Nonparametric estimates for conditional quantiles of time series

Jürgen Franke , Peter Mwita, Weining Wang*

November 14, 2013

Abstract

We consider the problem of estimating the conditional quantile of a time series $\{Y_t\}$ at time t given covariates $\mathbf{X}_t$, where $\mathbf{X}_t$ can either exogenous variables or lagged variables of Y_t . The conditional quantile is estimated by inverting a kernel estimate of the conditional distribution function, and we prove its asymptotic normality and uniform strong consistency. The performance of the estimate for light and heavy-tailed distributions of the innovations are evaluated by a simulation study. Finally, the technique is applied to estimate VaR of stocks in DAX, and its performance is compared with the existing standard methods using backtesting.

Keywords: conditional quantile, kernel estimate, quantile autoregression, time series, uniform consistency, value-at-risk

1 Introduction

We consider a stationary and α-mixing multivariate time series $\{V_t, t \in \mathbf{Z}\}$ adapted to the sequence $\mathbf{F}_t, -\infty < t < \infty$, of σ-algebras. Partition it as

*Hermann Otto Hirschfeld Junior Professor at the Ladislaus von Bortkiewicz chair of statistics and C.A.S.E., Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:wangwein@cms.hu-berlin.de

$V_t = (Y_t, \mathbf{X}_t)$ where the real-valued response variable $Y_t \in \mathbf{R}$ is $\mathbf{F}_t$-measurable and the covariate $\mathbf{X}_t \in \mathbf{R}^d$ is $\mathbf{F}_{t-1}$-measurable. For some $0 < \theta < 1$, we want to estimate the conditional θ-quantile of Y_t given the past $\mathbf{F}_{t-1}$ assuming that it is completely determined by $\mathbf{X}_t$, i.e. we have

$$Y_t = \mu_\theta(\mathbf{X}_t) + Z_t, \tag{1.1}$$

where the conditional θ-quantile of Z_t given $\mathbf{F}_{t-1}$ is 0. The quantile innovations Z_t are *not* assumed to be independent of $\mathbf{X}_t$. The conditional quantile function $\mu_\theta(\mathbf{x})$ may be rather arbitrary, apart from some regularity assumptions, and we want to estimate it nonparametrically. The model (1.1) includes the case of a nonparametric quantile regression where $(Z_t, \mathbf{X}_t)$, $-\infty < t < \infty$, are i.i.d., as well as the quantile autoregression (QAR) of order p

$$Y_t = \mu_\theta(Y_{t-1}, \ldots, Y_{t-p}) + Z_t,$$

where $\mathbf{X}_t = (Y_{t-1}, \ldots, Y_{t-p})$ is just part of the past of the univariate time series Y_t. If we choose $\mathbf{X}_t = (Y_{t-1}, \ldots, Y_{t-p}, \mathbf{U}_{t-1})$ where the random vector $\mathbf{U}_t$ consists of observations from other time series than Y_t available at time t, then (1.1) would become a quantile autoregressive model with exogeneous components. One main application, which we have in mind, is a flexible procedure for calculating the value-at-risk of a financial time series (compare, e.g., Jorion, 2000) which allows for including other information on the markets than just past data of the particular time series under consideration.

Considering other financial time series models, (1.1) can be seen, e.g., as a generalization of AR-ARCH-models, introduced in Weiss (1984), and their nonparametric generalizations reviewed by Härdle et al. (1997). For instance, consider a financial time series model of AR(p)-ARCH(p)-type

$$Y_t = \mu(\mathbf{X}_t) + \sigma(\mathbf{X}_t)\, e_t, \quad t = 1, 2, \ldots \tag{1.2}$$

where $\mathbf{X}_t = (Y_{t-1}, \ldots, Y_{t-p})$, μ and σ are arbitrary functions and $\{e_t\}$ is a sequence of iid random variables with mean 0 and variance 1. Then (1.2) can be written in the form (1.1) with $\mu_\theta(\mathbf{X}_t) = \mu(\mathbf{X}_t) + \sigma(\mathbf{X}_t) q_\theta^e$ and $Z_t = \sigma(\mathbf{X}_t)(e_t - q_\theta^e)$, where q_θ^e is the θ-quantile of e_t. The quantile innovations Z_t are independent of $\mathbf{X}_t$ only if $\{Y_t\}$ has a homoscedastic error, i.e. the volatility function $\sigma(\mathbf{x})$ is constant. For $p = 1$, (1.2) can be interpreted as a discrete-time version of the diffusion process $dY_t = m(Y_t)dt + \sigma(Y_t)dW_t$, where W denotes the standard Brownian motion, which includes the geometric Brownian motion as a stock price model in option pricing (for $\mu(x) = \mu x, \sigma(x) = \sigma x$) and the Vasicek model for interest rates.

The estimation of $\mu_\theta(\mathbf{X}_t)$ based on the model (1.2) usually involves the estimation of $\mu(\mathbf{X}_t)$ and $\sigma(\mathbf{X}_t)$ and the calculation of q^e_θ, for the latter assuming the distribution of the e_t to be known, using historical simulation procedures or a combination of both. Based on the more general model (1.1), we can, however, derive a more straightforward estimate, and, additionally, we do not have to assume the finiteness of the variance of Y_t which, for financial data, seems not always to be guaranteed. We get a nonparametric estimate of $\mu_\theta(\mathbf{x})$ directly by first estimating the conditional distribution function of Y_t given $\mathbf{X}_t$ and then inverting it. This type of estimate is related to local medians as considered by Truong and Stone (1992) and Boente and Fraiman (1995) - compare the discussion at the end of section 2. For estimating the conditional distribution, we use a kernel estimate of Nadaraya (1964) and Watson (1964) type. Apart from the disadvantages of not being adaptive and having some boundary effects, which can be fixed anyhow (see Hall et al., 1998), it has advantages of being a constrained estimator between 0 and 1 and a monotonically increasing function. This is an important property when deriving quantile function estimators by the inversion of a distribution estimator.

Our work is closely related to Cai (2002), and also the following up paper in Cai and Wang (2008), where they propose a conditional quantile estimation by inverting a weighted double kernel technique for serial dependent data and applied in estimating conditional VaR. However, our approach is much simpler than Cai (2002) and is much easier to be implemented as the calculations of weights are computationally demanding. Moreover the weights does not include our case as $1/n$ (only asymptotically) with an easy averaging. Theoretically, we have proved a stronger uniform consistency results for our estimation. Also the simulation performance is compared with Cai and Wang (2008), and we have shown that our method is competitive relative to Cai and Wang (2008).

Our paper consists of four sections, Section 2 we lay down the assumptions and state our major theoretically results. In Section 3, we illustrate the performance of the quantile function estimate with a small simulation study. The technical results and proofs are postponed to the appendix. Section 4 we demonstrate the empirical performance of our estimator by a VaR calculation for German stocks.

2 Asymptotic behaviour of quantile autoregressive kernel estimates

We consider kernel estimates of the quantile autoregressive function $\mu_\theta(\mathbf{x})$ based on a sample $(Y_t, \mathbf{X}_t), t = 1, \dots, n$, from the quantile autoregressive model (1.1). In a first step, we have to estimate the conditional distribution function $F_\mathbf{x}(y) = P(Y_t \le y \,|\, \mathbf{X}_t = \mathbf{x}) = E[\mathbf{I}_{t,y} \,|\, \mathbf{X}_t = \mathbf{x}]$ of Y_t given $\mathbf{X}_t = \mathbf{x}$, which can be written as the conditional expectation of $\mathbf{I}_{t,y} = \mathbf{I}_{\{Y_t \le y\}}$ and, therefore, may be estimated by the standard Nadaraya-Watson kernel estimate

$$\widehat{F}_\mathbf{x}(y) = \frac{\sum_{t=1}^n \mathbf{K}_h(\mathbf{x} - \mathbf{X}_t)\mathbf{I}_{t,y}}{\sum_{t=1}^n \mathbf{K}_h(\mathbf{x} - \mathbf{X}_t)}. \tag{2.1}$$

Here, $\mathbf{K}(\mathbf{u})$ is a d-dimensional kernel and $\mathbf{K}_h(\mathbf{u}) = h^{-d}\mathbf{K}(\mathbf{u}/h)$ is the rescaled kernel. For sake of simplicity, we assume that the bandwidth h is the same in all directions, but we could generalize our results in a straightforward manner to vectors $(h_1, \dots, h_d)^T$ of bandwidths.

For any $\theta \in (0, 1)$, the quantile autoregressive function $\mu_\theta(\mathbf{x})$ is given by

$$\mu_\theta(\mathbf{x}) = \inf\{y \in \mathbf{R} | F_\mathbf{x}(y) \ge \theta\}.$$

Therefore, we propose to estimate $\mu_\theta(\mathbf{x})$ by the following kernel estimate

$$\widehat{\mu}_\theta(\mathbf{x}) = \inf\{y \in \mathbf{R} | \widehat{F}_\mathbf{x}(y) \ge \theta\} \equiv \widehat{F}_\mathbf{x}^{-1}(\theta), \tag{2.2}$$

where $\widehat{F}_\mathbf{x}^{-1}(\theta)$ denotes the usual generalized inverse of the distribution function $\widehat{F}_\mathbf{x}(y)$ which is a pure jump function of y.

For our asymptotic considerations, we have to assume that the time series $(Y_t, \mathbf{X}_t)$ satisfies appropriate mixing conditions. There are a number of mixing conditions discussed, e.g., in the monographs of Doukhan (1994) and Bosq (1996). Among them α- or strong mixing is a reasonably weak one known to be fulfilled for many time series models. In particular, Masry (1995,1997) has demonstrated that under some mild conditions, both ARCH processes and nonlinear additive autoregressive models with exogeneous variables are stationary and α-mixing. Thus, choosing $\mathbf{X}_t = (Y_{t-1}, \dots, Y_{t-p})^T$ in (1.2) and assuming the time series Y_t to be α-mixing would be an example of a quantile autoregressive process (1.1) for which $(Y_t, \mathbf{X}_t)$ and $\mathbf{I}_{t,y}$ in (2.1) are α-mixing as well.

The following set of assumptions are required for proving asymptotic normality of $\widehat{\mu}_\theta(\mathbf{x})$. Here and in the following, $g(\mathbf{x})$ denotes the stationary probability density of $\mathbf{X}_t$.

(A1) The kernel $\mathbf{K} : \mathbf{R}^d \to \mathbf{R}$ is a nonnegative, Lipschitz continuous function, satisfying $|\mathbf{K}(\mathbf{u})| \le K_\infty$ for all $\mathbf{u}$, $\int \mathbf{K}(\mathbf{u})d\mathbf{u} = 1, \int \mathbf{u}\mathbf{K}(\mathbf{u})d\mathbf{u} = 0$ and $\int \|\mathbf{u}\|^2 \mathbf{K}(\mathbf{u})d\mathbf{u} < \infty$.
(A2) For all $y, \mathbf{x}$ satisfying $0 < F_\mathbf{x}(y) < 1, g(\mathbf{x}) > 0$
(i) $F_\mathbf{x}(y)$ and $g(\mathbf{x})$ are continuous and bounded in $y, \mathbf{x}$,
(ii) $g(\mathbf{x})$ is twice continuously differentiable, and, for fixed y, $F_\mathbf{x}(y)$ is twice continuously differentiable with respect to $\mathbf{x}$, where the derivatives are continuous functions of y and the second derivatives are Hölder-continuous in $\mathbf{x}$, i.e. for some $c, \beta > 0$ and all $\mathbf{x}, \mathbf{x}', y$

$$\Big|\frac{\partial^2}{\partial x_i x_j} F_\mathbf{x}(y) - \frac{\partial^2}{\partial x_i x_j} F_{\mathbf{x}'}(y)\Big| \le c \left\| \mathbf{x} - \mathbf{x}' \right\|^\beta, \quad i, j = 1, \dots, d,$$

and analogously for $g(\mathbf{x})$.
(iii) for fixed $\mathbf{x}$, $F_\mathbf{x}(y)$ has the conditional density $f_\mathbf{x}(y)$ which is continuous in $\mathbf{x}$ and Hölder-continuous in y: $|f_\mathbf{x}(y) - f_\mathbf{x}(y')| \le c|y - y'|^\beta$ for some $c, \beta > 0$.
(iv) $f_\mathbf{x}(\mu_\theta(\mathbf{x})) > 0$ for all $\mathbf{x}$.
(A3) The process $\{(Y_t, \mathbf{X}_t)\}$ is stationary and α-mixing with mixing coefficients satisfying $\alpha(s) = O(s^{-(2+\delta)})$, for some $\delta > 0$.

Theorem 2.1. *Assume that (A1)-(A3) hold. As $n \to \infty$, let the sequence of bandwidths $h > 0$ converge to 0 such that $nh^d \to \infty$. Then, the conditional quantile estimate is consistent, $\widehat{\mu}_\theta(\mathbf{x}) \to^p \mu_\theta(\mathbf{x})$, and asymptotically unbiased*

$$E\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x}) = h^2 B_\mu(\mu_\theta(\mathbf{x})) + o(h^2) \quad \text{where} \quad B_\mu(y) = -\frac{B(y)}{f_\mathbf{x}(y)}. \qquad (2.3)$$

If, additionally, the bandwidths are chosen such that nh^{d+4} is either 1 or converges to 0, $\widehat{\mu}_\theta(\mathbf{x})$ is asymptotically normal,

$$\sqrt{nh^d}\big(\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x}) - h^2 B_\mu(\mu_\theta(\mathbf{x}))\big) \to^D N\left(0, \frac{V^2(\mu_\theta(\mathbf{x}))}{f_\mathbf{x}^2(\mu_\theta(\mathbf{x}))}\right), \qquad (2.4)$$

Here, $B(y)$ and $V^2(y)$ are defined in the bias and variance expansion for the conditional distribution estimator in Lemma B.1 in the appendix.

As a step towards uniform consistency of the quantile autoregressive estimate, we first need uniform consistency of the Nadaraya-Watson kernel estimate

$\widehat{F}_{\mathbf{x}}(y)$ for the conditional distribution function. There are various versions of this well-known result, e.g. Theorem 5.4.2 of Abberger (1996), depending on the chosen set of assumptions. We impose the following conditions.

(B1) For some compact set G and some $\gamma > 0$, $g(\mathbf{x}) \geq \gamma$ for all $\mathbf{x} \in G$.
(B2) $(Y_t, \mathbf{X}_t)$ is stationary and α-mixing with mixing coefficients $\alpha(n), n \geq 1$, and there is an increasing sequence $s_n, n \geq 1$, of positive integers such that for some finite A

$$\frac{n}{s_n}\,\alpha^{2s_n/(3n)}(s_n) \leq A, \quad 1 \leq s_n \leq \frac{n}{2} \quad \text{for all} \quad n \geq 1. \tag{2.5}$$

Using these assumptions in addition to (A1), (A2) and remarking that $|\mathbf{I}_{t,y}| \leq 1$, we immediately get the following version of the uniform consistency as a special case of Theorem 3.3.5 of Györfi et al. (1989).

Theorem 2.2. *Assume (A1), (A2), (B1) and (B2). If, as $n \to \infty$, the bandwidth $h \to 0$ such that $\widetilde{S}_n = nh^d(s_n \log n)^{-1} \to \infty$, then $\widehat{F}_{\mathbf{x}}(y)$ is uniformly consistent on G in the strong sense*

$$\sup_{\mathbf{x}\in G} |\widehat{F}_{\mathbf{x}}(y) - F_{\mathbf{x}}(y)| \to 0 \quad a.s. \tag{2.6}$$

We remark that from the proof of Theorem 2.3 below we also get a rate of convergence of $\widehat{F}_{\mathbf{x}}(y)$. We also need uniform consistency of the Rosenblatt-Parzen kernel estimate for the density $g(\mathbf{x})$ of $\mathbf{X}_t$ given by

$$\widehat{g}(\mathbf{x}) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{K}_h(\mathbf{x} - \mathbf{X}_t). \tag{2.7}$$

The following lemma, which gives the uniform rate of convergence of $\widehat{g}(\mathbf{x})$ on the compact set G, follows immediately from the proof of Theorem 3.3.6 of Györfi et al. (1989).

Lemma 2.1. *Under the assumptions of Theorem 2.2*
(i) $\sup_{\mathbf{x}\in G} |\widehat{g}(\mathbf{x}) - E\widehat{g}(\mathbf{x})| = O(\widetilde{S}_n^{-\frac{1}{2}}) \quad a.s.$
(ii) $\sup_{\mathbf{x}\in G} |E\widehat{g}(\mathbf{x}) - g(\mathbf{x})| = O(h^2)$.

To show uniform convergence of the quantile estimator $\widehat{\mu}_\theta(\mathbf{x})$, we interpret the kernel estimate of the quantile autoregression in (2.2) in the light of concepts of M-estimation as in Huber (1981). As we assume that $F_{\mathbf{x}}(y)$

is absolutely continuous, we automatically have $F_{\mathbf{x}}(\mu_\theta(\mathbf{x})) = \theta$, and the conditional quantile function $\mu_\theta(\mathbf{x}), \mathbf{x} \in \mathbf{R}^d$, can be seen as a zero in the argument μ of the following function

$$\begin{aligned}\widetilde{H}(\mathbf{x},\mu) &= F_{\mathbf{x}}(\mu) - \theta = E[\mathbf{I}_{t,\mu} - \theta|\mathbf{X}_t = \mathbf{x}] \\ &= E[\Psi(Y_t - \mu)|\mathbf{X}_t = \mathbf{x}] \qquad (2.8)\end{aligned}$$

with $\Psi(u) = \mathbf{I}_{(-\infty,0]}(u) - \theta$. We define the estimator of $\widetilde{H}(\mathbf{x},\mu)$ as

$$\begin{aligned}\widetilde{H}_n(\mathbf{x},\mu) &= \widehat{F}_{\mathbf{x}}(\mu) - \theta = \frac{\sum_{t=1}^n \mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)(\mathbf{I}_{t,\mu}-\theta)}{\sum_{t=1}^n \mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)} \\ &= \frac{\sum_{t=1}^n \mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)\Psi(Y_t-\mu)}{\sum_{t=1}^n \mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)}. \qquad (2.9)\end{aligned}$$

Now, $\widehat{\mu}_\theta(\mathbf{x})$ is not necessarily a zero of $\widetilde{H}_n(\mathbf{x},\mu)$, as the latter is a pure jump function in μ, but we have at least in good approximation $\widetilde{H}_n(\mathbf{x},\widehat{\mu}_\theta(\mathbf{x})) \approx 0$. More precisely, the height of the jumps of $\widehat{F}_{\mathbf{x}}(\mu)$ and, therefore, of $\widetilde{H}_n(\mathbf{x},\mu)$ are $\frac{1}{n}\mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)/\widehat{g}(\mathbf{x})$ where $\widehat{g}(\mathbf{x})$ is the kernel estimate of $g(\mathbf{x})$ given by (2.7). By Lemma 2.1 and assumption (B1), $\widehat{g}(\mathbf{x}) \to g(\mathbf{x}) \geq \gamma > 0$ uniformly in $\mathbf{x} \in G$, and using additionally the boundedness of the kernel $\mathbf{K}$ by K_∞, the jumps of $\widehat{F}_{\mathbf{x}}(\mu)$ are bounded by $c_\gamma/(nh^d)$ with $c_\gamma = 2K_\infty/\gamma$ uniformly for $\mathbf{x} \in G$ for large enough n. The definition (2.2) of $\widehat{\mu}_\theta(\mathbf{x})$ immediately implies

$$0 \leq \widetilde{H}_n(\mathbf{x},\widehat{\mu}_\theta(\mathbf{x})) \leq \frac{c_\gamma}{nh^d} \quad \text{uniformly in } \mathbf{x} \in G. \qquad (2.10)$$

In addition to the assumptions (A2) on the conditional density, we need the following set of conditions for proving uniform convergence of $\widehat{\mu}_\theta(\mathbf{x})$.

(C1) The conditional density $f_{\mathbf{x}}(\mu)$ is uniformly bounded in $\mathbf{x}$ and μ by, say, c_f.
(C2) For the compact set G of (B1) and some compact neighborhood Θ_0 of 0, the set $\Theta = \{\nu = \mu_\theta(\mathbf{x}) + \mu;\ \mathbf{x} \in G, \mu \in \Theta_0\}$ is compact too, and for some constant $c_0 > 0$, $f_{\mathbf{x}}(\nu) \geq c_0$ for all $\mathbf{x} \in G, \nu \in \Theta$.

Theorem 2.3. *Assume (A1), (A2), (B1), (B2), (C1) and (C2). Suppose $h \to 0$ is a sequence of bandwidths such that $\widetilde{S}_n = nh^d(s_n \log n)^{-1} \to \infty$ for some $s_n \to \infty$. Let $S_n = h^2 + \widetilde{S}_n^{-\frac{1}{2}}$. Then we have*

$$\sup_{\mathbf{x}\in G} |\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x})| = O(S_n) + O(\frac{1}{nh^d}) \quad a.s. \qquad (2.11)$$

Usually, S_n will be much larger than $(nh^d)^{-1}$, and the rate of (2.11) will be $O(S_n)$. This is the case, e.g., if bias and variance are balanced and the mean-square error is asymptotically minimized which, by Theorem 2.1, requires $(nh^d)^{-1}$ to be of the order of h^4. We remark that Abberger (1996 - Corollary 5.4.2) has shown the pointwise consistency $\widehat{\mu}_\theta(\mathbf{x}) \to \mu_\theta(\mathbf{x})$ a.s. for given $\mathbf{x}$, assuming that $\widehat{F}_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x})) = \theta$ which, however, rarely happens to be satisfied in the light of the discussion leading to (2.10).

We have shown asymptotic normality and uniform consistency of the nonparametric quantile function estimate where, up to the usually neglible term $O((nh^d)^{-1})$-term in (2.11), the rates are the same as for the Nadaraya-Watson estimate of the corresponding conditional distribution function. Related results have been derived by Truong and Stone (1992) and Boente and Fraiman (1995) who consider local medians. These estimates correspond to the special choice $K(\mathbf{u}) = \mathbf{I}_{\{||u|| \leq 1\}}$ as the rectangular kernel and $\theta = \frac{1}{2}$. Under assumptions, which are similar to ours, Truong and Stone show pointwise and uniform consistency of this conditional median estimate. Their rates of convergence of $|\widehat{\mu}_{1/2}(\mathbf{x}) - \mu_{1/2}(\mathbf{x})|$ are different, e.g. for $d = 1$ and pointwise convergence, they have an optimal rate $n^{-1/3}$ whereas we have $n^{-2/5}$. The reason for that difference is the discontinuity of the rectangular kernel which does not satisfy our condition (A1). Boente and Fraiman consider the local median as a robust estimate of the conditional mean $r(\mathbf{x}) = E\{Y_t|\mathbf{X}_t = \mathbf{x}\}$ and prove asymptotic normality of $\widehat{\mu}_{1/2}(\mathbf{x}) - r(\mathbf{x})$

The uniform convergence of the nonparametric quantile function estimate allows for a detailed investigation of the quantile innovations Z_t of the model (1.1) based on the sample residuals $\widehat{Z}_t = Y_t - \widehat{\mu}_\theta(\mathbf{X}_t)$ which is not restricted to the iid case. Koenker (1999) has, e.g., considered a first order quantile autoregression model for daily temperatures with non-iid innovations, and the well-known conditional heteroscedasticity of financial time series also suggests that there is some scope for such general models.

In particular in heavy-tailed situations, scale provides a more natural concepts of dispersion than variance, compare Bickel (1978) or the recent results of Hall and Yao (2002) about the conventional quasi-maximum likelihood estimator. We remark that the nonparametric quantile estimate can be also used directly as scale function estimate in purely heteroscedastic models like ARCH and their derivatives. In general models of the form $Y_t = \mu_\theta(\mathbf{X}_t) + s_\theta(\mathbf{X}_t)e_t$ with iid innovations e_t the residuals $\widehat{Z}_t$ from (1.1) may be used to estimate the scale function $s_\theta(\mathbf{x})$ by a similar type of estimate as $\mu_\theta(\mathbf{x})$, just as in common nonparametric AR-ARCH-models like (1.2) where the estimate of the conditional variance is closely related to the estimate of the conditional mean. Such nonparametric scale function estimators will be investigated in a subsequent paper (compare also Mwita, 2003).

For sake of simplicity, we have restricted ourselves to kernel estimates of the conditional distribution function as the basis for the quantile function estimates. Our results may be modified in a straightforward manner to cover also the more general local polynomial estimates (Fan and Gijbels, 1995). Another approach for estimating $\mu_\theta(\mathbf{x})$ would be a nonparametric version of the quantile regression estimate of Koenker and Bassett (1978) which we shall consider in a subsequent paper.

3 Monte Carlo study

We illustrate the performance of kernel estimates for the quantile autoregressive function with a Monte Carlo study. For that purpose, we generate a sample $Y_t, t = 1, \ldots, n$ of size $n = 1000$, of the nonlinear AR(1)-ARCH(1)-process (1.2) with

$$\mu(x) = a + bx + \frac{1}{\sqrt{2\pi}dx} \exp\left(\frac{(x-c)^2}{d^2}\right), \quad \sigma^2(x) = \omega + \alpha x^2$$

where $a = 0.4$, $b = 0.3$, $c = 1.657$, $d = 0.1175$, $\omega = 0.007$, $\alpha = 0.2$, and where the innovations $e(t)$ have a standardized normal, exponential, t_4- and t_2-distribution resp. We estimate the conditional quantile $\mu_\theta(x)$ of Y_t given $Y_{t-1} = x$ by the kernel estimate of (2.2) for $\theta = 0.90$. Of course, in that particular case, $\mu_\theta(x) = \mu(x) + \sigma(x) q_\theta^e$ where q_θ^e denotes the θ-quantile of the distribution of the innovations $e(t)$.

Figure 1 shows for each of the four innovation distributions a typical sample with the true conditional quantile function $\mu_\theta(x)$ (black line) together with the estimate $\widehat{\mu}_\theta(x)$ (grey line) and observations on the left-hand side and the same picture without observations the right-hand side. For the kernel estimate, we used the bisquare kernel $K(u) = c\max(1-u^2, 0)^2$ with normalization constant $c = 15/16$. The estimate performs quite well for light- and heavy-tailed innovations apart from the areas at the extreme right and left of the support of the sample where data are scarce. Here, the performance could have been improved a bit by adapting the bandwidth to the

Distribution	AAE	AAE(Cai)
N	0.1104(0.0036)	0.1107(0.0036)
exp	0.1254(0.0092)	0.1260(0.0092)
t_4	0.1660(0.0066)	0.1710(0.0069)
t_2	0.3200(0.0139)	0.3070(0.0139)

Table 1: Averaged absolute error (standard deviation), calculated from 1000 independent samples of the simulated process, with $p = 0.95$.

local density of the observations, but for sake of simplicity we have chosen fixed bandwidths independent of x for the simulation by a data driven cross validation approach.

Figure 2 illustrates the sampling variability of the estimates $\widehat{\mu}_\theta(x)$. It shows for each of the four innovation distributions the function $\mu_\theta(x)$ together with a pointwise 95%-confidence interval. As expected, the estimates are in particular reliable where the stationary density of the data is large. It comes as a positive surprise that the performance of the conditional quantile estimate does not depend strongly on the innovation distributions and is quite reasonable for asymmetric (exponential), heavy-tailed (t_4) and even infinite variance (t_2) innovations.

To demonstrate the finite sample performance of the proposed nonparametric estimator, we evaluate it in terms of Average absolute error (AAE). We generated 1000 random samples of size $n = 1000$ and for each independent sample we calculated AAEs, based on the difference between the estimation and the true conditional 0.95-quantile. In Table 1 we can see that for all four distributions the AAE value is small, which indicates that the proposed estimator has a small bias. It also suggests that our method is competitive compared to Cai (2002).

As already mentioned, it can be seen that the performance of the conditional quantile estimate does not depend strongly on the innovation distributions and it performs reasonably well for asymmetric (exponential), heavy-tailed (t_4) and infinite variance (t_2) innovations.

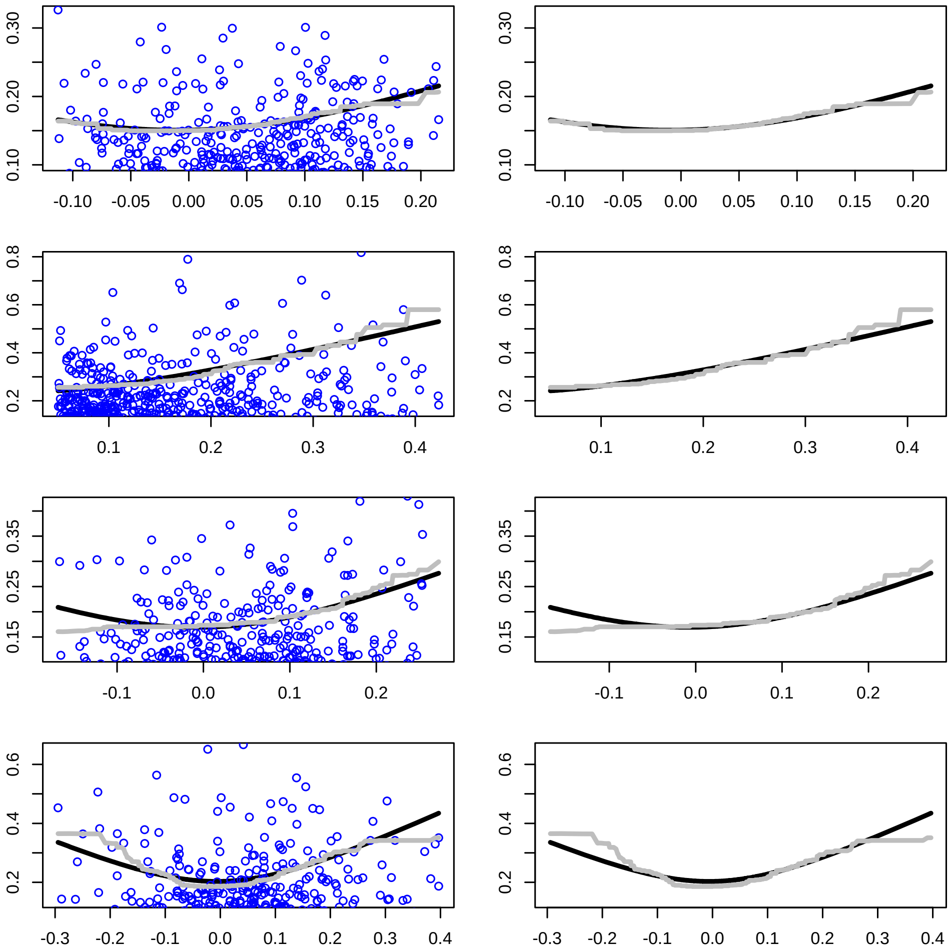
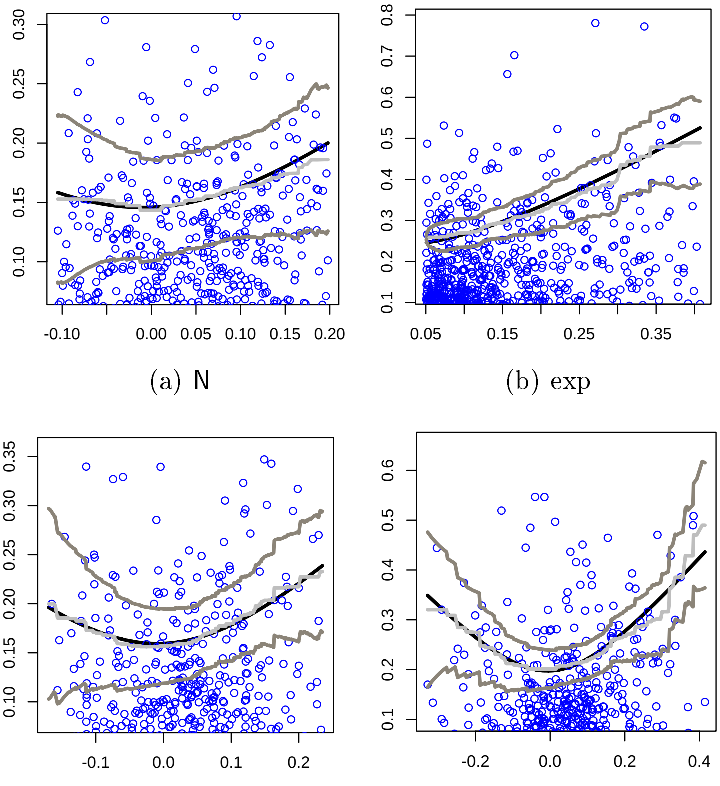


Figure 1: Simulated sample (left panel) with the true $\mu_\theta(\mathbf{X}_t)$ (black line) and $\hat{\mu}_\theta(\mathbf{X}_t)$ (grey line) (right panel) with different noise distributions, normal (N), exponential (exp) , t_2, t_4. (From up to below)



(a) N (b) exp

(c) t_2 (d) t_4

Figure 2: $\hat{\mu}_\theta(\mathbf{X}_t)$ (grey line) and its confidence intervals 95% (dark grey line) with different noise distributions

Stock	Mean	Std. Dev.	Skewness	Kurtosis	Min	Max
IBM	-0.0004	0.0149	-0.0375	5.5040	-0.1090	0.0866
HSBC	0.0002	0.0209	1.6152	37.1312	-0.1823	0.2764
Ford	-0.0001	0.0357	0.0975	12.1477	-0.2553	0.2897

Table 2: Summary statistics for daily returns. The period is from March 11, 2005 to February 10, 2011. The number of observations is 1512

4 Application

To see how the proposed nonparametric estimate for conditional quantiles of time series performs on a real data set, we will estimate the VaR of three different stocks and compare it with the CAViaR model (Engle and Manganelli, 2004) and the parametric linear quantile regression (Koenker and Basset, 1978). We examine the VaR forecasting performance for a portfolio that is short on IBM, HSBC and Ford. In this case, the holder of the portfolio suffers a loss when the value of the asset increases.

4.1 Data description

As an illustration, we have chosen three historical time series of returns for three stocks. The data set consists of 1512 daily adjusted closing prices from Yahoo Finance for the following stocks: IBM Corporation (component of S&P 500), HSBC Holding (component of FTSE 100 Index) and Ford Motor Company (component of S&P 500). The covered period is from March 1, 2005 to March 1, 2011. We computed the daily returns as the difference of the log of prices

$$R_t = \ln(P_t) - \ln(P_{t-1}) \tag{4.1}$$

Table 2 presents some relevant summary statistics for the calculated log returns of the chosen financial assets. It can be seen that IBM has negative skewness, while HSBC and Ford show positive skewness and across all three samples an excess kurtosis can be observed. Figure 3 shows the log returns of IBM. Therefore, the returns exhibit the typical behavior of financial time series: asymmetry in the data, violation of normality, which motivates nonparametric estimation of VaR.

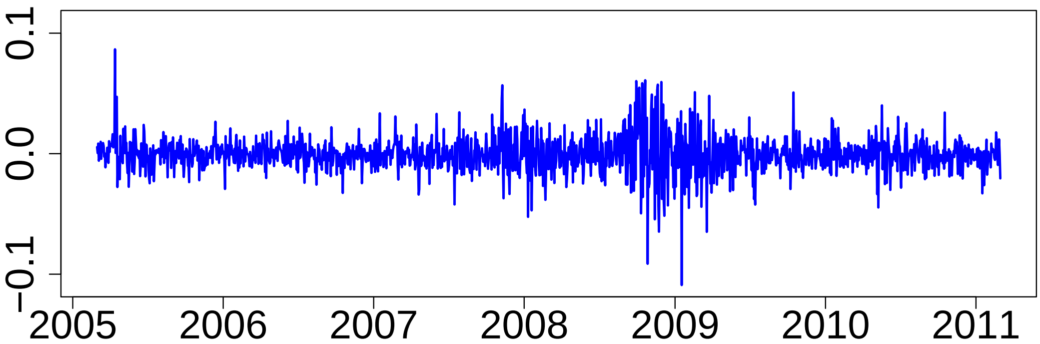


Figure 3: IBM log returns. The period is from March 1, 2005 to March 1, 2011. The number of observations is 1512

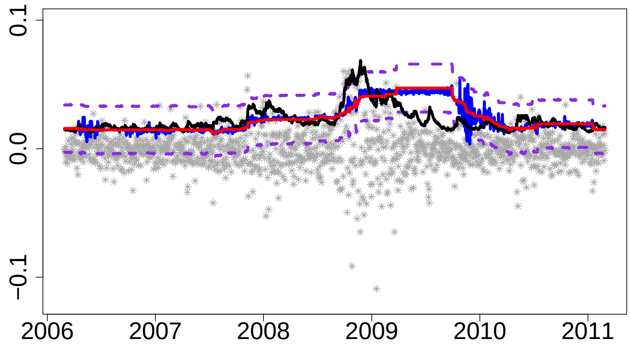
4.2 VaR estimation

In this section, we compare the performance of our method with other alternatives by applying all of them to VaR estimation. Three methods are used for estimating conditional quantiles, they have been implemented for each of the stock: 1, The estimation by our method ($\widehat{VaR}_{IBM}$); 2, The CaViaR model proposed by Engle and Manganelli (2004) ($\widehat{CAViaR}_{IBM}$); 3, The linear quantile regression technique proposed by Koenker and Basset (1978) ($\widehat{RQ\_VaR}_{IBM}$). We use a moving window of $N = 252$ (corresponding to approximately two years of trading data), which allows us to get an update for the estimator for each moving window with an increment of one trading day.

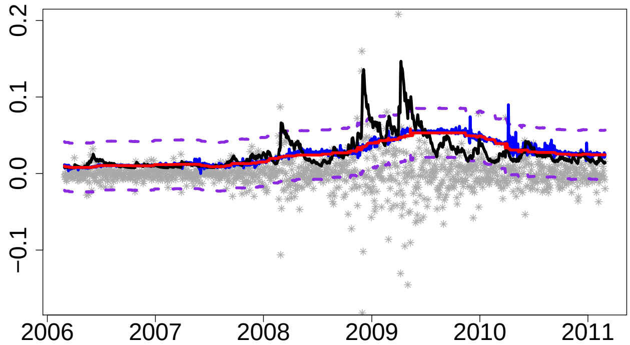
Figure 4 shows the forecasted 5% VaR sequence, estimated with the three techniques for IBM, HSBC and Ford. It can be seen that compared to CAViaR and linear quantile regression, the nonparametric VaR is much smoother, even for extreme values.

To check the accuracy of our estimator, we also constructed the 95% confidence interval. For all three stoks, the estimator lies inside the confidence interval.

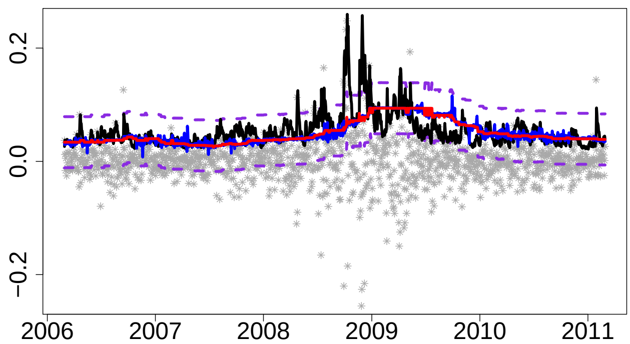
Table 3 shows the summary statistics of the 5% VaR estimates. Among three sequence of estimators, Ford has the highest mean and highest standard deviation, while IBM has the lowest mean and standard deviation, the CaViaR estimates have the highest maximum value, while the parametric quantile regression have the lowest minimum value, as compared to the other two implemented models.



(a) IBM



(b) HSBC



(c) Ford

Figure 4: The stars are daily returns, the blue line is the linear quantile regression, the black line is the $\widehat{CAViaR}$ and the red line shows the nonparametric estimate for conditional quantile $\widehat{VaR}$, with h=0.5 The violet dashed line is the 95% confidence interval. VaRInvq

Measure	Bandwidth	Mean	Std.Dev.	Min	Max
$\widehat{VaR}_{IBM}$	h=0.5	2.36	1.10	1.18	4.72
$\widehat{CAViaR}_{IBM}$		2.16	0.91	1.23	6.86
$\widehat{RQ\_VaR}_{IBM}$		2.35	1.10	0.36	5.30
$\widehat{VaR}_{HSBC}$	h=0.4	2.60	1.49	0.77	5.84
$\widehat{CAViaR}_{HSBC}$		2.38	2.10	0.74	14.70
$\widehat{RQ\_VaR}_{HSBC}$		2.70	1.58	0.01	8.99
$\widehat{VaR}_{Ford}$	h=0.3	4.83	1.83	2.53	14.17
$\widehat{CAViaR}_{Ford}$		5.03	2.77	2.11	25.96
$\widehat{RQ\_VaR}_{Ford}$		4.94	2.07	0.77	11.55

Table 3: VaR 5% summary statistics. The period is from March 1, 2005 to March 1, 2011. The numbers in the table are scaled up by 10^2

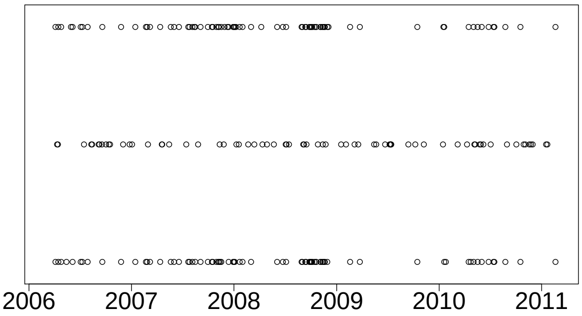
4.3 Forecast performance evaluated with backtesting

To evaluate the forecast performance of the proposed nonparametric estimator for conditional quantiles, we use an backesting procedure, namely, the CAViaR test in B.1 .

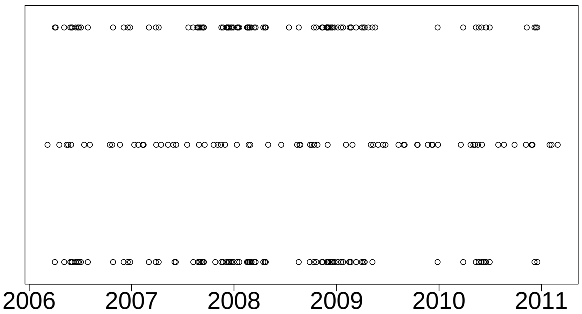
We calculate the violation sequence (as defined in Section B.1) for each stock. The estimated values of the VaR are compared with the actual returns, a violation occuring for each observation larger than the VaR estimate. Because we are interested in evaluating the forecast performance, each time we compare the ex post return to the VaR estimate. The violations are calculated using moving windows, with a window size of 252 days.

The upper panel of Figure 5 shows the timings of the violations $t : \hat{\mathbf{I}}_t = 1$ of $\widehat{VaR}_{IBM}$ (80 violations), $\widehat{CAViaR}_{IBM}$ (81 violations) and $\widehat{RQ\_VaR}_{IBM}$ (77 violations). The middle panel of Figure 5 shows the violations of $\widehat{VaR}_{HSBC}$ (89 violations), $\widehat{CAViaR}_{HSBC}$ (78 violations) and $\widehat{RQ\_VaR}_{HSBC}$ (84 violations). The lowest panel of Figure 5 depicts the violations of $\widehat{VaR}_{Ford}$ (77 violations), $\widehat{CAViaR}_{Ford}$ (81 violations) and $\widehat{RQ\_VaR}_{Ford}$ (78 violations).

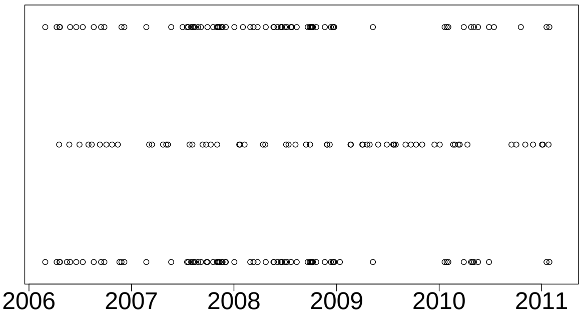
The backtesting procedure is performed separately for each sequence of $\hat{\mathbf{I}}_t$. The null hypothesis is that ideally each sequence $\hat{\mathbf{I}}_t$ forms a series of martingale difference. The out of sample CAViaR test has been applied. The results of the test are shown in Table 4. The highest p-values have been obtained by $\widehat{VaR}_{IBM}$, $\widehat{VaR}_{HSBC}$ and $\widehat{VaR}_{Ford}$. The best result is obtained



(a) IBM



(b) HSBC



(c) Ford

Figure 5: The timings of violations. The top circles are for $\widehat{VaR}$, the middle ones are for $\widehat{CAViaR}$ and the bottom ones are for $\widehat{RQ\_VaR}$. VaRInvq

for $\widehat{VaR}_{IBM}$. The $\widehat{CAViaR}_{Ford}$ and $\widehat{RQ\_VaR}_{Ford}$ are rejected at 5% and 1% significance level, respectively, by the CAViaR test. This indicates that overall, the nonparametric VaR performs better than CAViaR and parametric quantile regression.

Measure	Bandwidth	CAViaR test
$\widehat{VaR}_{IBM}$	$h = 0.5$	0.2147
$\widehat{CAViaR}_{IBM}$		0.1139
$\widehat{RQ\_VaR}_{IBM}$		0.1529
$\widehat{VaR}_{HSBC}$	$h = 0.4$	0.1572
$\widehat{CAViaR}_{HSBC}$		0.0865
$\widehat{RQ\_VaR}_{HSBC}$		0.0511
$\widehat{VaR}_{Ford}$	$h = 0.3$	0.0770
$\widehat{CAViaR}_{Ford}$		$0.0234*$
$\widehat{RQ\_VaR}_{Ford}$		$0.0010 **$

*, ** denotes significance at 5 and 1 percent level, respectively

Table 4: VaR, CAViaR and quantile regression estimates backtesting θ-values, obtained with CAViaR test

B Appendix

B.1 Backtesting

An backtesting procedure is used for assessing the accuracy and forecast performance of the VaR models, so that risk managers of financial institutions can use it in the decision-making process. More precisely, the quality of the forecast estimator is evaluated by comparing the actual observations to those estimated with the VaR model.

We follow the framework proposed by Berkowitz et al. (2009), which is designed for evaluating the accuracy of out-of-sample interval forecasts. The proposed procedure evaluates the VaR forecast by viewing them as one-sided interval forecasts. Each time the ex post return is lower than the VaR, a violation occurs. Formally, the violation time series can be defined as

$$\mathbf{I}_{t+1} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } R_{t+1} < \widehat{VaR_t^\theta}, \\ 0 & \text{otherwise.} \end{cases} \tag{B.1}$$

Berkowitz et al. (2009) note that if the VaR is estimated correctly, the probability that the VaR will be exceeded should be unpredictable, after using all past information.

The tests proposed by Berkowitz et al (2009) consider that the sequence of violations form a martingale difference, which means that the expectation of the violation at $t+1$, given the information set up to time t is zero. This property implies that the current violation is uncorrelated with any past variables. One of the ways they propose for testing the uncorrelatedness is by considering the CaViaR test of Engle and Manganelli (2004):

$$\mathbf{I}_t = \alpha + \beta_1 \mathbf{I}_{t-1} + \beta_2 VaR_t + u_t \tag{B.2}$$

Here, the error term u_t follows a Logistic distribution. Estimating the logit model, the coefficients $(\widehat{\beta}_1, \widehat{\beta}_2)^T$ are obtained. For testing the null hypothesis $\widehat{\beta}_1 = \widehat{\beta}_2 = 0$ the Wald's test is used.

Besides assesing the quality of the estimator, according to Lopez (1999), the backtesting technique can serve in establishing the required level of capital for market risk by including a multiplier based on the unconditional number of VaR violations.

B.2 Proof of Theorems

The following lemma gives the asymptotic bias and variance for $\widehat{F}_{\mathbf{x}}(y)$ which is a Nadaraya-Watson kernel estimate for the conditional expectation $F_{\mathbf{x}}(y)$ of $\mathbf{I}_{t,y}$ given $X_t = \mathbf{x}$. Therefore, we omit the proof of the lemma which follows standard lines of arguments. Details can be found in Mwita (2003). Under slightly different conditions, Abberger (1996) has also derived such a result.

Lemma B.1. *Suppose (A1)-(A3) hold. Then*

$$E[\widehat{F}_{\mathbf{x}}(y) - F_{\mathbf{x}}(y)] = h^2 B(y) + o(h^2) \tag{B.3}$$

$$var[\widehat{F}_{\mathbf{x}}(y)] = (nh^d)^{-1} V^2(y) + o((nh^d)^{-1}) \tag{B.4}$$

where

$$\begin{aligned} B(y) &= \frac{1}{g(\mathbf{x})}\nabla F_{\mathbf{x}}(y)^T \int \mathbf{u}\nabla g(\mathbf{x})^T \mathbf{u}\mathbf{K}(\mathbf{u})d\mathbf{u} + \frac{1}{2}\int \mathbf{u}^T \nabla^2 F_{\mathbf{x}}(y)\mathbf{u}\mathbf{K}(\mathbf{u})d\mathbf{u} \\ V^2(y) &= \frac{1}{g(\mathbf{x})}(F_{\mathbf{x}}(y) - F_{\mathbf{x}}^2(y))\int \mathbf{K}^2(\mathbf{u})d\mathbf{u}. \end{aligned}$$

The following lemma follows immediately from Lemma B.1, using the smoothness assumptions on $F_{\mathbf{x}}(y)$, and a Taylor expansion of $F_{\mathbf{x}}(y)$ around y.

Lemma B.2. *Suppose (A1)-(A3) hold. Then, for any $\delta_n \to 0$, we have*

$$\widehat{F}_{\mathbf{x}}(y+\delta_n) - \widehat{F}_{\mathbf{x}}(y) = \delta_n f_{\mathbf{x}}(y) + o_p(\delta_n) + o_p(h^2) + o_p((nh^d)^{-1/2}) \tag{B.5}$$

Proof of Theorem 2.1 :
First we prove consistency. By Lemma B.1, $\widehat{F}_{\mathbf{x}}(y) \to F_{\mathbf{x}}(y)$ in mean-square and, hence, in probability for all $\mathbf{x} \in \mathbf{R}^d$ and y. The Glivenko-Cantelli theorem in Krishnaiah (1990) for strongly mixing sequences implies

$$\sup_{y\in\mathbf{R}} |\widehat{F}_{\mathbf{x}}(y) - F_{\mathbf{x}}(y)| \to 0 \quad \text{in probability.} \tag{B.6}$$

By the uniqueness assumption (A2 iv) on $\mu_\theta(\mathbf{x})$, for any fixed $\mathbf{x} \in \mathbf{R}^d$, there exists an $\epsilon > 0$ such that

$$\delta = \delta(\epsilon) = \min\{\theta - F_{\mathbf{x}}(\mu_\theta(\mathbf{x}) - \epsilon), F_{\mathbf{x}}(\mu_\theta(\mathbf{x}) + \epsilon) - \theta\} > 0.$$

This implies, using the monotonicity of $F_{\mathbf{x}}$, that

$$\begin{aligned} P\{|\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x})| > \epsilon\} &\leq P\{|F_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x})) - F_{\mathbf{x}}(\mu_\theta(\mathbf{x}))| > \delta\} \\ &\leq P\{|F_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x})) - \widehat{F}_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x}))| > \delta - \frac{c_\gamma}{nh^d}\} \\ &\leq P\{\sup_y |\widehat{F}_{\mathbf{x}}(y) - F_{\mathbf{x}}(y)| > \delta'\} \end{aligned} \tag{B.7}$$

for arbitrary $\delta' < \delta$ and n large enough. Here, we have used $F_{\mathbf{x}}(\mu_\theta(\mathbf{x})) = \theta$ and $\theta \leq \widehat{F}_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x})) \leq \theta + c_\gamma/(nh^d)$ which follows from (2.10). Now, (B.7) tends to zero by (B.6). Hence the consistency follows.

To prove (2.4), let $b = -B(\mu_\theta(\mathbf{x}))f_{\mathbf{x}}^{-1}(\mu_\theta(\mathbf{x}))$ and $v = V(\mu_\theta(\mathbf{x}))f_{\mathbf{x}}^{-1}(\mu_\theta(\mathbf{x}))$. Let

$$\begin{aligned} q_n(z) &= P(\sqrt{nh^d}\,\frac{\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x}) - h^2 b}{v} \leq z) \\ &= P(\widehat{\mu}_\theta(\mathbf{x}) \leq \mu_\theta(\mathbf{x}) + h^2 b + (nh^d)^{-1/2} v z) \end{aligned}$$

As $\widehat{F}_{\mathbf{x}}(y)$ is increasing, but not necessarily strictly, we have

$$\begin{aligned} & P\big(\widehat{F}_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x})) < \widehat{F}_{\mathbf{x}}(\mu_\theta(\mathbf{x}) + h^2 b + (nh^d)^{-1/2} vz)\big) \\ \leq\ & q_n(z) \\ \leq\ & P\big(\widehat{F}_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x})) \leq \widehat{F}_{\mathbf{x}}(\mu_\theta(\mathbf{x}) + h^2 b + (nh^d)^{-1/2} vz)\big) \end{aligned}$$

By the same argument as in (B.7), we may replace $\widehat{F}_{\mathbf{x}}(\widehat{\mu}_\theta(\mathbf{x}))$ by $F_{\mathbf{x}}(\mu_\theta(\mathbf{x}))$ up to an error of $(nh^d)^{-1}$ at most, and we get, neglecting the $(nh^d)^{-1}$-term which is asymptotically negligible anyhow,

$$\begin{aligned} q_n(z) &\sim P(F_{\mathbf{x}}(\mu_\theta(\mathbf{x})) \leq \widehat{F}_{\mathbf{x}}(\mu_\theta(\mathbf{x}) + h^2 b + (nh^d)^{-1/2} vz) \\ &\sim P(-\delta_n f_{\mathbf{x}}(\mu_\theta(\mathbf{x})) \leq \widehat{F}_{\mathbf{x}}(\mu_\theta(\mathbf{x})) - F_{\mathbf{x}}(\mu_\theta(\mathbf{x}))) \end{aligned} \tag{B.8}$$

with $\delta_n = h^2 b + (nh^d)^{-1/2} vz$. Here, we have used Lemma B.2 and neglected the terms of order $o(\delta_n)$, $o(h^2)$ and $o((nh^d)^{-1/2})$ which are small compared to δ_n. Horvath and Yandell (1988) have shown that the conditional distribution estimator $\widehat{F}_{\mathbf{x}}(y)$ is asymptotically normal with asymptotic bias and variance given by Lemma B.1. This follows also under similar conditions from a functional central limit theorem for $\widehat{F}_{\mathbf{x}}(y)$ of Abberger (1996 - Corollary 5.4.1 and Lemma 5.4.1). Therefore, with $y_\theta = \mu_\theta(\mathbf{x})$, we get

$$\begin{aligned} q_n(z) &\sim P\Big(\sqrt{nh^d}\, \frac{\widehat{F}_{\mathbf{x}}(y_\theta) - F_{\mathbf{x}}(y_\theta) - h^2 B(y_\theta)}{V(y_\theta)} \geq \sqrt{nh^d}\, \frac{-f_{\mathbf{x}}(y_\theta)\delta_n - h^2 B(y_\theta)}{V(y_\theta)}\Big) \\ &\sim \Phi\Big(\sqrt{nh^d}\, \frac{f_{\mathbf{x}}(y_\theta).(h^2 b + (nh^d)^{-1/2} vz) + h^2 B(y_\theta)}{V(y_\theta)}\Big) \\ &= \Phi(z) \end{aligned}$$

by our choice of b and v and our condition on the rate of h. This proves the theorem. $\square$

The proof of Theorem 2.3 is close in lines with the proof in Collomb and Härdle (1986) and Györfi et al. (1989) chapter III. Note that it is difficult to deal with $\widetilde{H}_n(\mathbf{x}, \mu)$ directly, so we decompose the difference in the following manner. Let

$$H_n(\mathbf{x}, \mu) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{K}_h(\mathbf{x} - \mathbf{X}_t) \Psi(Y_t - \mu) \tag{B.9}$$

and let $H(\mathbf{x}, \mu) = \widetilde{H}(\mathbf{x}, \mu) g(\mathbf{x})$. Then the difference $\widetilde{H}_n - \widetilde{H}$ can be expressed as

$$\widetilde{H}_n - \widetilde{H} = \frac{H_n - H}{\widehat{g}} + \frac{H(g - \widehat{g})}{g\widehat{g}}$$

where, as before, $\widehat{g}(\mathbf{x})$ denotes the kernel density estimate of $g(\mathbf{x})$. Observe that

$$\begin{aligned}\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta}|\widetilde{H}_n-\widetilde{H}| &\le \sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta}\frac{|H_n-H|}{\widehat{g}}+\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta}\frac{|H|}{g\widehat{g}}|\widehat{g}-g| \\ &\le \sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta}\frac{|H_n-H|}{\widehat{g}}+\frac{\Gamma}{\gamma}\sup_{\mathbf{x}\in G}\frac{|\widehat{g}-g|}{\widehat{g}}, \qquad \text{(B.10)}\end{aligned}$$

as $|\widetilde{H}|\le 1$, g is bounded by, say, Γ, and $g\ge\gamma>0$ on G. Now, if $\sup_{\mathbf{x}\in G}|\widehat{g}-g|\le\epsilon$, we have

$$\frac{1}{\widehat{g}}=\frac{1}{g+(\widehat{g}-g)}\le\frac{1}{g-|\widehat{g}-g|}\le\frac{1}{\gamma-\epsilon}$$

on G. Therefore, to prove that $\widetilde{H}_n\to\widetilde{H}$ uniformly in $\mathbf{x}\in G,\mu\in\Theta$, it suffices to show that $H_n\to H$ and $\widehat{g}\to g$ uniformly in $\mathbf{x}\in G,\mu\in\Theta$, and the rate of convergence will be given by the slower of the two rates of convergence of $\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta}|H_n-H|$ and $\sup_{\mathbf{x}\in G}|\widehat{g}-g|$. We know the latter already from combining (i) and (ii) of Lemma 2.1. Therefore, we only have to investigate the convergence of $H_n(\mathbf{x},\mu)$.

Lemma B.3. *Under assumptions (A1), (A2), (B1), (B2) and (C1) we have for the compact set $G\subset\mathbf{R}^d$ of (B1) and for any compact $\Theta\subset\mathbf{R}$,*

$$\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta}|H_n(\mathbf{x},\mu)-EH_n(\mathbf{x},\mu)|=O(\widetilde{S}_n^{-\frac{1}{2}})\quad a.s.$$

with $\widetilde{S}_n=nh^d(s_n\log n)^{-1}\to\infty,\quad s_n\to\infty.$

Proof : We follow essentially the proof of Theorem 5.2.6 of Györfi et al. (1989) which gives a rate for the Glivenko-Cantelli theorem in the case of α-mixing random variables. Denote

$$\Delta_t=\frac{1}{n}\big(\mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)\Psi(Y_t-\mu)-E[\mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)\Psi(Y_t-\mu)]\big),$$

then $H_n(\mathbf{x},\mu)-EH_n(\mathbf{x},\mu)=\sum_{t=1}^n\Delta_t$ and $E\Delta_t=0$. As $|\Psi|$ is bounded by 1 and $\mathbf{K}$ is bounded by K_∞, we have

$$|\Delta_t|\le(nh^d)^{-1}2K_\infty<\infty.$$

We also have, using additionally the boundedness of $g(\mathbf{x})$ by Γ and the fact that $\mathbf{K}$ integrates to 1,

$$\begin{aligned}E[\Delta_t^2] &\le 2n^{-2}E[\mathbf{K}_h^2(\mathbf{x}-\mathbf{X}_t)\Psi_{\mathbf{x}}^2(Y_t-\mu)] \\ &\le \frac{2}{n^2h^d}E\big[\frac{1}{h^d}\mathbf{K}^2(\frac{\mathbf{x}-\mathbf{X}_t}{h})\big]\le 2(n^2h^d)^{-1}\Gamma K_\infty<\infty. \qquad \text{(B.11)}\end{aligned}$$

Therefore, we may apply the Bernstein inequality for strongly mixing time series of Carbon (1983) as in the proof of lemma 3.3.3 in Györfi et al. (1989) with the particular choice of $\alpha = c_2 nh^d s_n^{-1}$ and $C_\alpha = \alpha s_n (nh^d)^{-1} K_\infty > \frac{e}{4}$, and we get for any sequence $(\epsilon_n)_{n\in\mathbf{N}}$,

$$P(|\sum_{t=1}^{n} \Delta_t| > \epsilon_n) \leq c_1 \exp\{-c_2 nh^d \epsilon_n^2 s_n^{-1}\}, \tag{B.12}$$

uniformly in $\mathbf{x} \in \mathbf{R}^d$ and $\mu \in \mathbf{R}$ with some constants $c_1, c_2 > 0$.

Next, using the compactness of Θ, we cover it with M intervals $\mathbf{I}_m, m = 1, \ldots, M$, of length C_M:

$$\Theta \subset \cup_{m=1}^{M} \mathbf{I}_m, \quad \mathbf{I}_m = [\mu_{m-1}, \mu_m], \quad |\mu_m - \mu_{m-1}| = C_M, \quad m = 1, \ldots, M.$$

Mark that for all $m = 1, \ldots, M$, we have by monotonicity of $\Psi(Y_t - \mu)$ and, therefore, of $H_n(\mathbf{x}, \mu)$ as functions of μ

$$\begin{aligned} H_n(\mathbf{x}, \mu_{m-1}) &\leq \sup_{\mu\in\mathbf{I}_m} H_n(\mathbf{x}, \mu) = H_n(\mathbf{x}, \mu_m), \\ EH_n(\mathbf{x}, \mu_{m-1}) &\leq \sup_{\mu\in\mathbf{I}_m} EH_n(\mathbf{x}, \mu) = EH_n(\mathbf{x}, \mu_m). \end{aligned}$$

Therefore, we have for any $\mu \in \mathbf{I}_m$

$$\begin{aligned} H_n(\mathbf{x}, \mu) - EH_n(\mathbf{x}, \mu) &\leq H_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu_m) + EH_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu) \\ &\leq H_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu_m) + EH_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu_{m-1}). \end{aligned}$$

and, similarly,

$$EH_n(\mathbf{x}, \mu) - H_n(\mathbf{x}, \mu) \leq EH_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu_{m-1}) + EH_n(\mathbf{x}, \mu_{m-1}) - H_n(\mathbf{x}, \mu_{m-1}).$$

Recall that $E\{\Psi(Y_t - \mu) \,|\, \mathbf{X}_t = \mathbf{x}\} = F_{\mathbf{x}}(\mu) - \theta$, and that, by assumptions (A2 iii) and (C1), $F_{\mathbf{x}}(\mu)$ is Lipschitz continuous with constant c_f. Therefore,

$$\begin{aligned} EH_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu_{m-1}) &\leq c_f |\mu_m - \mu_{m-1}| \frac{1}{n} E[\sum_{t=1}^{n} \mathbf{K}_h(\mathbf{x} - \mathbf{X}_t)] \\ &= c_f C_M E[\hat{g}(\mathbf{x})]. \end{aligned}$$

Combining the last three equations, we get for all $\mu \in \mathbf{I}_m$,

$$\begin{aligned} & |H_n(\mathbf{x}, \mu) - EH_n(\mathbf{x}, \mu)| \\ \leq\ & \max\left\{|H_n(\mathbf{x}, \mu_{m-1}) - EH_n(\mathbf{x}, \mu_{m-1})|, |H_n(\mathbf{x}, \mu_m) - EH_n(\mathbf{x}, \mu_m)|\right\} + c_f C_M E\hat{g}(\mathbf{x}), \end{aligned}$$

and, therefore,

$$\sup_{\mu\in\Theta}|H_n(\mathbf{x},\mu)-EH_n(\mathbf{x},\mu)| \leq \max_{m=0,\ldots,M}|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)|+c_fC_ME\widehat{g}(\mathbf{x}) \tag{B.13}$$

We first consider the first term on the right-hand side and get, using (B.12),

$$\begin{aligned} & P(\max_{m=0,\ldots,M}|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)|>\epsilon_n) \\ \leq & \sum_{m=0}^{M}P(|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)|>\epsilon_n) \\ \leq & (M+1)c_1\exp\{-c_2nh^d\epsilon_n^2s_n^{-1}\}. \end{aligned}$$

We choose $C_M = c_3n^{-1}$ for some $c_3>0$ and, therefore, $M+1\leq c_4n$. Using the definition of $\widetilde{S}_n$, we have

$$\begin{aligned} P(\max_{m=0,\ldots,M}|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)|>\epsilon_n) &\leq c_1c_4n\exp\{-c_2\epsilon_n^2\widetilde{S}_n\log n\} \\ &= c_1c_4n^{1-c_2\widetilde{S}_n\epsilon_n^2} \\ &= c_1c_4n^{1-c_2\epsilon^2a_n^2} \\ &\leq c_5n^{-r} \end{aligned}$$

for arbitrary $r>0$ and some constant c_5 if n is large enough. Here, we have chosen $\epsilon_n = \epsilon\widetilde{S}_n^{-\frac{1}{2}}a_n$ for some arbitrary sequence $a_n\to\infty$ with $n\to\infty$. Choosing, e.g $r=2$, we get

$$\sum_{n=1}^{\infty}P(\frac{\widetilde{S}_n^{\frac{1}{2}}}{a_n}\max_{m=0,\ldots,M}|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)|>\epsilon)<\infty$$

which, by the Borel-Cantelli lemma, implies

$$\frac{\widetilde{S}_n^{\frac{1}{2}}}{a_n}\max_{m=0,\ldots,M}|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)|\to 0 \quad \text{a.s.}$$

As $a_n\to\infty$ arbitrarily slowly, this implies that

$$\widetilde{S}_n^{\frac{1}{2}}\max_{m=0,\ldots,M}|H_n(\mathbf{x},\mu_m)-EH_n(\mathbf{x},\mu_m)| \quad \text{is bounded a.s.} \tag{B.14}$$

Now, we consider the second term on the right-hand side of (B.13). As $h\to 0, \quad s_n\to\infty$, we have $\widetilde{S}_n^{\frac{1}{2}}C_M = \widetilde{S}_n^{\frac{1}{2}}c_3n^{-1}\to 0$ for $n\to\infty$. By Lemma 2.1, $E\widehat{g}(\mathbf{x})$ converges a.s. to $g(\mathbf{x})$ uniformly in $\mathbf{x}\in G$, and therefore, it is bounded. This implies $\widetilde{S}_n^{\frac{1}{2}} c_f C_M E\widehat{g}(\mathbf{x}) \to 0$ uniformly in $\mathbf{x} \in G$. Combining (B.13) with (B.14) we finally get

$$\sup_{\mu\in\Theta} |H_n(\mathbf{x},\mu) - EH_n(\mathbf{x},\mu)| = O(\widetilde{S}_n^{-\frac{1}{2}}) \quad \text{a.s. uniformly in} \quad \mathbf{x} \in G. \quad \text{(B.15)}$$

□

Lemma B.4. *In addition to the assumptions of Lemma B.3 assume (C2). Then,*

$$\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta_0} |H_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu) - EH_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu)| = O(\widetilde{S}_n^{-\frac{1}{2}}) \quad a.s.$$

Proof : Let Θ be the compact set of (C2). Then,

$$\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta_0} |H_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu) - EH_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu)| \le \sup_{\mathbf{x}\in G}\sup_{\nu\in\Theta} |H_n(\mathbf{x},\nu) - EH_n(\mathbf{x},\nu)|,$$

and the assertion follows from Lemma B.3. □

Lemma B.5. *Under the assumptions of Theorem 2.3, we have*

$$\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta_0} |EH_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu) - H(\mathbf{x},\mu_\theta(\mathbf{x})+\mu)| = O(h^2)$$

Proof : As

$$\begin{aligned} EH_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu) &= E\big[\mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)\Psi(Y_t-\mu_\theta(\mathbf{x})-\mu)\big] \\ &= E\big[\mathbf{K}_h(\mathbf{x}-\mathbf{X}_t)F_{\mathbf{X}_t}(\mu_\theta(\mathbf{x})+\mu)\big], \end{aligned}$$

the bias term does not depend on the dependence structure of the time series, and it can be treated exactly as in the well-known case of independent $\{Y_t, \mathbf{X}_t)\}_{t=1}^n$. Therefore, the result follows from standard arguments based on a Taylor expansion of $F_\mathbf{x}(y)$ up to order 2 with respect to $\mathbf{x}$, using in particular (A2 i-ii). □

Proof of Theorem 2.3 :
Combining Lemma B.3 and Lemma B.4, we have

$$\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta_0} |H_n(\mathbf{x},\mu_\theta(\mathbf{x})+\mu) - H(\mathbf{x},\mu_\theta(\mathbf{x})+\mu)| = O(\widetilde{S}_n^{-\frac{1}{2}}) + O(h^2) = O(S_n)$$

a.s. as $n \to \infty$. By Lemma 2.1, we have the same rate of convergence for $\sup_{\mathbf{x}\in G} |\widehat{g} - g|$. Hence, by the remarks following (B.10),

$$\sup_{\mathbf{x}\in G}\sup_{\mu\in\Theta_0} |\widetilde{H}_n(\mathbf{x}, \mu_\theta(\mathbf{x}) + \mu) - \widetilde{H}(\mathbf{x}, \mu_\theta(\mathbf{x}) + \mu)| = O(S_n) \to 0 \qquad \text{(B.16)}$$

a.s as $n \to \infty$. Using a similar technique as in Collomb and Härdle (1986), fix $\epsilon > 0$. By the definition (2.8) of $\widetilde{H}$, we have $\widetilde{H}(\mathbf{x}, \mu_\theta(\mathbf{x})) = 0$, and by assumptions (A2 iii-iv), (C2) we get from a Taylor expansion of $\widetilde{H}(\mathbf{x}, .)$ around $\mu_\theta(\mathbf{x})$

$$\widetilde{H}(\mathbf{x}, \mu_\theta(\mathbf{x}) - \epsilon) \leq -c_0\epsilon < 0 < c_0\epsilon \leq \widetilde{H}(\mathbf{x}, \mu_\theta(\mathbf{x}) + \epsilon)$$

for all $\mathbf{x} \in G$. The convergence in (B.16) and (2.10) imply that for any $0 < \delta < c_0\epsilon$ and all sufficiently large n we also have a.s. for all $\mathbf{x} \in G$

$$\widetilde{H}_n(\mathbf{x}, \mu_\theta(\mathbf{x}) - \epsilon) \leq \delta - c_0\epsilon < \widetilde{H}_n(\mathbf{x}, \mu_\theta(\mathbf{x})) < c_0\epsilon - \delta \leq \widetilde{H}_n(\mathbf{x}, \mu_\theta(\mathbf{x}) + \epsilon).$$

The monotonicity of $\widetilde{H}_n$ in μ implies $\mu_\theta(\mathbf{x}) - \epsilon < \widehat{\mu}_\theta(\mathbf{x}) < \mu_\theta(\mathbf{x}) + \epsilon$ a.s. for all $\mathbf{x} \in G$, i.e. we have

$$\sup_{\mathbf{x}\in G} |\hat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x})| \to 0 \quad \text{a.s. as} \quad n \to \infty. \qquad \text{(B.17)}$$

Again using $\widetilde{H}(\mathbf{x}, \mu_\theta(\mathbf{x})) = 0$ and (2.10), we have

$$\begin{aligned}
\widetilde{H}_n(\mathbf{x}, \widehat{\mu}_\theta(\mathbf{x})) - \widetilde{H}(\mathbf{x}, \widehat{\mu}_\theta(\mathbf{x})) &= \widetilde{H}(\mathbf{x}, \mu_\theta(\mathbf{x})) - \widetilde{H}(\mathbf{x}, \widehat{\mu}_\theta(\mathbf{x})) + O(\frac{1}{nh^d}) \\
&= -(\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x})) f_{\mathbf{x}}(\widetilde{\mu}_\theta(\mathbf{x})) + O(\frac{1}{nh^d}), \qquad \text{(B.18)}
\end{aligned}$$

by a Taylor expansion of $\widetilde{H}(\mathbf{x}, .) = F_{\mathbf{x}}(.) - \theta$, where $\widetilde{\mu}_\theta(\mathbf{x})$ is between $\mu_\theta(\mathbf{x})$ and $\widehat{\mu}_\theta(\mathbf{x})$. By (B.17), $\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x}) \in \Theta_0$ and, therefore, $\widetilde{\mu}_\theta(\mathbf{x}) \in \Theta$ a.s. for all $\mathbf{x} \in G$ and n large enough, and we get from (C2), (B.18) and (B.16)

$$|\widehat{\mu}_\theta(\mathbf{x}) - \mu_\theta(\mathbf{x})| \leq \frac{1}{c_0} |\widetilde{H}_n(\mathbf{x}, \widehat{\mu}_\theta(\mathbf{x})) - \widetilde{H}(\mathbf{x}, \widehat{\mu}_\theta(\mathbf{x})) - O(\frac{1}{nh^d})| = O(S_n) + O(\frac{1}{nh^d})$$

a.s. uniformly in $\mathbf{x} \in G$. $\square$

Acknowledgement: During this research, the second author has been working on a grant of Deutscher Akademischer Austauschdienst (DAAD).

References

[1] K. Abberger, *Nichtparametrische Schaetzung bedingter Quantile in Zeitreihen - Mit Anwendung auf Finanzmarktdaten*, Hartung-Gore Verlag, Konstanz, 1996.

[2] Berkowitz, J., Christoffersen, P. and Pelletier, D. (2009). Evaluating value-at-risk models with desk-level data, *Working Paper 010*, North Carolina State University, Department of Economics.

[3] P. J. Bickel and E. Lehmann, Descriptive statistics for nonparametric models.III: Dispersion *Ann. Statist.* **4**, (1976), 1139-1159.

[4] G. Boente and R. Fraiman, Asymptotic distribution of smoothers based on local means and local medians under dependence *Journal of Multivariate Analysis* **54**, (1995), 77-90.

[5] Z. Cai, Regression quantiles for time series *Econometric Theory* **18**, 01, (2002), pp 169-192.

[6] Z. Cai and X. Wang, Nonparametric method for estimating conditional value-at-risk and expected shorfall *Journal of Econometrics* **147**, (2008), pp 120-130.

[7] M. Carbon, Bernstein's inequality for strong mixing, non-stationary processes: applications *C. R. Acad. Scienc. Paris* **I**, 297, (1983), pp 303-306.

[8] G. Collomb and W. Härdle, Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations *Stochastic processes and their applications* **23**, (1986), 77-89.

[9] Engle, R. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles, *Journal of Business and Economic Statistics* **22**: 367-381.

[10] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications – Theory and Methodologies.* Chapman and Hall, New York, 1995.

[11] Lopez, J. A. and Walter, C. (2001). Evaluating Covariance Matrix Forecasts in a Value-at-Risk Framework, *Journal of Risk* **3**: 69-98.

[12] L. Györfi, W. Härdle, P. Sarda and P. Vieu, *Nonparametric curve estimation from time series*, Lecture Notes in Statistics, Springer-Verlag, Heidelberg, 1989.

[13] W. Härdle, H. Lütkepohl, and R. Chen, A review of nonparametric time series analysis. *International Statistical Review* **65**, (1997), 49-72.

[14] P. Hall and Q. Yao, Inference in ARCH and GARCH models with heavy-tailed errors *Econometrica* (2003), to appear.

[15] P. Hall, R. Wolff and Q. Yao. Methods for estimating a conditional distribution function *Journal of the American Statistical Association* **94**, (1999), 154-163.

[16] L. Horvath and B. Yandell, Asymptotics of conditional empirical processes *Journal of Multivariate Analysis* **26**, (1988), 184-206.

[17] P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[18] P. Jorion, *Value-at-Risk: The New Benchmark for Managing Financial Risk.* McGraw-Hill, New York, 2000.

[19] R. Koenker and G. Bassett, Regression quantiles *Econometrica* **46**, (1978), 33-50.

[20] R. Koenker, Galton, Edgeworth, Frisch and prospects for quantile regression in econometrics *Journal of Econometrics* **95**, (1999), 347-374.

[21] Y. S. Rama Krishnaiah, On the Glivenko-Cantelli theorem for generalized empirical processes based on strong mixing sequences *Stat. Probab. Letter* **47**(7), (1990), 2863-2875.

[22] E. Masry and D. Tjostheim, Nonparametric estimation and identification of nonlinear $ARCH$ time series: Strong convergence and asymptotic normality *Econometric Theory* **11**, (1995), 258-289.

[23] E. Masry and D. Tjostheim, Additive nonlinear ARX time series and projection estimates *Econometric Theory* **13**, (1997), 214-252.

[24] P. Mwita, *Semiparametric estimation of conditional quantiles for time series with applications in finance,* PhD Thesis, University of Kaiserslautern, 2003.

[25] E. A. Nadaraya, On estimating regression *Theory Probab. Appl.* **9**, (1964), 141-142.

[26] Y. K. Truong and C. J. Stone, Nonparametric function estimation involving time series *Ann. Statist.* **20**, (1992), 77-97.

[27] G. S. Watson, Smooth regression analysis *Sankya Ser. A* **26**, (1964), 359-372.

[28] A. A. Weiss, ARMA models with ARCH errors *J. Time Ser. Analysis* **3**, (1984), 129-143.

[29] A.A. Weiss, Estimating nonlinear dynamic models using least absolute error estimation *Econometric Theory* **7**, (1991), 46-68.