Probab. Theory Relat. Fields (2009) 144:53–77
DOI 10.1007/s00440-008-0141-2

Estimating the innovation distribution in nonparametric autoregression

Ursula U. Müller · Anton Schick · Wolfgang Wefelmeyer

Received: 4 July 2007 / Revised: 16 January 2008 / Published online: 9 February 2008

Abstract We prove a Bahadur representation for a residual-based estimator of the innovation distribution function in a nonparametric autoregressive model. The residuals are based on a local linear smoother for the autoregression function. Our result implies a functional central limit theorem for the residual-based estimator.

Keywords Residual-based empirical distribution function · Local linear smoother · Bahadur representation

Mathematics Subject Classification (2000) 62M05 · 62M10 · 62G30

1 Introduction

Regression models are described by their regression function and their error distribution, and possibly by their covariate distribution. The object of primary statistical

The research of A. Schick was supported in part by NSF Grant DMS0405791.

U. U. Müller
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
e-mail: uschi@stat.tamu.edu
URL: http://www.stat.tamu.edu/∼uschi/

A. Schick (✉)
Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA
e-mail: anton@math.binghamton.edu
URL: math.binghamton.edu/anton/

W. Wefelmeyer
Mathematical Institute, University of Cologne, Weyertal 86-90, 50931 Cologne, Germany
e-mail: wefelm@math.uni-koeln.de
URL: www.mi.uni-koeln.de/∼wefelm/

interest is the regression function. Estimators of the error distribution function are however also of interest, in particular for tests about the regression function and for prediction intervals about future observations. There is a large literature on estimating error distribution functions, but it is nearly exclusively concerned with cases in which the regression function is parametric, in particular with linear regression. We refer to [12,21,22,24,29,36], and for increasing dimension to [30,35]. Analogous results exist for autoregressive time series with parametric autoregression function, and for related time series models. For AR(p) models see [8,23,26]. For ARMA, ARCH and GARCH models we refer to [3,9,19,25,28]. See also Chapts. 7 and 8 in [24]. Empirical distribution functions of *powers* of residuals are studied by [4,18,27].

In these papers, the (auto-)regression function (and volatility) depends on a finite-dimensional parameter, which can be estimated at the root-n rate. If this function is nonparametric, different arguments are needed to obtain a stochastic expansion and hence the root-n rate and asymptotic normality for the residual-based empirical distribution function. For heteroscedastic nonparametric regression [2] give a functional central limit theorem for a residual-based empirical distribution function; see also [20]. A related result is in [10] who uses separate parts of the sample for estimating the regression function and the error distribution function. Müller et al. [32] consider the partly linear regression model $Y = \vartheta^\top U + \varrho(X) + \varepsilon$ with error ε independent of the covariate pair (U, X). They use a local linear smoother for the regression function ϱ and get by with weaker assumptions on the error distribution and the covariate distribution. In these results, the distribution of the covariate X is assumed to have bounded support.

We expect the results for nonparametric regression to have counterparts in nonparametric autoregression. Indeed [16] show that nonparametric autoregression is (locally) asymptotically equivalent, in the sense of Le Cam's deficiency distance, to certain nonparametric regression models. Below we study a stationary and ergodic nonparametric autoregressive model

$$X_t = r(X_{t-1}) + \varepsilon_t, \quad t \in \mathbb{Z},$$

with independent and identically distributed innovations ε_t, $t \in \mathbb{Z}$. We obtain a stochastic expansion ("Bahadur representation") and a functional central limit theorem for a residual-based empirical distribution function, using a local linear smoother for the function r. We assume that the innovations ε_t have mean zero, finite variance σ^2 and a distribution function F with positive density f. Compared to regression, two technical difficulties arise. One is that the observations are dependent. Another is that for regression we could assume that X is bounded, but the analogous assumption for the process X_t is ruled out by our requirement that f is positive.

We want to estimate F based on observations $X_0, X_1, \ldots, X_n$ of the autoregressive process. For this we need an estimator $\hat{r}$ of r. Then we can form the residuals $\hat{\varepsilon}_j = X_j - \hat{r}(X_{j-1})$, $j = 1, \ldots, n$. Typically, the performance of the estimator $\hat{r}(x)$ will be poor for large values of x. For this reason we shall use only the residuals $\hat{\varepsilon}_j$ for which X_{j-1} falls into an interval $I_n = [a_n, b_n]$ where $-a_n$ and b_n tend to infinity slowly.

We achieve this by using random weights

$$\bar{w}_j = \frac{w_{nj}}{\sum_{i=1}^n w_{ni}}, \quad j = 1, \ldots, n$$

with $w_{nj} = w_n(X_{j-1})$ based on a Lipschitz-continuous weight function w_n that vanishes off I_n, is 1 on $[a_n + \gamma, b_n - \gamma]$ for some fixed small positive γ and is linear on the intervals $[a_n, a_n + \gamma]$ and $[b_n - \gamma, b_n]$. Our estimator will be of the form

$$\hat{\mathbb{F}}(t) = \sum_{j=1}^n \bar{w}_j \mathbf{1}[\hat{\varepsilon}_j \le t], \quad t \in \mathbb{R}.$$

We shall compare this estimator with the empirical distribution function based on the true innovations,

$$\mathbb{F}(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}[\varepsilon_j \le t], \quad t \in \mathbb{R}.$$

We take $\hat{r}$ to be a local linear smoother. Recall that, for a fixed $x \in \mathbb{R}$, the *local linear smoother* $\hat{r}$ satisfies $\hat{r}(x) = \hat{\beta}_0$, where $(\hat{\beta}_0, \hat{\beta}_1)$ denotes a minimizer of

$$\sum_{j=1}^n \left(X_j - \beta_0 - \beta_1 \frac{X_{j-1} - x}{c_n} \right)^2 K\left(\frac{X_{j-1} - x}{c_n} \right).$$

Here c_n is a bandwidth and K is a kernel.

We impose the following conditions on the density f and the regression function r.

(F) The density f is positive, has mean zero and a finite moment of order greater than 8/3, and is Hölder with exponent ξ greater than $1/3$.

(R) The function r has a bounded second derivative and satisfies the growth condition $|r(x)| \le c|x| + d$ for some $c < 1$ and $d < \infty$.

Assumption (F) without positivity of f was already used in [32]. Positivity of f plays a role in guaranteeing ergodicity of the process. Indeed, together with the growth condition on r it guarantees geometric ergodicity of the autoregressive model. The growth condition could be replaced by any other condition on r that implies geometric ergodicity. Sufficient conditions for geometric ergodicity of nonlinear autoregressive models are in [1,5,6].

The above assumptions also guarantee the existence of a stationary density g that satisfies

$$g(y) = \int f(y - r(x)) g(x)\, dx, \quad y \in \mathbb{R}. \tag{1.1}$$

Thus positivity and the Hölder property of f carry over to g and guarantee that the latter is bounded and bounded away from zero on each compact subset of $\mathbb{R}$. This conforms with the customary assumption in nonparametric regression, namely that the covariate density is bounded and bounded away from zero on its compact support; see [32].

We impose the following conditions on the kernel K and the intervals I_n.

(K) The kernel K is a three times continuously differentiable density with mean zero and support $[-1, 1]$.

(I) The interval $I_n = [a_n, b_n]$ is such that $-a_n$ and b_n tend to infinity slowly enough so that $\log n \inf_{x \in I_n} g(x)$ stays bounded away from zero.

Assumption (I) is used to obtain uniform rates of convergence for $\hat{r}$ on the intervals I_n. This is analogous to [17] who proves uniform convergence rates for kernel estimators based on dependent data. Finally, in view of the inequality

$$\inf_{x \in I_n} g(x)(b_n - a_n) \le \int_{a_n}^{b_n} g(x)\,dx \le 1,$$

it follows from (I) that $b_n - a_n = O(\log n)$.

Theorem 1 *Suppose* (*F*), (*R*), (*K*) *and* (*I*) *hold and* $c_n \sim (n \log n)^{-1/4}$. *Then*

$$\sup_{t \in \mathbb{R}} \left| \hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t)\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \right| = o_p(n^{-1/2}).$$

In view of the differentiability assumptions on r, an optimal choice of bandwidth for $\hat{r}$ would be proportional to $n^{-1/5}$. Thus the present choice of bandwidth results in an undersmoothed estimator of r. Undersmoothing is needed in our proofs to guarantee that the bias is asymptotically negligible which amounts to the requirement $nc_n^4 \to 0$ on the bandwidth. The choice of bandwidth in the theorem is made to accomplish this and to make the bandwidth basically as large as possible. Actually, the choice $c_n \sim n^{-1/4} \log^{-\gamma} n$ works for any positive γ. We have taken $\gamma = 1/4$ for notational simplicity.

We set $X = X_0$ and $\varepsilon = \varepsilon_1$. By Theorem 1,

$$\sup_{t \in \mathbb{R}} \left| \hat{\mathbb{F}}(t) - F(t) - \frac{1}{n}\sum_{j=1}^{n} \big(\mathbf{1}[\varepsilon_j \le t] - F(t) + f(t)\varepsilon_j\big) \right| = o_p(n^{-1/2}).$$

The terms $\mathbf{1}[\varepsilon_j \le t] - F(t) + f(t)\varepsilon_j$ in this Bahadur representation of $\hat{\mathbb{F}}(t) - F(t)$ are martingale increments, and the density f is bounded under assumption (F). Hence by Corollary 7.7.1 of [24], the residual-based empirical process $n^{1/2}(\hat{\mathbb{F}} - F)$ converges weakly in $D[-\infty, \infty]$ to a centered Gaussian process with covariance function

$$(s, t) \mapsto F(s \wedge t) - F(s)F(t) + f(s)c(t) + f(t)c(s) + f(s)f(t)\sigma^2,$$

where

$$c(t) = \int_{-\infty}^{t} x f(x)\,dx$$

is the mean of $\varepsilon \mathbf{1}[\varepsilon \le t]$.

Paradoxically, the asymptotic variance

$$F(t)(1 - F(t)) + 2f(t)c(t) + f^2(t)\sigma^2$$

of the residual-based weighted empirical distribution function $\hat{\mathbb{F}}(t)$ can be smaller than the asymptotic variance $F(t)(1 - F(t))$ of the empirical distribution function $\mathbb{F}(t)$ based on the unobserved innovations. The explanation is that $\mathbb{F}(t)$ does not make use of the assumption that the innovations have mean zero, while the linear smoother $\hat{r}$ used for the residuals exploits this information (as do other nonparametric estimators for the autoregression function). For nonparametric regression, a similar observation is made in [31].

The estimator $\hat{\mathbb{F}}(t)$ is efficient. Efficiency can be proved similarly as for nonparametric regression in [31].

A result along the lines of Theorem 1 can be proved for higher lag nonparametric regression. This requires additional smoothness of the underlying regression function r of several variables and the use of appropriate multivariate local polynomial smoothers. We will pursue this somewhere else.

Note that the conclusions of Theorem 1 remain valid if we replace the endpoints of I_n by data-driven versions which take only finitely many values with high probability. This can be achieved by choosing $I_n = [a_n, b_n]$ at random from a collection $\mathscr{I}_n = \{[a, b] : a < b, a, b \in G_n\}$ of intervals with $G_n = \{k\eta : k = 0, 1, -1, 2, -2, \dots, |\eta k| \le C \log n\}$ for some small positive η and some constant C. For this let

$$\hat{g}(x) = \frac{1}{nc_n} \sum_{j=1}^{n} K\left(\frac{X_j - x}{c_n}\right), \quad x \in \mathbb{R},$$

be a kernel density estimator of g. Under the assumptions of Theorem 1 we have

$$\sup_{|x| \le C \log n} |\hat{g}(x) - g(x)| = o_p(n^{-1/12});$$

see (3.1) and (3.2) below with $i = 0$. Now we can choose I_n as the interval with largest length among the intervals I in $\mathscr{I}_n$ with $\log n \inf_{x \in I} \hat{g}(x) > \eta$.

The remainder of the paper is organized as follows. Section 2 describes some possible applications of Theorem 1. A proof of this theorem is presented in Sect. 3. Technical details needed in the proof are provided in Sects. 4 and 5.

2 Applications

In this section we describe some applications of residual-based empirical distribution functions. These applications have versions in nonparametric regression and have been extensively studied there.

Quantile functions. By Proposition 1 of [15] on compact differentiability of quantile functions we obtain from Theorem 1 the following uniform stochastic expansion for the residual-based empirical quantile function. For $0 < \alpha < \beta < 1$,

$$\sup_{\alpha \le u \le \beta} \left| \hat{\mathbb{F}}^{-1}(u) - F^{-1}(u) + \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\mathbf{1}[\varepsilon_j \le F^{-1}(u)] - u}{f(F^{-1}(u))} + \varepsilon_j \right) \right| = o_p(n^{-1/2}).$$

Prediction intervals. A predictor for X_{n+1} is $\hat{r}(X_n)$. By the above result on the quantile function, the probability that X_{n+1} lies in the prediction interval $[\hat{r}(X_n) + \hat{\mathbb{F}}^{-1}(\alpha/2), \hat{r}(X_n) + \hat{\mathbb{F}}^{-1}(1 - \alpha/2)]$ converges to $1 - \alpha$. For a related result in nonparametric (and heteroscedastic) *regression* see [2].

Goodness-of-fit tests for the innovation distribution. In order to test for a specific form of the innovation distribution function F, we can use, e.g. the Kolmogorov–Smirnov statistic

$$n^{1/2} \sup_{t \in \mathbb{R}} |\hat{\mathbb{F}}(t) - F(t)|$$

or the Cramér–von Mises statistic

$$n \int (\hat{\mathbb{F}}(t) - F(t))^2 d\hat{\mathbb{F}}(t).$$

Similarly, tests for parametric models F_ϑ can be based, e.g. on

$$n^{1/2} \sup_{t \in \mathbb{R}} |\hat{\mathbb{F}}(t) - F_{\hat{\vartheta}}(t)|$$

or

$$n \int (\hat{\mathbb{F}}(t) - F_{\hat{\vartheta}}(t))^2 d\hat{\mathbb{F}}(t)$$

for some estimator $\hat{\vartheta}$, for example the residual-based maximum likelihood estimator.

Goodness-of-fit tests for the autoregression function. Suppose we want to test the null hypothesis that we have a parametric form $r = r_\vartheta$ for the autoregression function. Let $\hat{\vartheta}$ denote the least squares estimator for ϑ, i.e. a minimizer of $\sum_{j=1}^{n} (X_j - r_\vartheta(X_{j-1}))^2$. Let $\hat{\varepsilon}_{0j} = X_j - r_{\hat{\vartheta}}(X_{j-1})$ denote the residuals under the null hypothesis, and let $\hat{\mathbb{F}}_0(t) = (1/n) \sum_{j=1}^{n} \mathbf{1}[\hat{\varepsilon}_{0j} \le t]$ denote the corresponding empirical distribution function. We can then base a test for the null hypothesis on the Kolmogorov–Smirnov statistic

$$n^{1/2} \sup_{t\in\mathbb{R}} |\hat{\mathbb{F}}(t) - \hat{\mathbb{F}}_0(t)|$$

or the Cramér–von Mises statistic

$$n \int (\hat{\mathbb{F}}(t) - \hat{\mathbb{F}}_0(t))^2 d\hat{\mathbb{F}}(t).$$

For a related approach in (heteroscedastic) *regression* see [38].

For other applications of residual-based empirical distribution functions we refer to [11,13,33,34].

3 Proof of Theorem 1

In this section we give the proof of our theorem. We will make repeated use of the following exponential inequality for martingales in [14].

Lemma 1 *Let $Y_1, \dots, Y_n$ be a sequence of martingale increments (with respect to a filtration $\mathcal{F}_0, \dots, \mathcal{F}_n$) bounded by c. Set $S_n = \sum_{j=1}^n Y_j$ and $T_n = \sum_{j=1}^n E(Y_j^2|\mathcal{F}_{j-1})$. Then for positive s and t one has*

$$P(S_n \geq s,\ T_n \leq t) \leq \exp\left(-\frac{s^2}{2sc + 2t}\right).$$

Throughout we assume that the assumptions of Theorem 1 are met. These imply that the innovation density f is bounded:

$$\|f\|_\infty = \sup_{t\in\mathbb{R}} f(t) < \infty.$$

The stationary density g of our nonparametric autoregression model can and will be chosen to satisfy (1.1) and is hence positive, bounded and Hölder with exponent ξ. For a continuous function h on $\mathbb{R}$ and an interval I we let

$$\|h\|_I = \sup_{x\in I} |h(x)|.$$

We begin by studying the behavior of the local linear smoother on the interval I_n. To this end we introduce for a non-negative integer i the function K_i by $K_i(u) = u^i K(u)$ and the random functions $\hat{p}_i$ and $\hat{q}_i$ by

$$\hat{p}_i(x) = \frac{1}{nc_n} \sum_{j=1}^n K_i\left(\frac{X_{j-1} - x}{c_n}\right), \quad x \in \mathbb{R},$$

and

$$\hat{q}_i(x) = \frac{1}{nc_n} \sum_{j=1}^{n} X_j K_i \left(\frac{X_{j-1} - x}{c_n} \right), \quad x \in \mathbb{R}.$$

It is easy to check that on the event $\{\hat{p}_2(x)\hat{p}_0(x) - \hat{p}_1^2(x) > 0\}$ we have the identity

$$\hat{r}(x) = \frac{\hat{p}_2(x)\hat{q}_0(x) - \hat{p}_1(x)\hat{q}_1(x)}{\hat{p}_2\hat{p}_0(x) - \hat{p}_1^2(x)}.$$

By the properties of f and K, we obtain from Lemmas 3 and 4 in Sect. 4 and the choice of bandwidth that

$$\sup_{x \in I_n} \left| \hat{p}_i(x) - E[\hat{p}_i(x)] \right| = O_p(n^{-1/3}), \quad i = 0, 1, 2, \ldots \tag{3.1}$$

Let us now set

$$\lambda_i = \int K_i(u)\, du = \int u^i K(u)\, du, \quad i = 0, 1, 2, \ldots$$

Since the density g is Hölder with exponent ξ and the kernel K has compact support, we obtain in view of the identity

$$\bar{p}_i(x) = E[\hat{p}_i(x)] = \int g(x - c_n u) u^i K(u)\, du, \quad x \in \mathbb{R},$$

that

$$\sup_{x \in \mathbb{R}} \left| E[\hat{p}_i(x)] - \lambda_i g(x) \right| = O(c_n^{\xi}), \quad i = 0, 1, 2, \ldots \tag{3.2}$$

It follows from (I), (3.1) and (3.2) that

$$\| \hat{p}_i / g - \lambda_i \|_{I_n} + \| \bar{p}_i / g - \lambda_i \|_{I_n} = o_p(n^{-1/12}), \quad i = 0, 1, 2, \ldots \tag{3.3}$$

As K is a density with mean zero, we have $\lambda_0 = 1$, $\lambda_1 = 0$ and $\lambda_2 > 0$ and obtain

$$\| \hat{p}_2 \hat{p}_0 - \hat{p}_1^2 - \lambda_2 g^2 \|_{I_n} = o_p(n^{-1/12}).$$

Since $\log n \inf_{x \in I_n} g(x)$ is bounded away from zero and λ_2 is positive, there exists an $\alpha > 0$ such that

$$P\left(\log^2 n \inf_{x \in I_n} \left| \hat{p}_2(x)\hat{p}_0(x) - \hat{p}_1^2(x) \right| > \alpha \right) \to 1. \tag{3.4}$$

We can write $\hat{q}_i = A_i + B_i$, where

$$A_i(x) = \frac{1}{nc_n} \sum_{j=1}^{n} \varepsilon_j K_i \left(\frac{X_{j-1} - x}{c_n} \right), \quad x \in \mathbb{R},$$

and

$$B_i(x) = \frac{1}{nc_n} \sum_{j=1}^{n} r(X_{j-1}) K_i \left(\frac{X_{j-1} - x}{c_n} \right), \quad x \in \mathbb{R}.$$

Since r has a bounded second derivative, a Taylor expansion shows that

$$\|(B_i - r\hat{p}_i - r' c_n \hat{p}_{i+1})/g\|_{I_n} \leq \sup_{x \in \mathbb{R}} |r''(x)| c_n^2 \|\hat{p}_0/g\|_{I_n} = O_p(c_n^2). \tag{3.5}$$

It follows from Lemma 5 in Sect. 4 that

$$\|A_i\|_{I_n} = O_p(n^{-3/8} \log^{5/8} n), \quad i = 0, 1. \tag{3.6}$$

Relations (3.1)–(3.6) imply that

$$\hat{\Delta} = \hat{r} - r = \hat{u} + \hat{v},$$

where

$$\hat{v}(x) = \frac{\bar{p}_2(x) A_0(x) - \bar{p}_1(x) A_1(x)}{\bar{p}_2(x)\bar{p}_0(x) - \bar{p}_1^2(x)}, \quad x \in \mathbb{R}, \tag{3.7}$$

and

$$\|\hat{u}\|_{I_n} = O_p((n \log n)^{-1/2}). \tag{3.8}$$

Since K is three times continuously differentiable, so are $\bar{p}_i$ and A_i. From Lemma 5 in Sect. 4 we derive the following rates for the derivatives of A_i,

$$\left\| A_i^{(\nu)} \right\| = O\left(c_n^{-\nu} n^{-3/8} \log^{5/8} n \right), \quad \nu = 0, 1, 2.$$

As K_i' integrates to zero, we can write

$$c_n \bar{p}_i'(x) = \int g(x - c_n u) K_i'(u)\, du = \int (g(x - c_n u) - g(x)) K_i'(u)\, du$$

and obtain $\|c_n \bar{p}_i'/g\|_{I_n} = O(c_n^{\xi} \log n)$ by (I) and the Hölder property of g. Similarly one verifies $\|c_n^2 \bar{p}_i''/g\|_{I_n} = O(c_n^{\xi} \log n)$. By (3.3) we have $\|\bar{p}_i/g\|_{I_n} = O(1)$.

We derive that $s_i = \bar{p}_{2-i}/(\bar{p}_2\bar{p}_0 - \bar{p}_1^2)$ satisfies

$$\|s_i\|_{I_n} = O(\log n), \quad \|c_n s_i'\|_{I_n} = o(1) \quad \text{and} \quad \|c_n^2 s_i''\|_{I_n} = o(1), \quad i = 0, 1.$$

As $\hat{v} = s_0 A_0 - s_1 A_1$, we conclude that

$$\|\hat{v}\|_{I_n} = o_p(n^{-3/8}\log^2 n), \tag{3.9}$$

$$\|\hat{v}'\|_{I_n} = o_p(n^{-1/8}\log^2 n), \tag{3.10}$$

$$\|\hat{v}''\|_{I_n} = o_p(n^{1/8}\log^3 n). \tag{3.11}$$

Moreover, it follows from Lemma 6 that

$$\frac{1}{n}\sum_{j=1}^{n} w_{nj}\hat{v}(X_{j-1}) = \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j + o_p(n^{-1/2}). \tag{3.12}$$

Let $\mathbb{F}_w$ denote the weighted empirical distribution function based on the unobserved innovations, defined by

$$\mathbb{F}_w(t) = \sum_{j=1}^{n}\bar{w}_j \mathbf{1}[\varepsilon_j \le t], \quad t \in \mathbb{R}.$$

It is easy to check that

$$\sup_{t\in\mathbb{R}} |\mathbb{F}_w(t) - \mathbb{F}(t)| = o_p(n^{-1/2})$$

and

$$\bar{W} = \frac{1}{n}\sum_{j=1}^{n} w_{nj} = 1 + o_p(1).$$

We have the identity

$$\bar{W}\left(\hat{\mathbb{F}}(t) - \mathbb{F}_w(t)\right) = H(t, \hat{\Delta}) - H(t, 0) + B(t, \hat{\Delta}),$$

where

$$B(t, \Delta) = \frac{1}{n}\sum_{j=1}^{n} w_{nj}\left(F(t + \Delta(X_{j-1})) - F(t)\right)$$

and

$$H(t, \Delta) = \frac{1}{n}\sum_{j=1}^{n} w_{nj}\left(\mathbf{1}[\varepsilon_j \le t + \Delta(X_{j-1})] - F(t + \Delta(X_{j-1}))\right)$$

for t in $\mathbb{R}$ and Δ in $C(\mathbb{R})$, the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. As f is Hölder of order ξ greater than $1/3$, we derive

$$\sup_{t\in\mathbb{R}}\left|B(t,\hat{\Delta})-f(t)\frac{1}{n}\sum_{j=1}^{n}w_{nj}\hat{\Delta}(X_{j-1})\right|\le\frac{1}{n}\sum_{j=1}^{n}w_{nj}L|\hat{\Delta}(X_{j-1})|^{1+\xi},$$

where L is the Hölder constant of f. In view of this, relations (3.8), (3.9) and (3.12) yield

$$\sup_{t\in\mathbb{R}}\left|B(t,\hat{\Delta})-f(t)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\right|=o_p(n^{-1/2}).$$

Thus we are left to show that

$$\sup_{t\in\mathbb{R}}\left|H(t,\hat{\Delta})-H(t,0)\right|=o_p(n^{-1/2}).$$

Since the innovations have a finite second moment, we have

$$\max_{1\le j\le n}|\varepsilon_j|=o_p(n^{1/2}).$$

Since $\|\hat{\Delta}\|_{I_n}=o_p(1)$, the probability of the event

$$\left\{\max_{1\le j\le n}|\varepsilon_j|\le n^{1/2}-1\right\}\cap\{\|\hat{\Delta}\|_{I_n}<1\}$$

tends to one. On this event we have

$$\begin{aligned}\sup_{|t|>n^{1/2}}|H(t,\hat{\Delta})-H(t,0)|&=\sup_{|t|>n^{1/2}}B(t,\hat{\Delta})\\&\le 2F(1-n^{1/2})+2(1-F(n^{1/2}-1)).\end{aligned}$$

Since F has a finite second moment, we have $F(t)=o(t^{-2})$ as $t\to-\infty$ and $1-F(t)=o(t^{-2})$ as $t\to\infty$. This shows that

$$\sup_{|t|>n^{1/2}}|H(t,\hat{\Delta})-H(t,0)|=o_p(n^{-1}).$$

Now fix a δ in the interval $(1/3,1/2)$. For an interval I, let $C_1^{1+\delta}(I)$ be the set of differentiable functions h on $\mathbb{R}$ that satisfy $\|h\|_{I,\delta}\le 1$ where

$$\|h\|_{I,\delta}=\|h\|_I+\|h'\|_I+\sup_{x,y\in I,x\ne y}\frac{|h'(x)-h'(y)|}{|y-x|^{\delta}}.$$

It follows from (3.9)–(3.11) that $\hat{v}$ belongs to $C_1^{1+\delta}(I_n)$ with probability tending to 1. Indeed from (3.10) we obtain

$$\sup_{x,y\in I_n,|y-x|>n^{-1/4}} \frac{|\hat{v}'(x)-\hat{v}'(y)|}{|y-x|^\delta} \le 2n^{\delta/4}\|\hat{v}'\|_{I_n} = o_p(n^{-1/8+\delta/4}\log^2 n),$$

and from (3.11) we obtain

$$\sup_{x,y\in I_n,|y-x|\le n^{-1/4}} \frac{|\hat{v}'(x)-\hat{v}'(y)|}{|y-x|^\delta} \le n^{-(1-\delta)/4}\|\hat{v}''\|_{I} = o_p(n^{-1/8+\delta/4}\log^3 n).$$

Since $-1/8+\delta/4<0$ by choice of δ, the above and relations (3.9) and (3.10) yield that

$$\|\hat{v}\|_{I_n,\delta} = o_p(1). \tag{3.13}$$

Now let $\mathscr{D}_n = \{u+v : u \in \mathscr{U}_n, v \in \mathscr{V}_n\}$, where

$$\mathscr{U}_n = \{h \in C(\mathbb{R}) : \|h\|_{I_n} \le n^{-1/2}\log^{-1/4} n\},$$
$$\mathscr{V}_n = \{h \in C_1^{1+\delta}(I_n) : \|h\|_{I_n} \le n^{-3/8}\log^2 n\}.$$

By (3.8), $\hat{u}$ belongs to $\mathscr{U}_n$ with probability tending to one; by (3.9) and (3.13), $\hat{v}$ belongs to $\mathscr{V}_n$ with probability tending to one. This shows that $\hat{\Delta}$ belongs to $\mathscr{D}_n$ with probability tending to one. In view of this we are left to show

$$\sup_{|t|\le n^{1/2},\Delta\in\mathscr{D}_n} |H(t,\Delta)-H(t,0)| = o_p(n^{-1/2}). \tag{3.14}$$

To this end set $\eta_n = n^{-1/2}\log^{-1/4} n$. Let $t_1,\dots,t_{M_n}$ be an η_n-net of $[-n^{1/2},n^{1/2}]$, and let $v_1,\dots,v_{N_n}$ denote an η_n-net for $\mathscr{V}_n$ for the pseudo-norm $\|\cdot\|_{I_n}$. We can choose the former net such that

$$M_n \le 2 + n\log^{1/4} n, \tag{3.15}$$

while we can take the latter net such that

$$N_n \le \exp\Big(K_*(2+b_n-a_n)(n\log^{1/2} n)^{1/(2+2\delta)}\Big) \tag{3.16}$$

for some constant K_*; see Theorem 2.7.1 in [37]. Note also that $v_1,\dots,v_{N_n}$ is an $2\eta_n$-net for $\mathscr{D}_n$. We have

$$\sup_{|t|\le n^{1/2},\Delta\in\mathscr{D}_n} |H(t,\Delta)-H(t,0)| \le \max_{i,l}|H(t_i,v_l)-H(t_i,0)| + \max_{i,l} D_{i,l},$$

where

$$D_{i,l} = \sup_{|t-t_i|\le\eta_n, \|\Delta-v_l\|_{I_n}\le 2\eta_n} \left(|H(t,\Delta) - H(t_i, v_l)| + |H(t,0) - H(t_i,0)|\right).$$

For $|t - t_i| \le \eta_n$ and $\|\Delta - v_l\|_{I_n} \le 2\eta_n$ we have

$$\mathbf{1}[y \le t_i - 3\eta_n + v_l(x)] \le \mathbf{1}[y \le t + \Delta(x)] \le \mathbf{1}[y \le t_i + 3\eta_n + v_l(x)]$$

and

$$F(t_i - 3\eta_n + v_l(x)) \le F(t + \Delta(x)) \le F(t_i + 3\eta_n + v_l(x))$$

for all $y \in \mathbb{R}$ and $x \in I_n$ and thus obtain

$$|H(t,\Delta) - H(t_i, v_l)| \le H(t_i + 3\eta_n, v_l) - H(t_i - 3\eta_n, v_l) + 2R_{i,l}$$

with

$$\begin{aligned} R_{i,l} &= \frac{1}{n}\sum_{j=1}^{n} w_{nj}\left(F(t_i + 3\eta_n + v_l(X_{j-1})) - F(t_i - 3\eta_n + v_l(X_{j-1}))\right) \\ &\le 6\|f\|_\infty \eta_n. \end{aligned}$$

Similarly, we derive the bound

$$|H(t,0) - H(t_i,0)| \le H(t_i + \eta_n, 0) - H(t_i - \eta_n, 0) + 4\|f\|_\infty \eta_n.$$

Thus we have the following bound:

$$\sup_{|t|\le n^{1/2}, \Delta\in\mathscr{D}_n} |H(t,\Delta) - H(t,0)| \le T_1 + T_2 + T_3 + 16\|f\|_\infty \eta_n,$$

where

$$\begin{aligned} T_1 &= \max_{i,l} |H(t_i, v_l) - H(t_i, 0)|, \\ T_2 &= \max_{i,l} H(t_i + 3\eta_n, v_l) - H(t_i - 3\eta_n, v_l), \\ T_3 &= \max_{i,l} H(t_i + \eta_n, 0) - H(t_i - \eta_n, 0). \end{aligned}$$

To continue we need the following lemma which follows from a simple application of Freedman's inequality.

Lemma 2 *Let s, t be real numbers and u and v be continuous functions. Then, for every $\beta > 0$ and every $\alpha \ge |t - s| + \|u - v\|_{I_n}$, we have*

$$P(|H(s,u)-H(t,v)|>\beta n^{-1/2})\le 2\exp\left(-\frac{\beta^2 n}{4\beta n^{1/2}+2n\alpha\|f\|_\infty}\right).$$

Proof We apply Lemma 1 with

$$\begin{aligned}Y_j = w_{nj}\big(&\mathbf{1}[\varepsilon_j \le s+u(X_{j-1})]-\mathbf{1}[\varepsilon_j\le t+v(X_{j-1})]\\ &-F(s+u(X_{j-1}))+F(t+v(X_{j-1}))\big).\end{aligned}$$

We have $|Y_j|\le 2$, $E(Y_j|X_0,\dots,X_{j-1})=0$ and

$$\begin{aligned}V_n &= \sum_{j=1}^n E(Y_j^2|X_0,\dots,X_{j-1}) \le \sum_{j=1}^n w_{nj}\left|F(s+u(X_{j-1})-F(t+v(X_{j-1})\right|\\ &\le n\|f\|_\infty(|t-s|+\|u-v\|_{I_n})\le n\alpha\|f\|_\infty.\end{aligned}$$

Since

$$P(|H(s,u)-H(t,v)|>\beta n^{-1/2}) = P\left(\left|\sum_{j=1}^n Y_j\right|>\beta n^{1/2},\ V_n\le n\|f\|_\infty\alpha\right),$$

the desired result follows from an application of Lemma 1. □

Note that $\|v_l\|_{I_n}\le n^{-3/8}\log^2 n+\eta_n$. Thus we obtain from Lemma 2 that

$$\begin{aligned}P(T_1>\beta n^{-1/2}) &\le \sum_{i,l} P(|H(t_i,v_l)-H(t_i,0)|>\beta n^{-1/2})\\ &\le 2M_nN_n\exp\left(-\frac{\beta^2 n}{4\beta n^{1/2}+2n\|f\|_\infty(n^{-3/8}\log^2 n+\eta_n)}\right).\end{aligned}$$

Similarly,

$$P(T_2>\beta n^{-1/2})\le 2M_nN_n\exp\left(-\frac{\beta^2 n}{4\beta n^{1/2}+12n\|f\|_\infty\eta_n}\right)$$

and

$$P(T_3>\beta n^{-1/2})\le 2M_nN_n\exp\left(-\frac{\beta^2 n}{4\beta n^{1/2}+4n\|f\|_\infty\eta_n}\right).$$

As $1/(2+2\delta)<3/8$, we obtain from the above and from relations (3.15) and (3.16) and the fact that $b_n-a_n=O(\log n)$ that

$$P(T_i>\beta n^{-1/2})\to 0,\quad i=1,2,3,\ \beta>0.$$

This completes the proof of (3.14) and hence the proof of Theorem 1.

4 Technical details

Let v be a measurable function and c_n a sequence of bandwidths. Let $t_1, t_2, \dots$ be measurable functions which are bounded by the same constant B. In this section we study the behavior of the processes

$$\hat{T}_n(x) = \frac{1}{nc_n} \sum_{j=1}^{n} t_n(X_j) v\left(\frac{X_j - x}{c_n}\right), \quad x \in \mathbb{R}, \tag{4.1}$$

and

$$U_n(x) = \frac{1}{nc_n} \sum_{j=1}^{n} \varepsilon_j v\left(\frac{X_{j-1} - x}{c_n}\right), \quad x \in \mathbb{R}, \tag{4.2}$$

on the interval I_n. For this we will use the following result.

Proposition 1 *For each x in $\mathbb{R}$, let h_{nx} be a bounded and measurable function from $\mathbb{R}^2$ into $\mathbb{R}$ such that*

$$E(h_{nx}(X_0, X_1)|X_0) = 0. \tag{4.3}$$

Suppose there are positive numbers κ_1, κ_2 and C such that

$$\sup_{x \in I_n} |h_{nx}(X_0, X_1)| \le C/\log n, \tag{4.4}$$

$$P\left(\sup_{x \in I_n} \sum_{j=1}^{n} E(h_{nx}^2(X_{j-1}, X_j)|X_{j-1}) > C/\log n\right) \to 0, \tag{4.5}$$

$$|h_{ny}(X_0, X_1) - h_{nx}(X_0, X_1)| \le Cn^{\kappa_2}|y - x|^{\kappa_1}, \quad x, y \in \mathbb{R}. \tag{4.6}$$

Then there is a constant A such that

$$P\left(\sup_{x \in I_n} \left|\sum_{j=1}^{n} h_{nx}(X_{j-1}, X_j)\right| > A\right) \to 0. \tag{4.7}$$

Proof Let us set $D_j(x) = h_{nx}(X_{j-1}, X_j)$. Then $M_n(x) = \sum_{j=1}^{n} D_j(x)$ is a sum of martingale differences with $|D_j(x)| \le C/\log n$. Set $W_n(x) = \sum_{j=1}^{n} E(D_j^2(x)|X_{j-1})$. It follows from Lemma 1 that

$$P\left(|M_n(x)| \ge \eta,\ W_n(x) \le \frac{C}{\log n}\right) \le 2\exp\left(-\frac{\eta^2 \log n}{2(1+\eta)C}\right), \quad \eta > 0.$$

Now let $x_{nk} = a_n + k(b_n - a_n)n^{-m}$ for $k = 0, 1, \ldots, n^m$, with m an integer greater than $(1 + \kappa_2)/\kappa_1$. We have

$$\sup_{x \in I_n} |M_n(x)| \le \max_{k=0,\ldots,n^m} |M_n(x_{nk})| + Q_n,$$

where, in view of (4.6),

$$\begin{aligned} Q_n &= \max_{k=0,\ldots,n^m} \sup_{|x - x_{nk}| \le (b_n - a_n)n^{-m}} |M_n(x) - M_n(x_{nk})| \\ &\le C n^{1+\kappa_2} (b_n - a_n)^{\kappa_1} n^{-m\kappa_1} \to 0. \end{aligned}$$

Now consider the events

$$A_n = \left\{ \max_{k=0,\ldots,n^m} |M_n(x_{nk})| > 1 + 2(m+2)C \right\}$$

and

$$B_n = \left\{ \sup_{x \in I_n} W_n(x) \le \frac{C}{\log n} \right\}.$$

The above yields, with $\eta = 1 + 2(m+2)C$,

$$\begin{aligned} P(A_n) &\le P(B_n^c) + P(A_n \cap B_n) \\ &\le P(B_n^c) + \sum_{k=0}^{n^m} P\left(|M_n(x_{nk})| > \eta,\ W_n(x_{nk}) \le \frac{C}{\log n} \right) \\ &\le P(B_n^c) + 2(1 + n^m) \exp\left(-\frac{(\eta - 1)\log n}{2C} \right) = o(1). \end{aligned}$$

Thus the desired result (4.7) holds with $A = 2 + 2C(m+2)$. □

Let us now compare $\hat{T}_n$ with $\tilde{T}_n$, where

$$\tilde{T}_n(x) = \frac{1}{nc_n} \sum_{j=1}^{n} E\left(t_n(X_j) v\left(\frac{X_j - x}{c_n} \right) \Big| X_{j-1} \right), \quad x \in \mathbb{R}.$$

Lemma 3 *Suppose f is bounded and v is integrable and Lipschitz. Let $c_n \to 0$ and $nc_n/\log n \to \infty$. Then*

$$\sup_{x \in I_n} |\hat{T}_n(x) - \tilde{T}_n(x)| = O_p\left(\left(\frac{\log n}{nc_n} \right)^{1/2} \right).$$

Proof We apply Proposition 1 with

$$h_{nx}(X_0, X_1) = \frac{1}{s_n}\left(t_n(X_1)v\left(\frac{X_1 - x}{c_n}\right) - E\left(t_n(X_1)v\left(\frac{X_1 - x}{c_n}\right)\Big|X_0\right)\right),$$

where $s_n = (nc_n \log n)^{1/2}$. Assumption (4.3) holds by construction. In order to show (4.4) note that the assumptions on v imply that v is bounded and square-integrable. We have

$$\sup_{x\in I_n} |h_{nx}(X_0, X_1)| \leq \frac{2B\|v\|_\infty}{\sqrt{nc_n \log n}}.$$

This is of the desired order $O(1/\log n)$ since $\log n/(nc_n) \to 0$ by assumption. Next, we have

$$\sum_{j=1}^{n} E(h_{nx}^2(X_j, X_{j-1})|X_{j-1}) \leq \frac{B^2}{s_n^2}\sum_{j=1}^{n} E\left(v^2\left(\frac{X_j - x}{c_n}\right)\Big|X_{j-1}\right), \quad x \in \mathbb{R}.$$

This yields the desired (4.5) in view of $n/s_n^2 = 1/(c_n \log n)$, stationarity, and the bound

$$\begin{aligned}\frac{1}{c_n}E\left(v^2\left(\frac{X_1 - x}{c_n}\right)\Big|X_0\right) &= \int \frac{1}{c_n}v^2\left(\frac{y + r(X_0) - x}{c_n}\right) f(y)\,dy\\ &= \int v^2(u) f(x - r(X_0) + c_n u)\,du\\ &\leq \|f\|_\infty \int v^2(u)\,du.\end{aligned}$$

Finally, relation (4.6) follows with $\kappa_1 = \kappa_2 = 1$ from the bound

$$\begin{aligned}|h_{ny}(X_0, X_1) - h_{nx}(X_0, X_1)| &\leq \frac{2B}{s_n}\sup_{z\in\mathbb{R}}\left|v\left(\frac{z - y}{c_n}\right) - v\left(\frac{z - x}{c_n}\right)\right|\\ &\leq \frac{2B\Lambda}{s_n c_n}|y - x|,\end{aligned}$$

where Λ is the Lipschitz constant of v, and the fact that $nc_n s_n \to \infty$. □

Lemma 4 *Suppose f is bounded and v is integrable and has a bounded derivative v' such that the integral $V = \int(1 + |u|)|v'(u)|\,du$ is finite. Suppose the functions $t_0 = f, t_1, t_2, \ldots$ satisfy*

$$|t_m(y) - t_m(x)| \leq H_m|y - x|^{\xi_0}, \quad x, y \in \mathbb{R}, m = 0, 1, 2\ldots,$$

for some exponent ξ_0, $0 \leq \xi_0 \leq 1$. Then

$$\sup_{x\in I_n} |\tilde{T}_n(x) - E(\tilde{T}_n(x))| = O_p\left((H_0 + H_n)(b_n - a_n)n^{-1/2}c_n^{\xi_0 - 1}\right).$$

Proof For $s \in \mathbb{R}$, let us define the function $\phi_{n,s}$ by

$$\phi_{n,s}(x) = t_n(x) f(x-s), \quad x \in \mathbb{R}.$$

By the properties of f and t_n, the functions $\phi_{n,s}$ are bounded by $B\|f\|_\infty$ and Hölder with exponent ξ_0 and constant $\Lambda_n = BH_0 + \|f\|_\infty H_n$,

$$|\phi_{n,s}(x) - \phi_{n,s}(y)| \le \Lambda_n |x-y|^{\xi_0}. \tag{4.8}$$

It is easy to see that

$$\tilde{T}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \psi_{n,r(X_{j-1})}(x), \quad x \in \mathbb{R},$$

where

$$\psi_{n,s}(x) = \int \frac{1}{c_n} v\left(\frac{y-x}{c_n}\right) \phi_{n,s}(y)\,dy = \int \phi_{n,s}(x + c_n u) v(u)\,du, \quad x \in \mathbb{R}.$$

By the properties of v, the functions $\psi_{n,s}$ are bounded by $B\|f\|_\infty \|v\|_1$ and differentiable with derivatives

$$\psi'_{n,s}(x) = -\frac{1}{c_n} \int \phi_{n,s}(x + c_n u) v'(u)\,du, \quad x \in \mathbb{R}.$$

In view of $\int v'(u)\,du = 0$ we obtain

$$\psi'_{n,s}(x) = -\frac{1}{c_n} \int (\phi_{n,s}(x + c_n u) - \phi_{n,s}(x)) v'(u)\,du, \quad x \in \mathbb{R}.$$

Thus (4.8) implies that

$$|\psi'_{n,s}(x)| \le \Lambda_n c_n^{\xi_0 - 1} \int |u|^{\xi_0} |v'(u)|\,du, \quad x \in \mathbb{R}.$$

Hence the functions $\psi_{n,s}$ are Lipschitz with constant $L_n = V \Lambda_n c_n^{\xi_0 - 1}$.

Since the autoregressive process is geometrically ergodic, there is a constant D such that

$$\operatorname{Var}\left(n^{-1/2} \sum_{j=1}^{n} h(X_j)\right) \le D\|h\|_\infty^2$$

for every bounded measurable function h. Since

$$|\psi_{n,r(y)}(s) - \psi_{n,r(y)}(t))| \le L_n |s-t|, \quad s, t, y \in \mathbb{R},$$

we obtain that

$$\operatorname{Var}\left(n^{1/2}(\tilde{T}_n(s) - \tilde{T}_n(t))\right) \le DL_n^2(s-t)^2, \quad s, t \in I_n. \tag{4.9}$$

Thus it follows from Theorem 12.3 in [7] that the sequence of $C([0, 1])$-valued processes

$$\frac{n^{1/2}}{L_n(b_n - a_n)}\left(\tilde{T}_n(a_n + (b_n - a_n)x) - E[\tilde{T}_n(a_n + (b_n - a_n)x)]\right), \quad 0 \le x \le 1$$

is tight. This is the desired result. □

Lemma 5 *Suppose the function* v *is as in Lemma 4. Let* f *be bounded and have a finite moment of order* $\beta > 2$*. Let* $c_n \to 0$, $n^{1/2}c_n/\log n \to \infty$ *and* $c_n^{-1}n^{-1+2/\beta}\log n$ *be bounded. Then*

$$\sup_{x \in I_n} |U_n(x)| = O_p\left(\left(\frac{\log n}{nc_n}\right)^{1/2}\right).$$

Proof Let $s_n = (nc_n \log n)^{1/2}$. Define

$$R_{nj}(x) = \frac{1}{s_n}\left(\varepsilon_j \mathbf{1}\left[|\varepsilon_j| \le n^{1/\beta}\right] - E\left[\varepsilon_j \mathbf{1}\left[|\varepsilon_j| \le n^{1/\beta}\right]\right]\right) v\left(\frac{X_{j-1} - x}{c_n}\right),$$
$$S_{nj}(x) = \frac{1}{s_n}\varepsilon_j \mathbf{1}\left[|\varepsilon_j| > n^{1/\beta}\right] v\left(\frac{X_{j-1} - x}{c_n}\right),$$
$$\bar{S}_{nj}(x) = \frac{1}{s_n} E\left[\varepsilon_j \mathbf{1}\left[|\varepsilon_j| > n^{1/\beta}\right]\right] v\left(\frac{X_{j-1} - x}{c_n}\right).$$

Since ε has mean zero, it suffices to show that

$$\sup_{x \in I_n}\left|\sum_{j=1}^n R_{nj}(x)\right| = O_p(1), \tag{4.10}$$

$$\sup_{x \in I_n}\left|\sum_{j=1}^n S_{nj}(x)\right| = o_p(1), \tag{4.11}$$

$$\sup_{x \in I_n}\left|\sum_{j=1}^n \bar{S}_{nj}(x)\right| = o_p(1). \tag{4.12}$$

We have

$$P\left(\max_{1 \le j \le n} |\varepsilon_j| > n^{1/\beta}\right) \le \sum_{j=1}^n P(|\varepsilon_j| > n^{1/\beta}) \le E\left[|\varepsilon|^\beta \mathbf{1}\left[|\varepsilon| > n^{1/\beta}\right]\right] \to 0$$

and thus

$$P\left(\sup_{x\in I_n}\left|\sum_{j=1}^{n} S_{nj}(x)\right| > 0\right) \le P\left(\max_{1\le j\le n}|\varepsilon_j| > n^{1/\beta}\right) \to 0.$$

The assumptions on v imply that v is bounded, say by B. Hence we also have

$$\begin{aligned}\sup_{x\in I_n}\left|\sum_{j=1}^{n}\bar{S}_{nj}(x)\right| &\le \frac{nB}{s_n}E\left[\varepsilon\mathbf{1}\left[|\varepsilon| > n^{1/\beta}\right]\right]\\ &\le E\left[|\varepsilon|^{\beta}\mathbf{1}\left[|\varepsilon| > n^{1/\beta}\right]\right]\frac{nB}{s_n n^{(\beta-1)/\beta}}\\ &= o\left(n^{1/\beta}s_n^{-1}\right) = o\left(\left(n^{-1+2/\beta}c_n^{-1}\log^{-1} n\right)^{1/2}\right)\\ &= o\left(\frac{1}{\log n}\right).\end{aligned}$$

To show (4.10) we apply Proposition 1 with $h_{nx}(X_{j-1}, X_j) = R_{nj}(x)$. We have

$$\sup_{x\in I_n}|h_{nx}(X_0, X_1)| \le \frac{2Bn^{1/\beta}}{s_n} = O\left(\frac{1}{\log n}\right).$$

Next, for x in $\mathbb{R}$, we have

$$\sum_{j=1}^{n} E\left(h_{nx}^2(X_{j-1}, X_j)|X_{j-1}\right) \le \frac{\sigma^2}{\log n}H_n(x) \tag{4.13}$$

with

$$H_n(x) = \frac{1}{nc_n}\sum_{j=1}^{n} v^2\left(\frac{X_{j-1}-x}{c_n}\right).$$

Note that v^2 inherits the properties imposed on v. Thus Lemmas 3 and 4, applied with v^2 in place of v and with $\xi_0 = 0$, yield

$$\sup_{x\in I_n}|H_n(x) - E[H_n(x)]| = o_p(1).$$

Finally,

$$E[H_n(x)] \le \|f\|_\infty \int v^2(u)\,du, \quad x\in\mathbb{R}.$$

This shows that $P(\sup_{x\in I_n} H_n(x) > C) \to 0$ for large enough C. This yields (4.5) in view of (4.13).

Since v is Lipschitz for some constant Λ, we obtain

$$|h_{ny}(X_0, X_1) - h_{nx}(X_0, X_1)| \le \frac{2\Lambda n^{1/\beta}}{s_n c_n}|y - x| \le Cn|y - x|.$$

Thus the assumptions of the Proposition 1 hold, and we obtain (4.10). □

5 Proof of (3.12)

In this section we provide the proof of (3.12). More precisely, we prove the following lemma.

Lemma 6 *Suppose* (F), (R), (K) *and* (I) *hold and* $c_n \sim (n \log n)^{-1/4}$. *Then* (3.12) *holds.*

Proof Let us set

$$s_i(x) = \frac{\bar{p}_{2-i}(x)}{\bar{p}_2(x)\bar{p}_0(x) - \bar{p}_1^2(x)}, \quad x \in \mathbb{R},\ i = 0, 1.$$

Then we can write $\hat{v} = s_0 A_0 - s_1 A_1$. Changing the order of summation leads to the identity

$$\frac{1}{n}\sum_{j=1}^{n} w_{nj}\hat{v}(X_{j-1}) = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \hat{h}(X_{k-1})$$

with $\hat{h} = \hat{h}_0 - \hat{h}_1$, where for $i = 0, 1$ and $x \in \mathbb{R}$,

$$\hat{h}_i(x) = \frac{1}{nc_n}\sum_{j=1}^{n} w_n(X_{j-1})s_i(X_{j-1})K_i\left(\frac{x - X_{j-1}}{c_n}\right).$$

Let $\bar{h}_n(x) = E[\hat{h}(x)]$. We calculate

$$\bar{h}_n(x) = \int w_n(x - c_n u)g(x - c_n u)\left(s_0(x - c_n u) - u s_1(x - c_n u)\right)K(u)\,du.$$

It follows from (3.3) that

$$\sup_{x\in I_n} |g(x)s_0(x) - 1| = o(n^{-1/12}) \quad \text{and} \quad \sup_{x\in I_n} |g(x)s_1(x)| = o(n^{-1/12}).$$

Using these properties it is easy to verify that $E[(\bar{h}_n(X)-1)^2] \to 0$. Therefore

$$\frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \left(\bar{h}_n(X_{k-1}) - 1\right) = o_p(n^{-1/2}).$$

Indeed a martingale argument shows that the second moment of the left-hand side is bounded by $E[\varepsilon^2]E[(\bar{h}_n(X)-1)^2]/n$.

Thus we are left to show that

$$\frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \left(\hat{h}(X_{k-1}) - \bar{h}_n(X_{k-1})\right) = o_p(n^{-1/2}). \tag{5.1}$$

Abbreviate $\hat{h} - \bar{h}_n$ by $\hat{h}_*$. Note that $\hat{h}_*(x) = 0$ for x outside the interval $J_n = [a_n - c_n, b_n + c_n]$ and that $w_n s_0/\log n$ and $w_n s_1/\log n$ are uniformly bounded and Hölder with exponent $\xi > 1/3$ and constant $H_n = O(\log n)$. Applying Lemmas 3 and 4 with I_n replaced by J_n, with $t_n = w_n s_i/\log n$ and with the choices $v = K_i$, $v = K_i'$ and $v = K_i''$ for $i = 0, 1$, we obtain

$$\|\hat{h}_*\|_\infty = o_p(n^{-1/3}), \quad \|\hat{h}_*'\|_\infty = o_p(n^{-1/12}) \quad \text{and} \quad \|\hat{h}_*''\|_\infty = o_p(n^{1/6}).$$

By (F), f has a finite moment of order $\beta > 8/3$. Hence we obtain $\max_k |\varepsilon_k| = o_p(n^{-1/\beta})$ and $\mu_n = E[\varepsilon 1[|\varepsilon| \le n^{1/\beta}]] = O_p(n^{-(\beta-1)/\beta}) = o_p(n^{-1/2})$ as shown in the proof of Lemma 5. Thus the desired (5.1) follows if we show that

$$\frac{1}{n}\sum_{k=1}^{n} \varepsilon_{n,k}\hat{h}_*(X_{k-1}) = o_p(n^{-1/2}), \tag{5.2}$$

where $\varepsilon_{n,k} = \varepsilon_k \mathbf{1}[|\varepsilon_k| \le n^{1/\beta}] - \mu_n$. To this end let us first show that $P(\hat{h}_* \in \mathscr{H}_n) \to 1$, where $\mathscr{H}_n$ is the set of all differentiable functions h on $\mathbb{R}$ which vanish off J_n and satisfy

$$\|h\|_\infty \le n^{-1/3} \quad \text{and} \quad \|h\|_\infty + \|h'\|_\infty + \sup_{y \ne x} \frac{|h'(x) - h'(y)|}{|x-y|^{1/3}} \le 1.$$

Indeed, by the properties of $\hat{h}_*$ we obtain

$$\sup_{|y-x| > n^{-1/4}} \frac{|\hat{h}_*'(x) - \hat{h}_*'(y)|}{|y-x|^{1/3}} \le 2n^{1/12}\|\hat{h}_*'\|_\infty = o_p(1)$$

and

$$\sup_{|y-x| \le n^{-1/4}} \frac{|\hat{h}_*'(x) - \hat{h}_*'(y)|}{|y-x|^{1/3}} \le n^{-1/6}\|\hat{h}_*''\|_\infty = o_p(1).$$

Thus (5.2) follows if we show that

$$S_n^* = \sup_{h\in\mathscr{H}_n} |S_n(h)| = o_p(n^{-1/2}), \tag{5.3}$$

where

$$S_n(h) = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_{n,k} h(X_{k-1}).$$

Let $\eta_n = (n\log n)^{-1/2}$. Let $h_1, \dots, h_{N_n}$ denote an η_n-net of $\mathscr{H}_n$. Then we have the bound

$$S_n^* \le \max_{1\le\nu\le N_n} |S_n(h_\nu)| + \frac{1}{n}\sum_{k=1}^{n} |\varepsilon_{n,k}|\eta_n = \max_{1\le\nu\le N_n} |S_n(h_\nu)| + o_p(n^{-1/2}).$$

If $\|h\|_\infty \le n^{-1/3}$, we derive from Lemma 1 that

$$\begin{aligned} P(|S_n(h)| > sn^{-1/2}) &\le 2\exp\left(-\frac{s^2 n}{4n^{1/\beta}n^{-1/3}sn^{1/2} + 2\sigma^2 nn^{-2/3}}\right) \\ &\le 2\exp\left(-\frac{s^2 n^{11/24}}{4s + 2\sigma^2}\right), \quad s > 0. \end{aligned}$$

In the last step we used the fact that $\beta > 8/3$. In view of Theorem 2.7.1 in [37], we can take

$$N_n \le \exp\left(K_*(2 + 2c_n + b_n - a_n)(n\log n)^{3/8}\right) \tag{5.4}$$

for some constant K_*. Thus we obtain

$$P\left(\max_{1\le\nu\le N_n} |S_n(h_\nu)| > sn^{-1/2}\right) \le 2N_n \exp\left(-\frac{s^2 n^{11/24}}{4s + 2\sigma^2}\right) \to 0, \quad s > 0.$$

This completes the proof of (5.3). □

References

1. An, H.Z., Huang, F.C.: The geometrical ergodicity of nonlinear autoregressive models. Statist. Sinica **6**, 943–956 (1996)
2. Akritas, M.G., Van Keilegom, I.: Non-parametric estimation of the residual distribution. Scand. J. Statist. **28**, 549–567 (2001)
3. Berkes, I., Horváth, L.: Empirical processes of residuals. In: Dehling, H., Mikosch, T., Sørensen, M. (eds.) Empirical Process Techniques for Dependent Data, pp. 195–209, Birkhäuser, Boston (2002)
4. Berkes, I., Horváth, L.: Limit results for the empirical process of squared residuals in GARCH models. Stoch. Process. Appl. **105**, 271–298 (2003)

5. Bhattacharya, R.N., Lee, C.: Ergodicity of nonlinear first order autoregressive models. J. Theoret. Probab. **8**, 207–219 (1995)
6. Bhattacharya, R., Lee, C.: On geometric ergodicity of nonlinear autoregressive models. Statist. Probab. Lett. **22**, 311–315 (1995b). Erratum: **41** 439–440 (1999)
7. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
8. Boldin, M.V.: Estimation of the distribution of noise in an autoregression scheme. Theory Probab. Appl. **27**, 866–871 (1982)
9. Boldin, M.V.: On residual empirical distribution functions in ARCH models with applications to testing and estimation. Mitt. Math. Sem. Giessen **235**, 49–66 (1998)
10. Cheng, F.: Asymptotic distributions of error density and distribution function estimators in nonparametric regression. J. Statist. Plann. Inference **128**, 327–349 (2005)
11. Dette, H., Neumeyer, N., Van Keilegom, I.: A new test for the parametric form of the variance function in nonparametric regression. J. Roy. Statist. Soc. Ser. B **69**, 903–917 (2007)
12. Durbin, J.: Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. **1**, 279–290 (1973)
13. Einmahl, J., Van Keilegom, I.: Specification tests in nonparametric regression. J. Econom. **143**, 88–102 (2008)
14. Freedman, D.A.: On tail probabilities for martingales. Ann. Probab. **3**, 100–118 (1975)
15. Gill, R.D.: Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. With a discussion by J. A. Wellner and J. Præstgaard and a reply by the author. Scand. J. Statist. **16**, 97–128 (1989)
16. Grama, I.G., Neumann, M.H.: Asymptotic equivalence of nonparametric autoregression and nonparametric regression. Ann. Statist. **34**, 1701–1732 (2006)
17. Hansen, B.E.: Uniform convergence rates for kernel estimation with dependent data. To appear in: Econom. Theory **24**, (2008)
18. Horváth, L., Kokoszka, P., Teyssière, G.: Empirical process of the squared residuals of an ARCH sequence. Ann. Statist. **29**, 445–469 (2001)
19. Kawczak, J., Kulperger, R., Yu, H.: The empirical distribution function and partial sum process of residuals from a stationary ARCH with drift process. Ann. Inst. Statist. Math. **57**, 747–765 (2005)
20. Kiwitt, S., Nagel, E.-R., Neumeyer, N.: Empirical likelihood estimators for the error distribution in nonparametric regression models. Technical Report, Faculty of Mathematics, University of Bochum (2005)
21. Koul, H.L.: Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. Ann. Math. Statist. **40**, 1950–1979 (1969)
22. Koul, H.L.: Some convergence theorems for ranks and weighted empirical cumulatives. Ann. Math. Statist. **41**, 1768–1773 (1970)
23. Koul, H.L.: A weak convergence result useful in robust autoregression. J. Statist. Plann. Inference **29**, 291–308 (1991)
24. Koul, H.L.: Weighted Empirical Processes in Dynamic Nonlinear Models. Lecture Notes in Statistics 166. Springer, New York (2002)
25. Koul, H.L., Ling, S.: Fitting an error distribution in some heteroscedastic time series models. Ann. Statist. **34**, 994–1012 (2006)
26. Koul, H.L., Ossiander, M.: Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. Ann. Statist. **22**, 540–562 (1994)
27. Kulperger, R., Yu, H.: High moment partial sum processes of residuals in GARCH models and their applications. Ann. Statist. **33**, 2395–2422 (2005)
28. Lee, S., Taniguchi, M.: Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. Statist. Sinica **15**, 215–234 (2005)
29. Loynes, R.M.: The empirical distribution function of residuals from generalised regression. Ann. Statist. **8**, 285–299 (1980)
30. Mammen, E.: Empirical process of residuals for high-dimensional linear models. Ann. Statist. **24**, 307–335 (1996)
31. Müller, U.U., Schick, A., Wefelmeyer, W.: Estimating linear functionals of the error distribution in nonparametric regression. J. Statist. Plann. Inference **119**, 75–93 (2004)
32. Müller, U.U., Schick, A., Wefelmeyer, W.: Estimating the error distribution function in semiparametric regression. Statist. Decisions **25**, 1–18 (2007)
33. Neumeyer, N., Dette, H.: A note on one-sided nonparametric analysis of covariance by ranking residuals. Math. Methods Statist. **14**, 80–104 (2005)

34. Pardo-Fernández, J.C., Van Keilegom, I., González-Manteiga, W.: Testing for the equality of k regression curves. Statist. Sinica **17**, 1115–1137 (2007)
35. Portnoy, S.: Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters. Ann. Statist. **14**, 1152–1170 (1986)
36. Shorack, G.R.: Empirical and rank processes of observations and residuals. Canad. J. Statist. **12**, 319–332 (1984)
37. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. With Applications to Statistics. Springer, New York (1996)
38. Van Keilegom, I., González Manteiga, W., Sánchez Sellero, C.: Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. To appear in: Test (2007)