UNIVERSITÄT HAMBURG

# The independence process in conditional quantile location-scale models and an application to testing for monotonicity

M. Birke, N. Neumeyer and S. Volgushev

Preprint–No. 2013-03 Juni 2013

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

# The independence process in conditional quantile location-scale models and an application to testing for monotonicity

Melanie Birke
Universität Bayreuth

Natalie Neumeyer*
Universität Hamburg

Stanislav Volgushev†
Ruhr-Universität Bochum

June 25, 2013

**Abstract**

In this paper the nonparametric quantile regression model is considered in a location-scale context. The asymptotic properties of the empirical independence process based on covariates and estimated residuals are investigated. In particular an asymptotic expansion and weak convergence to a Gaussian process are proved. The results can, on the one hand, be applied to test for validity of the location-scale model. On the other hand, they allow to derive various specification tests in conditional quantile location-scale models. In detail a test for monotonicity of the conditional quantile curve is investigated. For the test for validity of the location-scale model as well as for the monotonicity test smooth residual bootstrap versions of Kolmogorov-Smirnov type test statistics are suggested. We give rigorous proofs for bootstrap versions of the weak convergence results. The performance of the tests is demonstrated in a simulation study.

AMS Classification: 62G10, 62G08, 62G30
Keywords and Phrases: bootstrap, empirical independence process, Kolmogorov-Smirnov test, model test, monotone rearrangements, nonparametric quantile regression, residual processes, sequential empirical process

---

*corresponding author; e-mail: neumeyer@math.uni-hamburg.de

†Supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823), Teilprojekt (C1), of the Deutsche Forschungsgemeinschaft.

# 1 Introduction

Quantile regression was introduced by Koenker and Bassett (1978) as an extension of least squares methods focusing on the estimation of the conditional mean function. Due to its many attractive features as robustness with respect to outliers and equivariance under monotonic transformations that are not shared by the mean regression, it has since then become increasingly popular in many important fields such as medicine, economics and environment modelling [see Yu et al. (2003) or Koenker (2005)]. Another important feature of quantile regression is its great flexibility. While mean regression aims at modelling the average behaviour of a variable $Y$ given a covariate $X = x$, quantile regression allows to analyse the impact of $X$ in different regions of the distribution of $Y$ by estimating several quantile curves simultaneously. See for example Fitzenberger et al. (2008), who demonstrates that the presence of certain structures in a company can have different effects on upper and lower wages. For a more detailed discussion, we refer the interested reader to the recent monograph by Koenker (2005).

The paper at hand has a twofold aim. On the one hand it proves a weak convergence result for the empirical independence process of covariates and estimated errors in a nonparametric local-scale conditional quantile model. On the other hand it suggests a test for monotonicity of the conditional quantile curve. To the authors' best knowledge this is the first time that those problems are treated for the general nonparametric quantile regression model.

The empirical independence process results from the distance of a joint empirical distribution function and the product of the marginal empirical distribution functions. It can be used to test for independence; see Hoeffding (1948), Blum et al. (1961) and ch. 3.8 in van der Vaart and Wellner (1996). When applied to covariates $X$ and estimators of error terms $\varepsilon = (Y - q(X))/s(X)$ it can be used to test for validity of a location-scale model $Y = q(X) + s(X)\varepsilon$ with $X$ and $\varepsilon$ independent. Here the conditional distribution of $Y$, given $X = x$, allows for a location-scale representation $P(Y \leq y \mid X = x) = F_\varepsilon((y - q(x))/s(x))$, where $F_\varepsilon$ denotes the error distribution function. To the best of our knowledge, Einmahl and Van Keilegom (2008a) is the only paper that considers such tests for location-scale models in a very general setting (mean regression, trimmed mean regression,...). However, the assumptions made there rule out the quantile regression case, where $q$ is defined via $P(Y \leq q(x) \mid X = x) = \tau$ for some $\tau \in (0, 1)$, $\forall x$. The first part of our paper can hence be seen as extension and completion of the results by Einmahl and Van Keilegom (2008a). Plenty of technical effort was necessary to obtain the weak convergence result in the quantile context (see the proof of Theorem 3.1 below). Validity of a location-scale means that the covariates have influence on the trend and on the dispersion of the conditional distribution of $Y$, but otherwise do not affect the shape of the conditional distribution (such models are frequently used, see Shim et al., 2009, and Chen et al., 2005). Contrariwise if the test rejects independence of covariates and errors then there is evidence that the influence of the covariates on the response goes beyond location and scale effects. Note that our results easily can be adapted to test the validity of location models $P(Y \leq y \mid X = x) = F_\varepsilon(y - q(x))$; see also Einmahl and Van Keilegom (2008b) and Neumeyer (2009b) in the mean regression context.

Further if there is some evidence that certain quantile curves might be monotone one should check by a statistical test, that this assumption is reasonable. Such evidence can e.g. come from an economic, physical or biological background. In classical mean regression there are various methods for testing monotonicity. It has already been considered e.g. in Bowman et al. (1998), Gijbels et al. (2000), Hall and Heckman (2001), Goshal et al. (2000), Durot (2003), Baraud et al. (2003) or Domínguez-Menchero et al. (2005) and Birke and Dette (2007) who consider an $L_2$-distance of a monotone estimator and an unconstrained one. More recent work on testing monotonicity is given in Wang and Meyer (2011) who use regression splines and use the minimum slope in the knots as test criterion and Birke and Neumeyer (2013) who use empirical process techniques for residuals built from isotonized estimators. While most of the tests are very conservative and not powerful against alternatives with only a small deviation from monotonicity the method proposed by Birke and Neumeyer (2013) has in some situations better power than the other tests and can also detect local alternatives of order $n^{-1/2}$. While there are several proposals for monotone estimators of a quantile function (see e.g. Cryer et al. (1972) or Robertson and Wright (1973) for median regression and Casady and Cryer (1976) or Abrevaya (2005) for general quantile regression), the problem of testing whether a given quantile curve is increasing (decreasing) has received nearly no attention in the literature. Aside from the paper by Duembgen (2002) which deals with the rather special case of median regression in a location model, the authors - to the best of their knowledge - are not aware of any tests for monotonicity of conditional quantile curves. The method, which is introduced here is based on the independence process considered before. Note that the test is not the same as the one considered by Birke and Neumeyer (2013) for mean regression adapted to the quantile case. It turned out that in quantile regression the corresponding statistic would not be suitable for constructing a statistical test (see also Section 4).

The paper is organized as follows. In Section 2 we present the location-scale model, give necessary assumptions and define the estimators. In Section 3 we introduce the independence process, derive asymptotical results and construct a test for validity of the model. Bootstrap data generation and asymptotic results for a bootstrap version of the independence process are discussed as well. The results derived there are modified in Section 4 to construct a test for monotonicity of the quantile function. In Section 5 we present a small simulation study while we conclude in Section 6. All proofs are deferred to an appendix and supplementary material.

## 2 The location-scale model, estimators and assumptions

For some fixed $\tau \in (0,1)$, consider the nonparametric quantile regression model of location-scale type [see e.g. He (1997)],

$$Y_i = q_\tau(X_i) + s(X_i)\varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $q_\tau(x) = F_Y^{-1}(\tau|x)$ is the $\tau$-th conditional quantile function, $(X_i, Y_i)$, $i = 1, \ldots, n$, is a bivariate sample of i.i.d. observations and $F_Y(\cdot|x) = P(Y_i \leq \cdot|X_i = x)$ denotes the conditional distribution function of $Y_i$ given $X_i = x$. Further, $s(x)$ denotes the median of $|Y_i - q_\tau(X_i)|$, given $X_i = x$. We assume that $\varepsilon_i$ and $X_i$ are independent and, hence, that $\varepsilon_i$ has $\tau$-quantile zero and $|\varepsilon_i|$ has median one, because

$$\begin{aligned} \tau &= P\Big(Y_i \leq q_\tau(X_i) \,\Big|\, X_i = x\Big) = P(\varepsilon_i \leq 0) \\ \frac{1}{2} &= P\Big(|Y_i - q_\tau(X_i)| \leq s(X_i) \,\Big|\, X_i = x\Big) = P(|\varepsilon_i| \leq 1). \end{aligned}$$

Denote by $F_\varepsilon$ the distribution function of $\varepsilon_i$. Then for the conditional distribution we obtain a location-scale representation as $F_Y(y|x) = F_\varepsilon((y - q_\tau(x))/s(x))$, where $F_\varepsilon$ as well as $q_\tau$ and $s$ are unknown.

For example, consider the case $\tau = \frac{1}{2}$. Then we have a median regression model, which allows for heteroscedasticity in the sense, that the conditional median absolute deviation $s(X_i)$ of $Y_i$, given $X_i$, may depend on the covariate $X_i$. Here the median absolute deviation of a random variable $Z$ is defined as $\mathrm{MAD}(Z) = \mathrm{median}(|Z - \mathrm{median}(Z)|)$ and is the typical measure of scale (or dispersion), when the median is used as location measure. This heteroscedastic median regression model is analogous to the popular heteroscedastic mean regression model $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$, $i = 1, \ldots, n$, where $X_i$ and $\varepsilon_i$ are assumed to be independent, $E[\varepsilon_i] = 0$, $\mathrm{sd}(\varepsilon_i) = 1$, and hence, $m(x) = E[Y_i \mid X_i = x]$, $\sigma(x) = \mathrm{sd}(Y_i \mid X_i = x)$ (see among many others e.g. Efromovich (1999), chapter 4.2 for further details).

**Remark 2.1** Note that assuming $|\varepsilon_i|$ to have median one is not restrictive. More precisely, if the model $Y_i = q_\tau(X_i) + \tilde{s}(X_i)\eta_i$ with $\eta_i$ i.i.d. and independent of $X_i$ and some positive function $\tilde{s}$ holds, the model $Y_i = q_\tau(X_i) + s(X_i)\varepsilon_i$ with $s(X_i) := \tilde{s}(X_i)F_{|\eta|}^{-1}(1/2)$, $\varepsilon_i := \eta_i/F_{|\eta|}^{-1}(1/2)$ will also be true, where $F_{|\eta|}$ denotes the distribution function of $|\eta_i|$. Then in particular $P(|\varepsilon_i| \leq 1) = P(|\eta_i| \leq F_{|\eta|}^{-1}(1/2)) = 1/2$. ■

In the literature, several non-parametric quantile estimators have been proposed [see e.g. Yu and Jones (1997, 1998), Takeuchi et al. (2006) or Dette and Volgushev (2008), among others]. In this paper we follow the last-named authors who proposed non-crossing estimates of quantile curves using a simultaneous inversion and isotonization of an estimate of the conditional distribution function. To be precise, let

$$\hat{F}_Y(y|x) := (\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}\mathbf{Y} \tag{2.2}$$

with

$$\mathbf{X} = \begin{pmatrix} 1 & (x-X_1) & \dots & (x-X_1)^p \\ \vdots & \vdots & \dots & \vdots \\ 1 & (x-X_n) & \dots & (x-X_n)^p \end{pmatrix}, \qquad \mathbf{Y} := \Big(\Omega\Big(\frac{y-Y_1}{d_n}\Big), \dots, \Omega\Big(\frac{y-Y_n}{d_n}\Big)\Big)^t$$

$$\mathbf{W} = \mathrm{Diag}\Big(K_{h_n,0}(x-X_1), \dots, K_{h_n,0}(x-X_n)\Big),$$

denote a smoothed local polynomial estimate (of order $p \geq 2$) of the conditional distribution function $F_Y(y|x)$ where $\Omega(\cdot)$ is a smoothed version of the indicator function and we used the notation $K_{h_n,k}(x) := K(x/h_n)(x/h_n)^k$. Here $K$ denotes a nonnegative kernel and $d_n, h_n$ are bandwidths converging to 0 with increasing sample size. Note that the estimator $\hat{F}_Y(y|x)$ can be represented as weighted average

$$\hat{F}_Y(y|x) = \sum_{i=1}^n W_i(x)\Omega\Big(\frac{y-Y_i}{d_n}\Big). \tag{2.3}$$

Following Dette and Volgushev (2008) we consider a strictly increasing distribution function $G : \mathbb{R} \to (0,1)$, a nonnegative kernel $\kappa$ and a bandwidth $b_n$, and define the functional

$$H_{G,\kappa,\tau,b_n}(F) := \frac{1}{b_n}\int_0^1\int_{-\infty}^{\tau} \kappa\Big(\frac{F(G^{-1}(u))-v}{b_n}\Big)dvdu.$$

Note that it is intuitively clear that $H_{G,\kappa,\tau,b_n}(\hat{F}_Y(\cdot|x))$, where $\hat{F}_Y$ is the estimator of the conditional distribution function defined in (2.2), is a consistent estimate of $H_{G,\kappa,\tau,b_n}(F_Y(\cdot|x))$. If $b_n \to 0$, this quantity can be approximated as follows

$$\begin{aligned} H_{G,\kappa,\tau,b_n}(F_Y(\cdot|x)) &\approx \int_{\mathbb{R}} I\{F_Y(y|x) \leq \tau\}dG(y) \\ &= \int_0^1 I\{F_Y(G^{-1}(v)|x) \leq \tau\}dv \;=\; G \circ F_Y^{-1}(\tau|x), \end{aligned}$$

and as a consequence an estimate of the conditional quantile function $q_\tau(x) = F_Y^{-1}(\tau|x)$ can be defined by

$$\hat{q}_\tau(x) := G^{-1}(H_{G,\kappa,\tau,b_n}(F_Y(\cdot|x))).$$

Finally, note that the scale function $s$ is the conditional median of the distribution of $|e_i|$, given the covariate $X_i$, where $e_i = Y_i - q_\tau(X_i) = s(X_i)\varepsilon_i$, $i = 1, \dots, n$. Hence, we apply the quantile-regression approach to $|\hat{e}_i| = |Y_i - \hat{q}_\tau(X_i)|$, $i = 1, \dots, n$, and obtain the estimator

$$\hat{s}(x) = G_s^{-1}(H_{G_s,\kappa,1/2,b_n}(\hat{F}_{|e|}(\cdot|x)))\,. \tag{2.4}$$

Here $G_s : \mathbb{R} \to (0,1)$ is a strictly increasing distribution function and $\hat{F}_{|e|}(\cdot|x)$ denotes the estimator of the conditional distribution function $F_{|e|}(\cdot|x) = P(|e_i| \le \cdot|X_i = x)$ of $|e_i|$, $i = 1, \dots, n$, i. e.

$$\hat{F}_{|e|}(y|x) = \sum_{i=1}^{n} W_i(x) I\{|\hat{e}_i| \le y\} \tag{2.5}$$

with the same weights $W_i$ as in in (2.3).

For a better overview and for later reference, below we collect all the technical assumptions concerning the estimators needed throughout the rest of the paper. First, we collect the assumptions needed for the kernel functions and functions $G, G_s$ used in the construction of the estimators.

(**K1**) The function $K$ is a symmetric, positive, Lipschitz-continuous density with support $[-1, 1]$. Moreover, the matrix $\mathcal{M}(K)$ with entries

$$(\mathcal{M}(K))_{k,l} = \mu_{k+l-2}(K) := \int u^{k+l-2} K(u) du$$

is invertible.

(**K2**) The function $K$ is two times continuously differentiable, $K^{(2)}$ is Lipschitz continuous, and for $m = 0, 1, 2$ the set $\{x|K^{(m)}(x) > 0\}$ is a union of finitely many intervals.

(**K3**) The function $\Omega$ has derivative $\omega$ which has support $[-1, 1]$, is a kernel of order $p_\omega$, and is two times continuously differentiable with uniformly bounded derivatives.

(**K4**) The function $\kappa$ is a symmetric, uniformly bounded, and has one Lipschitz-continuous derivative.

(**K5**) The function $G : \mathbb{R} \to [0, 1]$ is strictly increasing. Moreover, it is two times continuously differentiable in a neighborhood of the set $Q := \{q_\tau(x)|x \in [0, 1]\}$ and its first derivative is uniformly bounded away from zero on $Q$.

(**K6**) The function $G_s : \mathbb{R} \to (0, 1)$ is strictly increasing. Moreover, it is two times continuously differentiable in a neighborhood of the set $S := \{s(x)|x \in [0, 1]\}$ and its first derivative is uniformly bounded away from zero on $S$.

The data-generating process needs to satisfy the following conditions.

(**A1**) The density $f_X$ has support [0,1], is uniformly bounded away from zero and infinity, and is Lipschitz-continuous.

(**A2**) The function $s$ is uniformly bounded and $\inf_{x \in [0,1]} s(x) = c_s > 0$.

(**A3**) The partial derivatives $\partial_x^j \partial_y^l F_Y(y|x), \partial_x^k \partial_y^l F_e(y|x)$ exist and are continuous and uniformly bounded on $\mathbb{R} \times [0,1]$ for $k \wedge l \leq 2$ or $k + l \leq d$ for some $d \geq 3$.

(**A4**) The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with strictly increasing distribution function $F_\varepsilon$ (independent of $X_i$) and density $f_\varepsilon$, which is continuously differentiable such that $f_\varepsilon(0) > 0$, $\sup_{y\in\mathbb{R}} |y f_\varepsilon(y)| < \infty$ and $\sup_{y\in\mathbb{R}} |y^2 f_\varepsilon'(y)| < \infty$. The $\varepsilon_i$ have $\tau$-quantile zero and $F_{|\varepsilon|}(1) = 1/2$, that is $|\varepsilon_1|$ has median one.

(**A5**) For some $\alpha > 0$ we have $\sup_{u,y} |y|^\alpha (F_Y(y|u) \wedge (1 - F_Y(y|u))) < \infty$.

**Remark 2.2** Note that by the implicit function theorem, assumptions (**A3**) and (**A4**) imply that $x \mapsto q_\tau(x)$ and $s(x)$ are $d$ times continuously differentiable with uniformly bounded derivatives. ■

Finally, we assume that the bandwidth parameters satisfy

$$(\mathbf{BW}) \quad \frac{\log n}{n h_n (h_n \wedge d_n)^4} = o(1), \quad \frac{\log n}{n h_n^2 b_n^2} = o(1), \quad d_n^{2(p_\omega \wedge d)} + h_n^{2((p+1)\wedge d)} + b_n^4 = o(n^{-1}),$$

with $p_\omega$ from (K3), $d$ from (A3) and $p$ the order of the polynomial estimator in (2.2).

**Remark 2.3** If for example $d = p_\omega = p = 3$ and we set $d_n = h_n = n^{-1/6-\beta}$ for some $\beta \in (0, 1/30)$, $b_n = h_n^{-1/4-\alpha}$ such that $\alpha + \beta \in (0, 1/12)$, condition (**BW**) holds. ■

# 3 The independence process, asymptotic results and testing for model validity

As estimators for the errors we build residuals

$$(3.1) \qquad \hat{\varepsilon}_i = \frac{Y_i - \hat{q}_\tau(X_i)}{\hat{s}(X_i)}, \quad i = 1, \ldots, n.$$

In the definition of the process on which test statistics are based we only consider those observations $(X_i, Y_i)$ such that $2h_n \leq X_i \leq 1 - 2h_n$ in order to avoid boundary problems of the estimators. For $y \in \mathbb{R}$, $t \in [2h_n, 1 - 2h_n]$ we define the joint empirical distribution function of pairs of covariates and residuals as

$$(3.2) \qquad \begin{aligned} \hat{F}_{X,\varepsilon,n}(t,y) &= \sum_{i=1}^n I\{\hat{\varepsilon}_i \leq y\} I\{2h_n < X_i \leq t\} \frac{1}{\sum_{i=1}^n I\{2h_n < X_i \leq 1 - 2h_n\}} \\ &= \frac{1}{n} \sum_{i=1}^n I\{\hat{\varepsilon}_i \leq y\} I\{2h_n < X_i \leq t\} \frac{1}{\hat{F}_{X,n}(1-2h_n) - \hat{F}_{X,n}(2h_n)}, \end{aligned}$$

where $\hat{F}_{X,n}$ denotes the usual empirical distribution function of the covariates $X_1, \dots, X_n$. The empirical independence process compares the joint empirical distribution with the product of the corresponding marginal distributions. We thus define

$$S_n(t, y) \;=\; \sqrt{n}\Big(\hat{F}_{X,\varepsilon,n}(t, y) - \hat{F}_{X,\varepsilon,n}(1 - 2h_n, y)\hat{F}_{X,\varepsilon,n}(t, \infty)\Big) \tag{3.3}$$

for $y \in \mathbb{R}$, $t \in [2h_n, 1 - 2h_n]$, and $S_n(t, y) = 0$ for $y \in \mathbb{R}$, $t \in [0, 2h_n) \cup (1 - 2h_n, 1]$. In the following theorem we state a weak convergence result for the independence process.

**Theorem 3.1** *Under the location-scale model (2.1) and assumptions (**K1**)-(**K6**), (**A1**)-(**A5**) and (**BW**) we have the asymptotic expansion*

$$\begin{aligned} S_n(t, y) \;=\; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( I\{\varepsilon_i \le y\} - F_\varepsilon(y) - \phi(y)\Big(I\{\varepsilon_i \le 0\} - \tau\Big) - \psi(y)\Big(I\{|\varepsilon_i| \le 1\} - \frac{1}{2}\Big)\Big) \\ & \times \Big(I\{X_i \le t\} - F_X(t)\Big) + o_P(1) \end{aligned}$$

*uniformly with respect to $t \in [0, 1]$ and $y \in \mathbb{R}$, where*

$$\phi(y) \;=\; \frac{f_\varepsilon(y)}{f_\varepsilon(0)}\Big(1 - y\frac{f_\varepsilon(1) - f_\varepsilon(-1)}{f_{|\varepsilon|}(1)}\Big), \qquad \psi(y) \;=\; \frac{y f_\varepsilon(y)}{f_{|\varepsilon|}(1)}$$

*and $f_{|\varepsilon|}(y) = (f_\varepsilon(y) + f_\varepsilon(-y))I_{[0,\infty)}(y)$ is the density of $|\varepsilon_1|$. The process $S_n$ converges weakly in $\ell^\infty([0, 1] \times \mathbb{R})$ to a centered Gaussian process $S$ with covariance*

$$\begin{aligned} \mathrm{Cov}(S(s, y), S(t, z)) \;=\; & (F_X(s \wedge t) - F_X(s)F_X(t)) \\ \times \Big[ & F_\varepsilon(y \wedge z) - F_\varepsilon(y)F_\varepsilon(z) + \phi(y)\phi(z)(\tau - \tau^2) + \frac{1}{4}\psi(y)\psi(z) \\ & - \phi(y)(F_\varepsilon(z \wedge 0) - F_\varepsilon(z)\tau) - \phi(z)(F_\varepsilon(y \wedge 0) - F_\varepsilon(y)\tau) \\ & - \psi(y)\Big((F_\varepsilon(z \wedge 1) - F_\varepsilon(-1))I\{z > -1\} - \frac{1}{2}F_\varepsilon(z)\Big) \\ & - \psi(z)\Big((F_\varepsilon(y \wedge 1) - F_\varepsilon(-1))I\{y > -1\} - \frac{1}{2}F_\varepsilon(y)\Big) \\ & + (\phi(y)\psi(z) + \phi(z)\psi(y))\Big(F_\varepsilon(0) - F_\varepsilon(-1) - \frac{1}{2}\tau\Big)\Big]. \end{aligned}$$

The proof is given in Appendix A.

**Remark 3.2** The result can easily be adapted for location models $Y_i = q_\tau(X_i) + \varepsilon_i$ with $\varepsilon_i$ and $X_i$ independent. To this end we just set $\hat{s} \equiv 1$ in the definition of the estimators. The asymptotic covariance in Theorem 3.1 then simplifies because the function $\phi$ reduces to $\phi(y) = f_\varepsilon(y)/f_\varepsilon(0)$ and $\psi(y) \equiv 0$. ■

In the remainder of this section we discuss how the asymptotic result can be applied to test for validity of the location-scale model, i. e. testing the null hypothesis of independence of error $\varepsilon_i$ and covariate $X_i$ in model (2.1).

**Remark 3.3** If the location-scale model is not valid, i. e. $X_i$ and $\varepsilon_i$ are dependent, but all other assumptions of Theorem 3.1 are valid, then one can show that $S_n(t,y)/n^{1/2}$ converges in probability to $P(\varepsilon_i \le y, X_i \le t) - F_\varepsilon(y)F_X(t)$, uniformly with respect to $y$ and $t$. ■

**Remark 3.4** If the location-scale model is valid for some $\tau$-th quantile regression function it is valid for every $\alpha$-th quantile regression function, $\alpha \in (0,1)$. This easily follows from $q_\alpha(x) = F_\varepsilon^{-1}(\alpha)s(x)+q_\tau(x)$ which is a consequence from the representation of the conditional distribution function $F_Y(y|x) = F_\varepsilon((y - q_\tau(x))/s(x))$ (compare Remark 2.1). A similar statement is even true for general location and scale measures, see e. g. Van Keilegom (1998), Prop. 5.1. Thus for testing the validity of the location-scale model one could restrict oneself to the median case $\tau = 0.5$, e. g. However, we consider the general case $\tau \in (0,1)$ here because in the next section we will apply our results to test for monotonicity of $q_\tau(x)$ in $x$, a property which is not universal in $\tau$. ■

**Remark 3.5** Einmahl and Van Keilegom (2008a) consider a process similar to $S_n$ for general location and scale models. They define $q(x) = \int_0^1 F^{-1}(s|x)J(s)\,ds$ and $s^2(x) = \int_0^1 (F^{-1}(s|x))^2 J(s)\,ds - q^2(x)$ with score function $J$, which rules out the quantile case $q(x) = F^{-1}(\tau|x)$. Einmahl and Van Keilegom (2008) show that estimation of the errors has no influence in their context, i. e. they obtain a scaled completely tucked Brownian sheet as limit process and thus asymptotically distribution-free tests. This is clearly not the case in Theorem 3.1. ■

To test for the validity of a location-scale model we reject the null hypothesis of independence of $X_i$ and $\varepsilon_i$ for large values of, e. g., the Kolmogorov-Smirnov statistic $K_n = \sup_{t\in[0,1],y\in\mathbb{R}} |S_n(t,y)|$. From Theorem 3.1 and the Continuous Mapping Theorem we obtain the following asymptotic distribution,

$$K_n \xrightarrow{d} \sup_{t\in[0,1],y\in\mathbb{R}} |S(t,y)| = \sup_{x\in[0,1],y\in\mathbb{R}} |S(F_X^{-1}(x),y)|,$$

which is independent from the covariate distribution $F_X$, but depends in a complicated manner on the error distribution $F_\varepsilon$. To overcome this problem we suggest a bootstrap version of the test. To this end let $\mathcal{Y}_n = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ denote the original sample. We generate bootstrap errors as $\varepsilon_i^* = \tilde{\varepsilon}_i^* + \alpha_n Z_i$ $(i = 1,\ldots,n)$, where $\alpha_n$ denotes a positive smoothing parameter, $Z_1,\ldots,Z_n$ are independent, standard normally distributed random variables (independent of $\mathcal{Y}_n$) and $\tilde{\varepsilon}_1^*,\ldots,\tilde{\varepsilon}_n^*$ are randomly drawn with replacement from the set of residuals $\{\hat{\varepsilon}_j \mid j \in \{1,\ldots,n\}, X_j \in (2h_n, 1-2h_n]\}$. Conditional on the original sample $\mathcal{Y}_n$ the random variables $\varepsilon_1^*,\ldots,\varepsilon_n^*$ are i.i.d. with distribution function

$$\tilde{F}_\varepsilon(y) = \frac{\frac{1}{n}\sum_{i=1}^n \Phi\left(\frac{y-\hat{\varepsilon}_i}{\alpha_n}\right) I\{2h_n < X_i \le 1-2h_n\}}{\hat{F}_{X,n}(1-2h_n) - \hat{F}_{X,n}(2h_n)},$$

where $\Phi$ denotes the standard normal distribution function. Note that the bootstrap error's $\tau$-quantile is not exactly zero, but vanishes asymptotically. We use a smooth distribution to generate new bootstrap errors because smoothness of the error distribution is a crucial assumption for the theory necessary to derive Theorem 3.1; see also Neumeyer (2009a).
Now we build new bootstrap observations,

$$Y_i^* \;=\; \hat{q}_\tau(X_i) + \hat{s}(X_i)\varepsilon_i^*, \quad i = 1, \ldots, n.$$

Let $\hat{q}_\tau^*$ and $\hat{s}^*$ denote the quantile regression and scale function estimator defined analogously to $\hat{q}_\tau$ and $\hat{s}$, but based on the bootstrap sample $(X_1, Y_1^*), \ldots, (X_n, Y_n^*)$. Analogously to (3.3) the bootstrap version of the independence process is defined as

$$S_n^*(t,y) \;=\; \sqrt{n}\Big(\hat{F}^*_{X,\varepsilon,n}(t,y) - \hat{F}^*_{X,\varepsilon,n}(1-4h_n, y)\hat{F}^*_{X,\varepsilon,n}(t,\infty)\Big)$$

for $t \in [4h_n, 1-4h_n]$, $y \in \mathbb{R}$, and $S_n^*(t,y) = 0$ for $t \in [0, 4h_n) \cup (1-4h_n, 1]$, $y \in \mathbb{R}$. Here, similar to (3.2),

$$\hat{F}^*_{X,\varepsilon,n}(t,y) \;=\; \frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}_i^* \le y\} I\{4h_n < X_i \le t\} \frac{1}{\hat{F}_{X,n}(1-4h_n) - \hat{F}_{X,n}(4h_n)},$$

with $\hat{\varepsilon}_i^* = (Y_i^* - \hat{q}_\tau^*(X_i))/\hat{s}^*(X_i)$, $i = 1, \ldots, n$.
To obtain the conditional weak convergence we need the following additional assumptions.

(**B1**) We have for some $\delta > 0$

$$\frac{nh_n^2\alpha_n^2}{\log h_n^{-1}\log n} \to \infty, \qquad \frac{n\alpha_n h_n}{\log n} \to \infty, \qquad \frac{h_n}{\log n} = O(\alpha_n^{8\delta/3}), \qquad n\alpha_n^4 = o(1)$$

and there exists a $\kappa > 0$ such that

$$\frac{nh_n^{1+\frac{1}{\kappa}}\alpha_n^{2+\frac{2}{\kappa}}}{\log h_n^{-1}(\log n)^{1/\kappa}} \to \infty.$$

(**B2**) Let $E[|\varepsilon_1|^{\max(\upsilon, 2\kappa)}] < \infty$ for some $\upsilon > 1 + 2/\delta$ and with $\delta$ and $\kappa$ from assumption (**B1**).

Here, (**B2**) can be relaxed to $E[|\varepsilon_1|^{2\kappa}] < \infty$ if the process is only considered for $y \in [-c, c]$ for some $c > 0$ instead of for $y \in \mathbb{R}$.

**Theorem 3.6** *Under the location-scale model (2.1) and assumptions (**K1**)-(**K6**), (**A1**)-(**A5**), (**BW**) and (**B1**)-(**B2**) conditionally on $\mathcal{Y}_n$, the process $S_n^*$ converges weakly in $\ell^\infty([0,1]\times\mathbb{R})$ to the Gaussian process $S$ defined in Theorem 3.1, in probability.*

The proof is given in Appendix B.

**Remark 3.7** Let the Kolmogorov-Smirnov test statistic be defined as $K_n = \sup_{t,y} |S_n(t,y)|$ and its bootstrap version as $K_n^* = \sup_{t,y} |S_n^*(t,y)|$. Let the critical value $k_{n,1-\alpha}^*$ be obtained from

$$P(K_n^* \geq k_{n,1-\alpha}^* \mid \mathcal{Y}_n) = 1 - \alpha,$$

and reject the location-scale model if $K_n \geq k_{n,1-\alpha}^*$. Then from Theorems 3.1 and 3.6 it follows that the test has asymptotic level $\alpha$. Moreover if the location-scale model is not valid by Remark 3.3 we have $K_n \to \infty$ in probability, whereas with the same methods as in the proof of Theorem 3.6 it can be shown that $k_{n,1-\alpha}^*$ converges to a constant. Thus the power of the test converges to one. The finite sample performance of the bootstrap version of the Kolmogorov-Smirnov test is studied in Section 6. ■

**Remark 3.8** Recently, Sun (2006) and Feng, He and Hu (2011) proposed to use wild bootstrap in the setting of quantile regression. To follow the approach of the last-named authors, one would define $\varepsilon_i^* = v_i \hat{\varepsilon}_i$ such that $P^*(v_i \hat{\varepsilon}_i \leq 0 | X_i) = \tau$, e. g.

$$v_i = \pm 1 \text{ with probability } \begin{cases} \frac{1-\tau}{\tau} & \text{if } \hat{\varepsilon}_i \geq 0 \\ \frac{\tau}{1-\tau} & \text{if } \hat{\varepsilon}_i < 0. \end{cases}$$

However, then when calculating the conditional asymptotic covariance (following the proof in Appendix B), instead of $\tilde{F}_\varepsilon(y)$ the following term appears

$$\frac{1}{n} \sum_{i=1}^{n} P(v_i \hat{\varepsilon}_i \leq y \mid \mathcal{Y}_n) \overset{n\to\infty}{\longrightarrow} (1-\tau)(F_\varepsilon(y) - F_\varepsilon(-y)) + \tau.$$

One obtains $F_\varepsilon(y)$ (needed to obtain the same covariance as in Theorem 3.1) only for $y = 0$ or for median regression ($\tau = 0.5$) with symmetric error distributions, but not in general. Hence, wild bootstrap cannot be applied in the general context of procedures using empirical processes in quantile regression. ■

**Remark 3.9** Under assumption of the location-scale model model (2.1) the result of Theorem 3.1 can be applied to test for more specific model assumptions (e. g. testing goodness-of fit of a parametric model for the quantile regression function). The general approach is to build residuals $\hat{\varepsilon}_{i,0}$ that only under $H_0$ consistently estimate the errors (e. g. using a parametric estimator for the conditional quantile function). Recall the definition of $\hat{F}_{X,\varepsilon,n}$ in (3.2) and define analgously $\hat{F}_{X,\varepsilon_0,n}$ by using the residuals $\hat{\varepsilon}_{i,0}$. Then, analogously to (3.3), define

$$S_{n,0}(t,y) \;=\; \sqrt{n}\Big(\hat{F}_{X,\varepsilon_0,n}(t,y) - \hat{F}_{X,\varepsilon,n}(1-2h_n,y)\hat{F}_{X,\varepsilon,n}(t,\infty)\Big)$$

for $y \in \mathbb{R}$, $t \in [2h_n, 1-2h_n]$, and $S_{n,0}(t,y) = 0$ for $y \in \mathbb{R}$, $t \in [0, 2h_n) \cup (1-2h_n, 1]$. With this process the discrepancy from the null hypothesis can be measured. This approach is

considered in detail for the problem of testing monotonicity of conditional quantile functions in the next section.
A related approach, which however does not assume the location-scale model, is suggested to test for significance of covariables in quantile regression models by Volgushev et al. (2013).
■

# 4 Testing for monotonicity of conditional quantile curves

In this section, we consider a test for the hypothesis

$$H_0 : q_\tau(x) \text{ is increasing in } x.$$

To this end we define an increasing estimator $\hat{q}_{\tau,I}$, which consistently estimates $q_\tau$ if the hypothesis $H_0$ is valid, and consistently estimates some increasing function $q_{\tau,I} \neq q_\tau$ under the alternative that $q_\tau$ is not increasing. For any function $h : [0,1] \to \mathbb{R}$ define the increasing rearrangement $h_I$ as the generalized inverse of $h_I^{-1}$, i. e.

$$h_I(x) = \inf\{z \in \mathbb{R} | h_I^{-1}(z) \geq x\},$$

where

$$h_I^{-1}(z) \;=\; \int_0^1 I\{h(t) \leq z\}\, dt.$$

Note that if $h$ is increasing, then $h^{-1}$ is the generalized inverse of $h$ and we have $h_I = h$. See Anevski and Fougères (2007) and Neumeyer (2007) who consider increasing rearrangements of curve estimators for the sake of obtaining monotone versions of unconstrained estimators [and also Dette, Neumeyer and Pilz (2006) or Birke and Dette (2008) for a smooth version of the increasing rearrangements in the regression context]. We define the increasing estimator $\hat{q}_{\tau,I}$ as increasing rearrangement of the unconstrained estimator $\hat{q}_\tau$ of $q_\tau$ that was defined in Section 2. The quantity $\hat{q}_{\tau,I}$ estimates the increasing rearrangement $q_{\tau,I}$ of $q_\tau$. Only under the hypothesis $H_0$ of an increasing regression function we have $q_\tau = q_{\tau,I}$. In Figure 1 (right part) a non-increasing function $q_\tau$ and its increasing rearrangement $q_{\tau,I}$ are displayed.
Now we build (pseudo-) residuals

$$\hat{\varepsilon}_{i,I} = \frac{Y_i - \hat{q}_{\tau,I}(X_i)}{\hat{s}(X_i)}, \tag{4.1}$$

which estimate pseudo-errors $\varepsilon_{i,I} = (Y_i - q_{\tau,I}(X_i))/s(X_i)$ that coincide with the true errors $\varepsilon_i = (Y_i - q_\tau(X_i))/s(X_i)$ $(i = 1, \ldots, n)$ in general only under $H_0$. Let further $\hat{\varepsilon}_i$ denote the unconstrained residuals as defined in (3.1). The idea for the test statistic we suggest is the

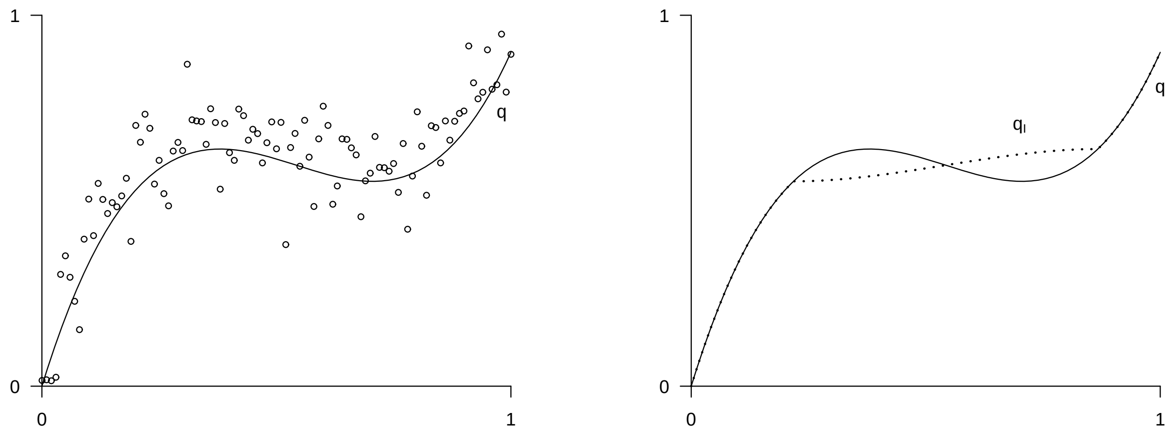


Figure 1: *Left part: True nonincreasing function $q_\tau$ for $\tau = 0.25$ with scatter-plot of a typical sample. Right part: $q_\tau$ (solid line) and increasing rearrangement $q_{\tau,I}$ (dotted line).*

following. Compared to the true errors $\varepsilon_1, \ldots, \varepsilon_n$, which are assumed to be i.i.d., the pseudo-errors $\varepsilon_{1,I}, \ldots, \varepsilon_{n,I}$ behave differently. If the true function $q_\tau$ is not increasing (e.g. like in Figure 1) and we calculate the pseudo-errors from $q_{\tau,I}$, they are no longer identically distributed. This effect is demonstrated in Figure 2 for a $\tau = 0.25$-quantile curve. Consider for instance the interval $[t, 1]$, where there are about 25% negative errors (left part) and in comparison too many negative pseudo-errors (right part). To detect such discrepancies from the null hypothesis, we estimate the pseudo-error distribution up to every $t \in [0, 1]$ (i. e. for the covariate values $X_i \leq t$) and compare with what is expected under $H_0$. To this end recall the definition of $\hat{F}_{X,\varepsilon,n}$ in (3.2) and define $\hat{F}_{X,\varepsilon_I,n}$ analogously, but using the constrained residuals $\hat{\varepsilon}_{i,I}$, $i = 1, \ldots, n$. Analogously to (3.3) define the process

$$S_{n,I}(t, y) = \sqrt{n}\Big(\hat{F}_{X,\varepsilon_I,n}(t, y) - \hat{F}_{X,\varepsilon,n}(1 - 2h_n, y)\hat{F}_{X,\varepsilon,n}(t, \infty)\Big) \tag{4.2}$$

for $y \in \mathbb{R}$, $t \in [2h_n, 1 - 2h_n]$, and $S_{n,I}(t, y) = 0$ for $y \in \mathbb{R}$, $t \in [0, 2h_n) \cup (1 - 2h_n, 1]$. For each fixed $t \in [0, 1]$, $y \in \mathbb{R}$, for $h_n \to 0$ the statistic $n^{-1/2} S_{n,I}(t, y)$ consistently estimates the expectation

$$\begin{aligned}
& E[I\{\varepsilon_{i,I} < y\} I\{X_i \leq t\}] - F_\varepsilon(y) F_X(t) \\
= \; & E\Big[I\Big\{\varepsilon_i < y + \frac{(q_{\tau,I} - q_\tau)(X_i)}{s(X_i)}\Big\} I\{X_i \leq t\}\Big] - F_\varepsilon(y) F_X(t) \\
= \; & \int_0^t \Big(F_\varepsilon\Big(y + \frac{(q_{\tau,I} - q_\tau)(x)}{s(x)}\Big) - F_\varepsilon(y)\Big) f_X(x)\, dx.
\end{aligned}$$

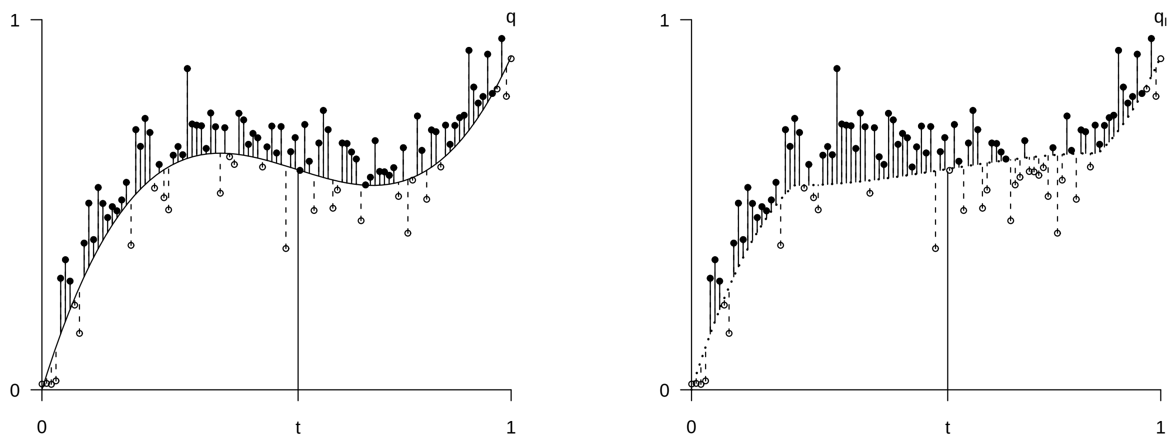


Figure 2: *Left part: True nonincreasing function* $q_\tau$ *for* $\tau = 0.25$ *and errors for the sample shown in Figure 1. Right part: Increasing rearrangement* $q_{\tau,I}$ *and pseudo-errors. (Positive errors are marked by solid points and solid lines, negative errors marked by circles and dashed lines.)*

If this term is zero for all $t \in [0,1]$ and all $y \in \mathbb{R}$, then also

$$\sup_{t\in[0,1]} \Big| \int_0^t \Big( F_\varepsilon\Big(\frac{(q_{\tau,I} - q_\tau)(x)}{s(x)}\Big) - F_\varepsilon(0)\Big) f_X(x)\, dx \Big| \;=\; 0$$

and from this it follows that $q_{\tau,I} = q_\tau$ is valid $F_X$-a. s. by the strict monotonicity of $F_\varepsilon$. We therefore use the Kolmogorov-Smirnov type statistic $K_n = \sup_{y\in\mathbb{R},t\in[0,1]} |S_{n,I}(t,y)|$ to obtain a consistent testing procedure, which rejects the null hypothesis for large values of $K_n$.
In the following theorem we state a weak convergence result for the process $S_{n,I}$ defined in (4.2).

**Theorem 4.1** *Under model (2.1) and assumptions **(K1)**-**(K6)**, **(A1)**-**(A5)** and **(BW)**, under the null hypothesis* $H_0$ *and the assumption* $\inf_{x\in[0,1]} q'_\tau(x) > 0$ *the process* $S_{n,I}$ *converges weakly in* $\ell^\infty([0,1]\times\mathbb{R})$ *to the Gaussian process* $S$ *defined in Theorem 3.1.*

The proof is given in Appendix A. For the application of the test for monotonicity we suggest a bootstrap version of the test analogously to the one considered in Section 3, but applying the increasing estimator to build new observations, i. e. $Y_i^* = \hat{q}_{\tau,I}(X_i) + \hat{s}(X_i)\varepsilon_i^*$, $i = 1, \ldots, n$. We have the following theoretical result.

**Theorem 4.2** *Under the assumptions of Theorem 4.1 and **(B1)**–**(B2)** the process* $S^*_{n,I}$*, conditionally on* $\mathcal{Y}_n$*, converges weakly in* $\ell^\infty([0,1]\times\mathbb{R})$ *to the Gaussian process* $S$ *defined in Theorem 3.1, in probability.*

The proof is given in Appendix B. A consistent asymptotic level-$\alpha$ test is constructed as in Remark 3.7.

**Remark 4.3** In the context of testing for monotonicity of mean regression curves Birke and Neumeyer (2013) based their tests on the observation that too many of the pseudo-errors are positive (see solid lines in Figure 2) on some subintervals of $[0,1]$ and too many are negative (see dashed lines) on other subintervals. Transferring this idea to the quantile regression model, one would consider a stochastic process

$$\tilde{S}_n(t,0) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_{i,I}\le 0\}I\{2h_n<X_i\le t\}-\hat{F}_{X,\varepsilon,n}(1-2h_n,0)I\{2h_n<X_i\le t\}\Big)$$

or alternatively (because $\hat{F}_{X,\varepsilon,n}(1-2h_n,0)$ estimates the known $F_\varepsilon(0)=\tau$)

$$R_n(t) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_{i,I}\le 0\}I\{X_i\le t\}-\tau I\{X_i\le t\}\Big)$$

where $t\in[0,1]$. For every $t\in[2h_n,1-2h_n]$ the processes count how many pseudo-residuals are positive up to covariates $\le t$. This term is then centered with respect to the estimated expectation under $H_0$ and scaled with $n^{-1/2}$. However, as can be seen from Theorem 4.1 the limit is degenerate for $y=0$, and hence we have under $H_0$ that

$$\sup_t|\tilde{S}_n(t,0)| = o_P(1). \tag{4.3}$$

Also, $\sup_{t\in[0,1]}|R_n(t)| = o_P(1)$ can be shown analogously. Hence, no critical values can be obtained for the Kolmogorov-Smirnov test statistics, and those test statistics are not suitable for our testing purpose. To explain the negligibility (4.3) heuristically, consider the case $t=1$ (now ignoring the truncation of covariates for simplicity of explanation). Then, under $H_0$, $n^{-1}\sum_{i=1}^n I\{\hat{\varepsilon}_{i,I}\le 0\}$ estimates $F_\varepsilon(0)=\tau$. But the information that $\varepsilon_i$ has $\tau$-quantile zero was already applied to estimate the $\tau$-quantile function $q_\tau$. Hence, one obtains $n^{-1}\sum_{i=1}^n I\{\hat{\varepsilon}_{i,I}\le 0\}-\tau = o_P(n^{-1/2})$. This observation is in accordance to the fact that $n^{-1}\sum_{i=1}^n\hat{\varepsilon}_i = o_P(n^{-1/2})$, when residuals are built from a mean regression model with centered errors [see Müller et al. (2004) and Kiwitt et al. (2008)].
Finally, consider the process

$$\tilde{S}_n(1-2h_n,y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_{i,I}\le y\}I\{2h_n<X_i\le 1-2h_n\} - \hat{F}_{X,\varepsilon,n}(1-2h_n,y)I\{2h_n<X_i\le 1-2h_n\}\Big)$$

i. e. the difference between the estimated distribution functions of pseudo-residuals $\hat{\varepsilon}_{i,I}$ and unconstrained residuals $\hat{\varepsilon}_i$ $(i=1,\ldots,n)$, respectively, scaled with $n^{1/2}$. An analogous process has been considered by Van Keilegom et al. (2008) for testing for parametric classes of mean regression functions. However, as can be seen from Theorem 4.1, in our case of testing for monotonicity the limit again is degenerate, i.e. $\mathrm{Var}(S(1,y)) = 0$ for all $y$, and hence $\sup_{y\in\mathbb{R}} |\tilde{S}_n(1,y)| = o_P(1)$. Similar observations can be made when typical distance based tests from lack-of-fit literature [for instance $L^2$-tests or residual process based procedures by Härdle and Mammen (1993) and Stute (1997), respectively] are considered in the problem of testing monotonicity of regression function, see Birke and Neumeyer (2013). The reason is that under $H_0$ the unconstrained and constrained estimators, $\hat{q}_\tau$ and $\hat{q}_{\tau,I}$, typically are first order asymptotically equivalent. This for estimation purposes very desirable property limits the possibilities to apply the estimator $\hat{q}_{\tau,I}$ for hypotheses testing. ■

# 5 Simulation results

In this section we show some simulation results for the bootstrap based tests introduced in this paper. If available we compare the results to already existing methods. Throughout the whole section we choose the bandwidths according to condition (**BW**) as $d_n = h_n = 2(\hat{\sigma}/n)^{1/7}$, $b_n = (\hat{\sigma}/n)^{2/7}$ and $\hat{\sigma}$ is the difference estimator proposed in Rice (1984) [see Yu and Jones (1997) for a related approach]. The degree of the local polynomial estimators of location and scale [see equation (2.2)] was chosen to be 3, the Kernel $K$ is the Gauss Kernel while $\kappa$ was chosen to be the Epanechnikov Kernel. The function $\Omega$ was defined through $\Omega(t) = \int_{-\infty}^t \omega(x)dx$ where $\omega(x) := (15/32)(3 - 10x^2 + 7x^4)I\{|x| \le 1\}$, which is a kernel of order 4 [see Gasser et al. (1985)]. For the choice of the distribution functions $G$ and $G_s$, we follow the procedure described in Dette and Volgushev (2008) who suggested a normal distribution such that the 5% and 95% quantiles coincide with the corresponding empirical quantities of the sample $Y_1, ..., Y_n$. Finally, the parameter $\alpha_n$ for generating the bootstrap residuals was chosen as $\alpha_n = \sqrt{2}n^{-1/4}\mathrm{median}(|\hat{\varepsilon}_1|, ..., |\hat{\varepsilon}_n|)$. In each of the 1000 simulation runs the quantiles of the test statistics were estimated as empirical quantiles from the bootstrap samples of 99 bootstrap repetitions.

## 5.1 Testing for location-scale models

The problem of testing the validity of location-scale models has previously been considered by Einmahl and van Keilegom (2008a), and we therefore compare the properties of our test statistic with theirs. To this end, we consider the following three models from Einmahl and van Keilegom (2008a)

$$\text{(model 1)} \quad Y|X = x \sim (x - 0.5x^2) + \frac{2+x}{10}\mathcal{N}(0,1), \quad X \sim U[0,1]$$

$$\text{(model 2)}\quad Y|X=x \sim (x-0.5x^2) + \Big(\frac{2+x}{10}\Big)\frac{\chi^2_{1/bx}-1/(bx)}{\sqrt{2/(bx)}},\quad X\sim U[0,1]$$

$$\text{(model 3)}\quad Y|X=x \sim (x-0.5x^2) + \Big(\frac{2+x}{10}\Big)\sqrt{1-(cx)^{1/4}}\,t_{2/(cx)^{1/4}},\quad X\sim U[0,1]$$

Observe that model 1 corresponds to a location-scale model while models 2 and 3 are not of location-scale type. The simulation results corresponding to different models and parameter setting are collected in Table 1. Model 1 corresponds to the null hypothesis, and we see that the test holds its level well. Models 2 and 3 represent alternatives. In model 2, our test outperforms all the tests in Einmahl and Van Keilegom (2008a), while in model 3 the power of the test is very low. An intuitive explanation of those differences is that Einmahl and Van Keilegom (2008a) scale their residuals to have the same variances while our residuals are scaled to have the same median absolute deviation. Under various alternative distributions, this leads to different power curves. In particular, the scaling with median absolute deviation leads to distribution functions whose supremum distance is large for chi-squared distributions but small for t-distributions with different parameters.

**Please insert Table 1 here**

## 5.2 Testing for monotonicity of quantile curves in a location-scale setting

Next, we considered the test for monotonicity of quantile curves that is introduced in Section 4. Here, we simulated the following two models that are both of location-scale type

$$\text{(model 4)}\qquad Y|X=x \sim 1+x-\beta e^{-50(x-0.5)^2}+0.2\mathcal{N}(0,1),\quad X\sim U[0,1]$$

$$\text{(model 5)}\qquad Y|X=x \sim \frac{x}{2}+2(0.1-(x-0.5)^2)\mathcal{N}(0,1),\quad X\sim U[0,1].$$

The results for models 4 and 5 are reported in Table 2 and Table 3, respectively. In model 4, all quantile curves are parallel and so all quantile curves have a similar monotonicity behavior. In particular, the parameter value $\beta = 0$ corresponds to strictly increasing quantile curves, for $\beta = 0.15$ the curves have a flat spot, and for $\beta > 0.15$ the curves have a small decreasing bump that gets larger for larger values of $\beta$. We simulated two different quantile ($\tau = 0.25$ and $\tau = 0.5$) values and see that in both cases the test has an increasing power for increasing values of $\beta$ and sample size. In particular, for a sample size of $n = 100$ the flat function corresponding to $\beta = 0.15$ are recognized as alternatives. Note that this is not in contradiction with our theory since all the results under $H_0$ require that the quantile curves be strictly increasing.

In model 5, the median is a strictly increasing function while the outer quantile curves are not increasing. In table 3, we report the simulation results for three different quantile values ($\tau = 0.25, \tau = 0.5$ and $\tau = 0.75$) and two sample sizes $n = 50$ and $n = 100$. For $n = 50$, the observed rejection probabilities are slightly above the nominal critical values (for $\tau = 0.5$), and the cases $\tau = 0.25$ and $\tau = 0.75$ are recognized as alternatives. For $n = 100$, the test holds its level for $\tau = 0.5$ and also shows a slow increase in power at the other quantiles. Overall, we can conclude that the proposed test shows a satisfactory behavior.

**Please insert Tables 2 and 3 here**

# 6 Conclusion

The paper at hand considered location-scale models in the context of nonparametric quantile regression. For the first time a test for model validity was investigated. It is based on the empirical independence process of covariates and residuals built from nonparametric estimators for the location and scale functions. The process converges weakly to a Gaussian process. A bootstrap version of the test was investigated in theory and by means of a simulation study. The theoretical results open a new toolbox to test for various model hypotheses in location-scale quantile models. As example we considered in detail the testing for monotonicity of a conditional quantile function in theory as well as in simulations. Similarly other structural assumptions on the location or the scale function can be tested. All weak convergence results are proved in the appendix and supplementary material in a detailed manner. A small simulation study demonstrated that the proposed method works well.

| | | | | | | |
|---|---|---|---|---|---|---|
| | model 1 | | | | | |
| | $n = 50$ | | | $n = 100$ | | |
| $\alpha$ | 0.05 | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 |
| | 0.04 | 0.085 | 0.173 | 0.037 | 0.08 | 0.176 |
| | model 2 | | | | | |
| | $n = 50$ | | | $n = 100$ | | |
| $b = 1$ | 0.084 | 0.145 | 0.267 | 0.146 | 0.255 | 0.41 |
| $b = 2.5$ | 0.275 | 0.383 | 0.542 | 0.643 | 0.743 | 0.851 |
| $b = 5$ | 0.408 | 0.539 | 0.689 | 0.823 | 0.888 | 0.945 |
| $b = 10$ | 0.459 | 0.602 | 0.726 | 0.796 | 0.872 | 0.935 |
| | model 3 | | | | | |
| | $n = 50$ | | | $n = 100$ | | |
| $c = 0.2$ | 0.045 | 0.094 | 0.171 | 0.052 | 0.089 | 0.169 |
| $c = 1$ | 0.061 | 0.102 | 0.215 | 0.052 | 0.103 | 0.221 |

Table 1: *Rejection probabilities for testing validity of the location-scale hypothesis [see Section 3] in models 1-3.*

| | $\tau = 0.25$ | | | | | | $\tau = 0.5$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n = 50$ | | | $n = 100$ | | | $n = 50$ | | | $n = 100$ | | |
| $\alpha$ | 0.05 | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 |
| $\beta = 0$ | 0.029 | 0.059 | 0.145 | 0.04 | 0.072 | 0.152 | 0.022 | 0.057 | 0.139 | 0.042 | 0.085 | 0.175 |
| $\beta = 0.15$ | 0.049 | 0.088 | 0.187 | 0.076 | 0.143 | 0.261 | 0.055 | 0.099 | 0.191 | 0.079 | 0.153 | 0.281 |
| $\beta = 0.25$ | 0.059 | 0.118 | 0.226 | 0.139 | 0.242 | 0.401 | 0.073 | 0.118 | 0.240 | 0.131 | 0.229 | 0.391 |
| $\beta = 0.45$ | 0.132 | 0.216 | 0.389 | 0.279 | 0.442 | 0.644 | 0.117 | 0.213 | 0.365 | 0.201 | 0.315 | 0.510 |

Table 2: *Rejection probabilities for the test for monotonicity of quantile curves in model 4.*

| | $n = 50$ | | | $n = 100$ | | |
|---|---|---|---|---|---|---|
| $\alpha$ | 0.05 | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 |
| $\tau = 0.25$ | 0.213 | 0.277 | 0.390 | 0.246 | 0.32 | 0.449 |
| $\tau = 0.5$ | 0.083 | 0.138 | 0.239 | 0.054 | 0.1 | 0.192 |
| $\tau = 0.75$ | 0.149 | 0.235 | 0.364 | 0.212 | 0.316 | 0.437 |

Table 3: *Rejection probabilities for the test for monotonicity of quantile curves in model 5. Different rows correspond to the* $0.25, 0.5$ *and* $0.75$ *quantile curves, respectively.*

# A Proof of weak convergence results

Before beginning with the proof, we give a brief overview of the results. The proofs of the main Theorems (Theorem 3.1 and 4.1) and the bootstrap versions (Theorems 3.6 and 4.2) are contained in Appendixes A and B, respectively. Technical details needed in the proofs of those results can be found in Appendix C.1 (in the supplementary material). Finally, Appendix C.2 (as well in the supplement) contains basic results on linearized versions and differentiability of the quantile estimator $\hat{q}_\tau$, scale estimator $\hat{s}$ and the corresponding bootstrap versions.

**Proof of Theorem 3.1.** For the joint empirical distribution function defined in (3.2) we have

$$\hat{F}_{X,\varepsilon,n}(t,y) \;=\; \frac{1}{n}\sum_{i=1}^n I\Big\{\varepsilon_i \le y\frac{\hat{s}(X_i)}{s(X_i)} + \frac{\hat{q}_\tau(X_i)-q_\tau(X_i)}{s(X_i)}\Big\} I\{2h_n < X_i \le t\}.$$

Note that by Lemma C.1 without changing the asymptotic distribution of the process the residuals $\hat{\varepsilon}_i$ can be replaced by their versions obtained from linearized estimators $\hat{q}_{\tau,L}$, $\hat{s}_L$ instead of $\hat{q}_\tau$, $\hat{s}$ (see Appendix C.2 for the definitions). Thus we have

$$\hat{F}_{X,\varepsilon,n}(t,y) \;=\; \frac{1}{n}\sum_{i=1}^n I\Big\{\varepsilon_i \le y\frac{\hat{s}_L(X_i)}{s(X_i)} + \frac{\hat{q}_{\tau,L}(X_i)-q_\tau(X_i)}{s(X_i)}\Big\} I\{2h_n < X_i \le t\} + o_p(\frac{1}{\sqrt{n}}).$$

From this we obtain the expansion

$$\begin{aligned}\hat{F}_{X,\varepsilon,n}(t,y) \;&=\; \frac{1}{n}\sum_{i=1}^n I\{\varepsilon_i \le y\} I\{2h_n < X_i \le t\} \\ &\quad + \int_{2h_n}^{1-2h_n} \Big(F_\varepsilon\Big(y\frac{\hat{s}_L(x)}{s(x)} + \frac{\hat{q}_{\tau,L}(x)-q_\tau(x)}{s(x)}\Big) - F_\varepsilon(y)\Big) I\{x \le t\} f_X(x)\,dx \\ &\quad + o_p(\frac{1}{\sqrt{n}})\end{aligned} \tag{A.1}$$

uniformly with respect to $t \in [2h_n, 1-2h_n]$ and $y \in \mathbb{R}$ by the following argumentation. Consider the empirical process

$$G_n(\varphi) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\varphi(X_i,\varepsilon_i) - E[\varphi(X_i,\varepsilon_i)]\Big), \quad \varphi \in \mathcal{F},$$

indexed by the following class of functions,

$$\begin{aligned}\mathcal{F} \;=\; \Big\{&(X,\varepsilon) \mapsto I\{\varepsilon \le y d_2(X) + d_1(X)\} I\{h < X\} I\{X \le t\} - I\{\varepsilon \le y\} I\{h < X\} I\{X \le t\} \\ &\Big|\; y \in \mathbb{R}, h, t \in [0,1], d_1 \in C_1^{1+\delta}[0,1], d_2 \in \tilde{C}_2^{1+\delta}[0,1]\Big\},\end{aligned}$$

for some arbitrary $\delta \in (0,1)$, where the function class $C_c^{1+\delta}[0,1]$ is defined as the set of differentiable functions $g : [0,1] \to \mathbb{R}$ with derivatives $g'$ such that

$$\max\Big\{\sup_{x\in[0,1]}|g(x)|, \sup_{x\in[0,1]}|g'(x)|\Big\} + \sup_{x,z\in[0,1]}\frac{|g'(x)-g'(z)|}{|x-z|^\delta} \le c$$

[see van der Vaart and Wellner (1996, p. 154)]. We further by slight abuse of notation define the subset $\tilde{C}_2^{1+\delta}([0,1])$ of $C_1^{1+\delta}([0,1])$ by the additional constraint $\inf_{x\in[0,1]} g(x) \ge 1/2$. Now $\mathcal{F}$ is a product of the uniformly bounded Donsker classes $\{(X,\varepsilon) \mapsto I\{h < X\}I\{X \le t\}|h,t \in [0,1]\}$ and $\{(X,\varepsilon) \mapsto I\{\varepsilon \le yd_2(X)+d_1(X)\} - I\{\varepsilon \le y\}|y \in \mathbb{R}, d_1 \in C_1^{1+\delta}[0,1], d_2 \in \tilde{C}_2^{1+\delta}[0,1]\}$ [the Donsker property for the second class is shown in Lemma 1 by Akritas and Van Keilegom (2001)] and is therefore Donsker as well (Ex. 2.10.8, van der Vaart and Wellner (1996), p. 192). The remaining part of the proof for equality (A.1) follows exactly the lines of the end of the proof of Lemma 1, Akritas and Van Keilegom (2001), p. 567, using the inequality

$$\begin{aligned}&\mathrm{Var}\Big(I\{\varepsilon_1 \le yd_2(X_1)+d_1(X_1)\}I\{h < X_1\}I\{X_1 \le s\} - I\{\varepsilon_1 \le y\}I\{h<X_1\}I\{X_1 \le s\}\Big)\\ \le\ &E\Big[\Big(I\{\varepsilon_1 \le yd_2(X_1)+d_1(X_1)\} - I\{\varepsilon_1 \le y\}\Big)^2\Big]\end{aligned}$$

and applying Lemmata C.3 and C.4 below. For $\varphi = \varphi_{h,t,y,d_1,d_2}$ we obtain

$$\sup_{\substack{y\in\mathbb{R},\\ t\in[2h_n,1-2h_n]}}\Big|G_n\Big(\varphi_{2h_n,t,y,\frac{\hat{q}_{\tau,L}-q_\tau}{s},\frac{\hat{s}_L}{s}}\Big)\Big| = o_P(1)$$

and thus (A.1).

Further, by a Taylor expansion we obtain from (A.1) together with assumption (**A4**) that

$$\begin{aligned}\hat{F}_{X,\varepsilon,n}(t,y) \ =\ &\frac{1}{n}\sum_{i=1}^n I\{\varepsilon_i \le y\}I\{2h_n < X_i \le t\} + yf_\varepsilon(y)\int_{2h_n}^{1-2h_n}\frac{\hat{s}_L(x)-s(x)}{s(x)}I\{x\le t\}f_X(x)\,dx\\ &+ f_\varepsilon(y)\int_{2h_n}^{1-2h_n}\frac{\hat{q}_{\tau,L}(x)-q_\tau(x)}{s(x)}I\{x\le t\}f_X(x)\,dx + o_P(\frac{1}{\sqrt{n}})\end{aligned}$$

uniformly with respect to $t \in [2h_n, 1-2h_n]$ and $y \in \mathbb{R}$. Applying Lemma C.2 below it follows that

$$\begin{aligned}\hat{F}_{X,\varepsilon,n}(t,y) \ =\ &\frac{1}{n}\sum_{i=1}^n I\{\varepsilon_i \le y\}I\{2h_n < X_i \le t\} - \phi(y)\frac{1}{n}\sum_{i=1}^n(I\{\varepsilon_i \le 0\}-\tau)I\{2h_n < X_i \le t\}\\ &- \psi(y)\frac{1}{n}\sum_{i=1}^n\Big(I\{|\varepsilon_i| \le 1\} - \frac{1}{2}\Big)I\{2h_n < X_i \le t\} + o_P(\frac{1}{\sqrt{n}}),\end{aligned}$$

where $\phi$ and $\psi$ are defined in the assertion of the theorem. Thus noting that $\hat{F}_{X,n}(1-2h_n) - \hat{F}_{X,n}(2h_n) = F_X(1-2h_n) - F_X(2h_n) + o_P(1) = 1 + o_P(1)$, from the definition (3.3) we obtain by Slutsky's lemma that

$$S_n(t,y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i \le y\} - F_\varepsilon(y) - \phi(y)(I\{\varepsilon_i \le 0\} - \tau) - \psi(y)\Big(I\{|\varepsilon_i| \le 1\} - \frac{1}{2}\Big)\Big)$$
$$\times \Big(I\{2h_n < X_i \le t\} - I\{2h_n < X_i \le 1-2h_n\}\frac{\hat F_{X,n}(t) - \hat F_{X,n}(2h_n)}{\hat F_{X,n}(1-2h_n) - \hat F_{X,n}(2h_n)}\Big)$$
$$+\, o_P(1).$$

uniformly with respect to $t \in [2h_n, 1-2h_n]$ and $y \in \mathbb{R}$. Note that the dominating part of this process vanishes in the boundary points $t = 2h_n$ and $t = 1-2h_n$. Further, from $\hat F_{X,n}(t) = F_X(t) + O_p(n^{-1/2})$ uniformly in $t \in [0,1]$ and $F_X(2h_n) \to 0$, $F_X(1-2h_n) \to 1$ we have

$$S_n(t,y) \;=\; S_{n,1}(t,y) + o_P(1),$$

uniformly with respect to $t \in [0,1]$, $y \in \mathbb{R}$, where $S_{n,1}(t,y) = 0$ for $t \in [0,2h_n) \cup (1-2h_n, 1]$ and

$$S_{n,1}(t,y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n g(\varepsilon_i, y)\Big(I\{2h_n < X_i \le t\} - I\{2h_n < X_i \le 1-2h_n\}F_X(t)\Big)$$

for $t \in [2h_n, 1-2h_n]$ and $y \in \mathbb{R}$, where $g(\varepsilon_i, y) = I\{\varepsilon_i \le y\} - F_\varepsilon(y) - \phi(y)(I\{\varepsilon_i \le 0\} - \tau) - \psi(y)(I\{|\varepsilon_i| \le 1\} - \frac{1}{2})$ is centered and independent of $X_i$. The first assertion of the theorem now follows if we show that for

$$S_{n,2}(t,y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n g(\varepsilon_i, y)\Big(I\{X_i \le t\} - F_X(t)\Big), \quad t \in [0,1], y \in \mathbb{R},$$

we have $\sup_{t\in[0,1], y\in\mathbb{R}} |S_{n,1}(t,y) - S_{n,2}(t,y)| = o_P(1)$, which is equivalent to

$$\text{(A.2)} \qquad \sup_{t\in[2h_n,1-2h_n], y\in\mathbb{R}} |S_{n,1}(t,y) - S_{n,2}(t,y)| \;=\; o_P(1)$$

together with

$$\text{(A.3)} \qquad \sup_{t\in[0,2h_n)\cup(1-2h_n,1], y\in\mathbb{R}} |S_{n,2}(t,y)| \;=\; o_P(1).$$

We will only show (A.2); (A.3) follows by similar arguments. Note that $S_{n,1}(t,y) - S_{n,2}(t,y) = G_n(h_n, t, y)$ for $t \in [2h_n, 1-2h_n]$, $y \in \mathbb{R}$, where the process

$$G_n(h,t,y) = \frac{-1}{\sqrt{n}}\sum_{i=1}^n g(\varepsilon_i, y)(I\{X_i \le t\} - F_X(t))I\{X_i \in [0,2h) \cup (1-2h, 1]\}$$

indexed in $h \in [0, \frac{1}{4}]$, $t \in [0,1]$, $y \in \mathbb{R}$, converges weakly to a centered Gaussian process $G$ with asymptotic variance

$$\begin{aligned}\operatorname{Var}(G(h,t,y)) &= E[g^2(\varepsilon_1, y)]\Big((F_X(t\wedge 2h) + F_X(t) - F_X(t\wedge(1-2h)))(1-2F_X(t)) \\ &\quad + F_X^2(t)(F_X(2h) + 1 - F_X(1-2h))\Big).\end{aligned}$$

For $h = h_n \to 0$ this asymptotic variance vanishes uniformly with respect to $y$ and $t$. From asymptotic equicontinuity of $G_n$ (confer van der Vaart and Wellner, 1996, p. 89/90), using the asymptotic variance as semi-metric, with $G_n(0,t,y) \equiv 0$ it follows that $\sup_{t,y} |G_n(h_n,t,y)| = o_P(1)$ and thus (A.2).
Hence, we have shown the first assertion of the theorem, i. e. $S_n = S_{n,2} + o_P(1)$ uniformly. Weak convergence of $S_{n,2}$ (and thus of $S_n$) to a centered Gaussian process with the asserted covariance structure follows by standard arguments. □

**Proof of Theorem 4.1.** Note that the process $S_{n,I}$ as defined in (4.2) reduces to $S_n$ defined in (3.3) when pseudo-residuals are replaced by unconstrained residuals, i. e. when $\hat{q}_{\tau,I}$ is replaced by $\hat{q}_\tau$ in the definition (4.1). Now let $c = \inf_{x\in[0,1]} q'_\tau(x)$ and note that by our assumptions $c > 0$ and by Lemmata C.3 and C.4 we have $P(\sup_{x\in[0,1]} |\hat{q}'_\tau(x) - q'_\tau(x)| > \frac{c}{2}) \to 0$ for $n \to \infty$. For every $\epsilon > 0$ it follows that

$$\begin{aligned} & P\Big(\sup_{t\in[0,1],y\in\mathbb{R}} |S_n(t,y) - S_{n,I}(t,y)| > \epsilon\Big) \\ \leq\ & P\Big(\sup_{t\in[0,1],y\in\mathbb{R}} |S_n(t,y) - S_{n,I}(t,y)| > \epsilon\ ,\ \sup_{x\in[0,1]} |\hat{q}'_\tau(x) - q'_\tau(x)| \leq \frac{c}{2}\Big) + o(1) \\ \leq\ & P\Big(\sup_{t\in[0,1],y\in\mathbb{R}} |S_n(t,y) - S_{n,I}(t,y)| > \epsilon\ ,\ \inf_{x\in[0,1]} \hat{q}'_\tau(x) > 0\Big) + o(1) \qquad \text{(A.4)} \\ =\ & o(1). \end{aligned}$$

Here the last equality is due to the following argumentation. If $\inf_{x\in[0,1]} \hat{q}'_\tau(x) > 0$, then $\hat{q}_\tau$ is strictly increasing, and for any increasing function $\hat{q}_\tau$ the increasing rearrangement, here $\hat{q}_{\tau,I}$, equals the original function function, i. e. $\hat{q}_{\tau,I} = \hat{q}_\tau$ (see Section 4). But then, $S_n(t,y) = S_{n,I}(t,y)$ for all $t, y$ and the probability in (A.4) is zero. Hence, we have shown that uniformly with respect to $t \in [0,1]$ and $y \in \mathbb{R}$, $S_n(t,y) = S_{n,I}(t,y) + o_P(1)$ and therefore the assertion follows from Theorem 3.1. □

# B Validity of bootstrap

**Proof of Theorem 3.6.**

In Lemma C.1 it is shown that in the process $\hat{F}^*_{X,\varepsilon,n}$ the residuals $\hat{\varepsilon}^*_i$ can be replaced by the linearized versions $\hat{\varepsilon}^*_{i,L}$ (see Appendix C.2 for the definitions). Now let $\tilde{f}_\varepsilon$ denote the density corresponding to $\tilde{F}_\varepsilon$. Then note that under assumptions (**B1**) and (**B2**), Lemma 2 and Proposition 4 in Neumeyer (2009a) are valid as well as

$$\text{(B.1)} \qquad \sup_{y\in\mathbb{R}} |\tilde{f}_\varepsilon(y) - f_\varepsilon(y)| = o((\frac{h_n}{\log n})^{1/2}), \quad \sup_{y\in\mathbb{R}} |y\tilde{f}_\varepsilon(y) - yf_\varepsilon(y)| = o(1)$$

almost surely, where those two additional results can be shown analogously to the first part of Lemma 2 in the aforementioned paper. The second result in (B.1) will also play an important role in the proof of Lemma C.2. Using this as well as Lemma C.3 (instead of Lemma 3 in the reference) we obtain analogously to the proofs of Lemma 1(i) and Theorem 2 in Neumeyer (2009a) that

$$\begin{aligned}
&\hat{F}^*_{X,\varepsilon,n}(t,y)\\
&= \frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}^*_{i,L} \le y\} I\{4h_n < X_i \le 1-4h_n\} + o_P(\frac{1}{\sqrt{n}})\\
&= \frac{1}{n}\sum_{i=1}^n I\{\varepsilon^*_i \le y\} I\{4h_n < X_i \le t\}\\
&\quad + \int \Big(\tilde{F}_\varepsilon\Big(y\frac{\hat{s}^*_L(x)}{\hat{s}_L(x)} + \frac{\hat{q}^*_{\tau,L}(x) - \hat{q}_{\tau,L}(x)}{\hat{s}_L(x)}\Big) - \tilde{F}_\varepsilon(y)\Big) I\{4h_n < x \le t\} f_X(x)\, dx\\
&\quad + o_P(\frac{1}{\sqrt{n}})
\end{aligned}$$

uniformly with respect to $t \in (4h_n, 1-4h_n]$, $y \in \mathbb{R}$. By a Taylor expansion for $\tilde{F}_\varepsilon$ and due to Lemma C.2 we obtain

$$\begin{aligned}
\hat{F}^*_{X,\varepsilon,n}(t,y) \;=\; & \frac{1}{n}\sum_{i=1}^n I\{4h_n < X_i \le t\}\Big(I\{\varepsilon^*_i \le y\} - \tilde{\psi}_n(y)\Big(I\{|\varepsilon^*_i| \le 1\} - \frac{1}{2}\Big)\\
&\qquad - \tilde{\phi}_n(y)\Big(I\{\varepsilon^*_i \le 0\} - \tau\Big)\Big)\\
& + o_P(\frac{1}{\sqrt{n}})
\end{aligned}$$

uniformly with respect to $t \in (4h_n, 1-4h_n]$, $y \in \mathbb{R}$, where

$$\tilde{\psi}_n(y) = \frac{y\tilde{f}_\varepsilon(y)}{f_{|\varepsilon|}(1)}, \quad \tilde{\phi}_n(y) = \frac{\tilde{f}_\varepsilon(y)}{f_\varepsilon(0)}\Big(1 - y\frac{f_\varepsilon(1) - f_\varepsilon(-1)}{f_{|\varepsilon|}(1)}\Big).$$

By the definition of the process $S_n^*$ one now directly has

$$\begin{aligned}
&S_n^*(t,y)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i^*\le y\} - \tilde{\psi}_n(y)\Big(I\{|\varepsilon_i^*|\le 1\}-\frac{1}{2}\Big) - \tilde{\phi}_n(y)\Big(I\{\varepsilon_i^*\le 0\}-\tau\Big)\Big)\\
&\quad \times\Big(I\{4h_n<X_i\le t\} - I\{4h_n<X_i\le 1-4h_n\}\frac{\hat{F}_{X,n}(t)-\hat{F}_{X,n}(4h_n)}{\hat{F}_{X,n}(1-4h_n)-\hat{F}_{X,n}(4h_n)}\Big)\\
&\quad + o_P(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n g_n(\varepsilon_i^*,y)\Big(I\{4h_n<X_i\le t\} - I\{4h_n<X_i\le 1-4h_n\}\frac{\hat{F}_{X,n}(t)-\hat{F}_{X,n}(4h_n)}{\hat{F}_{X,n}(1-4h_n)-\hat{F}_{X,n}(4h_n)}\Big)\\
&\quad + o_P(1)
\end{aligned}$$

uniformly with respect to $t\in(4h_n,1-4h_n]$, $y\in\mathbb{R}$, with

$$\begin{aligned}
&g_n(\varepsilon_i^*,y)\\
&= I\{\varepsilon_i^*\le y\} - \tilde{F}_\varepsilon(y) - \tilde{\phi}_n(y)\Big(I\{\varepsilon_i^*\le 0\}-\tilde{F}_\varepsilon(0)\Big) - \tilde{\psi}_n(y)\Big(I\{|\varepsilon_i^*|\le 1\}-\tilde{F}_\varepsilon(1)+\tilde{F}_\varepsilon(-1)\Big).
\end{aligned}$$

Note that $E[g_n(\varepsilon_i^*,y)\mid\mathcal{Y}_n]=0$ and the dominating part of the process $S_n^*$ vanishes in the boundary points $t=4h_n$ and $t=1-4h_n$, for all $y\in\mathbb{R}$. Similarly to the corresponding arguments in the proof of Theorem 3.1 (but with more technical effort) it can be shown that this process is equivalent in terms of conditional weak convergence in $\ell^\infty([0,1]\times\mathbb{R})$ in probability to the process

$$S_{n,2}^*(t,y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n g_n(\varepsilon_i^*,y)\Big(I\{X_i\le t\}-\hat{F}_{X,n}(t)\Big),\quad t\in[0,1], y\in\mathbb{R}.$$

Details are omitted for the sake of brevity.

It is easy to see that the conditional covariances $\mathrm{Cov}(S_{n,2}^*(s,y),S_{n,2}^*(t,z)\mid\mathcal{Y}_n)$ converge almost surely to $\mathrm{Cov}(S(s,y),S(t,z))$ as defined in Theorem 3.1. Thus it remains to show conditional tightness and conditional fidi convergence of $S_{n,2}^*$. To obtain the latter we use Cramér-Wold's device. Let $k\in\mathbb{N}$, $(y_1,t_1),\ldots,(y_k,t_k)\in\mathbb{R}\times[0,1]$, $a_1,\ldots,a_k\in\mathbb{R}$ and $Z_n=\sum_{j=1}^k a_j S_{n,2}^*(y_j,t_j)=n^{-1/2}\sum_{i=1}^n z_{n,i}$. Note that for some constant $c$, $|g_n(\varepsilon_i^*,y)(I\{X_i\le t\}-\hat{F}_{X,n}(t))|\le 1+c(1+y)\tilde{f}_\varepsilon(y)$, which converges almost surely to $1+c(1+y)f_\varepsilon(y)$ due to (B.1) and thus is almost surely bounded. From this the validity of the conditional Lindeberg condition easily follows, i. e.

$$L_n(\delta) \;=\; \frac{1}{n}\sum_{i=1}^n E[z_{n,i}^2 I\{|z_{n,i}|>n^{1/2}\delta\}\mid\mathcal{Y}_n]\to 0 \text{ almost surely, for all } \delta>0.$$

Finally, to prove conditional tightness we apply the quantile transformation $F_\varepsilon^{-1}$ to be able to consider the process in $\ell^\infty([0,1]^2)$ and use the decomposition $S_{n,2}^*(F_\varepsilon^{-1}(s),t)=U_n(s,t)-$

$\tilde{\phi}_n(F_\varepsilon^{-1}(s))V_{n,1}(t) - \tilde{\psi}_n(F_\varepsilon^{-1}(s))V_{n,2}(t) - \hat{F}_{X,n}(t)W_n(s)$, where

$$\begin{aligned}
U_n(s,t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i^* \le F_\varepsilon^{-1}(s)\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(s))\Big)I\{X_i \le t\} \\
V_{n,1}(t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i^* \le 0\} - \tilde{F}_\varepsilon(0)\Big)\Big(I\{X_i \le t\} - \hat{F}_{X,n}(t)\Big) \\
V_{n,2}(t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{|\varepsilon_i^*| \le 1\} - \tilde{F}_\varepsilon(1) + \tilde{F}_\varepsilon(-1)\Big)\Big(I\{X_i \le t\} - \hat{F}_{X,n}(t)\Big) \\
W_n(s) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i^* \le F_\varepsilon^{-1}(s)\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(s))\Big).
\end{aligned}$$

We show tightness seperately for the four processes in the decomposition. For $U_n$ we apply Bickel and Wichura's (1971) condition. To this end first consider neighbouring blocks $A = (x,y] \times (s,t]$ and $B = (y,z] \times (s,t]$ in $[0,1]^2$ and note that by a straightforward calculation

$$\begin{aligned}
&E\Big[\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{F_\varepsilon^{-1}(x) < \varepsilon_i^* \le F_\varepsilon^{-1}(y)\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(x))\Big)I\{s < X_i \le t\}\Big)^2 \\
&\quad \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{F_\varepsilon^{-1}(y) < \varepsilon_i^* \le F_\varepsilon^{-1}(z)\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(z)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y))\Big)I\{s < X_i \le t\}\Big)^2 \Big| \mathcal{Y}_n\Big] \\
\le\ & \frac{1}{n}E\Big[\Big(I\{x < F_\varepsilon(\varepsilon_1^*) \le y\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(x))\Big)^2 \\
&\quad \Big(I\{y < F_\varepsilon(\varepsilon_1^*) \le z\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(z)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y))\Big)^2 \Big| \mathcal{Y}_n\Big](\hat{F}_{X,n}(t) - \hat{F}_{X,n}(s)) \\
&+ E\Big[\Big(I\{x < F_\varepsilon(\varepsilon_1^*) \le y\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(x))\Big)^2 \Big| \mathcal{Y}_n\Big] \\
&\times E\Big[\Big(I\{y < F_\varepsilon(\varepsilon_1^*) \le z\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(z)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y))\Big)^2 \Big| \mathcal{Y}_n\Big](\hat{F}_{X,n}(t) - \hat{F}_{X,n}(s))^2 \\
&+ 2\Big(E\Big[\Big(I\{x < F_\varepsilon(\varepsilon_1^*) \le y\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(x))\Big) \\
&\quad \Big(I\{y < F_\varepsilon(\varepsilon_1^*) \le z\} - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(z)) + \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y))\Big) \Big| \mathcal{Y}_n\Big]\Big)^2(\hat{F}_{X,n}(t) - \hat{F}_{X,n}(s))^2.
\end{aligned}$$

Now calculating the conditional expectations one obtains the simple bound

$$\Big(\tilde{F}_\varepsilon(F_\varepsilon^{-1}(y)) - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(x))\Big)\Big(\tilde{F}_\varepsilon(F_\varepsilon^{-1}(z)) - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(y))\Big)(\hat{F}_{X,n}(t) - \hat{F}_{X,n}(s))^2 = \mu_n(A)\mu_n(B).$$

Here we have used that either $\hat{F}_{X,n}(t) - \hat{F}_{X,n}(s) = 0$ or $\frac{1}{n} \le \hat{F}_{X,n}(t) - \hat{F}_{X,n}(s)$. Further $\mu_n((a,b] \times (c,d]) = (\tilde{F}_\varepsilon(F_\varepsilon^{-1}(b)) - \tilde{F}_\varepsilon(F_\varepsilon^{-1}(a)))(\hat{F}_{X,n}(d) - \hat{F}_{X,n}(c))$ converges almost surely to $\mu((a,b] \times (c,d]) = (b-a)(F_X(d) - F_X(c))$. By a similar but simpler calculation for neighbouring blocks $A = (x,y] \times (s,t]$ and $B = (x,y] \times (t,u]$ one altogether obtains conditional tightness of $U_n$ almost surely.

For the other three processes first note that conditional weak convergence of $V_{n,1}$ and $V_{n,2}$ to centered Gaussian processes, almost surely, can be shown analogously to the proof of bootstrap validity in Birke and Neumeyer (2013), following the proof of Lemma A.3 by Stute et al. (1998). Now to prove tightness of the second process in the decomposition in terms of asymptotic stochastic equicontinuity the limit $\lim_{\delta\to 0}\lim_{n\to\infty}$ for the conditional probability

$$
\begin{aligned}
&P\Bigg(\sup_{\substack{|s_1-s_2|<\delta\\|t_1-t_2|<\delta}}|\tilde{\phi}_n(F_\varepsilon^{-1}(s_1))V_{n,1}(t_1)-\tilde{\phi}_n(F_\varepsilon^{-1}(s_2))V_{n,1}(t_2)|>\eta\;\Bigg|\;\mathcal{Y}_n\Bigg)\\
&\le P\Bigg(\sup_{y\in\mathbb{R}}|\tilde{\phi}_n(y)-\phi(y)|+\Big(\sup_{|s_1-s_2|<\delta}|\phi(F_\varepsilon^{-1}(s_1))-\phi(F_\varepsilon^{-1}(s_2))|\Big)\sup_{t\in[0,1]}|V_{n,1}(t)|>\frac{\eta}{2}\;\Bigg|\;\mathcal{Y}_n\Bigg)\\
&\quad+P\Bigg(\Big(\sup_{y\in\mathbb{R}}|\tilde{\phi}_n(y)-\phi(y)|+\sup_{y\in\mathbb{R}}|\phi(y)|\Big)\sup_{|t_1-t_2|<\delta}|V_{n,1}(t_1)-V_{n,1}(t_2)|>\frac{\eta}{2}\;\Bigg|\;\mathcal{Y}_n\Bigg)
\end{aligned}
$$

should be zero for almost all sequences $(X_1,Y_1),(X_2,Y_2),\ldots$, for all $\eta>0$. For the first probability this follows by uniform almost sure convergence of $\tilde{\phi}_n$ to $\phi$, uniform continuity of $\phi$ and conditional weak convergence of $\sup_{t\in[0,1]}|V_{n,1}(t)|$ almost surely. For the second probability it follows from uniform almost sure convergence of $\tilde{\phi}_n$, boundedness of $\phi$ and conditional tightness of $V_{n,1}$. The proof of conditional tightness of the third process in the decomposition, i. e. $\tilde{\psi}_n(F_\varepsilon^{-1}(s))V_{n,2}(t)$, is completely analogous. Also, conditional tightness of $\hat{F}_{X,n}(t)W_n(s)$ follows from almost sure uniform convergence of $\hat{F}_{X,n}$ to $F_X$, uniform continuity of $F_X$ and conditional weak convergence (and, thus, conditional tightness) of $W_n$. The latter is completely analogous to Theorem 4 by Neumeyer (2009a).
This completes the proof. □

**Proof of Theorem 4.2.**
Theorem 4.2 follows from Theorem 3.6 in the same manner as Theorem 4.1 follows from Theorem 3.1 by application of Lemmata C.3 and C.4. □

# References

**J. Abrevaya** (2005). Isotonic quantile regression: asymptotics and bootstrap. Sankhyā 67, 187–199.

**M. Akritas and I. Van Keilegom** (2001). *Nonparametric estimation of the residual distribution.* Scand. J. Statist. 28, 549–567.

**D. Anevski and A.-L. Fougères** (2007). *Limit properties of the monotone rearrangement for density and regression function estimation* arXiv:0710.4617v1

**Y. Baraud, S. Huet and B. Laurent** (2003). *Adaptive tests of qualitative hypotheses* ESAIM, Probab. Statist. 7, 147–159.

**P. J. Bickel and M. J. Wichura** (1971). *Convergence criteria for multiparameter stochastic processes and some applications.* Ann. Math. Statist. 42, 1656–1670.

**M. Birke and H. Dette** (2007). *Testing strict monotonicity in nonparametric regression.* Math. Meth. Statist. 16, p. 110–123.

**M. Birke and H. Dette** (2008). *A note on estimating a smooth monotone regression by combining kernel and density estimates.* J. Nonparam. Statist. 20, 679-690.

**M. Birke and N. Neumeyer** (2013). *Testing monotonicity of regression functions - an empirical process approach.* Scand. J. Statist., to appear.

**J.R. Blum, J. Kiefer and M. Rosenblatt** (1961). *Distribution free tests of independence based on the sample distribution functions.* Ann. Math. Stat. 32, 485–498.

**A. W. Bowman, M. C. Jones and I. Gijbels** (1998). *Testing monotonicity of regression.* J. Computational and Graphical Statist. 7, 489–500.

**R.J. Casady and J.D. Cryer** (1976). *Monotone percentile regression.* Ann. Stat. 4, 532–541.

**S. Chen, G.B. Dahl and S. Khan** (2005). *Nonparametric identification and estimation of a censored location-scale regression model.* J. Amer. Statist. Assoc. 100, 212–221.

**J.D. Cryer, T. Robertson, F.T. Wright and R.J. Casady** (1972). *Monotone median regression.* Ann. Math. Stat. 43, 1459–1469

**H. Dette, N. Neumeyer and K. F. Pilz** (2006). *A simple nonparametric estimator of a strictly monotone regression function.* Bernoulli 12 , 469–490.

**H. Dette, and S. Volgushev** (2008). *Non–crossing nonparametric estimates of quantile curves.* Journal of the Royal Statistical Society, Ser. B, 70(3):609–627.

**J. Domínguez-Menchero, G. González-Rodríguez and M. J. López-Palomo** (2005). *An L2 point of view in testing monotone regression* J. Nonparam. Statist. 17, 135–153.

**L. Duembgen** (2002). *Application of local rank tests to nonparametric regression.* J. Nonparametr. Stat. 14, 511–537.

**C. Durot** (2003). *A Kolmogorov-type test for monotonicity of regression.* Statist. Probab. Lett. 63, 425–433.

**S. Efromovich** (1999). Nonparametric curve estimation. Methods, theory, and applications. Springer Series in Statistics. New York, NY: Springer.

**J.H.J. Einmahl and I. Van Keilegom** (2008a). *Specification tests in nonparametric regression.* J. Econometr. 143, 88–102.

**J.H.J. Einmahl and I. Van Keilegom** (2008b). *Tests for independence in nonparametric regression.* Statist. Sinica 18, 601–616.

**X. Feng, X. He and J. Hu** (2011). *Wild bootstrap for quantile regression.* Biometrika 98, 995–999.

**B. Fitzenberger, K. Kohn, and A. Lembcke** (2008). *Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany.* IZA Working Paper No. 3356. Available at SSRN: http://ssrn.com/abstract=1135932

**Gasser, T., Müller, H., and Mammitzsch, V.** (1985). *Kernels for nonparametric curve estimation.* Journal of the Royal Statistical Society. Series B, pages 238-252.

**I. Gijbels, P. Hall, M. C. Jones and I. Koch** (2000). *Tests for monotonicity of a regression mean with guaranteed level.* Biometrika 87, 663–673.

**S. Ghosal, A. Sen, A. W. van der Vaart** (2000). *Testing monotonicity of regression.* Ann. Statist. 28, 1054–1082.

**P. Hall and N. E. Heckman** (2000). *Testing for monotonicity of a regression mean by calibrating for linear functions.* Ann. Statist. 28, 20–39.

**W. Härdle and E. Mammen** (1993). *Comparing nonparametric versus parametric regression fits.* Ann. Statist. 21, 1926–1947.

**X. He** (1997). *Quantile Curves without Crossing.* The American Statistician , Vol. 51, No. 2 (May, 1997), pp. 186-192

**W. Hoeffding** (1948). *A nonparametric test of independence.* Ann. Math. Statist. 19, 546–557.

**S. Kiwitt, E.-R. Nagel and N. Neumeyer** (2008). *Empirical Likelihood Estimators for the Error Distribution in Nonparametric Regression Models.* Math. Meth. Statist. 17, 241–260.

**R. Koenker** (2005) *Quantile Regresssion.* John Wiley & Sons, Ltd.

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

**U. U. Müller, A. Schick and W. Wefelmeyer** (2004). *Estimating linear functionals of the error distribution in nonparametric regression.* J. Statist. Plann. Inference 119, 75–93.

**N. Neumeyer** (2007). *A note on uniform consistency of monotone function estimators.* Statist. Probab. Lett. 77, 693–703.

**N. Neumeyer** (2009a). *Smooth residual bootstrap for empirical processes of nonparametric regression residuals.* Scand. J. Statist. 36, 204–228.

**N. Neumeyer** (2009b). *Testing independence in nonparametric regression.* J. Multiv. Anal. 100, 1551–1566.

**J. Rice** (1984). *Bandwidth choice for nonparametric regression.* Ann. Statist., 12(4):1215–1230.

**T. Robertson and F.T. Wright** (1973). *Multiple isotonic median regression.* Ann. Statist. 1, 422–432.

**J. Shim, C. Hwang and H.K. Seok** (2009). *Non-crossing quantile regression via doubly penalized kernel machine.* Comput. Statist. 24, 83–94.

**W. Stute** (1997). *Nonparametric model checks for regression.* Ann. Statist. 25, 613–641.

**W. Stute, W. González Manteiga and M. Presedo Quindimil** (1998). *Bootstrap Approximation in Model Checks for Regression.* J. Amer. Statist. Assoc. 93, 141–149.

**Y. Sun** (2006). *A consistent nonparametric equality test of conditional quantile functions.* Econometric Th. 22, 614–632.

**I. Takeuchi, Q.V. Le, T.D. Sears, and A.J. Smola** (2006). *Nonparametric quantile regression.* Journal of Machine Learning Research, 7:1231–1264.

**A. W. Van der Vaart and J. A. Wellner** (1996). *Weak convergence and empirical processes.* Springer, New York.

**I. Van Keilegom** (1998). *Nonparametric estimation of the conditional distribution in regression with censored data.* PhD thesis, University of Hasselt, Belgium. available at http.//www.ibiostat.be/publications/.

**I. Van Keilegom, W. González–Manteiga and C. Sánchez Sellero** (2008). *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution.* TEST 17, 401–415.

**S. Volgushev, M. Birke, H. Dette and N. Neumeyer** (2013). *Significance testing in quantile regression.* Electron. J. Stat. 7 , 105–145.

**J.C. Wang and M.C. Meyer** (2011). *Testing the monotonicity or convexity of a function using regression splines.* Can. J. Stat. 39, 89–107.

**K. Yu, and M.C. Jones** (1997). *A comparison of local constant and local linear regression quantile estimators.* Computational Statistics and Data Analysis, 25:159–166.

**K. Yu and M.C. Jones** (1998). *Local linear quantile regression.* Journal of the American Statistical Association, 93:228-237.

**K. Yu, Z. Lu, and J. Stander** (2003). *Quantile regression: applications and current research areas.* Journal of the Royal Statistical Society: Series D (The Statistician) 52(3) 331–350.

---

Melanie Birke, Universität Bayreuth, Fakultät für Mathematik, Physik und Informatik , 95440 Bayreuth, Germany, e-mail: Melanie.Birke@uni-bayreuth.de

Natalie Neumeyer, Universität Hamburg, Fachbereich Mathematik, Bundesstraße 55, 20146 Hamburg, Germany, e-mail: neumeyer@math.uni-hamburg.de

Stanislav Volgushev, Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstraße 150, 44780 Bochum, Germany, e-mail: stanislav.volgushev@rub.de

# C Supplement to “The independence process in conditional quantile location-scale models and an application to testing for monotonicity” by Melanie Birke, Natalie Neumeyer and Stanislav Volgushev — Technical results

## C.1 Main results for proofs

Define $\varepsilon_{i,L}$ as the estimated residuals based on linearized versions $\hat{q}_{\tau,L}, \hat{s}_L$ [see Appendix C.2 for their definition], i.e. $\hat{\varepsilon}_{i,L} := (Y_i - \hat{q}_{\tau,L}(X_i))/\hat{s}_L(X_i)$, and $\hat{\varepsilon}^*_{i,L}$ as the corresponding quantities in the bootstrap setting, that is

$$\hat{\varepsilon}^*_{i,L} = \frac{Y_i^* - \hat{q}^*_L(X_i)}{\hat{s}^*_L(X_i)} = \frac{\hat{s}_L(X_i)\varepsilon_i^* + \hat{q}_L(X_i) - \hat{q}^*_L(X_i)}{\hat{s}^*_L(X_i)}$$

The following Lemma demonstrates, that the sequential empirical process based on the residuals $\hat{\varepsilon}_i = (Y_i - \hat{q}_\tau(X_i))/\hat{s}(X_i)$ computed from the initial estimators $\hat{q}_\tau, \hat{s}$ and the sequential empirical process of residuals based on $\varepsilon_{i,L}$ have the same first order expansion.

**Lemma C.1** *Assume that **(K1)**-**(K6)**, **(A1)**-**(A5)**, **(BW)** hold. Then*

$$\sup_{t\in[2h_n,1-2h_n],y\in\mathcal{Y}} \Big|\frac{1}{\sqrt{n}}\sum_i I\{2h_n \le X_i \le t\}(I\{\hat{\varepsilon}_i \le y\} - I\{\hat{\varepsilon}_{i,L} \le y\})\Big| = o_P(1)$$

*uniformly in $t$. If additionally **(B1)**-**(B2)** hold we also have*

$$\sup_{t\in[4h_n,1-4h_n],y\in\mathcal{Y}} \Big|\frac{1}{\sqrt{n}}\sum_i I\{4h_n \le X_i \le t\}(I\{\hat{\varepsilon}^*_i \le y\} - I\{\hat{\varepsilon}^*_{i,L} \le y\})\Big| = o_P(1).$$

**Proof** We only proof the second assertion since the first one follows by similar but easier arguments. Start by observing that by assumption there exists a set $D_n$ whose probability tends to one such that on $D_n$ we have

$(i)$ $\sup_{x\in[4h_n,1-4h_n]}\Big(|\hat{q}(x) - \hat{q}_{\tau,L}(x)| + |\hat{q}^*(x) - \hat{q}^*_{\tau,L}(x)| + |\hat{s}(x) - \hat{s}_L(x)| + |\hat{s}^*(x) - \hat{s}^*_L(x)|\Big) \le \gamma_n$

$(ii)$ $\inf_{x\in[4h_n,1-4h_n]} \min(\hat{s}_L(x), \hat{s}^*_L(x)) \ge c > 0$

$(iii)$ $\sup_{y\in\mathcal{Y}} |y\tilde{f}_\varepsilon(y)| \le C$

for some deterministic sequence $\gamma_n = o(1/\sqrt{n})$ and finite constants $C, c > 0$. A standard Taylor expansion shows that on $D_n$

$$\Big|I\{\hat{\varepsilon}^*_i \le y\} - I\{\hat{\varepsilon}^*_{i,L} \le y\}\Big| \le I\Big\{\Big|U_i - \tilde{F}_\varepsilon\Big(y\frac{\hat{s}^*_L(X_i)}{\hat{s}_L(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\Big| \le C\gamma_n\Big\} =: Z_{n,y,C\gamma_n}(U_i, X_i).$$

In the same manner as the proof of Proposition 3 in Neumeyer (2009a) it follows from assumptions (**B1**) and (**B2**) that the classes of functions

$$\mathcal{G}_{n,2} := \Big\{ v \mapsto \tilde{F}_\varepsilon\Big( y\frac{\hat{s}_L^*(v)}{\hat{s}_L(v)} + \frac{\hat{q}^*_{\tau,L}(v) - \hat{q}_{\tau,L}(v)}{\hat{s}_L(v)}\Big)\Big| y \in \mathcal{Y} \Big\}$$

are with probability tending to one contained in classes of functions $\mathcal{G}_{n,3}$ with bracketing numbers satisfying the assumptions of part one of Lemma C.9. Thus the class of functions

$$\mathcal{F}_n := \Big\{ (u,v) \mapsto I\{s \le u \le t\}\Big( I\{|v - g(u)| \le z\}\Big)\Big| g \in \mathcal{G}_{n,2}, s, t \in [4h_n, 1-4h_n], |z| \le C\gamma_n \Big\}$$

satisfies $\log N_{[\,]}(\varepsilon, \mathcal{F}, L^2(P)) \le C\varepsilon^{-2a}$, see Lemma C.10, and moreover, standard arguments show that $\sup_{g\in\mathcal{F}_n} \int g dP = o(1/\sqrt{n})$ and $\sup_{g\in\mathcal{F}_n}\int g^2 dP = o(1)$. Here, $P$ denotes the probability distribution of $(X_i, U_i)$ and $g^2 = g$ for all $g \in \mathcal{F}_n$. Finally observe that, with probability tending to one,

$$\sup_{t\in[4h_n,1-4h_n], y\in\mathcal{Y}} \frac{1}{\sqrt{n}} \sum_i \Big( I\{h \le X_i \le t\} Z_{n,y,C\gamma_n}(U_i, X_i) - \int_h^t \int Z_{n,y,C\gamma_n}(v,u) f_X(u)dvdu]\Big)$$
$$\le \sqrt{n} \sup_{g\in\mathcal{F}_n} (\int g dP_n - \int g dP),$$

and the right-hand side of the inequality is of order $o_P(1)$ by part one of Lemma C.9 . Moreover, standard arguments yield

$$\int_h^t \int Z_{n,y,C\gamma_n}(v,u) f_X(u) dvdu = o_P(1/\sqrt{n}).$$

Summarizing, we have obtained the estimate

$$\sup_{t\in[4h_n,1-4h_n], y\in\mathcal{Y}} \frac{1}{\sqrt{n}} \sum_i I\{4h_n \le X_i \le t\} Z_{n,y,C\gamma_n}(U_i, X_i) = o_P(1).$$

and thus the proof is complete. □

**Lemma C.2** *Assume that the conditions (**K1**)-(**K6**), (**A1**)-(**A5**), (**BW**) hold. Then*

$$\int_{h_n}^t \frac{\hat{q}_{\tau,L}(x) - q_\tau(x)}{s(x)} f_X(x) f_\varepsilon(0) dx = -\frac{1}{n}\sum_{i=1}^n (I\{\varepsilon_i \le 0\} - \tau) I_{[h_n,t]}(X_i) + o_P(1/\sqrt{n})$$

*uniformly in* $t \in [h_n, 1-h_n]$ *and*

$$\int_{2h_n}^t \frac{\hat{s}_L(x) - s(x)}{\hat{s}(x)} f_X(x)\, dx$$
$$= -\frac{1}{n}\sum_{i=1}^n \frac{I_{[2h_n,t]}(X_i)}{f_{|\varepsilon|}(1)} \Big( I\{|\varepsilon_i| \le 1\} - \frac{1}{2} - \frac{(I\{\varepsilon_i \le 0\} - \tau)(f_\varepsilon(1) - f_\varepsilon(-1))}{f_\varepsilon(0)}\Big) + o_P(\frac{1}{\sqrt{n}})$$

*uniformly in $t \in [2h_n, 1-2h_n]$.*
*If additionally **(B1)**-**(B2)** hold*

$$\int_{3h_n}^t \frac{\hat{q}^*(x) - \hat{q}_{\tau,L}(x)}{\hat{s}_L(x)} f_X(x)dx = -\frac{1}{n}\sum_{i=1}^n \frac{I\{\varepsilon_i^* \le 0\} - \tau}{f_\varepsilon(0)} I_{[3h_n,t]}(X_i) + o_P(1/\sqrt{n})$$

*uniformly in $t \in [3h_n, 1-3h_n]$ and*

$$\int_{4h_n}^t \frac{\hat{s}^*(x) - \hat{s}(x)}{\hat{s}(x)} f_X(x)\,dx$$
$$= -\frac{1}{n}\sum_{i=1}^n \frac{I_{[4h_n,t]}(X_i)}{f_{|\varepsilon|}(1)}\Big(I\{|\varepsilon_i^*| \le 1\} - \frac{1}{2} - \frac{(I\{\varepsilon_i^* \le 0\} - \tau)(f_\varepsilon(1) - f_\varepsilon(-1))}{f_\varepsilon(0)}\Big) + o_P(\frac{1}{\sqrt{n}})$$

*uniformly in $t \in [4h_n, 1-4h_n]$.*

**Proof** We will only prove the representation for $\int_{3h_n}^t \frac{\hat{q}^*(x)-\hat{q}_{\tau,L}(x)}{\hat{s}_L(x)} f_X(x)dx$ since all other results can be derived by analogous arguments.
Observe the decomposition $\hat{q}^*(x) - \hat{q}_{\tau,L}(x) = \hat{q}^*(x) - q(x) + q(x) - \hat{q}_{\tau,L}(x)$. Moreover we have $\hat{q}^*(x) - \hat{q}^*_{\tau,L}(x) = o_P(1/\sqrt{n}), \hat{q}^*_{\tau,L}(x) - q_\tau(x) = O_P(r_n), \hat{s}_L(x) - s(x) = O_P(r_n)$ uniformly in $x \in [3h_n, 1-3h_n]$. It thus suffices to establish

$$\int_{3h_n}^t \frac{\hat{q}^*_{\tau,L}(x) - q_\tau(x)}{s(x)} f_X(x)dx = \int_{3h_n}^t \frac{\hat{q}_{\tau,L}(x) - q_\tau(x)}{s(x)} f_X(x)dx - \frac{1}{n}\sum_{i=1}^n \frac{I\{\varepsilon_i^* \le 0\} - \tau}{f_\varepsilon(0)} I_{[3h_n,t]}(X_i) + o_P(1/\sqrt{n})$$

uniformly in $t \in [3h_n, 1-3h_n]$. By definition we have

$$\frac{f_X(x)(\hat{q}^*_L(x) - q(x))}{s(x)}$$
$$= -\frac{f_X(x)u_1^t\mathcal{M}(K)^{-1}}{f_\varepsilon(0)} \int_{-1}^1 \kappa(v)\Big(\tilde{T}^*_{n,0,L,S}(x, q_{\tau+vb_n}(x)), \ldots, \tilde{T}^*_{n,p,L,S}(x, q_{\tau+vb_n}(x))\Big)^t dv + o_P(1\sqrt{n})$$

where

$$\tilde{T}^*_{n,k,L,S}(x,y) = \frac{1}{nh_n}\frac{1}{f_X(x)}\sum_{i=1}^n K_{h_n,k}(x - X_i)\Big(\Omega\Big(\frac{Y_i^* - y}{d_n}\Big) - F_Y(y|X_i)\Big)$$

Later we will establish the following two assertions

$$(A)\ \int_{3h_n}^t \tilde{T}^*_{n,k,L,S}(x, q_{\tau+vb_n}(x)) f_X(x)dx$$
$$= \frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\int_{-1}^1 K_{1,k}(u)\Big(\Omega\Big(\frac{Y_i^* - q_{\tau+vb_n}(X_i+uh_n)}{d_n}\Big) - F_Y(q_{\tau+vb_n}(X_i+uh)|X_i)\Big)du$$

$$+o_P(1/\sqrt{n})$$

$$(B)\ \ \frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\Big(I\{Y_i^*\leq q_{\tau+vb_n}(X_i+uh_n)+y\}-I\{\varepsilon_i^*\leq y/\hat{s}_L(X_i)\}\Big)$$

$$= \ \ \frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\Big(\bar{F}_\varepsilon\Big(\frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)+y}{\hat{s}_L(X_i)}\Big)-\bar{F}_\varepsilon\Big(\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)+y}{\hat{s}_L(X_i)}\Big)$$

$$+f_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)}\Big)\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)+o_P(1/\sqrt{n}).$$

uniformly in $t\in[3h_n,1-3h_n], u,v\in[-1,1], y\in\mathcal{Y}$ where $\bar{F}_\varepsilon$ is defined in Lemma C.5. Now convolving both sides of $(B)$ [with respect to the argument $y$] with $\frac{1}{d_n}\omega(\cdot/d_n)$ and evaluating the result in 0 yields the identity

$$\frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\Big(\Omega\Big(\frac{Y_i^*-q_{\tau+vb_n}(X_i+uh_n)}{d_n}\Big)-\Omega\Big(\frac{\varepsilon_i^*}{d_n}\Big)\Big)$$

$$= \frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\Big(\bar{F}_\varepsilon\Big(\frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)-\bar{F}_\varepsilon\Big(\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)$$

$$+f_\varepsilon(0)\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)+o_P(1/\sqrt{n}).$$

As we will show at the end of this proof, it holds that

$$(C)\quad \frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\Big(\Omega\Big(\frac{\varepsilon_i^*}{d_n}\Big)-I\{\varepsilon_i^*\leq 0\}\Big)=o_P(1/\sqrt{n}).$$

Moreover, the uniform rates of $\hat{q}_{\tau,L}-q_\tau$, $\hat{s}_L-s$ and the fact that $\hat{q}_{\tau,L}\in C_C^\delta$ with probability tending to one combined with Lemma C.9 yield

$$\begin{aligned}\frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)} &= \frac{1}{n}\sum_i I_{[3h_n,t-h_n]}(X_i)\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{s(X_i)}+o_P(1/\sqrt{n})\\ &= \int_{3h_n}^{t-h_n}\frac{q_\tau(u)-\hat{q}_{\tau,L}(u)}{s(u)}f_X(u)du+o_P(1/\sqrt{n})\\ &= \int_{3h_n}^{t}\frac{q_\tau(u)-\hat{q}_{\tau,L}(u)}{s(u)}f_X(u)du+o_P(1/\sqrt{n}),\end{aligned}$$

where the last equality follows for $t\in[6h_n,1-3h_n]$ (for $t<6h_n$, the indicator in the first line vanishes). Finally, observe that the smoothness properties of $\bar{F}_\varepsilon$ (defined in Lemma C.5) and $F_Y$ yield the representations

$$\bar{F}_\varepsilon\Big(\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)-\bar{F}_\varepsilon\Big(\frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)$$

$$=vb_n\gamma(X_i)+\sum_{j=1}^p\xi_j(X_i,v,n)(uh_n)^j+r_{n,1}$$

$$F_Y(q_{\tau+vb_n}(X_i+uh_n)|X_i)=\tau+vb_n+\sum_{j=1}^p\zeta_j(X_i,v,n)(uh_n)^j+r_{n,2}$$

where the remainder terms $r_{n,j}$ are of order $O(b_n^2 + h_n^{p+1}) = o(1/\sqrt{n})$ uniformly in $u, v$ and $\xi_j, \zeta_j$ denote some functions that do not depend on $u$ and $\gamma$ is a function not depending on $u$. Combining all the arguments so far, we obtain

$$\int_{3h_n}^{t} \tilde{T}^*_{n,k,L,S}(x, q_{\tau+vb_n}(x))dx$$
$$= \mu_k(K)\Big(f_\varepsilon(0)\int_{3h_n}^{t} \frac{q_\tau(u) - \hat{q}_{\tau,L}(u)}{s(u)} f_X(u)du + \frac{1}{n}\sum_{i=1}^{n} I_{[3h_n,t-h_n]}(X_i)(I\{\varepsilon_i^* \le 0\} - \tau)\Big)$$
$$+\frac{h_n^j}{n}\sum_{j=1}^{p}\mu_{k+j}(K)\sum_{i=1}^{n} I_{[3h_n,t-h_n]}(X_i)(\xi_j(X_i) + \zeta_j(X_i)) + o_P(1/\sqrt{n})$$

Noting that by definition

$$e_1^t \mathcal{M}(K)^{-1}(\mu_j(K), \ldots, \mu_{p+j}(K))^t = I\{j = 0\},$$

the proof thus will be complete once we establish (A)-(C) and

$$(D)\quad \frac{1}{n}\sum_{i=1}^{n} I_{[3h_n,t-3h_n]}(X_i)(I\{\varepsilon_i^* \le 0\} - \tau) = \frac{1}{n}\sum_{i=1}^{n} I_{[h_n,t]}(X_i)(I\{\varepsilon_i^* \le 0\} - \tau) + o_P(1/\sqrt{n})$$

uniformly in $t \in [3h_n, 1 - 3h_n]$.

**Proof of (A)** Remembering that $K$ has support $[-1, 1]$, we obtain for any $t \in [3h_n, 1-3h_n]$ the decomposition

$$K_{h_n,k}(x-X_i)I_{[3h_n,t]}(x) = K_{h_n,k}(x-X_i)I_{[3h_n,t]}(x)\Big(I_{[t-h_n,t+h_n]}(X_i)+I_{[2h_n,3h_n]}(X_i)+I_{[h_n,t-h_n]}(X_i)\Big).$$

We will now show that the contributions corresponding to the summands containing $I_{[2h_n,3h_n]}(X_i)$, $I_{[t-h_n,t+h_n]}(X_i)$ are negligible. Since both expressions can be treated analogously, we only provide the arguments for $I_{[t-h_n,t+h_n]}(X_i)$. By similar arguments as in the proof of Lemma C.6 it is easy to show that

$$\sup_{t,x\in[3h_n,1-3h_n],y\in\mathcal{Y}} \Big|\frac{1}{nh_n}\sum_{i=1}^{n} \frac{K_{h_n,k}(x - X_i)}{f_X(x)} I_{[t-h_n,t+h_n]}(X_i)\Big(\Omega\Big(\frac{Y_i^* - y}{d_n}\Big) - F_Y(y|X_i)\Big)\Big|$$
$$=: A_n = O_P(r_n)$$

for any compact $\mathcal{Y} \subset \mathbb{R}$. This yields uniformly in $t \in [3h_n, 1 - 3h_n]$ [remember that $K_{h_n,k}$ vanishes outside of $[-h_n, h_n]$ which yields the equality $K_{h_n,k}(x - X_i)I_{[3h_n,t]}(x) = K_{h_n,k}(x - X_i)I_{[3h_n,t]}(x)I_{[t-2h_n,t+2h_n]}(x)$]

$$\Big|\int_{h_n}^{t} \frac{1}{nh_n}\sum_{i=1}^{n} \frac{K_{h_n,k}(x - X_i)}{f_X(x)} I_{[t-h_n,t+h_n]}(X_i)\Big(\Omega\Big(\frac{Y_i^* - q_{\tau+vb_n}(x)}{d_n}\Big) - F_Y(q_{\tau+vb_n}(x)|X_i)\Big)dx\Big|$$
$$\le \int_{t-2h_n}^{t+2h_n} A_n dx = O_P(h_n r_n) = o_P(1/\sqrt{n})$$

uniformly in $t \in [3h_n, 1-3h_n], v \in [-1,1]$. This completes the proof of (A).

**Proof of (B)** Observe that

$$I\Big\{Y_i^* \le q_{\tau+vb_n}(X_i+uh_n)+y\Big\} = I\Big\{\varepsilon_i^* \le \frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}(X_i)+y}{\hat{s}(X_i)}\Big\}$$

and

$$\begin{aligned}
&\frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\Big(I\Big\{\varepsilon_i^* \le \frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}(X_i)+y}{\hat{s}(X_i)}\Big\} - I\Big\{\varepsilon_i^* \le \frac{y}{\hat{s}_L(X_i)}\Big\}\Big)\\
&= \frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\Big(I\Big\{\varepsilon_i^* \le \frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)+y}{\hat{s}_L(X_i)}\Big\} - I\Big\{\varepsilon_i^* \le \frac{y}{\hat{s}_L(X_i)}\Big\}\Big) + o_P(1/\sqrt{n})\\
&= \frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\Big(\tilde{F}_\varepsilon\Big(\frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)+y}{\hat{s}_L(X_i)}\Big) - \tilde{F}_\varepsilon(y/\hat{s}_L(X_i))\Big) + o_P(1/\sqrt{n})
\end{aligned}$$

uniformly in $t, v, u$, which follows from Lemma C.9, Lemma C.10 and the properties of $\tilde{F}_\varepsilon, \hat{q}_l, \hat{s}_L$ [the bracketing numbers can be bounded by observing that for $y$ from bounded sets the functions involved are all with probability tending to one contained in $C_C^{1+\delta}$]. Next, an application of Lemma C.5 yields

$$\begin{aligned}
&\frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\tilde{F}_\varepsilon\Big(\frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\\
&= \frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\tilde{F}_\varepsilon\Big(\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big) - \frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\bar{F}_\varepsilon\Big(\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\\
&\quad + \frac{1}{n}\sum_i I_{[3h_n,t-3h_n]}(X_i)\bar{F}_\varepsilon\Big(\frac{q_{\tau+vb_n}(X_i+uh_n)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big) + o_P(1/\sqrt{n}),
\end{aligned}$$

where $\bar{F}_\varepsilon$ is defined in Lemma C.5. Noting that

$$\tilde{F}_\varepsilon\Big(\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)+y}{\hat{s}_L(X_i)}\Big) - \tilde{F}_\varepsilon(y) = \tilde{f}_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)}\Big)\frac{q_\tau(X_i)-\hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)} + o_P(1/\sqrt{n}),$$

and remembering that $\tilde{f}_\varepsilon$ converges to $f_\varepsilon$ uniformly with rate $o_P((h_n/\log n)^{1/2})$ completes the proof of (B).

**Proof of (C) and (D)**. Define the sequence of sets $S(\delta_n) := \{(t, y_n, z_n) | x \in [3h_n, 1-3h_n], y_n, z_n \in \mathcal{Y}, |y_n - z_n| \le \delta_n\}$ for some $\delta_n = o(1)$ for some bounded $\mathcal{Y}$ containing zero. Observe that, with probability tending to one,

$$\begin{aligned}
&\sup_{(t,y_n,z_n)\in S(\delta_n)} \Big|\frac{1}{n}\sum_{i=1}^n I_{[3h_n,t-3h_n]}(X_i)\Big(I\{\varepsilon_i^* \le y_n\} - I\{\varepsilon_i^* \le z_n\} + \tilde{F}_\varepsilon(z_n) - \tilde{F}_\varepsilon(y_n)\Big)\Big|\\
&\le \sup_{(t,y_n,z_n)\in S(C\delta_n)} \Big|\frac{1}{n}\sum_{i=1}^n I_{[3h_n,t-3h_n]}(X_i)\Big(I\{U_i \le y_n\} - I\{U_i \le z_n\} + z_n - y_n\Big)\Big|\\
&= o_P(1/\sqrt{n})
\end{aligned}$$

by standard empirical, process arguments provided that $\delta_n = o(1)$. Assertion (D) can be established by a similar argument. For a proof of assertion (C), note further that

$$\tilde{F}_\varepsilon(z_n) - \tilde{F}_\varepsilon(y_n) = \bar{F}_\varepsilon(z_n) - \bar{F}_\varepsilon(y_n)$$

uniformly in $|y_n - z_n| = o(1)$ by Lemma C.5. Moreover, the smoothness of $\bar{F}_\varepsilon$ implies that we have [here, $*$ denotes convolution] $\frac{1}{d_n}(\bar{F}_\varepsilon(\cdot) * \omega(\cdot/d_n))(y) = \bar{F}_\varepsilon(y) + o_P(1/\sqrt{n})$. Combining this facts with the properties of convolution completes the proof of (C). A proof of (D) follows from

$$\begin{aligned} & \sup_{s,t\in[0,1]} \Big| \frac{1}{n} \sum_{i=1}^n \Big( I_{[3h_n,s]}(X_i) - I_{[3h_n,t]}(X_i) \Big) \Big( I\{U_i \leq \tilde{F}_\varepsilon(0)\} - F_\varepsilon(0) \Big) \Big| \\ \leq & \sup_{s,t,y\in[0,1]} \Big| \frac{1}{n} \sum_{i=1}^n \Big( I_{[3h_n,s]}(X_i) - I_{[3h_n,t]}(X_i) \Big) \Big( I\{U_i \leq y\} - y \Big) \Big| \end{aligned}$$

and standard empirical process arguments. □

## C.2 Properties of $\hat{q}_\tau$ and $\hat{s}$

We start this section by introducing some notation and giving an overview of the derived results. Define

$$T_{n,k,S}(x, y) := \frac{1}{nh_n} \sum_{i=1}^n K_{h,k}(x - X_i) \Omega\Big(\frac{y - Y_i}{d_n}\Big)$$

where $\Omega$ denotes a distribution function on $\mathbb{R}$. Also, set

$$T_{n,k,U}(x, y) := \frac{1}{nh_n} \sum_{i=1}^n K_{h,k}(x - X_i) I\{Y_i \leq y\}$$

and define $\hat{F}_{Y,U}$ with $T_{n,k,S}$ in the definition of $\mathbf{Y}$ replaced by $T_{n,k,U}$. Note that by definition, $T_{n,k,S}(x,\cdot)$ is a smoothed version of $T_{n,k,U}(x,\cdot)$ that is obtained by convolution with the function $\frac{1}{d_n}\omega(\cdot/d_n)$ where $\omega := \Omega'$. In particular, this implies

$$\begin{aligned} T_{n,k,S}(x, y) &= \int T_{n,k,U}(x, y - d_n u) \omega(u) du, \\ \partial_x^m \partial_y^l T_{n,k,S}(x, y) &= \frac{1}{d_n^l} \int \omega^{(l)}(u/d_n) \partial_x^m T_{n,k,U}(x, y - d_n u) du. \end{aligned}$$

Define the quantities $\hat{F}_{|e|}, \hat{F}_Y^*, \hat{F}_{|e|}^*$ analogously with $Y_i$ replaced by $|\hat{e}_i|, Y_i^*, |e_i^*|$, respectively where $\hat{e}_i := Y_i - \hat{q}(X_i)$, $\hat{e}_i^* := Y_i^* - \hat{q}^*(X_i)$. The structure of the estimator makes it rather complicated to directly analyze its derivatives with respect to $x$. However, it is possible to derive an asymptotic representation of the form

$$\begin{aligned} \hat{F}_Y(y|x) &= \hat{F}_{Y,L,S}(y|x) + o_P(1/\sqrt{n}), \qquad \hat{F}_{|e|}(y|x) = F_{|e|,L,S}(y|x) + o_P(1/\sqrt{n}), \\ \hat{F}_Y^*(y|x) &= \hat{F}_{Y,L,S}^*(y|x) + o_P(1/\sqrt{n}), \qquad \hat{F}_{|e|}^*(y|x) = F_{|e|,L,S}^*(y|x) + o_P(1/\sqrt{n}) \end{aligned}$$

holding uniformly over $x, y$ where $\hat{F}_{Y,L,S}(y|x)$ is defined as

$$\begin{aligned}
\hat{F}_{Y,L,S}(y|x) &:= F_Y(y|x) + u_1^t \mathcal{M}(K)^{-1}\Big(T_{n,0,L,S}(x,y), \dots, T_{n,p,L,S}(x,y)\Big)^t \\
\hat{F}_{|e|,L,S}(y|x) &:= F_{|e|}(y|x) + u_1^t \mathcal{M}(K)^{-1}\Big(T_{e,n,0,L,S}(x,y), \dots, T_{e,n,p,L,S}(x,y)\Big)^t \\
\hat{F}^*_{Y,L,S}(y|x) &:= F_Y(y|x) + u_1^t \mathcal{M}(K)^{-1}\Big(T^*_{n,0,L,S}(x,y), \dots, T^*_{n,p,L,S}(x,y)\Big)^t \\
\hat{F}^*_{|e|,L,S}(y|x) &:= F_{|e|}(y|x) + u_1^t \mathcal{M}(K)^{-1}\Big(T^*_{e,n,0,L,S}(x,y), \dots, T^*_{e,n,p,L,S}(x,y)\Big)^t
\end{aligned}$$

where $u_1^t := (1, 0, ..., 0)$ denotes the first unit vector in $\mathbb{R}^{p+1}$, $\mathcal{M}(K)$ denotes a $(p+1)\times(p+1)$ matrix with entries

$$\mathcal{M}(K)_{ij} = \mu_{i+j-2}(K) := \int u^{i+j-2}K(u)du,$$

and

$$\begin{aligned}
T_{n,k,L,S}(x,y) &:= \frac{1}{nh}\sum_{i=1}^n \frac{1}{f_X(X_i)} K_{h,k}(x - X_i)\Big(\Omega\Big(\frac{y - Y_i}{d_n}\Big) - F_Y(y|X_i)\Big) \\
T_{e,n,k,L,S}(x,y) &:= \frac{1}{nh}\sum_{i=1}^n \frac{1}{f_X(X_i)} K_{h,k}(x - X_i)\Big(\Omega\Big(\frac{y - |Y_i - \hat{q}_{\tau,L}(X_i)|}{d_n}\Big) - F_e(y|X_i)\Big) \\
T^*_{n,k,L,S}(x,y) &:= \frac{1}{nh}\sum_{i=1}^n \frac{1}{f_X(X_i)} K_{h,k}(x - X_i)\Big(\Omega\Big(\frac{y - Y^*_i}{d_n}\Big) - F_Y(y|X_i)\Big) \\
T^*_{e,n,k,L,S}(x,y) &:= \frac{1}{nh}\sum_{i=1}^n \frac{1}{f_X(X_i)} K_{h,k}(x - X_i)\Big(\Omega\Big(\frac{y - |Y^*_i - \hat{q}^*_{\tau,L}(X_i)|}{d_n}\Big) - F_e(y|X_i)\Big).
\end{aligned}$$

This, and further properties as differentiability and convergence rates of $\hat{F}_{Y,L,S}(y|x)$, $\hat{F}_{|e|}$, $\hat{F}^*_Y$, $\hat{F}^*_{|e|}$ is the subject of Lemma C.6.
Next, consider the functionals

$$H_{G,\kappa,\tau,b_n}(F) := \frac{1}{b_n}\int_0^1 \int_{-\infty}^{\tau} \kappa\Big(\frac{F(G^{-1}(u)) - v}{b_n}\Big) dv du$$

and $Q_{G,\kappa,\tau,b_n}(F) := G^{-1}(H_{G,\kappa,\tau,b_n}(F))$. Some properties of this functional are collected in Lemma C.8. With this definition, the quantiles estimators $\hat{q}, \hat{s}, \hat{q}^*, \hat{s}^*$ can be represented as

$$\begin{aligned}
\hat{q}(x) &= H_{G,\kappa,\tau,b_n}(\hat{F}_Y(\cdot|x)), \qquad \hat{q}^*(x) = H_{G,\kappa,\tau,b_n}(\hat{F}^*_Y(\cdot|x)), \\
\hat{s}(x) &= H_{G_s,\kappa,\tau,b_n}(\hat{F}_{|e|}(\cdot|x)), \qquad \hat{s}^*(x) = H_{G_s,\kappa,\tau,b_n}(\hat{F}^*_{|e|}(\cdot|x)).
\end{aligned}$$

The results in Lemma C.8 and properties of the estimators $\hat{F}_Y, \hat{F}_{|e|}, \hat{F}^*_Y, \hat{F}^*_{|e|}$ yield representations of the form

$$\begin{aligned}
\hat{q}_\tau(x) &= \hat{q}_{\tau,L}(x) + o_P(n^{-1/2}), \qquad \hat{s}(x) = \hat{s}_L(x) + o_P(n^{-1/2}), \\
\hat{q}^*_\tau(x) &= \hat{q}^*_{\tau,L}(x) + o_P(n^{-1/2}), \qquad \hat{s}^*(x) = \hat{s}^*_L(x) + o_P(n^{-1/2})
\end{aligned}$$

uniformly in $x$ [see Lemma C.4] where

$$\begin{aligned}
\hat{q}_{\alpha,L}(x) &:= q_\alpha(x) - \frac{1}{f_e(0|x)} \int_{-1}^{1} \Big(\hat{F}_{Y,L,S}(q_{\alpha+vb_n}(x)|x) - F_Y(q_{\alpha+vb_n}(x)|x)\Big)\kappa(v)dv \\
&= q_\alpha(x) - \frac{u_1^t \mathcal{M}(K)^{-1}}{f_e(0|x)} \int_{-1}^{1} \kappa(v)\Big(T_{n,0,L,S}(x, q_{\alpha+vb_n}(x)), \ldots, T_{n,p,L,S}(x, q_{\alpha+vb_n}(x))\Big)^t dv \\
\hat{s}_L(x) &:= s(x) - \frac{1}{f_{|\varepsilon|}(1|x)} \int_{-1}^{1} \Big(\hat{F}_{|e|,L,S}(s_{1/2+vb_n}(x)|x) - F_{|e|}(s_{1/2+vb_n}(x)|x)\Big)\kappa(v)dv \\
&= s(x) - \frac{u_1^t \mathcal{M}(K)^{-1}}{f_{|\varepsilon|}(1)} \int_{-1}^{1} \kappa(v)\Big(T_{e,n,0,L,S}(x, s_{1/2+vb_n}(x)), \ldots, T_{e,n,p,L,S}(x, s_{1/2+vb_n}(x))\Big)^t dv \\
\hat{q}^*_{\alpha,L}(x) &:= q_\alpha(x) - \frac{1}{f_e(0|x)} \int_{-1}^{1} \Big(\hat{F}^*_{Y,L,S}(q_{t+vb_n}(x)|x) - F_Y(q_{t+vb_n}(x)|x)\Big)\kappa(v)dv \\
&= q_\alpha(x) - \frac{u_1^t \mathcal{M}(K)^{-1}}{f_e(0|x)} \int_{-1}^{1} \kappa(v)\Big(T^*_{n,0,L,S}(x, q_{\alpha+vb_n}(x)), \ldots, T^*_{n,p,L,S}(x, q_{\alpha+vb_n}(x))\Big)^t dv \\
\hat{s}^*_L(x) &:= s(x) - \frac{1}{f_{|\varepsilon|}(1)} \int_{-1}^{1} \Big(\hat{F}^*_{|e|,L,S}(s_{1/2+vb_n}(x)|x) - F_{|e|}(s_{1/2+vb_n}(x)|x)\Big)\kappa(v)dv \\
&= s(x) - \frac{u_1^t \mathcal{M}(K)^{-1}}{f_{|\varepsilon|}(1)} \int_{-1}^{1} \kappa(v)\Big(T^*_{e,n,0,L,S}(x, s_{1/2+vb_n}(x)), \ldots, T^*_{e,n,p,L,S}(x, s_{1/2+vb_n}(x))\Big)^t dv
\end{aligned}$$

Differentiability properties and convergence rates of derivatives of these estimators can obviously be derived from the corresponding properties of the underlying distribution function estimators, see Lemma C.3.

**Lemma C.3** *Let **(K1)**-**(K6)**, **(A1)**-**(A5)**, **(BW)** hold. Then for any $k \leq p$*

$$\begin{aligned}
\sup_{x\in[h_n,1-h_n]} |\hat{q}^{(k)}_{\tau,L}(x) - q^{(k)}_\tau(x)| &= O_P\Big(\frac{\log h_n^{-1}}{nh_n(h_n \wedge d_n)^{2k}}\Big)^{1/2} = o_P(1), \\
\sup_{x\in[2h_n,1-2h_n]} |\hat{s}^{(k)}_L(x) - s^{(k)}(x)| &= O_P\Big(\frac{\log h_n^{-1}}{nh_n(h_n \wedge d_n)^{2k}}\Big)^{1/2} = o_P(1),
\end{aligned}$$

*and under **(B1)**-**(B2)** it follows that*

$$\begin{aligned}
\sup_{x\in[3h_n,1-3h_n]} |(\hat{q}^*_{\tau,L})^{(k)}(x) - q^{(k)}_\tau(x)| &= O_P\Big(\frac{\log h_n^{-1}}{nh_n(h_n \wedge d_n)^{2k}}\Big)^{1/2} = o_P(1), \\
\sup_{x\in[4h_n,1-4h_n]} |(\hat{s}^*_L)^{(k)}(x) - s^{(k)}(x)| &= O_P\Big(\frac{\log h_n^{-1}}{nh_n(h_n \wedge d_n)^{2k}}\Big)^{1/2} = o_P(1).
\end{aligned}$$

**Proof of Lemma C.3** Directly follows from the definitions of the linearized versions. □

**Lemma C.4** *Let **(K1)**-**(K6)**, **(A1)**-**(A5)**, **(BW)** hold. Then*

$$(i) \quad \sup_{x\in[h_n,1-h_n]} |\hat{q}(x) - \hat{q}_{\tau,L}(x)| = o_P(1/\sqrt{n}),$$

$$(ii) \quad \sup_{x\in[2h_n,1-2h_n]} |\hat{s}(x) - \hat{s}_L(x)| = o_P(1/\sqrt{n}),$$

*and if additionally (**B1**)-(**B2**) hold, we also have*

$$(iii) \quad \sup_{x\in[3h_n,1-3h_n]} |\hat{q}^*(x) - \hat{q}^*_{\tau,L}(x)| = o_P(1/\sqrt{n}),$$
$$(iv) \quad \sup_{x\in[4h_n,1-4h_n]} |\hat{s}^*(x) - \hat{s}^*_L(x)| = o_P(1/\sqrt{n}).$$

**Proof** Apply Lemma C.8 with $F_1 = \hat{F}_Y, \hat{F}_{|e|}, \hat{F}^*_Y, \hat{F}^*_{|e|}$, $F_2 = \hat{F}_{Y,L}, \hat{F}_{|e|,L}, \hat{F}^*_{Y,L}, \hat{F}^*_{|e|,L}$ and $F = F_Y, F_{|e|}, F_Y, F_{|e|}$, respectively. The corresponding assumptions are satisfied by Lemma C.8. □

**Lemma C.5** *Let $n\alpha_n^4 = o(1)$ and assume that the conditions of (i), (i)', (ii), (ii)' of Lemma C.6 hold. Then for any bounded $\mathcal{Y} \subset \mathbb{R}$ we have*

$$\sup_{a,b\in\mathcal{Y},|a-b|\le d_n} \left|\tilde{F}_\varepsilon(a) - \tilde{F}_\varepsilon(b) - \Big(\bar{F}_\varepsilon(a) - \bar{F}_\varepsilon(b)\Big)\right| = o_P(1/\sqrt{n})$$

*where*

$$\bar{F}_n(a) := \frac{\sum_k I_{[2h_n,1-2h_n]}(X_k) F_Y(\hat{q}_{\tau,L}(X_k) + a\hat{s}_L(X_k)|X_k)}{\sum_l I_{[2h_n,1-2h_n]}(X_l)}.$$

**Proof** Recalling the definition of $\tilde{F}_\varepsilon$, it is easy to see that $\tilde{F}_\varepsilon(y) = \frac{1}{\alpha_n}\Big(\hat{F}_\varepsilon(\cdot) * \phi(\cdot/\alpha_n)\Big)(y)$ where

$$\hat{F}_\varepsilon(y) := \frac{\sum_k I_{[2h_n,1-2h_n]}(X_k) I\{Y_k - \hat{q}(X_k) \le y\hat{s}(X_k)\}}{\sum_l I_{[2h_n,1-2h_n]}(X_l)}.$$

Standard calculations show that

$$\frac{1}{\alpha_n}\Big(\bar{F}_\varepsilon(\cdot) * \phi(\cdot/\alpha_n)\Big)(y) = \bar{F}_\varepsilon(y) + o_P(1/\sqrt{n})$$

uniformly in $y \in \mathcal{Y}$. Thus it suffices to establish that

$$\sup_{|a-b|\le d_n} \left|\hat{F}_\varepsilon(a) - \hat{F}_\varepsilon(b) - \Big(\bar{F}_\varepsilon(a) - \bar{F}_\varepsilon(b)\Big)\right| = o_P(1/\sqrt{n}).$$

Since $\frac{1}{n}\sum_l I_{[2h_n,1-2h_n]}(X_l) = 1 + o_P(1)$, we only need to consider the enumerator. Since $\mathcal{Y}$ is uniformly bounded we have, with probability tending to one, uniformly in $y \in \mathcal{Y}$

$$\begin{aligned} &\Big|I\{Y_k - \hat{q}(X_k) \le y\hat{s}(X_k)\} - I\{Y_k - \hat{q}_{\tau,L}(X_k) \le y\hat{s}_L(X_k)\}\Big| \\ \le\ & I\{Y_k - \hat{q}_{\tau,L}(X_k) - y\hat{s}_L(X_k) \le \gamma_n\} - I\{Y_k - \hat{q}_{\tau,L}(X_k) - y\hat{s}_L(X_k) \le -\gamma_n\} \end{aligned}$$

for some $\gamma_n = o(1/\sqrt{n})$. Moreover an application of Lemma C.10 shows that the functions

$$(u, v) \mapsto I\{v - \hat{q}_{\tau,L}(u) - y\hat{s}_L(u) \le \gamma_n\} - I\{v - \hat{q}_{\tau,L}(u) - y\hat{s}_L(u) \le -\gamma_n\}$$

are, with probability tending to one, contained in a class of functions satisfying the assumptions of the first part of Lemma C.9 with the additional property that each element has expectation of order $o(1/\sqrt{n})$. In particular, this implies

$$\sup_{y\in\mathcal{Y}}\Big|\sum_k I_{[2h_n,1-2h_n]}(X_k)\Big(I\{Y_k-\hat{q}(X_k)\le y\hat{s}(X_k)\}-I\{Y_k-\hat{q}_{\tau,L}(X_k)\le y\hat{s}_L(X_k)\}\Big)\Big| = o_P(1/\sqrt{n}),$$

and thus it remains to consider

$$\sup_{a,b\in\mathcal{Y},|a-b|\le d_n}\frac{1}{n}\sum_k I_{[2h_n,1-2h_n]}(X_i)\Big(I\{Y_k\le\hat{q}_{\tau,L}(X_k)+a\hat{s}_L(X_k)\}-I\{Y_k\le\hat{q}_{\tau,L}(X_k)+b\hat{s}_L(X_k)\}$$
$$-F_Y(\hat{q}_{\tau,L}(X_k)+a\hat{s}_L(X_k)|X_k)+F_Y(\hat{q}_{\tau,L}(X_k)+b\hat{s}_L(X_k)|X_k)\Big)$$

By arguments similar to those given above, it is easily seen that this quantity is of order $o_P(1/\sqrt{n})$ if one notes that the smoothness assumptions on $F_Y$ imply that with $\hat{q}_{\tau,L},\hat{s}_L\in C_C^{1+\delta}$ with probability tending to one the same holds for the function $u\mapsto F_Y(\hat{q}_{\tau,L}(u)+y\hat{s}_L(u)|u)$ uniformly in $y\in\mathcal{Y}$. This completes the proof. □

**Lemma C.6** *Assume that conditions **(K1)**-**(K6)**, **(A1)**-**(A5)** and **(BW)** hold. Then for any bounded $\mathcal{Y}\subset\mathbb{R}$ we have*

$$\begin{aligned}(i)'\quad \hat{F}_Y(y|x) &= F_Y(y|x)+u_1^t\mathcal{M}(K)^{-1}\Big(T_{n,0,L,S}(x,y),\dots,T_{n,p,L,S}(x,y)\Big)^t+o_P(1/\sqrt{n})\\ &=: \hat{F}_{Y,L,S}(y|x)+o_P(1/\sqrt{n})\end{aligned}$$

*uniformly in $y\in\mathcal{Y},x\in[h_n,1-h_n]$ and*

$$\begin{aligned}(ii)'\quad \hat{F}_e(y|x) &= F_e(y|x)+u_1^t\mathcal{M}(K)^{-1}\Big(T_{e,n,0,L,S}(x,y),\dots,T_{e,n,p,L,S}(x,y)\Big)^t+o_P(1/\sqrt{n})\\ &=: \hat{F}_{e,L,S}(y|x)+o_P(1/\sqrt{n}).\end{aligned}$$

*uniformly in $y\in\mathcal{Y},x\in[2h_n,1-2h_n]$. If additionally **(B1)**-**(B2)** hold,*

$$\begin{aligned}(iii)'\quad \hat{F}_Y^*(y|x) &= F_Y(y|x)+u_1^t\mathcal{M}(K)^{-1}\Big(T^*_{n,0,L,S}(x,y),\dots,T^*_{n,p,L,S}(x,y)\Big)^t+o_P(1/\sqrt{n})\\ &=: \hat{F}^*_{Y,L,S}(y|x)+o_P(1/\sqrt{n}).\end{aligned}$$

*uniformly in $y\in\mathcal{Y},x\in[3h_n,1-3h_n]$ and*

$$\begin{aligned}(iv)'\quad \hat{F}_e^*(y|x) &= F_e(y|x)+u_1^t\mathcal{M}(K)^{-1}\Big(T^*_{e,n,0,L,S}(x,y),\dots,T^*_{e,n,p,L,S}(x,y)\Big)^t+o_P(1/\sqrt{n})\\ &=: \hat{F}^*_{e,L,S}(y|x)+o_P(1/\sqrt{n})\end{aligned}$$

*uniformly in $y\in\mathcal{Y},x\in[4h_n,1-4h_n]$.*

*Assertions (i)'-(iv)' continue to hold under their respective assumptions for* $\tilde{T}_{n,k}, \tilde{T}_{e,n,k}, \tilde{T}^*_{n,k}, \tilde{T}^*_{e,n,k}$ *where the* $f_X(X_i)$ *in the definition of* $T_{n,k}, T_{e,n,k}, T^*_{n,k}, T^*_{e,n,k}$ *is replaced by* $f_X(x)$. *Moreover, (i)-(iv) hold under the assumptions of* $(i)' - (iv)'$, *respectively.*

$$(i) \quad \sup_{y\in\mathcal{Y},x\in[h_n,1-h_n]} |\partial_x^k \partial_y^l \hat{F}_{Y,L,S}(y|x) - \partial_x^k \partial_y^l F_Y(y|x)| = O_P\Big(\frac{\log n}{nh_n^{2k+1}d_n^{2l}}\Big)^{1/2}$$

$$(ii) \quad \sup_{y\in\mathcal{Y},x\in[2h_n,1-2h_n]} |\partial_x^k \partial_y^l \hat{F}_{e,L,S}(y|x) - \partial_x^k \partial_y^l F_e(y|x)| = O_P\Big(\frac{\log n}{nh_n^{2k+1}d_n^{2l}}\Big)^{1/2}$$

$$(iii) \quad \sup_{y\in\mathcal{Y},x\in[3h_n,1-3h_n]} |\partial_x^k \partial_y^l \hat{F}^*_{Y,L,S}(y|x) - \partial_x^k \partial_y^l F_Y(y|x)| = O_P\Big(\frac{\log n}{nh_n^{2k+1}d_n^{2l}}\Big)^{1/2}$$

$$(iv) \quad \sup_{y\in\mathcal{Y},x\in[4h_n,1-4h_n]} |\partial_x^k \partial_y^l \hat{F}^*_{e,L,S}(y|x) - \partial_x^k \partial_y^l F_e(y|x)| = O_P\Big(\frac{\log n}{nh_n^{2k+1}d_n^{2l}}\Big)^{1/2}$$

**Proof of Lemma C.6**

We will only provide the arguments for (iv) and (iv)' since all other assertions can be derived analogously. Define the quantity

$$T^*_{e,n,k,L}(x,y) := \frac{1}{nh_n}\sum_{i=1}^n \frac{1}{f_X(X_i)} K_{h_n,k}(x-X_i)\Big(I\{Y_i^* \le y + \hat{q}^*_{\tau,L}(X_i)\} - F_e(y|X_i)\Big)$$

and note that $T^*_{e,n,k,L,S}(x,y)$ is, up to an error of order $d_n^{p_\omega}$, the convolution of $T^*_{e,n,k,L}(x,\cdot)$ with $\frac{1}{d_n}\omega(\cdot/d_n)$. As we will now show, (iv)' follows from the following assertion

$$(iva)' \quad \hat{F}^*_{e,U}(y|x) = F_e(y|x) + u_1^t \mathcal{M}(K)^{-1}\Big(T^*_{e,n,0,L}(x,y),\dots,T^*_{e,n,p,L}(x,y)\Big)^t + o_P(1/\sqrt{n}).$$

Assertion $(iva)'$ implies assertion $(iv)'$ since

$$(F_e(\cdot|x) * \frac{1}{d_n}\omega(\cdot/d_n))(y) = F_e(y|x) + O(d_n^{p_\omega}) = F_e(y|x) + o(1/\sqrt{n})$$

uniformly in $x \in [4h_n, 1-4h_n], y \in \mathcal{Y}$. Similarly, in order to establish $(iv)$ it suffices to show that

$$(iva) \quad \sup_{y\in\mathcal{Y},x\in[4h_n,1-4h_n]} |\partial_x^k \hat{F}^*_{e,L,U}(y|x) - \partial_x^k F_e(y|x)| = O_P\Big(\frac{\log n}{nh_n^{2k+1}}\Big)^{1/2}.$$

This is due to the fact that

$$\begin{aligned}\partial_x^k \partial_y^l\Big(\hat{F}^*_{e,L,S}(y|x) - F_e(y|x)\Big) &= \frac{1}{d_n^{l+1}}\Big((\partial_x^k \hat{F}^*_{e,L,U}(\cdot|x) - \partial_x^k F_e(\cdot|x)) * \omega^{(l)}(\frac{\cdot}{d_n})\Big)(y) \\ &\quad + \frac{1}{d_n}((\partial_x^k \partial_y^l F_e(\cdot|x)) * \omega(\frac{\cdot}{d_n}))(y) - \partial_x^k \partial_y^l F_e(y|x).\end{aligned}$$

Now, since by assumption $\partial_x^k F_e(y|x)$ is $r$ times continuously differentiable with respect to $y$, the second summand is of order $d_n^{r-l} = O\Big(\frac{\log n}{nh_n^{2k+1}d_n^{2l}}\Big)^{1/2}$. The first summand can be estimated by $\frac{1}{d_n^l} O_P\Big(\frac{\log n}{n h_n^{2k+1}}\Big)^{1/2}$ [note that one $d_n$ can be absorbed into the integral with respect to $\kappa$]. We will now proceed by establishing $(iva)'$ and $(iva)$. Observe the identity

$$I\{Y_i^* \le y + \hat{q}^*_{\tau,L}(X_i)\} = I\Big\{U_i \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}(X_i)}{\hat{s}(X_i)}\Big)\Big\}.$$

Moreover, a Taylor expansion shows that, with probability tending to one,

$$\Big|I\Big\{U_i \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}(X_i)}{\hat{s}(X_i)}\Big)\Big\} - I\Big\{U_i \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\Big\}\Big|$$
$$\le I\Big\{\Big|U_i - \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\Big| \le C\gamma_n \sup_{y \in 2\mathcal{Y}/c_s} |y\tilde{f}_\varepsilon(y)|\Big\}$$

where $\gamma_n = o(1/\sqrt{n})$, and thus arguments similar to those in the proof of Lemma C.1 yield

$$\frac{1}{n}\sum_{i=1}^n \frac{K_{h_n,k}(x-u)}{h_n}\Big(\frac{1}{f(u)} - \frac{1}{f(x)}\Big)\Big(I\Big\{U_i \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}(X_i)}{\hat{s}(X_i)}\Big)\Big\}\Big)$$
$$= \frac{1}{n}\sum_{i=1}^n \frac{K_{h_n,k}(x-u)}{h_n}\Big(\frac{1}{f(u)} - \frac{1}{f(x)}\Big)\Big(I\Big\{U_i \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\Big\}\Big) + o_P(1/\sqrt{n}).$$

Next, observe that we have uniformly over $y \in \mathcal{Y}$

$$\Big|\tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big) - F_\varepsilon\Big(\frac{y}{s(X_i)}\Big)\Big| \le \|\tilde{F}_\varepsilon - F_\varepsilon\|_\infty + c_n \sup_{y \in 2\mathcal{Y}/c_s} |y f_\varepsilon(y)| \le C c_n$$

with probability tending to one for a deterministic sequence $c_n = O(r_n)$. Denote by $D_n$ the set where this inequality holds. Observe that on the set $D_n$ the classes of functions

$$\mathcal{F}_n := \Big\{(u,v) \mapsto \frac{K_{h_n,k}(x-u)}{h_n}\Big(\frac{1}{f(u)} - \frac{1}{f(x)}\Big) \times$$
$$\times\Big(I\Big\{v \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}_L(u)} + \frac{\hat{q}^*_{\tau,L}(u) - \hat{q}_{\tau,L}(u)}{\hat{s}_L(u)}\Big)\Big\} - F_e(y|u)\Big)\Big| x \in [4h_n, 1-4h_n], y \in \mathcal{Y}\Big\}$$

are, with probability tending to one contained in a class $\mathcal{G}_n$ of functions with bracketing numbers $\mathcal{N}_{[]}(\varepsilon, \mathcal{G}_n, L^2(P)) \le C\exp(C\varepsilon^{-a})$ for some $a < 2$ [see Lemma C.10] and moreover, every element $g$ of $\mathcal{F}_n$ satisfies $\mathbb{E}g(X_i,Y_i) = O(h_n c_n) = o(1/\sqrt{n}), \mathbb{E}g^2(X_i,Y_i) = O(h_n)$. Thus by Lemma C.9 $\sup_{g \in \mathcal{F}_n} |\sum_i g(X_i,Y_i)| = o_P(1/\sqrt{n})$, i.e. we have shown that the $1/f(X_i)$ in the definition of $T^*_{e,n,k}$ can be replaced with $1/f(x)$ with an error of order $o_p(1/\sqrt{n})$. For the rest of the proof, note that on $D_n$ we have the inequality

$$I\Big\{U_i \le F_\varepsilon\Big(\frac{y}{s(X_i)}\Big) - c_n\Big\} \le I\Big\{U_i \le \tilde{F}_\varepsilon\Big(\frac{y}{\hat{s}_L(X_i)} + \frac{\hat{q}^*_{\tau,L}(X_i) - \hat{q}_{\tau,L}(X_i)}{\hat{s}_L(X_i)}\Big)\Big\} \le I\Big\{U_i \le F_\varepsilon\Big(\frac{y}{s(X_i)}\Big) + c_n\Big\}$$

Consider the decomposition

$$T_{n,k,m}(x,y) := \partial_x^m T_{n,k}(x,y) = T^+_{n,k,m}(x,y) + T^-_{n,k,m}(x,y)$$

where

$$T^+_{n,k,m}(x,y) := \frac{1}{nh_n}\frac{1}{h_n^m}\sum_{i=1}^n \frac{K^{(m)}_{h_n,k}(x-X_i)}{f(X_i)} I\Big\{K^{(m)}_{h_n,k}(x-X_i)>0\Big\}\Big(I\{Y_i^*\leq y+\hat q^*_{\tau,L}(X_i)\}-F_e(y|X_i)\Big)$$

and $T^-_{n,k,m}$ is defined analogously. On the set $D_n$ we have

$$\begin{aligned}T^+_{n,k,m}(x,y) &\leq \frac{1}{nh_n^{m+1}}\sum_{i=1}^n \frac{K^{(m)}_{h_n,k}(x-X_i)}{f(X_i)} I\Big\{K^{(m)}_{h_n,k}(x-X_i)>0\Big\}\times \\ &\qquad \times\Big(I\Big\{U_i\leq F_\varepsilon\Big(\frac{y}{s(X_i)}\Big)+c_n\Big\}-F_\varepsilon\Big(\frac{y}{s(X_i)}\Big)\Big) \\ &=: \frac{1}{nh_n^{m+1}}\sum_{i=1}^n g^{(n,m,+)}_{x,y}(X_i,U_i,c_n).\end{aligned}$$

The expectation of each summand $g^{(n,m,+)}_{x,y}(X_i,U_i,c_n)$ in the above sum is of the order $O(hc_n)$. Moreover, the class of functions

$$\Big\{(u,v)\mapsto g^{(n,m,+)}_{x,y}(u,v,c_n)\Big| x\in[4h_n,1-4h_n], y\in\mathcal{Y}\Big\}$$

is with probability tending to one contained in a class that satisfies the assumptions of part 2 of Lemma C.9 with $\delta_n=h_n$. This yields the estimate

$$\frac{1}{nh^{m+1}}\sum_{i=1}^n g^{(n,m,+)}_{x,y}(X_i,U_i,c_n) = \frac{hc_n}{h^{m+1}} + O_P\Big(\frac{\log n}{nh^{2m+1}}\Big)^{1/2} = O_P\Big(\frac{\log n}{nh^{2m+1}}\Big)^{1/2}$$

uniformly in $x\in[4h_n,1-4h_n], y\in\mathcal{Y}$. Summarizing, we have obtained the bound $T^+_{n,k,m}(x,y)\leq O_P\Big(\frac{\log n}{nh_n^{2m+1}}\Big)^{1/2}$, and a corresponding lower bound can be obtained by similar arguments. Analogous reasoning yields a bound for $T^-_{n,k,m}(x,y)$ and altogether this implies $T_{n,k,m}(x,y) = O_P\Big(\frac{\log n}{nh_n^{2m+1}}\Big)^{1/2}$ uniformly in $x\in[4h_n,1-4h_n], y\in\mathcal{Y}$.
For a proof of (iva)', note that a Taylor expansion of $F_e(y|X_i)$ with respect to $X_i$ around the point $x$ combined with the fact that

$$\frac{1}{nh_n}e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\begin{pmatrix} h_n^k\sum_i K_{h_n,k}(x-X_i) \\ \vdots \\ h_n^{p+k}\sum_i K_{h_n,p+k}(x-X_i)\end{pmatrix} = I\{k=0\}$$

for $k=0,...,p$ yields the representation

$$\frac{e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}}{nh_n}\begin{pmatrix} \sum_i K_{h_n,0}(x-X_i)F_e(y|X_i) \\ \vdots \\ \sum_i h_n^p K_{h_n,p}(x-X_i)F_e(y|X_i)\end{pmatrix} = F_e(y|x)+O_P(h_n^{p+1}) = F_e(y|x)+o_P(n^{-1/2})$$

uniformly in $x \in [4h_n, 1-4h_n], y \in \mathcal{Y}$. Combining the arguments so far

$$\hat{F}_e^*(y|x) - F_e^*(y|x) = e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\Big(T_{n,0,0}(x,y), h_nT_{n,1,0}(x,y), \ldots, h_n^pT_{n,p,0}(x,y)\Big)^t + o_P(1/\sqrt{n})$$

and moreover $T_{n,j,0}(x,y) = O_P(r_n)$ for $j = 0, ..., p$ uniformly in $x, y$. Together with Lemma C.7 observing that $e_1^t\mathcal{H}^{-1} = e_1^t$ completes the proof of (iva)'.
For a proof of (iva), we recall that by definition of $\mathcal{M}(K)$ we have for $j = 0, ..., p$

$$e_1^t\mathcal{M}(K)^{-1}(\mu_j(K), \ldots, \mu_{p+j}(K))^t = I\{j = 0\}.$$

It thus suffices to show that

$$\frac{1}{nh_n^{m+1}}\sum_{i=1}^n K_{h_n,k}^{(m)}(x-X_i)\frac{F_e(y|X_i)}{f_X(X_i)} = \sum_{j=0}^{p+1-m-1}\mu_{k+j}(K)h^j\partial_x^{m+j}F_e(y|x) + O_P\Big(h_n^{p+1-m} + \sqrt{\frac{\log n}{nh_n^{2m+1}}}\Big).$$

To this end we first observe that for $x \in [4h_n, 1-4h_n]$ integration-by parts and a Taylor expansion yields

$$\begin{aligned}
\mathbb{E}\Big[K_{h_n,k}^{(m)}(x-X_i)\frac{F_e(y|X_i)}{f_X(X_i)}\Big] &= \int_{x-h_n}^{x+h_n} K_{h_n,k}^{(m)}(x-u)F_e(y|u)du \\
&= h_n^m\int_{x-h_n}^{x+h_n} K_{h,k}(x-u)\partial_2^m F_e(y|u)du \\
&= h_n^{m+1}\sum_{j=0}^{p+1-m-1}\mu_{k+j}(K)h_n^j\partial_2^{m+j}F_e(y|x) + O(h_n^{p+2})
\end{aligned}$$

Finally, the estimate

$$\sup_{x\in[4h_n,1-4h_n],y\in\mathcal{Y}}\Big|\frac{1}{nh_n^{m+1}}\sum_{i=1}^n K_{h_n,k}^{(m)}(x-X_i)\frac{F_e(y|X_i)}{f_X(X_i)} - \frac{1}{h^{m+1}}\mathbb{E}\Big[K_{h_n,k}^{(m)}(x-X_i)\frac{F_e(y|X_i)}{f_X(X_i)}\Big]\Big|$$
$$= O_P\Big(\sqrt{\frac{\log n}{nh^{2m+1}}}\Big).$$

follows from the fact that the sets of functions

$$\mathcal{F}_n := \Big\{u \mapsto K_{h_n,k}^{(m)}(x-u)\frac{F_e(y|u)}{f_X(u)}\Big|x \in [4h_n, 1-4h_n], y \in \mathcal{Y}\Big\}$$

satisfy the assumptions of the second part of Lemma C.9 with $\delta_n = h_n$. Now the proof is complete. □

**Lemma C.7** *Under assumptions **(K1)** and **(A1)** if additionally $(nh_n)^{-1} = o(h_n\sqrt{logn})$ we have the decomposition (holding uniformly in $x \in [h_n, 1-h_n]$)*

$$(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1} = \frac{1}{f_X(x)}\mathcal{H}^{-1}\mathcal{M}(K)^{-1}\mathcal{H}^{-1} + \mathcal{H}^{-1}1_{(p+1)\times(p+1)}O_P(h)\mathcal{H}^{-1}$$

*where $\mathcal{H} = diag(1, h_n, ..., h_n^p)$, and $1_{(p+1)\times(p+1)}$ is a matrix with 1 in every entry.*

**Proof** The elements of the matrix $\mathbf{X}^t\mathbf{W}\mathbf{X}$ are of the form

$$(\mathbf{X}^t\mathbf{W}\mathbf{X})_{k,l} = \frac{1}{nh_n}\sum_i K_{h_n,0}(x-X_i)(x-X_i)^m = \frac{h_n^m}{nh_n^d}\sum_i K_{h_n,m}(x-X_i)$$

where $m = k+l-2$. In particular, continuous differentiability of $f_X$ together with an application of Lemma C.9 and Lemma C.10 implies that

$$\frac{1}{nh_n}\sum_i K_{h_n,k}(x-X_i) = \mu_k f_X(x) + O_P\Big(\Big(\frac{\log n}{nh_n}\Big)^{1/2} + h_n\Big)$$

uniformly in $x$. Thus we obtain a representation of the form

$$\mathbf{X}^t\mathbf{W}\mathbf{X} = \mathcal{H}\Big(\mathcal{M}(K)f_X(x) + 1_{N\times N}O_P(h_n)\Big)\mathcal{H}$$

where $M_0 = \mathcal{M}(K)$ is invertible and $\mathcal{H}$ is a diagonal matrix with entries $1, h_n, ..., h_n^p$. Thus for $h_n$ sufficiently small an application of the Neumann series yields the assertion with probability tending to one. □

## C.3 Additional technical results

**Lemma C.8** *Assume that $\kappa$ is a symmetric, uniformly bounded density with support $[-1,1]$ and let $b_n = o(1)$.*
*(a) If the function $F : [0,1] \to \mathbb{R}$ is strictly increasing and $F^{-1}$ is $k$ times continuously differentiable in a neighborhood of the point $\tau$, we have for $b_n$ small enough*

$$H_{id,\kappa,\tau,b_n}(F) = F^{-1}(\tau) + \sum_{i=1}^{k}\frac{b_n^i}{i!}(F^{-1})^{(i)}(\tau)\mu_{i+1}(\kappa) + R_n(\tau)$$

*with $|R_n(\tau)| \leq C_k(\kappa)b_n^k \sup_{|s-\tau|\leq b_n}|(F^{-1})^{(k)}(\tau) - (F^{-1})^{(k)}(s)|$, $\mu_i(\kappa) := \int u^i\kappa(u)du$ and a constant $C_k$ depending only on $k$ and $\kappa$. In particular, if we assume that $F : \mathbb{R} \to [0,1]$ is strictly increasing and $F^{-1}$ is two times continuously differentiable in a neighborhood of $\tau$ and $G : \mathbb{R} \to (0,1)$ is two times continuously differentiable in a neighborhood of $F^{-1}(\tau)$ with $G'(F^{-1}(\tau)) > 0$ we have*

$$|F^{-1}(\tau) - Q_{G,\kappa,\tau,b_n}(F)| \leq Cb_n^2 \sup_{|s-G\circ F^{-1}(\tau)|\leq R_{n,1}}|(G^{-1})'(s)| \sup_{|s-\tau|\leq b_n}|(G\circ F^{-1})''(s)| =: R_{n,2}$$

*for some constant $C$ that depends only on $\kappa$ where $R_{n,1} := Cb_n^2\sup_{|s-\tau|\leq b_n}|(G\circ F^{-1})''(s)|$.*

*(b) Assume that $\kappa$ is additionally differentiable with Lipschitz-continuous derivative and that the functions $G, G^{-1}$ have derivatives that are uniformly bounded on any compact subset of*

$\mathbb{R}$ *[the bound is allowed to depend on the interval]. Then for any increasing function $F$ with uniformly bounded first derivative we have $|H(F_1) - H(F_2)| \leq R_{n,3} + R_{n,4}$ and*

$$|Q_{G,\kappa,\tau,b_n}(F_1) - Q_{G,\kappa,\tau,b_n}(F_2)| \leq \sup_{u \in \mathcal{U}(H(F_1),H(F_2))} |(G^{-1})'(u)|(R_{n,3} + R_{n,4})$$

*where $C$ is a constant that depends only on $\kappa$, $\mathcal{U}(a,b) := [a \wedge b, a \vee b]$, and*

$$R_{n,3} := \frac{Cc_n}{b_n}\|F_1 - F_2\|_\infty \sup_{|v-\tau|\leq c_n} |(G \circ F^{-1})'(v)|, \quad R_{n,4} := R_{n,3}\frac{\|F_1 - F\|_\infty + \|F_1 - F_2\|_\infty}{b_n}$$

*with $c_n := b_n + 2\|F_1 - F_2\|_\infty + \|F_1 - F\|_\infty$.*

*(c) If additionally to the assumptions made in (b), the function $F_1$ is two times continuously differentiable in a neighborhood of $F^{-1}(\tau)$ with $F_1'(F_1^{-1}(\tau)) > 0$ and $G$ is two times continuously differentiable in a neighborhood of $F_1^{-1}(\tau)$ with $G'(F^{-1}(\tau)) > 0$, we have*

$$\begin{aligned} Q_{G,\kappa,\tau,b_n}(F_1) - Q_{G,\kappa,\tau,b_n}(F_2) \;=\; & -\frac{1}{F_1'(F_1^{-1}(\tau))}\int_{-1}^{1} \kappa(v)\Big(F_2(F_1^{-1}(\tau + vb_n)) - F_1(F_1^{-1}(\tau + vb_n))\Big)dv \\ & + R_n, \end{aligned}$$

*where*

$$|R_n| \;\leq\; R_{n,5} + R_{n,6} + \frac{Cb_n \sup_{|s-\tau|\leq b_n}(G \circ F^{-1})''(s)\|F_1 - F_2\|_\infty + R_{n,4}}{G'(F_1^{-1}(\tau))}$$

*with a constant $C$ depending only on $\kappa$ and*

$$\begin{aligned} R_{n,5} \;&:=\; \frac{1}{2}\sup_{u \in \mathcal{U}(H(F_1),H(F_2))} |(G^{-1})''(u)|(H(F_1) - H(F_2))^2 \\ R_{n,6} \;&:=\; \sup_{u \in \mathcal{U}(H(F_1),G(F_1^{-1})(\tau))} |(G^{-1})''(u)| \cdot |H(F_1) - G(F_1^{-1})(\tau)| \cdot |H(F_1) - H(F_2)|. \end{aligned}$$

**Proof** See Volgushev et al. (2013).

**Lemma C.9 (Basic Lemma)**

1. *Assume that the classes of functions $\mathcal{F}_n$ consist of uniformly bounded functions (with the bound, say $D$, not depending on $n$) with $N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P)) \leq C\exp(-c\varepsilon^{-a})$ for every $\varepsilon \leq \delta_n$ for some $a < 2$ and constants $C, c$ not depending on $n$. Then we have*

   $$\sqrt{n}\sup_{f \in \mathcal{F}_n, \|f\|_{P,2}\leq\delta_n}\Big(\int fdP_n - \int fdP\Big) = o_P^*(1)$$

   *where the $^*$ denotes outer probability, see van der Vaart and Wellner (1996) for a more detailed discussion.*

2. *If under the the assumptions of part one we have $N_{[]}(\mathcal{F}_n,\varepsilon,L^2(P)) \le C\varepsilon^{-a}$ for every $\varepsilon \le \delta_n$, some $a>0$ and $C$ not depending on $n$, it holds that for any $\delta_n \sim n^{-b}$ with $b<1/2$*

$$\sqrt{n} \sup_{f\in\mathcal{F}_n, \|f\|_{P,2}\le\delta_n} \Big( \int f dP_n - \int f dP \Big) = O_P^*\Big(\delta_n |\log\delta_n|\Big)$$

**Proof** See Volgushev et al. (2013).

**Lemma C.10**

1. *Define $\mathcal{F}+\mathcal{G} := \{f+g | f\in\mathcal{F}, g\in\mathcal{G}\}, \mathcal{F}\mathcal{G} := \{fg | f\in\mathcal{F}, g\in\mathcal{G}\}$. Then*

$$N_{[]}(\mathcal{F}+\mathcal{G},\varepsilon,\rho) \le N_{[]}(\mathcal{F},\varepsilon/2,\rho) N_{[]}(\mathcal{G},\varepsilon/2,\rho)$$

*If additionally the classes $\mathcal{F},\mathcal{G}$ are uniformly bounded by the constant $C$, we have*

$$N_{[]}(\mathcal{F}\mathcal{G},\varepsilon,\|.\|) \le N_{[]}^2(\mathcal{F},\varepsilon/4C,\|.\|) N_{[]}^2(\mathcal{G},\varepsilon/4C,\|.\|)$$

*for any seminorm $\|.\|$ with the additional property that $|f_2| \le |f_2|$ implies $\|f_1\| \le \|f_2\|$.*

2. *Assume that the Kernel $K$ has compact support $[-1,1]$, that $K_{1,k}^{(m)}$ is uniformly bounded and Lipschitz-continuous, and that $f_X$ is uniformly bounded. Then the $L^2(P_X)$ bracketing numbers $N_{[]}(\mathcal{F}_n,\varepsilon,L^2(P_X))$ of the set*

$$\mathcal{F}_n := \Big\{ u \mapsto K_{h_n,k}^{(m)}(x-u) \Big| x\in[h_n,1-h_n] \Big\}$$

*are bounded by $C\varepsilon^{-3}$ for some constant $C$ independent of $n$.*

3. *Assume that the Kernel $K$ has compact support $[-1,1]$, that $K$ is uniformly bounded, and that $f_X$ is uniformly bounded away from zero on $[0,1]$ and Lipschitz-continuous. Then for the set of function*

$$\mathcal{F}_n := \Big\{ u \mapsto \frac{1}{h_n}\Big(\frac{1}{f_X(x)} - \frac{1}{f_X(u)}\Big) K_{h_n,k}(x-u) \Big| x\in[h_n,1-h_n] \Big\}$$

*we have $N_{[]}(\mathcal{F}_n,\varepsilon,L^2(P)) \le C\varepsilon^{-5}$ for some constant $C$ independent of $n$.*

4. *For any measure $P$ on the unit interval with uniformly bounded density $f$, the class of functions*

$$\mathcal{F} := \Big\{ u \mapsto I\{u\le s\} \Big| s\in[0,1] \Big\} \cup \Big\{ u \mapsto I\{u< s\} \Big| s\in[0,1] \Big\}$$

*can be covered by $C\varepsilon^{-2}$ brackets of $L^2(P)$ length $\varepsilon$.*

5. *Consider the class of distribution functions* $\mathcal{F} := \left\{u \mapsto F(y|u)\Big|y \in \mathbb{R}\right\}$ *with densities* $f(y|u)$ *and assume that* $\sup_{u,y} |y|^\alpha (F(y|u) \wedge (1 - F(y|u)) \leq D$ *for some* $\alpha > 0$ *and additionally* $\sup_{u,y} f(y|u) \leq D$. *Then we have* $N_{[]}(\mathcal{F}, \varepsilon, \| \ \|_\infty) \leq C\varepsilon^{-\frac{\alpha+1}{\alpha}}$ *for some constant* $C$ *independent of* $\alpha$.

6. *For any measure* $P$ *on* $\mathbb{R} \times \mathbb{R}^k$ *with uniformly bounded conditional density* $f_{V|U}$ *the class of functions*
$$\mathcal{G} := \left\{(u,v) \mapsto I\{v \leq f(u)\}\Big|f \in \mathcal{F}\right\}$$
*satisfies* $N_{[]}(\mathcal{G}, \varepsilon, \|.\|_{P,2}) \leq N_{[]}(\mathcal{F}, C\varepsilon^2, \|.\|_\infty)$ *for some constant* $C$ *independent of* $\varepsilon$.

**Proof**

**Part 1** The first assertion is obvious from the definition of bracketing numbers. For the second assertion, note that $\mathcal{F}\mathcal{G} = (\mathcal{F} + C)(\mathcal{G} + C) - C\mathcal{F} - C\mathcal{G} + C^2$. Moreover, all elements of the classes $\mathcal{F} + C, \mathcal{G} + C$ are by construction non-negative and thus it also is possible to cover them with brackets consisting of non-negative functions and amounts equal to the brackets of $\mathcal{F}, \mathcal{G}$, respectively. Finally, observe that if $0 \leq f_l \leq f \leq f_u$ and $0 \leq g_l \leq g \leq g_u$, we also have $f_l g_l \leq fg \leq f_u g_u$. Moreover $\|f_l g_l - f_u g_u\| \leq C\|f_u - f_l\| + C\|g_u - g_l\|$. Thus the class $(\mathcal{F} + C)(\mathcal{G} + C)$ can be covered by at most $\leq N_{[]}(\mathcal{F}, \varepsilon, \|.\|)N_{[]}(\mathcal{G}, \varepsilon, \|.\|)$ brackets of length $2C\varepsilon$. Finding brackets for the classes $C\mathcal{F}, C\mathcal{G}$ is trivial, and applying the first assertion of the Lemma completes the proof.

**Part 2+3** Without loss of generality, assume that $h = h_n < 1$. The respective assumptions imply that it suffices to establish that for any class of functions $\mathcal{F}$ with uniformly bounded (say by C) elements that have supports of the form $[x - h, x + h]$ with $x$ from $[h, 1 - h]$ and $\sup_{f \in \mathcal{F}} |f(x) - f(y)| \leq C|x - y|h^{-k}$ uniformly in $x, y$ we we have $N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P_X)) \leq C\varepsilon^{-(2k+1)}$ for some $C$ that does not depend on $h$. Observe that in particular, the $L^2(P)$ norm of elements from $\mathcal{F}$ is bounded by $Dh^{1/2}$. Now consider two cases.

1 $\varepsilon > 4h^{1/2}$

Divide $[0, 1]$ into $N := 2/\varepsilon^2$ subintervals of length $2\alpha := \varepsilon^2$ with centers $r\alpha$ for $r = 1, ..., N$ and call the intervals $I_1, ..., I_N$. Note that two adjunct intervals overlap by $\alpha > 2h$. This construction ensures that every set of the form $[x - h, x + h]$ with $x \in [h, 1 - h]$ is completely contained in at least one of the intervals defined above. Then a collection of $N$ brackets of $L^2$-length $D\varepsilon$ for some $D > 0$ independent of $h$ is given by $(-CI\{u \in I_j\}, CI\{u \in I_j\})$.

2 $\varepsilon \leq 4h^{1/2}$

Observe that by assumption any element $g$ of $\mathcal{F}$ satisfies $|g(x) - g(y)| \leq C|x - y|h^{-k}$. Consider the points $t_i := i/(N + 1), i = 1, ..., N$ with $N := 2^{2k+1}C/\varepsilon^{2k+1}$. By construction, to every $x \in [h, 1 - h]$ there exists $i(x)$ with $|t_{i(x)} - x| \leq \varepsilon^{2k+1}/(2^{2k+1}C)$. This implies

$$|g(x) - g(t_{i(x)})| \leq C\varepsilon^{2k+1}h^{-k}/2^{2k+1}C \leq \varepsilon/2$$

Then $N$ $\|.\|_\infty-$brackets of length covering $\mathcal{F}$ are given by $(g(t_i) - \varepsilon/2, g(t_i) + \varepsilon/2)$, $i = 1, ..., N$. From those one can easily construct $L^2(P_X)$-brackets.

**Part 4** Follows by standard arguments.

**Part 5** For any $\varepsilon > 0$, set $y_\varepsilon := \varepsilon^{-\alpha}/D$ and define $t_i := -1/y_\varepsilon + i\varepsilon/2D$ for $i = 1, ..., N$ with such that $t_N \geq 1/y_\varepsilon$. Note that $N \leq C\varepsilon^{-\frac{\alpha+1}{\alpha}}$ for some fixed, finite constant $C$. The collection of brackets $(f \equiv 0, f \equiv \varepsilon), (f \equiv 1 - \varepsilon, f \equiv 1), (F(y_{t_i}|.) - \varepsilon/2, F(y_{t_i}|.) + \varepsilon/2)$ with $i = 1, ..., N$ covers the class $\mathcal{F}$.

**Part 6** Follows from $|I\{v \leq g_1(u)\} - I\{v \leq g_2(u)\}| \leq I\{|v - g_1(u)| \leq 2\|g_1 - g_2\|_\infty\}$.

$\square$