Conditional Variance Estimation in Regression Models with Long Memory

Rafał Kulik[*]   Cornelia Wichelhaus [†]

October 11, 2011

Abstract

This paper studies asymptotic properties of a nonparametric kernel estimator of the conditional variance in a random design model with parametric mean and heteroscedastic errors, for a class of long memory errors and predictors. We establish small and large bandwidths asymptotics, which show a different behaviour compared to that of kernel estimators of the conditional mean. We distinguish between an oracle case (i.e., where the errors are directly observed) and a non-oracle case (where the errors are replaced with residuals), and show non-equivalence between the oracle and non-oracle case. We also discuss a practical problem of bandwidth choice. Theoretical results are justified by simulation studies. We apply our theory to DJA and FTSE indices.

1   Introduction

Consider the random design regression model,

$$Y_i = \beta_0 + \beta_1 X_i + \sigma(X_i)\varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

with intercept β_0 and slope β_1. We will study the model (1.1) in the presence of long memory behaviour of the errors and/or predictors. Our goal is to estimate the conditional variance $\sigma^2(\cdot)$ in a nonparametric way. To do so, we first estimate β_0 and β_1 by usual least squares estimators. Then, we estimate $\sigma^2(\cdot)$ by smoothing residuals with a kernel K and a bandwidth h:

$$\hat{\sigma}^2(x) = \hat{\sigma}_h^2(x) = \frac{1}{nh\hat{f}_h(x)} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 K\left(\frac{x - X_i}{h}\right), \tag{1.2}$$

---

[*]Department of Mathematics & Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa ON K1N 6N5, Canada; e-mail: rkulik@uottawa.ca

[†]Institute of Applied Mathematics, University of Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany; e-mail: wichelhaus@statlab.uni-heidelberg.de

where

$$\hat{f}_h(x) = \hat{f}_{h,X}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{1.3}$$

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 $Y_i = m(x_i) + \sigma(x_i)\varepsilon_i$, where $x_i = i/n$ and ε_i, $i = 1, \ldots, n$, are i.i.d. In this set-up it was shown in Wang *et al.* (2008) that if $m(\cdot)$ is differentiable, then there is no influence of the estimation of $m(\cdot)$ on the minimax rates of convergence for estimation of $\sigma(\cdot)$. In case of random design regression, a similar lack of influence was proven in Fan and Yao (1998) for weakly dependent data. However, very little is known in case of long memory errors and/or predictors. In case of long memory predictors and i.i.d. errors, it was shown in Zhao and Wu (2008) that there is no influence of estimating the mean, however, this is basically due to their imposed conditions on the bandwidth choice: for a *small* bandwidth nonparametric estimation behaves as if data were independent. On the other hand, in Guo and Koul (2008) the authors studied the model (1.1) and established a *large bandwidth* behaviour, i.e. when $h \to 0$ sufficiently slow.

The goal of this paper is to present the full asymptotic theory for the conditional variance estimation in the model (1.1), when errors and/or predictors have long memory. Such situations are very often encountered in financial time series. We will distinguish between an *oracle* and a *non-oracle* case. In the first situation, we assume that β_0 and β_1 are known, which amounts to estimation of $\sigma(\cdot)$ from direct observations $\sigma^2(X_i)\varepsilon_i^2$. In other words, the oracle estimator is defined as

$$\hat{\sigma}^2_{\text{oracle}}(x) = \frac{1}{nh\hat{f}_h(x)} \sum_{i=1}^{n} \sigma^2(X_i)\varepsilon_i^2 K\left(\frac{x - X_i}{h}\right). \tag{1.4}$$

By applying log-transformation, we may see that the problem is similar to nonparametric estimation of a conditional mean. In the latter situation, we refer to Mielniczuk and Wu (2004) for the most general results. One has to mention that the rates of convergence for the conditional variance differ when compared to the conditional mean, see Remark 3.4. The results for the oracle case are given in Section 3.1.

In Section 3.2 we state the results which show the influence of estimating β_0 and β_1 on estimation of $\sigma(\cdot)$. If the errors ε_i, $i \geq 1$, are i.i.d., the oracle and non-oracle case are the same, regardless whether the predictors are LRD. This agrees with the findings in Fan and Yao (1998). There, the predictors and errors are weakly dependent, and conditional mean is estimated in a nonparametric way. However, if the errors are LRD, there are two additional terms which may contribute to the limiting behavior of the conditional variance estimator. The first one comes from estimating β_1 and disappears if $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] = 0$. The second one comes from estimating β_0 and *always* contributes.

In Section 3.3 we discuss problem of bandwidth choice. We note that "large" bandwidths are not "practical", since they lead to a complicated LRD-type behaviour. Therefore, it is of practical interest to verify if a chosen bandwidth (for example, by implementing the plug-in method) leads to i.i.d. type behaviour.

We extend results in Guo and Koul (2008) in several directions. First, we establish a "practical", small bandwidth asymptotics. Furthermore, we allow predictors to be follow an EGARCH-type model. This allow us to implement our theory to "typical" financial time series (see Section 4), which cannot be modeled by linear LRD processes. In particular, we apply our theory do Dow Jones Composite Average and FTSE Indices, which have a "typical" behaviour: log-returns are uncorrelated, but residuals have long memory.

We would like to mention that our theory is further developed in Kulik and Wichelhaus (2011). There, we study the limiting behaviour of $\hat{\sigma}_h^2$, when the conditional mean is estimated in a nonparametric way. Furthermore, we estimate the error density.

2 Preliminaries

2.1 Predictors and the error sequence

Throughout the paper it is assumed that the predictors X_i, $i \geq 1$, are independent of the errors ε_i, $i \geq 1$. We consider the following assumptions on the predictors X_i, $i \geq 1$:

(P1) X_i, $i \geq 1$, are i.i.d. random variables with $\mathrm{E}X_1 = \mu < \infty$ and $\mathrm{Var}X_1 = 1$.

(P2) X_i, $i \geq 1$, is EGARCH model defined as $X_i = Z_i\varphi(U_i)$, $i \geq 1$. Here, Z_i is a centered sequence of random variables, $\varphi(\cdot)$ is a real valued function and U_i, $i \geq 1$, is a sequence of standard Gaussian random variables with covariance $\mathrm{E}(U_1U_{k+1}) = k^{-\alpha_U}L(k)$, where $\alpha_U \in (0,1)$ and L is slowly varying at infinity. Then U_i can be written as $\sum_{k=1}^{\infty} \tilde{c}_k\zeta_{-k}$. We will assume that (ζ_i, Z_i), $i \geq 1$, are i.i.d., with $\mathrm{Var}Z_1 = \mathrm{Var}\varphi(U_1) = 1$.

In particular, random variables Z_i can be degenerated with total mass at 1; in this case $X_i = \varphi(U_i)$ is a subordinated Gaussian sequence.

We shall consider similar assumptions on the error sequence:

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

(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 a finite fourth moment, $\mathrm{E}[\varepsilon_1^2] = 1$, and for some $\alpha_\varepsilon \in (0,1)$, $c_k \sim C_0 k^{-(\alpha_\varepsilon+1)/2}$ as $k \to \infty$.

3

According to different assumptions, we will apply different techniques. If (E1) and (P2) hold, then we will exploit martingale structure of particular sequences. In this case the normality assumption in (P2) is not necessary in fact. If (E2) and (P1) hold, we will decompose random quantities involved into a martingale part and a long memory part. Under (P2) and (E2), we will exploit normality and Hermite expansions. These different techniques seem to be most efficient.

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, such that $\kappa_0 = 1$, $\kappa_1 = 0$, $\kappa_2 \neq 0$. It is also assumed that $K(\cdot)$ is bounded and continuous. Denote for future use $K_h(\cdot) := K(\cdot/h)$.

For a given random variable V, let f_V denotes its density. Consider the following assumptions on $f = f_X$, σ and the bandwidth h.

(D1) f, σ are defined on $\mathbb{R}$ with $f, \sigma \in C^2(\mathcal{I})$ for some finite interval $\mathcal{I}$, where C^2 is the class of twice-differentiable functions, with bounded and continuous second order derivatives. Also, $\inf_{x \in \mathcal{I}} f(x) > 0$.

(D2) $\sigma(x) > 0$ for all $x \in \mathbb{R}$ and $\mathrm{E}[\sigma^4(X)] < \infty$.

(D3) $h + \sqrt{nh}h^2 \to 0$, $nh \to \infty$.

2.3 Long memory Gaussian sequences

To introduce further assumptions, let us recall some facts on LRD sequences and Hermite coefficients. A function $G(\cdot)$ in $L^2(\phi)$, $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, such that $\mathrm{E}[G(U_1)] = 0$, can be expanded as

$$G(\cdot) = \sum_{m=1}^{\infty} \frac{J_m}{m!} H_m(\cdot),$$

where $J_m = \mathrm{E}(G(U_1) H_m(U_1))$ and $H_m(\cdot)$ is the mth Hermite polynomial. The smallest $q \geq 1$ such that $J_q \neq 0$ is called the Hermite rank of G. In particular, let $J_m(x) = \mathrm{E}[1_{\{\varphi(U_1) \leq x\}} H_m(U_1)]$. Then

$$1_{\{\varphi(\cdot) \leq x\}} - F(x) = \sum_{m=q(x)}^{\infty} J_m(x) \frac{H_m(\cdot)}{m!},$$

with some $q(x) \in \mathbb{N}$. Let q be the Hermite rank of class of functions $1_{\{\varphi(\cdot) \leq x\}} - F(x)$, $x \in \mathbb{R}$, i.e.

$$q = \min\{q(x) : J_{q(x)}(x) \neq 0 \quad \text{for some } x \in \mathbb{R}\}. \tag{2.1}$$

The corresponding coefficient is denoted by $J_q(x)$. We shall assume that

(J1) $J_q'(\cdot)$ is uniformly bounded and uniformly continuous.

4

(J2) The following function is in $L^2(\phi)$:

$$e(\cdot; x) \rightarrow \frac{1}{\varphi(\cdot)} f_Z\left(\frac{x}{\varphi(\cdot)}\right). \tag{2.2}$$

Remark 2.1. Condition (J1) is related to the Gaussian sequence U_i and the function φ. It is imposed in Csörgő and Mielniczuk (1995) in case of density estimation for subordinated Gaussian sequences. It is fulfilled for example when $\varphi(u) = u$, since then $q = 1$, $J_1(x) = \int_{-\infty}^x s\phi(s)ds$, and $J_1'(x) = x\phi(x)$.

Assumption (J2) links the Gaussian sequence U_i, the function φ and random variables Z_i. In particular, (J2) implies that quantities

$$J_m^*(x) := \mathrm{E}\left[J_m'\left(\frac{x}{Z_1}\right)\frac{1}{Z_1}\right]$$

are well defined for all $m \geq q$.

It is clearly fulfilled if f_Z is degenerated, i.e. X_i is subordinated Gaussian sequence. Furthermore, if $\varphi(u) = u$ and f_Z does not vanish at 0, then (2.2) is equivalent to $\mathrm{E}[U^{-2}] < \infty$. However, the latter is not true. Thus, $x = 0$ has to be excluded from our analysis. On the other hand, if f_Z is bounded with bounded support, say, $(-M, M)$, then for $x \in (-M, M) \setminus \{0\}$,

$$\mathrm{E}\left[U^{-2}f_Z^2\left(\frac{x}{U}\right)\right] \leq C\frac{1}{x^2 M^2} \leq \frac{C}{M^4}.$$

Furthermore, if f_Z is the standard normal density, then (J2) is fulfilled. Indeed, as $u \rightarrow 0$, $u^{-2}f_Z^2(x/u) \sim Cu^{-2}\exp(-(x/u)^4/2) \rightarrow 0$, so that $u^{-2}f_Z^2(x/u)\phi(u)$ is integrable.

Further comments on (J2) and its relation to (J1) are given in Appendix.

Let us also recall that for if $q\alpha_U < 1$,

$$\frac{1}{a_{n,q}}\sum_{j=1}^n \frac{1}{q!}H_q(U_i) \xrightarrow{\mathrm{d}} L_q, \tag{2.3}$$

where $a_{n,q}^2 \sim n^{2-q\alpha_U}L^q(n)/C^2(q, \alpha_U)$, $C(q, \alpha_U) = \sqrt{(1 - q\alpha_U)(2 - q\alpha_U)/(2q!)}$,

$$L_q = (q!(1 - q\alpha_U)^{-1/2}Z_q(1) \tag{2.4}$$

and $Z_q(t)$ is the so-called Hermite or Rosenblatt process of order q, defined as a q-fold stochastic integral

$$Z_q(t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\mathrm{e}^{it(x_1 + \cdots + x_q)} - 1}{x_1 + \cdots + x_q}\prod_{i=1}^q x_i^{(1-\alpha_U)/2}\, W(dx_1)\ldots W(dx_q),$$

where W is an independently scattered Gaussian random measure with Lebesgue control measure. If $q\alpha_U > 1$, then $\sum_{i=1}^n H_q(U_i) = O_P(\sqrt{n})$.

For more details on Gaussian LRD sequences the reader is referred to Taqqu (2003).

5

2.4 Limit theorems for long memory linear processes

First, let us introduce the following σ-fields: let $\mathcal{X}_i$ be the σ-field generated by $(Z_i, \zeta_{i-1}, Z_{i-1}, \zeta_{i-2} \ldots)$ and let $\mathcal{H}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$. Under the assumption (E2) we have

$$b_n^2 := \mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i\right) \sim C_1^2 n^{2-\alpha_\varepsilon}, \qquad \alpha_\varepsilon \in (0,1), \tag{2.5}$$

with

$$C_1^2 = \frac{2C_0^2}{(1-\alpha_\varepsilon)(2-\alpha_\varepsilon)} \int_0^\infty (x+x^2)^{-(\alpha_\varepsilon+1)/2} \, dx. \tag{2.6}$$

Furthermore, (see Davydov (1970))

$$b_n^{-1} \sum_{i=1}^n \varepsilon_i \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{2.7}$$

Let $\xi_i = \varepsilon_i^2 - \mathrm{E}[\varepsilon_i^2] = \varepsilon_i^2 - 1$. One can verify that with some $C > 0$,

$$\mathrm{Cov}(\xi_0, \xi_i) \sim C^2 \mathrm{Cov}^2(\varepsilon_0, \varepsilon_j), \tag{2.8}$$

see e.g. Guo and Koul (2008, Lemma 4.1). Thus, for $d_n^2 := \mathrm{Var}\left(\sum_{i=1}^n \xi_i\right)$, we have

$$d_n^2 \sim \begin{cases} C_2^2 n^{2(1-\alpha_\varepsilon)}, & \text{if } \alpha_\varepsilon < 1/2, \\ C_3^2 n, & \text{if } \alpha_\varepsilon > 1/2, \end{cases} \tag{2.9}$$

with some constants C_2, C_3. Also,

$$\mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i^2\right) \sim \begin{cases} C_2^2 n^{2(1-\alpha_\varepsilon)}, & \text{if } \alpha_\varepsilon < 1/2, \\ C_4^2 n, & \text{if } \alpha_\varepsilon > 1/2, \end{cases} \tag{2.10}$$

with a possibly different constant C_4. Moreover, (see Avram and Taqqu (1987, Theorem 2),

$$d_n^{-1} \sum_{i=1}^n \xi_i \xrightarrow{\mathrm{d}} Z_2(1), \qquad d_n^{-1} \sum_{i=1}^n \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] \xrightarrow{\mathrm{d}} Z_2(1), \qquad \text{if } \alpha_\varepsilon < 1/2, \tag{2.11}$$

where H_2 is a Hermite-Rosenblatt random variable. If $\alpha_\varepsilon > 1/2$, then the above limits are standard normal with $\sqrt{n}$-normalization.

3 Results

3.1 Conditional variance: oracle case

Here, we consider behaviour of the oracle estimator defined in (1.4).

6

Proposition 3.1. *Assume* (E1) *and either* (P2) *or* (P1). *Suppose that assumptions of Section* 2.2 *and* (J1)-(J2) *are fulfilled. Then*

$$\sqrt{nhf(x)} \left(\hat{\sigma}_{\text{oracle}}^2(x) - \sigma^2(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^4(x)\left(\mathrm{E}[(\varepsilon_1^2 - 1)^2]\right)\int K^2\right). \quad (3.1)$$

Proposition 3.2. *Assume* (E2) *and either* (P2) *or* (P1). *Suppose that assumptions of Section* 2.2 *and* (J1)-(J2) *are fulfilled. We have:*

- *If $hn^{(1-2\alpha_\varepsilon)} \to 0$, then* (3.1) *holds.*

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

$$n^{\alpha_\varepsilon}\left(\hat{\sigma}_{\text{oracle}}^2(x) - \sigma^2(x)\right) \xrightarrow{d} C_2\sigma^2(x)Z_2(1), \quad (3.2)$$

  where the constant C_2 and random variable $Z_2(1)$ are defined in (2.10) *and* (2.11), *respectively.*

Remark 3.3. In Proposition 3.1 there is no influence of LRD in predictors. It is in contrast with results in Guo and Koul (2008). The reason for this is that we consider $\hat{f}_h$ in the definition of $\hat{\sigma}_h^2$ instead of f (or $\tilde{f}$, i.e. Gaussian density with estimated mean and variance). This effect is explained in details in Remark 5.2.

Remark 3.4. If $\alpha_\varepsilon \in (1/2, 1)$ there is no influence of LRD in errors on the rates of convergence, which is basically due to (2.9) below. If $\alpha_\varepsilon \in (0, 1/2)$ and h is *small* (i.e., when $hn^{(1-2\alpha_\varepsilon)} \to 0$ holds), then there is still no influence of LRD. However, if h is *big* (i.e., when $hn^{(1-2\alpha_\varepsilon)} \to \infty$ holds), LRD influences the limit. Note further that the meaning of *small* and *big* bandwidth is different than in case of estimating conditional mean in a nonparametric way. Namely, in the model $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$, for the standard kernel estimator of $m(\cdot)$, we have $\sqrt{nh}$ or $n^{\alpha_\varepsilon/2}$ rate of convergence if, respectively, $hn^{(1-\alpha_\varepsilon)} \to 0$ or $hn^{(1-\alpha_\varepsilon)} \to \infty$. We refer to Mielniczuk and Wu (2004) for more details.

3.2   Conditional variance: non-oracle case

In the non-oracle case, i.e., when β_0, β_1 are unknown and have to be estimated, the conditional variance is estimated using the kernel estimator (1.2). We will focus on cases (E1) and (E2)+(P1), to show an influence of LRD in the errors. The case of (E2)+(P2) is skipped here. It is a subject of Kulik and Wichelhaus (2011) in a more general context of nonparametric conditional mean.

Theorem 3.5. *Assume* (E1) *and either* (P2) *or* (P1). *Suppose that assumptions of Section* 2.2 *and* (J1)-(J2) *are fulfilled. Then*

$$\sqrt{nhf(x)} \left(\hat{\sigma}^2(x) - \sigma^2(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^4(x)\left(\mathrm{E}[(\varepsilon_1^2 - 1)^2]\right)\int K^2\right). \quad (3.3)$$

Theorem 3.6. *Assume* (E2) *and* (P1). *Suppose that assumptions of Section 2.2 and* (J1)-(J2) *are fulfilled.*

- *If $hn^{(1-2\alpha_\varepsilon)} \to 0$, then (3.3) holds.*

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

$$n^{\alpha_\varepsilon} \left(\hat{\sigma}^2(x) - \sigma^2(x)] \right) \xrightarrow{d} C_2 \sigma^2(x) Z_2(1) \tag{3.4}$$
$$+ C_1^2 \left(\mathrm{E}^2[\sigma(X_1)(X_1 - \mu)](x - \mu)^2 - 2\sigma(x)(x - \mu)\mathrm{E}[\sigma(X_1)(X_1 - \mu)] \right) \chi^2(1)$$
$$+ C_1^2 \left(\mathrm{E}^2[\sigma(X_1)] - 2\sigma(x)\mathrm{E}[\sigma(X_1)] \right) \chi^2(1),$$

where $\chi^2(1)$ is χ^2 random variable with 1 degree of freedom.

Remark 3.7. Comparing with Proposition 3.2, Theorem 3.6 reveals the influence of estimation of the linear regression parameters, when errors are LRD. If $\alpha_\varepsilon > 1/2$, then there is no difference between oracle and non-oracle case, and in fact the results are as in the case of i.i.d. errors. Second, there are two additional $O_P(n^{-\alpha_\varepsilon})$ terms in (3.4). The first one comes from estimating β_1 and vanishes if $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] = 0$. The second one comes from estimating β_0.

Remark 3.8. Once again, let us compare Theorem 3.6 with Guo and Koul (2008, Theorem 3.1), by setting $\alpha_X = 1$ there. Part (a) there is not applicable. Results in Part (b) there agree with (3.4) of Theorem 3.6.

Remark 3.9. The results of Propositions 3.1, 3.2 and Theorems 3.5, 3.6 can be formulated in a multivariate set-up. In case of (3.1) the limiting distribution of $(\hat{\sigma}^2(x_i) - \sigma^2(x_i), i = 1, \ldots, m)$ is asymptotically multivariate normal with independent components (it follows using the Cramer-Wold device). In case of (3.2) the limiting distribution is degenerate,

$$C_2(\sigma^2(x_i), i = 1, \ldots, m) Z_2(1),$$

in case of (3.4) the limiting distribution is also degenerate.

3.3 Bandwidth choice

Let us note first that from a practical point of view "large" bandwidths are not desirable. Indeed, in the context of (3.4) it is completely not clear how to estimate percentiles of the limiting distribution, since e.g. bootstrap does not work (see Lahiri (2003, Chapter 10)). Therefore, results in LRD zone (i.e., the results in (3.2) and (3.4)) are *not practical* and may serve as a warning against inappropriate choice of h.

In Kulik and Lorek (2011) the authors studied the problem of bandwidth choice for the conditional mean estimation. It was proven there that cross-validation may not be a valid procedure in case of long memory errors, and plug-in methods are preferable. In Section 4 we justify the latter statement

Table 1: Bandwidth choice for $\sigma(\cdot)$

d_ε		$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$
0	$\hat{h}$	0.1108	0.1725	0.2346	0.2639	0.2960
0.1	$\hat{h}$	0.1124	0.1725	0.2336	0.2660	0.2990
0.2	$\hat{h}$	0.1014	0.1828	0.2332	0.2666	0.3037
0.3	$\hat{h}$	0.1141	0.1750	0.2318	0.2633	0.2980
0.4	$\hat{h}$	0.1025	0.1754	0.2307	0.2613	0.2952

based on simulation studies. It is shown that there is little influence of LRD in errors on the plug-in bandwidth choice. Thus, in our implementation in Section 4 we apply the procedure:

- Estimate h using the plug-in method. Solve $n^{-\delta} = h$. Compare the obtained δ with $1 - 2\hat{\alpha}_\varepsilon$, where $\hat{\alpha}_\varepsilon$ is an estimator of α_ε; for example method from Guo and Koul (2008) leads to a consistent estimator. Verify if $hn^{1-2\alpha_\varepsilon} \to 0$ holds. In particular, there is nothing to verify if $\alpha_\varepsilon > 1/2$.

In this way the estimator of α_ε is not used to construct confidence intervals, rather to justify if we can use (3.1) to construct confidence intervals.

4 Numerical studies and data examples

4.1 Simulation studies

- We simulate $n = 1000$ observations from the models $Y_i = 0 + 2X_i + \sigma(X_i)\varepsilon_i$, with $\sigma(x) = \sqrt{x^2 + 1}$ 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 estimate β_0 and β_1 by LSE estimators. We compute the estimator $\hat{\sigma}_{\hat{h}}^2(\cdot)$ with $\hat{h}$ selected by the plug-in method.

- This procedure is repeated $M = 500$ times.

- Table 1 contains statistics for the bandwidth h selected by the plug-in method. There is little influence of the memory parameter.

4.2 Data Analysis

We study regression relationship between Dow Jones Composite Average (Y) and FTSE Index (predictor) from 1 January 2000 to 1 January 2010 (source: Yahoo Finance). From figure below we conclude that there is no correlation in log-returns, however, there is a strong correlation in squared log-returns. The memory parameter for the latter is estimated to be $d = 0.17$. Also, normality of the predictor can be assumed. Finally, the conditional variance is plotted. The bandwidth for the conditional variance was chosen using plug-in method.

Also, there is a little memory in residuals. This means that *pointwise* confidence intervals can be constructed using (3.1).

5 Proofs

5.1 Density estimation for SV models

Asymptotic behaviour of the kernel density estimator is well known in case of subordinated Gaussian or linear processes with LRD, see e.g. Csörgő and Mielniczuk (1995), Wu and Mielniczuk (2002). Here, we establish it under (P2). Consistency is needed later.

Lemma 5.1. *Suppose that assumptions of Section 2.2 and* (J1)-(J2) *are fulfilled.*

- *If* $h\, a_{n,q}^2/n \to 0$, *then*

$$\sqrt{nh}\left(\hat{f}_h(x) - f(x)\right) \overset{d}{\to} \mathcal{N}\left(0, f(x)\int K^2(u)\, du\right).$$

- *If* $h\, a_{n,q}^2/n \to \infty$, *then*

$$\frac{n}{a_{n,q}}\left(\hat{f}_h(x) - f(x)\right) \overset{d}{\to} J_q^*(x)L_q.$$

Proof. We decompose

$$\frac{1}{nh}\sum_{i=1}^n \left\{K_h(x - X_i) - f(x)\right\} = \left\{h^{-1}\mathrm{E}[K_h(x - X_1)] - f(x)\right\}$$

$$+\frac{1}{nh}\sum_{i=1}^n \left\{K_h(x - X_i) - \mathrm{E}\left[K_h(x - X_i)|\mathcal{X}_{i-1}\right]\right\} +$$

$$+\frac{1}{nh}\sum_{i=1}^n \left\{\mathrm{E}\left[K_h(x - X_i)|\mathcal{X}_{i-1}\right] - \mathrm{E}\left[K_h(x - X_i)\right]\right\} =: A_1 + A_2 + A_3. \, (5.1)$$

The first part in the above decomposition is the bias and is of order h^2, due to assumptions on K and f. The second part is a martingale, so that from martingale CLT (see Appendix) we conclude

$$\sqrt{nh}A_2 \overset{d}{\to} \mathcal{N}\left(0, f(x)\int K^2(u)\, du\right)$$

It remains to deal with LRD part A_3. Since K is bounded, a function

$$(\cdot) \to \int K_k(x - z\varphi(\cdot))f_Z(z)\, dz$$

10

is in $L^2(\phi)$. Applying the Hermite expansion,

$$\frac{1}{nh}\sum_{i=1}^{n}\left\{\mathrm{E}\left[K_h(x-X_i)|\mathcal{X}_{i-1}\right]-\mathrm{E}\left[K_h(x-X_i)\right]\right\} \tag{5.2}$$

$$= \int \frac{1}{nh}\sum_{i=1}^{n}\left\{K_h(x-z\varphi(U_i))-\mathrm{E}[K_h(x-z\varphi(U_i))]\right\}f_Z(z)\ dz$$

$$=: \int A_3(x;z)f_Z(z)\ dz.$$

The next few steps are similar to Csörgő and Mielniczuk (1995). Let $\hat{F}_n(\cdot)$ be the empirical distribution function associated with random variables $\varphi(U_1),\ldots,\varphi(U_n)$. Define the corresponding empirical processes

$$t_n(y)=\frac{n}{a_{n,q}}\left(\hat{F}_n(y)-F(y)\right),\quad \tilde{t}_n(y)=t_n(y)-J_q(y)\frac{1}{a_{n,q}}\sum_{i=1}^{n}\frac{1}{m!}H_m(U_i).$$

Then

$$\frac{n}{a_{n,q}}A_3(x;z)=\frac{1}{h}\int K\left(\frac{x-zy}{h}\right)dt_n(y)=\frac{z}{h^2}\int t_n(y)K'\left(\frac{x-zy}{h}\right)\ dy$$

$$= \frac{1}{h}\int \tilde{t}_n\left(\frac{x-hu}{z}\right)K'(u)\ du+\frac{1}{h}\frac{1}{a_{n,q}}\sum_{i=1}^{n}\frac{H_q(U_i)}{q!}\int J_q\left(\frac{x-hu}{z}\right)K'(u)\ du.$$

As in Csörgő and Mielniczuk (1995), the first part converges almost surely to 0 (uniformly in z, x). Thus

$$\sup_{x}\left|\frac{n}{a_{n,q}}\int A_3(x;z)f_Z(z)\ dv\right.$$

$$\left.-\frac{1}{a_{n,q}}\sum_{i=1}^{n}\frac{H_q(U_i)}{q!}\frac{1}{h}\iint J_q\left(\frac{x-hu}{z}\right)K'(u)\ f_Z(z)\ du\ dz\right|=o_P(1).$$

Noting that

$$\frac{1}{h}\iint J_q\left(\frac{x-hu}{z}\right)K'(u)f_Z(z)\ du\ dz$$

$$= \frac{1}{h}\int K(u)J_q'\left(\frac{x-hu}{z}\right)f_Z(z)\ du\ dz\sim\int\frac{1}{z}J_q'(x/z)f_Z(z)\ dz=J_q^*(x),$$

the result follows by (2.3) and assumption (J2). $\qquad\square$

5.2 Decomposition of the conditional variance estimator

For $i \geq 1$, let $\Delta_i = \left(\hat{\beta}_0 - \beta_0 \right) + \left(\hat{\beta}_1 - \beta_1 \right) X_i =: \Delta_0 + \Delta_{1,i}$. Decompose

$$
\begin{aligned}
\left\{ \hat{\sigma}^2(x) - \sigma^2(x) \right\} &= \left(\frac{1}{nh\hat{f}_h(x)} \sum_{i=1}^{n} \sigma^2(X_i) K_h(x - X_i) - \sigma^2(x) \right) \\
&\quad + \frac{1}{nh\hat{f}_h(x)} \sum_{i=1}^{n} \sigma^2(X_i) K_h(x - X_i) \left(\varepsilon_i^2 - 1 \right) \\
&\quad - \frac{2}{nh\hat{f}_h(x)} \sum_{i=1}^{n} \Delta_i \sigma(X_i) K_h(x - X_i) \varepsilon_i + \frac{1}{nh\hat{f}_h(x)} \sum_{i=1}^{n} \Delta_i^2 K_h(x - X_i) \\
&=: M_1 + M_2 - M_3 + M_4.
\end{aligned} \tag{5.3}
$$

In the oracle case, $\Delta_i = 0$ and thus, $M_3 = M_4 \equiv 0$. Therefore, for the oracle estimator we have the following decomposition:

$$
\left\{ \hat{\sigma}^2_{\text{oracle}}(x) - \sigma^2(x) \right\} = M_1 + M_2. \tag{5.4}
$$

It will be shown that under assumption (D3),

$$
\sqrt{nh} M_1 = o_P(1). \tag{5.5}
$$

Then, we will deal with the term M_2. Under (E1), it behaves as follows:

$$
\sqrt{nh f(x)} M_2 \xrightarrow{\mathrm{d}} \mathcal{N} \left(0, \sigma^4(x) \mathrm{E}[\xi_1^2] \int K^2 \right). \tag{5.6}
$$

Under (E2), (5.6) holds if $h n^{(1-2\alpha_\varepsilon)} \to 0$. If however (E2)+(P1) holds and $h n^{(1-2\alpha_\varepsilon)} \to \infty$,

$$
n^{\alpha_\varepsilon} M_2 \xrightarrow{\mathrm{d}} C_2 \sigma^2(x) Z_2(1). \tag{5.7}
$$

If (E2)+(P2) holds and $h n^{(1-2\alpha_\varepsilon)} \to \infty$,

$$
n^{\alpha_\varepsilon} M_2 \xrightarrow{\mathrm{d}} C_2 \sigma^2(x) \frac{J_0^*(x)}{f(x)} Z_2(1). \tag{5.8}
$$

The decomposition (5.4) together with asymptotic results (5.5)-(5.8) yield Propositions 3.1-3.2.

In the non-oracle case, under (E1) (for both (P1) and (P2)) the terms M_3 and M_4 are negligible. Under (E2) and (P1), we show that they contribute as follows:

$$
n^{\alpha_\varepsilon} M_3 \xrightarrow{\mathrm{d}} 2C_1^2 \sigma(x) \left(\mathrm{E}[\sigma(X_1)(X_1 - \mu)](x - \mu) + \mathrm{E}[\sigma(X_1)] \right) \times \chi^2(1), \tag{5.9}
$$

$$
n^{\alpha_\varepsilon} M_4 \xrightarrow{\mathrm{d}} C_1^2 \mathrm{E}^2[\sigma(X_1)] \chi^2(1) + (x - \mu)^2 \omega^2 \chi^2(1), \tag{5.10}
$$

where $\omega^2 := C_1^2 \mathrm{E}^2[\sigma(X_1)(X_1 - \mu)]$. Since the convergence in (5.9) and (5.10) is joint, these together with (5.5)-(5.8) and the decomposition (5.3) of the variance estimator, will imply the results of Theorem 3.5 and Theorem 3.6.

It is worth to point out that asymptotic behaviour of terms M_1, M_2, M_3, M_4 is not of particular use, unless we consider them as part of decompositions (5.3)-(5.4). The same apply to terms in different decompositions below.

5.3 Proof of Propositions 3.1 and 3.2

As noted above (see (5.4)), in order to prove Propositions 3.1 and 3.2 it is enough to deal with terms M_1 and M_2 (see (5.5) and (5.6), respectively). Write M_1 as

$$M_1 = \left(\frac{1}{\hat{f}_h(x)} - \frac{1}{f(x)} \right) \frac{1}{nh} \sum_{i=1}^{n} \left(\sigma^2(X_i) - \sigma^2(x) \right) K_h(x - X_i)$$

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

By Lemma 5.1, $\hat{f}_h$ is consistent estimator of f. Therefore, the first part in the expression above is dominated by the second one. Since (D1) holds, we have $\mathrm{E}|M_{11}| = O(h^2)$. Therefore, $\sqrt{nh}M_1 = o_P(1)$ by (D3) and (5.5) follows.

Next, we deal with M_2. We will prove (5.6), (5.7) and (5.8) under different sets of assumptions. In order to do this, we consider $\hat{f}_h$ replaced with f. This is allowed, since $\hat{f}_h$ is the weakly consistent estimator of f, and K has finite support.

Under (E1), M_2 is a martingale w.r.t. $\mathcal{G}_n$, $n \geq 1$, where $\mathcal{G}_n$ is a sigma field generated by $(Z_i, \eta_i, \zeta_{i-1}, Z_{i-1}, \eta_{i-1}, \ldots)$. Hence, from martingale CLT (see Appendix) we conclude that (5.6) holds.

Assume now (E2) and (P1). Recall that $\xi_i = \varepsilon_i^2 - 1$ and define $\mathcal{H}_i$ to be a sigma field generated by $(\eta_i, \eta_{i-1}, \ldots)$. Decompose M_2 as

$$M_2 = \frac{1}{nhf(x)} \sum_{i=1}^{n} \left\{ K_h(x - X_i)\sigma^2(X_i)\xi_i - \mathrm{E}[K_h(x - X_i)\sigma^2(X_i)\xi_i | \mathcal{G}_{i-1}] \right\}$$

$$+ \frac{1}{nhf(x)} \mathrm{E}[\sigma^2(X_1)K_h(x - X_1)] \sum_{i=1}^{n} \mathrm{E}[\xi_i | \mathcal{H}_{i-1}] =: M_{21} + M_{22}. \qquad (5.11)$$

Using the martingale CLT once again, we conclude that

$$\sqrt{nhf(x)}M_{21} \xrightarrow{\mathrm{d}} \mathcal{N} \left(0, \sigma^4(x)\mathrm{E}[\xi_1^2] \int K^2 \right). \qquad (5.12)$$

Moreover, $\mathrm{E}[M_{22}] = 0$ and with d_n defined in (2.9),

$$M_{22} = \sigma^2(x) \frac{d_n}{n} d_n^{-1} \sum_{i=1}^{n} \mathrm{E}[\xi_i | \mathcal{H}_{i-1}] + O(h) \frac{d_n}{n} d_n^{-1} \sum_{i=1}^{n} \mathrm{E}[\xi_i | \mathcal{H}_{i-1}].$$

13

If $\alpha_\varepsilon < 1/2$ we use (2.9) and (2.11) to get

$$n^{\alpha_\varepsilon} M_{22} \xrightarrow{\mathrm{d}} C_2 \sigma^2(x) Z_2(1). \tag{5.13}$$

Comparing (5.12) with (5.13) we note that either M_{21} or M_{22} dominates, according to the respective assumptions on h, as stated in the Proposition 3.2. If $hn^{(1-2\alpha_\varepsilon)} \to 0$ then M_{21} is of the higher order than M_{22} and the limiting behaviour of $\{\hat{\sigma}^2_{\mathrm{oracle}}(x) - \sigma^2(x)\}$ is determined by M_{21}. Otherwise, if $hn^{(1-2\alpha_\varepsilon)} \to \infty$, then M_{22} determines the limiting behaviour of the oracle estimator. On the other hand, if $\alpha_\varepsilon > 1/2$, $M_{22} = O_P(n^{-1/2})$ (see (2.10)), so that M_{22} is negligible w.r.t. M_{21}.

Under (E2) and (P2) we start once again with the following martingale decomposition (recall M_{21} from (5.11)):

$$M_{21} + \frac{1}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[K_h(x - X_i)\sigma^2(X_i)\xi_i|\mathcal{G}_{i-1}] \tag{5.14}$$

$$= M_{21} + \frac{1}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] \int K_h(x - z\varphi(U_i))\sigma^2(z\varphi(U_i))f_Z(z) \, dv.$$

Again, a function $\Lambda(\cdot; x) \to \int K_h(x - z\varphi(\cdot))\sigma^2(z\varphi(\cdot)))f_Z(z) \, dz$ is in $L^2(\phi)$. Indeed, its $L^2(\phi)$ norm can be bounded by $\sup_x |K^2(x)|\mathrm{E}[\sigma^2(X)]$, which is finite on account of (D2). Furthermore, let us assume that its centered version has Hermite rank $q' \geq 1$. Let $\tilde{J}^*_{m,h}(x)$, $m \geq q'$, be the corresponding Hermite coefficients, i.e.

$$\tilde{J}^*_{m,h}(x) = \mathrm{E}\left[H_m(U_1)\Lambda(U_1; x)\right],$$

where U_1 is standard normal. Put $\Psi(\varepsilon_i) := \mathrm{E}[\xi_i|\mathcal{H}_{i-1}]$. With this notation we can re-write the second part in (5.14) as

$$\begin{aligned}
M_{23} &:= \frac{1}{nhf(x)} \sum_{i=1}^{n} \mathrm{E}[\xi_i|\mathcal{H}_{i-1}]\Lambda(U_i; x) \\
&= \int \frac{1}{nh} \mathrm{E}[K_h(x - z\varphi(U_1))\sigma^2(v\varphi(U_1))]f_Z(z) \, dz \frac{1}{f(x)} \sum_{i=1}^{n} \Psi(\varepsilon_i) \\
&\quad + \frac{1}{nhf(x)} \sum_{i=1}^{n} \Psi(\varepsilon_i) \sum_{m=q'}^{\infty} \tilde{J}^*_{m,h}(x)\frac{H_m(U_i)}{m!}.
\end{aligned}$$

The first integral in the expression for J_{23} is evaluated as

$$\iint K_h(x - zy)\sigma^2(zy)f_Z(z)f_{\varphi(U)}(y) \, dy \, dz$$

$$\sim h\sigma^2(x)\mathrm{E}\left[f_{\varphi(U)}\left(\frac{x}{Z_1}\right)\frac{1}{Z_1}\right] = h\sigma^2(x)J^*_0(x).$$

Therefore, by (2.11), the first part in M_{23} behaves like

$$\frac{\sigma^2(x)J_0^*(x)}{nf(x)}\sum_{i=1}^n \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] = O_P\left(n^{-\alpha_\varepsilon}\vee n^{-1/2}\right). \tag{5.15}$$

Without loss of generality we may assume that $\mathrm{E}[U_iU_j] < 1$ for all $i \neq j$. With help of Mehler's formula and (2.8), the variance of M_{23} can be written as

$$\mathrm{Var}\left(\frac{1}{nhf(x)}\sum_{i=1}^n \Psi(\varepsilon_i)\Lambda(U_i;x)\right)$$

$$= \frac{1}{n^2h^2f^2(x)}\sum_{i,j=1}^n \mathrm{E}[\Psi(\varepsilon_i)\Psi(\varepsilon_j)]\mathrm{E}[\Lambda(U_i;x)\Lambda(U_j;x)]$$

$$= \frac{1}{n^2h^2f^2(x)}\sum_{i,j=1}^n \mathrm{E}[\Psi(\varepsilon_i)\Psi(\varepsilon_j)]\sum_{m=q'}^\infty \frac{\tilde{J}_{m,h}^{*2}(x)}{m!}\mathrm{E}^m[U_iU_j]$$

$$\leq \|\Lambda(\cdot;x)\|_{L^2(\phi)}^2 \frac{1}{n^2h^2f^2(x)}\sum_{i,j=1}^n \mathrm{E}[\Psi(\varepsilon_i)\Psi(\varepsilon_j)]|\mathrm{E}^{q'}[U_iU_j]|$$

$$= \frac{1}{h^2f^2(x)}\|\Lambda(\cdot;x)\|_{L^2(\phi)}^2 O(n^{-1}\vee n^{-(2\alpha_\varepsilon+q\alpha_U)}).$$

Applying (5.32) below, $\tilde{J}_{m,h}^{*2}(x) \sim h^2\sigma^4(x)J_m^{*2}(x)$, as $h \to 0$. Therefore,

$$\|\Lambda(\cdot;x)\|_{L^2(\phi)}^2 = O(h^2\sigma^4(x))\left\|\frac{1}{\varphi(\cdot)}f_Z\left(\frac{x}{\varphi(\cdot)}\right)\right\|_{L^2(\phi)}^2.$$

Consequently, we conclude that $M_{23} = O_P(n^{-1/2})$ if $\alpha_\varepsilon > 1/2$. Otherwise, if $\alpha_\varepsilon < 1/2$ then the limiting behaviour of $n^{\alpha_\varepsilon}M_{23}$ is the same as the term in (5.15), i.e.

$$n^{\alpha_\varepsilon}\frac{\sigma^2(x)J_0^*(x)}{nf(x)}\sum_{i=1}^n \mathrm{E}[\xi_i|\mathcal{H}_{i-1}] \overset{\mathrm{d}}{\to} C_2\sigma^2(x)\frac{J_0^*(x)}{f(x)}Z_2(1). \tag{5.16}$$

Now, again, asymptotic behaviour of M_2 in (5.8) is obtained by comparing M_{21} in (5.12) with (5.16) under the respective assumptions on h. □

Remark 5.2. If f is estimated by $\hat{f}_h$, then M_1 becomes basically the stochastic bias term. Consider for a moment the case of (P2) with $X_i = U_i$. Let us assume that the density f is known. Then,

$$\tilde{M}_1 := \left(\frac{1}{nhf(x)}\sum_{i=1}^n \sigma^2(X_i)K_h(x - X_i) - \sigma^2(x)\right)$$

can be decomposed in the analogous way as the kernel density estimator in (5.1). Since now $q = 1$, we conclude that under (D3) the proper normalization for $\tilde{M}_1$ is either $\sqrt{nhf(x)}$ if $hn^{(1-\alpha_X)} \to 0$, or $n^{\alpha_X/2}$ if $hn^{(1-\alpha_X)} \to \infty$. This explains difference between our results and those in Guo and Koul (2008). A similar effect was also mentioned in Mielniczuk and Wu (2004).

5.4 Proof of Theorems 3.5 and 3.6

Recall the decomposition (5.3). Asymptotic behaviour of M_1 and M_2 ((5.5) and (5.6), respectively) was proven in Section 3.1. Thus, in order to prove Theorem 3.5 (assuming that (E2) and either (P1) or (P2) holds) and Theorem 3.6 (assuming that (E2)+(P2) holds) it is enough to study M_3 and M_4. We will show that in the first case $\sqrt{nh}(M_3 + M_4) = o_P(1)$, whereas in the second case (5.9) and (5.10) hold. Specifically, the proof of theorems will be finished given that we validate the following lemmas.

Lemma 5.3. *Under the conditions of Theorem 3.5 we have $\sqrt{nh}M_3 = o_P(1)$. Under the conditions of Theorem 3.6, (5.9) holds.*

Lemma 5.4. *Under the conditions of Theorem 3.5 we have $\sqrt{nh}M_4 = o_P(1)$. Under the conditions of Theorem 3.6, (5.10) holds.*

Least squares estimation of regression parameters leads to the following expressions:

$$\hat{\beta}_1 - \beta_1 = \frac{1}{S_n}\left(\frac{1}{n}\sum_{j=1}^{n} X_j \sigma(X_j)\varepsilon_j - \bar{X}\bar{V}\right), \quad \hat{\beta}_0 - \beta_0 = \bar{V} - \bar{X}(\hat{\beta}_1 - \beta_1), (5.17)$$

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

Proof of Lemma 5.3. Again, we may replace $\hat{f}_h$ with f in the expression for M_3. Write

$$M_3 = (\hat{\beta}_0 - \beta_0)\frac{2}{nhf(x)}\sum_{i=1}^{n} K_h(x - X_i)\sigma(X_i)\varepsilon_i$$

$$+ (\hat{\beta}_1 - \beta_1)\frac{2}{nhf(x)}\sum_{i=1}^{n} K_h(x - X_i)X_i\sigma(X_i)\varepsilon_i$$

$$=: \quad L_3 + R_3 =: (\hat{\beta}_0 - \beta_0)\tilde{L}_3 + (\hat{\beta}_1 - \beta_1)\tilde{R}_3.$$

Under (E1) (regardless whether (P1) or (P2) holds) both $\tilde{L}_3$ and $\tilde{R}_3$ are martingales so that both are $O_P((nh)^{-1/2})$. Indeed, for example for $\tilde{L}_3$ we have

$$\text{Var}(\tilde{L}_3) = \frac{1}{n^2 h^2 f^2(x)}\sum_{i=1}^{n} \text{Var}\left(K_h(x - X_i)\sigma(X_i)\varepsilon_i\right)$$

$$\leq \frac{1}{nh^2 f^2(x)}\mathbb{E}[\varepsilon^2]\mathbb{E}[K_h^2(x - X_1)\sigma(X_1)] = O\left(\frac{1}{nh}\right).$$

Since $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent, we conclude that $\sqrt{nh}M_3 = o_P(1)$.

For the case of (E2)+(P1), let us define the quantity

$$\tilde{M}_3 := \frac{2}{n^2 hf(x)}\sum_{i=1}^{n}\sum_{j=1}^{n} K_h(x - X_i)\sigma(X_i)\sigma(X_j)(X_i - \mu)(X_j - \mu)\varepsilon_i\varepsilon_j.$$

We may decompose

$$M_3 = \frac{1}{S_n}\tilde{M}_3 + \bar{V}\tilde{L}_3 - \frac{1}{S_n}(\bar{X} - \mu)\bar{V}(\tilde{R}_3 - \mu\tilde{L}_3) - (\bar{X} - \mu)(\hat{\beta}_1 - \beta_1)\tilde{L}_3. \quad (5.18)$$

On account of Slutsky's lemma, we may ignore S_n when studying asymptotic of M_3. Under (E2) and (P1), it will be shown below that $\sqrt{nh}\tilde{M}_3 = o_P(1)$ if $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] = 0$ and

$$n^{\alpha_\varepsilon}\tilde{M}_3 \xrightarrow{\mathrm{d}} \left(2C_1^2\mathrm{E}[\sigma(X_1)(X_1 - \mu)]\sigma(x)(x - \mu)\right) \times \chi^2(1) \quad (5.19)$$

otherwise. For the second term in (5.18) we will prove

$$n^{\alpha_\varepsilon}\bar{V}\tilde{L}_3 \xrightarrow{\mathrm{d}} \left(2C_1^2\sigma(x)\mathrm{E}[\sigma(X_1)]\right) \times \chi^2(1). \quad (5.20)$$

Furthermore, the third term is $O_p(n^{-1/2}) = o_P((nh)^{-1/2})$, which is obvious since $\bar{X} - \mu = O_P(n^{-1/2})$ and $(\tilde{R}_3 - \mu\tilde{L}_3) = O_P(1)$ (the latter may be concluded by showing that $\mathrm{E}[|\tilde{R}_3|] < \infty$ and $\mathrm{E}[|\tilde{L}_3|] < \infty$. Indeed, for $|\tilde{R}_3|$ we have

$$\mathrm{E}\left[|\tilde{R}_3|\right] \leq \frac{2}{nhf(x)}n\mathrm{E}\left[|K_h(x - X_i)X_i\sigma(X_1)|\right]\mathrm{E}\left[|\varepsilon_1|\right] = O(1).)$$

Also,

$$n^{\alpha_\varepsilon/2}\bar{V} \xrightarrow{\mathrm{d}} \mathcal{N}(0, C_1^2\mathrm{E}^2[\sigma(X_1)]) \quad (5.21)$$

so that $\bar{V} = o_P(1)$. Likewise, the fourth term is $o_P(n^{-1/2}) = o_P((nh)^{-1/2})$. Equations (5.19), (5.20) and the fact that the convergence in (5.19) and (5.20) is joint (this follows from the proof below) imply (5.9).

Thus, let us deal with the contributing terms $\tilde{M}_3$ and $\bar{V}\tilde{L}_3$. Under (E2) and (P1), define

$$V_h(x, X_i, X_j) := K_h(x - X_i)\sigma(X_i)\sigma(X_j)(X_i - \mu)(X_j - \mu). \quad (5.22)$$

Let $\bar{V}_h(x, X_i, X_j) := V_h(x, X_i, X_j) - \mathrm{E}\left[V_h(x, X_i, X_j)\right]$. Consider the following decomposition

$$\begin{aligned}
\tilde{M}_3 &= \frac{2}{n^2hf(x)}\sum_{i=1}^{n}\varepsilon_i^2 V_h(x, X_i, X_i) + \frac{2}{n^2hf(x)}\sum_{\substack{i,j=1 \\ i \neq j}}^{n}\varepsilon_i\varepsilon_j\mathrm{E}\left[V_h(x, X_i, X_j)\right] \\
&\quad + \frac{2}{n^2hf(x)}\sum_{\substack{i,j=1 \\ i \neq j}}^{n}\varepsilon_i\varepsilon_j\bar{V}_h(x, X_i, X_j) =: M_{31} + M_{32} + M_{33}.
\end{aligned}$$

We note that $\mathrm{E}\left[V_h(x, X_1, X_1)\right]$ is proportional to h. Indeed, using the first order Taylor's expansion,

$$\mathrm{E}\left[V_h(x, X_1, X_1)\right] = \int K_h(x - s)\sigma^2(s)(s - \mu)^2 f_X(s)ds \sim h\sigma^2(x)(x - \mu)^2 f_X(x),$$

where f_X is the density of X_1. Therefore, the first term in $\tilde{M}_3$ is of order $O_P(1/n)$, so that $\sqrt{nh}M_{31} = o_P(1)$. In what follows, we will show that $\sqrt{nh}J_{33} = o_P(1)$ and the only part which *may* contribute is M_{32}.

The second part, M_{32}, is written as (recall the definition of b_n^2 in (2.5))

$$\frac{2\mathrm{E}\left[V_h(x, X_1, X_2)\right] b_n^2}{n^2 h f(x)} \left\{ \left(b_n^{-1} \sum_{j=1}^{n} \varepsilon_j \right)^2 - b_n^{-2} \sum_{j=1}^{n} \varepsilon_j^2 \right\}. \qquad (5.23)$$

The second part in the brackets is negligible. Furthermore, note that

$$\mathrm{E}[V_h(x, X_1, X_2)] = \mathrm{E}\left[K_h(x - X_i)\sigma(X_i)\sigma(X_j)(X_i - \mu)(X_j - \mu)\right]$$

$$= \int K_h(x-s)\sigma(s)\sigma(t)(s-\mu)(t-\mu)f(s)f(t)dsdt$$

$$= h \int K(v)\sigma(x - vh)\sigma(t)(x - vh - \mu)(t - \mu)f(x - vh)f(t)dvdt$$

$$\sim h\sigma(x)(x - \mu)f(x)\mathrm{E}[\sigma(X_1)(X_1 - \mu)]. \qquad (5.24)$$

Using (5.24) we obtain (5.19). If $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] = 0$, then $\mathrm{E}[V_h(x, X_1, X_2)] = 0$ and $M_{32} \equiv 0$.

Next, we establish a bound on M_{33}. We have

$$\mathrm{Var}\left(\sum_{\substack{j,i=1 \\ j \neq i}}^{n} \varepsilon_i \varepsilon_j \bar{V}_h(x, X_i, X_j) \right) \leq \sum_{\substack{i,j=1 \\ j \neq i}}^{n} \mathrm{E}[\varepsilon_i^2 \varepsilon_j^2] \mathrm{E}[\bar{V}_h^2(x, X_i, X_j)]$$

$$+ \sum_{i=1}^{n} \sum_{\substack{j,j'=1 \\ j' \neq j, j \neq i, j' \neq i}}^{n} \mathrm{E}[\varepsilon_i^2 \varepsilon_j \varepsilon_{j'}] \mathrm{E}[\bar{V}_h(x, X_i, X_j)\bar{V}_h(x, X_i, X_{j'})]$$

$$+ \sum_{j=1}^{n} \sum_{\substack{i,i'=1 \\ i' \neq i, i \neq j, i' \neq j}}^{n} \mathrm{E}[\varepsilon_j^2 \varepsilon_i \varepsilon_{i'}] \mathrm{E}[\bar{V}_h(x, X_i, X_j)\bar{V}_h(x, X_{i'}, X_j)]$$

$$+ \sum_{i=1}^{n} \sum_{\substack{j,j'=1 \\ i \neq j, j \neq j'}}^{n} \mathrm{E}[\varepsilon_i \varepsilon_j^2 \varepsilon_{j'}] \mathrm{E}[\bar{V}_h(x, X_i, X_j)\bar{V}_h(x, X_j, X_{j'})]$$

$$+ \sum_{\substack{i,i'=1 \\ i \neq j, i \neq i'}}^{n} \sum_{j=1}^{n} \mathrm{E}[\varepsilon_i^2 \varepsilon_j \varepsilon_{i'}] \mathrm{E}[\bar{V}_h(x, X_i, X_j)\bar{V}_h(x, X_{i'}, X_i)]$$

$$+ \sum_{\substack{i,i',j,j' \\ i \neq i' \neq j \neq j'}} \mathrm{E}[\varepsilon_i \varepsilon_{i'} \varepsilon_j \varepsilon_{j'}] \mathrm{E}[\bar{V}_h(x, X_i, X_j)\bar{V}_h(x, X_{i'}, X_{j'})]. \qquad (5.25)$$

The last part is equal to 0, since for different indices $\bar{V}_h(x, X_i, X_j)$, $\bar{V}_h(x, X_{i'}, X_{j'})$ are independent (and centered). Next, straightforward computation shows that

$$\mathrm{E}[V_h^2(x, X_i, X_j)] = O(h),$$

18

$$\mathrm{E}[V_h(x, X_i, X_j)V_h(x, X_{i'}, X_{j'})] = O(h^2), \qquad i \neq i',$$

and

$$\mathrm{E}[V_h(x, X_i, X_j)V_h(x, X_i, X_{j'})] = O(h), \qquad i \neq j.$$

Using the covariance bounds on the error sequence, we obtain that the variance above is of the order $O(hn^{3-\alpha_\varepsilon} + hn^2)$. Consequently,

$$\mathrm{Var}(M_{33}) = O\left(\frac{n^2 h}{n^4 h^2} + \frac{hn^{3-\alpha_\varepsilon}}{n^4 h^2}\right) = O\left(\frac{1}{n^2 h} + \frac{1}{hn^{1+\alpha_\varepsilon}}\right),$$

and so $\sqrt{nh}M_{33} = o_P(1)$.

The proof of (5.20) follows along the same lines as for (5.19). Furthermore, we note that the convergence in (5.19) follows from the asymptotic expansion of $\tilde{M}_3$ with the dominating term as in (5.23). The same applies to the convergence (5.20) - up to a deterministic constant, the term (5.23) appears in the asymptotic expansion for $\bar{V}\tilde{L}_3$. Therefore, the convergence in (5.19) and (5.20) is *joint*, so that we may conclude (5.9). This finishes the proof of Lemma 5.3. $\qquad\square$

Remark 5.5. Note that the second term in the decomposition (5.18) is the effect of the estimation of the intercept β_0 and contributes with rate $n^{-\alpha_\varepsilon}$. If the model (1.1) is considered without intercept β_0, then this term does not appear, so that $n^{-\alpha_\varepsilon}$ contribution appears only when $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] \neq 0$.

To prove Lemma 5.4, we establish limiting behaviour if $\hat{\beta}_1$ in a particular case.

Lemma 5.6. *Assume* (E2) *and* (P1). *If* $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] \neq 0$ *holds, then*

$$n^{\alpha_\varepsilon/2}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \mathcal{N}(0, \omega^2). \tag{5.26}$$

Proof of Lemma 5.6. Recall from (5.17) that

$$\hat{\beta}_1 - \beta_1 = \frac{1}{S_n}\left(\frac{1}{n}\sum_{j=1}^{n} X_j \sigma(X_j)\varepsilon_j - \bar{X}\bar{V}\right),$$

where S_n^2 is the sample variance associated with $X_1, \ldots, X_n$. Since $\mathrm{Var}[X_1^2] = 1$, $S_n \xrightarrow{P} 1$. Thus, via Slutsky's lemma, in order to show convergence of $\hat{\beta}_1 - \beta_1$, it suffices to study

$$I_1 - I_2 := \frac{1}{n}\sum_{j=1}^{n} X_j \sigma(X_j)\varepsilon_j - \bar{X}\bar{V}. \tag{5.27}$$

Consider the sum $\frac{1}{n}\sum_{j=1}^{n} \nu(X_j)\varepsilon_j$, where $\nu(\cdot)$ is a deterministic function. Decompose this sum into:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\nu(X_i)\varepsilon_i - \mathrm{E}\left[\nu(X_i)\varepsilon_i|\mathcal{G}_{i-1}\right]\right) + \mathrm{E}[\nu(X_1)]\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\varepsilon_i|\mathcal{H}_{i-1}\right]. \tag{5.28}$$

19

The first part is a martingale. If $\mathrm{E}[\nu(X_1)] \neq 0$, the second part is of order $O_P(n^{-\alpha_\varepsilon/2})$, and dominates the martingale part. In this case its convergence follows from (2.5) and (2.7). Otherwise, if $\mathrm{E}[\nu(X_1)] = 0$, the second part disappears and the sum is $O_P(n^{-1/2})$.

Such decomposition is applied to $I_1 - I_2$:

$$I_1 - I_2 = \frac{1}{n} \sum_{j=1}^n X_j \sigma(X_j)\varepsilon_j - I_2 \tag{5.29}$$

$$= \frac{1}{n} \sum_{j=1}^n \left(X_j \sigma(X_j)\varepsilon_j - \mathrm{E}\left[X_j \sigma(X_j)\varepsilon_j | \mathcal{G}_{j-1}\right]\right)$$

$$+ \mathrm{E}[X_1 \sigma(X_1)] \frac{1}{n} \sum_{j=1}^n \mathrm{E}\left[\varepsilon_j | \mathcal{H}_{j-1}\right] - I_2 =: A_1 + A_2 - I_2.$$

We have $\sqrt{n} A_1 \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_0^2)$. Convergence for A_2 follows from (2.7). The term I_2 is $O_P(n^{-(1+\alpha_\varepsilon)/2})$ (see (2.5)) so that it is negligible. $\qquad\square$

Proof of Lemma 5.4: Using (5.17), we decompose M_4 as follows:

$$M_4 = \frac{1}{nhf(x)}(\bar{V})^2 \sum_{i=1}^n K_h(x - X_i) + \frac{1}{nhf(x)}(\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n (X_i - \bar{X})^2 K_h(x - X_i)$$

$$+ \frac{2}{nhf(x)}\bar{V}(\hat{\beta}_1 - \beta_1) \sum_{i=1}^n (X_i - \bar{X}) K_h(x - X_i) =: B_1 + B_2 + B_3.$$

Under (E1), we have clearly $M_4 = O_P(n^{-1})$, which clearly implies $M_4 = O_P((nh)^{-1/2})$.

Under (E2) and (P1), we obtain via (5.21) that

$$n^{\alpha_\varepsilon} B_1 \xrightarrow{\mathrm{d}} C_1^2 \mathrm{E}^2[\sigma(X_1)]\chi^2(1). \tag{5.30}$$

Note that B_1 has a structure $(\bar{V})^2 \times V_n$, where $V_n = \frac{1}{nhf(x)} \sum_{i=1}^n K_h(x - X_i)$. We decompose the term as $(\bar{V})^2 \mathrm{E}[V_n] + (\bar{V})^2(V_n - \mathrm{E}[V_n])$. Then the second term is $o_P((\bar{V})^2)$ and $\mathrm{E}[V_n] \sim 1$. Likewise, from Lemma 5.6,

$$n^{\alpha_\varepsilon} B_2 \xrightarrow{\mathrm{d}} (x - \mu)^2 \omega^2 \chi^2(1) \tag{5.31}$$

if $\mathrm{E}[\sigma(X_1)(X_1 - \mu)] \neq 0$ and $B_2 = O_P(n^{-1})$ otherwise. For the term B_3 we have clearly that (recall (P1)) $(nh)^{-1} \sum_{i=1}^n (X_i - \bar{X}) K_h(x - X_i) = O_P((nh)^{-1/2})$, so that $B_3 = o_P((nh)^{-1/2})$. Furthermore, as in case of M_3, the convergence in (5.30) and (5.20) is concluded from the asymptotic expansions, so that it is *joint* and (5.10) may be concluded. Furthermore, the convergence in (5.9) and (5.10) is *joint*. $\qquad\square$

Appendix

Assumption (J2)

Let ζ be a real function. Consider a function

$$(\cdot; x, h, \zeta) \to \int K_h(x - z\varphi(\cdot))\zeta(z\varphi(\cdot))f_Z(z)\, dz$$

and its corresponding Hermite coefficient $J_{m,h}^*(x; \zeta)$. If ζ is in $\mathcal{C}^2(\mathcal{I})$, then one can evaluate

$$J_{m,h}^*(x; \zeta) \sim h\zeta(x)J_m^*(x) := h\zeta(x)\mathrm{E}\left[J_m'\left(\frac{x}{Z_1}\right)\frac{1}{Z_1}\right]. \qquad (5.32)$$

Indeed, to see this assume for simplicity that φ is invertible. Then changing variables $u \to s = (x - z\varphi(u))/h$ and applying Taylor expansion,

$$
\begin{aligned}
J_{m,h}^*(x) &= \iint K_h(x - z\varphi(u))\zeta(z\varphi(u))f_Z(z)\phi(u)H_m(u)\, dz\, du \\
&= h\int\int \phi\left(\varphi^{-1}\left(\frac{x - sh}{z}\right)\right)H_m\left(\varphi^{-1}\left(\frac{x - sh}{z}\right)\right)\frac{1}{z}\frac{K(s)\zeta(x - sh)f_Z(z)}{\varphi'\left(\varphi^{-1}\left(\frac{x - sh}{z}\right)\right)}dz\,ds \\
&\sim h\zeta(x)\int \phi\left(\varphi^{-1}\left(\frac{x}{z}\right)\right)H_m\left(\varphi^{-1}\left(\frac{x}{z}\right)\right)\frac{1}{\varphi'\left(\varphi^{-1}\left(\frac{x}{z}\right)\right)}\frac{1}{z}f_Z(z)\, dz \qquad (5.33) \\
&= h\zeta(x)\int J_m'\left(\frac{x}{z}\right)\frac{1}{z}f_Z(z)\, dz = h\zeta(x)\mathrm{E}\left[J_m'\left(\frac{x}{Z_1}\right)\frac{1}{Z_1}\right].
\end{aligned}
$$

Substituting further $u = \varphi^{-1}(x/z)$ we evaluate integral in (5.33) as

$$\int \phi(u)H_m(u)\frac{1}{\varphi(u)}f_Z\left(\frac{x}{\varphi(u)}\right)\, du.$$

The latter expression can be recognized as mth Hermite coefficient of function $(\cdot; x) \to \frac{1}{\varphi(\cdot)}f_Z\left(\frac{x}{\varphi(\cdot)}\right)$. Therefore, assumption (J2) guarantees that coefficients $J_m^*(x)$ are well defined for each m. We can also see that if f_Z is degenerated (i.e. it has mass one at $x = 1$, say), then (J2) reduces to (J1).

Martingale CLT

Let us recall from Hall and Heyde (1980) the following martingale central limit theorem.

Lemma 5.7. *Assume that $(R_i, \mathcal{G}_i)$, $i \geq 1$, is a martingale difference. Define $\bar{R}_i = R_i - \mathrm{E}[R_i|\mathcal{G}_{i-1}]$. If*

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

and

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

then

$$\sum_{i=1}^{n} R_i \xrightarrow{d} N(0,1).$$

To obtain the limit theorem for the term J_{21} in (5.11) we consider

$$R_i = (nh)^{-1/2} \sigma^2(X_i) K_h(x - X_i) \xi_i / \sqrt{f(x)}.$$

The proof is exactly the same as in Wu and Mielniczuk (2002, Lemma 2) and therefore it is omitted.

Acknowledgement

References

[1] Avram, F. and Taqqu, M. S. (1987) Noncentral limit theorems and Appell polynomials. *Annals of Probability* 15, 767–775.

[2] Csörgő, S. and Mielniczuk, J. (1995) Density estimation under long-range dependence. *Annals of Statistics* 23, 990–999.

[3] Davydov, Ju. A. (1970) The invariance principle for stationary processes. *Teor. Verojatnost. i Primenen.* 15, 498–509.

[4] Fan, J. and Yao, Q. (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85, 645–660.

[5] Guo, H. and Koul, H. L. (2008) Asymptotic inference in some heteroscedastic regression models with long memory design and errors. *Annals of Statistics* 36, 458–487.

[6] Kulik, R. and Lorek, P. (2011) Some results on random design regression with long memory errors and predictors. *Journal of Statistical Planning and Inference* 141, 508-523.

[7] Kulik, R. and Wichelhaus, C. (2011) Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors. *Electronic Journal of Statistics* 5, 856-898.

[8] Lahiri, S. N. (2003) *Resampling methods for dependent data.* Springer Series in Statistics.

[9] Mielniczuk, J. and Wu, W. B. (2004) On random-design model with dependent errors. *Statistica Sinica* 14, 1105–1126.

[10] Taqqu, M. S. (2003) Fractional Brownian motion and long-range dependence. *Theory and applications of long-range dependence*, 5–38. Birkhäuser.

[11] Wang, L., Brown, L. D., Cai, T. T. and Levine, M. (2008) Effect of mean on variance function estimation in nonparametric regression. *Annals of Statistics* 36, 646–664.

[12] Wu, W. B. and Mielniczuk, J. (2002) Kernel density estimation for linear processes. *Annals of Statistics* 30, 1441–1459.

[13] Zhao, Z. and Wu, W. B. (2008) Confidence bands in nonparametric time series regression. *Annals of Statistics* 36, 1854–1878.