1^2 = \frac{2C_0^2}{(2-\alpha_\varepsilon)(1-\alpha_\varepsilon)} \int_0^\infty (x+x^2)^{-(\alpha_\varepsilon+1)/2}\, dx. \tag{2.1}$$

Remark 2.1. Our results are in principle extendable to the case of weakly dependent (e.g. mixing) innovations η_i. In [25] the authors considered estimation of the conditional mean for a very general class of errors. In particular, it was assumed that the random variables η_i, $-\infty < i < \infty$, are modeled by a FARIMA-GARCH process introduced in [2]. However, such extension requires precise results on the limiting behaviour of, for example, $\sum_{i,j=1}^n \varepsilon_i\varepsilon_j$. We are not aware of such results in case of dependent innovations η_i.

On the other hand, the methods used in our paper rely strongly on the innovations ζ_i, $-\infty < i < \infty$, being independent.

2.2. Assumptions on bandwidths and functions

Let $\kappa_i = \int u^i K(u)du$. It is assumed that $K(\cdot)$ is symmetric and positive, and has a bounded support $[-T,T]$ with $K(T)=0$. Also, we assume that $\kappa_0 = 1$, $\kappa_2 \neq 0$ and that $K(\cdot)$ is bounded and continuous. Denote for future use $K_h(\cdot) := K(\cdot/h)$.

The limit theorems in this paper are obtained for a fixed, but arbitrary, point x. Accordingly, consider the following assumptions on $f = f_X$, σ and the bandwidth h. Let $\mathcal{I}$ be a compact interval such that $x \in \mathcal{I}$.

- (D1) f, σ are defined on the set R of real numbers with $f, \sigma \in C^2(\mathcal{I})$, where C^2 is the class of twice-differentiable functions, with bounded and continuous second order derivatives.
- (D2) $\inf_{x\in\mathcal{I}} f(x) > 0$.
- (D3) $\sigma(x) > 0$ for all $x \in R$.
- (H0) $nh^5 + \frac{\log n}{\sqrt{nh}} \to 0$.

Condition (H0) is standard also in i.i.d. or weakly dependent situations. In particular, when one considers CLT for $\hat{\sigma}^2(x) - \sigma^2(x)$, the condition $nh^5 \to 0$ makes the bias negligible.

3. Results and discussion

3.1. Conditional variance estimation

The first lemma describes the behaviour of $\hat{\sigma}_h^2(\cdot)$ in case of known $m(\cdot)$.

To state our result, we denote by H_2 a Hermite-Rosenblatt random variable. If $\alpha_\varepsilon < 1/2$ it is defined as a multiple Wiener-Itô integral (see e.g. [34]),

$$H_2 = C_2^{-1} \int_{\mathcal{J}} \int_0^1 \prod_{j=1}^s [\max(v-u_j, 0)]^{-(\alpha_\varepsilon+1)/2} dv dB(u_1) dB(u_2)\ , \tag{3.1}$$

where $\mathcal{J} = \{(u_1, u_2) : -\infty < u_1 < u_2 < 1\}$, $\{B(t), t \in \mathbb{R}\}$ is a two-sided standard Brownian motion and

$$C_2^2 = \{2[1-\alpha_\varepsilon][1-2\alpha_\varepsilon]\}^{-1} \left[\int_0^\infty (x+x^2)^{-(\alpha_\varepsilon+1)/2} dx\right]^2 . \tag{3.2}$$

The random variable H_2 is non-Gaussian and the constant C_2 assures that $\mathrm{E}[H_2^2] = 1$.

Proposition 3.1. *Assume* (P2) *and* (E2) *and that the conditions* (D1)-(D3) *and* (H0) *hold.*

- *If*
$$hn^{(1-2\alpha_\varepsilon)} \to 0, \tag{3.3}$$
then
$$\sqrt{nhf(x)}\left(\hat{\sigma}^2(x) - \sigma^2(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^4(x)\mathrm{E}[(\varepsilon_1 - 1)^2] \int K^2(u)\, du\right). \tag{3.4}$$
- *If* $hn^{(1-2\alpha_\varepsilon)} \to \infty$, *then*
$$n^{\alpha_\varepsilon}\left(\hat{\sigma}^2(x) - \sigma^2(x)\right) \xrightarrow{d} C_2\sigma^2(x)H_2 . \tag{3.5}$$

Remark 3.2. In the above lemma, in the borderline case $hn^{(1-2\alpha_\varepsilon)} \to 1$, say, we clearly have

$$\sqrt{nhf(x)}\left(\hat{\sigma}^2(x) - \sigma^2(x)\right) \xrightarrow{\mathrm{d}} \mathcal{N} + C_2\sqrt{f(x)}\sigma^2(x)H_2 ,$$

where $\mathcal{N}$ is the normal random variable with variance as in (3.4). It can be proven that the random variables $\mathcal{N}$ and H_2 are independent; see e.g. comments following Theorem 2 in [36].

Remark 3.3. Consider the case $\alpha_X = 1$, so that the predictors are weakly dependent. Then, we are automatically in scenario 2 of Proposition 3.1. If $\alpha_\varepsilon \in (1/2, 1)$ there is no influence of LRD of the errors on the rates of convergence. If $\alpha_\varepsilon \in (0, 1/2)$ and h is *small*, then there is still no influence of LRD of errors. However, if h is *big*, LRD of errors influences the limit. Note further that the meaning of *small* and *big* bandwidth is different than in case of estimating the conditional mean. Namely, in the model $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$, for the standard kernel estimator $\hat{m}_b(\cdot)$ of $m(\cdot)$, we have $\sqrt{nb}$ or $n^{\alpha_\varepsilon/2}$ rate of convergence if, respectively, $bn^{(1-\alpha_\varepsilon)} \to 0$ or $bn^{(1-\alpha_\varepsilon)} \to \infty$. We refer to [29] for more details.

Remark 3.4. In [15] it was established (see Theorem 3.1(a) with Assumption 5) that the scaling in the oracle case is $n^{\alpha_X/2}$, given that, in particular, $hn^{1-\alpha_X} \to \infty$. The reason for this discrepancy is that the normalization in [15, (1.4)] is $1/(nhf_n(x))$, instead of $1/(nh\hat{f}_h(x))$, where f_n is the normal density with estimated mean and variance. (This effect was also mentioned in [29]).

To deal with the non-oracle case, let us consider the following set of additional conditions:

(H1) $\sqrt{nh}(b^4 + 1/nb) \to 0$.
(H1a) $n^{\alpha_\varepsilon}(b^4 + 1/nb) \to 0$.
(H2) $\sqrt{nh}(b^2 + bh)n^{-\alpha_\varepsilon/2} \to 0$.
(H2a) $n^{\alpha_\varepsilon}(b^2 + bh)n^{-\alpha_\varepsilon/2} \to 0$.

Theorem 3.5. *Assume* (P2) *and* (E2) *and that the conditions* (D1)-(D3) *and* (H0) *hold.*

- *If* (3.3) *and* (H1)+(H2) *hold, then* (3.4) *holds.*
- *If* $hn^{(1-2\alpha_\varepsilon)} \to \infty$ *and* (H1a)+(H2a)*, then*

$$n^{\alpha_\varepsilon}\left(\hat{\sigma}^2(x) - \sigma^2(x)\right) \xrightarrow{d} C_2\sigma^2(x)H_2 - C_1^2\sigma^2(x)\chi^2(1), \tag{3.6}$$

where $\chi^2(1)$ *is* χ^2 *random variable with 1 degree of freedom,* H_2 *is the Hermite-Rosenblatt random variable defined in* (3.1) *and* C_1 *is defined in* (2.1).

Remark 3.6. In Theorem 3.5 as well as in Proposition 3.1, the results under (P1) and/or (E1) can be concluded by plug-ing in $\alpha_X = 1$ and/or $\alpha_\varepsilon = 1$, respectively.

Remark 3.7. The results of Proposition 3.1 and Theorem 3.5 can be formulated in a multivariate set-up. In case of (3.4) the limiting distribution of $(\hat{\sigma}^2(x_i) - \sigma^2(x_i), i = 1, \ldots, m)$ is asymptotically multivariate normal with independent components (this follows from the Cramer-Wold device). In case of (3.5) the limiting distribution is degenerate,

$$C_2(\sigma^2(x_i), i = 1, \ldots, m)H_2.$$

In case of (3.6) the limiting distribution is also degenerate.

Remark 3.8. Condition (H1) is the standard condition in the weakly dependent situation, see e.g. the proof of Theorem 8.5 in [11]. Without (H1), the estimator becomes inconsistent. This is intuitively clear: since the bias of $\hat{m}_b(x) - m(x)$ is $O(b^2)$, its contribution to $\hat{\sigma}_h^2(x)$ is $O(b^4)(nh\hat{f}_h(x))^{-1}\sum_{i=1}^n K_h(x - X_i)$ which is of order b^4. Since the term is scaled by $\sqrt{nh}$, the condition is sharp. Conditions (H1a) and (H1b) are versions of (H1), when a different scaling is applied. In fact, they are used in Theorem 3.5 in conjunction with $hn^{1-2\alpha_\varepsilon} \to \infty$ and $h^5n^{1-\alpha_X} \to \infty$, respectively, so that they become weaker than (H1). Note further, that if $h = b$ and $hn^{1-2\alpha_\varepsilon} \to \infty$, the condition $h^4n^{\alpha_\varepsilon} \to 0$ is automatically fulfilled.

Furthermore, under (P1), the condition (H2) may be replaced with the weaker one:

(H3) $\sqrt{nh}(b^3 + b^2h)n^{-\alpha_\varepsilon/2} \to 0$.

Also, let us note that in the conditions above b^2 may be replaced with b^4, if we adopt the following jackknife-type bias correction (see e.g. [37]):

$$\hat{m}_b^*(x) = 2\hat{m}_b(x) - \hat{m}_{\sqrt{2}b}(x).$$

Remark 3.9. We compare our results with [15]. First, the authors considered the parametric mean case with Gaussian predictors. Second, the authors assumed that $n^{1-\alpha_X}h(\ln n)^{-1} \to \infty$ if $\alpha_X \in (1/2, 1)$ and $n^{\alpha_X}h \to \infty$ when $\alpha_X \in (0, 1/2)$. Furthermore, they assumed $\alpha_\varepsilon > \alpha_X/2$ in conjunction with $n^{\alpha_X/2}h^2 \to 0$ to get $n^{\alpha_X/2}$ rates of convergence. This rate of convergence does not appear in our context, since we estimate the density of the predictors in a nonparametric way. On the other hand, if $\alpha_\varepsilon < \alpha_X/2$ and $n^{\alpha_\varepsilon/2}h \to 0$, then the authors obtained n^{α_ε} rate of convergence, the same as in our case. Let us further note, that if $\alpha_X = 1$, then they automatically assume in the latter case that $\alpha_\varepsilon < 1/2$, so that they exclude a wide range of memory parameters. In particular, they cannot obtain the small bandwidth asymptotics using their method.

3.2. Error density estimation

In this section we assume for simplicity that $\sigma(\cdot) \equiv 1$.

Theorem 3.10. *Consider the model* (1.1)*. Assume* (P2) *and* (E2) *and that conditions* (D1)-(D2) *hold. Let* $f'_\varepsilon(x) \neq 0$.

- *If* $hn^{1-\alpha_\varepsilon} \to 0$, $n^{\alpha_\varepsilon}h \to \infty$, $nh^5 \to 0$, *then*

$$(nh)^{1/2}\left(\hat{f}_{h,\Delta}(x) - f_\varepsilon(x)\right) \xrightarrow{\text{d}} \mathcal{N}\left(0, f_\varepsilon(x)\int K^2(u)du\right). \tag{3.7}$$

- *If* $hn^{1-\alpha_\varepsilon} \to \infty$, $nh^3 \to \infty$, $nh^5 \to 0$, *then*

$$n^{\alpha_\varepsilon/2}\left(\hat{f}_{h,\Delta}(x) - f_\varepsilon(x)\right) \xrightarrow{\text{p}} 0. \tag{3.8}$$

Let $\hat{f}_{h,\varepsilon}(x) = \frac{1}{nh}\sum_{i=1}^n K_h(x-\varepsilon_i)$ be the standard Parzen-Rosenblatt estimator based on ε_i, $i = 1, \ldots, n$. Then, assuming $nh^5 \to 0$ (i.e. imposing negligibility of the bias), we have the following smoothing dichotomy (see [36]):

- If $hn^{1-\alpha_\varepsilon} \to 0$, then

$$(nh)^{1/2}\left(\hat{f}_{h,\varepsilon}(x) - f_\varepsilon(x)\right) \xrightarrow{\text{d}} \mathcal{N}\left(0, f_\varepsilon(x)\int K^2(u)du\right). \tag{3.9}$$

- If $hn^{1-\alpha_\varepsilon} \to \infty$, then

$$n^{\alpha_\varepsilon/2}\left(\hat{f}_{h,\varepsilon}(x) - f_\varepsilon(x)\right) \xrightarrow{\text{d}} \mathcal{N}\left(0, C_1^2(f'_\varepsilon(x))^2\right). \tag{3.10}$$

Combining (6.39) with (3.9), yields (3.7). Now, let us note that the conclusion (3.10) is obtained by approximating

$$\left(\hat{f}_{h,\varepsilon}(x) - f_\varepsilon(x)\right) \approx -f'_\varepsilon(x)\frac{1}{n}\sum_{j=1}^n \varepsilon_j,$$

see [36, p.1454]. Combining this with (6.39) we obtain (3.8). In particular, the residual-based estimator converges faster than the errors-based one. We conjecture that for $\alpha_\varepsilon < 1/2$,

$$n^{\alpha_\varepsilon}\left(\hat{f}_{h,\Delta}(x) - f_\varepsilon(x)\right) \xrightarrow{\mathrm{d}} CH_2,$$

where C is a constant and H_2 is the Hermite-Rosenblatt distribution.

Such behaviour is somehow counterintuitive. However, a similar phenomenon occurs for empirical processes of residuals or empirical processes with estimated parameters, see [4] and [23].

Remark 3.11. In the situation of (3.7) besides the condition $hn^{1-\alpha_\varepsilon} \to 0$, coming from the oracle behaviour of the kernel estimator (3.9), we have the additional constraint $hn^{\alpha_\varepsilon} \to \infty$. This additional constraint comes from (6.39). In particular, these two conditions cannot be fulfilled simultaneously if $\alpha_\varepsilon < 1/2$. In other words, unlike in the oracle case, if $\alpha_\varepsilon < 1/2$ then we cannot conclude $\sqrt{nh}$-type behaviour of the kernel density estimator. This is confirmed by simulations below.

3.3. How to avoid LRD behaviour: Bandwidth choice

To apply the theoretical results for the estimation of the conditional variance, one has to choose bandwidths b and h, as well as to verify if the chosen bandwidth h is *small* (i.e. (3.3) holds), or *large* (i.e. (3.5), (3.6), (3.8) hold).

1. Choice of b: There is little available on theoretical and practical properties of different bandwidth selectors under long memory. In case of regression with fixed-design and density estimation the reader is referred to [16] and [17], respectively. In [25] the authors studied the problem in the model (1.1) with $\sigma(\cdot) \equiv 1$. In particular, it was established in the latter article that for

$$\mathrm{MISE}_f(b) := \int \mathrm{E}\left[(\hat{m}_b(x) - m(x))^2\right] f(x)\, dx$$

we have

$$\mathrm{MISE}_f(b) \quad \sim \quad C\frac{1}{nb} + Cb^4 + Cn^{-\alpha_\varepsilon} + Cb^2 n^{-\alpha_\varepsilon}, \tag{3.11}$$

where C is a generic constant, different at each appearance. Therefore, the optimal bandwidth choice (according to the quadratic loss function) for the kernel estimation of $m(\cdot)$ is

$$b_{\mathrm{opt}} \sim \begin{cases} Cn^{-1/5} & \text{if } \alpha_\varepsilon > 2/5; \\ Cn^{-(1-\alpha_\varepsilon)/3} & \text{if } \alpha_\varepsilon < 2/5. \end{cases}$$

Also, it is proven there that the cross validation (CV) produces a valid approximation to b_{opt}, however, the mean squared error computed with the cross validation bandwidth provides a valid approximation to the optimal MISE only if $\alpha_\varepsilon > 4/5$. Indeed, let

$$\mathrm{CV}(h) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{m}_{i,b}(X_i))^2,$$

where $m_{i,b}(\cdot)$ is the kernel estimator of $m(\cdot)$, where the summation is taken over $j \neq i$ (summation over different sets does not influence the asymptotic results). It was proven in [25] that uniformly over $[B_1 b_{\mathrm{opt}}, B_2 b_{\mathrm{opt}}]$ (with some $B_1 < B_2$),

$$\mathrm{CV}(b) - \mathrm{MISE}_f(b) - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 \approx \frac{1}{n^2}\sum_{\substack{j,j'=1\\ j\neq j'}}^{n}\varepsilon_j\varepsilon_{j'}. \tag{3.12}$$

Combining this with (3.11) and noting that $\sum_{\substack{j,j'=1\\ j\neq j'}}^{n}\varepsilon_j\varepsilon_{j'} = O_P(n^{2-\alpha_\varepsilon})$ we have

$$\mathrm{CV}(b) \approx \mathrm{E}[\varepsilon_1^2] + C\frac{1}{nb} + Cb^4 + Cn^{-\alpha_\varepsilon}. \tag{3.13}$$

Thus, if $1/(nb) = o(n^{-\alpha_\varepsilon})$, then the cross-validation curve should not depend on b and becomes "flat". The latter condition means that $bn^{1-\alpha_\varepsilon} \to \infty$ and as indicated in Remark 3.3, this is exactly the large bandwidth condition for $\hat{m}_b(\cdot)$. In other words, from the CV curve we should be able to read out whether a particular bandwidth is *small* or *large*.
However, our numerical studies indicate terrible performance of the CV procedure. In other words, the theoretical approximation (3.13) has very poor finite sample properties. Therefore, we shall choose b by using the plug-in method as described in [33]. Note that the method is based on the approximation $b_{\mathrm{opt}} \approx Cn^{-1/5}$ which is not correct in the LRD setting for $\alpha_\varepsilon < 2/5$.

2. Choice of h: From (3.11) it is easy to conclude that in *the oracle case* the following asymptotic formula holds:

$$\widetilde{\mathrm{MISE}}_f(h) \quad \sim \quad C\frac{1}{nh} + Ch^4 + Cn^{-2\alpha_\varepsilon} + Ch^2 n^{-2\alpha_\varepsilon}, \tag{3.14}$$

where now $\widetilde{\mathrm{MISE}}_f(h)$ is defined in terms of $\sigma_h^2(\cdot)$ and $\sigma^2(\cdot)$. Therefore, we have the standard bandwidth choice $h_{\mathrm{opt}} \sim Cn^{-1/5}$ if $\alpha_\varepsilon > 1/5$. This means that the plug-in procedure is still applicable and our numerical studies in Section 4.1 indicate little influence of LRD on the plug-in selector.

3. LRD or i.i.d. zone: Let us note first that from the practical point of view "large" bandwidths are not desirable. Indeed, in the context of (3.6) it is completely not clear how to estimate percentiles of the limiting distribution, since e.g. bootstrap does not work (see [27, Chapter 10]). Also, our simulation studies suggest that the LRD parameter α_ε tends to be

overestimated, which may lead to confidence intervals with incorrect coverage probabilities. For example, in our simulation in Section 4.1 we find out that for $\alpha_\varepsilon = 0.4$ ($d_\varepsilon = (1-\alpha_\varepsilon)/2 = 0.3$ there), the median of an errors-based estimator is 0.4118, whereas for the residuals-based we have 0.4512. In other words, recalling the scaling n^{α_ε} in (3.6), the resulting residuals-based CI will be about 20% shorter than the errors-based one. Therefore, large-bandwidths asymptotics are *not practical* and may serve as a warning against inappropriate choice of h. In our implementation in Section 5 we apply the following procedure:

- Estimate h using the plug-in method. Solve $n^{-\delta} = h$. Compare the obtained δ with $1 - 2\alpha_\varepsilon$. Verify if (3.3) holds.

In this way neither the estimator of α_ε nor α_X is used to construct the confidence intervals, rather to justify the i.i.d.-type behaviour only.

4. Numerical studies

We illustrate our theoretical results by some numerical experiments. All codes and data sets are available from the authors.

Simulation procedure:

- We simulate $n = 1000$ observations from the models $Y_i = 0 + 2X_i + \sigma(X_i)\varepsilon_i$, and $Y_i = \sin(2\pi X_i/3) + \sigma(X_i)\varepsilon_i$, where $\sigma(x) = 1$ (homoscedastic case) or $\sigma(x) = \sqrt{x^2+1}$ (heteroscedastic case), the predictors X_i are i.i.d. Gaussian and the errors ε_i are Gaussian FARIMA(0,d_ε,0). Here, $d_\varepsilon = (1-\alpha_\varepsilon)/2 \in (0, 1/2)$. We use the `fracdiff` R-package [12].
- We estimate β_0 and β_1 by LSE estimators. We estimate $m(\cdot)$ using the Nadaraya-Watson estimator with b selected by the plug-in method described in Section 3.3 (We have used the package `locpoly`). Also, we compute $\hat{m}_b(\cdot)$ using $b = 0.3$ (i.e. oversmoothing).
- We compute residuals. In the heteroscedastic case we compute the Nadaraya-Watson estimator $\hat{\sigma}^2_{\hat{h}}(\cdot)$ with $\hat{h}$ selected by the plug-in method.
- We estimate the LRD parameter based on the errors ε_i and the residuals. We use the `fracdiff` package, which implements the maximal likelihood method.
- We estimate the residual density using (1.5), for different choices of h.
- This procedure is repeated $M = 500$ times.

4.1. Bandwidth choice

- Table 1 contains statistics for the bandwidth b selected by the plug-in method in the model $Y_i = \sin(2\pi X_i/3) + \sigma(X_i)\varepsilon_i$ in both, homoscedastic and heteroscedastic case. Plug-in tends to select slightly larger bandwidths if the memory grows.
- Table 2 contains statistics for the bandwidth h selected by the plug-in method in the models $Y_i = 2X_i + \sqrt{X_i^2+1}\varepsilon_i$ ($\hat{h}_1$), $Y_i = \sin(2\pi X_i/3) +$

Table 1
Bandwidth choice for $m(\cdot)$

	Homoscedastic					Heteroscedastic				
d_ε	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$
0	0.1196	0.1596	0.1664	0.1739	0.1829	0.1223	0.1656	0.1790	0.1915	0.2073
0.1	0.1293	0.1620	0.1687	0.1741	0.1839	0.1270	0.1650	0.1779	0.1898	0.2076
0.2	0.1212	0.1623	0.1692	0.1767	0.1877	0.1347	0.1672	0.1808	0.1918	0.2135
0.3	0.1221	0.1656	0.1748	0.1828	0.1939	0.1276	0.1672	0.1824	0.1970	0.2182
0.4	0.1203	0.1719	0.1827	0.1918	0.2068	0.1270	0.1735	0.1909	0.2054	0.2319

Table 2
Bandwidth choice for $\sigma(\cdot)$

d_ε		$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$
0	$\hat{h}_1$	0.1108	0.1725	0.2346	0.2639	0.2960
	$\hat{h}_2$	0.1105	0.1749	0.2352	0.2653	0.3000
	$\hat{h}_3$	0.1153	0.1757	0.2347	0.2673	0.3029
0.1	$\hat{h}_1$	0.1124	0.1725	0.2336	0.2660	0.2990
	$\hat{h}_2$	0.1095	0.1738	0.2367	0.2683	0.3016
	$\hat{h}_3$	0.1072	0.1712	0.2368	0.2699	0.3042
0.2	$\hat{h}_1$	0.1014	0.1828	0.2332	0.2666	0.3037
	$\hat{h}_2$	0.1112	0.1909	0.2392	0.2689	0.3060
	$\hat{h}_3$	0.1101	0.1823	0.2368	0.2688	0.3081
0.3	$\hat{h}_1$	0.1141	0.1750	0.2318	0.2633	0.2980
	$\hat{h}_2$	0.1176	0.1842	0.2344	0.2664	0.3013
	$\hat{h}_3$	0.1153	0.1801	0.2340	0.2641	0.2998
0.4	$\hat{h}_1$	0.1025	0.1754	0.2307	0.2613	0.2952
	$\hat{h}_2$	0.1115	0.1807	0.2378	0.2656	0.3002
	$\hat{h}_3$	0.1149	0.1772	0.2370	0.2663	0.2970

$\sqrt{X_i^2+1}\varepsilon_i$ with $\hat{m}_{\hat{b}}(\cdot)$, where $\hat{b}$ is the plug-in bandwidth ($\hat{h}_2$), and the latter model with $\hat{m}_{0.3}(\cdot)$ ($\hat{h}_3$). There is little influence of the memory parameter as well as the type of the conditional mean.

4.2. Estimation of LRD parameter

- Table 3 contains statistics for the memory parameter d_ε. The estimator $\hat{d}_0$ is based on the errors $\sigma(X_i)\varepsilon_i$ for $\sigma(\cdot) \equiv 1$ and $\sigma(x) = \sqrt{x^2+1}$. Note the little difference between the homoscedastic and heteroscedastic case. Also, the variability remains the same as the memory increases. The remaining estimators are residuals-based: $\hat{d}_1$ in the model $Y_i = 2X_i + \sqrt{X_i^2+1}\varepsilon_i$; $\hat{d}_2$ in the model $Y_i = \sin(2\pi X_i/3) + \sqrt{X_i^2+1}\varepsilon_i$ with $\hat{m}_{\hat{b}}(\cdot)$, where $\hat{b}$ is the plug-in bandwidth; and ($\hat{d}_3$) in the latter model with $\hat{m}_{0.3}(\cdot)$. Note that the variability is similar across different LRD parameters and different estimators. However, in the heteroscedastic case, the residuals-based estimator tends to underestimate d_ε, especially if the conditional mean is oversmoothed.

Table 3
Estimation of LRD parameter

		Homoscedastic					Heteroscedastic				
d_ε		$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$
0.1	$\hat{d}_0$	0.0505	0.0770	0.0944	0.1096	0.1365	0.0551	0.0755	0.0941	0.1104	0.1351
	$\hat{d}_1$	0.0514	0.0774	0.0942	0.1098	0.1362	0.0482	0.0698	0.0870	0.1040	0.1278
	$\hat{d}_2$	0.0498	0.0767	0.0921	0.1081	0.1351	0.0468	0.0702	0.0868	0.1037	0.1289
	$\hat{d}_3$	0.0480	0.0756	0.0914	0.1075	0.1321	0.0478	0.0699	0.0862	0.1021	0.1279
0.2	$\hat{d}_0$	0.1554	0.1785	0.1935	0.2122	0.2373	0.1567	0.1773	0.1958	0.2121	0.2387
	$\hat{d}_1$	0.1554	0.1784	0.1936	0.2119	0.2377	0.1447	0.1639	0.1802	0.1980	0.2240
	$\hat{d}_2$	0.1541	0.1753	0.1921	0.2089	0.2356	0.1441	0.1643	0.1806	0.1977	0.2224
	$\hat{d}_3$	0.1503	0.1726	0.1891	0.2084	0.2313	0.1444	0.1630	0.1783	0.1965	0.2211
0.3	$\hat{d}_0$	0.2522	0.2767	0.2957	0.3130	0.3347	0.2495	0.2775	0.2941	0.3127	0.3357
	$\hat{d}_1$	0.2523	0.2765	0.2956	0.3125	0.3342	0.2281	0.2568	0.2729	0.2901	0.3124
	$\hat{d}_2$	0.2508	0.2747	0.2928	0.3094	0.3325	0.2282	0.2573	0.2739	0.2903	0.3127
	$\hat{d}_3$	0.2482	0.2708	0.2905	0.3056	0.3292	0.2270	0.2562	0.2725	0.2889	0.3116
0.4	$\hat{d}_0$	0.3508	0.3775	0.3950	0.4096	0.4288	0.3499	0.3761	0.3938	0.4098	0.4316
	$\hat{d}_1$	0.3508	0.3771	0.3948	0.4094	0.4284	0.3075	0.3356	0.3549	0.3721	0.3931
	$\hat{d}_2$	0.3461	0.3729	0.3896	0.4050	0.4246	0.3203	0.3453	0.3618	0.3767	0.4015
	$\hat{d}_3$	0.3427	0.3696	0.3863	0.4010	0.4220	0.3199	0.3426	0.3610	0.3754	0.3992

4.3. Error density

Here, we compare the mean square error (MSE) of the kernel estimators $\hat{f}_{h,\varepsilon}$ and $\hat{f}_{h,\Delta}$. We consider the homoscedastic model only with the parametric mean, i.e. $Y_i = 2X_i + \varepsilon_i$. We evaluate the squared errors $\mathrm{SE(errors)} = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_{h,\varepsilon}(u_i) - f(u_i))^2$ and $\mathrm{SE(residuals)} = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_{h,\Delta}(u_i) - f(u_i))^2$, where u_i, $i = 1,\ldots,n$, is an appropriately chosen deterministic grid. The estimator is evaluated for different values of h.

We note that if $d_\varepsilon = 0.1$, there is little difference between the errors-based and the residuals-based estimators if $m(x) = \beta_0 + \beta_1 x$, which is in line with results for weakly dependent random variables. The improvement is clearly visible for $d_\varepsilon = 0.4$.

4.4. Pathwise interpretation

Here, we explain heuristically why the estimation of the density may lead to better results when we use residuals instead of errors. We simulate just one sample from the parametric model $Y_i = 0 + 2 \cdot X_i + \varepsilon_i$ and plot the corresponding graphs for $d_\varepsilon = 0.4$, $h = 0.15$ and $h = 0.35$. According to Table 4, the latter bandwidth is close to optimal.

The results are displayed on Figure 1. We observe a big difference between errors-based and residuals-based density estimators. In fact, the errors-based estimator looks like *shifted to the left*. The reason for this is that LRD behaviour of errors leads to a poor performance of $\hat{f}_{h,\varepsilon}$. At the same time this LRD behaviour leads to a poor estimation of the intercept ($\hat{\beta}_0 = -0.6053$ in this case). However, when computing the residuals these two effects cancel out.

Table 4
MSE for kernel density estimator: $d_X = 0$

h	MSE(errors)	MSE(res)	MSE(errors)	MSE(res)
	$d_\varepsilon = 0.1$	$d_\varepsilon = 0.1$	$d_\varepsilon = 0.4$	$d_\varepsilon = 0.4$
0.05	0.0014	0.0013	0.0081	0.0024
0.10	0.0007	0.0006	0.0074	0.0017
0.20	0.0003	0.0003	0.0071	0.0014
0.30	0.0003	0.0002	0.0070	0.0015
0.40	0.0003	0.0002	0.0070	0.0016
0.50	0.0005	0.0004	0.0070	0.0019
0.60	0.0007	0.0006	0.0070	0.0022
0.70	0.0011	0.0010	0.0071	0.0026
0.80	0.0015	0.0014	0.0073	0.0031
0.90	0.0021	0.0020	0.0076	0.0037
1.00	0.0027	0.0026	0.0079	0.0043

Table 5
Memory parameters for electricity prices

Data	d
$\log(NSW)$	0.38
$\log(NSWpeak)$	0.34
$\log(QLD)$	0.32
$\log(QLDpeak)$	0.30

5. Data analysis

We consider electricity prices in Australia, from two states: NSW (New South Wales) and QLD (Queensland). The data are available at

http://www.aemo.org.au/data/avg_price/avgp_month2009.shtm.

They describe average regional reference prices per region for each month (0000-2400) and average peak prices (peak period covers 7:00am to 10:00pm EST weekdays excluding holidays) over the financial year.

Let us note first that these data have a completely different pattern than *typical* (i.e. stock prices, stock indices, exchange rates) financial data. Those financial time series have two patterns: they are uncorrelated, but squares have long memory, or they follow the unit root model, so that they are weakly dependent after differentiation.

In case of energy prices data several authors argued that differentiation leads to antipersistence. In particular, the original data are not stationary.

In our situation the *raw data* seem to follow a stationary model, see Figure 2. We estimate the memory parameter as indicated in Table 5.

To exclude a possible spikes effect on the long memory behaviour, we have considered the first 95 observations in each data set and computed log-prices. In fact, the memory parameters of NSW and QLD data remain almost the same.

- We set $X_i = \log \mathrm{NSW}_i$, $Y_i = \log \mathrm{QLD}_i$, $i = 1, \ldots, 95$. The estimated memory parameters for X_i and Y_i are, respectively, $\alpha_X = 0.36$ and $\alpha_\varepsilon = 0.26$.

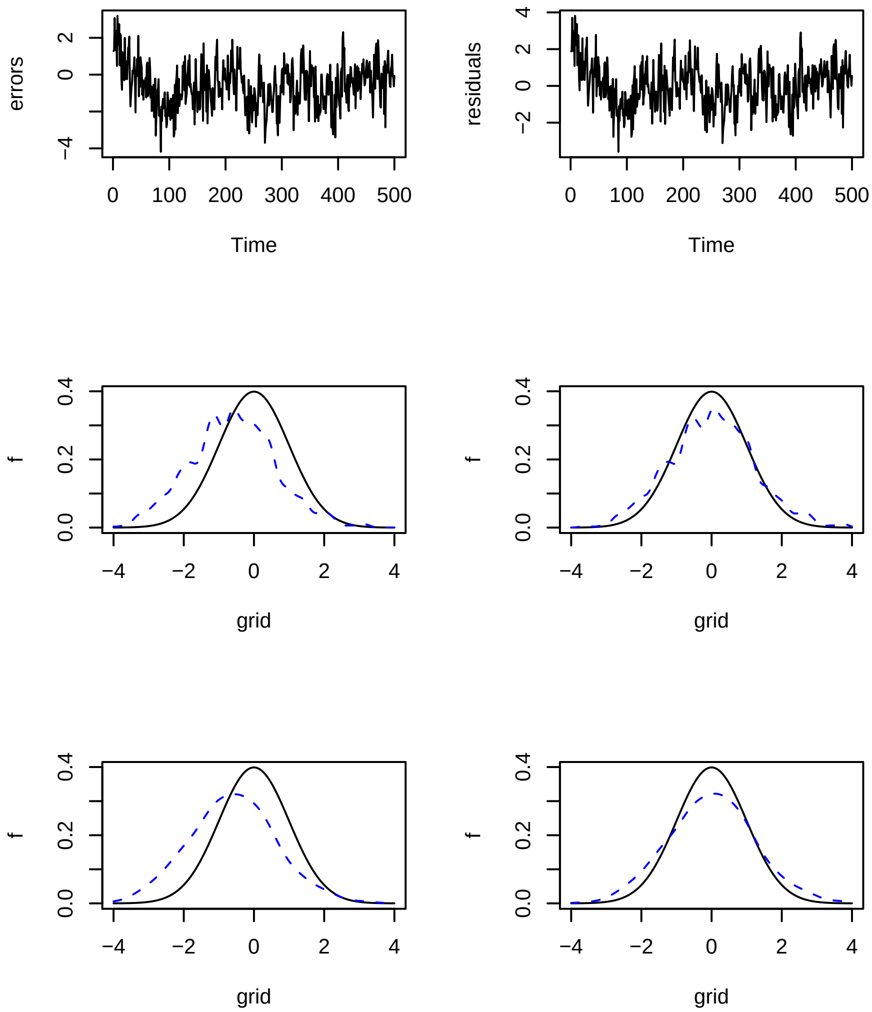

FIG 1. *Top line: simulated LRD errors (left panel;* $d_\varepsilon = 0.4$*) and corresponding residuals obtained by fitting the conditional mean. The dependence parameter for predictors is* $d_X = 0$*. Kernel estimators: middle line -* $h = 0.15$*, bottom line -* $h = 0.35$*; the solid line - true density, the dashed line - the kernel estimator; left panel - the kernel estimator based on errors, right panel - the kernel estimator based on residuals.*

- The histogram of log-prices indicates that the predictors are non-normal, i.e. results of [15] are not applicable (see Figure 3).
- We fitted both, the linear and the non-linear model.
- Following our discussion in Section 3.3, the plug-in method yields $b = 0.17$. A scatter plot (Figure 4) indicates that there is not too much difference between the linear and the non-linear model.
- We have computed two sets residuals for both models and estimated h using the plug-in method. We obtain $h = 0.18$ (linear case) and $h = 0.17$ (nonparametric one).

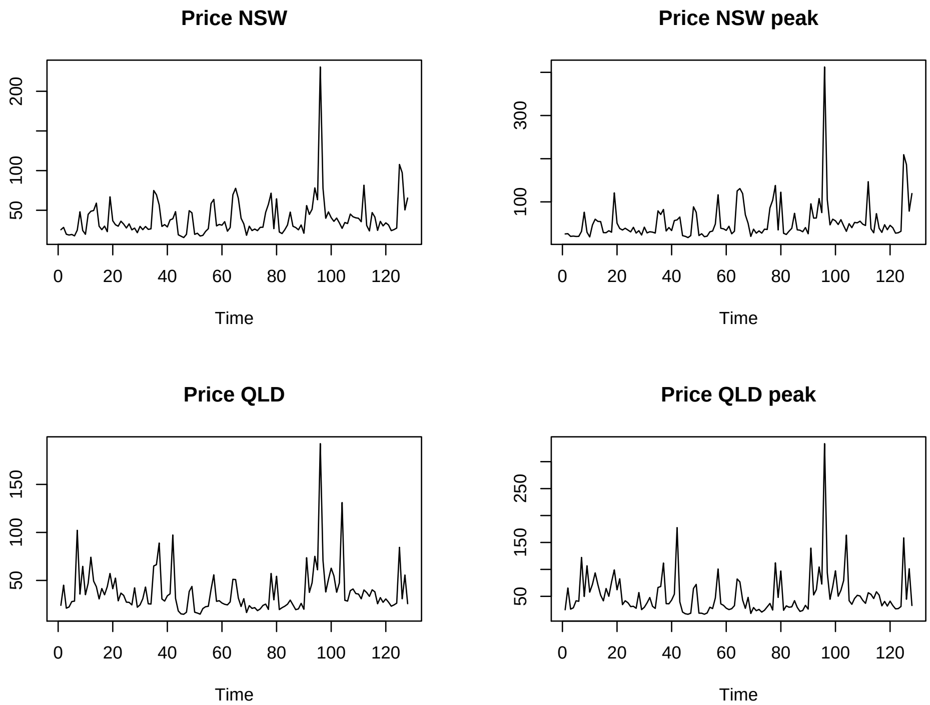

FIG 2. *Energy prices in NSW and QLD*

- The estimated LRD parameters of the residuals are, respectively, $\alpha_\varepsilon = 0.5$ and $\alpha_\varepsilon = 0.52$.
- We solve $n^{-\delta} = 0.18$ to get $\delta = 0.37$. We verify that $hn^{1-2\alpha_\varepsilon} \to 0$. Also, with $\delta = 0.37$, $\alpha_X = 0.26$ and $\alpha_\varepsilon = 0.5$ we verify that $h^2 n^{-\alpha_X/2} = o(1/\sqrt{nh} + n^{-\alpha_\varepsilon})$. In other words, i.i.d.-type asymptotics (3.6) for $\sigma^2(\cdot)$ is permitted.
- The normal reference rule bandwidth for the error density estimation yields $h = 0.14$. Thus, $\delta = 0.43$ and with $\alpha_\varepsilon \approx 0.5$ the condition $hn^{1-\alpha_\varepsilon} \to 0$ does not hold, i.e. i.i.d.-type asymptotics (3.7) is not permitted.

6. Technical details

We refer to the Appendices for technical results involving partial sums of LRD random variables, covariance bounds for long memory linear sequences and some asymptotic expressions for kernel functions.

Let

$$\Delta_i = \Delta_{i,b} = \hat{m}_b(X_i) - m(X_i) = R_b(X_i) + \frac{1}{nb\hat{f}_b(X_i)} \sum_{j=1}^{n} \sigma(X_j) K_b(X_i - X_j)\varepsilon_j, \tag{6.1}$$

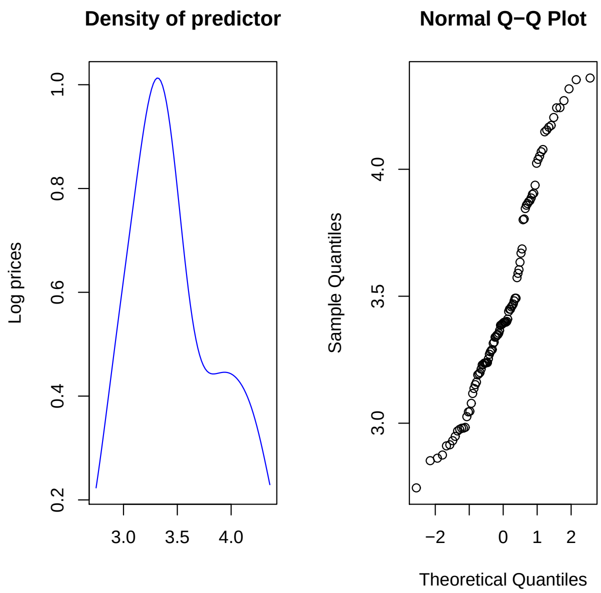

FIG 3. *Kernel estimate of the density of predictors with the plug-in bandwidth b=0.17 and QQ plot of log(NSW). Normality of predictors is rejected.*

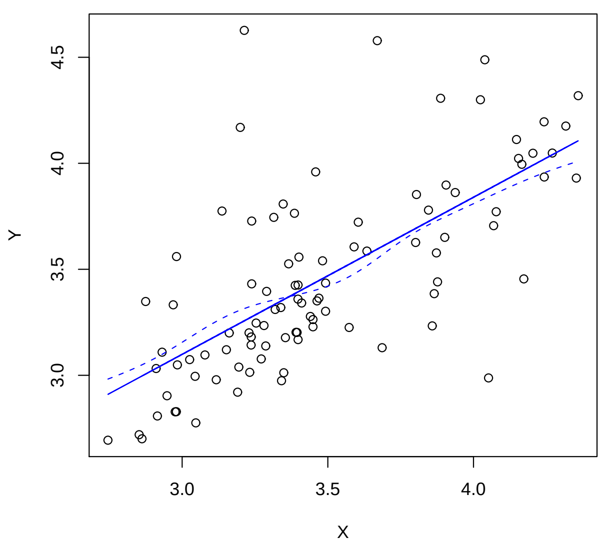

FIG 4. *Conditional mean fitting: dashed curve - nonparametric fitting with b=0.17; solid curve - parametric fitting.*

where $R_b(y) := R_b(y; m(\cdot))$ and for a given function g,

$$R_b(y; g(\cdot)) = \frac{1}{nb\hat{f}_b(y)} \sum_{j=1}^{n} (g(X_j) - g(y)) K_b(y - X_j). \tag{6.2}$$

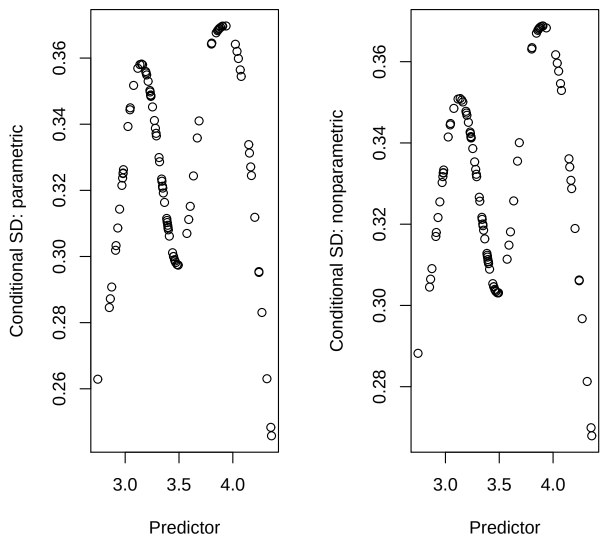

FIG 5. *Conditional standard deviation fitting: with h=0.2 (parametric case) and h=0.3 (nonparametric case).*

Denote $\rho(y;g(\cdot)) = (gf)''(y) - g(y)f''(y)$, $\rho(y) := \rho(y;f(\cdot))$. Then uniformly over $\{y : f(y) > 0\}$,

$$R_b(y) = R_b(y;m(\cdot)) - \frac{b^2\kappa_2}{2}\frac{\rho(y)}{f(y)} = O(b^4(1+o_P(1))). \tag{6.3}$$

6.1. Decomposition of the conditional variance estimator

We have

$$\begin{aligned}\left\{\hat\sigma^2(x) - \sigma^2(x)\right\} &= \left(\frac{1}{nh\hat f_h(x)}\sum_{i=1}^n \sigma^2(X_i)K_h(x-X_i) - \sigma^2(x)\right)\\ &+\frac{1}{nh\hat f_h(x)}\sum_{i=1}^n \sigma^2(X_i)K_h(x-X_i)\left(\varepsilon_i^2 - 1\right)\\ &-\frac{2}{nh\hat f_h(x)}\sum_{i=1}^n \Delta_i\sigma(X_i)K_h(x-X_i)\varepsilon_i\\ &+\frac{1}{nh\hat f_h(x)}\sum_{i=1}^n \Delta_i^2K_h(x-X_i) =: J_1 + J_2 - J_3 + J_4. \end{aligned} \tag{6.4}$$

For these four terms we obtain the following asymptotic behaviour. For the proofs of the lemmas see below.

Lemma 6.1. *Under the conditions of Proposition 3.1 we have*

$$\sqrt{nh}J_1 = o_P(1)\ . \tag{6.5}$$

Lemma 6.2. *Suppose the assumptions of Proposition 3.1 hold.*

- *If $hn^{(1-2\alpha_\varepsilon)} \to 0$ then*

$$\sqrt{nh\hat{f}_h(x)}J_2 \xrightarrow{d} \mathcal{N}\left(0, \sigma^4(x)\mathrm{E}[(\varepsilon_1^2 - 1)^2]\int K^2(u)\,du\right). \tag{6.6}$$

- *If $hn^{(1-2\alpha_\varepsilon)} \to \infty$ then*

$$n^{\alpha_\varepsilon}J_2 \xrightarrow{d} C_2\sigma^2(x)H_2. \tag{6.7}$$

Note that if $m(\cdot)$ is known, then $\Delta_i \equiv 0$, so that the limiting behavior of the oracle estimator of $\sigma^2(\cdot)$ is determined by that of J_1 and J_2. Therefore, Proposition 3.1 follows directly from Lemmas 6.1 and 6.2.

The next two lemmas are needed in the non-oracle case only.

Lemma 6.3. *Under the conditions of Theorem 3.5 we have*

$$n^{\alpha_\varepsilon}J_3 \xrightarrow{d} 2C_1^2\sigma^2(x) \times \chi^2(1). \tag{6.8}$$

Lemma 6.4. *Under the conditions of Theorem 3.5 we have*

$$n^{\alpha_\varepsilon}J_4 \xrightarrow{d} C_1^2\sigma^2(x) \times \chi^2(1). \tag{6.9}$$

However, from Lemmas 6.3 and 6.4 one cannot conclude yet the asymptotics of Theorem 3.5. One needs a precise expansions for the terms J_3 and J_4, as presented in (6.17) and (6.18).

6.2. Proof of Proposition 3.1

Proposition 3.1 will be justified if we prove Lemmas 6.1 and 6.2.

6.2.1. Proof of Lemma 6.1

Write J_1 as

$$J_1 = \left(\frac{1}{\hat{f}_h(x)} - \frac{1}{f(x)}\right)\frac{1}{nh}\sum_{i=1}^{n}\left(\sigma^2(X_i) - \sigma^2(x)\right)K_h(x - X_i)$$
$$+\frac{1}{nhf(x)}\sum_{i=1}^{n}\left(\sigma^2(X_i) - \sigma^2(x)\right)K_h(x - X_i) := \left(\frac{1}{\hat{f}_h(x)} - \frac{1}{f(x)}\right)J_{11} + J_{11}.$$

By Lemma [37, Lemma 2(i)], $\hat{f}_h$ is the consistent estimator of f. Therefore, the first part in the expression above is dominated by the second one. Since (D1) holds, we have $\mathrm{E}|J_{11}| = O(h^2)$. Therefore, $\sqrt{nh}J_1 = o_P(1)$ by (H0). □

6.2.2. Proof of Lemma 6.2

Since $\hat{f}_h$ is a consistent estimator of f, to deal with J_2, we just consider $\hat{f}_h$ replaced by f.

Let $\mathcal{F}_i =: \mathcal{H}_i \vee \mathcal{X}_i$. Let $\xi_i = \varepsilon_i^2 - 1$, and decompose

$$\frac{1}{nhf(x)} \sum_{i=1}^{n} \left\{ \sigma^2(X_i) K_h(x - X_i)\xi_i - \mathrm{E}[\sigma^2(X_i)K_h(x - X_i)\xi_i|\mathcal{F}_{i-1}] \right\} + \frac{1}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[\sigma^2(X_i)K_h(x - X_i)|\mathcal{X}_{i-1}]\mathrm{E}[\xi_i|\mathcal{H}_{i-1}] =: J_{21} + J_{22} \quad (6.10)$$

It is shown below, that either J_{21} or J_{22} dominates, according to different assumptions on h and α_ε, however, long memory of the predictors does not play any role.

Furthermore, using the martingale CLT, we can easily check (see Appendix C) that

$$\sqrt{nhf(x)} J_{21} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sigma^4(x)\mathrm{E}[(\varepsilon_1^2 - 1)^2] \int K^2(u)\, du\right). \quad (6.11)$$

Moreover, under (P1)

$$J_{22} = \frac{\mathrm{E}[\sigma^2(X_1)K_h(x - X_1)]}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] = \sigma^2(x)\frac{\sigma_{n,\varepsilon}}{n}\frac{1}{\sigma_{n,\varepsilon}} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] + O(h)\frac{\sigma_{n,\varepsilon}}{n}\sigma_{n,\varepsilon}^{-1} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] =: J_{221} + O(h)J_{221}.$$

Recall that for any random variable V with a finite mean, we write $\bar{V} = V - \mathrm{E}[V]$. Under (P2) we write

$$\begin{aligned} J_{22} &= \frac{\mathrm{E}[\sigma^2(X_1)K_h(x - X_1)]}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] \\ &\quad + \frac{1}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}]\overline{\mathrm{E}[\sigma^2(X_i)K_h(x - X_i)|\mathcal{X}_{i-1}]} \\ &=: J_{221} + O(h)J_{221} + J_{222}. \end{aligned}$$

Remark 6.5. In both representations for J_{22}, all the terms will be negligible w.r.t. J_{221}. Furthermore, in the latter term only LRD of the errors plays a role. This will be a common feature of the decompositions for J_3 and J_4 below. We will call this *LE/N* decomposition (LRD of errors part + negligible part).

We can easily verify that for a fixed x,

$$\begin{aligned}
&\mathrm{E}[\sigma^2(X_1)K_h(x-X_1)|\mathcal{X}_0] = \mathrm{E}[\sigma^2(X_{1,0}+\zeta_1)K_h(x-(X_{1,0}+\zeta_1))|\mathcal{X}_0] \\
&= \int \sigma^2(X_{1,0}+u)K_h(x-(X_{1,0}+u))f_\zeta(u)\,du \\
&= h\int \sigma^2(x-vh)K(v)f_\zeta(x-vh-X_{1,0})dv \\
&\le \|f_\zeta\|_\infty h\int \sigma^2(x-vh)K(v)dv = O\left(h\sigma^2(x)\right) = O(h),
\end{aligned}$$

so that $\mathrm{Var}\left(\mathrm{E}[\sigma^2(X_1)K_h(x-X_1)|\mathcal{X}_0]\right) = O(h^2)$. Therefore, via (B.1) and (A.8) in the Appendix,

$$\begin{aligned}
\mathrm{Var}(J_{222}) &= O(1)\frac{nh^2}{(nh)^2f^2(x)} + O(1)\frac{h^2\sigma^4(x)}{(nh)^2f^2(x)}\sum_{\substack{i,j=1\\ i\neq j}}^{n}|j-i|^{-\alpha_X/2}\mathrm{Cov}^2(\varepsilon_i,\varepsilon_j) \\
&= \frac{O(1)}{n} + \frac{O(1)}{n^2f^2(x)}\sigma^4(x)\sum_{\substack{i,j=1\\ i\neq j}}^{n}|j-i|^{-(\alpha_X/2+2\alpha_\varepsilon)} = O(n^{-(\alpha_X/2+2\alpha_\varepsilon)}\vee n^{-1}).
\end{aligned}$$

Using (A.11) we obtain that (6.7) holds for J_{221} given that $\alpha_\varepsilon < 1/2$ and $J_{221} = O_P(n^{-1/2})$ if $\alpha_\varepsilon > 1/2$. Furthermore, J_{222} is either $O_P(n^{-1/2})$ or of a smaller order than J_{221}.

Now, in (6.10), either J_{21} or J_{22} dominates, according to the respective assumptions on h.

6.3. Hoeffding decomposition of weighted quadratic forms under (P1) and its consequences

In this section we will work under the condition (P1), so that $X_i = X_i^*$, $i\ge 1$, are i.i.d. Let $T(X_i^*,X_j^*)$ be a measurable, real-valued function defined on X_i^*,X_j^* such that appropriate moment conditions are satisfied. Recall that $\overline{T(X_i^*,X_j^*)} = T(X_i^*,X_j^*) - \mathrm{E}[T(X_j^*,X_j^*)]$ and define for $i\neq k$,

$$Z_{1,i} = \mathrm{E}\left[\overline{T(X_i^*,X_k)}|X_i^*\right], \qquad Z_{2,i} = \mathrm{E}\left[\overline{T(X_k,X_i^*)}|X_i^*\right].$$

Note that $Z_{1,i}$ and $Z_{2,i}$ do not depend on k and $Z_{1,i}, i\ge 1$, are uncorrelated and centered. The same applies to $Z_{2,i}, i\ge 1$. Furthermore, denote

$$\underline{Z}_{i,j} = \overline{T(X_i^*,X_j^*)} - Z_{1,i} - Z_{2,j}.$$

Let G_1 and G_2 be two measurable functions and consider

$$\frac{1}{t_n}\sum_{i,j=1}^{n} T(X_i^*,X_j^*)G_1(\varepsilon_i)G_2(\varepsilon_j),$$

where t_n is a sequence of real numbers. To this term we apply the following Hoeffding-type decomposition:

$$\frac{1}{t_n}\sum_{i=1}^{n} G_1(\varepsilon_i)G_2(\varepsilon_i)T(X_i^*,X_i^*) + \frac{1}{t_n}\sum_{\substack{i,j=1\\ i\neq j}}^{n} G_1(\varepsilon_i)G_2(\varepsilon_j)\mathrm{E}\left[T(X_i^*,X_j^*)\right] + \\ +\frac{1}{t_n}\sum_{\substack{i,j=1\\ i\neq j}}^{n} G_1(\varepsilon_i)G_2(\varepsilon_j)(Z_{1,i}+Z_{2,j}) + \frac{1}{t_n}\sum_{\substack{i,j=1\\ i\neq j}}^{n} G_1(\varepsilon_i)G_2(\varepsilon_j)\underline{Z}_{i,j} \\ =: \quad B_1+B_2+B_3+B_4. \tag{6.12}$$

We have

$$B_1 = O_P\left(\frac{1}{t_n} n\mathrm{E}[T(X_1^*,X_1^*)]\mathrm{E}[G_1(\varepsilon_1)G_2(\varepsilon_1)]\right), \tag{6.13}$$

$$B_2 = \frac{1}{t_n}\mathrm{E}[T(X_1^*,X_2^*)]\left\{\sum_{i=1}^{n} G_1(\varepsilon_i)\sum_{j=1}^{n} G_2(\varepsilon_j) - \sum_{i=1}^{n} G_1(\varepsilon_i)G_2(\varepsilon_i)\right\}. \tag{6.14}$$

Next, since $Z_{1,i}$, $i\geq 1$, are uncorrelated and centered, we compute

$$\mathrm{Var}\Bigg(\sum_{\substack{j,i=1\\ j\neq i}}^{n} G_1(\varepsilon_i)G_2(\varepsilon_j)Z_{1,i}\Bigg) \\ = \sum_{\substack{i,j=1\\ j\neq i}}^{n} \mathrm{E}[G_1^2(\varepsilon_i)G_2^2(\varepsilon_j)]\mathrm{E}[Z_{1,i}^2] + \sum_{i=1}^{n}\sum_{\substack{j,j'=1\\ j'\neq j, j\neq i, j'\neq i}}^{n} \mathrm{E}[G_1^2(\varepsilon_i)G_2(\varepsilon_j)G_2(\varepsilon_{j'})]\mathrm{E}[Z_{1,i}^2] \\ = \mathrm{E}[Z_{1,1}^2]\Bigg(\sum_{\substack{i,j=1\\ i\neq j}}^{n} \mathrm{E}[G_1^2(\varepsilon_i)G_2^2(\varepsilon_j)] + \sum_{i=1}^{n}\sum_{\substack{j,j'=1\\ j'\neq j, j\neq i, j'\neq i}}^{n} \mathrm{E}[G_1^2(\varepsilon_i)G_2(\varepsilon_j)G_2(\varepsilon_{j'})]\Bigg). \tag{6.15}$$

A similar computation is valid if $Z_{1,i}$ are replaced with $Z_{2,j}$.

Finally, we deal with B_4. Define

$$Z_j^{\#} = G_2(\varepsilon_j)\sum_{i=1}^{j-1} G_1(\varepsilon_i)\underline{Z}_{i,j}\,, \qquad j\geq 2.$$

Then

$$\sum_{\substack{i,j=1\\ i<j}}^{n} G_1(\varepsilon_i)G_2(\varepsilon_j)\underline{Z}_{i,j} = \sum_{j=2}^{n} Z_j^{\#},$$

and note that to show negligibility of B_4, it suffices to consider the term above.

We will show that $Z_j^{\#}, j\geq 2$, are uncorrelated. To do this, recall that $\mathcal{F}_j = \mathcal{H}_j \vee \mathcal{X}_j$. Clearly, $Z_j^{\#}$ is $\mathcal{F}_j$-measurable. Recall that for $i\neq j$,

$$Z_{2,j} = \mathrm{E}[\overline{T(X_i^*,X_j^*)}|X_j^*] = \phi(X_j^*)$$

is independent of i. Here, ϕ is a measurable function such that $\mathrm{E}[\phi(X_j^*)] = 0$. Since X_j^*, $j \geq 1$, are independent, we also have $\mathrm{E}[\phi(X_j^*)|\mathcal{F}_{j-1}] = 0$. Thus,

$$\begin{aligned}
\mathrm{E}[Z_j^{\#}|\mathcal{F}_{j-1}] = & \sum_{i=1}^{j-1} \mathrm{E}\left[\overline{T(X_i^*, X_j^*)} G_1(\varepsilon_i) G_2(\varepsilon_j) | \mathcal{F}_{j-1}\right] \\
& - \sum_{i=1}^{j-1} \mathrm{E}\left[\mathrm{E}[\overline{T(X_i^*, X_j^*)}|X_i^*] G_1(\varepsilon_i) G_2(\varepsilon_j) | \mathcal{F}_{j-1}\right] \\
& - \sum_{i=1}^{j-1} \mathrm{E}\left[\mathrm{E}[\overline{T(X_i^*, X_j^*)}|X_j^*] G_1(\varepsilon_i) G_2(\varepsilon_j) | \mathcal{F}_{j-1}\right] \\
= & \sum_{i=1}^{j-1} \mathrm{E}\left[G_1(\varepsilon_i) G_2(\varepsilon_j) | \mathcal{F}_{j-1}\right] \left\{ \mathrm{E}\left[\overline{T(X_i^*, X_j^*)}|X_i^*\right] - \mathrm{E}\left[\overline{T(X_i^*, X_j^*)}|X_i^*\right] \right\} \\
& - \sum_{i=1}^{j-1} \mathrm{E}\left[G_1(\varepsilon_i) G_2(\varepsilon_j) | \mathcal{F}_{j-1}\right] \mathrm{E}\left[\phi(X_j^*)|\mathcal{F}_{j-1}\right] = 0.
\end{aligned}$$

In fact we showed that $(Z_j^{\#}, \mathcal{F}_j), j \geq 2$, is a martingale difference sequence. In particular,

$$\begin{aligned}
\mathrm{Var}\left(\sum_{j=2}^{n} Z_j^{\#}\right) &= \sum_{j=2}^{n} \mathrm{Var}\left(Z_j^{\#}\right) = \sum_{j=2}^{n} \mathrm{Var}\left(G_2(\varepsilon_j) \sum_{i=1}^{j-1} G_1(\varepsilon_i) \underline{Z}_{i,j}\right) \\
&= \sum_{j=2}^{n} \left\{ \sum_{i=1}^{j-1} \mathrm{E}\left[G_1^2(\varepsilon_i) G_2^2(\varepsilon_j)\right] \mathrm{E}[\underline{Z}_{i,j}^2] \right. \\
&\quad \left. + \sum_{\substack{i,i'=1 \\ i \neq i'}}^{j-1} \mathrm{E}[G_2^2(\varepsilon_j) G_1(\varepsilon_i) G_1(\varepsilon_{i'})] \mathrm{E}[\underline{Z}_{i,j} \underline{Z}_{i',j}] \right\}. \qquad (6.16)
\end{aligned}$$

□

6.4. Proof of Theorem 3.5

Let us start with some heuristic for J_3 and J_4. Assume for a moment that (P1) holds, so that $X_i = X_i^*$. Combining the definition of J_3 (see (6.4)) with the formula for Δ_i (cf. (6.1)) we may write

$$J_3 = \frac{2}{nh\hat{f}_h(x)} \sum_{i,j=1}^{n} \frac{2}{nb\hat{f}_b(X_i^*)} \sigma(X_i^*) \sigma(X_j^*) K_h(x - X_i^*) K_b(X_i^* - X_j^*) \varepsilon_i \varepsilon_j.$$

We will apply the Hoeffding decomposition and we will conclude that

$$J_3 \approx \frac{2\mathrm{E}\left[T_{h,b}(x, X_1^*, X_2^*)\right] \sigma_{n,\varepsilon}^2}{n^2 hbf(x)} \left\{ \left(\sigma_{n,\varepsilon}^{-1} \sum_{j=1}^{n} \varepsilon_j\right)^2 - \sigma_{n,\varepsilon}^{-2} \sum_{j=1}^{n} \varepsilon_j^2 \right\}, \qquad (6.17)$$

where $T(\cdot,\cdot)$ is a deterministic function defined in (6.22) below, such that $\mathrm{E}\left[T_{h,b}(x,X_1^*,X_2^*)\right] \sim \sigma^2(x)f(x)hb$. Therefore, the asymptotics for J_3 will follow from (A.10).

Likewise, again under (P1),

$$J_4 \approx \frac{\mathrm{E}\left[S_{h,b}(x,X_1^*,X_2^*)\right]\sigma_{n,\varepsilon}^2}{n^3hb^2f(x)}\left\{\left(\sigma_{n,\varepsilon}^{-1}\sum_{j=1}^n \varepsilon_j\right)^2 - \sigma_{n,\varepsilon}^{-2}\sum_{j=1}^n \varepsilon_j^2\right\}, \tag{6.18}$$

where S is a deterministic function defined in (6.37) below, such that $\mathrm{E}[S_{h,b}(x,X_1^*,X_2^*)] \sim \sigma^2(x)f(x)nhb^2$. Thus, the asymptotics for J_4 will also follow from (A.10).

Similar approximations to (6.17) and (6.18) are valid under (P2) as well. The asymptotics for $-J_3+J_4$ in the decomposition (6.4) will follow.

6.5. Proof of Lemma 6.3

Define

$$K_{h,b}(x,X_i,X_j) = K_h(x-X_i)K_b(X_i-X_j). \tag{6.19}$$

Then

$$\begin{aligned} J_3 &= \frac{2}{nh\hat{f}_h(x)}\sum_{i=1}^n R_b(X_i)\sigma(X_i)K_h(x-X_i)\varepsilon_i \\ &\quad + \frac{2}{nh\hat{f}_h(x)}\sum_{i,j=1}^n \frac{1}{nb\hat{f}_b(X_i)}\sigma(X_i)\sigma(X_j)K_{h,b}(x,X_i,X_j)\varepsilon_i\varepsilon_j =: J_{31}+J_{32}. \end{aligned}$$

In what follows, we will show that J_{32} can be written as $J_{322}+$remainder, where

$$n^{\alpha_\varepsilon}J_{322} \xrightarrow{\mathrm{d}} \left(2C_1^2\sigma^2(x)\right)\times\chi^2(1), \tag{6.20}$$

and the remainder is $o_P(1/\sqrt{nh}) + O_P(1/(nb))$. This will be done in Sections 6.5.1, 6.5.2 under (P1) and (P2), respectively.

Furthermore, for the term J_{31} we will show (see Section 6.5.3)

$$J_{31} = \frac{\kappa_2\rho(x)\sigma(x)}{f(x)}\frac{b^2}{n}\sum_{i=1}^n \varepsilon_i + O_P\left(\frac{b^2h^2}{n^{\alpha_\varepsilon/2}}\right) + O_P\left(\frac{b^2}{\sqrt{nh}}\right) + O_P\left(\frac{b^2}{n^{\alpha_\varepsilon/2+\alpha_X/4}}\right), \tag{6.21}$$

where the last term is present under (P2) only.

6.5.1. Behaviour of J_{32} under (P1)

Recall that under (P1) we denote $X_i = X_i^*$. To deal with J_{32}, we replace $\hat{f}_b(X_i)$ with $f(X_i)$. This is allowed in view of the consistency of $\hat{f}_b$ and the finite support of $K(\cdot)$.

The term is dealt with the help of the Hoeffding decomposition (6.12), by setting

$$T(X_i^*, X_j^*) = T_{h,b}(x, X_i^*, X_j^*) := \frac{\sigma(X_i^*)\sigma(X_j^*)}{f(X_i^*)} K_{h,b}(x, X_i^*, X_j^*), \tag{6.22}$$

$t_n = n^2hb$, and $G_1(u) = G_2(u) = u$ (Recall the definition of $K_{h,b}$ in (6.19)). Denote by J_{321}, J_{322}, J_{323} and J_{324} the terms in the decomposition (6.12), which correspond to B_1, B_2, B_3 and B_4, respectively.

By (6.13) and (B.3) below, $\mathrm{E}\left[|J_{321}|\right] = O(1/(nb))$. On account of (6.15), (B.4), (A.3),

$$\mathrm{Var}\Bigg(\sum_{\substack{j,i=1 \\ j \neq i}}^{n} \varepsilon_i \varepsilon_j Z_{1,i} \Bigg) \quad = \quad O(hb^2) \left(n^2 + n^{3-\alpha_\varepsilon} \right).$$

A similar computation is valid if $Z_{1,i}$ is replaced with $Z_{2,j}$. Therefore,

$$\mathrm{Var}(J_{323}) = O(h^{-1} n^{-(1+\alpha_\varepsilon)}), \tag{6.23}$$

and thus $J_{323} = o_P(1/\sqrt{nh})$.

Now, in order to deal with J_{324}, we refer to (6.16). Recalling the definition of $\underline{Z}_{i,j}$, in order to evaluate $\mathrm{E}[\underline{Z}_{i,j}]$ we have to bound $\mathrm{E}[Z_{1,i}^2]$, $\mathrm{E}[Z_{2,j}^2]$ and $\mathrm{Var}[\overline{T_{h,b}(x, X_i^*, X_j^*)}]$. For the latter in fact it suffices to bound $\mathrm{E}[T_{h,b}^2(x, X_i^*, X_j^*)]$. By (B.2) with $l = 2$, (B.4), (B.5) we conclude that the first term in (6.16) is

$$\sum_{j=2}^{n} \sum_{i=1}^{j-1} \mathrm{E}[\varepsilon_i^2 \varepsilon_j^2] \mathrm{E}[\underline{Z}_{i,j}^2] = O(hbn^2).$$

Similarly, we bound $\mathrm{E}[\underline{Z}_{i,j} \underline{Z}_{i',j}]$ as follows:

$$\begin{aligned} \mathrm{E}[\underline{Z}_{i,j} \underline{Z}_{i',j}] = {} & \mathrm{E}[\overline{T_{h,b}(x, X_i^*, X_j^*)}\ \overline{T_{h,b}(x, X_{i'}^*, X_j^*)}] + \mathrm{E}[(Z_{1,i} + Z_{2,j})(Z_{1,i'} + Z_{2,j})] \\ & - \mathrm{E}\left[\overline{T_{h,b}(x, X_i^*, X_j^*)} (Z_{1,i'} + Z_{2,j}) \right] - \mathrm{E}\left[\overline{T_{h,b}(x, X_{i'}^*, X_j^*)} (Z_{1,i} + Z_{2,j}) \right]. \end{aligned}$$

Via (B.9), the first part is of order $O(hb^2)$. To the second part we apply the Cauchy-Schwartz inequality, so that it is bounded by $C\mathrm{E}[Z_{1,i}^2] + C\mathrm{E}[Z_{2,i}^2] = O(hb^2)$ (we use (B.4) and (B.5)). The third and the fourth part are also $O(hb^2)$ by applying (B.6), (B.7). This together with (A.3) yields that the second part in (6.16) is of the order $O(hb^2 n^{3-\alpha_\varepsilon})$.

Consequently, $\mathrm{Var}(J_{324}) = O(1/(n^2hb) + 1/(n^{1+\alpha_\varepsilon} h))$. This yields $J_{324} = o_P(1/\sqrt{nh})$.

Finally, recalling (6.14), the part J_{322} is written as

$$\frac{2\mathrm{E}\left[T_{h,b}(x, X_1^*, X_2^*)\right] \sigma_{n,\varepsilon}^2}{n^2 hb f(x)} \left\{ \left(\sigma_{n,\varepsilon}^{-1} \sum_{j=1}^{n} \varepsilon_j \right)^2 - \sigma_{n,\varepsilon}^{-2} \sum_{j=1}^{n} \varepsilon_j^2 \right\}.$$

The second part in the brackets is negligible. Consequently, via (B.2) below, we obtain (6.20) for this term.

6.5.2. Behaviour of J_{32} under (P2)

The goal of this section is to show that (6.20) is still valid when (P1) is replaced with (P2). Let X_i^*, $i \geq 1$, be an independent version of X_i, $i \geq 1$. Recall Lemma A.3 and its consequence for $f_{j|i}(\cdot|x_i)$, the conditional expectation of X_j given $X_i = x_i$. For $i < j$, write

$$
\begin{aligned}
\mathrm{E}[T_{h,b}(x,X_i,X_j)|\mathcal{X}_i] &= \frac{\sigma(X_i)}{f(X_i)}K_h(x-X_i)\mathrm{E}[\sigma(X_j)K_b(X_i-X_j)|\mathcal{X}_i] \\
&= \frac{\sigma(X_i)}{f(X_i)}K_h(x-X_i)\int K_b(X_i-u)\sigma(u)f_{j|i}(u|X_i)\,du \\
&= \frac{\sigma(X_i)}{f(X_i)}K_h(x-X_i)\int \sigma(u)K_b(X_i-u)f(u)\,du \\
&\quad + O\left(\gamma_X^{1/2}(j-i)\right)\frac{\sigma(X_i)}{f^2(X_i)}K_h(x-X_i)\int \sigma(u)K_b(X_i-u)\,du \\
&= \mathrm{E}[T_{h,b}(x,X_i^*,X_j^*)|X_i^*] + O\left(\gamma_X^{1/2}(j-i)b\right)\frac{\sigma(X_i)}{f^2(X_i)}K_h(x-X_i) \\
&= \mathrm{E}[T_{h,b}(x,X_i^*,X_j^*)|X_i^*] + O\left(\gamma_X^{1/2}(j-i)b\right)O_P(h),
\end{aligned}
$$

where $O_P(\cdot)$ is uniform in $i \neq j$. Likewise,

$$
\begin{aligned}
\mathrm{E}[T_{h,b}(x,X_i,X_j)] &= \mathrm{E}[T_{h,b}(x,X_i^*,X_j^*)] + O\left(\gamma_X^{1/2}(j-i)b\right)\mathrm{E}\left[\frac{\sigma(X_i)}{f^2(X_i)}K_h(x-X_i)\right] \\
&= \mathrm{E}[T_{h,b}(x,X_i^*,X_j^*)] + O\left(\gamma_X^{1/2}(j-i)hb\right).
\end{aligned}
$$

Combining the two expressions,

$$
\mathrm{E}[\overline{T_{h,b}(x,X_i,X_j)}|\mathcal{X}_i] = \mathrm{E}[\overline{T_{h,b}(x,X_i^*,X_j^*)}|X_i^*] + O\left(\gamma_X^{1/2}(j-i)\right)O_P(hb).
$$

Therefore, under (P2), J_{322} can be written as

$$
J_{322} = \frac{2\mathrm{E}\left[T_{h,b}(x,X_1^*,X_2^*)\right]}{n^2hbf(x)}\sum_{\substack{i,j=1\\ i\neq j}}^n \varepsilon_i\varepsilon_j + \frac{O(hb)}{n^2hbf(x)}\sum_{\substack{i,j=1\\ i\neq j}}^n \varepsilon_i\varepsilon_j\gamma_X^{1/2}(j-i).
$$

The expected value of the second part is $O(n^{-(\alpha_\varepsilon+\alpha_X/2)} \vee n^{-1})$, so that it is negligible w.r.t. the first part.

Now, for J_{323} we proceed in the same way. Recall (6.23). Thus,

$$
\begin{aligned}
&\frac{2}{n^2hbf(x)}\sum_{\substack{i,j=1\\ i\neq j}}^n \varepsilon_i\varepsilon_j Z_{1,i} = \\
&= \frac{2\mathrm{E}[T_{h,b}(x,X_1^*,X_2^*)|X_1^*]}{n^2hbf(x)}\sum_{\substack{i,j=1\\ i\neq j}}^n \varepsilon_i\varepsilon_j + O_P(hb)\frac{2}{n^2hbf(x)}\sum_{\substack{i,j=1\\ i\neq j}}^n \varepsilon_i\varepsilon_j\gamma_X^{1/2}(j-i),
\end{aligned}
$$

and once again the second part is of a smaller order than the first one.

From the above computations for J_{322} and J_{323} it is also clear that J_{324} has the same asymptotic behaviour under (P1) and (P2), and thus it is negligible.

6.5.3. Behaviour of J_{31}

Recall (6.2) and denote

$$L_h(X_i) = \frac{\rho(X_i)}{f(X_i)}\sigma(X_i)K_h(x - X_i).$$

Using (6.3), LE/N decomposition (see Remark 6.5) and since $\hat{f}_h(\cdot)$ is a weakly consistent estimator of f,

$$\begin{aligned}
J_{31} &= \frac{2}{nh\hat{f}_h(x)}\sum_{i=1}^{n} R_b(X_i)\sigma(X_i)K_h(x - X_i)\varepsilon_i \\
&= \frac{b^2\kappa_2}{nh\hat{f}_h(x)}\sum_{i=1}^{n} \frac{\rho(X_i)}{f(X_i)}\sigma(X_i)K_h(x - X_i)\varepsilon_i(1 + o_P(1)) \\
&= \frac{b^2\kappa_2}{nhf(x)}\sum_{i=1}^{n} L_h(X_i)\varepsilon_i(1 + o_P(1)) \\
&= \frac{b^2\kappa_2}{nhf(x)}\mathrm{E}\left[L_h(X_1)\right]\sum_{i=1}^{n} \varepsilon_i(1 + o_P(1)) + \frac{b^2\kappa_2}{nhf(x)}\sum_{i=1}^{n} \overline{L_h(X_i)}\varepsilon_i(1 + o_P(1)) \\
&=: J_{311}(1 + o_P(1)) + J_{312}(1 + o_P(1)).
\end{aligned}$$

Since

$$\mathrm{E}[L_h(X_1)] = h\rho(x)\sigma(x) + \frac{h^3}{2}\kappa_2(\rho\sigma)''(x) + o(h^3),$$

we get

$$J_{311} = \frac{\kappa_2\rho(x)\sigma(x)}{f(x)}\frac{b^2}{n}\sum_{i=1}^{n}\varepsilon_i + O_P\left(\frac{b^2h^2}{n^{\alpha_\varepsilon/2}}\right). \tag{6.24}$$

It remains to deal with J_{312}. We have

$$\begin{aligned}
\mathrm{Var}(J_{312}) &= O\left(\frac{b^4}{n^2h^2}\right)\sum_{i,i'=1}^{n} \mathrm{E}\varepsilon_i\varepsilon_{i'}\mathrm{E}\left[\overline{L_h(X_i)}\;\overline{L_h(X_{i'})}\right] \\
&= O\left(\frac{b^4}{n^2h^2}\right)\left\{n\mathrm{E}[\varepsilon_1^2]\mathrm{E}\left[L_h^2(X_1)\right] + \sum_{\substack{i,i'=1\\ i\neq i'}}^{n} \mathrm{E}\varepsilon_i\varepsilon_{i'}\mathrm{E}\left[\overline{L_h(X_i)}\;\overline{L_h(X_{i'})}\right]\right\}.
\end{aligned}$$

Under (P1), the second part vanishes and the above variance is of order

$$O\left(\frac{b^4}{n^2h^2}\right)nh = O\left(\frac{b^4}{nh}\right). \tag{6.25}$$

Under (P2) we do the same trick as in case of J_{321}:

$$\begin{aligned}
&\mathrm{E}\left[\overline{L_h(X_i)}\ \overline{L_h(X_{i'})}\right] = \\
&\quad = \mathrm{E}\left[\overline{L_h(X_i^*)}\ \overline{L_h(X_{i'}^*)}\right] + \gamma^{1/2}(|i-i'|)\left(\int \frac{\rho(u)\sigma(u)}{f(u)}K_h(x-u)\,du\right)^2 \\
&\quad = 0 + \gamma^{1/2}(|i-i'|)O(h^2).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathrm{Var}(J_{312}) &= O\left(\frac{b^4}{n^2h^2}\right)\left(nh + h^2\mathrm{Var}\left(\sum_{\substack{i,i'=1\\ i\neq i'}}^{n} \mathrm{Cov}(\varepsilon_i\varepsilon_{i'})\gamma^{1/2}(|i-i'|)\right)\right) \\
&= O(b^4 n^{-(\alpha_\varepsilon+\alpha_X/2)}),
\end{aligned}$$

which means that this term is of a smaller order than J_{311}. This together with (6.24) and (6.25) yields (6.21). □

6.6. Proof of Lemma 6.4

We work simulateneously under (P1) and (P2). As for J_4, the general idea is similar to J_3. Using the previous notation (cf. (6.1), (6.2), (6.19)), and introducing

$$K_{h,b,b}(x, X_i, X_j, X_l) = K_h(x-X_i)K_b(X_i-X_j)K_b(X_i-X_l), \tag{6.26}$$

we have

$$\begin{aligned}
J_4 &= \frac{1}{nh\hat{f}_h(x)}\sum_{i=1}^{n} R_b^2(X_i)K_h(x-X_i) \\
&\quad + \frac{2}{nh\hat{f}_h(x)}\sum_{i=1}^{n}\frac{1}{nb\hat{f}_b(X_i)}R_b(X_i)\sum_{j=1}^{n}\sigma(X_j)K_{h,b}(x,X_i,X_j)\varepsilon_j \\
&\quad + \frac{1}{n^3hb^2\hat{f}_h(x)}\sum_{i=1}^{n}\frac{1}{\hat{f}_b^2(X_i)}\sum_{j,j'=1}^{n}K_{h,b,b}(x,X_i,X_j,X_{j'})\varepsilon_j\varepsilon_{j'} =: J_{41}+J_{42}+J_{43}.
\end{aligned}$$

Using (6.2), (6.3) and the weak consistency of $\hat{f}_h$ we write

$$J_{41} = \frac{b^4\kappa_2^2}{4nhf(x)}\sum_{i=1}^{n}\frac{\rho(X_i)}{f(X_i)}K_h(x-X_i)(1+o_P(1)) = O_P(b^4). \tag{6.27}$$

Furthermore, we show below (see Section 6.6.1) that

$$J_{42} = \frac{\kappa_2\rho(x)\sigma(x)}{f(x)}\frac{b^2}{n}\sum_{i=1}^{n}\varepsilon_i + O_P\left(\frac{b^2+bh}{n^{\alpha_\varepsilon/2}}\right), \tag{6.28}$$

and

$$J_{42} = \frac{\kappa_2 \rho(x)\sigma(x)}{f(x)} \frac{b^2}{n} \sum_{i=1}^{n} \varepsilon_i + O_P\left(\frac{b^2(h^2+b^2)}{n^{\alpha_\varepsilon}}\right), \tag{6.29}$$

under (P2) and (P1), respectively. Furthermore (see Section 6.6.2),

$$J_{43} = J_{4341} + O_P(1/(nb)) + o_P(n^{-\alpha_\varepsilon} + 1/\sqrt{nh}), \tag{6.30}$$

where

$$n^{\alpha_\varepsilon} J_{4341} \xrightarrow{\text{d}} C_1^2 \sigma^2(x) \times \chi^2(1). \tag{6.31}$$

Combining (6.21) and (6.28), we find out that under (P2) we see that the term $-J_{31} + J_{42}$ is negligible if either (H2), (H2a) or (H2b) hold. Under (P1), combining (6.21) and (6.29) we see that the leading term in $-J_{31} + J_{42}$ is of order

$$O_P\left(\frac{b^3 + b^2 h}{n^{\alpha_\varepsilon/2}} + \frac{b^2}{nh}\right).$$

6.6.1. *Behaviour of J_{42}*

The term J_{42} is treated in a similar way to J_{31}:

$$\begin{aligned}
J_{42} &= \frac{2}{nhf(x)} \sum_{i=1}^{n} \frac{R_b(X_i)}{nb\hat{f}(X_i)} \sum_{j=1}^{n} \sigma(X_j) K_{h,b}(x, X_i, X_j)\varepsilon_j(1 + o_P(1)) \\
&= \quad \frac{b^2\kappa_2}{nhf(x)} \sum_{i=1}^{n} \frac{\rho(X_i)}{nbf^2(X_i)} \sum_{j=1}^{n} \sigma(X_j) K_{h,b}(x, X_i, X_j)\varepsilon_j(1 + o_P(1)) \\
&\overset{LE/N}{=} \quad \frac{b\kappa_2}{n^2hf(x)} \sum_{j=1}^{n} \mathrm{E}[\tilde{L}_{h,b}(X_j)]\varepsilon_j(1 + o_P(1)) \\
&\quad + \frac{b\kappa_2}{n^2hf(x)} \sum_{j=1}^{n} \overline{\tilde{L}_{h,b}(X_j)}\varepsilon_j(1 + o_P(1)) \\
&=: \quad J_{421}(1 + o_P(1)) + J_{422}(1 + o_P(1)),
\end{aligned}$$

where (recall (6.19))

$$\tilde{L}_{h,b}(X_j) = \sigma(X_j) \sum_{i=1}^{n} \frac{\rho(X_i)}{f^2(X_i)} K_{h,b}(x, X_i, X_j). \tag{6.32}$$

Under (P1), we use (B.11) to get

$$J_{421} = \frac{\kappa_2 \sigma(x)\rho(x)}{f(x)} \frac{b^2}{n} \sum_{j=1}^{n} \varepsilon_j + O_P\left(b^2(h^2 + b^2)n^{-\alpha_\varepsilon/2}\right). \tag{6.33}$$

Under (P2), using the trick as for J_{312},

$$\mathrm{E}[\tilde{L}_{h,b}(X_j)] = \mathrm{E}[\tilde{L}_{h,b}(X_j^*)] + \sum_{i=1}^{n} \gamma^{1/2}(|j - i|)O(hb).$$

Now,

$$\mathrm{Var}\left(\sum_{j=1}^{n}\varepsilon_j\sum_{i=1}^{n}\gamma^{1/2}(|j-i|)\right)=O(n^{1-\alpha_X}+n^{2-(\alpha_X+\alpha_\varepsilon)})=o(n^{2-\alpha_\varepsilon}).$$

Consequently, the behaviour of J_{421} is the same under (P1) and (P2).

Finally, using (B.12) and (B.10) we have:

$$\begin{aligned}\mathrm{Var}(J_{422})&=O\left(\frac{b^2}{n^4h^2}\right)\sum_{j,j'=1}^{n}\mathrm{E}[\varepsilon_j\varepsilon_{j'}]\mathrm{E}\left[\overline{\tilde{L}_{h,b}(X_j)}\,\overline{\tilde{L}_{h,b}(X_{j'})}\right]\\ &= O\left(\frac{b^2}{n^4h^2}\right)\left(\underbrace{n^2h^2b^2n^{2-\alpha_\varepsilon}+nhb^2n^{2-\alpha_\varepsilon}}_{j\neq j'}+\underbrace{nhbn+n^2h^2bn}_{j=j'}\right)\\ &= O\left(\frac{b^4+b^2h^2}{n^{\alpha_\varepsilon}}+\frac{b^4}{hn^{1+\alpha_\varepsilon}}+\frac{b^3}{n^2h}+\frac{b^3}{n}\right).\end{aligned} \tag{6.34}$$

Combining (6.33) and (6.34), we get (6.28). Under (P1), we use (B.13) instead of (B.12) to get (6.29).

6.6.2. *Behaviour of J_{43}*

Recall (6.26). Using the weak consistency of $\hat{f}_h$ and the finite support of $K(\cdot)$, we may write J_{43} as

$$\frac{1}{n^3hb^2f(x)}\sum_{i,j,l=1}^{n}\frac{1}{f^2(X_i)}\sigma(X_j)\sigma(X_l)K_{h,b,b}(x,X_i,X_j,X_l)\varepsilon_j\varepsilon_l(1+o_P(1)). \tag{6.35}$$

The expression (6.35) can be decomposed as

$$\begin{aligned}&\frac{1}{n^3f(x)hb^2}K^2(0)\sum_{i=1}^{n}\frac{\sigma^2(X_i)}{f^2(X_i)}K_h(x-X_i)\varepsilon_i^2\\ &\quad+\frac{1}{n^3f(x)hb^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{\sigma^2(X_j)}{f^2(X_i)}K_h(x-X_i)K_b^2(X_i-X_j)\varepsilon_j^2\\ &\quad+\frac{2}{n^3f(x)hb^2}K(0)\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{\sigma(X_i)\sigma(X_j)}{f^2(X_i)}K_h(x-X_i)K_b(X_i-X_j)\varepsilon_i\varepsilon_j\\ &\quad+\frac{1}{n^3hb^2f(x)}\sum_{\substack{i,j,l=1:\\ \text{all different}}}^{n}\frac{1}{f^2(X_i)}\sigma(X_j)\sigma(X_l)K_{h,b,b}(x,X_i,X_j,X_l)\varepsilon_j\varepsilon_l\\ =:&\quad J_{431}+J_{432}+J_{433}+J_{434}.\end{aligned}$$

Under (P1), $\mathrm{E}[|J_{433}|] = O(1/(nb))$. The latter is also valid under (P2) by proceeding in the very same way as in case of J_3. Using this we get

$$\mathrm{E}[|J_{431}| + |J_{432}| + |J_{433}|] = O(1/(nb)^2) + O(1/(nb)) + O(1/(nb)) = O(1/(nb)). \tag{6.36}$$

We deal with J_{434}. Let

$$S_{h,b}(x, X_j, X_l) := \sigma(X_j)\sigma(X_l) \sum_{\substack{i=1 \\ i\neq j,\ i\neq l}}^{n} \frac{1}{f^2(X_i)} K_{h,b,b}(x, X_i, X_j, X_l). \tag{6.37}$$

We use again LE/N decomposition:

$$\begin{aligned} J_{434} \quad &= \quad \frac{1}{n^3hb^2 f(x)} \sum_{\substack{j,l=1 \\ j\neq l}}^{n} \varepsilon_j\varepsilon_l \mathrm{E}[S_{h,b}(x, X_j, X_l)] \\ &\quad + \frac{1}{n^3hb^2 f(x)} \sum_{\substack{j,l=1 \\ j\neq l}}^{n} \varepsilon_j\varepsilon_l \left\{\overline{S_{h,b}(x, X_j, X_l)}\right\} =: J_{4341} + J_{4342}. \end{aligned}$$

Assume (P1). One can verify that for $j \neq l$,

$$a_n := \mathrm{E}[S_{h,b}(x, X_j, X_l)] \sim nhb^2 f(x)\sigma^2(x)$$

we have

$$\frac{a_n}{n^3hb^2 f(x)} \sum_{\substack{j,l=1 \\ j\neq l}}^{n} \varepsilon_j\varepsilon_l = \frac{a_n\sigma^2_{n,\varepsilon}}{n^3hb^2 f(x)} \left\{ \left(\sigma^{-1}_{n,\varepsilon} \sum_{j=1}^{n} \varepsilon_j \right)^2 - \sigma^{-2}_{n,\varepsilon} \sum_{j=1}^{n} \varepsilon_j^2 \right\},$$

so that (6.31) holds. The same is valid under (P2), using Lemma A.3 and the same trick as for $\tilde{L}_{h,b}(X_j)$.

Now, using the covariance bound (A.4), (B.14) and Lemma A.1 (for all different indices),

$$\begin{aligned} \mathrm{Var}(J_{4342}) &= O\left(\frac{1}{n^6h^2b^4}\right) \sum_{\substack{j,l,j',l'=1 \\ j\neq l, j'\neq l'}}^{n} \mathrm{E}[\varepsilon_j\varepsilon_l\varepsilon_{j'}\varepsilon_{l'}]\mathrm{E}[\overline{S_{h,b}(x, X_j, X_l)}\ \overline{S_{h,b}(x, X_{j'}, X_{l'})}] \\ &= O\left(\frac{1}{n^6h^2b^4}\right) \Bigg\{ \underbrace{o(n^2h^2b^4)n^{4-2\alpha_\varepsilon} + O(n^2h^2b^4)o(n^{4-2\alpha_\varepsilon})}_{\text{all indices different}} + \underbrace{(n^2hb^4 + nhb^3)n^{3-\alpha_\varepsilon}}_{\text{two indices agree}} \Bigg\}. \end{aligned}$$

The contribution from *all indices different* is $o(n^{-2\alpha_\varepsilon})$, so that this contribution is smaller than of J_{4341}. The estimates coming from *two indices agree* are $o(1/(nh))$. From this and (6.36), we obtain (6.35).

6.7. Proof of Theorem 3.10

The proofs are sketched here, since most of the steps are similar to the proof for the conditional variance estimation. The proof is conducted for (P1) only. Recall the definition of Δ_i in (6.1). We write the two term Taylor expansion

$$\begin{aligned}
&\frac{1}{nh}\sum_{i=1}^{n} K_h(x-\hat{\varepsilon}_i) - \frac{1}{nh}\sum_{i=1}^{n} K_h(x-\varepsilon_i) = \\
&= \quad \frac{1}{nh^2}\sum_{i=1}^{n} K'\left(\frac{x-\varepsilon_i}{h}\right)\Delta_i + \frac{1}{2nh^3}\sum_{i=1}^{n} K''\left(\frac{x-\varepsilon_i^{\#}}{h}\right)\Delta_i^2 \\
&=: \quad I_1 + I_2,
\end{aligned} \tag{6.38}$$

where $\varepsilon_i^{\#}$ is a random term which lies between ε_i and $\hat{\varepsilon}_i$.

In what follows, we will show that

$$\begin{aligned}
\frac{1}{nh}\sum_{i=1}^{n} K_h(x-\hat{\varepsilon}_i) - \frac{1}{nh}\sum_{i=1}^{n} K_h(x-\varepsilon_i) &= f_\varepsilon'(x)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \\
+o_P(n^{-\alpha_\varepsilon/2}) + O_P&\left(\frac{1}{nh^{3/2}b^{1/2}} + \frac{1}{n^{(1+\alpha_\varepsilon)/2}h^{3/2}} + b^4 + \frac{1}{nb}\right).
\end{aligned} \tag{6.39}$$

Since we assumed here $\sigma(\cdot) \equiv 1$, we have uniformly in i, $\Delta_i = O_P(b^2) + O_P(n^{-\alpha_\varepsilon/2})$, see (6.3) together with estimations for J_{31} in Section 6.5.3. Also, it is easy to verify that

$$\mathrm{E}K''((x-\varepsilon_i^{\#})/h) = O(h^3). \tag{6.40}$$

Therefore,

$$I_2 = O_P(b^4 + n^{-\alpha_\varepsilon} + b^2 n^{-\alpha_\varepsilon/2}).$$

In fact, in this situation it will be seen that the first term I_1 is of higher order than the second one.

We write the first term as (recall that we may replace $\hat{f}_h(X_i)$ with $f(X_i)$),

$$\begin{aligned}
I_1 &= \frac{1}{nh^2}\sum_{i=1}^{n} K'\left(\frac{x-\varepsilon_i}{h}\right) R_b(X_i) \\
&\quad + \frac{1}{n^2h^2b}\sum_{i,j=1}^{n} \frac{1}{f(X_i)} K_b(X_i-X_j) K'\left(\frac{x-\varepsilon_i}{h}\right)\varepsilon_j =: O_P(b^2) + I_{12},
\end{aligned}$$

where the latter follows from (6.3) and

$$\mathrm{E}K'\left(\frac{x-\varepsilon_i}{h}\right) = -h^2 f_\varepsilon'(x)\int vK'(v)dv + o(h^2) = h^2 f_\varepsilon'(x) + o(h^2). \tag{6.41}$$

For I_{12} we apply the Hoeffding decomposition (6.12) with

$$T(X_i, X_j) = T_b(X_i, X_j) = \frac{1}{f(X_i)} K_b(X_i - X_j),$$

$t_n = n^2h^2b$ and $G_1(u) = K'((x-u)/h)$, $G_2(u) = u$. Denote the terms in the decomposition by I_{121}, I_{122}, I_{123} and I_{124}. Via (6.13),

$$I_{121} = O_P\left(\frac{1}{n^2h^2b} n\mathrm{E}[K'\left(\frac{x-\varepsilon_1}{h}\right)\varepsilon_1]\right) = O_P(1/(nb)).$$

For I_{123}, I_{124} we use (6.15), (6.16), respectively. We note that $\mathrm{E}[Z_{1,i}^2 + Z_{2,j}^2] = O(b^2)$. Furthermore, $\mathrm{E}[T^2(X_i, X_j)] = O(b)$. Also,

$$\mathrm{E}[K'^2((x-\varepsilon_i)/h)\varepsilon_j^2] = O(h), \qquad \mathrm{E}[K'^2((x-\varepsilon_i)/h)\varepsilon_j\varepsilon_{j'}] = O(1)\mathrm{E}[\varepsilon_j\varepsilon_{j'}].$$

where in the second we used that $K'(\cdot)$ is bounded, with finite support. We conclude that the term I_{123} is of order

$$O_P\left(\frac{1}{n^4h^4b^2}\left(n^2hb + n^{3-\alpha_\varepsilon}b^2\right)\right).$$

A similar estimation is valid for I_{124}.

Finally, for the second term I_{122}, via (6.41) and $\mathrm{E}[T(X_1, X_2)] = b + o(b)$, we conclude that the asymptotic behaviour of $I_1 + I_2$ is

$$f'_\varepsilon(x)\frac{1}{n}\sum_{j=1}^n \varepsilon_j + o_P(n^{-\alpha_\varepsilon/2}) + O_P\left(\frac{1}{nh^{3/2}b^{1/2}} + \frac{1}{n^{(1+\alpha_\varepsilon)/2}h^2} + b^4 + 1/(nb)\right).$$

Appendix A: LRD processes

Moment bounds

Recall that under (E2) we have

$$\sigma_{n,\varepsilon}^2 := \mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i\right) \sim C_1^2 n^{2-\alpha_\varepsilon}, \qquad \alpha_\varepsilon \in (0,1), \tag{A.1}$$

with C_1 defined in (2.1). Furthermore, one can verify that with some $C > 0$ (see e.g. [15, Lemma 4.1, Appendix A]),

$$\mathrm{Cov}(\varepsilon_i^2, \varepsilon_j^2) \sim C^2\gamma_\varepsilon^2(|i-j|), \tag{A.2}$$

$$\mathrm{E}[\varepsilon_i^2\varepsilon_j\varepsilon_{j'}] = O\left(\gamma_\varepsilon(|i-j|) + \gamma_\varepsilon(|i-j'|) + \gamma_\varepsilon(|j-j'|)\right), \tag{A.3}$$

$$\begin{aligned}\mathrm{E}[\varepsilon_i\varepsilon_j\varepsilon_{i'}\varepsilon_{j'}] = O\big(&\gamma_\varepsilon(|i-j|)\gamma_\varepsilon(|i'-j'|) + \gamma_\varepsilon(|i-j'|)\gamma_\varepsilon(|i'-j|) \\ &+ \gamma_\varepsilon(|i-i'|)\gamma_\varepsilon(|j-j'|)\big),\end{aligned} \tag{A.4}$$

if the differences $|i-j|$, $|i-j'|$, $|i'-j|$, $|i'-j'|$ are large. From (A.2), (A.3) and (A.4) we obtain, in particular,

$$d_{n,\varepsilon}^2 := \mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i^2\right) \sim \begin{cases} C_2^2 n^{2(1-\alpha_\varepsilon)}, & \text{if } \alpha_\varepsilon < 1/2, \\ C_3^2 n, & \text{if } \alpha_\varepsilon > 1/2, \end{cases} \tag{A.5}$$

$$\sum_{\substack{i,j,j' \\ \text{all indices different}}}^{n} \mathrm{E}[\varepsilon_i \varepsilon_j \varepsilon_{j'}^2] = O(n^{3-\alpha_\varepsilon}), \qquad \sum_{\substack{i,j,i',j'=1 \\ \text{all indices different}}}^{n} \mathrm{E}[\varepsilon_j \varepsilon_j \varepsilon_{i'} \varepsilon_{j'}] = O(n^{4-2\alpha_\varepsilon}). \tag{A.6}$$

where C_2 is defined in (3.2) and C_3 is a finite and positive constant. Moreover:

Lemma A.1. *Assume* (E2) *and* (P2)*. Let*

$$c(i,i',j,j') = (\max(|i-j|,|i-j'|,|i-i'|,|i'-j|,|i'-j'|,|j'-j|))^{-\alpha_X/2}).$$

Then

$$\sum_{\substack{i,i',j,j'=1 \\ \text{all indices different}}}^{n} \mathrm{E}[\varepsilon_i \varepsilon_j \varepsilon_{i'} \varepsilon_{j'}] c(i,i',j,j') = o(n^{4-2\alpha_\varepsilon}).$$

Projections

Recall that under (E2) we have the following σ-field: $\mathcal{H}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$. Define

$$\varepsilon_{i,i-1} := \mathrm{E}[\varepsilon_i | \mathcal{H}_{i-1}] = \sum_{k=1}^{\infty} c_k \eta_{i-k}, \quad i \geq 1. \tag{A.7}$$

This linear process has the same memory parameter as ε_i, $i \geq 1$, and is introduced for technical reasons. In particular, (A.1) is also valid for $\mathrm{Var}\,(\sum_{i=1}^n \varepsilon_{i,i-1})$.

Let $\xi_i = \varepsilon_i^2 - \mathrm{E}[\varepsilon_i^2] = \varepsilon_i^2 - 1$. Also, in the same spirit as in (A.7), we obtain

$$\mathrm{E}[\xi_i | \mathcal{H}_{i-1}] = \varepsilon_{i,i-1}^2 - \mathrm{E}[\varepsilon_{i,i-1}^2],$$

The sequence $\mathrm{E}[\xi_i | \mathcal{H}_{i-1}]$, $i \geq 1$, is defined in terms of squares of the linear process. Therefore (cf. (A.2))

$$\mathrm{Cov}\,(\mathrm{E}[\xi_i | \mathcal{H}_{i-1}], \mathrm{E}[\xi_j | \mathcal{H}_{j-1}]) \sim C^2 \gamma_\varepsilon^2(|i-j|), \tag{A.8}$$

so that

$$\mathrm{Var}\left(\sum_{i=1}^{n} \mathrm{E}[\xi_i | \mathcal{H}_{i-1}]\right) \sim \begin{cases} C_2^2 n^{2(1-\alpha_\varepsilon)}, & \text{if } \alpha_\varepsilon < 1/2, \\ C_4^2 n, & \text{if } \alpha_\varepsilon > 1/2, \end{cases} \tag{A.9}$$

with a possibly different constant C_4.

Limit theorems

We have (see e.g. [1, Theorem 2])

$$\sigma_{n,\varepsilon}^{-1} \sum_{i=1}^{n} \varepsilon_i \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \qquad \sigma_{n,\varepsilon}^{-1} \sum_{i=1}^{n} \varepsilon_{i,i-1} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \tag{A.10}$$

$$d_{n,\varepsilon}^{-1} \sum_{i=1}^{n} \xi_i \xrightarrow{\mathrm{d}} H_2, \qquad d_{n,\varepsilon}^{-1} \sum_{i=1}^{n} \mathrm{E}[\xi_i | \mathcal{H}_{i-1}] \xrightarrow{\mathrm{d}} H_2, \qquad \text{if } \alpha_\varepsilon < 1/2, \tag{A.11}$$

where H_2 is the Hermite-Rosenblatt random variable defined in (3.1). If $\alpha_\varepsilon > 1/2$, then $\sum_{i=1}^n \xi_i$ converges to a normal random variable with $\sqrt{n}$-normalization.

Remark A.2. If $\alpha_\varepsilon = 1/2$, then the expressions for the variances in (A.5) or (A.9) involve slowly varying functions. Furthermore, the limiting results like in (A.11) would involve a linear combination of H_2 and a standard normal random variable. For simplicity, we do not include this case in our computations.

For more details on limit theorems for linear processes and its functionals we refer to [19].

Density expansions

Lemma A.3. *Assume* (P2) *and that* $\|f_\zeta + f'_\zeta\| < \infty$. *Let* $\mathbf{i}_m = \mathbf{i} = (i_1, \ldots, i_m)$ *and let* $f_\mathbf{i}$ *be the joint density of* $X_\mathbf{i} = (X_{i_1}, \ldots, X_{i_m})$. *Then*

$$\sup_{x_\mathbf{i}} \left| f_\mathbf{i}(x_\mathbf{i}) - \prod_{l=1}^m f(x_{i_l}) \right| = O\left(\max_{l,l'=1,\ldots,m} |l - l'|^{-\alpha_X/2} \right). \tag{A.12}$$

Remark A.4. Let $f_{j|i}$ be the conditional density of X_j given X_i. Let $\gamma_X(i) = \mathrm{Cov}(X_0, X_i)$. As a consequence, we obtain that for a given x such that $f(x) \neq 0$

$$\sup_y |f_{j|i}(y|x) - f(y)| = O(|j-i|^{-\alpha_X/2}/f(x)) = O(\gamma_X^{1/2}(|i-j|)/f(x)).$$

Remark A.5. Furthermore, let $\mathbf{i}_r = (i_1, \ldots, i_r)$, $\mathbf{i}_{r,m} = (i_{r+1}, \ldots, i_m)$. From (A.12) we can also conclude

$$\sup_{x_\mathbf{i}} \left| f_{\mathbf{i}_m}(x_{\mathbf{i}_m}) - f_{\mathbf{i}_r}(x_{\mathbf{i}_r}) f_{\mathbf{i}_{r,m}}(x_{\mathbf{i}_{r,m}}) \right| = O\left(\max_{l,l'=1,\ldots,m} |l - l'|^{-\alpha_X/2} \right).$$

Remark A.6. In [14] the authors established

$$\sup_{x,y} |f_{i,j}(x,y) - f(x)f(y) - \gamma(|j-i|)f'(x)f'(y)| = o(|j-i|^{-\alpha_X}),$$

if there exist $\delta > 0$ and $C < \infty$ such that $|\mathrm{E}\exp(iu\zeta_0)| \leq C/(1+|u|)^\delta$. If $\|f'\|_\infty < \infty$, then the bound leads to

$$\sup_{x,y} |f_{i,j}(x,y) - f(x)f(y)| = O(|j-i|^{-\alpha_X}).$$

Therefore, our Lemma A.3 yields a less precise bound, which is though appropriate for our purposes. On the other hand, the method in [14] does not seem to be suitable for multivariate densities.

Proof of Lemma A.3. Let us start with $m = 2$ and set $(i_1, i_2) = (i, j)$. W.l.o.g. assume that $i < j$. Split

$$X_j = \sum_{k=0}^\infty a_k \zeta_{j-k} = \sum_{k=0}^{j-i-1} a_k \zeta_{i-k} + \sum_{k=j-i}^\infty a_k \zeta_{i-k} =: \tilde{X}_{j,i} + X_{j,i}$$

and note that $X_{j,i}$ is $\mathcal{X}_i = \sigma(\zeta_i, \zeta_{i-1}, \ldots)$-measurable, whereas $\tilde{X}_{j,i}$ is independent of $\mathcal{X}_i$. Note further that

$$F(y) = P(X_j \le y) = \mathrm{E}P(X_j \le y | \mathcal{X}_i) = F_{\tilde{X}_{j,i}}(y - X_{j,i}),$$

so that $f(y) = f_{\tilde{X}_{j,i}}(y - X_{j,i})$, where $F_{\tilde{X}_{j,i}}$ and $f_{\tilde{X}_{j,i}}$ are the distribution and density of $\tilde{X}_{j,i}$. We claim that if f_ζ and f'_ζ are bounded, then, respectively, $f_{\tilde{X}_{j,i}}$ and $f'_{\tilde{X}_{j,i}}$ are bounded as well. Indeed,

$$f_{\tilde{X}_{j,i}}(x) = \int \cdots \int f_{\zeta_j}\left(\frac{x - \sum_{k=1}^{j-i-1} a_k u_{j-k}}{a_0}\right) \prod_{l=1}^{j-i-1} f_{\zeta_l}(u_{j-l})\, du_{j-1} \cdots du_{i+1},$$

and clearly $\|f_{\tilde{X}_{j,i}}\|_\infty \le \|f_{\zeta_j}\|_\infty = \|f_\zeta\|_\infty$. A similar argument works for the derivative.

Bearing this in mind,

$$\begin{aligned} P(X_i \le x, X_j \le y) &= \mathrm{E}[1_{\{X_i \le x\}} F_{\tilde{X}_{j,i}}(y - X_{j,i})] \\ &= F_{\tilde{X}_{j,i}}(y)\mathrm{E}[1_{\{X_i \le x\}}] - \mathrm{E}[1_{\{X_i \le x\}} f_{\tilde{X}_{j,i}}(\theta) X_{j,i}] \end{aligned}$$

where $|\theta - y| \le X_{i,j}$. By the Cauchy-Schwartz inequality, the second term is bounded by

$$\|f\|_\infty F^{1/2}(x) \left(\mathrm{E}X_{i,j}^2\right)^{1/2} = O(1) \left(\sum_{k=j-i}^{\infty} a_k^2\right)^{1/2} = O(|j-i|^{-\alpha_X/2}).$$

Furthermore, by the Markov inequality

$$\begin{aligned} |F_{\tilde{X}_{j,i}}(y) - F(y)| &= |F_{\tilde{X}_{j,i}}(y) - \mathrm{E}F_{\tilde{X}_{j,i}}(y - X_{j,i})| \\ &= |\mathrm{E}f_{\tilde{X}_{j,i}}(\theta) X_{j,i}| \le \|f^2\|_\infty^{1/2} \left(\mathrm{E}X_{j,i}^2\right)^{1/2} = O(|j-i|^{-\alpha_X/2}). \end{aligned}$$

We conclude that

$$P(X_i \le x, X_j \le y) = F(x)F(y) + O(|j-i|^{-\alpha_X/2}).$$

The same argument applies to the joint density, given that $\|f'_\zeta\|_\infty < \infty$. Thus, the Lemma is valid for $m = 2$.

Now, we show the induction step, from $m = 2$ to $m = 3$. Of course, the same holds from arbitrary $m - 1$ to m. Assume that $i < j < l$. Repeating the same argument as above

$$\begin{aligned} P(X_i \le x, X_j \le y, X_l \le z) &= \mathrm{E}[1_{\{X_i \le x, X_j \le y\}} \tilde{F}_{X_{l,j}}(z - X_{l,j})] \\ &= \tilde{F}_{X_{l,j}}(z)\mathrm{E}[1_{\{X_i \le x, X_j \le y\}}] - \mathrm{E}[1_{\{X_i \le x, X_j \le y\}} \tilde{f}_{X_{l,j}}(\xi) X_{l,j}] \\ &\overset{\text{induction}}{=} \tilde{F}_{X_{l,j}}(z)F(x)F(y) + \tilde{F}_{X_{l,j}}(y) O(|j-i|^{-\alpha_X/2}) + \|f\|_\infty \left(\mathrm{E}X_{j,i}^2\right)^{1/2} \\ &= F(x)F(y)F(z) + F(x)F(y)(\tilde{F}_{X_{l,j}}(z) - F(z)) + O(|j-i|^{-\alpha_X/2}) \\ &= F(x)F(y)F(z) + O(|l-j|^{-\alpha_X/2}) + O(|j-i|^{-\alpha_X/2}). \end{aligned}$$

□

Appendix B: Integrals

Covariance bound

Assume that (P2) holds. Recall that $\gamma_X(i) = \mathrm{Cov}(X_0, X_i)$ and $X_{i,i-1} = X_i - \zeta_i$. Note that Lemma A.3 is applicable to $\tilde{f}_{i,j}$, the joint density of $X_{i,i-1}$, $X_{j,j-1}$; we have to replace $f(\cdot)$ there with $f_{1,0}(\cdot)$, the density of $X_{i,i-1}$, $i \geq 1$. Let $r(s,u) = \sigma^2(s) f_\zeta(s-u)$. For $i \neq j$,

$$\begin{aligned}
&\mathrm{Cov}\left(\mathrm{E}[\sigma^2(X_i)K_h(x-X_i)|\mathcal{X}_{i-1}], \mathrm{E}[\sigma^2(X_j)K_h(x-X_j)|\mathcal{X}_{j-1}]\right) \\
&\quad= h^2 \mathrm{Cov}\left(\int K(s) r(x-sh, X_{i,i-1})\, ds, \int K(t) r(x-th, X_{j,j-1})\, dt\right) \\
&\quad= h^2 O\left(\gamma_X^{1/2}(|j-i|)\right) \iiiint K(s)K(t) r(x-sh,u) r(x-th,v)\, du\, dv\, ds\, dt. \\
&\qquad = h^2 O\left(\gamma_X^{1/2}(|j-i|)\right). \qquad (B.1)
\end{aligned}$$

Bounds on $T_{h,b}$

For appropriately smooth functions $r_1(\cdot)$, $r_2(\cdot)$, let

$$V_{h,b}(x, X_i, X_j) := r_1(X_i) r_2(X_j) K_h(x - X_i) K_b(X_i - X_j).$$

Note that $T_{h,b}$ in (6.22), when multiplied by $n^2 hb$, is the special case of $V_{h,b}$.

For $V_{h,b}$ we have the following bounds which are valid under (P2) - below, it is assumed that all indices are different. Also, recall from Lemma A.3 that $f_{i,j,i',j'}$ is the joint density of $(X_i, X_j, X_{i'}, X_{j'})$.

$$\mathrm{E}\left[V_{h,b}^l(x, X_i, X_j)\right] \sim hb r_1^l(x) r_2^l(x) f_{i,j}(x,x) \left(\int K^l(s)\, ds\right)^2, \quad l = 1, 2. \qquad (B.2)$$

$$\mathrm{E}\left[V_{h,b}(x, X_i, X_i)\right] \sim hK(0) r_1(x) r_2(x) f(x). \qquad (B.3)$$

$$\mathrm{E}\left[\left(\mathrm{E}\left[V_{h,b}(x, X_i, X_j)|X_i\right]\right)^2\right] = O(hb^2). \qquad (B.4)$$

$$\mathrm{E}\left[\left(\mathrm{E}\left[V_{h,b}(x, X_i, X_j)|X_j\right]\right)^2\right] = O(h^2 b^2). \qquad (B.5)$$

$$\mathrm{E}\left[V_{h,b}(x, X_i, X_j) \mathrm{E}[V_{h,b}(x, X_{i'}, X_k)|X_{i'}]\right] = O(h^2 b^2). \qquad (B.6)$$

$$\mathrm{E}\left[V_{h,b}(x, X_i, X_j) \mathrm{E}[V_{h,b}(x, X_k, X_j)|X_j]\right] = O(hb^2). \qquad (B.7)$$

$$\begin{aligned}
&\mathrm{E}[V_{h,b}(x, X_i, X_j) V_{h,b}(x, X_{i'}, X_{j'})] = \\
&\quad = h^2 b^2 r_1^2(x) r_2^2(x) f_{i,j,i',j'}(x,x,x,x) + h^2 b^2 O(h^2 + b^2). \qquad (B.8)
\end{aligned}$$

$$\mathrm{E}[V_{h,b}(x, X_i, X_j) V_{h,b}(x, X_{i'}, X_j)] = O(hb^2). \qquad (B.9)$$

Bounds on $\tilde{L}_{h,b}$

Recall the definitions of $\tilde{L}_{h,b}(X_j)$ and $K_{h,b}(X_i, X_j)$ in (6.32) and (6.19), respectively. Let

$$r_0(x_i, x_j) := \frac{\sigma(x_j)\rho(x_i)}{f^2(x_i)}.$$

If $j = j'$ we have

$$\mathrm{E}[\tilde{L}^2_{h,b}(X_j)] = O(n^2h^2b + nhb), \tag{B.10}$$

and

$$\mathrm{E}[\tilde{L}_{h,b}(X_j)] = hbr_0(x,x)\sum_{i=1}^{n} f_{i,j}(x,x) + O(nhb(b^2 + h^2)). \tag{B.11}$$

Likewise, assuming (P2) we obtain for $j \neq j'$,

$$\mathrm{E}[\overline{\tilde{L}_{h,b}(X_j)}\ \overline{\tilde{L}_{h,b}(X_{j'})}] = O(n^2h^2b^2) + O(n^2h^2b^2c(j,j')) + O(n^{2-\alpha_X/2}h^2b^2). \tag{B.12}$$

Under (P1), for $j \neq j'$,

$$\mathrm{E}[\overline{\tilde{L}_{h,b}(X_j)}\ \overline{\tilde{L}_{h,b}(X_{j'})}] = O(n^2h^2b^2(h^2 + b^2)), \tag{B.13}$$

where $c(j,j') = |j - j'|^{-\alpha_X/2}$ (cf. Lemma A.1).

Bounds on $S_{h,b}$

Recall the definition of $S_{h,b}(x, X_j, X_l)$ in (6.37). Let

$$r_1(x_i, x_j, x_l) := \frac{\sigma(x_j)\sigma(x_l)}{f^2(x_i)}.$$

Let $f_{i,j,l,i',j',l'}$ be the joint density of $(X_i, X_j, X_l, X_{i'}, X_{j'}, X_{l'})$.

If all indices j, j', l, l' are different, then recalling the notation from Lemma A.1 and using Lemma A.3 we have,

$$\mathrm{Cov}\,(S_{h,b}(x, X_j, X_l), S_{h,b}(x, X_{j'}, X_{l'})) = o(n^2h^2b^4) + O(n^2h^2b^4c(j,l,j',l')). \tag{B.14}$$

Appendix C: Martingale CLT

Here, we prove (6.11). Note that

$$\mathrm{Var}(J_{21}) \sim \frac{1}{nhf(x)}\sigma^4(x)\mathrm{E}[\xi_1^2]\int K^2(u)\,du.$$

The proof is similar to [36, Lemma 2] and [22, Lemma 3.1]. Let $R_i = (nh)^{-1/2}\sigma^2(X_i)K_h(x-X_i)\xi_i/\sqrt{f(x)}$ and $\bar{R}_i = R_i - \mathrm{E}[R_i|\mathcal{F}_{i-1}]$. From the martingale central limit theorem it suffices to show the Lindeberg condition

$$\sum_{i=1}^{n} \mathrm{E}\left[\bar{R}_i^2 1_{\{|\bar{R}_i|>\delta\}}\right] \to 0 \quad \text{for each } \delta > 0$$

and the convergence of the conditional variances

$$\sum_{i=1}^{n} \mathrm{E}[\bar{R}_i^2|\mathcal{F}_{i-1}] \xrightarrow{\mathrm{P}} 1.$$

Let f_X and g_ξ be the density of X_1 and ξ_1, respectively. As for the Lindeberg condition we have

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{E}\left[\bar{R}_i^2 1_{\{|\bar{R}_i|>\delta\}}\right] &\le 4\sum_{i=1}^{n} \mathrm{E}\left[R_i^2 1_{\{|R_i|>\delta\}}\right] \\
&= C_0 \frac{1}{nh}\sum_{i=1}^{n} \iint \sigma^2(u)K_h^2(x-u)f_X(u)v^2 g_{\varepsilon_1}(v)1_{\{|v|>C_1\delta\sqrt{nh}\}} \\
&\le C_2 \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\xi_i^2 1_{\{|\xi_i|>C_1\delta\sqrt{nh}\}}\right] \to 0,
\end{aligned}$$

where $C_0 = 1/(f(x))$, $C_1 = \left(\sqrt{C_0}\sup K(x)\right)^{-1}$ and $C_2 = C_0 \int K^2$.

As for the conditional variances note first that

$$\mathrm{E}[\bar{R}_i^2|\mathcal{F}_{i-1}] = \mathrm{E}[R_i^2|\mathcal{F}_{i-1}] - \mathrm{E}\left[(\mathrm{E}[R_i|\mathcal{F}_{i-1}])^2\right]$$

and note that the second term is of a smaller order than the first one. Now,

$$\begin{aligned}
&\sum_{i=1}^{n} \left\{\mathrm{E}[R_i^2|\mathcal{F}_{i-1}] - \mathrm{E}[R_i^2]\right\} \\
&= \frac{1}{nhf(x)}\mathrm{E}[\sigma^2(X_1)K_h^2(x-X_1)] \sum_{i=1}^{n} \left\{\mathrm{E}[\xi_i^2|\mathcal{F}_{i-1}] - \mathrm{E}[\xi_i^2]\right\} \\
&= \left(\frac{1}{f(x)}\int \sigma^2(v)K^2(v)f(x-vh)dv\right)\frac{1}{n}\sum_{i=1}^{n} \left\{\mathrm{E}[\xi_i^2|\mathcal{F}_{i-1}] - \mathrm{E}[\xi_i^2]\right\}.
\end{aligned}$$

Now, the deterministic term in the bracket is asymptotically equal to 1. The second part converges to 0 in probability from ergodicity. Consequently, the second Lindeberg condition is proven. □

Acknowledgement

This work was started while the second author visited the Department of the University of Ottawa in March 2009. The visit was partially supported by the Robert Bosch Stiftung. The work of the first author was supported by a NSERC Canada grant.

References

[1] Florin Avram and Murad S. Taqqu. Noncentral limit theorems and Appell polynomials. *Ann. Probab.*, 15(2):767–775, 1987. MR0885142

[2] J. Beran and Y. Feng. Local polynomial estimation with a FARIMA-GARCH error process. *Bernoulli*, 7(5):733–750, 2001. MR1867080

[3] T. Tony Cai and Lie Wang. Adaptive variance function estimation in heteroscedastic nonparametric regression. *Ann. Statist.*, 36(5):2025–2054, 2008. MR2458178

[4] Ngai Hang Chan and Shiqing Ling. Residual empirical processes for long and short memory time series. *Ann. Statist.*, 36(5):2453–2470, 2008. MR2458194

[5] Bing Cheng and P. M. Robinson. Density estimation in strongly dependent nonlinear time series. *Statist. Sinica*, 1(2):335–359, 1991. MR1130123

[6] Fuxia Cheng. Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *J. Statist. Plann. Inference*, 128(2):327–349, 2005. MR2102762

[7] Sándor Csörgő and Jan Mielniczuk. Random-design regression under long-range dependent errors. *Bernoulli*, 5(2):209–224, 1999. MR1681695

[8] Sándor Csörgő and Jan Mielniczuk. The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. *Statist. Sinica*, 10(3):771–787, 2000. MR1787779

[9] Sam Efromovich. Estimation of the density of regression errors. *Ann. Statist.*, 33(5):2194–2227, 2005. MR2211084

[10] Jianqing Fan and Qiwei Yao. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660, 1998. MR1665822

[11] Jianqing Fan and Qiwei Yao. *Nonlinear time series*. Springer Series in Statistics. Springer-Verlag, New York, 2003. Nonparametric and parametric methods. MR1964455

[12] C. Fraley, F. Leisch, and M. Maechler. `fracdiff`*: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models*, 2006. `R` package version 1.3-1.

[13] Jiti Gao and Qiying Wang. Long-range dependent time series specification. 2009.

[14] Liudas Giraitis and Donatas Surgailis. Central limit theorem for the empirical process of a linear sequence with long memory. *J. Statist. Plann. Inference*, 80(1-2):81–93, 1999. MR1713796

[15] Hongwen Guo and Hira L. Koul. Asymptotic inference in some heteroscedastic regression models with long memory design and errors. *Ann. Statist.*, 36(1):458–487, 2008. MR2387980

[16] Peter Hall, Soumendra Nath Lahiri, and Jörg Polzehl. On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.*, 23(6):1921–1936, 1995. MR1389858

[17] Peter Hall, Soumendra Nath Lahiri, and Young K. Truong. On bandwidth choice for density estimation with dependent data. *Ann. Statist.*, 23(6):2241–2263, 1995. MR1389873

[18] Hwai-Chung Ho and Tailen Hsing. On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.*, 24(3):992–1024, 1996. MR1401834

[19] Hwai-Chung Ho and Tailen Hsing. Limit theorems for functionals of moving averages. *Ann. Probab.*, 25(4):1636–1669, 1997. MR1487431

[20] Hira Koul. *Weighted Empirical and Residual Processes in Dynamical Nonlinear Models*. Springer Series in Statistics. Springer-Verlag, New York, 2002. MR1911855

[21] Hira L. Koul and Donatas Surgailis. Goodness-of-fit testing under long memory. *J. Statist. Plann. Inference*, 140(12):3742–3753, 2010. MR2674162

[22] Rafał Kulik. Nonparametric deconvolution problem for dependent sequences. *Electron. J. Statist.*, 2:722–740, 2008. MR2430252

[23] Rafal Kulik. Empirical process of long-range dependent sequences when parameters are estimated. *J. Statist. Plann. Inference*, 139(2):287–294, 2009. MR2474005

[24] Rafal Kulik and Pawel Lorek. Empirical process of residuals for regression models with long memory errors. *Preprint*, 2010.

[25] Rafal Kulik and Pawel Lorek. Some results on random design regression with long memory errors and predictors. *J. Statist. Plann. Inference*, 141(1):508–523, 2011. MR2719514

[26] Rafal Kulik and Cornelia Wichelhaus. Conditional variance estimation in regression models with long memory. 2010.

[27] S. N. Lahiri. *Resampling methods for dependent data*. Springer Series in Statistics. Springer-Verlag, New York, 2003. MR2001447

[28] Han-Ying Liang and Si-Li Niu. Asymptotic properties of error density estimator in regression model under α-mixing assumptions. *Comm. Statist. Theory Methods*, 38(6-7):761–783, 2009. MR2522528

[29] Jan Mielniczuk and Wei Biao Wu. On random-design model with dependent errors. *Statist. Sinica*, 14(4):1105–1126, 2004. MR2126343

[30] Ursula U. Müller, Anton Schick, and Wolfgang Wefelmeyer. Estimating the error distribution function in semiparametric regression. *Statist. Decisions*, 25(1):1–18, 2007. MR2370101

[31] Ursula U. Müller, Anton Schick, and Wolfgang Wefelmeyer. Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.*, 79(7):957–964, 2009. MR2509488

[32] Natalie Neumeyer and Ingrid Van Keilegom. Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.*, 101(5):1067–1078, 2010. MR2595293

[33] D. Ruppert, S. J. Sheather, and M. P. Wand. An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, 90(432):1257–1270, 1995. MR1379468

[34] Murad S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:287–302, 1974/75. MR0400329

[35] Lie Wang, Lawrence D. Brown, T. Tony Cai, and Michael Levine. Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.*, 36(2):646–664, 2008. MR2396810

[36] Wei Biao Wu and Jan Mielniczuk. Kernel density estimation for linear processes. *Ann. Statist.*, 30(5):1441–1459, 2002. MR1936325

[37] Zhibiao Zhao and Wei Biao Wu. Confidence bands in nonparametric time series regression. *Ann. Statist.*, 36(4):1854–1878, 2008. MR2435458