Bernoulli **7**(5), 2001, 733–750

Local polynomial estimation with a FARIMA-GARCH error process

JAN BERAN* and YUANHUA FENG**

Department of Mathematics and Statistics, University of Konstanz, Universitätsstr. 10, Postfach 5560, D-78457 Konstanz, Germany. E-mail: **jan.beran@uni-konstanz.de;* ***yuanhua.feng@uni-konstanz.de*

This paper considers estimation of the trend function g as well as its νth derivative $g^{(\nu)}$ in a so-called semi-parametric FARIMA-GARCH model by local polynomial fits. The focus is on the derivation of the asymptotic normality of $\hat{g}^{(\nu)}$. A central limit theorem based on martingale theory is developed. Asymptotic normality of the sample mean of a FARIMA-GARCH process is proved. These results are then used to show the asymptotic normality of $\hat{g}^{(\nu)}$. As an auxiliary result, the weak consistency of a weighted sum is obtained for second-order stationary time series with short or long memory under very weak conditions. Formulae for the mean integrated square error and the asymptotically optimal bandwidth of $\hat{g}^{(\nu)}$ are also given.

Keywords: asymptotic normality; FARIMA-GARCH process; local polynomial estimation; long memory; martingales; semi-parametric models

1. Introduction

Consider the semi-parametric regression model

$$Y_i = g(t_i) + X_i, \tag{1.1}$$

where g: $[0, 1] \to \mathbb{R}$ is a smooth function, $t_i = (i/n)$ and

$$X_i = (1 - B)^{-\delta}\phi^{-1}(B)\psi(B)\epsilon_i, \tag{1.2}$$

with

$$\epsilon_i = z_i h_i^{1/2}, \qquad h_i = \alpha_0 + \sum_{j=1}^{r} \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^{s} \beta_k h_{i-k}. \tag{1.3}$$

The z_i are independent and identically distributed (i.i.d.) standard normal random variables, $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s \geqslant 0$, $\delta \in (-0.5, 0.5)$, B is the backshift operator, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_l B^l$ and $\psi(B) = 1 + \psi_1 B + \ldots + \psi_m B^m$ are polynomials in B with no common factors and all roots outside the unit circle. Here, the fractional difference $(1 - B)^{\delta}$ introduced by Granger and Joyeux (1980) and Hosking (1981) – see also the monograph of Beran (1994) – is defined by

1350–7265

$$(1 - B)^{\delta} = \sum_{k=0}^{\infty} b_k B^k, \tag{1.4}$$

with

$$b_k(\delta) = \frac{\Gamma(k - \delta)}{\Gamma(k + 1)\Gamma(-\delta)}. \tag{1.5}$$

The innovation process defined in (1.3) follows a GARCH model (Bollerslev 1986), which is a generalization of the autoregressive conditional heteroscedastic (ARCH) model proposed by Engle (1982). It is assumed that there is a strictly stationary solution of (1.3) such that $\mathrm{E}(\epsilon_i^4) < \infty$. Necessary and sufficient conditions which guarantee this may be found in Ling (1999) and Ling and McAleer (2000) (see also Ling and Li 1997; Chen and An 1998; He and Teräsvirta 1999a). The error process X_i has short memory if $\delta = 0$, long memory if $0 < \delta < 0.5$ and antipersistence if $-0.5 < \delta < 0$ (see Beran 1994). Model (1.1)–(1.3) is an extension of the semi-parametric fractional autoregressive (SEMIFAR) model introduced by Beran (1999) (see also Beran 1995), which will be called a semi-parametric fractional autoregressive integrated moving average (FARIMA)-GARCH model – see Ling and Li (1997) and Ling (1998) for a FARIMA-GARCH model without trend. Such a model allows for simultaneous estimation of trend, long memory as well as conditional heteroscedasticity in a time series. Estimation of $g^{(\nu)}$, the νth derivative of g, leads to a nonparametric regression problem with different dependence structures.

The current paper focuses on investigating the asymptotic properties of the local polynomial fits of $g^{(\nu)}$. It is shown that $\hat{g}^{(\nu)}$ converges uniformly on the whole support [0, 1] for errors with short or long memory as well as for errors with antipersistence. Under given conditions, the rate of convergence of $\hat{g}^{(\nu)}$ is $n^{(2\delta-1)(p+1-\nu)/(2p+3-2\delta)}$ for all $\delta \in (-0.5, 0.5)$ and $n^{(1-2\delta)(p+1-\nu)/(2p+3-2\delta)}(\hat{g}^{(\nu)} - g^{(\nu)})$ is asymptotically normal, where $p \geqslant \nu$ is the order of the local polynomial with $p - \nu$ odd.

The paper is organized as follows. The proposed local polynomial estimator is described in Section 2. Section 3 gives some auxiliary results, including a central limit theorem for stationary processes being a weighted sum of a square-integrable martingale difference. Our main results are given in Section 4. Section 5 contains some final remarks. Proofs of theorems are to be found in the Appendix.

2. The estimator

Kernel estimation for nonparametric regression with long-memory errors is investigated by Hall and Hart (1990), Csörgő and Mielniczuk (1995) and Beran (1999). Beran and Feng (1999) proposed to estimate $g^{(\nu)}$ in nonparametric regression with long-memory errors by local polynomial fitting, introduced by Stone (1977) and Cleveland (1979). See Ruppert and Wand (1994) and Fan and Gijbels (1996) for recent development in the context of local polynomial fits. Assume that g is at least $(p + 1)$-times differentiable at a point t_0. Let K be a symmetric density having compact support $[-1, 1]$, called the weight function. Let

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 - t_0 & \dots & (t_1 - t_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_n - t_0 & \dots & (t_n - t_0)^p \end{bmatrix},$$

and let $\mathbf{e}_j$, $j = 1, \dots, p+1$, denote the jth $(p+1)\times 1$ unit vector. Also let $\mathbf{K}$ denote the diagonal matrix with

$$k_i = K\left(\frac{t_i - t_0}{b}\right)$$

as its ith diagonal entry, where b is the bandwidth. Finally, let $\mathbf{y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$ be the vector of observations. Then $\hat{g}^{(\nu)}(t_0)$ $(\nu \leqslant p)$ is obtained by solving the locally weighted least-squares problem

$$Q = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} b_j (t_i - t_0)^j \right\}^2 K\left(\frac{t_i - t_0}{b}\right) \Rightarrow \min, \tag{2.6}$$

which leads to

$$\hat{g}^{(\nu)}(t_0) = \nu! \mathbf{e}_{\nu+1}^{\mathrm{T}} (\mathbf{X}^{\mathrm{T}} \mathbf{K} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{K} \mathbf{y}. \tag{2.7}$$

Here $\hat{g}^{(\nu)}(t_0)$ is a linear smoother with the weighting system $\mathbf{w}^{\nu}(t_0) = \nu!(\mathbf{e}_{\nu+1}^{\mathrm{T}} (\mathbf{X}^{\mathrm{T}} \mathbf{K} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{K})^{\mathrm{T}} = (w_1^{\nu}, \dots, w_n^{\nu})^{\mathrm{T}}$, where $w_i^{\nu} \neq 0$ only for $|t_i - t_0| \leqslant b$. The weighting system does not depend on the dependence structure of the error process. For any interior point $t_0 \in [b, 1-b]$ the non-zero part of $\mathbf{w}^{\nu}(t_0)$ is the same, that is, $\hat{g}^{(\nu)}$ works as a moving average in the interior. Furthermore, $\mathbf{w}^{\nu}(t_0)$ satisfies

$$\sum_{i=1}^{n} w_i^{\nu} (t_i - t_0)^{\nu} = \nu!$$

$$\sum_{i=1}^{n} w_i^{\nu} (t_i - t_0)^{j} = 0 \qquad \text{for } j = 0, \dots, p, \, j \neq \nu. \tag{2.8}$$

Property (2.8) ensures that $\hat{g}^{(\nu)}$ is exactly unbiased if g is a polynomial of order not larger than p.

3. Auxiliary results

In the following a central limit theorem will be developed for the sum of random variables $S_n = \sum_{i=1}^{n} X_i$, where X_i is a weighted sum of $(0, \sigma^2)$ random variables ϵ_k forming a square-integrable martingale difference. For the definitions of martingales and martingale differences see, for example, Hall and Heyde (1980) and Shiryaev (1996). A martingale difference with finite variance is called a square-integrable martingale difference. Let $(\Omega, \mathscr{F}, P)$ be a probability space, where $\mathscr{F}$ is a σ-field of subsets of Ω. Denote by $\{\mathscr{F}_i, \, i \in \mathbf{I} = \{-\infty, \dots, -1, 0, 1, \dots, \infty\}\}$ a non-decreasing sequence of σ-fields of $\mathscr{F}$ sets. We have:

Theorem 1. *Let the sequence* $(\epsilon_i, \mathscr{F}_i, i \in \mathbf{I})$ *be a square-integrable martingale difference with constant variance, that is,* $\mathrm{E}(\epsilon_i|\mathscr{F}_{i-1}) = 0$, $\mathrm{E}(\epsilon_i^2) = \mathrm{E}(\epsilon_0^2) < \infty$. *Furthermore, assume either one of the following cases*

Case 1.

$$X_i = \sum_{k=-\infty}^{\infty} c_{k-i}\epsilon_k, \tag{3.1}$$

where

$$\sum_{k=-\infty}^{\infty} c_k^2 < \infty, \tag{3.2}$$

with $\sigma_n^2 = \mathrm{E}(X_1 + \ldots + X_n)^2 \to \infty$ *as* $n \to \infty$, *and the conditional variance of* ϵ_i *is equal to the unconditional one, that is,* $h_i = \mathrm{E}(\epsilon_i^2|\mathscr{F}_{i-1}) \equiv \mathrm{E}(\epsilon_0^2)$,

or that

Case 2. *X is a FARIMA process with strictly stationary innovations* ϵ_i, *the process* $|\epsilon_i|$ *has an extreme index* $\theta \in (0, 1]$ *and the square process* ϵ_i^2 *is second-order stationary with autocovariances* $\gamma_{\epsilon^2}(k) = \mathrm{cov}(\epsilon_i^2, \epsilon_{i+k}^2) \to 0$ *as* $k \to \infty$.

Then

$$\frac{X_1 + \ldots + X_n}{\sigma_n} \xrightarrow{\mathscr{D}} \mathrm{N}(0, 1).$$

The proof of Theorem 1 is given in the appendix.

Remark 1. Note that i.i.d. $(0, \sigma^2)$ random variables form a square-integrable martingale difference with $h_i \equiv \sigma^2$, Theorem 1, case 1, includes Theorem 18.6.5 of Ibragimov and Linnik (1971) as a special case.

Remark 2. Theorem 1, case 2, is developed for a FARIMA process with the FARIMA-GARCH model as a special case. Such processes are special cases of those defined by (3.1)–(3.2) with $c_k = 0$ for $k > 0$ and $c_k \sim \gamma k^{-1+\delta}$ for $k < 0$ and $|k|$ large, where γ is a constant. In this case we have $\sigma_n^2 = \mathrm{E}(X_1 + \ldots + X_n)^2 = O(n^{1+2\delta})$ for $\delta \in (-0.5, 0.5)$.

Remark 3. Some conditions of Theorem 1 are made here in order to simplify the proof. For instance, in case 2 it is assumed that the process $|\epsilon_i|$ has an extreme index $\theta \in (0, 1]$ so that the extreme value of $|\epsilon_i|$ is asymptotically of the same order as that for an associated independent sequence (with the same marginal distribution) (see Leadbetter 1983; Embrechts *et al*. 1997, Section 8.1). The degenerate case with $\theta = 0$ is excluded in this paper.

Remark 4. The assumption of constant variance for ϵ_i implies that it is an uncorrelated white noise (see Shiryaev 1999, p. 42). Hence, the assumptions on ϵ_i given in Theorem 1 are stronger than that ϵ_i is an uncorrelated white noise. For long-memory processes the assumption that the ϵ_i are uncorrelated $(0, \sigma^2)$ random variables is not sufficient for the derivation of asymptotic normality of the sample mean (see, for example, Taqqu 1975).

We will now develop a theorem on the convergence of the variance (to zero) of a general linear filter. The corollary of this on the weak consistency of a general weighted sum will be used to prove the weak consistency of a weighted sum of the square process ϵ_i^2 in Theorem 1, case 2. Both of them are given for second-order stationary time series.

Theorem 2. *Let $(X_{i,n})$, $1 \leqslant i \leqslant n$, $n = 1, 2, \ldots,$ be a triangular array of random variables from a second-order stationary time series with zero mean, variance σ^2 and autocovariances $\gamma(k)$ such that $\gamma(k) \to 0$ as $k \to \infty$. Let $(w_{i,n})$ be a triangular array of weights such that $\sum_{i=1}^n |w_i| < \infty$ and $\max_{1 \leqslant i \leqslant n} |w_i| \to 0$ as $n \to \infty$. Then $\mathrm{var}(\sum_{i=1}^n w_i X_i) \to 0$ as $n \to \infty$.*

The proof of Theorem 2 is given in the appendix. The weighting system w_i is 'formless': the w_i are also allowed to be negative. Localized weighting systems are included by setting $w_i \equiv 0$ for all i outside a given interval. Hence, all weighting systems generated by common kernel or local polynomial estimators of $g^{(\nu)}$ are special cases of Theorem 2. This means that the variances of these estimators converge to zero for any second-order stationary time series with $\gamma(k) \to 0$ as $k \to \infty$. Furthermore, if X_i is a process with unknown mean μ, we have:

Corollary 1. *Let $(X_{i,n})$ and $(w_{i,n})$, $1 \leqslant i \leqslant n$, $n = 1, 2, \ldots,$ be the triangular array as defined in Theorem 2. Suppose now that the mean μ of X_i is unknown and is estimated by $\hat{\mu} = \sum_{i=1}^n w_i X_i$. If $\sum_{i=1}^n w_i \to 1$ as $n \to \infty$ and the other conditions of Theorem 2 are satisfied, then $\hat{\mu}$ is weakly consistent.*

4. Main results

4.1. Properties of the error process

We consider first the asymptotic normality of the sample mean $\overline{X} = (1/n)\sum_{i=1}^n X_i$ of a FARIMA-GARCH process defined by (1.2)–(1.3). To our knowledge, there are no detailed results on this topic in the literature. Under the condition $\mathrm{E}(\epsilon_i^4) < \infty$ we have $\sum_{j=1}^r \alpha_j + \sum_{k=1}^s \beta_k < 1$ (see Lemma 2.2 in Chen and An 1998). Under this condition ϵ_i is a square-integrable martingale difference with respect to $(\mathscr{F}_i, i \in \mathbf{I})$, where $\mathscr{F}_i$ is the σ-field generated by the information in the past (Shiryaev 1999) and $\mathbf{I}$ is as in Theorem 1. And hence ϵ_i is an uncorrelated white noise. The autocovariance function $\gamma x(k)$ of the FARIMA-GARCH process X has the same form as given in Brockwell and Davis (1991) and Beran (1994). Furthermore, He and Teräsvirta (1999a) show that, under the condition $\mathrm{E}(\epsilon_i^4) < \infty$, the autocorrelation function of the square process ϵ_i^2 decays exponentially. More detailed results on this may be found in He and Teräsvirta (1999b) for second-order GARCH models. The existence of an extreme index for the process $|\epsilon_i|$ follows from Davis *et al.* (1999) – see de Haan *et al*. (1989) and Mikosch and Stărică (2000) for explicit results in some special cases. Whence ϵ_i fulfils the conditions of Theorem 1, case 2.

Based on Theorem 1 we obtain the following theorem, which extends the results of Theorem 8(ii) in Hosking (1996).

Theorem 3. *Let X_i be generated by model (1.2)–(1.3) with $\delta \in (-0.5, 0.5)$. Assume that there is a strictly stationary solution of (1.3) such that $\mathrm{E}(\epsilon_i^4) < \infty$. And suppose that $\phi(B)$ and $\psi(B)$ have no common factors, and all roots of $\phi(B)$ and $\psi(B)$ lie outside of the unit circle. Then*

$$n^{1/2-\delta}\overline{X} \xrightarrow{\mathscr{D}} N(0, V_\delta),$$

where

$$V_\delta = \sigma_\epsilon^2 \frac{|\psi(1)|^2}{|\phi(1)|^2} \frac{\Gamma(1-2\delta)}{(2\delta+1)} \frac{\sin(\pi\delta)}{\pi\delta}. \tag{4.1}$$

The proof of Theorem 3 is given in the Appendix.

4.2. An extension of Theorem 1

Theorem 1 is a central limit theorem for the sample mean of a second-order stationary time series. In the following we will extend it to a central limit theorem for a linear filter of such a process, which can be directly used to derive asymptotic normality of a kernel or a local polynomial estimator.

Theorem 4. *Let $(X_{i,n})$, $1 \leqslant i \leqslant n$, $n = 1, 2, \ldots,$ be a triangular array of random variables and let $(w_{i,n})$ be a triangular array of weights such that $\sigma_n^2 := \mathrm{var}(\sum_{i=1}^n w_i X_i) > 0$ for all n. If*

$$\max_{1\leqslant i\leqslant n} |w_i|/\sigma_n \to 0 \qquad \text{as } n \to \infty, \tag{4.2}$$

$$\sup_k \left| \sum_{i=1}^n w_i c_{k-i} \right| / \sigma_n \to 0 \qquad \text{as } n \to \infty \tag{4.3}$$

and the conditions as given in cases 1 and 2 of Theorem 1 hold respectively, then

$$\left[\sum_{i=1}^n w_i X_i \right] / \sigma_n \xrightarrow{\mathscr{D}} N(0, 1).$$

Condition (4.2) means that the weights w_i are uniformly negligible. If $\max|w_i| = O(1)$, then it implies that $\sigma_n^2 \to \infty$ as $n \to \infty$. Condition (4.3) on the weighted sum $\sum w_i c_{k-i}$ is often not independent of (4.2). Theorem 1 is a special case of Theorem 4 with $w_i \equiv 1$, in which case (4.3) can be derived from (4.2). Theorem 4.2 in Müller (1988) on the asymptotic normality of a weighted sum of i.i.d. random variables is also a very special case of Theorem 4. Based on Theorem 4, the asymptotic normality of $\hat{g}^{(\nu)}$ is easily proved.

4.3. Pointwise asymptotic results

What follows gives asymptotic normality of $\hat{g}^{(\nu)}(t)$ at any point $t \in [0, 1]$ for $\delta \in (-0.5, 0.5)$. Asymptotic results on $\hat{g}^{(\nu)}(t)$ as given in Beran and Feng (1999) are included without proof, since they also hold for the current model. It is assumed that $p - \nu$ is odd, and we put $k = p + 1$. Note that it is enough to give formulae at points $t = cb$ with $c \in [0, 1]$. Here $c < 1$ corresponds to a left boundary point and $c = 1$ to a point in the interior. Following Müller (1987) and Feng (1999), $\hat{g}^{(\nu)}$ is asymptotically equivalent to a kernel estimation in the interior as well as at the boundary. For any ν, k and c, denote by $K_{(\nu,k,c)}$ the asymptotically equivalent kernel (or boundary kernel, respectively) for $\hat{g}^{(\nu)}$ (see, for example, Ruppert and Wand 1994); it is easy to show that

$$\int_{-c}^{1} u^j K_{(\nu,k,c)}(u)\mathrm{d}u = \begin{cases} 0, & j = 0, \ldots, \nu - 1, \nu + 1, \ldots, k - 1, \\ \nu!, & j = \nu, \\ \beta_{(\nu,k,c)}, & j = k, \end{cases} \tag{4.4}$$

where $\beta_{(\nu,k,c)}$ is the (non-zero) kernel constant.

To drive the asymptotic results given below additional assumptions are required:

Assumption 1. *g is an at least k-times continuously differentiable function on* $[0, 1]$.

Assumption 2. *The weight function* $K(u)$ *is a symmetric density* (*a kernel of order* 2) *with compact support* $[-1, 1]$, *having the polynomial form*

$$K(u) = \sum_{l=0}^{r} \alpha_l u^{2l} \mathbb{1}_{[-1,1]}(u)$$

(see, for example, Gasser and Müller 1979).

Assumption 3. *The bandwidth satisfies* $b \to 0$, $(nb)^{1-2\delta} b^{2\nu} \to \infty$ *as* $n \to \infty$.

It can be shown that, under Assumption 3

$$\max_{1 \leqslant i \leqslant n} |w_i| = O([nb^{1+\nu}]^{-1}) = o(1).$$

Let $n_0 = [nt + 0.5]$, $n_1 = [nb]$, $n_c = [ncb]$, where $[\cdot]$ denotes the integer part. Let

$$V_n(c, \delta, b) = (nb)^{-1-2\delta} \sum_{i,j=n_0-n_c}^{n_0+n_1} K_{(\nu,k,c)}\left(\frac{t_i - t}{b}\right) K_{(\nu,k,c)}\left(\frac{t_j - t}{b}\right) \gamma(i - j). \tag{4.5}$$

We obtain the following theorem.

Theorem 5. *Let* Y_i *be generated by model* (1.1)–(1.3). *Suppose that the assumptions of*

Theorem 3 *hold. Suppose that Assumptions* 1–3 *hold, and let* $t = cb$ *with* $0 \leqslant c \leqslant 1$. *Then, for* $\delta \in (-0.5, 0.5)$:

(*i*) *we have bias*

$$\mathrm{E}[\hat{g}^{(\nu)} - g^{(\nu)}] = b^{(k-\nu)} \frac{g^{(k)}(t)\beta_{(\nu,k,c)}}{k!} + o(b^{(k-\nu)}); \tag{4.6}$$

(*ii*)

$$\lim_{n\to\infty} V_n(c, \delta, b) = V(c, \delta), \tag{4.7}$$

where $0 < V(c, \delta) < \infty$ *is a constant;*

(*iii*) *the variance of* $g^{(\nu)}$ *is given by*

$$\mathrm{var}(\hat{g}^{(\nu)}(t)) = (nb)^{-1+2\delta} b^{-2\nu}[V(c, \delta) + o(1)]; \tag{4.8}$$

(*iv*) *when the bias is given by* (4.6), *assuming that* $nb^{(2k+1-2\delta)/(1-2\delta)} \to d^2$ *as* $n \to \infty$, *for some* $d > 0$, *then*

$$(nb)^{1/2-\delta} b^{\nu}(\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)) \xrightarrow{\mathscr{D}} N(d\Delta, V(c, \delta)), \tag{4.9}$$

where $\Delta = g^{(k)}(t)\beta_{(\nu,k,c)}/k!$ *and* $V(c, \delta)$ *is the constant defined in* (4.7).

All of the results of parts (i)–(iii) and those of Theorem 6 below also hold when X_i in (1.1)–(1.3) is a FARIMA process with uncorrelated innovations. Denote $V(c, \delta)$ with $c = 1$ by $V(\delta)$. The formula for $V(\delta)$ with $\nu = 0$ and $\delta > 0$ may be found in Hall and Hart (1990). In the special case when g is estimated by an unweighted local linear fit, then we have, for all $\delta \in (-0.5, 0.5)$,

$$V(\delta) = 2^{(2\delta-1)} \sigma_\epsilon^2 \frac{|\psi(1)|^2}{|\phi(1)|^2} \frac{\Gamma(1-2\delta)}{(2\delta+1)} \frac{\sin(\pi\delta)}{\pi\delta} \tag{4.10}$$

(see Corollary 1 in Beran 1999). Note that the only difference between V_δ given in (4.1) and $V(\delta)$ in (4.10) is the factor $2^{(2\delta-1)}$.

Remark 5. The condition $nb^{(2k+1-2\delta)/(1-2\delta)} \to d^2$ as $n \to \infty$ implies bandwidth $b = O(n^{(2\delta-1)/(2k+1-2\delta)})$ with $(nb)^{1/2-\delta} b^{\nu} = O(n^{(1-2\delta)(k-\nu)/(2k+1-2\delta)})$. Theorem 6 below shows that such a bandwidth is of the optimal order. In this case the squared asymptotic bias and the asymptotic variance are of the same order. If the bandwidth b is of higher order that is, with a small bandwidth – the result in Theorem 5(iv) also holds with $\Delta = 0$. Now the asymptotic bias is negligible. On the other hand, the asymptotic result will be dominated by the bias part, if the bandwidth b is of a smaller order. In this case, $b^{-k+\nu}(\hat{g}^{(\nu)}(t) - g^{(\nu)}(t))$ has a degenerate asymptotic distribution with a constant mean and variance zero.

4.4. The MISE

A well-known criterion for the quality of a nonparametric regression estimator is the mean integrated square error (MISE) defined by

$$\mathrm{MISE}(\hat{g}^{(\nu)}(x)) = \int_0^1 \mathrm{E}\{[\hat{g}^{(\nu)}(x) - g^{(\nu)}(x)]^2\}\mathrm{d}x. \tag{4.11}$$

For $p-\nu$ even, $\mathrm{MISE}(\hat{g}^{(\nu)}(x))$ is dominated by the estimation in the boundary area. For $p-\nu$ odd, the MISE due to the estimation in the boundary area is negligible. Let

$$I(g^{(k)}) = \int_0^1 [g^{(k)}(t)]^2\,\mathrm{d}t, \tag{4.12}$$

and denote by $K_{(\nu,k)}$ and $\beta_{(\nu,k)}$ respectively the equivalent kernel and the kernel constant for the interior points with $c = 1$. Then the following result holds.

Theorem 6. *Under the assumptions of Theorem* 5 *and for* $\delta \in (-0.5, 0.5)$*:*

(*i*) *the MISE of* $\hat{g}^{(\nu)}$ *is given by*

$$\begin{aligned}\int_0^1 &\mathrm{E}\{[\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)]^2\}\mathrm{d}t\\ &= \mathrm{MISE}_{\mathrm{asympt}}(n, b) + o(\max(b^{2(k-\nu)}, (nb)^{2\delta-1}b^{-2\nu}))\\ &= b^{2(k-\nu)}\frac{I(g^{(k)})\beta^2_{(\nu,p)}}{k!} + (nb)^{2\delta-1}b^{-2\nu}V(\delta) + o(\max(b^{2(k-\nu)}, (nb)^{2\delta-1}b^{-2\nu}));\end{aligned} \tag{4.13}$$

(*ii*) *the optimal bandwidth that minimizes the asymptotic MISE is given by*

$$b_{\mathrm{opt}} = C_{\mathrm{opt}}n^{(2\delta-1)/(2k+1-2\delta)}, \tag{4.14}$$

where

$$C_{\mathrm{opt}} = \left[\frac{2\nu+1-2\delta}{2(k-\nu)}\,\frac{(k!)^2V(\delta)}{I(g^{(k)})\beta^2_{(\nu,p)}}\right]^{1/(2k+1-2\delta)}, \tag{4.15}$$

in which it is assumed that $I(g^{(k)}) > 0$.

The proof of Theorem 6 will be omitted, since it is the same as for the case when the ϵ_i are an i.i.d. sequence (see Beran and Feng 1999).

Note that by inserting b_{opt} in (4.13), Theorem 2 implies that for $p-\nu$ odd the optimal MISE is of order

$$\int_0^1 \mathrm{E}\{[\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)]^2\}\mathrm{d}t = O(n^{2(2\delta-1)(k-\nu)/(2k+1-2\delta)}). \tag{4.16}$$

The rate of convergence of $\hat{g}^{(\nu)}$ is $n^{(2\delta-1)(k-\nu)/(2k+1-2\delta)} = n^{(2\delta-1)(p+1-\nu)/(2p+3-2\delta)}$. For $\nu = 0$ with $\delta \geqslant 0$, Hall and Hart (1990) show that this is the optimal convergence rate.

5. Final remarks

The ARCH and GARCH models proposed by Engle (1982) and Bollerslev (1986) have become a widely used model for analysing financial time series. Ling and Li (1997) showed the potential usefulness of the FARIMA-GARCH model. The semi-parametric FARIMA-GARCH model proposed in this paper is expected to become a useful tool for modelling stochastic processes with trends, long memory as well as conditional heteroscedasticity. Examples for modeling financial time series with the related SEMIFAR model proposed by Beran (1999) may be found in, for example, Beran and Ocker (1999).

To estimate the whole model one has to combine the proposal here and the approach for estimating the parameters, which determine the stochastic structure of the model, as proposed in Beran (1995; 1999) and Ling and Li (1997). This will be discussed elsewhere.

Acknowledgements

This research was supported in part by the Center of Finance and Econometrics at the University of Konstanz and by a National Science Foundation (SBIR, phase 2) grant to MathSoft, Inc. We would like to thank Dr Jonathan R.M. Hosking (IBM Research Division, Yorktown Heights) for allowing us to cite his helpful research report. We would also like to thank Dr Shiqing Ling (Hong Kong University of Science and Technology) for sending us his thesis. Finally, our thanks go to one of the editors and two referees for their useful comments, which helped to improve the quality of the paper.

Appendix: Proofs of theorems

Suppose that on the probability space $(\Omega, \mathscr{F}, P)$ there are given martingale differences

$$\xi^n = (\xi_{nk}, \mathscr{F}_{nk}), \qquad 0 \leqslant k \leqslant n,\ n \geqslant 1,$$

with $\xi_{n0} = 0$, $\mathscr{F}_{n0} = (\Phi, \Omega)$, $\mathscr{F}_{nk} \subseteq \mathscr{F}_{n,k+1} \subseteq \mathscr{F}$. Set

$$S_{nk} = \sum_{i=0}^{k} \xi_{ni}, \qquad 1 \leqslant k \leqslant n.$$

The double sequence $\{S_{nk}, \mathscr{F}_{nk}, 1 \leqslant k \leqslant n, n \geqslant 1\}$ will be called a martingale array.

To prove cases 1 and 2 of Theorem 1 we will use the following Lemmas A.1 and A.2, respectively. Lemma A.1 is a special case of Theorem 4 of Shiryaev (1996, Section VII.8) – see also Corollary 3.1 of Hall and Heyde (1980) and Corollary 6 of Liptser and Shiryaev (1980). Denote by $I(A)$ the indicator function of a set A.

Lemma A.1. *Let the square-integrable martingale differences* $\xi^n = (\xi_{nk}, \mathscr{F}_{nk})$, $n \geqslant 1$, *satisfy the conditional Lindeberg condition that for each* $\epsilon > 0$,

$$\sum_{k=0}^{n} \mathrm{E}[\xi_{nk}^2 I(|\xi_{nk}| > \epsilon)|\mathscr{F}_{n,k-1}] \xrightarrow{P} 0, \tag{A.1}$$

and the condition

$$\sum_{k=0}^{n} \mathrm{E}(\xi_{nk|\mathscr{F}_{n,k-1}}^2) \xrightarrow{P} 1. \tag{A.2}$$

Then $S_{nn} \xrightarrow{\mathscr{D}} N(0, 1)$.

Lemma A.2 is a special case of Theorem 3.2 of Hall and Heyde (1980). This lemma is used here to avoid the need to check the conditional Lindeberg condition (A.1).

Lemma A.2. *Let* $\{S_{nk}, \mathscr{F}_{nk}, 1 \leqslant k \leqslant n, n \geqslant 1\}$ *be a zero-mean, square-integrable martingale array with differences* ξ_{nk}. *Suppose that*

$$\max_{1 \leqslant k \leqslant n} |\xi_{nk}| \xrightarrow{P} 0, \tag{A.3}$$

$$\sum_{k=1}^{n} \xi_{nk}^2 \xrightarrow{P} 1 \tag{A.4}$$

and

$$\mathrm{E}(\max_{1 \leqslant k \leqslant n} \xi_{nk}^2) \textit{ is bounded in } n. \tag{A.5}$$

Then $S_{nn} \xrightarrow{\mathscr{D}} N(0, 1)$.

Proof of Theorem 1. Let $\sigma_n^2 = \mathrm{E}(X_1 + \ldots + X_n)^2$. Suppose that $\mathrm{E}(\epsilon_0^2) = 1$ for simplicity. Following the proof of Theorem 18.6.5 in Ibragimov and Linnik (1971), we have

$$\begin{aligned} \sigma_n^2 &= \mathrm{E}(X_1 + \ldots + X_n)^2 \\ &= \sum_{k=-\infty}^{\infty} (c_{k-1} + \ldots + c_{k-n})^2 \\ &= \sum_{k=-\infty}^{\infty} c_{k,n}^2, \end{aligned}$$

say. Hosking (1994) gave some corrections of the proof of Ibragimov and Linnik (1971) and showed that

$$\frac{|c_{k,n}|}{\sigma_n} \leqslant a_n := \left[8\sigma_n^{-1} \left\{ \left(\sum_{i=-\infty}^{\infty} c_i^2 \right)^{1/2} + \frac{1}{2}\sigma_n^{-1} \sum_{i=-\infty}^{\infty} c_i^2 \right\} \right]^{1/2}, \tag{A.6}$$

that is, $a_{k,n} := c_{k,n}/\sigma_n$ tends to zero uniformly in k as $n \to \infty$. Following Hosking (1994), we have

$$\sigma_n^{-1}(X_1 + \ldots + X_n) = \sum_{k=-\infty}^{\infty} a_{k,n}\epsilon_k$$

with

$$\sum_{k=-\infty}^{\infty} a_{k,n}^2 = 1.$$

For each $n \geqslant 1$ let $n_1 = -[(n-1)/2]$, $n_2 = [n/2]$ such that $n_1 \leqslant 0$, $n_2 \geqslant 0$ and $n_1 + n_2 + 1 = n$, where $[\cdot]$ denotes the integer part. For $k = 1, \ldots, n$ denote ϵ_{n_1+k-1} by ϵ_{nk}, $\mathscr{F}_{n_1+k-1}$ by $\mathscr{F}_{nk}$ and $a_{n_1+k-1,n}$ by b_{nk} for convenience. Define $\epsilon_{n0} = 0$. Also define $\xi^n = (\xi_{nk}, \mathscr{F}_{nk})$ with $\xi_{n0} = 0$ and $\xi_{nk} = b_{nk}\epsilon_{nk}$ for $1 \leqslant k \leqslant n$. Then we have

$$\begin{aligned}\sigma_n^{-1}(X_1 + \ldots + X_n) &= \sum_{k=-\infty}^{\infty} a_{k,n}\epsilon_k \\ &= \sum_{k=n_1}^{n_2} a_{k,n}\epsilon_k + \sum_{\substack{k<n_1\\k>n_2}} a_{k,n}\epsilon_k \\ &= S_{nn} + \eta_n, \end{aligned} \tag{A.7}$$

where

$$S_{nk} = \sum_{i=n_1}^{n_1+k-1} a_{i,n}\epsilon_i = \sum_{i=0}^{k} \xi_{ni}, \qquad k = 1, \ldots, n,$$

and

$$\eta_n = \sum_{\substack{k<n_1\\k>n_2}} a_{k,n}\epsilon_k.$$

It is clear that $\eta_n = o_p(1)$, since $\mathrm{E}(\eta_n) = 0$ and

$$\mathrm{var}(\eta_n) = \sum_{\substack{k<n_1\\k>n_2}} a_{k,n}^2 \to 0, \qquad \text{as } n \to \infty.$$

Here, $\xi^n = (\xi_{nk}, \mathscr{F}_{nk})$ is a square-integrable martingale difference and $\{S_{nk}, \mathscr{F}_{nk}, 1 \leqslant k \leqslant n, n \geqslant 1\}$ is a zero-mean square-integrable martingale array. It remains to show that, in Theorem 1, case 1, $\xi^n = (\xi_{nk}, \mathscr{F}_{nk})$ fulfil the conditions of Lemma A.1 and, in Theorem 1, case 2, $\{S_{nk}, \mathscr{F}_{nk}, 1 \leqslant k \leqslant n, n \geqslant 1\}$ fulfil those of Lemma A.2, respectively.

Case 1. In this case it is easy to show that the square-integrable martingale differences $\xi^n = (\xi_{nk}, \mathscr{F}_{nk})$ satisfy conditions (A.1) and (A.2). We have $\mathrm{E}(\xi_{nk}^2 | \mathscr{F}_{n,k-1}) = b_{nk}$ and hence

$$\sum_{k=0}^{n} \mathrm{E}(\xi_{nk}^2 | \mathscr{F}_{n,k-1}) = \sum_{k=n_1}^{n_2} b_{nk}^2 = \sum_{k=-\infty}^{\infty} a_{k,n}^2 + o(1) \to 1,$$

and (A.2) is satisfied. Furthermore, using (A.6) and noting that $\sum b_{nk}^2 \leqslant 1$, we have

$$\begin{aligned}\sum_{k=0}^{n} \mathrm{E}[\xi_{nk}^2 I(|\xi_{nk}| > \epsilon) | \mathscr{F}_{n,k-1}] &= \sum_{k=0}^{n} b_{nk}^2 \mathrm{E}[\epsilon_0^2 I(|\epsilon_0| > \epsilon / b_{nk})] \\ &\leqslant \sum_{k=0}^{n} b_{nk}^2 \mathrm{E}[\epsilon_0^2 I(|\epsilon_0| > \epsilon / a_n)] \\ &\leqslant \mathrm{E}[\epsilon_0^2 I(|\epsilon_0| > \epsilon / a_n)] \to 0.\end{aligned}$$

This shows that $\xi^n = (\xi_{nk}, \mathscr{F}_{nk})$ satisfy (A.1).

Case 2. Now, we have to check that $\{S_{nk}, \mathscr{F}_{nk}, 1 \leqslant k \leqslant n, n \geqslant 1\}$ fulfils conditions (A.3)–(A.5). Let $\tilde{\epsilon}_i$ denote an associate independent sequence for ϵ_i, that is, an i.i.d. sequence with the same marginal distribution. Using the Chebyshev's inequality we can obtain the tail behaviours of $|\tilde{\epsilon}_i|$ and $\tilde{\epsilon}_i^2$, that is, $P(|\tilde{\epsilon}_i| > x) \leqslant \mathrm{E}(\tilde{\epsilon}^4)/x^4$ and $P(\tilde{\epsilon}_i^2 > x) \leqslant \mathrm{E}(\tilde{\epsilon}^4)/x^2$ for every x. Following results in extreme value theory (see, for example, Embrechts *et al.* 1997) we have $\max_{1 \leqslant i \leqslant n} |\tilde{\epsilon}_i| = O_p(n^{1/4})$ and $\mathrm{E}(\max_{1 \leqslant i \leqslant n} \tilde{\epsilon}_i^2) = O(n^{1/2})$. Under the assumption that $|\epsilon_i|$ has an extreme index $\theta \in (0, 1]$, the same results hold for $|\epsilon_i|$ and ϵ_i^2, respectively: $\max_{1 \leqslant i \leqslant n} |\epsilon_i| = O_p(n^{1/4})$ and $\mathrm{E}(\max_{1 \leqslant i \leqslant n} \epsilon_i^2) = O(n^{1/2})$.

For $\delta > 0$ we have $a_n = O(\sigma_n^{-1/2}) = O(n^{-1/4-\delta/2})$. Using (A.6),

$$\begin{aligned}\max_{1 \leqslant k \leqslant n} |\xi_{nk}| &\leqslant a_n \max_{1 \leqslant k \leqslant n} |\epsilon_{nk}| \\ &= O(n^{-1/4-\delta/2}) O_p(n^{1/4}) = o_p(1).\end{aligned}$$

If $\delta = 0$ we have $c_{k,n} = O(1)$, $a_{k,n} = c_{k,n}/\sigma_n = O(\sigma_n^{-1}) = O(n^{-1/2})$. If $\delta < 0$, it can be shown that, for $n_1 \leqslant k \leqslant n_2$, $c_{k,n} = O(n^{\delta})$ and $a_{k,n} = O(n^{\delta}) O(\sigma_n^{-1}) = O(n^{-1/2})$. Hence we have, for $\delta \leqslant 0$,

$$\begin{aligned}\max_{1 \leqslant k \leqslant n} |\xi_{nk}| &= O(n^{-1/2}) \max_{1 \leqslant k \leqslant n} |\epsilon_{nk}| \\ &= O(n^{-1/2}) O_p(n^{1/4}) = o_p(1).\end{aligned}$$

This shows that (A.3) holds for $\delta \in (-0.5, 0.5)$.

In this case ϵ_i^2 is a second-order stationary process with $\mathrm{E}(\epsilon_0^2) = 1$ and $\gamma_{\epsilon^2}(k) = \mathrm{cov}(\epsilon_i^2, \epsilon_{i+k}^2) \to 0$ as $k \to \infty$. Observing that the weights b_{nk}^2 satisfy the conditions of Corollary 1, we have

$$\sum_{k=1}^{n} \xi_{nk}^2 = \sum_{k=1}^{n} b_{nk}^2 \epsilon_{nk}^2 \xrightarrow{P} \mathrm{E}(\epsilon_0^2) = 1,$$

Thus (A.4) is fulfilled.

We will now show that (A.5) holds. In fact we have $\mathrm{E}(\max_{1\leqslant k\leqslant n}\xi_{nk}^2) = o(1)$. This will only be shown for $\delta > 0$:

$$\mathrm{E}(\max_{1\leqslant k\leqslant n} \xi_{nk}^2) \leqslant a_n^2 \mathrm{E}(\max_{1\leqslant k\leqslant n} \epsilon_{nk}^2)$$

$$= O(n^{-1/2-\delta})O(n^{1/2}) = o(1).$$

This completes the proof. □

Proof of Theorem 2. Again, we put $\mathrm{var}(X_i) = 1$ for convenience. In this case we have that $|\gamma(k)| \leqslant 1$. It is clear that Theorem 2 holds in the naive case with $w_i \equiv 0$. Otherwise we have $\max_{1\leqslant i\leqslant n}|w_i| > 0$. Now let $N = N_n = (\max_{1\leqslant i\leqslant n}|w_i|)^{-1/2}$, such that $N \to \infty$, $N \cdot \max_{1\leqslant i\leqslant n} |w_i| \to 0$ as $n \to \infty$. Then

$$\mathrm{var}\left(\sum_{i=1}^{n} w_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(i-j)$$

$$= \sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \gamma(i-j)$$

$$= \sum_{i=1}^{n} w_i \left[\sum_{|i-j|\leqslant N} w_j \gamma(i-j) + \sum_{|i-j|>N} w_j \gamma(i-j)\right]$$

$$\leqslant \sum_{i=1}^{n} |w_i| \left[\sum_{|i-j|\leqslant N} |w_j|\,|\gamma(i-j)| + \sum_{|i-j|>N} |w_j|\,|\gamma(i-j)|\right].$$

Observing that $\sum_{i=-\infty}^{\infty}|w_i| < \infty$, it is sufficient to show that

$$\sum_{|i-j|\leqslant N} |w_j|\,|\gamma(i-j)| + \sum_{|i-j|>N} |w_j|\,|\gamma(i-j)| = o(1) \tag{A.8}$$

holds uniformly in i. Consider the first part of the left-hand side of (A.8):

$$\sum_{|i-j|\leqslant N} |w_j|\,|\gamma(i-j)| \leqslant (2N+1) \max_{1\leqslant i\leqslant n} |w_i|$$

$$= O((\max_{1\leqslant i\leqslant n} |w_i|)^{1/2}) = o(1). \tag{A.9}$$

For the second part we have

$$\sum_{|i-j|>N} |w_j|\,|\gamma(i-j) \leqslant \sup_{|k|>N} \gamma(k)| \sum_{|i-j|>N} |w_j|$$

$$= O(\sup_{|k|>N} |\gamma(k)|) = o(1), \tag{A.10}$$

completing the proof of Theorem 2. □

Proof of Corollary 1. By Theorem 2, we have $\mathrm{var}(\hat{\mu}) \to 0$ as $n \to \infty$. Since $\sum_{k=1}^{n} w_k \to 1$ as $n \to \infty$, we have

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}\left(\sum_{k=1}^{n} w_k \epsilon_k\right) \to \mu \qquad \text{as } n \to \infty.$$ □

Proof of Theorem 3. The formula for the asymptotic variance of $\overline{X}$ remains unchanged from case to case only if ϵ_i are uncorrelated $(0, \sigma^2)$ random variables. Hence, it is the same as that for i.i.d. innovations given by Theorems 1 and 8 of Hosking (1996), that is, $\mathrm{var}(\overline{X}) = n^{2\delta-1} V_\delta$ for $-\frac{1}{2} < \delta < \frac{1}{2}$, where

$$V_\delta = \sigma_\epsilon^2 \frac{|\psi(1)|^2}{|\phi(1)|^2} \frac{\Gamma(1-2\delta)}{(2\delta+1)} \frac{1}{\Gamma(1+\delta)\Gamma(1-\delta)}.$$

Using the relationships $\Gamma(1+\delta) = \delta\Gamma(\delta)$ and $\Gamma(\delta)\Gamma(1-\delta) = \pi/\sin(\pi\delta)$ (for $0 < \delta < 0.5$), we obtain the alternative representation of V_δ,

$$V_\delta = \sigma_\epsilon^2 \frac{|\psi(1)|^2}{|\phi(1)|^2} \frac{\Gamma(1-2\delta)}{(2\delta+1)} \frac{\sin(\pi\delta)}{\pi\delta},$$

which is used in this paper.

Since X_i defined in (1.2)–(1.3) is a zero-mean FARIMA process with innovations ϵ_i following a GARCH model, we have

$$X_i = \sum_{k=0}^{\infty} c_k \epsilon_{i-k} \tag{A.11}$$

with $c_k \sim (|\psi(1)|/|\phi(1)|) k^{\delta-1}$ as for large K (see, for example, Brockwell and Davis 1991; Beran 1994). Hence, for $-0.5 < \delta < 0.5$, $\sum_{k=0}^{\infty} c_k^2 < \infty$. This, together with the context above Theorem 3, shows that X_i fulfils the conditions of Theorem 1, case 2, and so $(X_1 + \ldots + X_n)/\sigma_n \xrightarrow{D} N(0, 1)$. Observing that $[n^{1/2-\delta}\overline{X} - (X_1 + \ldots + X_n)/\sigma_n] \xrightarrow{P} 0$, we have $n^{1/2-\delta}\overline{X} \xrightarrow{D} N(0, 1)$. □

Proof of Theorem 4. In order to prove Theorem 4 we only need to show that the decomposition (A.7) holds for proper $a_{k,n}$. Following (3.1), the weighted sum can be rewritten as

$$\sum_{i=1}^{n} w_i X_i = \sum_{i=1}^{n} w_i \left(\sum_{k=-\infty}^{\infty} c_{i-k}\epsilon_k \right)$$
$$= \sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{n} w_i c_{k-i} \right) \epsilon_k$$
$$=: \sum_{k=-\infty}^{\infty} c_{k,n}\epsilon_k, \tag{A.12}$$

where $c_{k,n} = (w_1 c_{k-1} + \ldots + w_n c_{k-n})$. Noting that ϵ_k are uncorrelated random variables, we have

$$\sigma_n^2 := \mathrm{E}\left(\sum_{i=1}^{n} w_i X_i \right)^2 = \sum_{k=-\infty}^{\infty} c_{k,n}^2. \tag{A.13}$$

Define $a_{k,n} = c_{k,n}/\sigma_n$, we have

$$\sigma_n^{-1} \sum_{i=1}^{n} w_i X_i = \sum_{k=-\infty}^{\infty} a_{k,n}\epsilon_k,$$

with

$$\sum_{k=-\infty}^{\infty} a_{k,n}^2 = 1.$$

The uniform negligibility of $a_{k,n}$ – which means that it tends to zero uniformly in k as $n \to \infty$ – is guaranteed by condition (4.3). The rest part of the proof of Theorem 4 is the same as that of Theorem 1. □

Proof of Theorem 5. The proof of the first three parts will be omitted (see Beran and Feng 1999). Note that

$$\hat{g}^{(\nu)}(t) - g^{(\nu)}(t) = \sum_{i=1}^{n} w_i X_i.$$

The weights of $\hat{g}^{(\nu)}$ generated by local polynomial fitting have the properties that $\max_{1\leqslant i\leqslant n}|w_i| = O[(nb^{1+\nu})^{-1}]$ and $w_i \equiv 0$ outside an interval with length of order nb. Using the result given in part (iii) we have $\max|w_i|/\sigma_n = O[(nb)^{-1/2-\delta}] \to 0$ as $n \to \infty$. Noting that $c_k \sim k^{\delta-1}$,

$$\left| \sum_{i=1}^{n} w_i c_{k-i} \right| / \sigma_n \leqslant \max_{1\leqslant i\leqslant n} |w_i| \left[\sum_{i=1}^{n} |c_{k-i}| \Big| w_i \neq 0 \right] / \sigma_n$$
$$= O[(nb)^{-1/2-\delta}] O[(nb)^{\delta}]$$
$$= O[(nb)^{-1/2}] \to 0 \qquad \text{as } n \to \infty.$$

Conditions (4.2) and (4.3) are satisfied by the weights of $\hat{g}^{(\nu)}(t)$. We have

$$\frac{\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)}{\operatorname{var}(\hat{g}^{(\nu)}(t))} \xrightarrow{D} N(0, 1).$$

When the bias has the representation (4.6), and assuming that $nb^{(2k+1-2\delta)/(1-2\delta)} \to d^2$, for some $d > 0$, we obtain

$$(nb)^{1/2-\delta} b^{\nu}(\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)) \xrightarrow{\mathscr{D}} N(d\Delta, V(c, \delta)),$$

where $\Delta = g^{(k)}(t)\beta_{(\nu,k,c)}/k!$ and $V(c, \delta)$ is the constant defined in (4.7). □

References

Beran, J. (1994) *Statistics for Long Memory Processes*. London: Chapman & Hall.

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B*, **57**, 659–672.

Beran, J. (1999) SEMIFAR models – a semi-parametric fractional framework for modelling trends, long-range dependence and nonstationarity. Preprint, University of Konstanz.

Beran, J. and Feng, Y. (1999) Local polynomial fitting with long-memory, short-memory and anti-persistent errors. Preprint, University of Konstanz.

Beran, J. and Ocker, D. (1999) SEMIFAR forecasts, with applications to foreign exchange rates. *J. Statist. Plann. and Inference*, **80**, 137–153.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *J. Econometrics*, **31**, 307–327.

Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods*, 2nd edn. Berlin: Springer-Verlag.

Chen, M. and An, H.Z. (1998) A note on the stationarity and the existence of moments of the GARCH model. *Statist. Sinica*, **8**, 505–510.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, **74**, 829–836.

Csörgő, S. and Mielniczuk, J. (1995) Nonparametric regression under long-range dependent normal errors. *Ann. Statist.*, **23**, 1000–1014.

Davis, R.A., Mikosch, T. and Basrak, B. (1999) Sample ACF of multivariate stochastic recurrence equations with application to GARCH. Preprint, University of Groningen.

de Haan, L., Resnick, S.I., Rootzén, H. and de Vries, C.G. (1989) Extremal behaviour of stochastic difference equation with application to ARCH processes. *Stochastic Process. Appl.*, **32**, 213–224.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimation of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1008.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

Feng, Y. (1999) *Kernel- and Locally Weighted Regression*. Berlin: Verlag für Wissenschaft und Forschung.

Gasser, T. and Müller, H.G. (1979) Kernel estimation of regression functions. In T. Gasser and M. Rosenblatt (eds), *Smoothing Techniques for Curve Estimation*, Lecture Notes in Math. 757, pp. 23–68. Berlin: Springer-Verlag.

Granger, C.W.J. and Joyeux, R. (1980) An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.*, **1**, 15–30.

Hall, P. and Hart, J.D. (1990) Nonparametric regression with long-range dependence. *Stochastic Process. Appl.*, **36**, 339–351.

Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.

He, C. and Teräsvirta, T. (1999a) Fourth moment structure of the GARCH(p, q) process. *Econometric Theory*, **15**, 824–846.

He, C. and Teräsvirta, T. (1999b) Properties of the autocorrelation function of squared observations for second-order GARCH processes under two sets of parameter constraints. *J. Time Ser. Anal.*, **20**, 23–30.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika*, **68**, 165–176.

Hosking, J.R.M. (1994) Supplement to 'Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series'. Research Report, RC 19831, IBM Research Division, Yorktown Heights, NY.

Hosking, J.R.M. (1996) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econometrics*, **73**, 261–284.

Ibragimov, I.A. and Linnik, Yu.V. (1971) *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Leadbetter, M.R. (1983) Extremes and local dependence in stationary sequences. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **65**, 291–306.

Ling, S. (1998) Stationary and nonstationary time series models with conditional heteroskedasticity. Ph.D. thesis, University of Hong Kong.

Ling, S. (1999) On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.*, **36**, 688–705.

Ling, S. and Li, W.K. (1997) On fractional integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, **92**, 1184–1194.

Ling, S. and McAleer, M. (2000) Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory*. To appear.

Liptser, R.S. and Shiryaev, A.N. (1980) A functional central limit theorem for semimartingales. *Theory Probab. Appl.*, **25**, 667–688.

Mikosch, T. and Stărică, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process. *Ann. Statist.*, **28**, 1427–1451.

Müller, H.-G. (1987) Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.*, **82**, 231–238.

Müller, H.-G. (1988) *Nonparametric Regression Analysis of Longitudinal Data*. Berlin: Springer-Verlag.

Ruppert, D. and Wand, M.P. (1994) Multivariate locally weighted least squares regression. *Ann. Statist.*, **22**, 1346–1370.

Shiryaev, A.N (1996) *Probability*, 2nd edn. New York: Springer-Verlag.

Shiryaev, A.N (1999) *Essentials of Stochastic Finance – Facts, Models, Theory.* Singapore: World Scientific.

Stone, C.J. (1977) Consistent nonparametric regression (with discussion). *Ann. Statist.*, **5**, 595–620.

Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **31**, 287–302.

Received May 1999 and revised June 2001.