Journal of Statistical and Econometric Methods, vol.2, no.3, 2013, 175-196
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2013

Bootstrap of Kernel Smoothing in Quantile Autoregression Process

Peter N. Mwita[1] **and Jürgen Franke**[2]

Abstract

The paper considers the problem of bootstrapping kernel estimator of conditional quantiles for time series, under independent and identically distributed errors, by mimicking the kernel smoothing in non-parametric autoregressive scheme. A quantile autoregression bootstrap generating process is constructed and the estimator given. Under appropriate assumptions, the bootstrap estimator is shown to be consistent.

Mathematics Subject Classification: 62F40; 62G05; 62G08; 62G20; 62M10
Keywords: Asymptotic; Bootstrap; Conditional Quantile; Convergence; Kernel Estimate; Quantile Autoregression; Time Series; Uniform Consistency; Value-at-Risk

1 Introduction

We consider a partitioned stationary α-mixing time series (X_{t+1}, Z_t) where

[1] Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi, Kenya.
[2] Department of Mathematics, University of Kaiserslautern, D-67653 Kaiserslautern, Germany.

Article Info: *Received* : June 9, 2013. *Revised* : July 10, 2013
Published online : September 1, 2013

the real-valued response variable $X_{t+1} \in \mathbf{R}$ is $\mathbf{F}_{t+1}$-measurable and the variate $Z_t \in \mathbf{R}^p$ is $\mathbf{F}_t$-measurable. For some $0 < \theta < 1$, we are interested in estimating the conditional θ-quantile of X_{t+1} given the past $\mathbf{F}_t$, assuming that it is completely determined by Z_t. For that purpose, we consider the model

$$X_{t+1} = \mu_\theta(Z_t) + \xi_{t+1}, \quad t = 1, 2, \dots \tag{1}$$

where $\mu_\theta(Z_t)$ is the conditional quantile function providing the conditional θ-quantile of X_{t+1} given $Z_t = z$. The quantile innovations ξ_{t+1} are assumed to be independent and identically distributed (i.i.d) with θ-quantile 0. Model (1) includes the case of the quantile autoregression (QAR) of order p,

$$X_{t+1} = \mu_\theta(X_t, \dots, X_{t-p-1}) + \xi_{t+1},$$

where $Z_t = (X_t, \dots, X_{t-p-1})$ is just part of the past of the univariate time series X_{t+1}. Model (1) also includes autoregressive (AR) models:

$$X_{t+1} = \mu(X_t, \dots, X_{t-p-1}) + \epsilon_{t+1}, \quad t = 1, 2, \dots \tag{3}$$

where μ is an arbitrary autoregressive function and $\{\epsilon_{t+1}\}$ is a sequence of (i.i.d) random variables with mean 0. In the latter special case, we have $\mu_\theta(Z_t) = \mu(Z_t) + \epsilon^\theta$, where ϵ^θ is the θ-quantile of ϵ_t. Correspondingly, the quantile innovations are given by $X_{t+1} - \mu_\theta(Z_t) = \epsilon_{t+1} - \epsilon^\theta$. For Z_t independent of X_{t+1}, the parametric quantile regression model was first introduced in Koenker and Basset (1978).

Another example, leading to non identically distributed quantile residuals, is AR-ARCH processes of say, common order p:

$$X_{t+1} = \mu(X_t, \dots, X_{t-p-1}) + \sigma(X_t, \dots, X_{t-p-1})\epsilon_{t+1}, \quad t = 1, 2, \dots \tag{4}$$

where, now, $\{\epsilon_t\}$ is a normalized sequence of i.i.d random variables with mean 0 and variance 1. In this special case of (1), we have $\mu_\theta(Z_t) = \mu(Z_t) + \sigma(Z_t)\epsilon^\theta$, where again, ϵ^θ is the θ-quantile of ϵ. In this case

$$\xi_{t+1} = \sigma(Z_t)(\epsilon_{t+1} - \epsilon^\theta).$$

In model (3), Nadaraya (1964) and Watson (1964) kernel smoothing provides nonparametric estimates of μ. Assuming finite moments up to order 4, Franke and Wendel (1992) and Kreutzberger (1993) proposed and autoregression bootstrap resampling scheme that approximates the laws of kernel estimator of μ and σ. Franke et al. (2002) considered two estimators for the estimation of conditional variance, and gave consistency of the residual based and Wild bootstrap procedures for μ and σ. Although the errors in model (1), are homoscedastic, we follow similar lines as in Franke et al. (2002) for bootstraping μ. Furthermore, in this paper, properties of the estimators are derived without assuming the finiteness of the variance of X_{t+1}.

We get a nonparametric estimator of $\mu_\theta(Z_t)$ and its bootstrap version by first estimating respective conditional distribution functions and then inverting. We use the standard kernel estimate of Nadaraya-Watson type for the conditional distribution. Apart from the disadvantages of not being adaptive and having some boundary effects, which can be fixed by well-known techniques (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 present the estimator of the QAR which we are considering, and formulate a bootstrap algorithm for approximating the distribution of the estimator. In section 3, we state our main results on consistency properties of the bootstrap procedure, where the proofs are deferred to the appendix in section 5. Section 4 gives the conclusion.

2 Quantile Autoregression Bootstrap Estimator

For simplicity, we will assume that $Z_t = X_t \in \mathbf{R}$. Let $\rho_\theta(\mu) = \mu(\theta - \mathbf{I}_{\{\mu \le 0\}})$ be the so-called check function. We define the true objective function of μ as

$$Q(x,\mu) = E[\rho_\theta(X_{t+1}-\mu)|X_t = x] = E[(X_{t+1}-\mu)(\theta - \mathbf{I}_{\{X_{t+1}-\mu \le 0\}})|X_t = x]$$

The QAR function may be obtained by minimizing $Q(X_t,\mu)$ with respect to μ, i.e.,

$$\mu_\theta(x) = \arg\min_{\mu \in \mathbf{R}} Q(x,\mu) \tag{6}$$

The kernel estimator of (6) at point $X_t = x$ is obtained as

$$\hat{\mu}_\theta(x) = \arg\min_{\mu \in \mathbf{R}} \hat{Q}(x,\mu)$$

where, assuming that the data $X_1, \ldots, X_{n+1}$ are available,

$$\hat{Q}(x, \mu) = \frac{1}{n} \sum_{t=1}^{n} K_g(x - X_t) \rho_\theta(X_{t+1} - \mu)$$

and $K_g(u) = \frac{1}{g} K(\frac{u}{g})$, is 1-dimensional rescaled kernel with bandwidth $g > 0$.

The properties of $\hat{Q}(x, \mu)$ are well investigated in Mwita (2003). The objective function is strictly convex and continuous as a function of μ. This implies the minimizer, $\hat{\mu}_\theta(x)$, of $\hat{Q}(X_t, \mu)$ at point x exists uniquely and is practically the same as the solution to the following differential equation;

$$\frac{d}{d\mu} \hat{Q}(x, \mu) \approx 0 \tag{9}$$

Rearranging (9), we get the standard Nadaraya-Watson kernel estimator,

$$\begin{aligned} \widehat{F}_x(\mu) &= \frac{\sum_{t=1}^{n} K_g(x - X_t) \mathbf{I}_{\{X_{t+1} \le \mu\}}}{\sum_{t=1}^{n} K_g(x - X_t)} \\ &= \frac{(\hat{p}_g(x))^{-1}}{n} \sum_{t=1}^{n} K_g(x - X_t) \mathbf{I}_{\{X_{t+1} \le \mu\}} \\ &\approx \theta \end{aligned}$$

where $\hat{p}_g(x)$ is the kernel estimator of the density function $p(x)$ of X_t at x.

For any $\theta \in (0, 1)$, the quantile autoregression function $\mu_\theta(x)$ is given by

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

Therefore, $\mu_\theta(x)$ can be estimated by the following kernel estimator

$$\hat{\mu}_\theta(x, g) = \inf\{\mu \in \mathbf{R} | \hat{F}_x(\mu) \ge \theta\} \equiv \hat{F}_x^{-1}(\theta),$$

where $\widehat{F}_x^{-1}(\theta)$ denotes the inverse of the distribution function $\widehat{F}_x(\mu)$, which is a pure jump function of μ.

To construct the bootstrap generating process, we mimic the scheme of the autoregressive or residual-based bootstrap proposed in Franke and Wendel (1992), Kreutzberger (1993) and Franke et al.(2002). We choose the pilot estimate of $\mu_\theta(x)$, denoted by $\tilde{\mu}_\theta(x, g)$, in the following way; Let $\mathbf{1} = [-\gamma_n, \gamma_n]$ be a growing interval with $\gamma_n \to \infty$ for $n \to \infty$. The pilot estimator is chosen as $\tilde{\mu}_\theta(x, g) = \hat{\mu}_\theta(x, g) \mathbf{1}_{\{|x| \le \gamma_n\}}$ such that, outside $\mathbf{1}$, the estimator $\hat{\mu}_\theta(x, g)$ is replaced by constants. This ensures that $\hat{\mu}_\theta(x, g)$ is fairly reliable estimator for $|x|$ large.

The resampling algorithm, then, runs as follows:

Initialization:

- Calculate the pilot estimates $\tilde{\mu}_\theta(X_t), t = 1, \ldots, n$.
- Obtain the residuals as $\hat{\xi}_{t+1} = X_{t+1} - \tilde{\mu}_\theta(X_t), t = 1, \ldots, n$.
- Let $T_n = \{t \leq n; X_t \notin \mathbf{I}_n\}$; calculate the θ-sample quantile $\hat{\xi}^\theta$ of the corresponding residuals, i.e. of $\{\hat{\xi}_{t+1}, t \in T_n\}$.
- Reshift the remaining residuals to ensure that their empirical distribution has θ-quantile at 0:
$$\tilde{\xi}_{t+1} = \hat{\xi}_{t+1} - \hat{\xi}^\theta, t \in T_n.$$
- Let $\tilde{F}_n$ be the empirical distribution of $\tilde{\xi}_{t+1}, t \in T_n$.
- Smooth $\tilde{F}_n$ by convolving it with some probability density $H_b(u) = \frac{1}{b}H\left(\frac{u}{b}\right)$, where H is the density of a distribution whose θ-quantile is zero. Denote the resulting smoothed empirical law of $\tilde{\xi}_{t+1}$ as $\tilde{F}_{n,b} = \tilde{F}_n * H_b$.

Resampling:

- The bootstrap quantile residuals $\xi^*_{t+1}, t = 1, \ldots, n$ are generated as i.i.d. variables from the smoothed distribution $\tilde{F}_{n,b}$.
- The bootstrap data $X_1^*, X_2^*, \ldots, X_n^*$ are generated analogously to the QAR scheme (1).
$$X^*_{t+1} = \tilde{\mu}(X^*_t) + \xi^*_{t+1}, t = 1, 2, \ldots, n+1.$$
- From the bootstrap data, the Nadaraya-Watson estimate of the conditional distribution function $F^*_x(\mu)$ of X^*_{t+1} given $X^*_t = x$ is calculated:

$$\begin{aligned}
\hat{F}^*_x(\mu, h) &= \frac{1}{n\hat{p}^*_h(x)} \sum_{t=1}^n K_h(x - X^*_t)\mathbf{I}_{\{X^*_{t+1} - \mu \leq 0\}} \\
\text{with} & \\
\hat{p}^*_h(x) &= \frac{1}{n} \sum_{t=1}^n K_h(x - X^*_t).
\end{aligned}$$

- Finally, we get the quantile kernel estimator $\hat{\mu}^*_\theta(x,h)$, in the bootstrap world by inversion, i.e. by solving

$$\hat{F}^*_x(\hat{\mu}^*_\theta(x,h)) = \theta$$

Note that the pilot estimator of $\mu_\theta(x)$ may be taken as $\tilde{\mu}_\theta(x) = \hat{\mu}_\theta(x,g)$.

3 Consistency of the bootstrap estimator

For our convergence considerations, we have to assume that the time series (X_{t+1}, X_t) satisfies appropriate mixing conditions. There are a number of mixing conditions discussed, e.g., in the monograph of Doukhan (1995). Among them strong or α- mixing is a reasonably weak one known to be fulfilled for many time series models. In particular, for a stationary solution of (4), geometric ergodicity implies the process is strongly mixing with geometrically decreasing mixing coefficient. Because model (3) is a special case of (4), then the time series is an example of a quantile autoregressive process (1) for which (X_{t+1}, X_t) is α-mixing as well.

Let π be the unique stationary distribution. We will assume that process (1) and hence, (3) satisfy the following assumptions.

(M1) The distribution of innovations ξ_t possesses a density p_ξ, which satisfies $\inf_{x\in C} p_\xi(x) > 0$ for all compact sets C.
(M2) μ_θ is bounded on compact sets and $\limsup_{|x|\to\infty} |x|^{-1}|\mu_\theta(x)| < 1$

The two conditions imply stationarity and geometric ergodicity of the process, see Theorems 1 and 2 in Diebold and Guegan (1990), in the case of autoregressive mean model. The assumptions ensure that the stationary distribution π possesses an everywhere positive Lebesgue density, here denoted as p. From (1), we have

$$p(y) = \int_{\mathbf{R}} p_\xi\left(\frac{y-\mu_\theta(x)}{\sigma_\theta}\right)p(x)dx$$

with $\sigma_\theta > 0$ being some constant scale at θ. For a stationary solution of (1), geometric ergodicity implies that the process is α-mixing with geometrically decreasing mixing coefficients.

Let us denote by $\ell_F(x)$, the law of $\sqrt{ng}(\hat{F}_x(\mu, g) - F_x(\mu))$ and, in the bootstrap world, $\ell_{\tilde{F}}(x)$, the conditional distribution of $\sqrt{nh}(\hat{F}_x^*(\mu, h) - \tilde{F}_x(\mu))$ given the original data. Likewise, denote the distribution of $\sqrt{ng}(\hat{\mu}_\theta(x, g) - \mu_\theta(x))$ by $\ell_\mu(x)$ and the conditional distribution of the estimator $\sqrt{nh}(\hat{\mu}_\theta^*(x, h) - \tilde{\mu}_\theta(x, g))$ by $\ell_{\tilde{\mu}}(x)$. The following assumptions are now made.

(A1) The kernel K is nonnegative Lipschitz continuous function with compact support $[-1, 1]$, satisfying $\int K(u)du = 1, \quad \int uK(u)du = 0$. The bandwidth h satisfies $h \to 0, nh \to \infty$ as $n \to \infty$. As abbreviations, we use $K_\infty = max_u|K(u)|, \quad s_K^2 = \int K^2(u)du, \quad \sigma_K^2 = \int u^2 K(u)du$.
(A2) For all μ, x satisfying $0 < F_x(\mu) < 1$

(i) $F_x(\mu)$ and $p(x)$ are twice continuously differentiable with continuous and bounded derivatives in x, μ.

(ii) The distribution P_ξ of the innovations ξ_{t+1} has a density p_ξ with the following properties: $\inf_{x \in C} p_\xi(x) > 0$ for all compact sets C, $\quad p_\xi$ is twice continuously differentiable, and $\sup_{x \in \mathbf{R}} |xp_\xi'(x)| < \infty$.

(iii) for fixed x, $F_x(\mu)$ has the conditional density, $f_x(\mu)$, which is continuous in x and Hölder-continuous in μ: $|f_x(\mu) - f_x(\mu')| \leq c|\mu - \mu'|^\beta$ for some $c, \beta > 0$.

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

(v) The conditional density $f_x(\mu)$ is uniformly bounded in x and μ by, say, c_f and has at least two bounded derivatives.

(A3) The quantile innovations ξ_{t+1} and $\rho_\theta(\xi_{t+1}, 0) - 1$, have a continuous positive density in the neighborhood of 0.
(A4) $\sqrt{nh}(\hat{\mu}_\theta(x, h) - \mu_\theta(x))$ has asymptotic normal distribution.
(A5) The bootstrap innovations ξ_{t+1}^* have (conditional) θ-quantile zero and $d_K(\tilde{P}_\xi, P_\xi) = o_p(1)$, where d_K denotes the Kolmogorov distance between the probability measures.
(A6) For some compact set G,

(i) there are $\epsilon > 0, \gamma > 0$, such that $p(x) \geq \gamma$ for all x in the ϵ-neighborhood $\{x; |x - u| < \epsilon \text{ for some } u \in G\}$ of G.

(ii) In addition to (i) above, there exists some compact neighborhood Θ_0

of 0, for which the set $\Theta = \{\nu = \mu_\theta(x) + \mu;\ x \in G, \mu \in \Theta_0\}$ is compact too, and for some constant $c_0 > 0$, $f_x(\nu) \geq c_0$ for all $x \in G, \nu \in \Theta$.

(A7) (X_{t+1}, X_t) is stationary and α-mixing with mixing coefficients satisfying $\alpha(s_n) = O(s_n^{-(2+\delta)}), n \geq 1$, $\delta > 0$ and s_n is an increasing sequence of positive integers such that for some finite A

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

(A8) There exists a sequence $\gamma_n \to \infty$ such that $\sup_{-\gamma_n \leq x \leq \gamma_n} |\tilde{\mu}_\theta(x)| = O_p(1)$ and the function $\mu_\theta(x)$ is Lipschitz continuous with constant L_μ, and $\sup_{-\gamma_n \leq x \leq \gamma_n} |\tilde{\mu}_\theta(x) - \mu_\theta(x)| = o_p(1)$

(A9) Let $F_x^{ij}(\mu) = \frac{\partial^{i+j} F_x(\mu)}{dx^i d\mu^j}$ and correspondingly for $\tilde{F}_x(\mu)$. Moreover, let $p_\xi^{(i)}$ denote the i-th derivative of p_ξ, and correspondingly for $\tilde{p}_\xi$. With γ_n as in (A8), $\sup_{-\gamma_n \leq x \leq \gamma_n} |\tilde{F}_x^{ij}(\mu) - F_x^{ij}(\mu)| = o_p((\gamma_n)^{-1})$, and for $C > 0$, $\sup_{|x| \leq C} |\tilde{p}_\xi^{(i)}(x) - p_\xi^{(i)}(x)| = o_p(1)$, for $i, j = 0, 1, 2$.

Consistency of the bootstrap nonparametric quantile autoregression function estimator is now stated in the following Theorem.

Theorem 3.1. *Assume (A1)-(A9) hold for $x \in \mathbf{R}$. Then*

$$d_K(\ell_{\tilde{F}}(x), \ell_F(x)) \rightarrow 0 \quad \textit{in probability, and} \tag{16}$$

$$d_K(\ell_{\tilde{\mu}}(x), \ell_\mu(x)) \rightarrow 0 \quad \textit{in probability.} \tag{17}$$

4 Conclusion

In this paper, we have given the bootstrap resampling algorithm for estimating the quantile autoregression function. Rigorous proof has shown that the distribution of the estimator of the function, in the bootstrap world, converges uniformly to the one in the real world. However, simulation study may be necessary to ascertain the accurracy of the two methods. This is the topic of the next paper.

5 Appendix

For simplicity, identical bandwidths are assumed and henceforth dropped from various estimators. To facilitate comparisons of various terms, the following definitions and proposition are important.

Definition B.1. *For two random variables X and Y, the Mallows distance is defined as*

$$d_q^q(X,Y) = d_q^q(\ell(X),\ell(Y)) = \inf\{E|U-V|^q|\ell(U)=\ell(X),\ell(V)=\ell(Y)\}$$

For $q=1$, we have the absolute loss function. In current work, we modify the distance to cater for asymmetries, that depend on the value of θ, and call the distance, the Mallows-Check distance, for q=1.

Definition B.2. *For two random variables, X and Y, the Mallows-Check distance is defined as*

$$\begin{aligned} d_{1,\theta}(X,Y) &= d_{1,\theta}(\ell(X),\ell(Y)) \\ &= \inf\{E\rho_\theta(U-V)|\ell(U)=\ell(X),\ell(V)=\ell(Y)\} \end{aligned}$$

where ρ_θ is the usual check function defined in section 2.

Let F_{ξ_θ} denote the law of the innovations ξ_{t+1}. Suppose $\hat{F}_n$ is the empirical distribution of $\xi_{j,\theta}, j = 1,2,\ldots,n$ and $\hat{F}_{n,b} = \hat{F}_n * H_b$ is the empirical smoothed version of the empirical law with bandwidth b. Also, let $\tilde{F}_n$ and $\tilde{F}_{n,b}$ be the empirical distribution and convoluted (smoothed) distribution of $\tilde{\xi}_{j,\theta}, j=1,2,\ldots,n$. The following proposition shows that the bootstrap innovations ξ^*_{t+1} approximate the true residuals ξ_{t+1} in the Mallows-Check distance.

Proposition B.1. *Under assumption in Theorem 3.1*

$$d_{1,\theta}(\xi_{t+1},\xi^*_{t+1}) \to 0, \quad \mathit{if} \quad b\to 0 \quad \mathit{for} \quad n\to\infty.$$

Proof *of Proposition B.1*
We have

$$d_{1,\theta}(\xi_{t+1},\xi^*_{t+1}) = d_{1,\theta}(F_{\xi_\theta},\tilde{F}_{n,b}) \le d_{1,\theta}(F_{\xi_\theta},\hat{F}_{n,b}) + d_{1,\theta}(\hat{F}_{n,b},\tilde{F}_{n,b})$$

For the second term, let J be laplace distributed on a set $\{1, 2, \ldots n\}$, i.e $J = j$ with probability n^{-1} for each $j = 1, 2, \ldots, n$. Consider the random variables ξ_J and $\tilde{\xi}_J$ which have marginals $\hat{F}_n$ and $\tilde{F}_n$ respectively. Let χ be a random variable with density H_b. Then $\xi_J + \chi$ and $\tilde{\xi}_J + \chi$ have marginals $\hat{F}_{n,b}$ and $\tilde{F}_{n,b}$ respectively and

$$\begin{aligned} & d_{1,\theta}(\hat{F}_{n,b}, \tilde{F}_{n,b}) \leq E^* \rho_\theta(\xi_J + \chi - \tilde{\xi}_J - \chi) \\ = & \frac{1}{n} \sum_j \rho_\theta(\xi_{j,\theta} - \hat{\xi}_{j,\theta} + \hat{\xi}_\theta^\theta), \quad \text{where} \quad \hat{\xi}_\theta^\theta \quad \text{is} \quad \theta - \text{quantile} \quad \text{of} \quad \hat{\xi}_{j,\theta}, \quad j = 1, 2, \ldots n \\ = & \frac{1}{n} \sum_j [\rho_\theta(\hat{\mu}_\theta(X_j - \mu_\theta(X_j), 0) + \rho_\theta(\hat{\xi}_\theta^\theta, 0)] \\ \leq & \frac{1}{n} \sum_j [|\hat{\mu}_\theta(X_j) - \mu_\theta(X_j)| + |\hat{\xi}_\theta^\theta|] \to 0 \end{aligned}$$

These terms go to zero by law of large numbers. For the first term, we have $d_{1,\theta}(F_{\xi_\theta}, \hat{F}_{n,b}) \leq d_{1,\theta}(F_{\xi_\theta}, \hat{F}_n) + d_{1,\theta}(\hat{F}_n, \hat{F}_{n,b})$. As $n \to \infty$, the first term converges to 0 by lemma 8.4 in Bickel and Freedman (1981). And

$$\begin{aligned} d_{1,\theta}(\hat{F}_n, \hat{F}_{n,b}) & \leq E^* \rho_\theta(\xi_J - \xi_J - \chi) \\ & = E^* \rho_\theta(\chi, 0) \end{aligned}$$

which is $O(b) \to 0$ as $n \to \infty$. $\square$

The infimum of the Mallows-Check distance is attainable in the Mallows L_1 distance sense, as shown below;

$$\begin{aligned} d_{1,\theta}(\xi_{t+1}, \xi^*_{t+1}) & = \inf_{\ell(\xi'_{t+1}) = \ell(\xi_{t+1}), \ell(\xi'^*_{t+1}) = \ell(\xi^*_{t+1})} E\rho_\theta(\xi'_{t+1} - \xi'^*_{t+1}) \\ & \leq \inf_{\ell(\xi'_{t+1}) = \ell(\xi_{t+1}), \ell(\xi'^*_{t+1}) = \ell(\xi^*_{t+1})} E|\xi'_{t+1} - \xi'^*_{t+1}| \\ & = d_1(\xi'_{t+1}, \xi'^*_{t+1}) \end{aligned}$$

where the last expression is the L_1 Mallows distance. From assumption(A5), $E^*(\xi^*_{t+1})^2 = E(\xi^2_{t+1})$, and Proposition B.1, we have

$$d_{1,\theta}(\xi_{t+1}, \xi^*_{t+1}) = d_{1,\theta}(P_{\xi_\theta}, \tilde{P}_{\xi_\theta}) = E^* \rho_\theta(\tilde{\xi}_{t,\theta} - \xi^*_{t+1}, 0) = o_p(1),$$

Let $\tilde{X}_o$ be such that $\ell^*(\tilde{X}_o) = \ell(X_o)$ and define $\tilde{X}_t = \mu_\theta(\tilde{X}_{t-1}) + \sigma_\theta \tilde{\xi}_{t,\theta}$. The following three Lemmas show that for the distributions of $\tilde{X}_t$ and X^*_t, $|X^*_t| \leq \gamma_n$, for $\theta = 0.5$, holds with probability one.

Lemma B.1. *Under assumptions (A1)-(A9),*

$$\max_{0\le t\le n} P^*(|X_t^*| \ge \gamma_n) \to 0 \quad \textit{in probability.}$$

Lemma B.2. *Under assumptions (A1)-(A9), for a constant $0 < \lambda < 1$ and for random variables $S_1 = o_p(1)$, $\quad S_2 = O_p(1)$ and $L < \lambda + o_p(1)$ that do not depend on t, we have for $1 \le t \le n$, that*

$$\sup_{1\le t\le n} E^*|X_t^* - \tilde{X}_t| = S_1 + L^{t-1} S_2.$$

Lemma B.3. *Under the assumptions of Theorem 3.1, we have*

$$E^*\Big\{\frac{1}{n}\sum_{t=1}^{n} |\tilde{X}_t - X_t^*|\Big\} = o_p(1)$$

The proofs of Lemmas B.1, B.2 and B.3 are found in Franke et al. (2002). In the proofs, X_t is assumed to have approximately the same distribution as X_t^*. For strong approximation of this property, we can construct samples of errors $\tilde{\xi}_{1,\theta}, \tilde{\xi}_{2,\theta}, \ldots, \tilde{\xi}_{n,\theta}$ that have conditional distribution P_{ξ_θ}, given $X_1, X_2, \ldots, X_n$. The errors are then used to construct a process X_t^* with conditional distribution equal to unconditional distribution of X_t. The samples of the error variables are choosen such that they have the following properties:

(i) $\tilde{\xi}_{1,\theta}, \tilde{\xi}_{2,\theta}, \ldots, \tilde{\xi}_{n,\theta}$ are conditionally *iid*, given the original data $(X_1, \ldots, X_n)$

(ii)$\tilde{\xi}_{t,\theta}$ has a conditional distribution, given the original data $(X_1, \ldots, X_n)$, which is identical to unconditional distribution $\xi_{t,\theta}$, i.e., $\ell^*(\tilde{\xi}_{t,\theta}) = \ell(\xi_{t,\theta})$

(iii) $E^*\rho_\theta(\tilde{\xi}_{t,\theta} - \xi^*_{t,\theta}) = d_{1,\theta}(\tilde{\xi}_{t,\theta}, \xi^*_{t,\theta}) = d_{1,\theta}(\xi_{t,\theta}, \xi^*_{t,\theta})$ where $d_{1,\theta}$ is the Mallows-Check distance.

Theorem B.1. *Under assumptions (A1), (A2) and (A7),*

$$\sqrt{nh}(\hat{F}_x(\mu) - F_x(\mu) - B(\mu) + o_p(h^2)) \to^D N(0, V^2(\mu))$$

where

$$\begin{aligned} B(\mu) &= h^2\mu_K^2\Big[\frac{p^{'}(x)}{p(x)}F_x^{10}(\mu) + \frac{1}{2}F_x^{20}(\mu)\Big], \quad \textit{and} \\ V^2(\mu) &= \frac{1}{p(x)}[F_x(\mu)(1 - F_x(\mu))\sigma_K^2] \end{aligned}$$

Proof *of Theorem B.1*
The Theorem can easily be proved by employing Doob's small- and large-block techniques, see Ibragimov and Linnik (1971) page 316. First show that the summands over residuals and in large blocks are asymptotically negligible in probability and asymptotically independent respectively. Then show that the standard Lindeberg-Feller conditions for asymptotic normality of the summands in large blocks hold under independence assumptions. See Mwita (2003) for a complete proof. □

Theorem B.2 below gives the asymptotic distribution with bias and variance of the estimator $\hat{\mu}_\theta(x)$.

Theorem B.2. *Assume that (A1)-(A3) and (A7) hold. As* $n \to \infty$*, let the sequence of bandwidths* $h > 0$ *converge to 0 such that* $nh \to \infty$*. Then, the QAR function estimator is consistent,* $\hat{\mu}_\theta(x) \to^p \mu_\theta(x)$*, and asymptotically unbiased,*

$$\begin{aligned} E\sqrt{nh}(\hat{\mu}_\theta(x) - \mu_\theta(x)) &= \sqrt{nh}B_\mu(\mu_\theta(x)) + o(\sqrt{nh^5}) \quad \textit{where} \\ B_\mu(\mu) &= -\frac{B(\mu)}{f_x(\mu)}. \end{aligned}$$

If additionally, the bandwidths are chosen such that nh^5 *converges to 0 as* $n \to \infty$*, then the* $\widehat{\mu}_\theta(x)$ *is asymptotically normal,*

$$\sqrt{nh}\big(\hat{\mu}_\theta(x) - \mu_\theta(x) - B_\mu(\mu_\theta(x))\big) \to^D N\left(0, \frac{V^2(\mu_\theta(x))}{f_x^2(\mu_\theta(x))}\right), \tag{26}$$

where, $B(y)$ *and* $V^2(y)$ *are defined as in Theorem B.1 above.*

Proof *of Theorem B.2*
Theorem B.1 implies that, $\widehat{F}_x(\mu) \to F_x(\mu)$ in probability for all $x \in \mathbf{R}$ and y. The Glivenko-Cantelli Theorem in Krishnaiah (1990) for strongly mixing sequences implies

$$\sup_{\mu \in \mathbf{R}} |\widehat{F}_x(\mu) - F_x(\mu)| \to 0 \quad \text{in probability.} \tag{27}$$

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

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

This implies, using the monotonicity of F_x, that

$$\begin{aligned} P\{|\widehat{\mu}_\theta(x) - \mu_\theta(x)| > \epsilon\} &\leq P\{|F_x(\widehat{\mu}_\theta(x)) - F_x(\mu_\theta(x))| > \delta\} \\ &\leq P\{|F_x(\widehat{\mu}_\theta(x)) - \widehat{F}_x(\widehat{\mu}_\theta(x))| > \delta - \frac{c_\gamma}{nh}\} \\ &\leq P\{\sup_\mu |\widehat{F}_x(\mu) - F_x(\mu)| > \delta'\} \qquad (28) \end{aligned}$$

for arbitrary $\delta' < \delta$ and n large enough. Here, we have used $F_x(\mu_\theta(x)) = \theta$ and $\theta \leq \widehat{F}_x(\widehat{\mu}_\theta(x)) \leq \theta + 2K_\infty/(\gamma nh)$. Now, (28) tends to zero by (27). Hence the consistency follows.

To prove (26), let $b = -B(\mu_\theta(x)) f_x^{-1}(\mu_\theta(x))$ and $v = V(\mu_\theta(x)) f_x^{-1}(\mu_\theta(x))$. Let

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

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

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

By the same argument as in (28), we may replace $\widehat{F}_x(\widehat{\mu}_\theta(x))$ by $F_x(\mu_\theta(x))$ up to an error of $(nh)^{-1}$ at most, and we get, by Taylor expansion and neglecting the $(nh)^{-1}$-term which is asymptotically negligible anyhow,

$$\begin{aligned} q_n(z) &\sim P(F_x(\mu_\theta(x)) \leq \widehat{F}_x(\mu_\theta(x) + b + (nh)^{-1/2} vz) \\ &\sim P(-\delta_n f_x(\mu_\theta(x)) \leq \widehat{F}_x(\mu_\theta(x)) - F_x(\mu_\theta(x))) \end{aligned}$$

with $\delta_n = b + (nh)^{-1/2} vz$. Horvath and Yandell (1988) have also shown that the conditional distribution estimator $\widehat{F}_x(y)$ is asymptotically normal with asymptotic bias and variance as in Theorem B.1. This follows also under similar conditions from a functional central limit Theorem for $\widehat{F}_x(\mu)$ of Abberger (1996) - Corollary 5.4.1 and Lemma 5.4.1. Therefore, with $y_\theta = \mu_\theta(x)$, we get

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

by our choice of b and v and our condition on the rate of h. This proves the Theorem. □

The following Theorem gives the uniform convergence of $\hat{\mu}_\theta(x)$.

Theorem B.3. *Assume (A1), (A2), (A3) (A6) and (A7) hold. Suppose $h \to 0$ is a sequence of bandwidths such that $\tilde{S}_n = nh(s_n \log n)^{-1} \to \infty$ for some $s_n \to \infty$. Let $S_n = h^2 + \tilde{S}_n^{-\frac{1}{2}}$. Then we have*

$$\sup_{x\in G} |\widehat{\mu}_\theta(x) - \mu_\theta(x)| = O(S_n) + O(\frac{1}{nh}) \quad a.s. \tag{29}$$

The proofs of Theorem B.3 can be found in Franke and Mwita (2003). Usually, S_n will be much larger than $(nh)^{-1}$, and the rate of (29) will be $O(S_n)$. In particular, bias and variance are balanced and the mean-square error is asymptotically minimized.

Proof *of Theorem 3.1*
For the prove of (16), we first split $\hat{F}_x(\mu)$ into variance and bias terms

$$\sqrt{nh}(\hat{F}_x(\mu) - F_x(\mu)) = \frac{\sqrt{nh}\hat{r}_{V,h}(x,\mu)}{\hat{p}_h(x,\mu)} + \frac{\sqrt{nh}\hat{r}_{B,h}(x)}{\hat{p}_h(x)}$$

where

$$\begin{aligned}\hat{r}_{V,h}(x,\mu) &= \frac{1}{n}\sum_t K_h(x - X_t)(\mathbf{I}_{\{X_{t+1}\le\mu\}} - F_{X_t}(\mu))\\ \hat{r}_{B,h}(x,\mu) &= \frac{1}{n}\sum_t K_h(x - X_t)(F_{X_t}(\mu) - F_x(\mu))\end{aligned}$$

Similarly, we decompose the bootstrap distribution estimate, $\hat{F}^*_x(\mu)$ as

$$\sqrt{nh}(\hat{F}^*_x(\mu) - \tilde{F}_x(\mu) = \frac{\sqrt{nh}\hat{r}^*_{V,h}(x,\mu)}{\hat{p}^*_h(x)} + \frac{\sqrt{nh}\hat{r}^*_{B,h}(x,\mu)}{\hat{p}^*_h(x)}$$

where

$$\begin{aligned}\hat{r}^*_{V,h}(x,\mu) &= \frac{1}{n}\sum_t K_h(x - X_t)(\mathbf{I}_{\{X^*_{t+1}\le\mu\}} - \tilde{F}_{X^*_t}(\mu))\\ \hat{r}^*_{B,h}(x,\mu) &= \frac{1}{n}\sum_t K_h(x - X_t)(\tilde{F}_{X^*_t}(\mu) - \tilde{F}_x(\mu))\end{aligned}$$

We now compare random variables $\hat{r}_{V,h}(x)$, $\hat{r}_{B,h}(x)$ and $\hat{p}_h(x)$ with $\hat{r}^*_{V,h}(x)$, $\hat{r}^*_{B,h}(x)$ and $\hat{p}^*_h(x)$ and show that they have the same asymptotic behaviour.

Lemma B.4. *Under assumptions (A1)-(A9),*

$$\begin{aligned} d_K[\ell(\sqrt{nh}\hat{r}_{V,h}(x,\mu)), N(0,\tau^2(x)] &= o_p(1) \\ d_K[\ell(\sqrt{nh}\hat{r}^*_{V,h}(x,\mu)), N(0,\tau^2(x)] &= o_p(1) \end{aligned} \quad (32)$$

where $\tau^2(x) = F_x(\mu)(1-F_x(\mu))p(x)\int K^2(v)dv$.

Proof *of Lemma B.4*

We will verify the assumptions of a version of the central limit Theorem for martingale difference arrays, Brown (1971), also used in Franke et al.(2002). That is

$$E(\sqrt{nh}\hat{r}^*_{V,h}(x,\mu))^2 = \frac{h}{n}\sum_t E^*[K_h^2(x-X_t^*)(\eta^*_{t+1,\theta})^2|\mathbf{F}_t^*] \rightarrow_p \tau^2(x)$$

where $\eta^*_{t+1,\theta} = \mathbf{I}_{\{X^*_{t+1}\leq\mu\}} - \tilde{F}_{X_t^*}(\mu)$ and $\mathbf{F}_t^* = \sigma(X_1^*,\ldots,X_t^*)$, and for all $\epsilon > 0$,

$$\frac{h}{n}\sum_t E^*[K_h^2(x-X_t^*)(\eta^*_{t+1,\theta})^2\mathbf{I}_{\{K_h^2(x-X_t^*)(\eta^*_{t+1,\theta})^2>\epsilon\}}|\mathbf{F}_t^*] \rightarrow_p 0$$

Since K is bounded in the neighbourhood of x, we have

$$\begin{aligned} &\frac{1}{nh}\sum_t E^*[(\eta^*_{t+1,\theta})^2\mathbf{I}_{\{(\eta^*_{t+1,\theta})^2>K^2_\infty nh\epsilon\}}|\mathbf{F}_t^*] \\ \leq\ & n^{-1}(K^2_\infty\epsilon nh)^{-(\gamma_n-2)/2}E^*(\mathbf{I}_{\{(\eta^*_{t+1,\theta})^2)^{\gamma_n}\}} = o_p(1) \end{aligned}$$

For (33), we have

$$\begin{aligned} &\frac{h}{n}\sum_t E^*[K_h^2(x-X_t^*)(\eta^*_{t+1,\theta})^2|\mathbf{F}_t^*] \\ =\ & \frac{1}{nh}\sum_t \big(K^2(\frac{x-X_t^*}{h})(\tilde{F}_{X_t^*}(\mu)-\tilde{F}^2_{X_t^*}(\mu)) \\ -\ & E^*[K^2(\frac{x-X_t^*}{h})(\tilde{F}_{X_t^*}(\mu)-\tilde{F}^2_{X_t^*}(\mu))|\mathbf{F}^*_{t-1}]\big) \\ +\ & \frac{1}{nh}\sum_t\int K^2(\frac{x-\tilde{\mu}_\theta(X^*_{t-1})-\tilde{\sigma}_\theta u}{h})(\tilde{F}_{\tilde{\mu}_\theta(X^*_{t-1})+\sigma_\theta u}(\mu) \\ -\ & \tilde{F}_{\tilde{\mu}_\theta(X^*_{t-1})+\sigma_\theta u}(\mu)^2)\tilde{P}_{\xi_\theta}(du) \end{aligned}$$

Note that because K and $F_{X_t^*}(\mu)$ are bounded, the first summand is of order $O_p((nh)^{-2}) = o_p(1)$. We only consider the second summand. Letting $v = -(x - \tilde{\mu}_\theta(X_t^*) - \tilde{\sigma}_\theta u)/h$, the second summand is equal to

$$\frac{1}{n}\sum_t \int_{[-1,1]} K^2(v)(\tilde{F}_{x+hv}(\mu) - \tilde{F}^2_{x+hv}(\mu))\tilde{p}_{\xi_\theta}\big(\frac{x - \tilde{\mu}_\theta(X^*_{t-1})}{\sigma_\theta} + \frac{hv}{\sigma_\theta}\big)\frac{1}{\sigma_\theta}dv$$

$\tilde{p}_{\xi_\theta}$ is bounded in absolute value by

$$(|x| + \sup_{|x|\le\gamma_n} |\tilde{\mu}_\theta(x) - \mu_\theta(x)| + \sup_{|x|>\gamma_n} |\tilde{\mu}_\theta(x)| \sup_{|x|\le\gamma_n} |\mu_\theta(x)| + h)/\sigma_\theta \tag{36}$$

The fact that the time series $X_1, \ldots, X_n$ is a realization of stationary process with *iid* innovations and by assumptions (A8), the order of the (36) is $O_p(\gamma_n)$. Since σ_θ is constant, we can replace $\tilde{p}_{\xi_\theta}$ by p_{ξ_θ} and using uniform convergence of $\tilde{F}_x$ to F_x on a compact set, see assumptions (A9), we obtain

$$\frac{1}{n}\sum_t \int K^2(v)(F_{x+hv}(\mu) - F^2_{x+hv}(\mu))p_{\xi_\theta}\big(\frac{x - \tilde{\mu}_\theta(X^*_{t-1})}{\sigma_\theta} + \frac{hv}{\sigma_\theta}\big)\frac{1}{\sigma_\theta}dv + o_p(1). \tag{37}$$

Assume that F_x and p_{ξ_θ} have bounded derivatives and that $\sigma_\theta > 0$, then expression (37) is the equal to

$$\int K^2(v)dv(F_x(\mu) - F^2_x(\mu))\frac{1}{n}\sum_t p_{\xi_\theta}\big(\frac{x - \tilde{\mu}_\theta(X^*_{t+1})}{\sigma_\theta}\big)\frac{1}{\sigma_\theta}dv + o_p(1).$$

We now have to show that

$$\frac{1}{n}\sum_t p_{\xi_\theta}\big(\frac{x - \tilde{\mu}_\theta(X^*_{t-1})}{\sigma_\theta}\big)\frac{1}{\sigma_\theta} = \frac{1}{n}\sum_t p_{\xi_\theta}\big(\frac{x - \mu_\theta(\tilde{X}_{t-1})}{\sigma_\theta}\big)\frac{1}{\sigma_\theta} + o_p(1) \tag{39}$$

$$\frac{1}{n}\sum_t p_{\xi_\theta}\big(\frac{x - \mu_\theta(\tilde{X}_{t-1})}{\sigma_\theta}\big)\frac{1}{\sigma_\theta} = p(x) + o_p(1) \tag{40}$$

Note that $\{\tilde{X}_t\}$ is a process with conditional distribution equal to unconditional distribution of $\{X_t\}$ and the expected value of the left-hand side of (40) equals to $p(x)$. Hence (40) follows from the ergodicity of the process $\{X_t\}$. To show (39), we have by splitting,

$$\frac{1}{n}\sum_t \frac{1}{\sigma_\theta} p_{\xi_\theta}\big(\frac{x-\tilde{\mu}_\theta(X^*_{t-1})}{\sigma_\theta}\big)\mathbf{I}_{\{|X^*_{t-1}|>\gamma_n\}} = o_p(1) \tag{41}$$

$$\frac{1}{n}\sum_t \big\{\frac{1}{\sigma_\theta} p_{\xi_\theta}\big(\frac{x-\tilde{\mu}_\theta(X^*_{t-1})}{\sigma_\theta}\big) - \frac{1}{\sigma_\theta} p_{\xi_\theta}\big(\frac{x-\mu_\theta(X^*_{t-1})}{\sigma_\theta}\big)\big\}\mathbf{I}_{\{|X^*_{t-1}|\le\gamma_n\}}$$
$$= o_p(1) \tag{42}$$

$$\frac{1}{n}\sum_t \frac{1}{\sigma_\theta} p_{\xi_\theta}\big(\frac{x-\mu_\theta(X^*_{t-1})}{\sigma_\theta}\big)\mathbf{I}_{\{|X^*_{t-1}|>\gamma_n\}} = o_p(1) \tag{43}$$

$$\frac{1}{n}\sum_t \big|\frac{1}{\sigma_\theta} p_{\xi_\theta}\big(\frac{x-\tilde{\mu}_\theta(X^*_{t-1})}{\sigma_\theta}\big) - \frac{1}{\sigma_\theta} p_{\xi_\theta}\big(\frac{x-\mu_\theta(\tilde{X}_{t-1})}{\sigma_\theta}\big)\big|$$
$$= O\big(\frac{1}{n}\sum_t |X^*_{t-1} - \tilde{X}_{t-1}|\big) \tag{44}$$

Because of the boundness of p_ξ and the fact that $\sigma_\theta < \infty$, then the left-hand side of (41) is bounded by $\frac{1}{n}\sum_t \mathbf{I}_{\{|X^*_{t-1}|>\gamma_n\}} O_p(1)$, which is $o_p(1)$ by Lemma B.1. Likewise (43) follows from Lemma B.1 and the boundedness of p_{ξ_θ} and σ_θ^{-1}. Claim (42) follows from the boundedness of p_{ξ_θ} and p'_{ξ_θ} and from the fact that $\tilde{\mu}_\theta$ converges uniformly on $[-\gamma_{n,\theta-1}, \gamma_{n,\theta}]$. For the proof of (44), note that from the stationarity of the process, the boundedness of p'_{ξ_θ} and σ_θ^{-1} and assumptions in (A2)(ii), the function $\sigma_\theta^{-1} p_{\xi_\theta}(x - \mu_\theta(x))/\sigma_\theta(x))$ is Lipschitz continuous.Therefore claim (39) follows from Lemma B.3. This completes the proof of (32).

The following Lemma shows that the bootstrap density estimator converges to the true stationary density in probability.

Lemma B.5. *Under assumptions (A1)-(A9),*

$$\hat{p}_h(x) \;\to\; p(x) \quad \textit{in probability}$$
$$\hat{p}^*_h(x) \;\to\; p(x) \quad \textit{in probability}$$

Proof *of Lemma B.5*
The prove of Lemma B.5 can be found in Franke et al (2002). □

Lemma B.6. *Under assumptions (A1)-(A9),*

$$\sqrt{nh}\,\hat{r}_{B,h}(x,\mu) \;\to\; b(x) \quad \textit{in probability} \tag{45}$$
$$\sqrt{nh}\,\hat{r}^*_{B,h}(x,\mu) \;\to\; b(x) \quad \textit{in probability} \tag{46}$$

where $b(x) = \sqrt{nh^5}\int v^2K(v)dv[p^{'}(x)F_x^{10}(\mu) + \frac{1}{2}p(x)F_x^{20}(u)]$.

Proof *of Lemma B.6*
The proof of (46) is given. Similar lines can be used for (45). Now,

$$\sqrt{nh}\hat{r}^*_{B,h}(x,\mu) = \sqrt{\frac{h}{n}}\sum_t K_h(x - X_t^*)(\tilde{F}_{X_t^*}(\mu) - \tilde{F}_x(\mu)).$$

The Taylor expansion about x gives

$$\begin{aligned}\sqrt{nh}\hat{r}^*_{B,h}(x) &= \sqrt{\frac{h}{n}}\sum_t K_h(x - X_t^*)(X_t^* - x)\tilde{F}_x^{10}(\mu) \\ &+ \frac{1}{2}\sqrt{\frac{h}{n}}\sum_t K_h(x - X_t^*)(X_t^* - x)^2 F^{20}_{\hat{X}_t}(\mu). \qquad (48)\end{aligned}$$

Here, $\hat{X}_t$ denotes a suitable value between x and X_t^*. We will show that

$$\sqrt{\frac{h}{n}}\sum_t E^*[K_h(x - X_t^*)(X_t^* - x)|\mathbf{F}^*_{t-1}] \rightarrow Bp'(x)\int v^2K(v)dv \qquad (49)$$

$$\sqrt{\frac{h}{n}}\sum_t E^*[K_h(x - X_t^*)(X_t^* - x)^2|\mathbf{F}^*_{t-1}] \rightarrow Bp(x)\int v^2K(v)dv \qquad (50)$$

and then (46) will follow from the convergence of $\tilde{F}_x^{10}$ and $\tilde{F}_x^{20}$, see assumption (A9), and from the fact that the conditional variance of both terms on the right-hand side of (48) are of order $o_p(1)$.

Now, we can express (49) as

$$\begin{aligned}&\sqrt{\frac{h}{n}}\sum_t E^*[K_h(x - X_t^*)(X_t^* - x)|\mathbf{F}^*_{t-1}] \\ = &\sqrt{\frac{h}{n}}\sum_t \int \frac{1}{h}K(\frac{x - \tilde{\mu}_\theta(X^*_{t-1}) - \sigma_\theta u}{h})(\tilde{\mu}_\theta(X^*_{t-1}) + \sigma_\theta u - x)\tilde{P}_{\xi_\theta}(du) \\ = &\sqrt{\frac{h^3}{n}}\sum_t \int vK(v)\tilde{p}_{\xi_\theta}(\frac{x - \tilde{\mu}_\theta(X^*_{t-1}) + vh}{\sigma_\theta})\frac{1}{\sigma_\theta}dv\end{aligned}$$

A Taylor expansion for $\tilde{p}_{\xi_\theta}$ yiels

$$\sqrt{\frac{h^5}{n}}\sum_t \int v^2K(v)\tilde{p}^{'}_\xi(\hat{Z}_t^*)\frac{1}{\sigma_\theta}dv + o_p(1)$$

where $\hat{Z}_t^*$ is a suitable value between $(x - \tilde{\mu}_\theta(X_{t-1}^*))/\sigma_\theta$ and $(x - \tilde{\mu}_\theta(X_{t-1}^*) + hv)/\sigma_\theta$ as argument of $\tilde{p}'_{\xi_\theta}$. From Lemma B.5, $\tilde{p}_{\xi_\theta}$ converges uniformly to p_{ξ_θ} for all $C > 0$ (see (A9)), and since $p''_{\xi_\theta}(x)$ is bounded, the left hand-side of (49) is asymptotically equal to

$$\sqrt{nh^5}\int v^2K(v)dv\frac{1}{n}\sum_t p'_{\xi_\theta}(\frac{x - \tilde{\mu}_\theta(X_{t-1}^*)}{\sigma_\theta})\frac{1}{\sigma_\theta}$$

and the left hand side of (50) is asymptotically equal

$$\sqrt{nh^5}\int v^2K(v)dv\frac{1}{n}\sum_t p_{\xi_\theta}(\frac{x - \tilde{\mu}_\theta(X_{t-1}^*)}{\sigma_\theta})\frac{1}{\sigma_\theta}.$$

Similar arguments, to those of claim (39) and (40) are then used to complete the proof.

To show (17), we mainly use Taylor expansion. Now from Lemmas B.4 and B.5, the variance is

$$E(\sqrt{nh}(\hat{F}_x^*(\mu) - \tilde{F}_x(\mu))^2 \to_p \frac{F_x(\mu)(1 - F_x(\mu))\sigma_K^2}{p(x)}$$

Setting $\mu = \hat{\mu}_\theta^*(x)$, and expanding the left side, we get

$$E(\sqrt{nh}(\hat{\mu}_\theta^*(x) - \tilde{\mu}_\theta(x))(\tilde{F}_x^{01}(\tilde{\mu}_\theta(x)))^2 \to_p \frac{\theta(1-\theta)\sigma_K^2}{p(x)}$$

where terms small in probability have been ommited. By assumption (A9) and uniform convergence of $\hat{\mu}_\theta(x)$ to $\mu_\theta(x)$, in Theorem B.3, we get $\tilde{f}_x(\tilde{\mu}_\theta(x)) \to_p f_x(\mu_\theta(x))$ and

$$E(\sqrt{nh}(\hat{\mu}_\theta^*(x) - \tilde{\mu}_\theta(x))^2 \to_p \frac{\theta(1-\theta)\sigma_K^2}{p(x)f_x^2(\mu_\theta(x))}$$

From Lemmas B.5 and B.6, we have the bias is

$$\begin{aligned}&E\sqrt{nh}(\hat{F}_x^*(\mu) - \tilde{F}_x(\mu))\\ &\to_p \frac{\sqrt{nh^5}\mu_K^2}{p(x)}[p'(x)F_x^{10}(\mu) + \frac{1}{2}p(x)F_x^{20}(\mu)]\end{aligned}$$

Again, setting $\mu = \hat{\mu}_\theta^*(x)$, we get

$$\begin{aligned}&-E\sqrt{nh}(\hat{\mu}_\theta^*(x) - \tilde{\mu}_\theta(x))\tilde{F}_x^{01}(\tilde{\mu}_\theta(x)) - \sqrt{nh}(\hat{\mu}_\theta^*(x) - \tilde{\mu}_\theta(x))^2\tilde{F}_x^{02}(\tilde{\mu}_\theta(x)))\\ &\to_p \frac{\sqrt{nh^5}\mu_K^2}{p(x)}[p'(x)F_x^{10}(\mu_\theta(x)) + \frac{1}{2}p(x)F_x^{20}(\mu_\theta(x))]\end{aligned}$$

Hence, by assumption (A9) on convergence of $\tilde{F}_x$ and uniform convergence of $\tilde{\mu}_\theta(x)$, we get

$$\sqrt{nh}(\hat{\mu}^*_\theta(x) - \tilde{\mu}_\theta(x)) \\ \to_p -\frac{\sqrt{nh^5}\mu_K^2}{f_x(\mu_\theta(x))p(x)}[p^{'}(x)F_x^{10}(\mu_\theta(x)) + \frac{1}{2}p(x)F_x^{20}(\mu_\theta(x))]$$

where again, terms of smaller order in probability have been left out. This completes the proof of Theorem 3.1. □

ACKNOWLEDGEMENTS. During this research, the first 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 Financemarketdaten*, Hartung-Gore Verlag, Konstanz, 1996.

[2] P.J. Bickel and D. A. Freedman, Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9**, (1981), 1196-1217.

[3] B.M. Brown, Martingale Central Limit Theorems, *Ann. Math. Statist.*, **42**, (1971), 59-66.

[4] J. Diebold and D. Guegan, *Probabilistic Properties of the General Nonlinear Markovian Process of Order One and Applications to Time Series Modelling*, Rapport Technique de L. S. T. A., **125**, Université Paris XIII, 1990.

[5] P. Doukhan, Mixing: Properties and Examples, *Lecture notes in Statistics*, Springer-Verlag, New-York, 1995.

[6] J. Franke, J-P Kreiss and E. Mammen, Bootstrap of Kernel Smoothing in Nonlinear Time Series, *Bernoulli*, **8**(1), (2002), 1-37.

[7] J. Franke and P. Mwita, Nonparametric Estimates for Conditional Quantiles of time series, *Technical Report in Wirtschaft Mathematik*, **87**, University of Kaiserslautern, 2003.

[8] J. Franke and M. Wendel, A Bootstrap Approach for Nonlinear Autoregression– Some Preliminary Results. In: Jäckel, K.–H., Rothe, G. and Sendler, W. eds: Bootstrapping and Related Techniques. *Lecture Notes in Economics and Mathematical System*, **376**, Springer, Berlin–Heidelberg–New York, 1992.

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

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

[11] I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff Publishing, Groningen, 1971. With a supplimentary chapter by I. A. Ibragimov and V. V. Petror, Translation from the Rassian edited by J.F.C. Kingman

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

[13] E. Kreutzberger, *Bootstrap für Nichtlinear AR(1)–Prozesse*, Thesis, University of Kaiserslautern, Germany, 1993.

[14] Y.S. Rama Krishnaiah, On the Glivenko-Cantelli Theorem for generalized empirical processes based on strong mixing sequences, *Stat. Probab. Letter*, **10**(5), (1990), 439-447.

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

[16] E.A. Nadaraya, On estimating regression, *Theory of Probab. and its Applications*, **9**, (1964), 141-142.

[17] G.S. Watson, Smooth regression analysis, *Sankhya: The Indian Journal of Statistics*, (Series A), **26**(4), (1964), 359-372.