QUANTILE SIEVE ESTIMATES FOR TIME SERIES

Jürgen Franke*
Jean-Pierre Stockis*
Joseph Tadjuidje*

σt

* University of Kaiserslautern, Germany

BERLIN

ECONOMIC RISK

SFB 649

QUANTILE SIEVE ESTIMATES FOR TIME SERIES

JÜRGEN FRANKE, JEAN-PIERRE STOCKIS, JOSEPH TADJUIDJE

University of Kaiserslautern, Germany

ABSTRACT. We consider the problem of estimating the conditional quantile of a time series at time t given observations of the same and perhaps other time series available at time $t - 1$. We discuss sieve estimates which are a nonparametric versions of the Koenker-Bassett regression quantiles and do not require the specification of the innovation law. We prove consistency of those estimates and illustrate their good performance for light- and heavy-tailed distributions of the innovations with a small simulation study. As an economic application, we use the estimates for calculating the value at risk of some stock price series.

1

1. Introduction

Reliable measures of market risk are crucial tools for an effective risk management which financial institutions have to employ for internal and regulatory purposes. There are now established procedures for modelling asset returns and for subsequent calculation of risk measures, but there is still room for improvement and more flexibility dealing with the shortcomings of standard methodology. An extensive discussion of these issues has been given recently by (Engle and Manganelli 2002).

The standard measure of market risk is currently the *value at risk* (VaR). If Y_t is the return of an asset a time t, the value at risk of level $1 - \alpha$ at time t (VaR_t) is determined by the conditonal α-quantile q_t of Y_t given information up to time $t - 1$ represented by the σ-algebra $\mathbf{I}_{t-1}$, i.e.

$$pr(Y_t \leq q_t | \mathbf{I}_{t-1}) = \alpha.$$

Then, $VaR_t = -q_t$ where we follow the convention that the value at risk is commonly reported as a positive number. It provides a bound on future losses at time t which is not exceeded with high probability $1 - \alpha$ given currently available information.

We concentrate on the VaR as a risk measure, but our exposition can be easily extended to the *expected shortfall* as another popular measure of risk, i.e. the conditional expectation of the loss given that it exceeds the value at risk

$$(1) \qquad es_t = E\big\{ - Y_t \, \big| \, \mathbf{I}_{t-1}, -Y_t \geq VaR_t \big\}.$$

The expected shortfall provides more information than VaR_t about the size of extreme losses and, moreover, is a coherent risk measure as shown by (Artzner *et al.* 1997).

A broad class of approaches to estimating VaR is volatility based, i.e. the distribution of the return time series $\{Y_t, -\infty < t < \infty\}$ is assumed to conform with some form of stochastic volatility model of the general form

$$(2) \qquad Y_t = \mu_t + \sigma_t \eta_t$$

where μ_t, σ_t denote the conditional mean and volatility of Y_t given information on the past up to time $t - 1$, and the innovations η_t are i.i.d. with mean 0 and variance 1. If q_α^η denotes the α-quantile of η_t, then for a time series following (2), we immediately have $VaR_t = \mu_t + \sigma_t q_\alpha^\eta$. Typical examples are based on the classical ARCH- and GARCH-models for financial returns and their extensions, compare (Engle 1982), (Bollerslev 1986). For the GARCH(1,1) model, which currently is quite popular for market risk analysis, we have, e.g., $\mu_t = 0, \sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2$. To calculate an estimate of VaR_t given data $Y_1, \ldots, Y_{t-1}$, we only need estimates of the model parameters ω, α, β and some initial value for the first volatility σ_1, where the influence of the latter will be neglible for large t under short memory asumptions like (A1) below.

As pointed out by (Engle and Manganelli 2002), volatility based estimates of VaR assume that the extreme negative returns follow the same process as the remainder of typical returns. Additionally, it is assumed that the standardized returns $\eta_t = (Y_t - \mu_t)/\sigma_t$ are i.i.d., and, moreover, their distribution has to be specified, frequently as standard normal. To avoid relying on those assumptions which may well be not satisfied for real data, (Engle and Manganelli 2002) propose to estimate the VaR directly without taking a detour over estimating volatility and without having to make an ad-hoc choice of an innovation distribution. They consider a class of models where the conditional quantile q_t is specified as a function of finitely many of its own past values as well as of past returns. The different functions which they consider are specified up to finitely many parameters and are quite similar to the manner how volatility is given as a function of the past in GARCH models and their modifications like, e.g., TGARCH (Rabemananjara and Zakoian 1993), (Glosten *et al.* 1993). (Engle and Manganelli 2002) call such models CAViaR, i.e. conditional autoregressive value at risk, and discuss how to estimate the parameters following the regression quantiles approach of (Koenker and Bassett 1978).

Models like GARCH for volatility or CAViaR for value at risk have a particular parametric form to be chosen in advance. A more flexible alternative is given by nonparametric approaches. For volatility based models, these have been extensively studied in the last years. E.g., we may choose $\mu_t = m(X_t), \sigma_t = \sigma(X_t)$ in the volatility based model (2), where X_t is a vector of finitely many $Y_{t-1}, \ldots, Y_{t-p}$ and perhaps some additional financial data known at time $t-1$, we get a nonparametric AR(d)-ARCH(d)-model including exogeneous variables. Local smoothing estimates of the trend and volatility functions m, σ and their use for market risk management have been studied by (Härdle and Tsybakov 1997), (Hafner 1998) and (Franke *et al.* 2004) among many others. Nonparametric sieve estimates of m, σ based on neural networks or on other approximating function classes are discussed in (Gouriéroux and Montfort 1992), (Franke 1998), (Franke 2000) and, in a similar general context as in this paper, in (Franke and Diagne 2006). If the main interest in fitting such nonparametric models to financial data is estimating the value at risk, then they suffer from similar problems as the volatility-based parametric approaches. The innovation distribution has to be specified somewhat arbitrarily, and the crucial volatility estimate is mainly influenced by the bulk of the data, i.e. by small and medium returns which are not important for managing extreme risks. The latter is only partially true for estimates based on local smoothing, e.g. kernel or local polynomial estimates, but those suffer from the curse of dimensionality which in particular leads to highly unreliable estimates in regions with low data density, in particular in the regions of the few extreme data.

In this paper, we try to combine both approaches, i.e. nonparametric estimation to get flexibility and the Koenker-Bassett method of accessing regression quantiles directly which does not require the specification of the innovation law and focusses on the extreme data which are important for quantifying risk. We study general sieve estimates as, in assessing the risk of an asset, we want to allow for incorporating not

only past asset prices, but also other available information on the market. This leads to the problem of estimating functions on higher-dimensional spaces than local smoothers can easily handle.

In the following, the information available at time $t-1$ is represented by an observable random vector $X_t \in \mathbb{R}^d$ which may consist of past observations $Y_{t-1}, \ldots, Y_{t-p}$ of the time series of interest but also on past observations of other time series. Our goal is to estimate the conditional α-quantile function $q(x)$ given by

$$
(3) \qquad\qquad pr(Y_t \leq q(X_t) \,|\, X_t = x) = \alpha.
$$

$q(x)$ solves the minimization problem

$$
(4) \qquad E\{|Y_t - q(X_t)|_\alpha \,|\, X_t = x\} = \min_{f \in L^1(\mu)} E\{|Y_t - f(X_t)|_\alpha \,|\, X_t = x\}
$$

Here, $|u|_\alpha$ denotes the skew absolute value given by

$$
(5) \qquad |u|_\alpha = \alpha u^+ + (1-\alpha) u^- = u(\, \alpha - 1_{(-\infty,0)}(u)\,) = \alpha u + u^-
$$

where u^+, u^- denote the positive and negative part of u.

In section 2 we introduce general nonparametric sieve estimates for $q(x)$ and formulate a nonparametric consistency result. In the following two sections, we consider two special case: qualitative threshold quantile estimates similar to the trend and volatility estimates proposed by (Gouriéroux and Montfort 1992) and neural network based quantile estimates. In section 5 we present some simulations and application to quantifying market risk. Some technical results and all the proofs are deferred to the last section 6.

2. CONSISTENCY OF SIEVE QUANTILE ESTIMATES

We need the following assumptions on the time series of interest.

(A1) (Y_t, X_t) is α-mixing with geometrically decreasing mixing coefficients, i.e. the mixing coefficients α_s satisfy

$$
\alpha_s \leq a_1 e^{-a_2 s}, \; s \geq 1,
$$

for some $a_1, a_2 > 0$. Furthermore, $E|Y_t| < \infty$.

(A2) Let $p_\varepsilon(z|x)$ denote the conditional density of $\varepsilon_t = Y_t - q(X_t)$ given $X_t = x$. There are functions $\pi(x), \zeta(x)$ and a constant γ_0 such that for all x

a) $p_\varepsilon(z|x) \geq \pi(x) > 0$ for all $|z| \leq \zeta(x)$,

b) $\pi(x)\zeta(x) \geq \gamma_0 > 0$.

(A1) is a standard short-memory condition. (A2) corresponds to the usual assumption for quantile asymptotics that the probability density of Y_t is uniformly bounded away from 0 in a neighbourhood of the quantile - in our case conditional on x with a certain

degree of uniformity w.r.t. x. The condition is not very strong. For example, consider the case

$$(6) \qquad\qquad Y_t = q(X_t) + s(X_t)\,\eta_t,$$

where $q(x)$ is the conditional α-quantile of Y_t given $X_t = x$, $s(x)$ is the conditional α-scale of Y_t given $X_t = x$, i.e. the conditional α-quantile of $|Y_t - q(X_t)|_\alpha$ given $X_t = x$, and η_t, $-\infty < t < \infty$, are i.i.d. real random variables with α-quantile 0, α-scale 1 and density p_η. If we assume that p_η is bounded away from 0 in a neighbourhood of its α-quantile 0, i.e. for some π_0, $\zeta_0 > 0$

$$p_\eta(u) \geq \pi_0 > 0 \text{ for } |u| \leq \zeta_0,$$

then (A2) is satisfied with $\pi(x) = \pi_0/s(x)$, $\zeta(x) = \zeta_0 s(x)$ and $\gamma_0 = \pi_0\zeta_0$ as $\varepsilon_t = s(X_t)\eta_t$ and, therefore,

$$p_\varepsilon(z|x) = \frac{1}{s(x)}p_\eta\left(\frac{z}{s(x)}\right) \geq \frac{\pi_0}{s(x)} \quad \text{for } |z| \leq \zeta_0 s(x).$$

We remark that for the special case $X_t = (Y_{t-1}, \ldots, Y_{t-p})^T$, (6) is a quantile AR($p$) - ARCH($p$)-process as discussed in (Franke and Mwita 2003).

Let $\mathbf{F}_n, n \geq 1$, denote an increasing sequence of subsets of $L^1(\mu)$, and let $\mathbf{F}_\infty$ denote their union. We estimate the conditional quantile function $q(x)$ by solving the sample version of (4) restricted to functions in $\mathbf{F}_n$, i.e.

$$(7) \qquad\qquad q_n = \mathrm{argmin}_{f \in \mathbf{F}_n} \frac{1}{n} \sum_{t=1}^{n} \left|Y_t - f(X_t)\right|_\alpha.$$

Estimating q by q_n belongs to the broad class of nonparametric regression estimates based on Grenander's method of sieves (Grenander 1981). To get consistency of these estimates we have to assume that $\mathbf{F}_\infty$ is dense in $L^1(\mu)$, the space of integrable functions on $\mathbb{R}^d$ w.r.t. μ. Mark that $q \in L^1(\mu)$ as we have assumed $E|Y_t| < \infty$.

Examples for $\mathbf{F}_n$ are given by piecewise constant functions or by feedforward neural networks which we discuss in detail in sections 3 resp. 4.

Typically, the functions in $\mathbf{F}_n$ are parametrized by some parameter vector with finite dimension increasing with n. For proving consistency of the estimate q_n of (7), we could assume uniform boundedness of the functions in $\mathbf{F}_n$ which usually is achieved by bounding the parameter vector or, in the case of feedforward neural networks, like in Theorem 3.3 of (White 1990) or Theorem 3.2 of (Franke and Diagne 2006). This procedure has some computational drawbacks discussed in section 10.1 of (Györfy et al. 2002) where, as an alternative to bounding the functions in $\mathbf{F}_n$ in advance, the original estimate q_n is replaced by a truncated version instead, i.e. for some sequence $\Delta_n \to \infty$ we consider

$$(8) \qquad\qquad \hat{q}_n(x) = T_{\Delta_n} q_n(x),$$

where the truncation operator T_L is defined as

$$T_L u = u, \text{ if } |u| \leq L, \text{ and } T_L u = L \operatorname{sgn}(u), \text{ else.}$$

Let

$$\widehat{\mathbf{F}}_n = \{T_{\Delta_n} f \; ; \; f \in \mathbf{F}_n\}$$

denote the truncated functions of $\mathbf{F}_n$. We assume that $\widehat{\mathbf{F}}_n$ satisfies the following assumption on bounded real-valued functions.

(A3) G is a class of bounded, real-valued measurable functions on $\mathbb{R}^d$ such that for all $\delta > 0$, $N \geq 1$, there exists $k_N(\delta)$ such that for all $z_1, \ldots, z_N \in \mathbb{R}^d$ there are functions $g_k^* : \mathbb{R}^d \to \mathbb{R}$, $k = 1, \ldots, k_N(\delta)$, with:

for any $g \in \mathbf{G}$ there is a $k \leq k_N(\delta)$ such that $\frac{1}{N} \sum_{j=1}^N |g(z_j) - g_k^*(z_j)| < \delta$.

$k_N(\delta)$ is a bound on the δ-covering number of $\mathbf{G}$ w.r.t. the L^1-norm of the discrete measure with point masses $1/N$ in $z_1, \ldots, z_N$, assumed to hold uniformly in $z_1, \ldots, z_N$, compare ch.9 of (Györfy *et al.* 2002). Let $K_N(\delta)$ denote the size of the smallest δ-cover, i.e. the minimal value of $k_N(\delta)$ in (A2).

Assumption (A2) is satisfied for many function classes $\mathbf{G}$. By Lemma 9.2 and Theorem 9.4 of (Györfy *et al.* 2002), we have, e.g., for all N and some bound B on the absolute value of functions in $\mathbf{G}$

$$(9) \qquad\qquad K_N(\delta) \leq 3 \left(\frac{4eB}{\delta} \log \frac{6eB}{\delta} \right)^{\mathbf{V}(\mathbf{G}^+)}$$

if the Vapnik-Chervonenkis dimension $\mathbf{V}(\mathbf{G}^+)$ of $\mathbf{G}^+ = \{(z,t); \; t \leq g(z) + B, g \in \mathbf{G}\}$ is at least 2 and if $\delta < B/2$. Mark that (9) differs slightly from the version in (Györfy *et al.* 2002) as we do not assume that $\mathbf{G}$ contains only nonnegative functions.

For later reference, we remark that each δ-covering of $\mathbf{G}$ w.r.t. $z_1, \ldots, z_{2N}$ is automatically a 2δ-covering w.r.t. $z_1, \ldots, z_N$ as

$$\frac{1}{N} \sum_{j=1}^N |g(z_j) - g_k^*(z_j)| \leq 2 \frac{1}{2N} \sum_{j=1}^{2N} |g(z_j) - g_k^*(z_j)|,$$

which immediately implies

$$(10) \qquad\qquad K_N(2\delta) \leq K_{2N}(\delta) \quad \text{for all } N \geq 1, \delta > 0.$$

Theorem 1. *Let $\{(Y_t, X_t)\}$ be a stationary stochastic process satisfying (A1) and (A2). Let $\mathbf{F}_n$ be increasing classes of bounded functions in $L^1(\mu)$, such that their union $\mathbf{F}_\infty$ is dense in $L^1(\mu)$, and, for $\Delta_n \to \infty$, the corresponding classes of truncated functions $\widehat{\mathbf{F}}_n$ satisfy (A3). Let*

$$\kappa_n(\epsilon) = \log K_{2n} \left(\frac{\epsilon}{32} \right).$$

Let $\hat{q}_n = T_{\Delta_n} q_n$, given by (7) and (8) be the truncated sieve estimate for the conditional α-quantile $q(z)$ given by (3).

a) *If, for $n \to \infty$, $\quad \Delta_n \kappa_n(\epsilon)/\sqrt{n} \to 0$ for all $\epsilon > 0$, then $\hat{q}_n$ is a consistent estimate of q in the mean sense, i.e. for $n \to \infty$*

$$E \int |\hat{q}_n(z) - q(z)| \mu(dz) \to 0.$$

b) *Let, additionally, $\{Y_t\}$ satisfy Cramér's condition, i.e. $E|Y_t|^j \leq c^{j-2}j!EY_t^2$, $j = 3, 4, \ldots$ for some $c > 0$. If, for some $\beta > 0$ and some sequence $\delta_n \to 0$ we have $\Delta_n \kappa_n(\epsilon\delta_n)/\sqrt{n} \to 0$ and $\Delta_n/(\delta_n n^{\frac{1}{2}-\beta}) \to 0$, then $\hat{q}_n$ is even strongly $L^1(\mu)$-consistent, i.e. for $n \to \infty$*

$$\int |\hat{q}_n(z) - q(z)| \mu(dz) \to 0 \quad a.s.$$

By this result, proving consistency of the truncated sieve estimate of the conditional quantile $q(z)$ for specific function classes $\mathbf{F}_n$ reduces to finding bounds on the covering numbers. In the next two sections, we consider two specific examples.

3. Qualitative threshold quantile estimates

(Gouriéroux and Montfort 1992) have introduced the class of qualitative threshold ARCH models for financial time series. For order d, they have the form

$$Y_t = \sum_{j=1}^{H} a_j 1_{A_j}(Y_{t-1}, \ldots, Y_{t-d}) + \sum_{j=1}^{H} b_j 1_{A_j}(Y_{t-1}, \ldots, Y_{t-d})\eta_t$$

where $A_1, \ldots, A_H$ is a given partition of $\mathbb{R}^d$, i.e. the sets are pairwise disjoint and their union is $\mathbb{R}^d$, and the η_t are white noise with zero mean and unit variance. A straightforward extension would allow the conditional mean and volatility of Y_t given the past to depend on a general random vector X_t observable at time $t-1$ including past values $Y_s, s < t$ as well as other market data. The elements A_j of the partition may correspond to phases of increasing and decreasing prices, to phases of low and high volatility, etc.

Based on this intuition, we consider approximating the conditional quantile function $q(x)$ of (3) by a simple function from

(11) $$\mathbf{P}(H) = \{f(x) = \sum_{j=1}^{H} c_j 1_{A_j}(x); c_1, \ldots, c_H \in \mathbb{R}\}.$$

Applying this approach to VaR-calculation is based on the assumption that approximately the market can be in H different states characterized by the value of the risk variable X_t observable at time $t-1$ and that the VaR of the asset of interest is approximately constant in each state. If H is chosen large enough and the $A_1, \ldots, A_H$ provide a suitable partition of $\mathbb{R}^d$, then we get a reasonable approximation of $q(x)$ even

if it is not locally constant. This follows from the following consistency result which is a special case of Theorem 1 for the function classes

$$\mathbf{F}_n = \mathbf{P}(H_n) = \{f(x) = \sum_{j=1}^{H_n} c_j 1_{A_{nj}}(x); \ c_1, \ldots, c_{H_n} \in \mathbb{R}\}.$$

We have to assume that $\mathbf{F}_n$ is increasing in n and that $\mathbf{F}_\infty$ is dense in $L^1(\mu)$ which follows from $H_n \to \infty$ and the following assumptions on the partitioning:

(A4) For all n, $\mathbf{A}_n = \{A_{n1}, \ldots, A_{nH_n}\}$ is a partition of $\mathbb{R}^d$, such that
   a) for $m > n$ and any $i \le H_m$, $A_{mi} \subset A_{nj}$ for some $j \le H_n$,
   b) for all bounded subsets B of $\mathbb{R}^d$, $\sup_{j \le H_n} diam(A_{nj} \cap B) \to 0$ for $n \to \infty$.

a) states that $\mathbf{A}_{n+1}$ is a subpartition of $\mathbf{A}_n$, and b) guarantees that the partitions become finer and finer with increasing n except for the extreme part of $\mathbb{R}^d$. For given H_n, we get as a nonparametric quantile estimate of $q(x)$:

$$(12) \qquad\qquad q_n(x) = \sum_{j=1}^{H_n} c_{nj} 1_{A_{nj}}(x) \qquad \text{where}$$

$$\mathbf{c_n} = \mathrm{argmin}_{b_1,\ldots,b_{H_n}} \frac{1}{n} \sum_{t=1}^{n} \left| Y_t - \sum_{j=1}^{H_n} b_j 1_{A_{nj}}(X_t) \right|_\alpha$$

with $\mathbf{c_n} = (c_{n1}, \ldots, c_{nH_n}) \in \mathbb{R}^{H_n}$. As only one term in the sum does not vanish, truncating $q_n(x)$ is equivalent to just truncating the coefficients c_{nj}, and we get

$$(13) \qquad \hat{q}_n(x) = T_{\Delta_n} q_n(x) = \sum_{j=1}^{H_n} \hat{c}_{nj} 1_{A_{nj}}(x) \qquad \text{with} \quad \hat{c}_{nj} = T_{\Delta_n} c_{nj}.$$

Theorem 2. *Let $\{(Y_t, X_t)\}$ be a stationary process satisfying (A1) and (A2). For $H_n \to \infty, \Delta_n \to \infty$, let $\hat{q}_n$ be the truncated qualitative threshold quantile estimate for q given by (12) and (13). Assume that the sequence of partitions $\mathbf{A}_n$ satisfies (A4).*
   a) *If for $n \to \infty, \Delta_n H_n \log(\Delta_n)/\sqrt{n} \to 0$, then*

$$E \int |\hat{q}_n(x) - q(x)| \mu(dx) \to 0 \quad (n \to \infty)$$

   b) *If, additionally, $\{Y_t\}$ satisfies Cramér's condition and $\Delta_n^2/n^{1-\beta} \to 0$ for some $\beta > 0$, then*

$$\int |\hat{q}_n(z) - q(z))| \mu(dz) \to 0 \quad a.s. \quad (n \to \infty).$$

4. Neural networks

As a second example, we now consider estimates for $q(z)$ based on fitting neural networks to the data. Given an input variable $x = (x_1, x_2, ..., x_d)^T \in \mathbb{R}^d$, a feedforward neural network with one hidden layer consisting of $H \geq 1$ neurons defines a function $f(x) = f_H(x, \theta)$ of the following form

$$f_H(x; \theta) = v_0 + \sum_{h=1}^{H} v_h \Psi(x^T w_h + w_{h0})$$

where $w_h = (w_{h1}, ..., w_{hd})^T$. The so-called activation function Ψ is fixed in advance, whereas the network weights $v_0, ..., v_H, w_{hi}, h = 1, ..., H, i = 0, ..., d$, which we combine to a $M(H)$-dimensional parameter vector θ with $M(H) = 1 + H + H(1 + d)$, may be chosen appropriately. We denote the class of such neural network output functions by

(14) $$\mathbf{O} = \left\{ f_H(x; \theta); \theta \in \mathbb{R}^{M(H)}, H \geq 1 \right\}.$$

In the following, we consider only sigmoid activation functions satisfying

(A5) Ψ is continuous and strictly increasing, $0 < \lim_{x \to \infty} \Psi(x) = \Psi(\infty) \leq 1$ and
$0 \geq \lim_{x \to -\infty} \Psi(x) = \Psi(-\infty) \geq -1$.

Assuming $|\Psi(u)| \leq 1$ is no restriction but only a convenient standardization. A typical example of such a function is the hyperbolic tangent or symmetrized logistic function

(15) $$\Psi(u) = \tanh(u) = \frac{2}{1 + \exp(-2u)} - 1.$$

We also consider neural networks of finite complexity characterized by subclasses of $\mathbf{O}$ of the form

(16) $$\mathbf{O}(H, \Delta) = \left\{ f_H(x; \theta); \theta \in \mathbb{R}^{M(H)}, \sum_{h=0}^{H} |v_h| \leq \Delta \right\}$$

for some given number $H \geq 1$ of neurons and some bound Δ on the ℓ^1-norm of the output weights. We consider the increasing function classes

$$\mathbf{F}_n = \mathbf{O}(H_n, \Delta_n) \quad \text{for some increasing sequences } H_n, \Delta_n \to \infty.$$

Their union $\mathbf{F}_\infty = \mathbf{O}$ is dense in $L^2(\mu)$ by Theorem 1 of (Hornik 1991), compare also Lemma 16.2 of (Györfy *et al.* 2002), if Ψ satisfies (A3). But $\mathbf{O} \subset L^1(\mu)$ too, as, by (A5), it consists of bounded functions, and for any $f \in L^1(\mu), g \in \mathbf{O}, L > 0$ we have by the triangular and by Jensen's inequality

$$\int |f(x) - g(x)| \mu(dx) \leq \int |f(x) - T_L f(x)| \mu(dx) + \left(\int |T_L f(x) - g(x)|^2 \mu(dx) \right)^{1/2},$$

which implies denseness of $\mathbf{O}$ in $L^1(\mu)$ too.

Now, we consider the estimate $q_n(x)$ of $q(x)$ based on feedforward neural networks, i.e.

$$(17) \qquad q_n(x) = f_{H_n}(x; \hat{\theta}_n), \qquad \hat{\theta}_n = \mathrm{argmin}_{\theta \in \Theta_n} \frac{1}{n} \sum_{t=1}^{n} \left| Y_t - f_{H_n}(X_t; \theta) \right|_\alpha$$

with $\Theta_n = \{\theta \in \mathbb{R}^{M(H_n)}; \sum_{h=0}^{H_n} |v_h| \leq \Delta_n\}$. From Theorem 1, we get immediately

Theorem 3. *Let $\{(Y_t, X_t)\}$ be a stationary process satisfying (A1) and (A2). For $H_n \to \infty, \Delta_n \to \infty$, let q_n be the neural network quantile estimate for $q(x)$ given by (17). Assume that Ψ satisfies (A5).*

a) *If for $n \to \infty, \Delta_n H_n \log(\Delta_n H_n)/\sqrt{n} \to 0$, then*

$$E \int |q_n(x) - q(x)| \mu(dx) \to 0 \quad (n \to \infty)$$

b) *If, additionally, $\{Y_t\}$ satisfies Cramér's condition and $\Delta_n^2/n^{1-\beta} \to 0$ for some $\beta > 0$, then*

$$\int |q_n(z) - q(z))| \mu(dz) \to 0 \quad a.s. \quad (n \to \infty).$$

5. Simulations and applications

In this section, we first apply nonparametric quantile sieve estimates to some artificially generated data. As approximating function classes, we use feedforward neural networks as in section 4. For an easy graphical comparison of the function estimate with the true quantile function, we restrict ourselves to the case of a one-dimensional regressor. Finally, we use the quantile sieve approach for estimating the conditional VaR of some real stock price series. In each case, we have chosen the size of the network such that a further increase of the number of neurons did not change the visual impression significantly.

For simulation, we consider a nonlinear AR-ARCH processes of order 1, i.e.
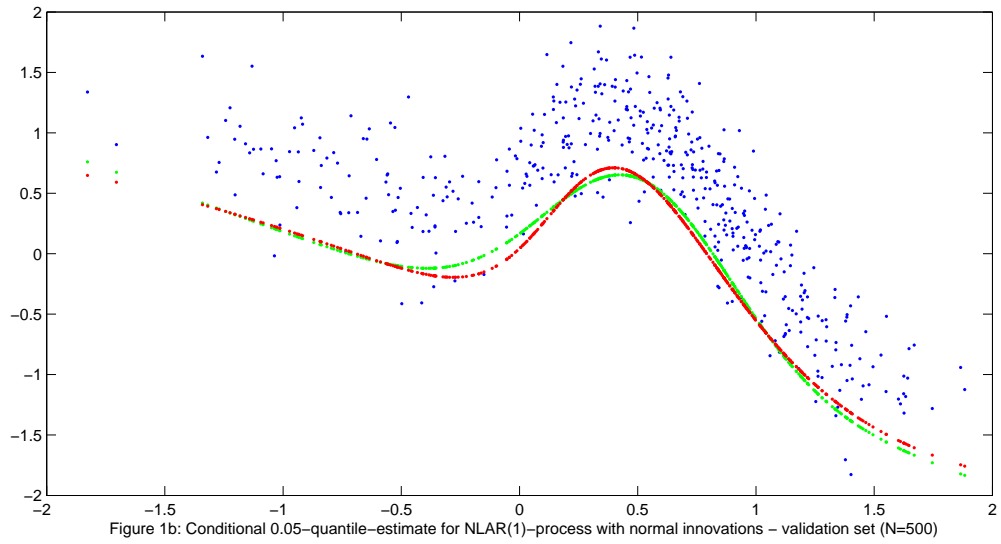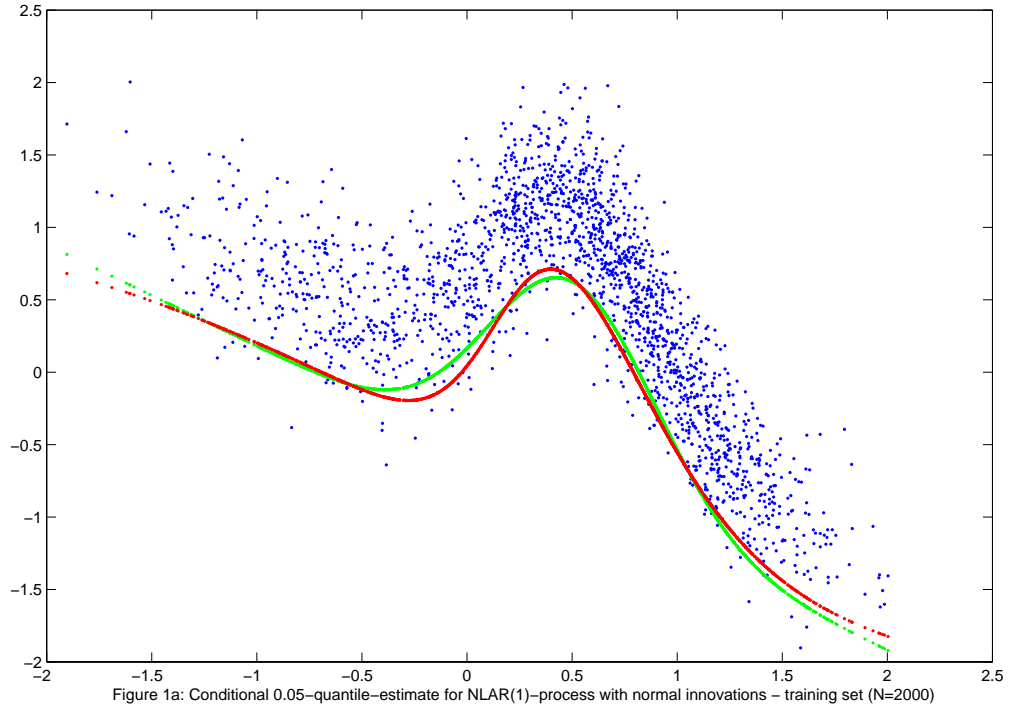
$$Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\eta_t$$

with i.i.d. innovations η_t having mean 0 and variance 1. In each case, we generate a sample of size 2500, use the first 2000 data as a training set from which we get the estimates of the network parameters. The last 500 observations are set aside as a validation set to check the out-of-sample performance of the estimate.

In the first two examples, we consider pure autoregressive processes with a bump function as the autoregressive function
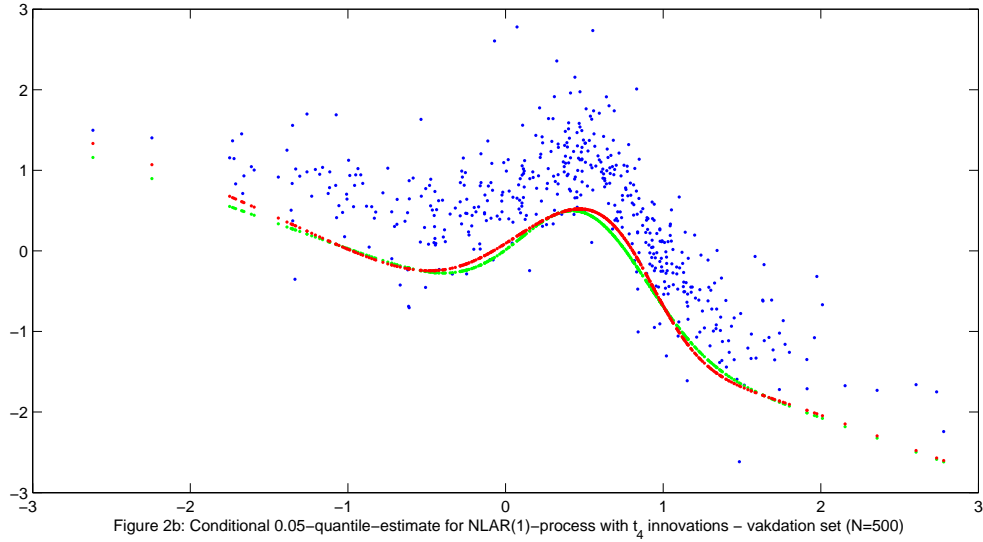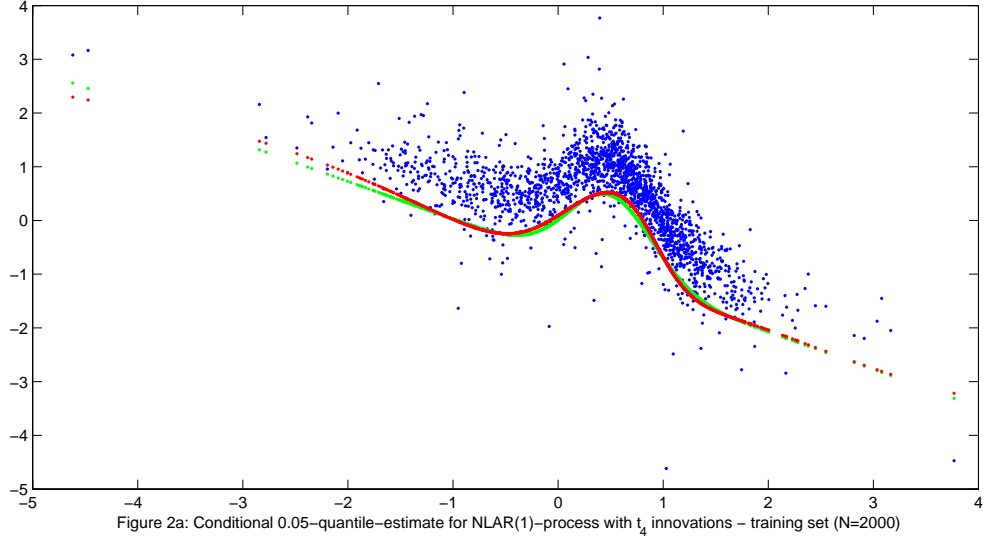
$$m(x) = -0.7x + 1.5\varphi_{0.5,0.4}(x), \qquad \sigma(x) = 0.2,$$

where $\varphi_{\mu,v}$ denotes the density of the normal law with mean μ and variance v. We use a feedforward neural network with H=7 neurons to estimate the conditional 5%-quantile

Figure 1a: Conditional 0.05–quantile–estimate for NLAR(1)–process with normal innovations – training set (N=2000)

Figure 1b: Conditional 0.05–quantile–estimate for NLAR(1)–process with normal innovations – validation set (N=500)
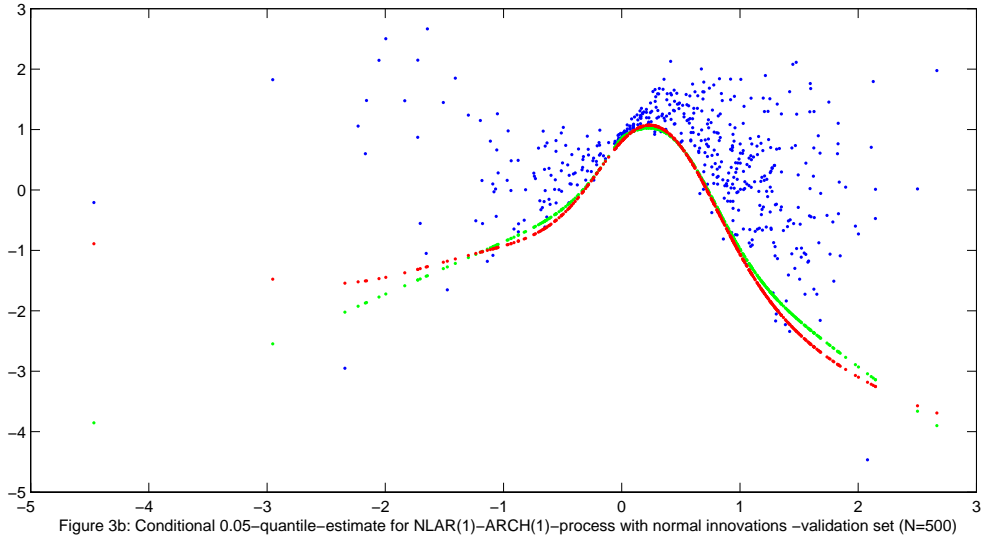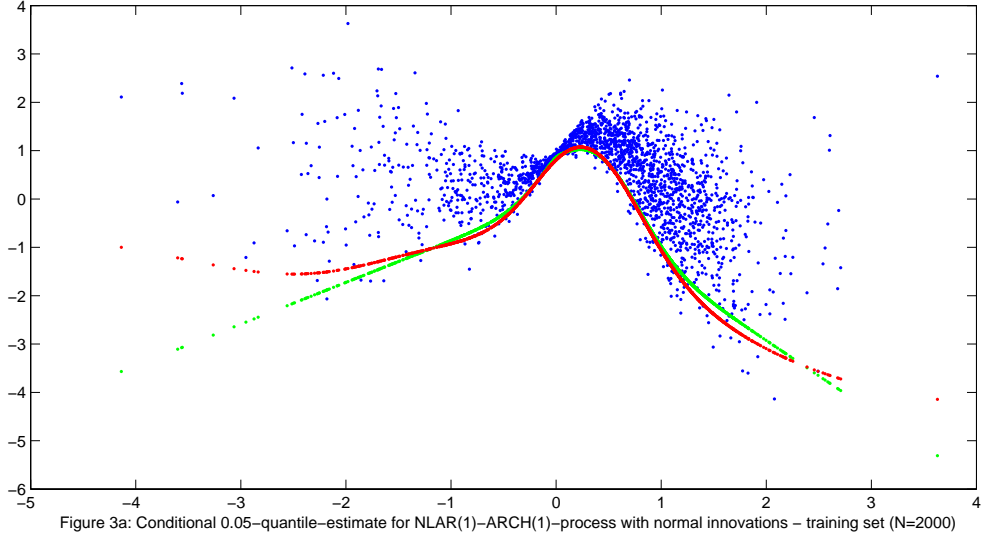
function $q(x)$.

For standard normal innvoations, Figure 1a shows the scatter plot Y_t against $Y_{t-1}, t = 2, \ldots, 2000$, of the training set as well as the true quantile function $q(x)$ (green curve)

Figure 2a: Conditional 0.05–quantile–estimate for NLAR(1)–process with t_4 innovations – training set (N=2000)

Figure 2b: Conditional 0.05–quantile–estimate for NLAR(1)–process with t_4 innovations – vakdation set (N=500)

and the neural-network based quantile estimate $q_n(x)$ (red curve). Mark that for pure autoregressive processes, the conditional quantile function is just a shifted version of the conditional mean $m(x)$. Figure 1b shows the same picture for the data of the validation set. On the training set, we get an empirical level of 4.95%, i.e. a fraction of 0.0495 of the data Y_t are below the estimated conditional quantile $q_n(Y_{t-1})$. For the validation set, the empirical level is 5.61%.

For the second example, we consider the same autoregressive process but with heavy-tailed (t_4-distributed) innovations η_t. Figures 2a and 2b show the corresponding

Figure 3a: Conditional 0.05–quantile–estimate for NLAR(1)–ARCH(1)–process with normal innovations – training set (N=2000)

Figure 3b: Conditional 0.05–quantile–estimate for NLAR(1)–ARCH(1)–process with normal innovations –validation set (N=500)

results for the training and the validation set. The empirical levels are 5.00% and 6.81% resp.

As the last example, we generate data from a genuine AR-ARCH-process, again with a bump function as autoregressive function $m(x)$ and with a volatility function $\sigma(x)$ as in the parametric ARCH(1)-model of (Engle 1982):

$$m(x) = -0.2x + 1.5\varphi_{0.5,0.4}(x), \qquad \sigma^2(x) = 0.01 + 0.5x^2$$

As innovations, we use again standard normal variables, such that the conditional law of Y_t given $Y_{t-1} = x$ is normal with mean $m(x)$ and variance $\sigma^2(x)$. Therefore, the true

conditional quantile function is no longer just a shifted conditional mean. As the basis for the nonparametric quantile estimate we use a neural network with $H = 9$ neurons. Figures 3a and 3b show the scatter plots for training and validation set and the true and estimated quantile function. The empirical levels are 5.00% and 4.21% resp.

Finally, we consider the problem of estimating the conditional 5%-VaR for the BASF-stock for the period 1990 to 1992 (N=745) which covers the first Gulf War as a phase of high volatility and the attempted coup d'etat in Moscow as an example of an isolated event, having a strong, but very local effect on the market.
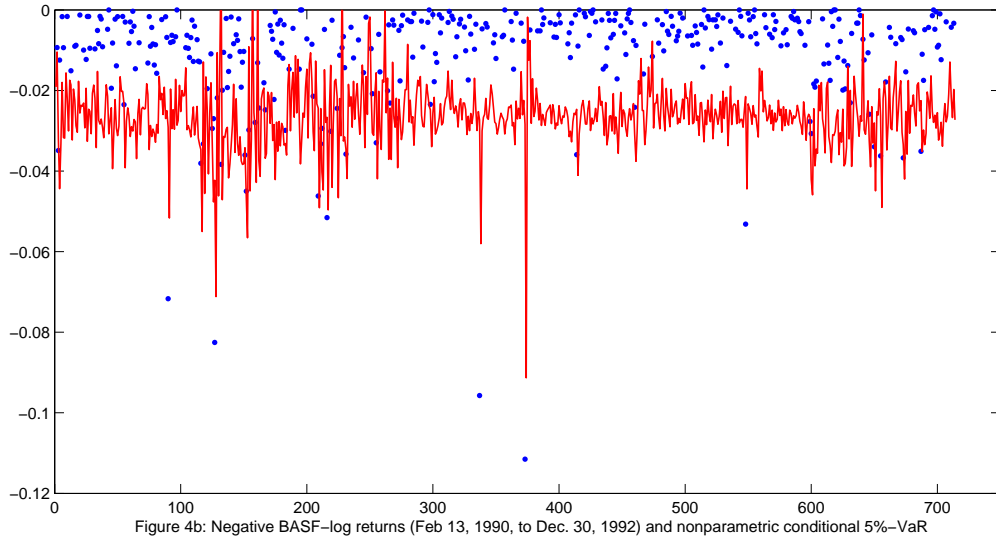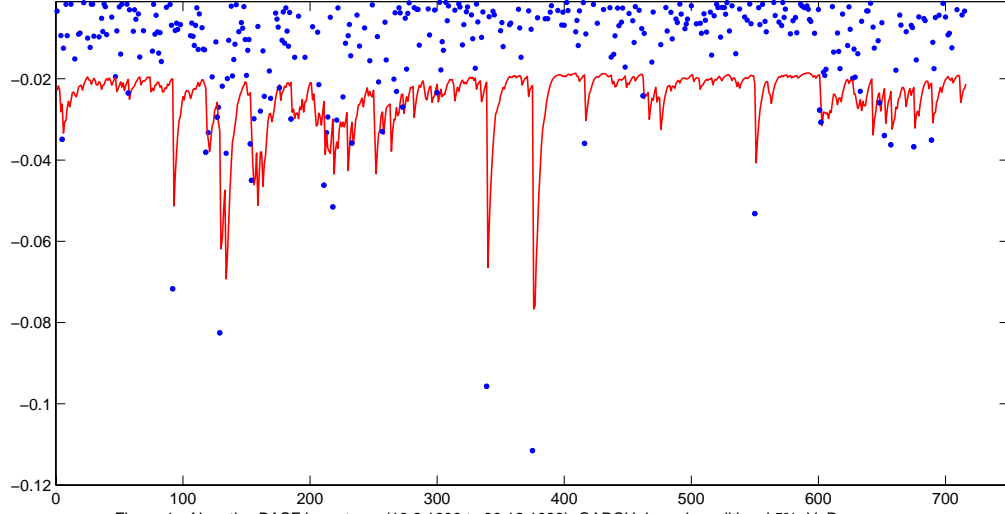
The figures show only data starting with February 13, 1990 (N=716), as only then, the exogenous variables discussed below are available. As a benchmark, we first consider the VaR calculated from fitting a GARCH(1,1)-model with standard normal innovations to the data, where the model parameters are estimated by conditional maximum likelihood. Figure 4a shows the usual backtesting plot, i.e. the actual log returns Y_t (dots), where for better visibility only the negative values are plotted, and the (negative) VaR (solid line), i.e. the conditional quantile of Y_t given the last log return Y_{t-1} and the last volatility σ_{t-1}.

Figure 4b shows the corresponding backtesting plot with value at risk based on a neural network quantile estimate as described in section 4. As input, we have chosen the last log return Y_{t-1}, the corresponding log return D_{t-1} of the market index, i.e. the DAX, a 30-days moving average $M_{t-1} = \{D_{t-1} + ... + D_{t-30}\}/30$ as a local market trend indicator, and an exponentially weighted 30-days historical variance of Y_t:

$$V_{t-1} = \frac{1 - \rho}{1 - \rho^{30}} \sum_{k=1}^{30} \rho^{k-1} \big(Y_{t-k} - \bar{Y}_{t-1}\big)^2$$

with $\rho = 0.95$ and $\bar{Y}_{t-1} = \{Y_{t-1} + ... + Y_{t-30}\}/30$. The neural network used in calculating the conditional quantile estimate $q_n(Y_{t-1}, D_{t-1}, M_{t-1}, V_{t-1})$ had $H = 4$ neurons and the symmetrized logistic function (15) as activation function.

The neural network based VaR shows somewhat better than the GARCH-VaR. The empirical levels are 5.04% and 3.91% resp., i.e. the GARCH-fit leads to a rather conservative view of risk whereas the nonparametric approach leads to a rather good agreement with the nominal level 5%. Moreover, the network-based risk measure recovers much faster form the shock of an isolated extreme event in a phase of otherwise stable volatility like the Moscow coup (t=376) compared to the GARCH-procedure. The neural network based VaR has, therefore, some kind of robustness, but still reacts as fast to significant increases in volatility as the GARCH-VaR. On the other hand, an advantage of GARCH is the more stable visual appearance of the backtesting plot in Figure 4a; the nonparametric quantile estimate leads to considerably larger fluctuations of the corresponding VaR from day to day.

Figure 4a: Negative BASF log returns (13.2.1990 to 30.12.1992), GARCH–based conditional 5%–VaR

Figure 4b: Negative BASF–log returns (Feb 13, 1990, to Dec. 30, 1992) and nonparametric conditional 5%–VaR

6. Technical results and proofs

In this section we formulate some auxiliary results needed for the proof of the main Theorem 1. The first result is a variant of the Vapnik-Chervonenkis inequality (Vapnik and Chervonenkis 1971) which holds for dependent data from a stationary process. The proof can be found in (Franke and Diagne 2006).

Theorem 4. *Let $\{Z_t, -\infty < t < \infty\}$ be a $\mathbb{R}^d$-valued stationary stochastic process satisfying an α-mixing condition with exponentially decreasing mixing coefficients. Let*

G *be a set of measurable functions* $g : \mathbb{R}^d \to [0, B]$ *satisfying (A3). Then, for any* $\epsilon > 0, \ n \geq 1$

$$(18) \qquad pr\left(\sup_{g \in \mathbf{G}} \Big| \frac{1}{n} \sum_{t=1}^{n} g(Z_t) - Eg(Z_1) \Big| > \epsilon \right) \leq K_{2n}\left(\frac{\epsilon}{32}\right) c_1 e^{-c_2 \sqrt{n}\epsilon/B}$$

where $c_1, c_2 > 0$ *are some constants not depending on* n.

Lemma 1. *Let* q *denote the* α*-quantile of the real random variable* Y. *Let* $F_\varepsilon, p_\varepsilon$ *denote the distribution function and density of* $\varepsilon = Y - q$. *Then, for any* $f \in \mathbb{R}$

$$E|Y - f|_\alpha - E|Y - q|_\alpha = \int_0^{f-q} (F_\varepsilon(z) - F_\varepsilon(0))dz$$

Proof. Using $|u|_\alpha = \alpha \, u + u^-$, $F_\varepsilon(0) = \alpha$ and distinguishing the two cases $d = f - q > 0$ and $d = f - q < 0$, we get using integration by parts

$$
\begin{aligned}
E|Y - f|_\alpha \quad - \quad & E|Y - q|_\alpha = E|\varepsilon - d|_\alpha - E|\varepsilon|_\alpha \\
= \ & 1_{(0,\infty)}(d) \int_0^d (d - z)p_\varepsilon(z)dz + 1_{(-\infty,0)}(d) \int_d^0 (z - d)p_\varepsilon(z)dz \\
= \ & 1_{(0,\infty)}(d) \int_0^d (F_\varepsilon(z) - F_\varepsilon(0))dz + 1_{(-\infty,0)}(d) \int_d^0 (F_\varepsilon(0) - F_\varepsilon(z))dz. \quad \square
\end{aligned}
$$

A corresponding relation holds analogously for the conditional quantile $q(x)$ of Y given $X = x$ where $\varepsilon = Y - q(X)$, $F_\varepsilon(\cdot|x), p_\varepsilon(\cdot|x)$ denote the conditional distribution function and density of ε given $X = x$, expectation E is replaced by conditional expectation $E\{.|X = x\}$, and $f(x)$ is an arbitrary function in $L^1(\mu)$.

Theorem 5. *Let* $(Y_t, X_t), -\infty < t < \infty$, *be a stationary time series with* $Y_t \in \mathbb{R}, X_t \in \mathbb{R}^d$ *satisfying assumption (A2). Let* $E|Y_t| < \infty$, *and let* μ *denote the stationary distribution of* X_t.

Let $\mathbf{F}_n \subset L^1(\mu), n \geq 1$, *be increasing classes of functions* $f : \mathbb{R}^d \to \mathbb{R}$ *such that* $\mathbf{F}_\infty = \bigcup_{n=1}^{\infty} \mathbf{F}_n$ *is dense in* $L^1(\mu)$. *Let* $q_n \in \mathbf{F}_n$ *denote the regression quantile given by (7), and* $\hat{q}_n = T_{\Delta_n} q_n$ *the truncated version for some sequence* $\Delta_n > 0, \lim_{n \to \infty} \Delta_n = \infty$. *Let* $\widehat{\mathbf{F}}_n = \{\hat{f}_n = T_{\Delta_n} f; f \in \mathbf{F}_n\}$. *Assume furthermore*

$$(19) \qquad \lim_{n \to \infty} \inf_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} \int |f(z) - m(z)|\mu(dz) = 0.$$

a) If for all $L > 0$

$$(20) \qquad \lim_{n \to \infty} E \sup_{f \in \widehat{\mathbf{F}}_n} \Big| \frac{1}{n} \sum_{t=1}^{n} |T_L Y_t - f(X_t)|_\alpha - E|T_L Y_1 - f(X_1)|_\alpha \Big| = 0,$$

with $T_L Y_t$ denoting the random variable Y_t truncated at $\pm L$, then

$$\lim_{n \to \infty} E \int |\hat{m}_n(z) - m(z)| \mu(dz) = 0. \tag{21}$$

b) If there is a sequence $\delta_n \to 0$ such that for all $L > 0$

$$\frac{1}{\delta_n} \left(\frac{1}{n} \sum_{t=1}^{n} |Y_t - T_{L_n} Y_t| - E|Y_1 - T_{L_n} Y_1| \right) \to 0 \quad a.s. \tag{22}$$

$$\frac{1}{\delta_n} \sup_{f \in \widehat{\mathbf{F}}_n} \left| \frac{1}{n} \sum_{t=1}^{n} |T_{L_n} Y_t - f(X_t)|_\alpha - E|T_{L_n} Y_1 - f(X_1)|_\alpha \right| \to 0 \quad a.s. \tag{23}$$

then

$$\lim_{n \to \infty} \int |\hat{m}_n(z) - m(z)| \mu(dz) = 0 \quad a.s.$$

Proof. We use Lemma 1 and assumption (A2) to relate $||\hat{q}_n - q||_1$ to $E|Y - \hat{q}_n(X)|_\alpha - E|Y - q(X)|_\alpha$ where, here, E is taken conditional on the data, i.e. $\hat{q}_n(x)$ is given. In the first part of the proof we bound this term from above by terms converging to 0.

i) By definition of q as conditional quantile function we have

$$
\begin{aligned}
0 \;\leq\; & E|Y - \hat{q}_n(X)|_\alpha - E|Y - q(X)|_\alpha \\
=\; & E|Y - \hat{q}_n(X)|_\alpha - \inf_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} E|Y - f(X)|_\alpha \\
& + \inf_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} E|Y - f(X)|_\alpha - E|Y - q(X))|_\alpha \\
\leq\; & \sup_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} \{ E|Y - \hat{q}_n(X)|_\alpha - E|Y - f(X)|_\alpha \} + \inf_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} E|q(X) - f(X)|_\alpha \\
\leq\; & \sup_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} \{ E|Y - \hat{q}_n(X)|_\alpha - E|Y - f(X)|_\alpha \} + \inf_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} ||f - q||_1
\end{aligned}
$$

where we have used the triangular inequality for $|u|_\alpha$ and $|u|_\alpha \leq |u|$. For a yet arbitrary $L \leq \Delta_n$, let Y_L, Y_{tL} denote $T_L Y, T_L Y_t$. We decompose the first term on the right-hand side.

$$
\sup_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} \{ E|Y - \hat{q}_n(X)|_\alpha - E|Y - f(X)|_\alpha \}
$$

$$\leq \sup_{f \in \mathbf{F}_n, ||f||_\infty \leq \Delta_n} \Big\{ E|Y - \hat{q}_n(X)|_\alpha - E|Y_L - \hat{q}_n(X)|_\alpha \tag{24}$$

$$+ \; E|Y_L - \hat{q}_n(X)|_\alpha - \frac{1}{n} \sum_{t=1}^{n} |Y_{tL} - \hat{q}_n(X_t)|_\alpha \tag{25}$$

$$+ \; \frac{1}{n} \sum_{t=1}^{n} |Y_{tL} - \hat{q}_n(X_t)|_\alpha - \frac{1}{n} \sum_{t=1}^{n} |Y_{tL} - q_n(X_t)|_\alpha \tag{26}$$

$$+ \; \frac{1}{n} \sum_{t=1}^{n} |Y_{tL} - q_n(X_t)|_\alpha - \frac{1}{n} \sum_{t=1}^{n} |Y_{tL} - q_n(X_t)|_\alpha \tag{27}$$

$$(28) \qquad + \quad \frac{1}{n}\sum_{t=1}^{n}|Y_t - q_n(X_t)|_\alpha - \frac{1}{n}\sum_{t=1}^{n}|Y_t - f(X_t)|_\alpha$$

$$(29) \qquad + \quad \frac{1}{n}\sum_{t=1}^{n}|Y_t - f(X_t)|_\alpha - \frac{1}{n}\sum_{t=1}^{n}|Y_{tL} - f(X_t)|_\alpha$$

$$(30) \qquad + \quad \frac{1}{n}\sum_{t=1}^{n}|Y_{tL} - f(X_t)|_\alpha - E|Y_L - f(X)|_\alpha$$

$$(31) \qquad + \quad E|Y_L - f(X)|_\alpha - E|Y - f(X)|_\alpha \Big\}.$$

By definition of q_n, (28) is bounded from above by 0. As $L \le \Delta_n$ and as $Y_{tL}, \hat{q}_n(X_t) \le \Delta_n$, we immediately get $|Y_{tL} - \hat{q}_n(X_t)|_\alpha - |Y_{tL} - q_n(X_t)|_\alpha \le 0$, and (26) is bounded from above by 0, too. By definition of $\hat{q}_n$ and $\hat{\mathbf{F}}_n$, (25) and (30) both are bounded by

$$\sup_{f \in \hat{\mathbf{F}}_n} \left| \frac{1}{n}\sum_{t=1}^{n}|Y_{tL} - f(X_t)|_\alpha - E|Y_L - f(X)|_\alpha \right|.$$

Again using the triangular inequality for $|u|_\alpha$ and $|u|_\alpha \le |u|$, (24), (31) are bounded by

$$\frac{1}{n}\sum_{t=1}^{n}|Y_t - Y_{tL}|.$$

Therefore, we have

$$\begin{aligned} 0 \quad &\le \quad E|Y - \hat{q}_n(X)|_\alpha - E|Y - q(X)|_\alpha \\ &\le \quad 2\sup_{f \in \hat{\mathbf{F}}_n}\left|\frac{1}{n}\sum_{t=1}^{n}|Y_{tL} - f(X_t)|_\alpha - E|Y_L - f(X)|_\alpha\right| \end{aligned}$$

$$(32) \qquad + \inf_{f \in \mathbf{F}_n, ||f||_\infty \le \Delta_n}||f - q||_1 + 2\,E|Y - Y_L| + \frac{2}{n}\sum_{t=1}^{n}|Y_t - Y_{tL}|.$$

ii) By Lemma 5.1, applied to conditional quantiles and expectations, we have for any $f \in L^1(\mu)$, $0 \le \delta \le 1$,

$$\begin{aligned} E|Y - f(X)|_\alpha &- E|Y - q(X)|_\alpha = E\big[E\{|Y - f(X)|_\alpha|X\} - E\{|Y - q(X)|_\alpha|X\}\big] \\ &= E\Big[1_{(0,\infty)}(f(X) - q(X))\int_0^{f(X)-q(X)}(F_\varepsilon(z|X) - F_\varepsilon(0|X))\,dz \\ &\quad + 1_{(-\infty,0)}(f(X) - q(X))\int_{f(X)-q(X)}^0(F_\varepsilon(0|X) - F_\varepsilon(z|X))\,dz\Big] \\ &\ge E\Big[\pi(X)1_{(0,\infty)}(f(X) - q(X))\int_0^{f(X)-q(X)}\min(\delta\zeta(X), z)\,dz \\ &\quad + \pi(X)1_{(-\infty,0)}(f(X) - q(X))\int_{f(X)-q(X)}^0\min(\delta\zeta(X), -z)\,dz\Big] \end{aligned}$$

as, by assumption (A2a)

$$F_\varepsilon(z|X) - F_\varepsilon(0|X) = \int_0^z p_\varepsilon(z|X)dz \geq \begin{cases} \pi(X)z & \text{for} \quad 0 \leq z \leq \delta\zeta(X) \\ \pi(X)\delta\zeta(X) & \text{for} \quad \delta\zeta(X) \leq z \end{cases}$$

and, analogously, $F_\varepsilon(0|X) - F_\varepsilon(z|X) \geq \pi(X)\min(\delta\zeta(X), -z)$ for $z \leq 0$.

Using (A2b) and distinguishing the cases $f(X) - q(X) > \delta\zeta(X)$, $f(X) - q(X) < -\delta\zeta(X)$ and $|f(X) - q(X)| \leq \delta\zeta(X)$, we get

$$\begin{aligned}
E|Y - f(X)|_\alpha - E|Y - q(X)|_\alpha \quad \geq \quad & \delta\ \gamma_0\ E(f(X) - q(X) - \delta\zeta(X))^+ \\
& + \delta\ \gamma_0\ E(q(X) - f(X) - \delta\zeta(X))^+ \\
& + \frac{1}{2}\ E\ \pi(X)|f(X) - q(X)|^2 \cdot 1_{[0,\delta\zeta(X)]}\ (|f(X) - q(X)|) \\
\geq \quad & \delta\ \gamma_0\big[E(f(X) - q(X) - \delta\zeta(X))^+ + E(q(X) - f(X) - \delta\zeta(X))^+\big]
\end{aligned}$$

Replacing f by $\hat{q}_n$ and taking expectations with respect to $\hat{q}_n$ we get by a monotone convergence argument for $\delta = \delta_n \to 0$ $(n \to \infty)$, as $\zeta(x) > 0$,

$$\begin{aligned}
E\ & \int |\hat{q}_n(x) - q(x)|\mu(dx) \\
= \quad & \lim_{n\to\infty} E\ \int \left[(\hat{q}_n(x) - q(x) - \delta_n\zeta(x))^+ + (q(x) - \hat{q}_n(x) - \delta\zeta(x))^-\right]\mu(dx) \\
\leq \quad & \lim_{n\to\infty} \frac{1}{\delta_n\gamma_0} E\left[E|Y - \hat{q}_n(X)|_\alpha - E|Y - q(X)|_\alpha\right] \longrightarrow 0
\end{aligned}$$

by i), by choosing $\delta_n \to 0$ slowly enough and by letting $L \to \infty$.

We get part b) of the assertion by applying the same argument without expectations using assumptions (22) and (23). $\quad\square$

Proof of Theorem 1: The assertion follows from Theorem 5 if we show that (19) and (20) resp. (22)-(23) are satisfied.

i) For arbitrary $\epsilon > 0$ there are n_1 and $f \in \mathbf{F}_{n_1}$ such that $\|f - q\|_1 \leq \epsilon$ by denseness of $\mathbf{F}_\infty \in L^1(\mu)$. As f is bounded and $\Delta_n \to \infty$, there is a $n \geq n_1$ such that $f \in \mathbf{F}_{n_1} \subseteq \mathbf{F}_n$ and $\|f\|_\infty \leq \Delta_n$. (19) follows.

ii) Let $L > 0$ be arbitrary and n large enough such that $L \leq \Delta_n$. We use as abbreviation $Z_t = (Y_t, X_t)$, $t = 1, \ldots, n$, and we set for $z = (y, x)$

$$g(z) = |T_L y - f(x)|_\alpha.$$

Let $\hat{\mathbf{G}}_n$ denote the class of such functions $g : \mathbb{R}^{d+1} \longrightarrow \mathbb{R}$ which we get if f ranges over $\hat{\mathbf{F}}_n$.

Let $\delta > 0$, $N > 0$, $z_1, \ldots, z_N \in \mathbb{R}^{d+1}$ be arbitrary. We write $z_j = (y_j, x_j)$, $j = 1, \ldots, N$. By (A3), there are f_k^*, $k = 1, \ldots, k_N(\delta)$ such that for all $f \in \hat{\mathbf{F}}_n$

$$\frac{1}{N}\sum_{j=1}^{N}|f(x_j) - f_k^*(x_j)| < \delta \qquad \text{for some } k.$$

Let $g_k^*(z) = |T_L y - f_k^*(x)|_\alpha$, $k = 1, \ldots, k_N(\delta)$, and let $g(z) = |T_L y - f(x)|_\alpha \in \hat{\mathbf{G}}_n$, i.e. $f \in \hat{\mathbf{F}}_n$. We have

$$\frac{1}{N}\sum_{j=1}^{N}|g(z_j) - g_k^*(z_j)| = \frac{1}{N}\sum_{j=1}^{N}\left| |T_L y_j - f(x_j)|_\alpha - |T_L y_j - f_k^*(x_j)|_\alpha \right|$$

$$\leq \frac{1}{N}\sum_{j=1}^{N}|f(x_j) - f_k^*(x_j)| < \delta.$$

i.e. for every δ-covering of $\hat{\mathbf{F}}_n$ we get a corresponding δ-covering of $\hat{\mathbf{G}}_n$ with the same size $k_N(\delta)$. We conclude $\tilde{K}_N(\delta) \leq K_N(\delta)$ for the minimal δ-covering numbers $\tilde{K}_N(\delta), K_N(\delta)$ of $\hat{\mathbf{G}}_n$ resp. $\hat{\mathbf{F}}_n$.

As $0 \leq g(z) \leq 2\Delta_n$ for $g \in \hat{\mathbf{G}}_n$, we have by Theorem 4 for arbitrary $\epsilon > 0$

$$\text{pr}\left\{ \sup_{f \in \hat{F}_n} \left|\frac{1}{n}\sum_{t=1}^{n}|T_L Y_t - f(X_t)|_\alpha - E|T_L Y_1 - X_1|_\alpha\right| > \epsilon \right\}$$

$$= \text{pr}\left\{ \sup_{g \in \hat{\mathbf{G}}_n} \left|\frac{1}{n}\sum_{t=1}^{n}g(Z_t) - Eg(Z_1)\right| > \epsilon \right\}$$

(33) $$\leq K_{2n}\left(\frac{\epsilon}{32}\right)c_1 e^{-c_2\sqrt{n}\epsilon/(2\Delta_n)}$$

iii) As for any nonnegative variable V and $\epsilon > 0$

$$EV = \int_0^\infty \text{pr}(V > u)\, du \leq \epsilon + \int_\epsilon^\infty \text{pr}(V > u)\, du$$

we get by (33), using that $K_N(\delta)$ is decreasing in δ,

$$E \sup_{f \in \hat{\mathbf{F}}_n} \left|\frac{1}{n}\sum_{t=1}^{n}|T_L Y_t - f(X_t)|_\alpha - E|T_L Y_1 - X_1|_\alpha\right|$$

$$\leq \epsilon + K_{2n}\left(\frac{\epsilon}{32}\right)\int_\epsilon^\infty c_1 e^{-c_2\sqrt{n}u/(2\Delta_n)}$$

$$= \epsilon + K_{2n}\left(\frac{\epsilon}{32}\right)\frac{2\Delta_n}{c_2\sqrt{n}}e^{-c_2\sqrt{n}\epsilon/(2\Delta_n)} \longrightarrow \epsilon \qquad (n \to \infty)$$

by our assumptions. For $\epsilon \to 0$, (20) follows.

iv) Under assumption (A1), $Y_t - T_L Y_t$ is α-mixing with geometrically mixing coefficients. Furthermore, it satisfies Cramér's condition by our assumption on Y_t. Therefore, by Theorem 1.6 of (Bosq 1996), we have

$$\frac{1}{\sqrt{n \log n \log \log n}} \sum_{t=1}^{n} (|Y_t - T_L Y_t| - E|Y_1 - T_L Y_1|) \to 0 \qquad \text{a.s.,}$$

and (22) is satisfied for any $\delta_n = O(\sqrt{\log n \log \log n}/\sqrt{n})$.

v) Again by (33) we have for any $\epsilon > 0$

$$\sum_{n=1}^{\infty} \mathrm{pr} \left(\frac{1}{\delta_n} \sup_{f \in \hat{\mathbf{F}}_n} \left| \frac{1}{n} \sum_{t=1}^{n} |T_L Y_t - f(X_t)|_\alpha - E|T_L Y_1 - f(X_1)|_\alpha \right| > \varepsilon \right)$$

$$\leq \sum_{n=1}^{\infty} K_{2n} \left(\frac{\varepsilon \delta_n}{32} \right) c_1 e^{-c_2 \sqrt{n} \epsilon \delta_n / (2\Delta_n)}$$

$$= \sum_{n=1}^{\infty} c_1 \exp\left\{ -n^\beta \frac{n^{\frac{1}{2}-\beta}}{\Delta_n} \left(\frac{c_2 \epsilon \delta_n}{2} - \frac{\kappa_n(\epsilon \delta_n)\Delta_n}{\sqrt{n}} \right) \right\} < \infty$$

as, by our assumptions, $n^{\frac{1}{2}-\beta}\delta_n/\Delta_n \longrightarrow \infty$ and $\kappa_n(\epsilon \delta_n)\Delta_n/\sqrt{n} \longrightarrow 0$ for $n \to \infty$. (23) follows from the Borel-Cantelli-Lemma. $\square$

To prove Theorem 2 we need an auxiliary result which is closely related to Lemma 9.3 of (Györfy et $al.$ 2002) and proven in an analogous manner. Let $\mathbf{G}$, $\delta > 0$, $z_1, \ldots, z_N \in \mathbb{R}^d$ and $g_1^*, \ldots, g_{k_N}^*$ be as in assumption (A3). For given $\mathbf{z} = (z_1, \ldots, z_N)$, let $K_N(\delta, \mathbf{G}, \mathbf{z})$ denote the smallest value of k_N such that for any $g \in \mathbf{G}$ there is a $k \leq k_N$ with $\frac{1}{N} \sum_{j=1}^{N} |g(z_j) - g_k^*(z_j)| < \delta$.

Lemma 2. *Let* $A_1, \ldots, A_H$ *be disjoint subsets of* $\mathbb{R}^d$, *and let* $\mathbf{P}(H)$ *be the class of corresponding simple functions given by (11). For arbitrary* $r > 0$, $N \geq 1$, $z_1, \ldots, z_N \in \mathbb{R}^d$ *let*

$$\mathbf{G} = \{f \in \mathbf{P}(H); \frac{1}{N} \sum_{i=1}^{N} |f(z_i)| \leq r\}.$$

Then, $K_N(\delta, \mathbf{G}, \mathbf{z}) \leq (1 + 4r/\delta)^H$.

Proof. For $f \in \mathbf{P}(H)$, $c \in \mathbb{R}^H$ we write

$$\|f\|_N = \frac{1}{N} \sum_{j=1}^{N} |f(z_j)|, \quad \|c\| = \sum_{j=1}^{H} |c_j|.$$

Set $d_j = \|1_{A_j}\|_N$, and let D denote the diagonal matrix with entries $d_1, \ldots, d_H$. As $1_{A_j}(z_i)$, $j = 1, \ldots, H$, vanish except for at most one j, we have for $f = \sum c_j 1_{A_j} \in \mathbf{G}$

$$\|f\|_N \;=\; \frac{1}{N}\sum_{i=1}^{N}|\sum_{i=1}^{H} c_j\,1_{A_j}(z_i)| = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{H}|c_j|1_{A_j}(z_i)$$

(34)
$$=\; \sum_{j=1}^{H}|c_j|\,\|1_{A_j}\|_N = \|Dc\|$$

Let $f_1,\ldots,f_K \in \mathbf{G}$ be an arbitrary δ-packing of $\mathbf{G}$, i.e. $\|f_k - f_l\|_N \geq \delta$ for all $1 \leq k < l \leq K$. From Lemma 9.2 of (Györfy *et al.* 2002), it suffices to show $K \leq (1 + 4r/\delta)^H$. As $f_k \in \mathbf{G}$, we have, using (34), for $1 \leq k < l \leq K$

$$f_k = \sum_{j=1}^{H} a_j^{(k)} 1_{A_j}, \;\; \|Da^{(k)}\| = \|f_k\|_N \leq r, \;\; \|Da^{(k)} - Da^{(l)}\| = \|f_k - f_l\|_N \geq \delta.$$

Therefore, by the triangular inequality, the sets

$$B_k = \{u \in \mathbb{R}^H; \; \|u - Da^{(k)}\| \leq \frac{\delta}{4}\}, \;\; k = 1,\ldots,K$$

are disjoint subsets of $\{u \in \mathbb{R}^H; \|u\| \leq r + \frac{\delta}{4}\}$, and we have for the volumes

$$K\,c_H(\frac{\delta}{4})^H \leq c_H(r + \frac{\delta}{4})^H$$

where c_H denotes the volume of the l^1-unit ball $\{u, \|u\| \leq 1\}$ in $\mathbb{R}^H$. $\qquad\square$

Proof of Theorem 2: It only remains to check the rate conditions of Theorem 1. As $A_1,\ldots,A_{H_n}$ are disjoint we have for all $c_1,\ldots,c_{H_n} \in \mathbb{R}$

$$T_{\Delta_n}\left(\sum_{j=1}^{H_n} c_j 1_{A_j}\right) = \sum_{j=1}^{H_n} T_{\Delta_n} c_j 1_{A_j} \in \{f \in \mathbf{P}(H_n); \; \|f\|_\infty \leq \Delta_N\}.$$

Therefore, for any $z_1,\ldots,z_{2n} \in \mathbb{R}^d$, $\mathbf{z} = (z_1,\ldots,z_{2n})$, we have

$$\hat{\mathbf{F}}_n = T_{\Delta_n}\mathbf{P}(H_n) \subseteq \{f \in \mathbf{P}(H_n); \frac{1}{2n}\sum_{j=1}^{2n}|f(z_j)| \leq \Delta_n\} = \mathbf{G}.$$

From Lemma 2 we conclude

$$K_{2n}(\frac{\epsilon}{32}) \leq K_{2n}(\frac{\epsilon}{32}, \mathbf{G}, \mathbf{z}) \leq (1 + \frac{128\Delta_n}{\epsilon})^{H_n},$$

or $\kappa_n(\epsilon) \leq H_n \log(1 + 128\Delta_n/\epsilon)$. a) follows.

For b), choose $\delta_n \to 0$ such that $\delta_n^{-1} = O(\Delta_n^\gamma)$ with $0 < \gamma < \beta$. Setting $\beta' = (\beta - \gamma)/2$, we have $\Delta_n(\delta_n n^{\frac{1}{2}-\beta'}) \longrightarrow 0$ and, using the same type of upper bound for $\kappa_n(\epsilon\delta_n)$ as above, we also get $\Delta_n\kappa_n(\epsilon\delta_n)/\sqrt{n} \longrightarrow 0$. $\qquad\square$

Proof of Theorem 3: By (A5), we have $|q_n(z)| \leq \Delta_n$, and q_n and the truncated estimate $\hat{q}_n$ given by (8) coincide in this case. Therefore, we only have to check the assumptions of Theorem 1. By (16.19) in the proof of Theorem 16.1 of Györfy et al. (Györfy *et al.* 2002), $\mathbf{F}_n$ satisfies (A3) with

$$K_{2n}\left(\frac{\epsilon}{32}\right) = e^{\kappa_n(\epsilon)} \leq \left(\frac{12\ e\ \Delta_n(H_n+1)}{\epsilon}\,32\right)^{(2d+5)H_n+1}$$

with $\kappa_n(\epsilon) \leq \{(2d+5)H_n + 1\}\log(384\ e\ \Delta_n(H_n+1)/\epsilon)$. Neglecting constant factors and terms of smaller order, a) follows immediately from Theorem 1.

For b), choose $\delta_n \to 0$ such that $\delta_n^{-1} = O((\Delta_n H_n)^\gamma)$ with $0 < \gamma < \beta$. Setting $\beta' = (\beta - \gamma)/2$, we have $\Delta_n/(\delta_n n^{1/2-\beta'}) \to 0$, and using the same upper bound on the log covering number $\kappa_n(\epsilon\delta_n)$ as for showing a), the other rate condition of Theorem 1 b) follows too. $\square$

REFERENCES

Artzner, P., Delbaen, F., Eber, F.-J., Heath, D., 1997. Thinking Coherently, Risk Magazine.

Bollerslev, T.P., 1986. Generalized autoregressive conditional heteroscedasticity, Journal of Econometrics 31, 307-327.

Bosq, D., 1996. Nonparametric Statistics for Stochastic Processes, Lecture Notes in Statistics 110 (Springer-Verlag, Heidelberg).

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987-1008.

Engle, R.F., Manganelli, S., 2002. CAViaR: Conditional autoregressive value at risk by regression quantiles, to appear in Journal of Business and Economic Statistics.

Franke, J., 1998. Nonlinear and nonparametric methods for analyzing financial time series, in: P. Kall and H.-J. Luethi, eds., Operation Research Proceedings 98 (Springer-Verlag, Heidelberg).

Franke, J., 2000. Portfolio management and market risk quantification using neural networks, in: W.S. Chan, W.K. Li and H. Tong, eds., Statistics and Finance: An Interface (Imperial College Press, London).

Franke, J., Mwita, P., 2003. Nonparametric estimates for conditional quantiles of time series, Report in Wirtschaftsmathematik, University of Kaiserslautern.

Franke, J., Diagne, M., 2006. Estimating market risk with neural networks, Statistics and Decisions 24, 233-253.

Franke, J., Neumann, M.H., Stockis, J.P., 2004. Bootstrapping nonparametric estimates of the volatility function, Journal of Econometrics 118, 189-218.

Glosten, L., Jagannathan, R., Runkle, D., 1993. Relationship between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48, 1779-1801.

Gouriéroux, C., Montfort, A., 1992. Qualitative Threshold ARCH Models, Journal of Econometrics 52, 159-200.

Grenander, U., 1981. Abstract Inference, (Wiley, New York).

Györfy, L., Kohler, M., Krzyzak, A., Walk, H., 2002. A Distribution-Free Theory of Nonparametric Regression (Springer-Verlag, Heidelberg).

Härdle, W., Tsybakov, A.B., 1997. Local polynomial estimation of the volatility function, Journal of Econometrics 81, 223-242.

Hafner, C.M., 1998. Nonlinear Time Series Analysis with Applications to Foreign Exchange Volatility, (Physica-Verlag, Heidelberg).

Hornik, K., 1991. Approximation Capabilities of Multilayer Feedforward Networks, Neural Networks 4, 251-257.

Koenker, R., Bassett, G., 1978. Regression quantiles, Econometrica 46, 33-50.

Pollard, D., 1984. Convergence of Stochastic Processes, (Springer-Verlag, Heidelberg).

Rabemananjara, R., Zakoian, J.M., 1993. Threshold ARCH Models and asymmetries in volatility, Journal of Applied Econometrics 8, 31-49.

Vapnik, V.N., Chervonenkis, A.Y., 1971. On the uniform convergence of relative frequencies of events to their probabilities, Theory of Probability and its Applications 16, 264-280.

White, H., 1990. Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings, Neural Networks 3, 535-549.

SFB 649 Discussion Paper Series 2007

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Trade Liberalisation, Process and Product Innovation, and Relative Skill Demand" by Sebastian Braun, January 2007.

002 "Robust Risk Management. Accounting for Nonstationarity and Heavy Tails" by Ying Chen and Vladimir Spokoiny, January 2007.

003 "Explaining Asset Prices with External Habits and Wage Rigidities in a DSGE Model." by Harald Uhlig, January 2007.

004 "Volatility and Causality in Asia Pacific Financial Markets" by Enzo Weber, January 2007.

005 "Quantile Sieve Estimates For Time Series" by Jürgen Franke, Jean-Pierre Stockis and Joseph Tadjuidje, February 2007.