Local Polynomial Estimation with a FARIMA-GARCH Error Process

By Jan Beran and Yuanhua Feng

University of Konstanz

Abstract

This paper considers a class of semiparametric models being the sum of a nonparametric trend function g and a FARIMA-GARCH error process. Estimation of $g^{(\nu)}$, the νth derivative of g, by local polynomial fitting is investigated. The focus is on the derivation of the asymptotic normality of $\hat{g}^{(\nu)}$. At first a central limit theorem based on martingale theory is developed and asymptotic normality of the sample mean of a FARIMA-GARCH process is proved. The central limit theorem is then extended from the case of an unweighted sum to a weighted sum in order 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. Asymptotic results on $\hat{g}^{(\nu)}$ in the presentation of long memory as well as antipersistence are also given for the current model.

Keywords. Local polynomial estimation, FARIMA-GARCH process, semiparametric models, long memory, square-integrable martingale-difference, asymptotic normality.

1 Introduction

In this paper a semiparametric model of the form $Y_i = g(t_i) + X_i$ is considered, where $g(t)$ is a smooth nonparametric regression function and the error process X_i follows a FARIMA-GARCH (fractional autoregressive integrated moving average - generalized autoregressive conditional heteroskedastic) model. We call this a semiparametric FARIMA-GARCH model. Such a model allows for simultaneous estimation of trend, long memory as well as conditional heteroskedasticity in a time series (see Beran 1994 for the definition of long memory processes and see Engle 1982 and Bollerslev 1986 for time series models with conditional heteroskedasticity). Estimation of $g^{(\nu)}$, the

νth derivative of g, leads to a nonparametric regression problem with a specified long memory error process. Recent research on the topic of nonparametric regression with long memory errors may be found in Hall and Hart (1990), Csörgö and Mielniczuk (1995), Beran (1999) and Beran and Feng (1999).

The most popular stationary long memory process is the FARIMA model proposed by Granger and Joyeux (1980) and Hosking (1981). On the other hand, to analyze time series with conditional heteroskedasticity, Engle (1982) introduced the ARCH (autoregressive conditional heteroskedastic) model, which was generalized by Bollerslev (1986) to the so-called GARCH model. This class of models has important applications, in particular to the analysis of financial time series. Ling and Li (1997) (see also Ling 1998, 1999) proposed a so-called FARIMA-GARCH model so that long memory and conditional heteroskedasticity may be analyzed in a unified approach. Their approach generalizes FARIMA introduced in Beran (1995). Following Ling and Li (1997) and Ling (1998, 1999) a FARIMA(l, δ, m)-GARCH(r, s) model is defined to be a discrete time series Y_i that satisfies the following equation:

$$\phi(B)(1-B)^{\delta}\{Y_i - \mu\} = \psi(B)\epsilon_i, \tag{1.1}$$

$$\epsilon_i = z_i h_i^{\frac{1}{2}}, \quad h_i = \alpha_0 + \sum_{j=1}^{r} \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^{s} \beta_k h_{i-k}, \tag{1.2}$$

where z_i are iid random variables with zero mean and variance 1, $\alpha_0 > 0$, $\alpha_1, ..., \alpha_r, \beta_1, ..., \beta_s \geq 0$, r and s are nonnegative integers, $\delta \in (-0.5, 0.5)$ is a real number, B is the backshift operator, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_l B^l$ and $\psi(B) = 1 + \psi_1 B + \cdots + \psi_m B^m$ are polynomials in B with no common factors and all roots outside the unit circle and, l and m are nonnegative integers. Furthermore, it is assumed that $\sum_{j=1}^{r} \alpha_j + \sum_{k=1}^{s} \beta_k < 1$. Here, the fractional difference $(1-B)^{\delta}$ introduced by Granger and Joyeux (1980) and Hosking (1981) is defined by

$$(1-B)^{\delta} = \sum_{k=0}^{\infty} b_k B^k \tag{1.3}$$

with

$$b_k(\delta) = \frac{\Gamma(k-\delta)}{\Gamma(k+1)\Gamma(-\delta)}. \tag{1.4}$$

Remark 1. Note that (1.2) defines a GARCH process in a sense wider than the original GARCH model defined by Bollerslev (1986), where ϵ_i are assumed to be conditionally normal. Conditional normality is also assumed in Ling and Li (1997).

In our paper this is not required. The definition (1.2) follows (6.1)-(6.2) in Ling (1999). If z_i are iid standard normal random variables, then (1.1)-(1.2) reduces to the definition as given in Ling and Li (1997).

Some time series, in particular financial time series, may however exhibit trend, long memory and conditional heteroskedasticity at the same time. Following the proposal of the SEMIFAR (semiparametric fractional autoregressive) model (see Beran 1995, 1999) we obtain the semiparametric FARIMA-GARCH model by introducing a nonparametric trend $g(t)$ in (1.1),

$$\phi(B)(1-B)^{\delta}\{Y_i - g(t_i)\} = \psi(B)\epsilon_i, \tag{1.5}$$

$$\epsilon_i = z_i h_i^{\frac{1}{2}}, \quad h_i = \alpha_0 + \sum_{j=1}^{r} \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^{s} \beta_k h_{i-k}, \tag{1.6}$$

where $t_i = (i/n)$ and $g : [0,1] \to \Re$ is a smooth function, l, m, r, s, α_0, $\alpha_1, \cdots, \alpha_r$, $\beta_1, \cdots, \beta_s$, δ, B, $\phi(B)$, $\psi(B)$ and z_i are as before. Model (1.5)-(1.6) provides an analytic tool for modeling time series with long-range dependence and conditional heteroskedasticity which is nonstationary in the mean.

The current paper focuses on investigating the asymptotic properties of local polynomial estimation of $g^{(\nu)}$ in (1.5)-(1.6). Unified consistency and asymptotic normality of $\hat{g}^{(\nu)}$ on the whole support $[0,1]$ are obtained for errors with short- or long memory as well as for errors with antipersistence. The rate of convergence of a pth order local polynomial estimator $\hat{g}^{(\nu)}$ with $p-\nu$ odd is shown to be $n^{(2\delta-1)(p+1-\nu)/(2p+3-2\delta)}$ for all $\delta \in (-0.5, 0.5)$. The asymptotic normality of $\hat{g}^{(\nu)}(t)$ is derived based on a central limit theorem for stationary processes with short- or long memory, which is an extension of theorem 18.6.5 in Ibragimov and Linnik (1971). Asymptotic results on the estimator $\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ under model (1.1)-(1.2) are also given. Similar to the SEMIFAR model, model (1.5)-(1.6) can also be extended to include stochastic trends by allowing $\delta > 0.5$.

The paper is organized as follows. The proposed local polynomial estimators are described in section 2. Section 3 gives some auxiliary results including a central limit theorem for stationary processes being a 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 put in the appendix.

2 The estimator

Assume that $\sum_{j=1}^{r}\alpha_j + \sum_{k=1}^{s}\beta_k < 1$. Then the process $X_i := Y_i - g(t_i)$ defined by model (1.5)-(1.6) is stationary, causal and invertible. Hence, model (1.5)-(1.6) can be rewritten as

$$Y_i = g(t_i) + X_i, \tag{2.1}$$

where

$$X_i = (1-B)^{-\delta}\phi^{-1}(B)\psi(B)\epsilon_i \tag{2.2}$$

with

$$\epsilon_i = z_i h_i^{\frac{1}{2}}, \quad h_i = \alpha_0 + \sum_{j=1}^{r}\alpha_j\epsilon_{i-j}^2 + \sum_{k=1}^{s}\beta_k h_{i-k}, \tag{2.3}$$

where $t_i = (i/n)$ and $g : [0,1] \to \Re$ is a smooth function, l, m, r, s, α_0, $\alpha_1, \cdots, \alpha_r$, $\beta_1, \cdots, \beta_s$, δ, B, $\phi(B)$, $\psi(B)$ and z_i are as before. The error process X_i can be rewritten as $X_i = (1-B)^{-\delta}U_i$ with $U_i = \phi^{-1}(B)\psi(B)\epsilon_i$. U_i is a stationary process having short memory so that $0 < V_0 := \sum_{k=-\infty}^{\infty}\mathrm{cov}\,(U_i, U_{i+k}) < \infty$. Hence, model (2.1)-(2.3) defines an equidistant nonparametric regression with short memory ($\delta = 0$), long memory ($0 < \delta < 0.5$) and antipersistence ($-0.5 < \delta < 0$).

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). The proposed approach in this paper follows the idea in Beran and Feng (1999). Local polynomial fitting has some advantages in comparison with kernel estimation. It is an automatic kernel method. A kernel estimator $\hat{g}$ is just a local constant estimator. Also, the estimation of the νth derivative of g by local polynomial fitting is very simple. For recent developments in this context of we refer the readers to Ruppert and Wand (1994), Wand and Jones (1995), and Fan and Gijbels (1995, 1996) and references therein.

Assume that g is at least $(p+1)$-times differentiable at a point t_0. Then $g(t)$ can be approximated locally by a polynomial of order p:

$$g(t) = g(t_0) + g'(t_0)(t-t_0) + ... + g^{(p)}(t_0)(t-t_0)^p/p! + R_p \tag{2.4}$$

for t in a neighborhood of t_0, where R_p is a remainder term. Let K be a symmetric density (a kernel of order two without boundary correction) having compact support $[-1,1]$. Giving n observations Y_1, ..., Y_n, we can obtain an estimator of $g^{(\nu)}$ $(\nu \leq p)$

by solving the locally weighted least squares problem

$$Q = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} b_j (t_i - t_0)^j \right\}^2 K\left(\frac{t_i - t_0}{b}\right) \Rightarrow \min, \tag{2.5}$$

where b is the bandwidth and K is called the weight function. Let $\hat{b} = (\hat{b}_0, \hat{b}_1, ..., \hat{b}_p)^{\mathrm{T}}$, then it is clear from (2.4) that $\nu!\hat{b}_\nu$ estimates $g^{(\nu)}(t_0)$, $\nu = 0, 1, ..., p$. Let

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

and let $\mathbf{e}_j$, $j = 1, ..., p+1$, 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. Finally, let $\mathbf{y} = (Y_1, ..., Y_n)^{\mathrm{T}}$. Then $\hat{g}^{(\nu)}(t_0)$ can be written as

$$\begin{aligned} \hat{g}^{(\nu)}(t_0) &= \nu!\mathbf{e}_{\nu+1}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{y} \\ &=: \{\mathbf{w}^{\nu}(t_0)\}^{\mathrm{T}}\mathbf{y}, \end{aligned} \tag{2.6}$$

where $\{\mathbf{w}^{\nu}(t_0)\}^{\mathrm{T}} = \nu!\mathbf{e}_{\nu+1}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}$ is called the weighting system. We see that $\hat{g}^{(\nu)}(t_0)$ is a linear smoother with the weighting system $\mathbf{w}^{\nu}(t_0) = (w_1^{\nu}, ..., w_n^{\nu})^{\mathrm{T}}$, where $w_i^{\nu} \neq 0$ only if $|t_i - t_0| \leq b$. 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, i.e. $\hat{g}^{(\nu)}$ works as a moving average in the interior. Furthermore, $\mathbf{w}^{\nu}(t_0)$ satisfies:

$$\sum_{i=1}^{n} w_i^{\nu}(t_i - t_0)^{\nu} = \nu! \text{ and } \sum_{i=1}^{n} w_i^{\nu}(t_i - t_0)^j = 0 \text{ for } j = 0, ..., p, j \neq \nu. \tag{2.7}$$

The property (2.7) ensures that $\hat{g}^{(\nu)}$ is exactly unbiased if g is a polynomial of order not larger than p.

Local polynomial fitting is asymptotically equivalent to some kernel estimation. In the interior, the difference in finite samples is also not large (see e.g. Müller 1987 and Feng 1998). These results also hold for the current case (see Beran and Feng, 1999), since the weights do not depend on the dependent structure of the data. This provides a powerful tool for deriving the asymptotic results of local polynomial fitting. In particular, we have $\max|w_i| = O[(nb^{1+\nu})]$, which will be used in the appendix.

3 Auxiliary results

To prove the asymptotic normality of $\hat{g}^{(\nu)}$ we need some auxiliary results. In the following we will develop a central limit theorem 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. If ϵ_k are iid $(0, \sigma^2)$ random variables, the result is given by theorem 18.6.5 of Ibragimov and Linnik (1971). However, their result (and hence theorems 2 and 8 in Hosking (1996) based on this result) can not be applied to the case when ϵ_i defined in (2.1)-(2.3) follow a GARCH or ARCH process, since now they are conditionally heteroskedastic and not independent.

Here, martingales and martingale-differences are defined as follows (see e.g. Hall and Heyde 1980 and Shiryaev 1996). Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $\mathcal{F}$ is a σ-field of subsets of Ω. Let $\mathbf{I}$ be any interval of the form (a, b), $[a, b)$, $(a, b]$ or $[a, b]$ of the ordered set $\{-\infty, ..., -1, 0, 1, ..., \infty\}$. Let $\{\mathcal{F}_i, i \in \mathbf{I}\}$ be a nondecreasing sequence of σ-fields of $\mathcal{F}$ sets. A stochastic sequence $M = (M_i, \mathcal{F}_i, i \in \mathbf{I})$ is said to be a martingale (with respect to $\mathcal{F}_i$), if

$$E(|M_i|) < \infty \quad \text{and} \quad E(M_i|\mathcal{F}_{i-1}) = M_{i-1} \text{ (a.s.).}$$

M is said to be square-integrable, if $E(M_i^2) < \infty$. A stochastic sequence $\xi = (\xi_i, \mathcal{F}_i, i \in \mathbf{I})$ is said to be a martingale-difference, if

$$E(|\xi_i|) < \infty \quad \text{and} \quad E(\xi_i|\mathcal{F}_{i-1}) = 0 \text{ (a.s.).}$$

A martingale-difference with finite variance is called a square-integrable martingale-difference.

The following central limit theorem extends theorem 18.6.5 of Ibragimov and Linnik (1971) to the case when the innovations form a square-integrable martingale-difference satisfying given conditions, which includes the GARCH process as a special case.

Theorem 1. *Let the sequence $(\epsilon_i, F_i, i \in \mathbf{I} = \{-\infty, ..., -1, 0, 1, ..., \infty\})$ be a square-integrable martingale-difference with identical variance, i.e. $E(\epsilon_i|\mathcal{F}_{i-1}) = 0$, $E(\epsilon_i^2) = E(\epsilon_0^2) < \infty$. Let*

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

where

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

Assume that $\sigma_n^2 = E(X_1 + ... + X_n)^2 \to \infty$ *as* $n \to \infty$.

Case 1. *Let* $h_i = E(\epsilon_i^2 | \mathcal{F}_{i-1})$ *be the conditional variance of* ϵ_i. *If* $h_i \equiv E(\epsilon_0^2)$, *then*
$(X_1 + ... + X_n)/\sigma_n \overset{\mathcal{D}}{\to} N(0,1)$.

Case 2. *Suppose that* ϵ_i^2 *is a second order stationary process with autocovariances* $\gamma_{\epsilon^2}(k) = cov(\epsilon_i^2, \epsilon_{i+k}^2)$. *If* $\gamma_{\epsilon^2}(k) \to 0$ *as* $k \to \infty$, *then*
$(X_1 + ... + X_n)/\sigma_n \overset{\mathcal{D}}{\to} N(0,1)$.

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

Although the conditions of theorem 1 are much weaker than those used by Ibragimov and Linnik (1971), we do not declare that they are necessary. Some of them are made here for simplifying the proof. Noting that a square-integrable martingale-difference is a sequence of uncorrelated random variables (see Shiryaev 1999, p. 42), the assumption that ϵ_i have identical variance implies that it is an uncorrelated white noise. Hence, the assumptions on ϵ_i given in theorem 1 are stronger than that ϵ_i is an uncorrelated white noise but much weaker than that ϵ_i are iid random variables. For long memory process the assumption of an uncorrelated $(0, \sigma^2)$ random variables is not sufficient for the derivation of asymptotic normality of the sample mean (see e.g. Taqqu 1975).

The proof of theorem 1 is given in the appendix. To prove this theorem under the condition given in case 2, we need to prove the weak consistency of a weighted sum of the second order stationary time series ϵ_i^2. This is just a very special case of the following results on the convergence of the variance (to zero) of a general linear filter and the weak consistency of a general weighted sum, both are given for second order stationary time series.

Theorem 2. *Let* $(X_{i,n})$, $1 \le i \le n$, $n = 1, 2, ...$ *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* $\max_{1\le i\le n}|w_i| \to 0$ *as* $n \to \infty$, *then* $var(\sum_{i=1}^{n} w_i X_i) \to 0$ *as* $n \to \infty$.

The proof of theorem 2 is given in the appendix. The weighting system w_i is "formless"; 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 of the weighting systems generated by commonly used kernel estimators or local polynomial estimators of $g^{(\nu)}$ are included as special cases of theorem 2. This shows 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* $(w_{i,n})$, $1 \le i \le n$, $n = 1, 2, ...$ *be the triangular array as defined in theorem 2. Suppose that now 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

In the following we will show that the GARCH process fulfills the conditions on ϵ_i given in theorem 1 (case 2). Bollerslev (1986) showed that, under the condition $\sum_{j=1}^{r}\alpha_j + \sum_{k=1}^{s}\beta_k < 1$, a GARCH process is second order stationary with $E(\epsilon_i) = 0$, $\text{var}(\epsilon_i) = \alpha_0(1 - \sum_{j=1}^{r}\alpha_j - \sum_{k=1}^{s}\beta_k)^{-1}$ and $\text{cov}(\epsilon_i, \epsilon_{i+k}) = 0$ for $k \ne 0$. Note that the FARIMA process X_i can also be defined by taking the innovations ϵ_i to be an uncorrelated white noise, i.e. uncorrelated $(0, \sigma^2)$ random variables, called a FARIMA process in the wide sense (see e.g. Brockwell and Davis, 1991). Hence, the FARIMA-GARCH process is a special case of a FARIMA process in the wide sense, since a GARCH process is an uncorrelated white noise. It is easy to show that under the condition $\sum_{j=1}^{r}\alpha_j + \sum_{k=1}^{s}\beta_k < 1$ the GARCH process is also a square-integrable martingale-difference with respect to $(\mathcal{F}_i, i \in \{-\infty, ..., -1, 0, 1, ..., \infty\})$, where $\mathcal{F}_i$ is the σ-field generated by the information in the past, i.e. $\mathcal{F}_i = \sigma\{\epsilon_i, \epsilon_{i-1}, ...\}$ (see Shiryaev 1999). Note that, if ϵ_i follows a GARCH model, then ϵ_i^2 has an ARMA (autoregressive moving average) representation in the wide sense (see (A.11) in the appendix). Hence, for a GARCH process, the condition $E(\epsilon_i^4) < \infty$ on ϵ_i is sufficient for theorem 1, since this condition guarantees that ϵ_i^2 is a second order stationary process with summable autocovariances.

A general condition for the existence of $2m$th moments of a GARCH process was given by Ling (1999) (see also Ling and Li 1997). Let

$$\mathbf{A}_i = \left(\begin{array}{ccc|ccc} \alpha_1 z_i^2 & \cdots & \alpha_r z_i^2 & \beta_1 z_i^2 & \cdots & \beta_s z_i^2 \\ & \mathbf{I}_{(r-1)\times(r-1)} & \mathrm{O}_{(r-1)\times 1} & & \mathrm{O}_{(r-1)\times s} & \\ \hline \alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\ & \mathrm{O}_{(s-1)\times r} & & & \mathbf{I}_{(s-1)\times(s-1)} & \mathrm{O}_{(s-1)\times 1} \end{array} \right), \qquad (4.1)$$

where $\mathbf{I}$ denotes an identity matrix, $\mathbf{O}$ denotes a matrix of zeros and z_i are as defined before. Denote by $\mathbf{A}_i^{\otimes n}$ the Kronecker product of n matrices $\mathbf{A}_i$ and by $\rho(\mathbf{A})$ the spectral radius of a matrix $\mathbf{A}$. Suppose that $E(z_i^{2m})$ is finite for some positive integer m. If

$$\rho[E(\mathbf{A}_i^{\otimes m})] < 1, \qquad (4.2)$$

then the GARCH process ϵ_i is strictly stationary and ergodic, and its $2m$th moments are finite (see theorem 6.2 in Ling 1999). The examples in Ling (1999) show that condition (4.2) is equivalent to known conditions. For $m = 1$, condition (4.2) reduces to the above mentioned one, i.e. $\sum_{j=1}^r \alpha_j + \sum_{k=1}^s \beta_k < 1$. This shows that theorem 1 in Bollerslev (1986) holds without the assumption of conditional normality. For a GARCH(1,1) model and $m = 2$, let $E(z_i^4) = 3 + \kappa$, where κ denotes the kurtosis of z_i, condition (4.2) becomes $(3+\kappa)\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. When z_i are iid standard normal random variables, i.e. $\kappa = 0$, this condition reduces to $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. This is the one given by theorem 2 in Bollerslev (1986) with $m = 2$. We see that the condition for the existence of fourth moments of a GARCH model depends on the kurtosis κ of the iid random variables z_i.

Denote by π the stationary distribution of the GARCH process. It is assumed for simplicity that the process starts infinitely far in the past with this stationary distribution π. Based on theorem 6.2 in Ling (1999) and the above mentioned properties of a GARCH process we obtain the following lemma.

Lemma 1. *Let ϵ_i be the GARCH process generated by model (2.1)-(2.3). Suppose that $\alpha_0 > 0$, $\alpha_j \geq 0$, $\beta_k \geq 0$ and $\sum_{j=1}^r \alpha_j + \sum_{k=1}^s \beta_k < 1$. If $\rho[E(\mathbf{A}_i^{\otimes 2})] < 1$, then:*

1. *ϵ_i is strictly stationary having finite fourth moments and forming a square-integrable martingale-difference.*

2. *The process ϵ_i^2 is second order stationary with $\sum_{k=-\infty}^{\infty} \gamma_{\epsilon^2}(k) < \infty$.*

Lemma 1 ensures that ϵ_i generated by (2.1)-(2.3) fulfill the conditions on the innovations given in theorem 1. Let $X_i := Y_i - g(t_i)$ as defined in (2.1)-(2.3). Under the assumptions of lemma 1, Ling and Li (1997) further showed that X_i is also strictly stationary having finite fourth moments.

The covariance structure of a FARIMA-GARCH process is given by the following lemma, which also holds for a FARIMA process in the wide sense (see theorem 13.2.2 of Brockwell and Davis 1991).

Lemma 2. *Let Y_i be generated by (1.1)-(1.2). Suppose that $\delta \in (-0.5, 0.5)$, $\phi(B)$ and $\psi(B)$ have no common factors, all roots of $\phi(B)$ and $\psi(B)$ lie outside of the unit circle and $\sum_{j=1}^{r} \alpha_j + \sum_{k=1}^{s} \beta_k < 1$.*

1. *Denoting by $\gamma(k)$ the autocovariances, then we have*

$$\begin{cases} \gamma(k) \sim c_\gamma |k|^{-1+2\delta} & \text{with } c_\gamma > 0, \sum\limits_{k=-\infty}^{\infty} \gamma(k) = \infty \\ |\gamma(k)| \le c_0 \rho^{-k} & \text{with } c_0 > 0, 0 < \rho < 1, \sum\limits_{k=-\infty}^{\infty} \gamma(k) = V_0 > 0 \\ \gamma(k) \sim c_\gamma |k|^{-1+2\delta} & \text{with } c_\gamma < 0, \sum\limits_{k=-\infty}^{\infty} \gamma(k) = 0 \end{cases} \tag{4.3}$$

   for $\delta > 0$, $= 0$ and < 0 respectively, where "$\sim$" means that the ratio of the left and right hand sides converges to one.

2. *And denoting by $f(\lambda)$ the spectral density, we have*

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{|\psi(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} |1 - e^{-i\lambda}|^{-2\delta} \sim \frac{\sigma_\epsilon^2}{2\pi} [\psi(1)/\phi(1)]^2 \lambda^{-2\delta} \tag{4.4}$$

   as $\lambda \to 0$, where $\sigma_\epsilon^2 := \alpha_0 (1 - \sum_{j=1}^{s} \alpha_j - \sum_{k=1}^{r} \beta_k)^{-1}$ denotes the finite variance of ϵ_i.

The proof of lemma 2 will be omitted, since these results are just known facts on long memory processes (see e.g. Beran 1994 and Brockwell and Davis 1991).

Now, we consider the asymptotic properties of the sample mean of a FARIMA-GARCH process. To our knowledge, there are no detailed results on this topic in the literature. Based on theorem 1 and lemma 1 we have:

Theorem 3. *Let X_i be generated by model (2.1)-(2.3) with $\delta \in (-0.5, 0.5)$. Suppose that the assumptions of lemmas 1 and 2 hold. Then*

a) $\text{var}(\bar{X}) \sim n^{2\delta-1} V_\delta$ *as* $n \to \infty$,

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

where

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

As shown in Hosking (1996), similar result holds for $\delta = -\frac{1}{2}$ but with a different formula. Theorem 3 b) is the basis for the derivation of the asymptotic normality of parametric and nonparametric regression estimators with a FARIMA-GARCH error process. The importance of theorem 3 is shown by theorem 4 below, which gives a connection between the asymptotic normality of $\bar{X}$ and that of a weighted sum whose weights satisfy given conditions. In particular, these conditions are satisfied by the weights of a kernel or a local polynomial estimator $\hat{g}^{(\nu)}(t)$ under (2.1)-(2.3).

4.2 An extension of theorem 1

Theorem 1 is a central limit theorem on 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 $(X_{i,n})$, $1 \le i \le n$, $n = 1, 2, ...$ be a triangular array of random variables as defined in (3.1)-(3.2) and let $(w_{i,n})$ be a triangular array of weights such that $\sigma_n^2 := \text{var}(\sum_{i=1}^n w_i X_i) > 0$ for all n. If*

$$\max_{1 \le i \le n} |w_i|/\sigma_n \to 0 \text{ as } n \to \infty, \tag{4.5}$$

$$\sup_k |\sum_{i=1}^n w_i c_{k-i}|/\sigma_n \to 0 \text{ as } n \to \infty \tag{4.6}$$

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

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

Condition (4.5) means that the weights w_i are uniformly negligible. If $\max |w_i| = O(1)$, then it implies the condition given by theorem 1 on σ_n^2, i.e. $\sigma_n^2 \to \infty$ as $n \to \infty$. Condition (4.6) on the weighted sum $\sum w_i c_{k-i}$ is often not independent of (4.5). Theorem 1 is a special case of theorem 4 with $w_i \equiv 1$, in which case (4.6) can be derived from (4.5) (see Hosking 1994 and the proof of theorem 1 in the appendix). The central limit theorem given by Müller (1988) (theorem 4.2 therein) for the derivation of asymptotic normality of kernel or local polynomial estimators with iid errors is also a very special case of theorem 4 with ϵ being iid $(0, \sigma^2)$ random variables and $c_0 = 1$, $c_k = 0$ for $k \neq 1$. Based on theorem 4 the asymptotic normality of $\hat{g}^{(\nu)}$ is easy to prove.

4.3 Pointwise asymptotic results

What follows gives unified formulas for pointwise asymptotic bias and asymptotic variance for interior points $t \in [b, 1-b]$ as well as for boundary points $t \in [0, b) \cup (1-b, 1]$. The discussion will only be carried out for the region $[0, b]$. The results at point $t \in (b, 1-b]$ are the same as those at $t = b$. Formulas for $t \in (1-b, 1]$ are symmetric to those for $t = 1-t$. Note, however, that any fixed point $t \in (0, 1)$ will asymptotically not be a boundary point, since $b \to 0$ as $n \to \infty$. A standard definition of a left boundary point is $t = cb$ with $0 \leq c < 1$.

In the following it is assumed that $p - \nu$ is odd and $k = p + 1$. Here k denotes the order of the asymptotically equivalent kernel. It will be shown that $\hat{g}^{(\nu)}(t)$ converges to $g^{(\nu)}(t)$ at the same rate in the interior as well as at the boundary, if $p - \nu$ is odd. However, the convergence rate of $\hat{g}^{(\nu)}(t)$ at the boundary is slower than in the interior, if $p - \nu$ is even (see Beran and Feng 1999). Hence, a local polynomial approach with $p - \nu$ odd is more preferable. To derive the asymptotic results given below additional assumptions are required:

A1. g is an at least k times continuously differentiable function on $[0, 1]$.

A2. The weight function $K(u)$ is a symmetric density (a kernel of order two) with compact support $[-1, 1]$ having the polynomial form

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

(see e.g. Gasser and Müller 1979).

A3. The bandwidth satisfies: $b \to 0$, $(nb)^{1-2\delta} b^{2\nu} \to \infty$ as $n \to \infty$.

For $0 \le c \le 1$ we define the truncated kernel $K_c(u)$ as

$$K_c(u) = \left(\int_{-c}^{1} K(x)dx \right)^{-1} K(u) 1\!\!1_{[-c,1]}(u) \,. \tag{4.7}$$

For $c = 1$, i.e. in the interior, we have $K_1(u) = K(u)$. The truncated kernel used for the estimation at the left end point with $c = 0$ is $K_0(u) = 2K(u)1\!\!1_{[0,1]}$.

Let

$$\mathbf{N}_{pc} = \begin{bmatrix} 1 & \mu_{1,c} & \cdots & \mu_{p,c} \\ \mu_{1,c} & \mu_{2,c} & \cdots & \mu_{p+1,c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{p,c} & \mu_{p+1,c} & \cdots & \mu_{2p,c} \end{bmatrix}, \tag{4.8}$$

where $\mu_{j,c} = \int_{-c}^{1} u^j K_c(u)du$ is the jth moment of K_c. For $i, j = 1, ..., p+1$, let $(\alpha_{i,j,c}) = \mathbf{N}_{pc}^{-1}$ and define

$$K_{(\nu,k,c)}(u) = \nu! Q_{(\nu,k,c)}(u) K_c(u), \tag{4.9}$$

where

$$Q_{(\nu,k,c)}(u) = \sum_{j=1}^{p+1} \alpha_{\nu+1,j,c} u^{(j-1)}.$$

It is easily established that the function $K_{(\nu,k,c)}$ defined in (4.9) satisfies

$$\int_{-c}^{1} u^j K_{(\nu,k,c)}(u)du = \begin{cases} 0, & j = 0, ..., \nu-1, \nu+1, ..., k-1, \\ \nu!, & j = \nu, \\ \beta_{(\nu,k,c)}, & j = k, \end{cases} \tag{4.10}$$

where $\beta_{(\nu,k,c)}$ is a non-zero constant. Therefore, $K_{(\nu,p,c)}$ is a boundary kernel of order k for estimating the νth derivative, which will be called an "equivalent kernel". In the interior with $c = 1$ it is the same as defined by Gasser, Müller and Mammitzsch (1985) up to a $(-1)^\nu$ sign. It is clear that $K_c(u)$ and the equivalent kernel $K_{(\nu,k,c)}(u)$ are both polynomial Lipschitz-continuous kernels. Beran and Feng (1999) show that $\hat{g}^{(\nu)}$ is asymptotically equivalent to a kernel estimator using the kernel function defined in (4.9).

In the following we denote $x_i = (t_i - t)/b$, $y_j = (t_j - t)/b$. Let $n_0 = [nt + 0.5]$, $n_1 = [nb]$, $n_c = [ncb]$, where $[\cdot]$ denotes the integer part. The notation

$$V_n(c, b) = (nb)^{-1-2\delta} \sum_{i,j=n_0-n_c}^{n_0+n_1} K_{(\nu,k,c)}(x_i) K_{(\nu,k,c)}(y_i) \gamma(i-j)\,, \tag{4.11}$$

will be used for convenience. We obtain

Theorem 5. *Let Y_i be generated by model (2.1)-(2.3). Suppose that the assumptions of theorem 3 hold. Under the assumptions A1 to A3, and let $t = cb$ with $0 \le c \le 1$. Then for $\delta \in (-0.5, 0.5)$, we have*

i) Bias:

$$E[\hat{g}^{(\nu)} - g^{(\nu)}] = b^{(k-\nu)} \frac{g^{(k)}(t)\beta_{(\nu,k,c)}}{k!} + o(b^{(k-\nu)})\,; \tag{4.12}$$

ii):

$$\lim_{n\to\infty} V_n(c, b) = V(c), \tag{4.13}$$

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

iii) Variance:

$$var\,(\hat{g}^{(\nu)}(t)) = (nb)^{-1+2\delta} b^{-2\nu} [V(c) + o(1)\,], \tag{4.14}$$

iv) Asymptotic normality: In the case that the bias has the representation (4.12), assuming that $nb^{(2k+1-2\delta)/(1-2\delta)} \to d^2$ as $n \to \infty$, for some $d > 0$, then

$$(nb)^{1/2-\delta} b^{\nu} (\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)) \xrightarrow{\mathcal{D}} N(d\Delta, V(c)), \tag{4.15}$$

where $\Delta = \frac{g^{(k)}(t)\beta_{(\nu,k,c)}}{k!}$ and $V(c)$ is the constant defined in (4.13).

Remark 3. All of the results of parts i) to iii) and those of theorem 6 given below also hold in the case when X_i in (2.1)-(2.3) is a FARIMA process in the wide sense. In order that iv) holds, the innovations ϵ_i have just to satisfy the conditions given in theorem 1. Hence, iv) holds of cause in the case when ϵ_i are iid random variables (cf. the results in theorems 1 and 2 in Hosking 1996).

Remark 4. The formula for the asymptotic bias is the same as that for nonparametric regression with independent errors. The formula for the asymptotic variance with $\delta \neq 0$ is different from that for nonparametric regression with independent errors or short memory errors, i.e. the case of $\delta = 0$. In this case the asymptotic variance converges to zero at the rate $(nb)^{-1}b^{2\nu}$. When $\delta > 0$, var $(\hat{g}^{(\nu)})$ converges to zero at a slower rate, while the rate of convergence of var $(\hat{g}^{(\nu)})$ for $\delta < 0$ is higher than that for $\delta = 0$.

Remark 5. Theorem 6 below shows that a bandwidth $b = O(n^{(2\delta-1)/(2k+1-2\delta)})$ is of the optimal order. In this case the asymptotic bias and the asymptotic variance are of the same order. If the bandwidth b is of higher order, i.e. 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 MISE (mean integrated squared error) defined by

$$\text{MISE}(\hat{g}^{(\nu)}(x)) = \int_0^1 E\{[\hat{g}^{(\nu)}(x) - g^{(\nu)}(x)]^2\}dx. \tag{4.16}$$

For $p - \nu$ even $\text{MISE}(\hat{g}^{(\nu)}(x))$ 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

$$I(g^{(k)}) = \int_0^1 [g^{(k)}(t)]^2 dt\,, \tag{4.17}$$

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

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

i) *The mean integrated squared error (MISE) of* $\hat{g}^{(\nu)}$ *is given by*

$$\begin{aligned}
&\int_0^1 E\{[\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)]^2\}dt \\
&= \text{MISE}_{\text{asympt}}(n,b) + o(\max(b^{2(k-\nu)}, [(nb)^{2\delta-1}b^{-2\nu}])) \\
&= b^{2(k-\nu)}\frac{I(g^{(k)})\beta^2_{(\nu,p)}}{k!} + (nb)^{2\delta-1}b^{-2\nu}V \\
&+ o(\max(b^{2(k-\nu)}, [(nb)^{2\delta-1}b^{-2\nu}]))\,;
\end{aligned} \tag{4.18}$$

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

$$b_{\text{opt}} = C_{\text{opt}} n^{(2\delta-1)/(2k+1-2\delta)}, \tag{4.19}$$

where

$$C_{\text{opt}} = \left[\frac{2\nu+1-2\delta}{2(k-\nu)} \frac{[k!]^2 V}{I(g^{(k)})\beta^2_{(\nu,p)}} \right]^{1/(2k+1-2\delta)}, \tag{4.20}$$

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

The proof of theorem 6 and the following formulas will be omitted, since they are they same as the case that only uncorrelated $(0, \sigma^2)$ innovations are assumed (see Beran and Feng 1999).

Note that by inserting b_{opt} in (4.18), theorem 2 implies that for $p - \nu$ odd the optimal MISE is of the order

$$\int_0^1 E\{[\hat{g}^{(\nu)}(t) - g^{(\nu)}(t)]^2\}dt = O(n^{2(2\delta-1)(k-\nu)/(2k+1-2\delta)}). \tag{4.21}$$

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 $\delta \geq 0$, Hall and Hart (1990) show that this is the optimal convergence rate. The following remarks clarify the results given above.

Remark 6. For bandwidth selection with the plug-in method one has to calculate the value of V. Let $c_f = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \text{cov}\,(U_i, U_{i+k}) = (2\pi)^{-1} V_0$. Under the assumptions of theorem 1 we have:

$$V = 2\pi c_f \int_{-1}^{1} K^2_{(\nu,k)}(x)dx \tag{4.22}$$

for $\delta = 0$, and

$$V = 2c_f,\ (1-2\delta)\sin(\pi\delta) \int_{-1}^{1}\int_{-1}^{1} K_{(\nu,k)}(x) K_{(\nu,k)}(y)|x-y|^{2\delta-1} dxdy \tag{4.23}$$

for $\delta > 0$ (see Hall and Hart 1990 and Beran 1999). The explicit form of V for $\delta < 0$ is more complex, since the integral $\int_{-1}^{1} K_{(\nu,k)}(y)|x-y|^{2\delta-1}dy$ does not exist. However, at any point x the kernel $K_{(\nu,k)}(y)$ may be written as $K_{(\nu,k)}(y) = \sum_{l=0}^{r} \beta_l(x)(x-y)^l =: K_0(x) + K_1(x-y)$, where r is an integer, $K_0(x) = \beta_0(x)$ and $K_1(x-y) = \sum_{l=1}^{r} \beta_l(x)(x-y)^l$. Note that, in the case of antipersistence it holds $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$. We have, for $\delta < 0$ (see Beran and Feng 1999),

$$\begin{aligned} V \ = \ & 2c_f,\ (1-2\delta)\sin(\pi\delta) \int_{-1}^{1} K_{(\nu,k)}(x) \times \\ & \left\{ \int_{-1}^{1} K_1(x-y)|x-y|^{2\delta-1}dy - K_0(x) \int_{|y|>1} |x-y|^{2\delta-1}dy \right\} dx. \end{aligned} \tag{4.24}$$

If g is estimated by a first order local polynomial with the uniform kernel as the weight function, then we have, in the interior, $K_{(0,2)}(x) = K(x) = \mathbb{1}_{\{|x|\leq 1\}}/2$. In this case we have $K_0(x) = \mathbb{1}_{\{|x|\leq 1\}}/2$ and $K_1 \equiv 0$. The formulas (4.22), (4.23) and (4.24) give the same result

$$V(\delta) = \frac{2^{2\delta} c_f,\ (1-2\delta)\sin(\pi\delta)}{\delta(2\delta+1)} \tag{4.25}$$

with $V(0) = \lim_{\delta\to 0} V(\delta) = \pi c_f$ (see corollary 1 in Beran 1999).

5 Final remarks

In this paper we introduced a class of semiparametric FARIMA-GARCH models with short- or long memory allowing the conditional variance of the innovations to change. Asymptotic results on the nonparametric estimation of $g^{(\nu)}$ are investigated in detail. The ARCH and GARCH models proposed by Engle (1982) and Bollerslev (1986) have become a widely used model for analyzing financial time series. Ling and Li (1997) showed the potential usefulness of the FARIMA-GARCH model. As a semiparametric extension of the FARIMA-GARCH model, the model proposed in this paper is expected to become a useful tool for modeling stochastic processes with trends, long memory as well as conditional heteroskedasticity. Particularly, it provides a more general class of models for analyzing volatility in financial time series. In this paper we did not given any application example. However, examples for modeling financial time series with the related SEMIFAR model proposed by Beran (1999) (see e.g. Beran and Ocker, 1999, Beran et al. 1999 and Ocker 1999) illustrate the potential usefulness of semiparametric long memory time series models.

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. Although the given results on the asymptotic behaviors of the proposed estimators do not depend on the exact distribution of z_i in (1.1), conditionally normal distribution is required, if one wants to estimate the unknown parameters determining the whole model by maximum likelihood.

6 Acknowledgements

This research was supported in part by the Center of Finance and Econometrics at the University of Konstanz and by an NSF (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. Finally, our thanks go to Dr. Shiqing Ling (University of Western Australia) for sending us his thesis and an important preprint.

Appendix: Proofs of theorems

To prove theorem 1 we need the following lemmas A.1 and A.2. Suppose that on the probability space $(\Omega, \mathcal{F}, P)$ there are given stochastic sequences

$$\xi^n = (\xi_{nk}, \mathcal{F}_{nk}), 0 \le k \le n, n \ge 1,$$

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

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

The double sequence $\{S_{nk}, \mathcal{F}_{nk}, 1 \le k \le n, n \ge 1\}$ will be called a martingale array.

For the proof of case 1 in theorem 1 we will use lemma A.1, which is a special case of theorem 4 of Shiryaev (1996, Chapter 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, then we have:

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

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

and the condition

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

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

See Shiryaev (1996) for the proof.

The proof of case 2 of theorem 1 is based on lemma A.2, a special case of theorem 3.2 of Hall and Heyde (1980). Theorem 24.3 of Davidson (1994) is very similar to lemma A.2, where slightly different conditions are used.

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

$$\max_k |\xi_{nk}| \xrightarrow{P} 0, \tag{A.3}$$

$$\sum_{k=0}^{n} \xi_{nk}^2 \xrightarrow{P} 1 \tag{A.4}$$

and

$$E(\max_k \xi_{nk}^2) \qquad \text{is bounded in } n. \tag{A.5}$$

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

For the proof see Hall and Heyde (1980).

Remark A.1. Lemma A.2 is a special case of theorem 3.2 in Hall and Heyde (1980) by setting k_n there equal to n and by replacing the a.s. finite random variable η^2 with the constant 1. The "nested σ-field" condition (3.21) in Hall and Heyde (1980) is now not necessary due to the latter specification. This lemma is used here to avoid checking the conditional Lindeberg condition (A.1).

Proof of theorem 1. Let $\sigma_n^2 = E(X_1 + ... + X_n)^2$ as defined in theorem 1. We first show that, following Ibragimov and Linnik (1971) and Hosking (1994), $(X_1 + ... + X_n)/\sigma_n$ can be rewritten as $(X_1 + ... + X_n)/\sigma_n = S_{nn} + \eta_n$, where $\eta_n \xrightarrow{P} 0$ and

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

where ξ_{nk} form a square-integrable martingale-difference $\xi^n = (\xi_{nk}, \mathcal{F}_{nk})$ and $\{S_{nk}, \mathcal{F}_{nk}, 1 \leq k \leq n, n \geq 1\}$ is a zero mean square-integrable martingale array with respect to $\mathcal{F}_{nk}$ as defined below. Then we show that they satisfy the conditions of lemmas A.1 and A.2, respectively, for cases 1 and 2 of theorem 1.

Suppose that $E(\epsilon_0^2) = 1$ for simplicity. Following the proof of their theorem 18.6.5 in Ibragmov and Linnik (1971), we have

$$\begin{aligned}\sigma_n^2 &= E(X_1 + \cdots + X_n)^2 = \cdots \\ &= \sum_{k=-\infty}^{\infty} (c_{k-1} + \cdots + c_{k-n})^2 \\ &= \sum_{k=-\infty}^{\infty} c_{k,n}^2 , \qquad \text{say.}\end{aligned}$$

Hosking (1994) gave some corrections of the proof of Ibragmov and Linnik (1971) and showed that

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

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

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

with

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

For each $n \ge 1$ let $n_1 = -[\frac{n-1}{2}]$, $n_2 = [\frac{n}{2}]$ such that $n_1 \le 0$, $n_2 \ge 0$ and $n_1 + n_2 + 1 = n$, where $[\cdot]$ denotes the integer part. Define the square integrable martingale-difference $\xi^n = (\xi_{nk}, \mathcal{F}_{nk})$ with

$$\xi_{nk} = a_{n_1+k-1,n}\epsilon_{n_1+k-1}, \qquad \text{and } \mathcal{F}_{nk} = \mathcal{F}_{n_1+k-1}, \quad k = 1, ..., n.$$

Denote $a_{n_1+k-1,n}$ by b_{nk} for convenience. Then we have

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

with

$$S_{nk} = \sum_{i=n_1}^{n_1+k-1} a_{i,n}\epsilon_i = \sum_{i=1}^{k} b_{ni}\epsilon_{n_1+i-1} , \qquad k = 1, ..., n,$$

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 $E(\eta_n) = 0$ and

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

Here, $\xi^n = (\xi_{nk}, \mathcal{F}_{nk})$ is a square-integrable martingale-difference and $\{S_{nk}, \mathcal{F}_{nk}, 1 \leq k \leq n, n \geq 1\}$ is a zero mean square-integrable martingale array. It remains to show that, in case 1 of theorem 1, $\xi^n = (\xi_{nk}, \mathcal{F}_{nk})$ fulfill the conditions of lemma A.1 and, in case 2 of theorem 1, $\{S_{nk}, \mathcal{F}_{nk}, 1 \leq k \leq n, n \geq 1\}$ fulfill those of lemma A.2, respectively.

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

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

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

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

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

Case 2. Now, we have to check that the zero mean square-integrable martingale array $\{S_{nk}, \mathcal{F}_{nk}, 1 \leq k \leq n, n \geq 1\}$ fulfills the conditions (A.3)-(A.5). It is clear that $E(\max_k \xi_{nk}^2)$ is bounded in n, i.e. (A.5) is satisfied. For (A.3) we have, by using (A.6),

$$\begin{aligned}
\max_k |\xi_{nk}| &\leq a_n \max_k |\epsilon_{n_1+k-1}| \\
&= o(1) \max_k |\epsilon_{n_1+k-1}| \xrightarrow{P} 0.
\end{aligned}$$

To see $\{S_{nk}, \mathcal{F}_{nk}, 1 \le k \le n, n \ge 1\}$ fulfills (A.4) noting that

$$\sum_{k=0}^{n} \xi_{nk}^2 = \sum_{k=1}^{n} \xi_{nk}^2 = \sum_{k=1}^{n} b_{nk}^2 \epsilon_{n_1+k-1}^2,$$

since $\xi_{n0} = 0$ by definition. Under the condition of case 2 ϵ_i^2 is a second order stationary process with $E(\epsilon_0^2) = 1$ and $\gamma_k(\epsilon_i, \epsilon_{i+k}) \to 0$ as $n \to \infty$. And, observing that the weights b_{nk}^2 satisfy the condition of corollary 1 of theorem 2, we haven

$$\sum_{k=0}^{n} \xi_{nk}^2 = \sum_{k=1}^{n} b_{nk}^2 \epsilon_{n_1+k-1}^2 \xrightarrow{P} E(\epsilon_0^2) = 1.$$

This completes the proof. □

Proof of theorem 2.

Again, we put $\operatorname{var}(\epsilon_i) = 1$ for convenience. In this case we have that $|\gamma(k)| \le 1$. For simplicity, assume that $\max_{1\le i\le n} |w_i| = O(n^{-\alpha})$, with $\alpha > 0$, since $\max_{1\le i \le n} |w_i| = o(1)$. The following proof can be, of cause, easily extended to general cases. Let $N = n^{\alpha/2}$ such that $N \to \infty$, $N \cdot \max\limits_{1\le i\le n} |w_i| \to 0$ as $n \to \infty$. We have

$$\begin{aligned}
\operatorname{var}(\sum_{i=1}^{n} w_i \epsilon_i) &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(i-j) \\
&= \sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \gamma(i-j) \\
&= \sum_{i=1}^{n} w_i \left[\sum_{|i-j|\le N} w_j \gamma(i-j) + \sum_{|i-j|>N} w_j \gamma(i-j) \right] \\
&\le \sum_{i=1}^{n} |w_i| \left[\sum_{|i-j|\le N} |w_j|\, |\gamma(i-j)| + \sum_{|i-j|>N} |w_j|\, |\gamma(i-j)| \right].
\end{aligned}$$

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

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

holds uniformly in i. Consider the first part on the left side of (A.8),

$$\begin{aligned}
\sum_{|i-j|\le N} |w_j|\, |\gamma(i-j)| &\le (2N+1) \max_{1\le i\le n} |w_i| = O(Nn^{-\alpha}) \\
&= O(n^{-\alpha/2}) = o(1). \qquad \text{(A.9)}
\end{aligned}$$

For the second part we have

$$\sum_{|i-j|>N} |w_j|\,|\gamma(i-j)| \leq \max_{|k|>N}(|\gamma(k)|) \sum_{|k|>N} |w_k| = O(\max_{|k|>N} |\gamma(k)|) = o(1), \tag{A.10}$$

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

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

□

Proof of lemma 1. Under the condition of this lemma, Ling (1998 1999) showed that ϵ_i is strictly stationary with finite fourth moments. In this case, the statement that ϵ_i is a square-integrable martingale-difference follows immediately from the definitions. To show that $\sum_{k=-\infty}^{\infty} \gamma_{\epsilon^2}(k) < \infty$ define $v_i = \epsilon_i^2 - h_i$. Under the condition $\sum_{j=1}^{r} \alpha_j + \sum_{k=1}^{s} \beta_k < 1$, we have $E|v_i| < \infty$ and $E(v_i|\mathcal{F}_{i-1}) = 0$, i.e. v_i is a martingale-difference (see Shiryaev 1999, p. 106ff). Noting that, under the condition $\rho[E(\mathbf{A}_i^{\otimes 2})] < 1$, the fourth moments of ϵ_i, and hence the second moments of ϵ_i^2, exist. v_i is a martingale-difference with finite variance $\mathrm{var}\,(v_i) < \infty$ and hence v_i is again a square-integrable martingale-difference as well as an uncorrelated white noise. Furthermore, ϵ_i^2 has the ARMA(max(r, s), s) representation in the uncorrelated white noise v_i

$$\epsilon_i^2 = \alpha_0 + \sum_{j=1}^{r} \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^{s} \beta_k \epsilon_{i-k}^2 - \sum_{k=1}^{s} \beta_k v_{i-k} + v_i \tag{A.11}$$

(see e.g. Bollerslev 1986 and Shiryaev 1999). By applying the well known fact that the autocovariances of a second order stationary ARMA(p, q) process is absolutely summable, the proof of lemma 1 is completed. □

Proof of theorem 3. The formula of the asymptotic variance of $\bar{X}$ stays unchanged from case to case, if only ϵ_i are uncorrelated $(0, \sigma^2)$ random variables. Hence, it is the same as that for iid innovations given by theorems 1 and 8 of Hosking (1996), i.e. $\mathrm{var}\,(\bar{X}) = n^{2\delta-1} V_\delta$ 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 , $(1+\delta) = \delta$, (δ) and , (δ), $(1-\delta) = \frac{\pi}{\sin(\pi\delta)}$ (for δ is not an integer), we obtain the alternative representation of V_δ

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

which is used in this paper.

Since X_i defined in (2.1)-(2.3) is a zero mean FARIMA process with innovations ϵ_i following a GARCH model. We have

$$X_i = \sum_{k=0}^{\infty} c_k \epsilon_{i-k} \tag{A.12}$$

with $c_k \sim \frac{|\psi(1)|}{|\phi(1)|} k^{\delta-1}$ as $n \to \infty$. Hence, for $-0.5 < \delta < 0.5$, $\sum_{k=0}^{\infty} c_k^2 < \infty$. This together with lemma 1 shows that X_i fulfills the conditions of theorem 1, and so $(X_1 + \cdots + X_n)/\sigma_n \xrightarrow{D} N(0,1)$. Observing that $[n^{1/2-\delta}\bar{X} - (X_1 + \cdots + X_n)/\sigma_n] \xrightarrow{P} 0$, we have $n^{1/2-\delta}\bar{X} \xrightarrow{D} N(0,1)$. □

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

$$\begin{aligned}
\sum_{i=1}^{n} w_i X_i &= \sum_{i=1}^{n} w_i \Big(\sum_{k=-\infty}^{\infty} c_{i-k}\epsilon_k\Big) \\
&= \sum_{k=-\infty}^{\infty} \Big(\sum_{i=1}^{n} w_i c_{k-i}\Big)\epsilon_k \\
&=: \sum_{k=-\infty}^{\infty} c_{k,n}\epsilon_k.
\end{aligned} \tag{A.13}$$

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

$$\begin{aligned}
\sigma_n^2 &:= E\Big(\sum_{i=1}^{n} w_i X_i\Big)^2 \\
&= \sum_{k=-\infty}^{\infty} c_{k,n}^2.
\end{aligned} \tag{A.14}$$

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 $a_{k,n}$, i.e. it tends to zero uniformly in k as $n \to \infty$, is guaranteed by condition (4.6). 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 $\max_i |w_i| = O[(nb^{1+\nu})^{-1}]$ and $w_i \equiv 0$ outside an interval with length of order b. 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 $c_k \sim k^{\delta-1}$,

$$\begin{aligned} |\sum_{i=1}^{n} w_i c_{k-i}|/\sigma_n &\leq \max_i |w_i| [\sum_{i=1}^{n} |c_{k-i}| \Big| w_i \neq 0]/\sigma_n \\ &= O[(nb)^{-1/2-\delta}] O[(nb)^{\delta}] \\ &= O[(nb)^{-1/2}] \to 0 \text{ as } n \to \infty. \end{aligned}$$

Conditions (4.5) and (4.6) are satisfied by the weights of $\hat{g}^{(\nu)}(t)$. We have

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

In case that the bias has the representation (4.12), and assuming that $nb^{\frac{2k+1-2\delta}{1-2\delta}} \to d^2$, for some $d > 0$, we obtain,

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

where $\Delta = \frac{g^{(k)}(t)\beta_{(\nu,k,c)}}{k!}$ and $V(c)$ is the constant defined in (4.13). □

References

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

Beran, J. (1995). Maximum likelihood of 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 semiparametric fractional framework for modelling trends, long-range dependence and nonstationarity. Preprint, University of Konstanz.

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

Beran, J., Feng, Y., Franke, G., Hess, D. and Ocker, D. (1999). SEMIFAR models, with applications to commodities, exchange rates and the volatility of stock market indices. Preprint, University of Konstanz.

Beran, J., and Ocker, D. (1999). Volatility of stock market indices - an analysis based on SEMIFAR models. Preprint, University of Konstanz.

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

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* Springer, New York.

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

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

Davidson, J. (1994). *Stochastic Limit Theory.* Advanced Texts in Econometrics. Oxford University Press, New York.

Engel, R.F. (1982). Autoregressive conditional heteroskedasticity with estimation of U.K. inflation. *Econometrica*, **50** 987–1008.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B* **57** 371–394.

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

Feng, Y. (1998). *Kernel and Locally Weighted Regression – with Application to Time Series Decomposition.* Ph.D. thesis. University of Konstanz.

Gasser, T. and Müller, H.G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.) 23–68. Springer, Heidelberg.

Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

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.* Academic Press, New York.

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.

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 Variables Random.* Wolters-Noordhoff Publishing Groningen, The Netherlands.

Ling, S. (1998). *Stationary and Nonstationary Time Series Models with Conditional Heteroskedasticity.* Ph.D. thesis, The University of Hong Kong.

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. To appear in *J. Appl. Probab.*.

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

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 Analysis of Longitudinal Data.* Springer, Berlin.

Ocker, D. (1999). *Stationary and Nonstationary Fractional ARIMA Models - Model Choice, Forecasting, Aggregation and Intervention.* Ph.D. thesis. University of Konstanz.

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

Shiryaev, A.N (1996). *Probability.* Springer, New York.

Shiryaev, A.N (1999). *Essentials of Stochastic Finance – Facts, Models, Theory.* Advanced Series on Statistical Science & Applied Probability, **Vol. 3**. World Scientific, Singapore.

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 Roseblatt processes. *Z. Wahrsch. verw. Geb.*, **31** 287–302.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman & Hall, London.