African Journal of Science and Technology (AJST)
Science and Engineering Series Vol. 6, No. 1, pp. 50 - 58

CONDITIONAL SCALE FUNCTION ESTIMATE IN THE PRESENCE OF UNKNOWN CONDITIONAL QUANTILE FUNCTION

Mwita, P.N. and Otieno, R.O.

Department of Mathematics and Statistics, Jomo Kenyatta University of Agriculture & Technology, P.O. Box 62,000 Nairobi-00200, KENYA

ABSTRACT:- *Standard approach for modeling and understanding the variability of statistical data or, generally, dependant data, is often based on the mean variance regression models. However, the assumptions employed on standardized residuals may be too restrictive, in particular, when the data follows heavy-tailed distribution with probably infinite variance. This paper considers the problem of nonparametric estimation of conditional scale function of time series, based on quantile regression methodology of Koenker and Bassett (1978). We use a flexible model introduced in Mwita (2003), that makes no moment assumptions, and discuss an estimate which we get by inverting a kernel estimate of the conditional distribution function. We finally prove the consistency and asymptotic normality for the estimate.*

Key word and phrases. Conditional quantile, kernel estimate, quantile autoregression, ARCH, QARCH, time series, consistency, asymptotic normality, value-at-risk.

AMS subject classification: 62G05, 62G20, 62M10

INTRODUCTION

Let $\{V_t, t \in Z\}$ be a stationary and α-mixing multivariate time series adapted to the sequence $F_t, -\infty < t < \infty$, of σ-algebras. Partition it as $V_t = (Y_t, X_t)$ where the real-valued response variable $Y_t \in R$ is F_t -measurable and the covariate $X_t \in R^d$ is F_{t-1}-measurable. For $0 < \theta < 1$, we want to estimate the conditional scale function of Y_t given the pasts F_{t-1} assuming that it is completely determined by X_t, i.e. we have

$$Y_t = \mu_\theta(X_t) + \sigma_\theta(X_t) Z_t \quad (1.1)$$

where $\mu_\theta(X_t)$ is the conditional θ-quantile of Y_t given X_t. The function $\sigma_\theta(X_t)$ is the conditional scale function of Y_t given X_t. This function is a product of a constant and a variable, i.e., $\sigma_\theta(X_t) = b\sigma(X_t)$ where $\sigma(X_t)$ is the so called conditional volatility, see Bollerslev et al. (1994), Shephard (1996) among others for review of models containing such functions and their many variants and *b*, a positive constant independent of time but depends on θ. The standardized residuals Z_t are assumed independent and identically distributed (i.i.d.) with zero θ-quantile and unit scale. The conditional functions $\mu_\theta(X_t)$ and $\sigma_\theta(X_t)$ may be rather arbitrary, apart from some regularly assumptions, and we want to estimate the conditional scale function, $\sigma_\theta(X_t)$, nonparametrically, given that $\mu_\theta(X_t)$ is unknown.

The model (1.1) includes the case of a nonparametric quantile-scale regression where $(Z_t, X_t), -\infty < t < \infty$ are i.i.d. as well as the quantile autoregressive-quantile autoregressive conditional heteroscedastic (QAR-QARCH), introduced in Mwita (2003). The corresponding QAR-QARCH of order *d* takes the form

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

where $X_t = (Y_{t-1}, \ldots, Y_{t-d})$ is just part of the univariate time series Y_t. If we choose $X_t = (Y_{t-1}, \ldots, Y_{t-d}, U_{t-1})$, where the random vector U_t consists of observations from other time series than Y_t available at time *t*, then (1.1), would become a quantile autoregressive-quantile autoregressive conditional heteroscedastic model with exogeneous components. Two main applications we have in mind are a flexible procedure for estimating indicators for financial market volatility as well as for use in the calculation of extreme value-at-risk in a heavy-tailed financial time series data, compare for example, Jorion (2000). In the latter, the QARCH function estimate could act as a link between the estimations in the interior and extreme parts of data, as discussed in Mwita (2003), see also, McNeil and Frey (2000).

Considering other financial time series models, (1.1) can be seen as a robust 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(d)-ARCH(d)-type,

$$Y_t = \mu(X_t) + \sigma(X_t) e_t, \quad t = 1, 2, \ldots \qquad (1.2)$$

where $X_t = (Y_{t-1}, \ldots, Y_{t-d})$, μ and α are are arbitrary and $\{e_t\}$ is a sequence of i.i.d. random variables with mean 0 and variance 1. Then (1.2) can be written in the form (1.1) with $\mu_\theta(X_t) = \mu(X_t) + \sigma(X_t) q_\theta^e$, $\sigma_\theta(X_t) = \sigma(X_t) M_\theta^e$ and $Z_t = (e_t - q_\theta^e)(M_\theta^e)^{-1}$, where q_θ^e ans M_θ^e are θ-quantiles of e_t and $M_\theta(e_t, q_\theta^e)$ respectively, and

$M_\theta(Z, \mu) = (Z - \mu)(\theta - I_{\{Z-\mu \le 0\}})$, being a function of any real random variable Z with distribution function F_Z and a real value $\mu \in R$, is the asymmetric absolute value function whose amount of asymmetry depends on θ, see Koenker and Bassett (1978). When Y_t is symmetric and θ=0.5, then $2M_\theta(Z, \mu)$ is an absolute value function and $\sigma_{0.5}(X_t)$ is the conditional median absolute deviation (CMAD) of Y_t. When $\mu(X_t) = 0$ in (1.2), we have a purely heteroscedastic ARCH model introduced in Engle (1982) and $\mu_\theta(X_t)$ for $\theta > 0.5$, in this particular case, can be regarded as a conditional scale function at θ-level.

The concept of scales is well discussed in Huber (1981) and conditional scale models in the case of heteroscedastic regression with independent variables, in Welsh et al. (1994) and Welsh (1996). Because quantiles are readily interpretable in location-scale models and are robustly estimable than moments, Koenker and Zhao (1996) has exploited regression quantiles idea of Koenker and Bassett (1978) to ARCH settings. Instead of modeling conditional variance, it focuses on ARCH models for conditional scale, where the standardized errors are assumed to be i.i.d. random variables with mean zero and finite variance.

The nonparametric estimation of $\mu_\theta(X_t)$ in models such as (1.1) has been carried out in among others, Franke and Mwita (2003), which gives the uniform strong consistency properties of the estimator. The estimation of $\sigma_\theta(X_t)$, under the assumption that $\mu_\theta(X_t)$ is known, has been carried out in Mwita (2004), which also gives the asymptotic properties of the estimator. Based on model (1.1) and given that an estimator of $\mu_\theta(X_t)$ is $\hat{\mu}_\theta(X_t)$, we get a nonparametric estimator of $\sigma_\theta(X_t)$ directly by first estimating the conditional distribution function of $M_\theta(Y_t, \hat{\mu}_\theta(X_t))$ given X_t and then inverting it. We use a kernel estimate of Nadaraya

(1964) and Watson (1964) type for the conditional distribution. Apart from the disadvantages of not being adaptive and having some boundary effects, which can be fixed anyhow (see Hall et al., 1999), 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.

In the following section, we propose an estimate of the QARCH function in (1.1), when $\mu_\theta(X_t)$ is unknown, and derive its consistency and asymptotic normality properties which are important for inferences. The technical results and proofs are postponed to the third section.

QARCH FUNCTION ESTIMATE AND ITS ASYMPTOTIC PROPERTIES

We propose an estimate of the QARCH function in model (1.1) and establish the weak consistency and asymptotic normality. For that purpose, express (1.1) as

$$M_\theta(Y_t, \mu_\theta(X_t)) = \sigma_\theta(X_t) + \sigma_\theta(X_t)(M_\theta(Z_t, 0) - 1) \quad (2.1)$$

and observe that (2.1) is again of the form (1.1), now with the QAR, $\mu_\theta(X_t) = \sigma_\theta(X_t) \in R_+$ and $Z_t = M_\theta(Z_t, 0) - 1$. If $\mu_\theta(X_t)$ in model (2.1) is known, we could consider (2.1) as a quantile autoregressive model with errors not necessarily independent, and employ same procedure as in Mwita (2004) to estimate $\sigma_\theta(X_t)$. That is we find the function σ such that the following equation is simultaneously satisfied,

$$P(Y_t \le \mu_\theta(X_t) | X_t = x) = P(M_\theta(Y_t, \mu_\theta(X_t)) \le \sigma | X_t = x) = \theta \quad (2.2)$$

Since $\mu_\theta(X_t)$ is unknown, our first task involves estimating it. Then we compute the quantile residuals and pass them through the loss function M_θ as in (2.1) and finally, estimate the QARCH function.

The kernel estimates of the autoregressive function $\mu_\theta(X_t)$ at point x based on a sample $(Y_t, X_t), t = 1, \ldots, n$ from model (1.1), is obtained in two steps. In the first one, we have to estimate the conditional distribution function,

$$F_x(y) = P(Y_t \le y | X_t = x) = E\left[I_{\{Y_t \le y\}} | X_t = x\right], \quad (2.3)$$

of Y_t given $X_t = x$, which can be written as the conditional expectation of $I_{\{Y_t \le y\}}$ and, therefore, may be estimated by the standard Nadaraya-Watson kernel estimate

$$\hat{F}_x(y) = \frac{\sum_{t=1}^n K_h(x - X_t) I_{\{Y_t \le y\}}}{\sum_{t=1}^n K_h(x - X_t)} \quad (2.4)$$

Here, $K(u)$ is a d-dimensional kernel and $K_h(u) = h^{-d} K(u/h)$ is the rescaled kernel. For any $\theta \in (0,1)$, the QAR function $\mu_\theta(x)$ is given by $\mu_\theta(x) = \inf\{y \in R | F_x(y) \ge \theta\}$.

Therefore, we estimate $\mu_\theta(x)$ by the following kernel estimator

$$\hat{\mu}_\theta(x) = \inf\{y \in R | \hat{F}_x(y) \ge \theta\} \equiv \hat{F}_x^{-1}(\theta) \quad (2.5)$$

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

Let $R_t = M_\theta(Y_t, \mu_\theta(X_t))$ be the true residuals, when $\mu_\theta(X_t)$ is known, and denote the conditional distribution function of R_t given X_t as $F_x(r)$, with r being a fixed value on R_+. The true conditional scale function of Y_t given X_t can be approximated locally by

$$\sigma_\theta(x) = \inf\{r \in R_+ | F_x(r) \ge \theta\} \quad (2.6)$$

and its kernel estimate obtained as

$$\hat{\sigma}_\theta(x) = \inf\{r \in R_+ | \hat{F}_x(r) \ge \theta\} \quad (2.7)$$

with $\hat{F}_x(r)$ being an estimator for $F_x(r)$. The asymptotic properties for estimators (2.5) and (2.7) are given in Franke and Mwita (2003) and Mwita (2004), respectively.

When $\mu_\theta(X_t)$ is unknown in model (1.1), we get the estimated quantile residuals as $(Y_t - \hat{\mu}_\theta(X_t))$ and the estimated transformed residuals as $\hat{R}_t = M_\theta(Y_t, \hat{\mu}_\theta(X_t))$. We therefore define the conditional distribution function of $\hat{R}_t$ given X_t as

$$\hat{F}_x(r') = \frac{\sum_{t=1}^{n} K_h(x - X_t) I_{\{\hat{R}_t \le r'\}}}{\sum_{t=1}^{n} K_h(x - X_t)} \qquad (2.8)$$

with r' being a fixed-real value on R+ in the neighborhood of *r*. Therefore we propose to estimate $\sigma_\theta(x)$, in this case, by the following kernel estimate,

$$\hat{\sigma}_\theta(x) = \inf\{r' \in R_+ \,|\, \hat{F}_x(r') \ge \theta\} \qquad (2.9)$$

For sake of simplicity, we have assumed that the bandwidth *h* is the same in all directions, but we could generalize our results in a straightforward manner to vectors $(h_1, \ldots, h_d)^T$ of bandwidths. For our asymptotic considerations, we have to assume that the time series (Y_t, X_t) satisfies appropriate mixing conditions. There are a number of mixing conditions discussed in literature, for example, 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 and Tjostheim (1995,1997) have demonstrated that under some mild conditions, both ARCH processes and nonlinear additive autoregressive models with exogeneous variables are stationary and α-mixing. Thus, choosing $X_t = (Y_{t-1}, \ldots, Y_{t-d})^T$ 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 $(Y_t, X_t), (M_\theta(Y_t, \mu_\theta(X_t)), X_t)$ and $I_{\{Y_t \le y\}}$ are α-mixing as well.

The following set of assumptions are required for proving consistency and asymptotic normality of $\hat{\sigma}_\theta(x)$. Here and in the following, $g(x)$ denotes the stationary probability density of X_t.

(A1) For some compact subset *G* of R^d there are $\varepsilon > 0, \gamma > 0,$ such that $g(x) \ge \gamma$ for all x in the $\varepsilon -$neighborhood $\{x; \|x - u\| < \varepsilon$ for some $u \in G\}$.

(A2) (Y_t, X_t) is stationary and α-mixing with mixing coefficients $\alpha(n), n \ge 1,$ and there is an increasing sequence $s_n, n \ge 1,$ of positive integers such that for some finite *A*,

$$\frac{n}{s_n} \alpha^{2s_n/(3n)}(s_n) \le A, 1 \le s_n \le \frac{n}{2} \text{ for all } n \ge 1 \qquad (2.10)$$

(B1) The conditional density $f_x(\mu)$ is uniformly bounded in *x* and μ by, say, *cf*.

(B2) For the compact set *G* of (B1) and some compact neighborhood Θ_0 of 0. The set

$$\Theta = \{v = \mu_\theta(x) + \mu; x \in G, \mu \in \Theta_0\}$$

is compact too, and for some constant $c_0 > 0, f_x(v) \ge c_0$ for all $x \in G, v \in \Theta$.

(C1) The kernel $K : R^d \to R$ is a nonnegative, Lipschitz continuous function, satisfying $|K(u)| \le K_\infty$ for all u, $\int K(u)du = 1, \int uK(u)du = 0$ and $\int \|u\|^2 K(u)\,du < \infty$

(C2) For all r, x satisfying $0 < F_x(r) < 1, g(x) > 0$:

i) $F_x(r)$ and $g(x)$ are continuous and bounded in r, x

ii) $g(x)$ is twice continuously differentiable, and, for fixed *r*, $F_x(r)$ is twice continuously differentiable with respect to x, where derivatives are continuous functions of *r* and the second derivatives are Hölder-continuous in x for some $c, \beta > 0$ and all x, x', r i.e.,

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

and analogously for $g(x)$

iii) For fixed $x, F_x(r)$ has the conditional density, $f_x(r)$, which is continuous in x and Hölder-continuous in $r: \left|f_x(r) - f_x(r')\right| \le c|r - r'|^\beta$ for some $c, \beta > 0$

iv) $f_x(\sigma_\theta(x)) > 0$ for all x.

(C3) The process $\{(Y_t, X_t)\}$ is stationary and α-mixing with mixing coefficients satisfying $\alpha(n) = O\left(n^{-(2+\delta)}\right)$, for some $\delta > 0$.

Assumptions (A1)-(A2), (B1)-(B2) and (C1)-(C2) are required for proving the uniform convergence of the QAR function estimate, $\mu_\theta(x)$. This result is a pre-requisite for the investigation of the behavior of the QAR residual based on model (1.1). Assumption (A1)-(A2), (B1)-(B2) and (C1)-(C3) are required for proving the consistency and asymptotic normality of $\hat{\sigma}_\theta(x)$

Theorem 2.1. *Assume that (A1)-(A2), (B1)-(B2) and (C1)-(C3) hold. Suppose the sequence of bandwidths h>0 converge to 0 such that* $\tilde{S}_n = nh^d(s_n \log n)^{-1} \to \infty$ *for some* $s_n \to \infty$. *Let* $S_n = h^2 + \tilde{S}_n^{-\frac{1}{2}}$. *Then the QARCH function estimate is consistent,* $\hat{\sigma}_\theta(x) \to^p \sigma_\theta(x)$, *and asymptotically unbiased,*

$$E\hat{\sigma}_\theta(x) - \sigma_\theta(x) = h^2 B_\sigma(\sigma_\theta(x)) + O(S_n)\, f_x(\sigma_\theta(x)) + o(h^2) \quad (2.11)$$

where $B_\sigma(r) = -\dfrac{B(r)}{f_x(r)}$

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

$$\sqrt{nh^d}\left(\hat{\sigma}_\theta(x) - \sigma_\theta(x) - h^2 B_\sigma(\sigma_\theta(x)) - O(S_n)\, f_x(\sigma_\theta(x))\right) \to^D N\left(0, \frac{V^2(\sigma_\theta(x))}{f_x^2(\sigma_\theta(x))}\right) \quad (2.12)$$

Here, B(r) and V^2 *(r) are defined in the bias and variance expansion for the conditional distribution estimator in Lemma 3.1 of the following section.*

Proof of Theorem 2.1

In order to prove Theorem 2.1, we require the following result on uniform rate of convergence of the QAR function estimate, $\hat{\mu}_\theta(x)$, on a compact set G. The proof can be found in Franke and Mwita (2003).

Theorem 3.1. *Assume (A1)-(A2), (B1)-(B2) and (C1)-(C2). Suppose* $h \to 0$ *is a sequence of bandwidths such that* $\tilde{S}_n = nh^d(s_n \log n)^{-1} \to \infty$ *for some* $s_n \to \infty$.

Let $S_n = h^2 + \tilde{S}_n^{-\frac{1}{2}}$. *Then we have*

$$\sup_{x \in G}\left|\hat{\mu}_\theta(x) - \mu_\theta(x)\right| = O(S_n) + O\left(\frac{1}{nh^d}\right) \quad a.s. \quad (3.1)$$

S_n will be much larger than $(nh^d)^{-1}$ and therefore the rate of convergence of $\mu_\theta(x)$ will be $O(S_n)$ particularly, if the bias and variance are balanced.

The following Lemma gives the asymptotic bias and variance for $\hat{F}_x(r)$ which is a Nadaraya-Watson kernel estimate for the conditional distribution, $F_x(r)$, of $I_{\{R_t \le r\}}$ given $X_t = x$. Therefore, we omit the proof of the Lemma which follows standard lines of arguments.

Lemma 3.1 *Suppose* (C1)-(C3) *hold*. *Then*

$$E\left[\hat{F}_x(r) - F_x(r)\right] = h^2 B(r) + o(h^2) \quad (3.2)$$

$$\text{var}\left[\hat{F}_x(r)\right] = (nh^d)^{-1} V^2(r) + o((nh^d)^{-1})$$

where

$$B(r) = \frac{1}{g(x)} \nabla F_x(r)^T \int u \nabla g(x)^T u K(u) du + \frac{1}{2} \int u^T \nabla^2 F_x(r) u K(u) du$$
$$V^2(r) = \frac{1}{g(x)} (F_x(r) - F_x^2(r)) \int K^2(u) du \quad (3.3)$$

The following Lemma follows immediately from Lemma 3.1, using the smoothness assumptions on $F_x(r)$, $F_x(r)$, and a Taylor expansion of $\hat{F}_x(r)$ around r.

Lemma 3.2. *Suppose* (C1)-(C3) *hold. Then, for any* $\delta_n \to 0$, *we have*

$$\hat{F}_x(r+\delta) - \hat{F}_x(r) = \delta_n f_x(r) + o_p(\delta_n) + o_p(h^2) + o_p((nh^d)^{-1/2}) \quad (3.4)$$

We also need consistency of the kernel estimate of the density of X_t. The following Lemma, which gives the uniform rate of convergence of the Rosenblatt (1959b) - Parzen (1962) kernel estimate, $\hat{g}(x)$ for the density g(x) of X_t on the compact set G, follows immediately from the proof of Theorem 3.3.6 of Györfi et al. (1989).

Lemma 3.3. *Assume* (A1) *and* (C1)-(C3). *If, as* $n \to \infty$, *the bandwidth* $h \to 0$ *such that* $\tilde{S}_n = nh^d (s_n \log n)^{-1} \to \infty$, *then*

$$\sup_{x \in G} \left|\hat{g}(x) - g(x)\right| = O(\tilde{S}_n^{-\frac{1}{2}}) \quad a.s \quad (3.5)$$

The following Lemma shows that $M_\theta(Y_t, \mu)$ is not only convex but also continuous in $\mu \in R$.

Lemma 3.4. *Let* (y, μ) *be real –valued variables, then for all y,* $M_\theta(y,\mu)$ is *Lipschitz continuous in* μ *with Lipschitz constant 1, i.e,*

$$\left|M_\theta(y,\mu) - M_\theta(y,\mu')\right| \le \left|\mu - \mu'\right| \text{ for all } y, \mu, \mu'$$

Proof: Note that

$$M_\theta(y,\mu) - M_\theta(y,\mu') = \theta(\mu' - \mu) - ((y-\mu) I_{\{y-\mu \le 0\}} - (y-\mu') I_{\{y-\mu' \le 0\}}) \quad (3.6)$$

for $\mu < y < \mu'$, we have $I_{\{y-\mu \le 0\}} = 0$, $I_{\{y-u' \le 0\}} = 1$, and (3.6) becomes

$$M_\theta(y,\mu) - M_\theta(y,\mu') = \theta(\mu' - \mu) + (y - \mu') = (y-\mu) - (1-\theta)(\mu' - \mu) \quad (3.7)$$

For $(y - \mu') > 0$ and $(y - \mu) > 0$, the last two expressions on the right of (3.7) both imply

$$-(1-\theta)(\mu' - \mu) \le M_\theta(y,\mu) - M_\theta(y,\mu') \le \theta(\mu' - \mu),$$

and therefore $\left|M_\theta(y,\mu) - M_\theta(y,\mu')\right|$ is bounded from above by at least one of $\theta(\mu' - \mu)$ and $(1-\theta)(\mu' - \mu)$. Similarly, for $\mu \le \mu' < y$ and $y < \mu \le \mu'$, we have respectively $I_{\{y-\mu \le 0\}} = 0$ implying $M_\theta(y,\mu) - M_\theta(y,\mu') = \theta(\mu' - \mu)$ and $I_{\{y-\mu \le 0\}} = 1$, $I_{\{y-\mu' \le 0\}} = 1$ implying $M_\theta(y,\mu) - M_\theta(y,\mu') = (1-\theta)(\mu - \mu')$. Hence $\left|M_\theta(y,\mu) - M_\theta(y,\mu')\right| \le \max(\theta, 1-\theta)\left|\mu - \mu'\right| \le \left|\mu - \mu'\right|$ which immediately implies the assertion.

Proof of Theorem 2.1:

Making use of the uniform convergence result in Theorem 3.1 and the boundedness in Lemma 3.4, we express $\hat{R}_t$ in terms of R_t as $\hat{R}_t = \hat{R}_t - R_t + R_t = R_t + O(S_n)$ a.s., with the latter term being the bound for $\hat{\mu}_\theta(x) - \mu_\theta(x)$.

The indicator function $I_{\{\hat{R}_t \le r'\}}$ can then be expressed as, $I_{\{R_t + O(S_n) \le r + O(S_n)\}} = I_{\{\hat{R}_t \le r + O(S_n)\}}$, with $r' = r + O(S_n)$. We first proof that $\hat{F}_x(r')$ is a consistent estimator for $F_x(r)$. Let the bound $O(S_n) \to 0$ as $n \to \infty$, then by Lemma 3.2, we get

$$\hat{F}_x(r') - F_x(r) = \hat{F}_x(r) - F_x(r) + O(S_n) f_x(r) \quad (3.8)$$

Since the last term depends only on n, the asymptotic behavior of $\hat{F}_x(r')$ will largely depend on the behavior of $\hat{F}_x(r)$. Taking expectation and variance on both sides of (3.8) and using Lemma 3.1, we get the bias $E\left[\hat{F}_x(r') - F_x(r)\right] \approx h^2 B(r) + O(S_n) f_x(r)$, and $\text{var}\left[\hat{F}_x(r') - F_x(r)\right] \approx (nh^d)^{-1} V^2(r)$. In both the bias and variance, terms of smaller order in probability have been left out. Because the bias is of order $O(h^2) + O(S_n)$ and the variance, of order $O((nh^d)^{-1})$, the mean squared error is seen to go to zero as *n* goes to infinity. Hence $\hat{F}_x(r') \to F_x(r)$ in probability, for all $x \in R^d$ and *r*, with a rate which implies $\hat{F}_x(r')$ is consistent. To show that

$$\sqrt{nh^d}\,(\hat{F}_x(r') - F_x(r) - h^2 B(r) - O(S_n) f_x(r)) \quad (3.9)$$

is asymptotically normal, we proceed as in Theorem 2.1 in Franke and Mwita (2003), by replacing $\hat{F}_x(y)$ by $\hat{F}_x(r')$ and $\text{var}(\hat{F}_x(y))$ by $\text{var}(\hat{F}_x(r'))$. For consistency of $\hat{\sigma}_\theta(x)$, note that the Glivenko-Cantelli Theorem in Krishnaiah (1990) for strongly mixing sequences implies

$$\sup\left|F_x(r') - F_x(r)\right| \to 0, r \in R_+ \text{ in probability} \quad (3.10)$$

By the uniqueness assumption (C2 iv) on $\sigma_\theta(x)$, for any fixed $x \in R^d$, there exists an $\varepsilon > 0$ such that $\delta = \delta(\varepsilon)$

$$= \min\left\{\theta - F_x(\sigma_\theta(x) - \varepsilon), F_x\left(\sigma_\theta(x) + \varepsilon\right) - \theta\right\} > 0.$$

This implies, using the monotonicity of F_x, that

$$P\left\{\left|\hat{\sigma}_\theta(x) - \sigma_\theta(x)\right| > \varepsilon\right\}$$
$$\le P\left\{\left|F_x(\hat{\sigma}_\theta(x)) - F_x(\sigma_\theta(x))\right| > \delta\right\}$$
$$\le P\left\{\ \left|F_x(\hat{\sigma}_\theta(x)) - \hat{F}_x(\hat{\sigma}_\theta(x))\right| > \delta - \frac{2K_\infty}{\gamma nh^d}\ \right\} \quad (3.11)$$
$$\le P\left\{\ \sup_r\left|\hat{F}_x(r) - F_x(r)\right| > \delta'\ \right\}$$

for arbitrary $\delta' < \delta$ and *n* large enough. Here, because $\hat{F}_x(r)$ is a pure jump function in r with heights equal to $\frac{1}{n} K_h(x - X_t) / \hat{g}(x)$, we have used $F_x(\sigma_\theta(x)) = \theta$ and $\theta \le \hat{F}_x(\hat{\sigma}_\theta(x)) \le \theta + \dfrac{2K_\infty}{\gamma nh^d}$, which follows from Lemma 3.3, assumption (A1) and the boundedness of *K* by K_∞. Now, (3.11) tends to zero by (3.10). Hence the consistency follows.

Finally, to prove that the left hand side of (2.12) is asymptotically normally distributed with mean zero, let $b = -B(\sigma_\theta(x)) f_x^{-1}(\sigma_\theta(x)) - O(S_n) h^{-2} f_x(\sigma_\theta(x))$ and $v = V(\sigma_\theta(x)) f_x^{-1}(\sigma_\theta(x))$. Let

$$q_n(z) = P(\sqrt{nh^d}\ \frac{\hat{\sigma}_\theta(x) - \sigma_\theta(x) - h^2 b}{v} \le z)$$
$$= P(\hat{\sigma}_\theta(x) \le \sigma_\theta\left(x\right) + h^2 b + (nh^d)^{-1/2} vz)$$

As $\hat{F}_x(r')$ is increasing, but not necessarily strictly, we have

$$P\left(\hat{F}_x(\hat{\sigma}_\theta(x)) \le \hat{F}_x(\sigma_\theta(x) + h^2 b + (nh^d)^{-1/2} vz)\ \right)$$
$$\le q_n(z)$$
$$\le P\left(\hat{F}_x(\hat{\sigma}_\theta(x)) \le \hat{F}_x(\sigma_\theta(x) + h^2 b + (nh^d)^{-1/2} vz)\ \right)$$

By the same argument as in (3.11), we may replace $\hat{F}_x(\hat{\sigma}_\theta(x))$ by $F_x(\sigma_\theta(x))$ up to an error of $(nh^d)^{-1}$

at most, and we get, neglecting the $(nh^d)^{-1}$ –term which is asymptotically negligible anyhow,

$$q_n(z) \sim P(F_x(\sigma_\theta(x)) \\ \leq F_x(\sigma_\theta(x) + h^2 b + (nh^d)^{-1/2} vz)) \\ \sim P(-\delta_n f_x(\sigma_\theta(x)) \leq \hat{F}_x(\sigma_\theta(x)) - F_x(\sigma_\theta(x))) \tag{3.12}$$

with $\delta_n = h^2 b + (nh^d)^{-1/2} vz$. Here we have used Lemma 3.2 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 Yondell (1988), see also Mwita (2003) have shown that the empirical conditional distribution estimator is asymptotically normal. This follows also under similar conditions from a functional central limit theorem for $\hat{F}_x(y)$ of Abberger (1996-Corollary 5.4.1 and Lemma 5.4.1). Therefore, with $r_\theta = \sigma_\theta(x)$, we get

$$\sqrt{nh^d}\,\frac{\hat{F}_x(r_\theta + O(S_n)) - F_x(r_\theta) - h^2 B(r_\theta) - O(S_n) f_x(r_\theta)}{V(r_\theta)}$$

$$\geq \sqrt{nh^d}\,\frac{-f_x(r_\theta)\delta_n - h^2 B(r_\theta) - O(S_n) f_x(r_\theta)}{V(r_\theta)}$$

$$\sim \Phi\left(\sqrt{nh^d}\,\frac{f_x(r_\theta).(h^2 b + (nh^d)^{-1/2} vz) + h^2 B(r_\theta) - O(S_n) f_x(r_\theta)}{V(r_\theta)}\right)$$

$$= \Phi(z)$$

By our choice of *b* and *v* and our condition on the rate of *h*. this proves the theorem.

CONCLUSION

We have shown weak consistency and asymptotic normality of the nonparametric QARCH function estimate where, up to the term $O(S_n) f_x(\sigma_\theta(x))$, the form of asymptotic variance is the same as for the Nadaraya-Watson estimator for $\mu_\theta(x)$. For sake of simplicity, we have restricted ourselves to Nadaraya- Watson estimate of the conditional distribution function as the basis for the QARCH function estimates. Our results may be modified in straightforward manner to cover also the more general local polynomial estimates (Fan and Gijbels, 1996).

ACKNOWLEDGEMENTS

This research was partially supported by the Graduiertenkolleg "Mathematik und Praxis", Technische Universität Kaiserslautern, Germany and the DAAD. The authors also acknowledge the assistance provided by the Department of Mathematics and Statistics, JKUAT, Kenya.

REFERENCES

Abbeger, K. (1996). Nichtparametrische Schätzung bedingter Quantile in Zeitreihen- Mit Anwedung auf Financemarketdaten, *Hartung-Gore Verlag*, *Konstanz.*

Bollerslev, T., Engle, R. F and Nelson, D. (1994). ARCH. Models, In R.F. Engle and D.L. McFaden (eds.), *Handbook of Economic, vol. 4. Amsterdam: North-Holland*.

Bosq, D. (1996). Nonparametric statistics for Stochastic Processes: Estimation and Prediction, *Springer-Verlag*.

Doukhan, P. (1994). Mixing: Properties and Examples, Lecture notes in statistics, *Springer-Verlag*.

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

Fan, J. and I. Gijbels (1996). Local polynomial modeling and its Applications – Theory and methodologies; Monographs on Statistics and Applied Probability, Nr. 66, *Chapman and Hall*.

Franke, J. and Mwita, P. (2003). Nonparametric estimates for conditional quantiles of time series, *Technical Report in Wirtschaftsmathematik*, Nr. 87, universität Kaiserslautern, Germany.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). Nonparametric curve estimation from time series; Lecture note in statistics, *Springer-Verlag*.

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

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

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

Huber, P. J. (1981). Robust statistics, *John Wiley and sons*.

Jorion, P. (2000). Value-at- Risk. The New Benchmark for Managing Financial Risk, *McGraw-Hill, NewYork*.

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

Koenker, R. and Zhao, Q. (1996). Conditional Quantile estimation and inference for ARCH models, *Econometric Theory* **12**, 793-813.

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

Masry, E. and Tjostheim, O. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality, *Econometric Theory* **11**, 258-289.

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

McNeil, A. J. and Frey, R. (2000). Estimation of Tail-Related Risk measures for Heteroscedastic Financial Time Series: An Extreme Approach, *Journal of Empirical Finance* **7**, 271-300.

Mwita, P. (2003). Semiparametric estimation of conditional quantiles for time series with applications in finance, *PhD Thesis*, University of Kaiserslautern-Germany.

Mwita, P. (2004). On conditional scale function: Estimate and asymptotic properties, *African Diaspora Journal of Mathematics*, accepted to appear in vol. 2, No. 2, in press.

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

Parzen, E. (1962). On estimation of a probability density function and mode, *Annal of Mathematical statistics* **33**, 1065-1076.

Rosenblatt, M. (1956b). Remarks on some nonparametric estimates of a density function, *Annal of Mathematical Statistics* **27**, 832-835.

Shepherd, N. (1996). Statistical aspects of ARCH and stochastic volatility, in time series models in Econometrics, Finance and other Fields; ed. D. Cox, Hinkley and O. Barndorff-Nielen, *Chapman and Hall*, pp. 1-55.

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

Weiss, A. A. (1984). ARMA models with ARCH errors, *Journal of Time series Analysis* **3**, 129-143.

Welsh, A.H., Carrol, R. J. and Ruppert, D. (1994). Fitting heteroscedastic regression model, *Journal of the American Statistical Association* **89** (425), 100-116.

Welsh, A. H. (1996), Robust estimation of smooth regression and spread function and their derivatives, *Statistical Sinica* **6**, 347-366.