Ann Inst Stat Math (2015) 67:595–619
DOI 10.1007/s10463-014-0467-8

Smooth change point estimation in regression models with random design

Maik Döring · Uwe Jensen

Received: 4 June 2013 / Revised: 21 December 2013 / Published online: 18 May 2014

Abstract We consider the problem of estimating the location of a change point θ_0 in a regression model. Most change point models studied so far were based on regression functions with a jump. However, we focus on regression functions, which are continuous at θ_0. The degree of smoothness q_0 has to be estimated as well. We investigate the consistency with increasing sample size n of the least squares estimates $(\hat{\theta}_n, \hat{q}_n)$ of (θ_0, q_0). It turns out that the rates of convergence of $\hat{\theta}_n$ depend on q_0: for q_0 greater than $1/2$ we have a rate of $\sqrt{n}$ and the asymptotic normality property; for q_0 less than $1/2$ the rate is $n^{1/(2q_0+1)}$ and the change point estimator converges to a maximizer of a Gaussian process; for q_0 equal to $1/2$ the rate is $\sqrt{n \cdot \ln(n)}$. Interestingly, in the last case the limiting distribution is also normal.

Keywords Regression · Change points · M-estimates · Rate of consistency · Asymptotic distribution

1 Introduction

The problem to estimate the location of a change point in a regression model has been studied in the literature to some extent, see, among others, Hinkley (1971), Feder (1975), Müller (1992), Bai (1997), Csörgö and Horváth (1997), Müller and Song (1997), Rukhin and Vajda (1997), Müller and Stadtmüller (1999), Hušková (1999, 2001), Dempfle and Stute (2002), Koul et al. (2003) and Lan et al. (2009) and the

M. Döring (✉) · U. Jensen
Institut für Angewandte Mathematik und Statistik, Universität Hohenheim,
Stuttgart 70599, Germany
e-mail: maik.doering@uni-hohenheim.de

U. Jensen
e-mail: jensen@uni-hohenheim.de

cited references therein. In most cases locating a jump discontinuity is considered and properties of the estimators are studied. Müller (1992) investigates the problem of estimating a jump change point in the derivative of some order $\nu \geq 0$ of the regression function. His change point estimators are based on one-sided kernels. This includes the case of continuous regression functions with a change in the derivative at some point which we call smooth change point. In a number of applications one would rather model a smooth change point than a jump in the regression function. In particular, in the recently published article by Lan et al. (2009) the plotted dataset would suggest to fit a regression function with a smooth change instead of the proposed jump model. Hušková (1999, 2001) considers a least squares type estimator of the parameters in a location models with gradual changes in a fixed design setup. Here we focus on a similar approach in a random design regression model.

Let for $n \in \mathbb{N}$ the observations $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. $\mathbb{R}^2$-valued random variables. We assume that the distribution of X is absolutely continuous with a density function d_X, which is uniformly bounded on the unit interval $[0, 1]$. Further we assume that the response variables Y_i are given by the following regression model with an unknown change point $\theta_0 \in [0, 1)$, an unknown exponent $q_0 \in (0, \infty)$ and an unknown nuisance parameter vector $a_0 \in \mathbb{R}^d$

$$Y_i = f_{\theta_0, q_0, a_0}(X_i) + \epsilon_i, \quad 1 \leq i \leq n, \ n \in \mathbb{N}.$$

For $(\theta, q, a) \in [0, 1] \times [0, \infty) \times \mathbb{R}^d$ the regression function $f_{\theta,q,a} : \mathbb{R} \to \mathbb{R}$ is given by

$$f_{\theta,q,a}(x) := g(x, a) + h(x, a) \cdot (x - \theta)^q \mathbb{1}_{(\theta,1]}(x),$$

where $\mathbb{1}_A$ is the indicator function of a set A and the functions $g : \mathbb{R}^{d+1} \to \mathbb{R}$ and $h : \mathbb{R}^{d+1} \to \mathbb{R}$ are two times continuously differentiable. For example, g can be a polynomial and h a constant factor. Let $\epsilon, \epsilon_1, \dots, \epsilon_n$ for $n \in \mathbb{N}$ be i.i.d. real-valued random variables. We assume that $E(\epsilon|X) = 0$ a.s. and that the random variable ϵ is suitably integrable.

We will study here the limit behavior of the least squares estimators for (θ_0, q_0) and analyze the influence of the exponent on the estimation of the change point. The treatment of the nuisance parameter we defer to a forthcoming publication. There it can be shown, that the estimation of the nuisance parameter vector a_0 has no influence on the rate of convergence of the least squares estimator for (θ_0, q_0). To reduce the complexity, we assume here that the nuisance parameter a_0 is known. In this case without limiting the generality we can assume that $g \equiv 0$ and $h \equiv 1$, hence we look at regression functions

$$f_{\theta,q}(x) := (x - \theta)^q \mathbb{1}_{(\theta,1]}(x).$$

To make sure that the parameters are identifiable, we assume that

$$P(X \in (\theta_0, 1)) > 0. \tag{1}$$

We consider the least squares error for any possible change point and any exponent. For $(\theta, q) \in [0, 1] \times [0, \infty)$ and $n \in \mathbb{N}$ we define

$$M_n(\theta, q) := -\frac{1}{n}\sum_{i=1}^{n}(Y_i - f_{\theta,q}(X_i))^2.$$

For $n \in \mathbb{N}$ our estimator is defined as the maximizing point of M_n:

$$(\hat{\theta}_n, \hat{q}_n) := \underset{(\theta,q)\in[0,1]\times[0,\infty)}{\operatorname{argmax}} M_n(\theta, q).$$

Observe that $M_n(\theta, q) = \tilde{M}_n(\theta, q) - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2$, where

$$\begin{aligned}\tilde{M}_n(\theta, q) &:= \frac{1}{n}\sum_{i=1}^{n} m_{\theta,q}(\epsilon_i, X_i),\\ m_{\theta,q}(\epsilon, x) &:= -2\epsilon(f_{\theta_0,q_0}(x) - f_{\theta,q}(x)) - (f_{\theta_0,q_0}(x) - f_{\theta,q}(x))^2. \qquad (2)\end{aligned}$$

It follows that M_n and $\tilde{M}_n$ have the same maximizers. To analyze the asymptotic behavior of our estimator, we use the theory of M-estimators and empirical processes. For a fuller treatment we refer, for example, to Van der Vaart and Wellner (1996), Van der Vaart (1998), Van de Geer (2000) or Kosorok (2008).

Several authors consider the case, where q_0 is known. The case of two straight lines, i.e. $q_0 = 1$, has been studied before, see for example Hinkley (1971) and Feder (1975). The case that $q_0 = 0$, i.e. the regression function has a jump at θ_0, can be found for example in Kosorok (2008, Chap. 14.5.1), whereas the case that $q_0 > 2$, i.e. the regression function is twice differentiable at θ_0, has been considered by Rukhin and Vajda (1997) in a fixed design model. In these cases the rates of convergence of $\hat{\theta}_n$ are known (under suitable conditions) to be n for $q_0 = 0$ and $\sqrt{n}$ for $q_0 > 2$, respectively. In Döring and Jensen (2010) we have considered the problem in a fixed design model with equidistant design points. In the jump case it was shown that the rate is n and that the asymptotic distribution is that of a maximizer of a certain two-sided random walk. For $1 < q_0 \leq 2$ we proved that the rate is $\sqrt{n}$.

The obvious question what happens between these cases, i.e. for $0 < q_0 < 1$, has not been answered for the random design so far. We will show in our paper that with $\alpha := \min\{2q_0 + 1, 2\}$ for $q_0 \neq 1/2$ the rate of convergence of the change point estimator is $n^{1/\alpha}$. Surprisingly, we have found for $q_0 = 1/2$ a different rate, namely $\sqrt{n \ln n}$. This rate is in a sense monotone in q_0 and continuous in $q_0 = 0$, but discontinuous in $q_0 = 1/2$. Also, concerning the asymptotic distribution of the suitably scaled sequence $\hat{\theta}_n$ the point $q_0 = 1/2$ plays the role of a transition point: for $q_0 \geq 1/2$ the asymptotic distribution is normal, whereas for $0 < q_0 < 1/2$ the asymptotic distribution is that of a maximizer of a certain Gaussian process.

Hušková (1999, 2001) considers the problem in a fixed design model with equidistant design points. Therein, a least squares type estimator, where the change point is searched on a grid, is proposed and studied for $0 \leq q_0 \leq 1$. It turns out that the asymptotic behavior in the fixed design of Huškovás estimator is the same as in a random design setup of our estimator. The methods used to prove these similar results are completely different. In Hušková (2001) the extension to the case $q_0 = 0$ was obtained by a certain localization method.

Several authors studied smooth changes in other contexts. For example, Aue and Steinebach (2002) discuss an extension of Huškovás approach to certain stochastic processes. Ibragimov and Has'minskii (1981, Chap. 6) considered an estimator for a singularity z of a density function f with

$$f(x) = p(x)(z-x)^{\alpha}\mathbb{1}_{(-\infty,z)}(x) + q(x)(x-z)^{\alpha}\mathbb{1}_{[z,\infty)}(x),$$

where p and q are continuous functions and $\alpha \in (0, 1)$ is the order of singularity. Their proposed estimator for the singularity shows a similar asymptotic behavior as our estimator for the change point. Wang (1995) dealt with the detection of a sharp cusp in a white noise model. Therein the regression function f satisfies the so-called α-cusp property $|f(x_0+h) - f(x_0)| \geq K|h|^{\alpha}$ with positive constants K and α. This is the two-sided version of the cusp property, which is similar to the one-sided cusp property, that our regression function satisfies. Wang (1995) used a nonparametric approach by wavelets to estimate the sharp cusp.

The paper is organized as follows. In Sect. 2 we show consistency and in Sect. 3 we consider the rates of convergence of our estimator. In Sect. 4 the asymptotic distribution of the sequence of estimators is characterized. Section 5 contains numerical results. In Sect. 6 some conclusions are given. For improved readability all proofs of the lemmas are moved to the appendix.

2 Consistency

We prove that our estimators $(\hat{\theta}_n, \hat{q}_n)$ are strongly consistent. For $1 \leq i \leq n$ the random variables ϵ_i and X_i are i.i.d. and $E(\epsilon|X) = 0$ a.s., hence

$$E(\tilde{M}_n(\theta, q)) = E(m_{\theta,q}(\epsilon, X)) = -E((f_{\theta_0,q_0}(X) - f_{\theta,q}(X))^2) =: \tilde{M}(\theta, q).$$

Note that the functions $f_{\theta,q}$ are integrable. By (1) it follows that the deterministic function $\tilde{M} : [0, 1] \times [0, \infty) \to \mathbb{R}$ has a unique maximizer at (θ_0, q_0). By definition our estimator $(\hat{\theta}_n, \hat{q}_n)$ is a maximizer of $\tilde{M}_n$. We use the Glivenko–Cantelli theorem to show that $\tilde{M}_n$ converges for $n \to \infty$ uniformly to $\tilde{M}$.

Lemma 1 *Let $E(|\epsilon||X) < C$ a.s. for some $C > 0$. Then*

$$\lim_{n\to\infty} \sup_{(\theta,q)\in[0,1]\times[0,\infty)} |\tilde{M}_n(\theta, q) - \tilde{M}(\theta, q)| = 0 \ \ a.s.$$

By a corresponding argmax theorem we can transfer this convergence to the maximizing points.

Theorem 1 *Let $E(|\epsilon||X) < C$ a.s. for some $C > 0$. Then*

$$\lim_{n\to\infty} (\hat{\theta}_n, \hat{q}_n) \to (\theta_0, q_0) \ \ a.s.$$

Proof We need to check whether all assumptions of the well-known argmax theorem are satisfied. Let $\delta > 0$ and $(\theta_n, q_n)_{n\in\mathbb{N}}$ be a sequence with $(\theta_n, q_n) \in ([0, 1] \times [0, \infty)) \setminus ([\theta_0 - \delta, \theta_0 + \delta] \times [q_0 - \delta, q_0 + \delta])$. Since the function $f_{\theta,q}$ is monotone in θ and in q, it follows that

$$\liminf_{n\to\infty} \tilde{M}(\theta_n, q_n) \le \max\{\tilde{M}(\theta_0 - \delta, q_0 + \delta), \tilde{M}(\theta_0 + \delta, q_0 - \delta)\} < \tilde{M}(\theta_0, q_0).$$

The definition of our estimator yields $\tilde{M}_n(\hat{\theta}_n, \hat{q}_n) = \sup_{(\theta,q)\in[0,1]\times[0,\infty)} \tilde{M}_n(\theta, q)$ directly and by Lemma 1 we get $\lim_{n\to\infty} \sup_{(\theta,q)\in[0,1]\times[0,\infty)} |\tilde{M}_n(\theta, q) - \tilde{M}(\theta, q)| = 0$ a.s. Hence all assumptions of Theorem 2.12 in Kosorok (2008) are satisfied and the assertion follows. □

The assumption that the random variable X follows a distribution with a uniformly bounded density function can be relaxed. In the proof of Lemma 1 we only use that a constant $\tilde{C} > 0$ exists for all $0 \le \theta_1 < \theta_2 \le 1$ and for all $0 \le q_1 < q_2 < \infty$ with

$$E(f_{\theta_1,q_1}(X) - f_{\theta_2,q_2}(X)) \le \tilde{C}(\theta_2 - \theta_1 + q_2 - q_1).$$

Lemma 1 is also valid in the jump case, i.e. $q_0 = 0$. But to identify the change point and to get that the deterministic function $\tilde{M}$ has a unique maximizer, one has to assume instead of $P(X \in (\theta_0, 1)) > 0$ that $P(X \in [\theta_0 - \eta, \theta_0)) > 0$ and $P(X \in [\theta_0, \theta_0 + \eta)) > 0$ for any $\eta > 0$.

Misspecification. If one chooses the exponent in advance and estimates only the change point, then the least square estimator is not consistent for $q \neq q_0$. More precisely, let

$$\hat{\theta}_n(q) := \operatorname*{argmax}_{\theta\in[0,1]} M_n(\theta, q).$$

Also in this case the deterministic function $\tilde{M}$ has a unique maximizer

$$\tilde{\theta}(q) := \operatorname*{argmax}_{\theta\in[0,1]} \tilde{M}(\theta, q).$$

We have that $\tilde{\theta}(q) \neq \theta_0$ for $q \neq q_0$ and that $\tilde{\theta}(q_0) = \theta_0$. By the same arguments as in the proofs of Lemma 1 and Theorem 1 it follows that $\lim_{n\to\infty} \hat{\theta}_n(q) \to \tilde{\theta}(q)$ P-a.s.

3 Rate of convergence

From now on we additionally assume that:

$$\inf_{x\in[\theta_0,1]} d_X(x) > C_0 \text{ for some constant } C_0 > 0, \tag{3}$$

$$E(\epsilon^2|X) < C_1 \text{a.s. for some constant } C_1 > 0. \tag{4}$$

We are going to show that $\sqrt{n}(\hat{q}_n - q_0) = O_P(1)$ and that the sequence $\hat{\theta}_n$ is consistent at a rate which depends on q_0. We will show that $r_n(\hat{\theta}_n - \theta_0) = O_P(1)$ as $n \to \infty$, where the sequence $(r_n)_{n\in\mathbb{N}}$ is defined by

$$r_n := \begin{cases} n^{1/(2q_0+1)} & 0 \le q_0 < 1/2 \\ (n\ln(n))^{1/2} & q_0 = 1/2 \\ n^{1/2} & 1/2 < q_0. \end{cases}$$

For that purpose we define for $q \ge 0$ the functions $g_q : [0, (1-\theta_0)/\sqrt{e}] \to \mathbb{R}$ for $x > 0$ by

$$g_q(x) := \begin{cases} x^{(2q+1)/2} & 0 \le q \le 1/2 \\ x(\ln(1-\theta_0) - \ln(x))^{1/2} & q = 1/2 \\ x & 1/2 < q, \end{cases}$$

and $g_q(0) = 0$. Observe that $d_q(x, y) := g_q(\min\{|x-y|, (1-\theta_0)/\sqrt{e}\})$ is a metric. Let the function g_q^{-1} be the inverse function of g_q, which exists since g_q is continuous and strictly increasing. Let $H(\theta_0, q_0, \delta)$ be a δ-environment of (θ_0, q_0):

$$H(\theta_0, q_0, \delta) := \{(\theta, q) \in [0, 1] \times [0, \infty) : g_{q_0}^2(|\theta - \theta_0|) + (q - q_0)^2 < \delta^2\}.$$

Lemma 2 *Let the Assumption* (3) *be satisfied. Then constants* $\delta > 0$ *and* $C > 0$ *exist such that for all* $(\theta, q) \in H(\theta_0, q_0, \delta)$

$$E(m_{\theta,q}(\epsilon, X) - m_{\theta_0,q_0}(\epsilon, X)) \le -C(g_{q_0}^2(|\theta - \theta_0|) + (q - q_0)^2).$$

The next lemma provides an inequality, which is the main step, to get the rate of convergence.

Lemma 3 *Let the Assumption* (4) *be satisfied. Then there exists a constant* $C > 0$ *such that for all* $0 < \delta < \min\{g_{q_0}((1-\theta_0)/\sqrt{e}), q_0/2\}$

$$E\left(\sup_{(\theta,q)\in H(\theta_0,q_0,\delta)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (m_{\theta,q}(\epsilon_i, X_i) - E(m_{\theta,q}(\epsilon_i, X_i))) \right| \right) \le C\delta.$$

The next theorem states that $r_n(\hat{\theta}_n - \theta_0)$ is stochastically bounded.

Theorem 2 *Let the assumptions* (3) *and* (4) *are satisfied. Then for* $q_0 > 0$

$$r_n(\hat{\theta}_n - \theta_0) = O_p(1) \;\; and \;\; \sqrt{n}(\hat{q}_n - q_0) = O_p(1) \quad as \quad n \to \infty.$$

Proof The crucial part of the proof is to show that all assumptions of a corresponding rate of convergence theorem are satisfied. The condition (14.2) of Theorem 14.4 in Kosorok (2008) is satisfied by Lemma 2. The condition (14.3) with the functions $\phi_n(\delta) = \delta$ follows by Lemma 3. The sequence $\hat{\theta}_n$ is consistent by Theorem 1 and by the definition of $\hat{\theta}_n$ it follows that $(1/n)\sum_{i=1}^n m_{\hat{\theta}_n,\hat{q}_n}(\epsilon_i, X_i) \geq (1/n)\sum_{i=1}^n m_{\theta_0,q_0}(\epsilon_i, X_i)$. Therefore, by Theorem 14.4 in Kosorok (2008) we have that $\sqrt{n}(g_{q_0}(|\hat{\theta}_n - \theta_0|) + |\hat{q}_n - q_0|) = O_p(1)$. For $q_0 \neq 1/2$ it follows obviously that $r_n(\hat{\theta}_n - \theta_0) = O_p(1)$. For $q_0 = 1/2$ we get

$$\begin{aligned}
&\lim_{x\to\infty} \limsup_{n\to\infty} P(\sqrt{n\ln(n)}|\hat{\theta}_n - \theta_0| > x) \\
&\leq \lim_{x\to\infty} \limsup_{n\to\infty} P\left(\frac{x}{\sqrt{n\ln(n)}} < |\hat{\theta}_n - \theta_0| < \frac{1-\theta_0}{\sqrt{e}}\right) + P\left(\frac{1-\theta_0}{\sqrt{e}} \leq |\hat{\theta}_n - \theta_0|\right) \\
&= \lim_{x\to\infty} \limsup_{n\to\infty} P\left(g_{1/2}\left(\frac{x}{\sqrt{n\ln(n)}}\right) < g_{1/2}(|\hat{\theta}_n - \theta_0|) < g_{1/2}\left(\frac{1-\theta_0}{\sqrt{e}}\right)\right) \\
&= \lim_{x\to\infty} \limsup_{n\to\infty} P\left(x\left(\frac{\ln(1-\theta_0) - \ln(x) + \ln(\sqrt{n\ln(n)})}{\ln(n)}\right)^{1/2}\right. \\
&\qquad \left. < \sqrt{n}g_{1/2}(|\hat{\theta}_n - \theta_0|) < \sqrt{n}g_{1/2}\left(\frac{1-\theta_0}{\sqrt{e}}\right)\right) \\
&\leq \lim_{x\to\infty} \limsup_{n\to\infty} P(x/\sqrt{2} < \sqrt{n}g_{1/2}(|\hat{\theta}_n - \theta_0|)) = 0.
\end{aligned}$$

□

The case $q_0 = 0$, i.e. the regression function has a jump at θ, can be found for example in Kosorok (2008, Chap. 14.5.1). Therein it was shown by the same method that $r_n = n$ for $q_0 = 0$. Hence for fixed n the rate r_n as a function of the power q_0 is right continuous at $q_0 = 0$, but r_n is discontinuous at $q_0 = 1/2$.

4 Convergence in distribution

In this section we additionally assume that:

$$\text{the density } d_X \text{ of the random variable } X \text{ is continuous at } \theta_0, \tag{5}$$

$$E(\epsilon^2) < \infty \text{ and the random variables } \epsilon \text{ and } X \text{ are uncorrelated.} \tag{6}$$

We next show that $(r_n(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{q}_n - q_0))$ converges for $n \to \infty$ in distribution to a maximizer of a Gaussian process. In particular for $q_0 \geq 1/2$ our estimator is asymptotically normal. Let $\theta(s) := \theta_0 + sr_n^{-1}$ and $q(t) := q_0 + tn^{-1/2}$. For that purpose we define a sequence of stochastic processes $(Z_n)_{n\in\mathbb{N}}$ with $Z_n = \{Z_n(s,t) : (s,t) \in \mathbb{R}^2\}$ by

$$Z_n(s,t) := n\tilde{M}_n(\theta(s), q(t))$$

for $(s,t) \in [-r_n\theta_0, r_n(1-\theta_0)] \times [-\sqrt{n}q_0, \infty)$ and $Z_n(s,t) < 0$ otherwise. Observe that $(r_n(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{q}_n - q_0))$ is a maximizer of the process Z_n for $n \in \mathbb{N}$. We show that the sequence Z_n converges for $n \to \infty$ to a Gaussian process Z in some sense. Subsequently, we can transfer this convergence to the maximizing points by a continuous mapping theorem for the argmax functional. In the following we assume that $\theta_0 \in (0,1)$. Later on we will come back to the case of $\theta_0 = 0$.

Similar to the regression function we define for $(s,t) \in \mathbb{R}^2$ the function $\tilde{f}_{s,t} : \mathbb{R} \to \mathbb{R}$ by $\tilde{f}_{s,t}(y) := y^t \mathbb{1}_{(0,\infty)}(y) - (y-s)^t \mathbb{1}_{(s,\infty)}(y)$. For $q > 0$ we define the following constants

$$a_{11}(q) := \begin{cases} d_X(\theta_0) \int_{-\infty}^{\infty} \tilde{f}_{1,q}^2(y)\,\mathrm{d}y & 0 < q < 1/2 \\ d_X(\theta_0)/8 & q = 1/2 \\ E(q^2(X-\theta_0)^{2q-2} 1_{(\theta_0,1]}(X)) & 1/2 < q < \infty \end{cases}$$

$$a_{12}(q) := \begin{cases} 0 & 0 < q \leq 1/2 \\ E(q(X-\theta_0)^{2q-1}(-\ln(X-\theta_0))1_{(\theta_0,1]}(X)) & 1/2 < q < \infty \end{cases}$$

$$a_{22}(q) := E((X-\theta_0)^{2q}(\ln(X-\theta_0))^2 1_{(\theta_0,1]}(X)).$$

We define the symmetric 2×2 matrix $A(q)$ by $A(q)_{ij} := a_{ij}(q)$. We begin to show that the sequence of the mean value functions $(E(Z_n))_{n\in\mathbb{N}}$ converges uniformly. Let the function $z : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$z(s,t) := \begin{cases} -a_{11}(q_0) \cdot |s|^{2q_0+1} - a_{22}(q_0) \cdot t^2 & 0 < q_0 < 1/2 \\ -(s,t) \cdot A(q_0) \cdot (s,t)^T & 1/2 \leq q_0 < \infty. \end{cases}$$

Lemma 4 *Let the assumption* (5) *be satisfied and let* $\theta_0 \in (0,1)$. *Then we have for all* $\tilde{\delta} > 0$ *that*

$$\lim_{n\to\infty} \sup_{(s,t)\in[-\tilde{\delta},\tilde{\delta}]\times[-\tilde{\delta},\tilde{\delta}]} |E(Z_n(s,t)) - z(s,t)| = 0.$$

Next we look at the covariance function. Let the function $K : \mathbb{R}^4 \to \mathbb{R}$ be defined by $K((s_1,t_1),(s_2,t_2))$

$$:= 4E(\epsilon^2) \begin{cases} d_X(\theta_0) \int_{-\infty}^{\infty} \tilde{f}_{s_1,q_0}(y)\tilde{f}_{s_2,q_0}(y)\,\mathrm{d}y + a_{22}(q_0)t_1t_2 & 0 < q_0 < 1/2 \\ (s_1,t_1) \cdot A(q_0) \cdot (s_2,t_2)^T & 1/2 \leq q_0 < \infty. \end{cases}$$

Lemma 5 *Let the assumptions* (5) *and* (6) *are satisfied and let* $\theta_0 \in (0, 1)$. *Then we have for all* $(s, t) \in \mathbb{R}^2$ *that*

$$\lim_{n\to\infty} E((Z_n(s_1,t_1) - E(Z_n(s_1,t_1)))(Z_n(s_2,t_2) - E(Z_n(s_2,t_2))))$$
$$= K((s_1,t_1),(s_2,t_2)).$$

Let $Z = \{Z(s,t) : (s,t) \in \mathbb{R}^2\}$ for $\theta_0 \in (0, 1)$ be a continuous Gaussian process with the mean value function z and the covariance function K. We will denote by $X_n \rightsquigarrow X$ that a sequence of random elements $(X_n)_{n\in\mathbb{N}}$ converges in distribution to a random element X as $n \to \infty$. The following Lemma states that the sequence of processes $(Z_n)_{n\in\mathbb{N}}$ converges in distribution to the Gaussian process Z on each closed interval.

Lemma 6 *Let the assumptions* (5) *and* (6) *are satisfied and let* $\theta_0 \in (0, 1)$. *Then we have for all* $\tilde{\delta} > 0$ *that*

$$\{Z_n(s,t):\ (s,t) \in [-\tilde{\delta}, \tilde{\delta}]^2\} \rightsquigarrow \{Z(s,t):\ (s,t) \in [-\tilde{\delta}, \tilde{\delta}]^2\} \quad \textit{as } n \to \infty.$$

The case $\theta_0 = 0$ can be handled in much the same way. For $\theta_0 = 0$ we define the stochastic processes $(Z_n)_{n\in\mathbb{N}}$ and Z only on the index set $[0, \infty)$. Then the assertions of the Lemmas 4, 5 and 6 remain true. We can now formulate one of our main results.

Theorem 3 *Let the assumptions* (3), (5) *and* (6) *are satisfied. Then:*

(i) *The trajectories of* Z *possess a unique maximizer* (τ, κ) *almost surely.*
(ii)

$$(r_n(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{q}_n - q_0)) \rightsquigarrow (\tau, \kappa) \quad \textit{as } n \to \infty.$$

Proof Since $E(Z(s,t)) \to -\infty$ as $|s| + |t| \to \infty$, the process Z has a maximizer a.s. We have for $0 < q_0 < 1/2$ that

$$Var(Z(s_1,t_1) - Z(s_2,t_2)) = 4E(\epsilon^2)\left(d_X(\theta_0)\int_{-\infty}^{\infty} (\tilde{f}_{s_1,q_0}(y) - \tilde{f}_{s_2,q_0}(y))^2\,\mathrm{d}y\right.$$
$$\left. + a_{22}(q_0)(t_1 - t_2)^2\right),$$

for $q = 1/2$ that

$$Var(Z(s_1,t_1) - Z(s_2,t_2)) = 4E(\epsilon^2)(d_X(\theta_0)(s_1 - s_2)^2 + a_{22}(1/2)(t_1 - t_2)^2),$$

and for $q > 1/2$ that

$$Var(Z(s_1,t_1) - Z(s_2,t_2)) = 4E(\epsilon^2)\cdot$$
$$E((q_0(X-\theta_0)^{q_0-1}(s_1 - s_2) - (x-\theta_0)^{q_0}\ln(X-\theta_0)(t_1 - t_2))^2 1_{(\theta_0,1]}(X)).$$

Hence $Var(Z(s_1,t_1)-Z(s_2,t_2)) \neq 0$ for all $(s_1,t_1) \neq (s_2,t_2)$. Thus the first assertion follows by Lemma 2.6 of Kim and Pollard (1990). For the proof of the second assertion we use a continuous mapping theorem for the argmax functional. By Theorem 2 it follows that the sequence $(r_n(\hat{\theta}_n-\theta_0), \sqrt{n}(\hat{q}_n-q_0))$ is uniformly tight. By Lemma 6 we have for all compact sets $K \subset \mathbb{R}$ that

$$\{Z_n(t) : t \in K\} \rightsquigarrow \{Z(t) : t \in K\} \quad \text{as } n \to \infty.$$

Therefore, all conditions of Theorem 3.2.2 in Van der Vaart and Wellner (1996) are satisfied, which proves the theorem:

$$(r_n(\hat{\theta}_n-\theta_0), \sqrt{n}(\hat{q}_n-q_0)) = \underset{(s,t)\in\mathbb{R}^2}{\operatorname{argmax}}\, Z_n(s,t) \rightsquigarrow \underset{(s,t)\in\mathbb{R}^2}{\operatorname{argmax}}\, Z(s,t) = (\tau,\kappa).$$

□

An important consequence of the last theorem is that the rate r_n is the optimal rate of convergence. Observe that $Var(Z(s,t)) = -4E(\epsilon^2)\ E(Z(s,t))$. This allows us to compute the distribution of the random variable (τ,κ) for $q_0 \geq 1/2$. Since $A(q_0)$ is a positive definite and symmetric matrix, its inverse matrix $A(q_0)^{-1}$ and its square root $B(q_0)$ exist, i.e. $A(q_0) = B(q_0)\cdot B(q_0)^T$. Let N_1 and N_2 be two independent and standard normally distributed random variables. Then we have for $q_0 \geq 1/2$ the following representation of the distribution of the process Z:

$$\begin{aligned}
Z(s,t) &= -(s,t)\cdot A(q_0)\cdot(s,t)^T + \sqrt{4E(\epsilon^2)}\cdot(s,t)\cdot B(q_0)\cdot(N_1,N_2)^T \\
&= -\left((s,t)\cdot B(q_0) - \sqrt{E(\epsilon^2)}\cdot(N_1,N_2)\right) \\
&\quad \cdot\left((s,t)\cdot B(q_0) - \sqrt{E(\epsilon^2)}\cdot(N_1,N_2)\right)^T \\
&\quad + E(\epsilon^2)\cdot(N_1,N_2)\cdot(N_1,N_2)^T \\
&\leq E(\epsilon^2)\cdot(N_1,N_2)\cdot(N_1,N_2)^T = Z\left(\sqrt{E(\epsilon^2)}\cdot(N_1,N_2)\cdot B(q_0)^{-1}\right),
\end{aligned} \tag{7}$$

where the equalities hold in distribution. Hence the process Z has a unique maximizer $\sqrt{E(\epsilon^2)}\cdot(N_1,N_2)\cdot B(q_0)^{-1}$, which gives us the following result.

Theorem 4 *Let the assumptions (3), (5) and (6) be satisfied. Then for $n \to \infty$:*

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}_n-\theta_0)\\ \sqrt{n}(\hat{q}_n-q_0)\end{pmatrix} \rightsquigarrow \begin{pmatrix}\tau\\ \kappa\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, E(\epsilon^2)\cdot A(q_0)^{-1}\right) \quad 1/2 < q_0 < \infty$$

τ *and* κ *are independent* $\quad 0 < q_0 \leq 1/2$

$$\sqrt{n}(\hat{q}_n-q_0) \rightsquigarrow \kappa \sim N\left(0, \frac{E(\epsilon^2)}{a_{22}(q_0)}\right) \quad 0 < q_0 \leq 1/2$$

$$\sqrt{n\ln(n)}(\hat{\theta}_n-\theta_0) \rightsquigarrow \tau \sim N\left(0, \frac{8E(\epsilon^2)}{d_X(\theta_0)}\right) \quad q_0 = 1/2$$

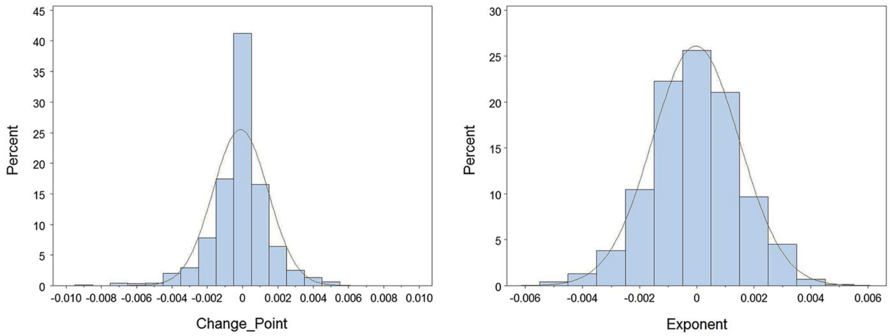

Fig. 1 Histogram for $n^{2/3}(\hat{\theta}_n - 0.2)$ *left*, and $\sqrt{n}(\hat{q}_n - 0.25)$ *right*, with $(\theta_0, q_0) = (0.2, 0.25)$

$$n^{1/(2q_0+1)}(\hat{\theta}_n - \theta_0) \rightsquigarrow \tau = \operatorname*{argmax}_{s\in\mathbb{R}} Z(s, 0) \quad 0 < q_0 < 1/2,$$

where the equality holds in distribution.

Proof The first assertion follows by (7). For $0 < q_0 \leq 1/2$ we can decompose the process $Z(s,t) = Z(s,0) + Z(0,t)$, hence the random variables τ and κ are independent. Analysis similar to (7) gives the other assertions. □

5 Numerical results

For $0 < q_0 < 1/2$ we have only a representation of the limit distribution of $n^{1/(2q_0+1)}(\hat{\theta}_n - \theta_0)$ as a unique maximizer of a Gaussian process. With the statistic software SAS we have simulated this distribution, where we have chosen X as uniformly distributed on the unit interval, ϵ as normally distributed with $E(\epsilon) = 0$ and $E(\epsilon^2) = 0.000001$ and $n = 1{,}000$. We have chosen a small variance $E(\epsilon^2)$, since the focus of this simulation is to show the type of the limit distribution of our estimator, rather than to study the influence of the variance. In Fig. 1 two histograms are displayed for 1,000 estimations of $n^{2/3}(\hat{\theta}_n - 0.2)$ and $\sqrt{n}(\hat{q}_n - 0.25)$ with $(\theta_0, q_0) = (0.2, 0.25)$. The line represents an estimated density for a normal distribution. Apparently $n^{2/3}(\hat{\theta}_n - 0.2)$ does not follow a normal distribution.

In comparison to Fig. 1 in Fig. 2 two histograms are displayed for 1,000 estimations of $\sqrt{n}(\hat{\theta}_n - 0.2)$ and $\sqrt{n}(\hat{q}_n - 1.5)$ with $(\theta_0, q_0) = (0.2, 1.5)$. In this case both asymptotic distributions are normal since $q_0 > 1/2$.

The variation of the dependence structure is shown in Fig. 3. One can clearly recognize that for $q_0 = 1.5$ the estimators $\hat{\theta}_n$ and $\hat{q}_n$ are dependent.

Further we have tested for normality of the change point estimator, where we simulated and estimated 200 times $\hat{\theta}_n$ with X as uniformly distributed on the unit interval, ϵ as normally distributed with $E(\epsilon) = 0$ and $E(\epsilon^2) = 0.000\,001$ and $n = 1{,}000$. In Table 1 the test values and the p values are given for the Shapiro–Wilk test and for the Kolmogorov–Smirnov test for several choices of (θ_0, q_0). As to be expected the simulations yield that the limit distribution of $n^{1/(2q_0+1)}(\hat{\theta}_n - \theta_0)$ is not normal for $0 < q_0 < 1/2$.

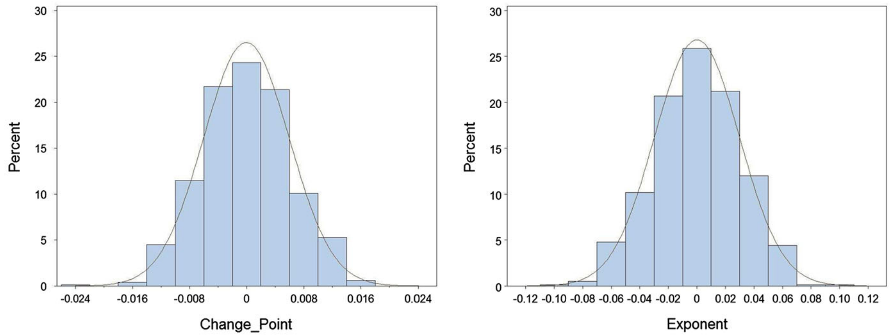

Fig. 2 Histogram for $\sqrt{n}(\hat{\theta}_n - 0.2)$ *left*, and $\sqrt{n}(\hat{q}_n - 1.5)$ *right*, with $(\theta_0, q_0) = (0.2, 1.5)$

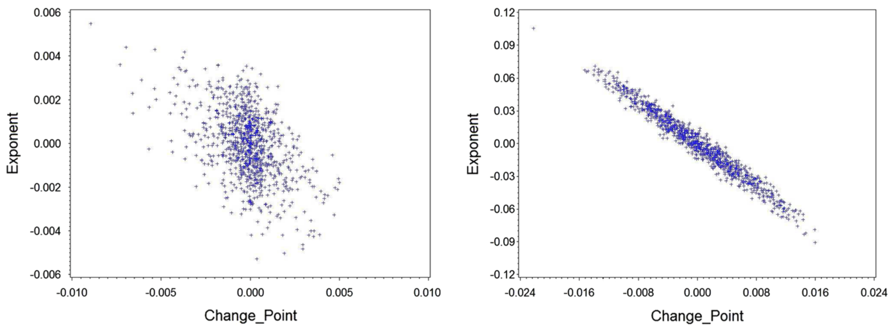

Fig. 3 Scatter plot for $(\hat{\theta}_n, \hat{q}_n)$ with $(\theta_0, q_0) = (0.2, 0.25)$ *left*, and with $(\theta_0, q_0) = (0.2, 1.5)$ *right*

Table 1 Tests for normality of $r_n(\hat{\theta}_n - \theta_0)$

θ_0	q_0	Shapiro–Wilk		Kolmogorov–Smirnov	
		Test value	p value	Test value	p value
0.8	0.25	≈0.917808	<0.0001	≈0.127415	<0.01
0.8	0.4	≈0.970939	≈0.0004	≈0.090642	<0.01
0.8	0.6	≈0.994188	≈0.6286	≈0.055055	≈0.1428
0.8	1	≈0.996312	≈0.9154	≈0.030766	>0.15
0.2	0.25	≈0.958407	<0.0001	≈0.108736	<0.01
0.2	0.4	≈0.964585	<0.0001	≈0.087246	<0.01
0.2	0.6	≈0.992458	≈0.3930	≈0.041445	>0.15
0.2	1	≈0.995179	≈0.7759	≈0.043467	>0.15

As a real-life example, we have applied our model to the weight/height charts (y in pounds/inch) as a function of the age (x in months) from 72 boys in the pre-school age, see Gallant (1987, p. 143/144), Dufner et al. (2004, p. 98–102). We use the slightly enlarged model

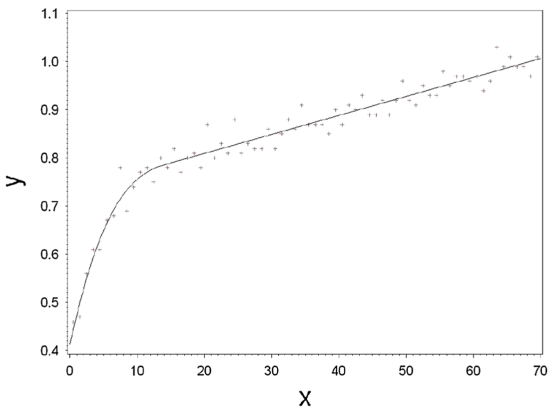

Fig. 4 Weight/height charts (y) as a function of the age (x)

$$y = a_1 + a_2 \cdot x + a_3 \cdot (\theta - x)^q \cdot 1_{[0,\theta)}(x),$$

where a_1, a_2, a_3, θ and q are unknown parameters. As above we remark that the estimation of the additional parameters a_1, a_2 and a_3 does not influence the rate of convergence of the estimates $\hat{\theta}_n$ and $\hat{q}_n$. By the least squares method we get the estimated regression function

$$y = 0.7308 + 0.0039 \cdot x - 0.0002 \cdot (15.06 - x)^{2.833} \cdot 1_{[0,\ 15.06)}(x).$$

That means the estimated value of the change point is $\hat{\theta} = 15.06$ and of the exponent is $\hat{q} = 2.833$. The data and the estimated regression function are plotted in Fig. 4.

6 Conclusions

We have shown that our least squares estimator for the change point and the exponent is for $q_0 \geq 1/2$ asymptotically normal. The variance of the limit distribution depends on (θ_0, q_0) and can be estimated consistently by a plug-in method for $q_0 \geq 1/2$. For example, the quantiles of the limit distribution could be used for the construction of confidence intervals. For $0 < q_0 < 1/2$ we have only a representation of the limit distribution as a unique maximizer of a Gaussian process. In the jump case, i.e. $q_0 = 0$, it turns out that $n(\hat{\theta}_n - \theta_0)$ converges to a maximizer of a two-sided random walk, see Kosorok (2008, Chap 14.5.1) for more details. An open problem is the derivation of an explicit representation of the limit distribution of the change point estimator for $0 \leq q_0 < 1/2$. Some simulations suggest that the limit distribution is not normal.

In comparison with Hušková (2001) we get the same structure of the limit distribution. But in our case the quantities $a_{ij}(q_0)$ are different compared to the more involved representation in Hušková (2001). If we assume that X is uniformly distributed on the unit interval, which comes the closest to the model of Hušková (2001), then we get in our model the following representations:

$$a_{11}(q) := \begin{cases} \int_{-\infty}^{\infty} \tilde{f}_{1,q}^2(y)\, \mathrm{d}y & 0 < q < 1/2 \\ 1/8 & q = 1/2 \\ \frac{q^2}{2q-1} \cdot (1-\theta_0)^{2q-1} & 1/2 < q < \infty \end{cases}$$

$$a_{12}(q) := \begin{cases} 0 & 0 < q \leq 1/2 \\ \frac{1}{2} \cdot (-\ln(1-\theta_0) + 1/2q) \cdot (1-\theta_0)^{2q} & 1/2 < q < \infty \end{cases}$$

$$a_{22}(q) := \frac{1}{2q+1} \cdot \left((\ln(1-\theta_0) - \frac{1}{2q+1})^2 + \frac{1}{(2q+1)^2} \right) \cdot (1-\theta_0)^{2q+1}.$$

It turns out that compared to the fixed design model of Hušková (2001) we obtain in our random design model a different covariance matrix of the asymptotic distribution of the least squares estimators.

Future work is to consider a more complex regression function

$$f_\theta(x) = g(x)(\theta - x)^p \mathbb{1}_{[0,\theta)}(x) + h(x)(x-\theta)^q \mathbb{1}_{[\theta,1]}(x),$$

where g and h are twice continuously differentiable unknown functions defined on [0, 1]. Such models were considered in a different context, see for example Ibragimov and Has'minskii (1981, Chap. 6). If one assumes that g and h belong to a parametric family of functions, then it is in a forthcoming paper shown that under suitable conditions the least squares estimators of the additional parameters are asymptotically normal.

Appendix

Proof of Lemma 1 Let the set $\mathscr{M}$ of functions $m_{\theta,q} : \mathbb{R}^2 \to \mathbb{R}$, see (2),

$$m_{\theta,q}(\mathfrak{e}, x) := -2\mathfrak{e}(f_{\theta_0,q_0}(x) - f_{\theta,q}(x)) - (f_{\theta_0,q_0}(x) - f_{\theta,q}(x))^2,$$

be defined by

$$\mathscr{M} := \{m_{\theta,q} \;:\; (\theta, q) \in [0,1] \times [0,\infty)\}.$$

Thus we have the following representation.

$$\sup_{(\theta,q)\in[0,1]\times[0,\infty)} |\tilde{M}_n(\theta,q) - \tilde{M}(\theta,q)| = \sup_{m\in\mathscr{M}} \left| (1/n)\sum_{i=1}^n m(\epsilon_i, X_i) - Em(\epsilon, X) \right|.$$

For $\eta > 0$ let $N_{[\,]}(\eta, \mathscr{M}, L_1)$ be the bracketing number of the class of functions $\mathscr{M}$ related to the L_1-norm, i.e. $\|m\|_{L_1} = E(|m(\epsilon, X)|)$, for details see Van der Vaart (1998, Chap. 19). We will show that the bracketing number $N_{[\,]}(\eta, \mathscr{M}, L_1)$ of the class $\mathscr{M}$ is finite for any $\eta > 0$, hence $\mathscr{M}$ is a Glivenko–Cantelli class by Theorem 19.4 in Van der Vaart (1998) and the assertion follows.

Let $0 \leq \theta_1 < \theta_2 \leq 1$ and $0 \leq q_1 < q_2 < \infty$ and fix $\mathfrak{e}$ and x. Then the function $m_{\cdot,\cdot}(\mathfrak{e}, x) : [\theta_1, \theta_2] \times [q_1, q_2] \to \mathbb{R}$ has at least one minimizer $(\underline{\theta}, \underline{q})$ and one maximizer $(\overline{\theta}, \overline{q})$. Since $f_{\theta,q}$ is nonnegative and decreasing in θ and in q we have by (2) that

$$\begin{aligned}
&\sup_{(\theta,q)\in[\theta_1,\theta_2]\times[q_1,q_2]} m_{\theta,q}(\mathfrak{e}, x) - \inf_{(\theta,q)\in[\theta_1,\theta_2]\times[q_1,q_2]} m_{\theta,q}(\mathfrak{e}, x) \\
&\quad = -2\mathfrak{e}(f_{\theta_0,q_0}(x) - f_{\overline{\theta},\overline{q}}(x)) - (f_{\theta_0,q_0}(x) - f_{\overline{\theta},\overline{q}}(x))^2 \\
&\qquad + 2\mathfrak{e}(f_{\theta_0,q_0}(x) - f_{\underline{\theta},\underline{q}}(x)) + (f_{\theta_0,q_0}(x) - f_{\underline{\theta},\underline{q}}(x))^2 \\
&\quad = (2\mathfrak{e} + 2f_{\theta_0,q_0}(x) - f_{\overline{\theta},\overline{q}}(x) - f_{\underline{\theta},\underline{q}}(x)) \cdot (f_{\overline{\theta},\overline{q}}(x) - f_{\underline{\theta},\underline{q}}(x)) \\
&\quad \leq |(2\mathfrak{e} + 2f_{\theta_0,q_0}(x) - f_{\overline{\theta},\overline{q}}(x) - f_{\underline{\theta},\underline{q}}(x)| \cdot |(f_{\overline{\theta},\overline{q}}(x) - f_{\underline{\theta},\underline{q}}(x)| \\
&\quad \leq |2\mathfrak{e} + 2|(f_{\theta_1,q_1}(x) - f_{\theta_2,q_2}(x)).
\end{aligned} \tag{8}$$

Since the density d_X of the distribution X is uniformly bounded on the unit interval and by $E(|\epsilon||X) < C$ a.s. it follows that

$$\begin{aligned}
&E\left(\sup_{(\theta,q)\in[\theta_1,\theta_2]\times[q_1,q_2]} m_{\theta,q}(\epsilon, X) - \inf_{(\theta,q)\in[\theta_1,\theta_2]\times[q_1,q_2]} m_{\theta,q}(\epsilon, X)\right) \\
&\quad \leq \tilde{C} \int_0^1 (f_{\theta_1,q_1}(x) - f_{\theta_2,q_2}(x)) d_X(x) \mathrm{d}x \\
&\quad \leq \tilde{C} \left(\int_{\theta_1}^1 (x-\theta_1)^{q_1} \mathrm{d}x - \int_{\theta_2}^1 (x-\theta_2)^{q_2} \mathrm{d}x\right) \\
&\quad \leq \tilde{C}((1-\theta_1)^{q_1+1}/(q_1+1) - (1-\theta_2)^{q_2+1}/(q_2+1)) \leq \tilde{C}(\theta_2 - \theta_1 + q_2 - q_1),
\end{aligned}$$

where $\tilde{C}$ is a positive generic constant. The last inequality follows by the convexity of the function $(1-\theta)^{q+1}/(q+1)$. Using standard methods it follows for any $\eta > 0$ that $N_{[\,]}(\eta, \mathscr{M}, L_1) \leq \tilde{C}/\eta^2 < \infty$. □

We present two technical lemmas, which give us a lower and an upper bound of certain integrals. Let $h_1(q) := q$ and for $k \in \mathbb{N} \setminus \{1\}$ let $h_k(q) := (q/k!) \prod_{i=2}^{k}(i - 1 - q)$. By Taylor's expansion we have

$$-(x-\theta_2)^q = -(x-\theta_1)^q + \sum_{k=1}^{\infty} h_k(q)(\theta_2 - \theta_1)^k (x-\theta_1)^{q-k}. \tag{9}$$

This representation will be useful for the calculation of some integrals. Observe that $h_k(q) \geq 0$ for $q \leq 1$ and, setting $x = \theta_2$ in (9), that $\sum_{k=1}^{\infty} h_k(q) = 1$.

Lemma 7 *For* $0 < c < 1$ *there exists a positive constant* $C > 0$*, such that for all* $0 \leq \theta_1 < \theta_2 < 1 - c$

$$\int_{\theta_2}^{1} (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x \geq C \begin{cases} (\theta_2 - \theta_1)^{2q+1} & 0 < q < 1/2 \\ (\theta_2 - \theta_1)^2 \ln\left(\frac{1-\theta_1}{\theta_2-\theta_1}\right) & q = 1/2 \\ (\theta_2 - \theta_1)^2 & 1/2 < q. \end{cases}$$

Proof For $0 < q \leq 1/2$ by (9) we have that

$$\begin{aligned} &\int_{\theta_2}^{1} (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x \\ &= \sum_{(k,l)\in\mathbb{N}^2} h_k(q)h_l(q)(\theta_2 - \theta_1)^{k+l} \int_{\theta_2}^{1} (x - \theta_1)^{2q-k-l} \, \mathrm{d}x \\ &\geq \sum_{(k,l)=(1,1)} h_k(q)h_l(q)(\theta_2 - \theta_1)^{k+l} \int_{\theta_2}^{1} (x - \theta_1)^{2q-k-l} \, \mathrm{d}x \\ &= q^2(\theta_2 - \theta_1)^2 \int_{\theta_2}^{1} (x - \theta_1)^{2q-2} \, \mathrm{d}x, \end{aligned}$$

where the inequality holds true, since $h_k(q) > 0$. Hence for $0 < q < 1/2$

$$\begin{aligned} &\int_{\theta_2}^{1} (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x \geq q^2 \frac{(\theta_2 - \theta_1)^{2q+1}}{1 - 2q} \left(1 - \left(\frac{\theta_2 - \theta_1}{1 - \theta_1}\right)^{1-2q}\right) \\ &\geq q^2 \frac{(\theta_2 - \theta_1)^{2q+1}}{1 - 2q} (1 - (1 - c)^{1-2q}) \geq C(\theta_2 - \theta_1)^{2q+1}. \end{aligned}$$

For $q = 1/2$ we have that

$$\int_{\theta_2}^{1} (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x \geq \frac{(\theta_2 - \theta_1)^2}{4} \ln\left(\frac{1 - \theta_1}{\theta_2 - \theta_1}\right).$$

For $\frac{1}{2} < q$ it follows by the mean value theorem with $\theta_1 < \xi < \theta_2 < 1 - c$

$$\begin{aligned} &\int_{\theta_2}^{1} (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x = (\theta_2 - \theta_1)^2 \int_{\theta_2}^{1} q^2(x - \xi)^{2q-2} \, \mathrm{d}x \\ &\geq (\theta_2 - \theta_1)^2 \left\{ \begin{array}{ll} \int_{1-c}^{1} q^2(x - 1 + c)^{2q-2} \, \mathrm{d}x & q \geq 1 \\ \int_{1-c}^{1} q^2 \, \mathrm{d}x & \frac{1}{2} < q < 1 \end{array} \right\} \geq C(\theta_2 - \theta_1)^2. \end{aligned}$$

□

Lemma 8 *There exists a positive constant* $C > 0$*, such that for all* $0 \leq \theta_1 < \theta_2 < 1$

$$\int_{0}^{1} (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x \leq C \begin{cases} (\theta_2 - \theta_1)^{2q+1} & 0 < q < 1/2 \\ (\theta_2 - \theta_1)^2 \ln\left(\frac{1-\theta_0}{\theta_2-\theta_1}\right) & q = 1/2 \\ (\theta_2 - \theta_1)^2 & 1/2 < q, \end{cases}$$

where in the case of $q = 1/2$ it is additionally assumed that $\theta_2 - \theta_1 < 1 - \theta_0$. For $q = 1/2$ the constant C may depend on θ_0.

Proof Similar to the proof of Lemma 7 by (9) we have that

$$\begin{aligned}\int_0^1 (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \,\mathrm{d}x &= \int_{\theta_1}^{\theta_2} (x-\theta_1)^{2q} \mathrm{d}x \\ &\quad + \int_{\theta_2}^1 ((x-\theta_1)^q - (x-\theta_2)^q)^2 \,\mathrm{d}x \\ &= \frac{1}{2q+1}(\theta_2-\theta_1)^{2q+1} \\ &\quad + \sum_{(k,l)\in\mathbb{N}^2} h_k(q)h_l(q)(\theta_2-\theta_1)^{k+l} \int_{\theta_2}^1 (x-\theta_1)^{2q-k-l} \,\mathrm{d}x.\end{aligned}$$

For $0 \le q < 1/2$ we have that $2q+1 < 2$, hence by $\sum_{k=1}^{\infty} h_k(q) = 1$

$$\begin{aligned}\int_0^1 (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \,\mathrm{d}x &= \frac{(\theta_2-\theta_1)^{2q+1}}{2q+1} \\ &\quad + \sum_{(k,l)\in\mathbb{N}^2} \frac{h_k(q)h_l(q)(\theta_2-\theta_1)^{k+l}}{2q+1-k-l}((1-\theta_1)^{2q+1-k-l} - (\theta_2-\theta_1)^{2q+1-k-l}) \\ &= (\theta_2-\theta_1)^{2q+1}\left(\frac{1}{2q+1} + \sum_{(k,l)\in\mathbb{N}^2} \frac{h_k(q)h_l(q)}{k+l-(2q+1)}\left(1-\left(\frac{\theta_2-\theta_1}{1-\theta_1}\right)^{k+l-(2q+1)}\right)\right) \\ &\le C(\theta_2-\theta_1)^{2q+1}.\end{aligned}$$

For $q = 1/2$ we get similarly

$$\begin{aligned}\int_0^1 (f_{\theta_1,1/2}(x) - f_{\theta_2,1/2}(x))^2 \,\mathrm{d}x &= \frac{(\theta_2-\theta_1)^2}{2} + \frac{(\theta_2-\theta_1)^2}{4}\ln\left(\frac{1-\theta_1}{\theta_2-\theta_1}\right) \\ &\quad + \sum_{\substack{(k,l)\in\mathbb{N}^2 \\ k+l>2}} \frac{h_k(q)h_l(q)(\theta_2-\theta_1)^2}{2-k-l}\left(1-\left(\frac{\theta_2-\theta_1}{1-\theta_1}\right)^{k+l-2}\right) \\ &\le C(\theta_2-\theta_1)^2 \ln\left(\frac{1-\theta_0}{\theta_2-\theta_1}\right).\end{aligned}$$

For $1/2 < q$ we have $(\theta_2-\theta_1)^{2q+1} < (\theta_2-\theta_1)^2$ and it follows by the mean value theorem with $\theta_1 < \xi < \theta_2$

$$\int_0^1 (f_{\theta_1,q}(x) - f_{\theta_2,q}(x))^2 \, \mathrm{d}x = \frac{(\theta_2-\theta_1)^{2q+1}}{2q+1} + (\theta_2-\theta_1)^2 \int_{\theta_2}^1 q^2 (x-\xi)^{2q-2} \, \mathrm{d}x$$
$$\leq \frac{(\theta_2-\theta_1)^{2q+1}}{2q+1} + (\theta_2-\theta_1)^2 \left\{ \begin{matrix} q^2 & q \geq 1 \\ \int_{\theta_2}^1 q^2 (x-\theta_2)^{2q-2} \, \mathrm{d}x & \frac{1}{2} < q < 1 \end{matrix} \right\}$$
$$\leq C(\theta_2-\theta_1)^2.$$

□

Proof of Lemma 2 $\delta > 0$ can be chosen such that for all $(\theta, q) \in H(\theta_0, q_0, \delta)$ we have that $\theta < 1-\delta$. Let $a \vee b = \max\{a, b\}$. By the mean value theorem, where $\tilde{\theta}$ is a point between θ_0 and θ, $\tilde{q}$ is a point between q_0 and q and by Lemma 7 we have that

$$E(m_{\theta,q}(\epsilon, X) - m_{\theta_0,q_0}(\epsilon, X)) = -\int_0^1 (f_{\theta_0,q_0}(x) - f_{\theta,q}(x))^2 d_X(x) \, \mathrm{d}x$$
$$\leq -C_0 \int_{\theta_0 \vee \theta}^1 (f_{\theta_0,q_0}(x) - f_{\theta,q}(x))^2 \, \mathrm{d}x$$
$$= -C_0 \int_{\theta_0 \vee \theta}^1 (f_{\theta_0,q_0}(x) - f_{\theta,q_0}(x) - (q-q_0) f_{\theta,\tilde{q}}(x) \ln(x-\theta))^2 \, \mathrm{d}x$$
$$= -C_0 \int_{\theta_0 \vee \theta}^1 (f_{\theta_0,q_0}(x) - f_{\theta,q_0}(x))^2 \, \mathrm{d}x - C_0 (q-q_0)^2 \int_{\theta_0 \vee \theta}^1 (f_{\theta,\tilde{q}}(x) \ln(x-\theta))^2 \, \mathrm{d}x$$
$$- 2C_0(\theta-\theta_0)(q-q_0) \int_{\theta_0 \vee \theta}^1 (-q_0 (x-\tilde{\theta})^{q_0-1} (x-\theta)^{\tilde{q}} \ln(x-\theta)) \, \mathrm{d}x$$
$$\leq -C_2 (g_{q_0}^2(|\theta-\theta_0|) + (q-q_0)^2) - C_3(\theta-\theta_0)(q-q_0),$$

where $C_2 > 0$ and $C_3 > 0$. For $(\theta-\theta_0)(q-q_0) > 0$ the assertion follows directly. For $0 < q_0 < 1/2$ and $(\theta-\theta_0)(q-q_0) < 0$ and since $(\theta, q) \in H(\theta_0, q_0, \delta)$ we can choose δ small, such that

$$-C_3(\theta-\theta_0)(q-q_0) = C_3 \cdot (\, g_{q_0}(|\theta-\theta_0|)\,)^{2/(2q_0+1)} \cdot |q-q_0|$$
$$\leq (C_3/2) \cdot \delta^{2/(2q_0+1)-1} \cdot (g_{q_0}^2(|\theta-\theta_0|) + (q-q_0)^2)$$
$$\leq (C_2/2) \cdot (g_{q_0}^2(|\theta-\theta_0|) + (q-q_0)^2).$$

By the same arguments the assertion holds true for $q_0 = 1/2$ and $(\theta-\theta_0)(q-q_0) < 0$. For $q_0 > 1/2$ the assertion follows by a second-order Taylor expansion. □

Proof of Lemma 3 For $0 < \delta < \min\{g_{q_0}((1-\theta_0)/\sqrt{e}), q_0/2\}$ let $\bar{\delta} := g_{q_0}^{-1}(\delta)$, then by the first-order Taylor expansion in q

$$
\begin{aligned}
&E\left(\sup_{(\theta,q)\in H(\theta_0,q_0,\delta)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m_{\theta,q}(\epsilon_i,X_i)-E(m_{\theta,q}(\epsilon_i,X_i)))\right|\right)\\
&\leq E\left(\sup_{\theta\in[\theta_0-\bar{\delta},\theta_0+\bar{\delta}]\cap[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m_{\theta,q_0}(\epsilon_i,X_i)-E(m_{\theta,q_0}(\epsilon_i,X_i)))\right|\right)\\
&\quad+\delta\cdot E\left(\sup_{(\theta,q)\in H(\theta_0,q_0,\delta)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{m}_{\theta,q}(\epsilon_i,X_i)-E(\tilde{m}_{\theta,q}(\epsilon_i,X_i)))\right|\right):\\
&=E_1+\delta E_2,
\end{aligned}
$$

where

$$
\tilde{m}_{\theta,q}(\mathfrak{e},x):=2\cdot(\mathfrak{e}+f_{\theta_0,q_0}(x)-f_{\theta,q}(x))\cdot f_{\theta,q}(x)\cdot\ln(x-\theta).
$$

We only state the case for E_1. By the same method we have that $E_2<\infty$. Let $\mathscr{M}_\delta$ be the following class of measurable functions:

$$
\mathscr{M}_\delta:=\{m_{\theta,q_0}\ :\ \theta\in[\theta_0-\bar{\delta},\theta_0+\bar{\delta}]\cap[0,1]\}. \tag{10}
$$

Further let $M_\delta:\mathbb{R}^2\to\mathbb{R}$ be a measurable function defined by

$$
M_\delta(\mathfrak{e},x):=|2\mathfrak{e}+2|(f_{\theta_0-\bar{\delta},q_0}(x)-f_{\theta_0+\bar{\delta},q_0}(x)).
$$

Similar to (8) we have that M_δ is an envelope function of the class $\mathscr{M}_\delta$, i.e. $|m(\mathfrak{e},x)|\leq M_\delta(\mathfrak{e},x)$ for all $x\in\mathbb{R}$, $\mathfrak{e}\in\mathbb{R}$ and for all $m\in\mathscr{M}_\delta$. For $\eta>0$ let $N_{[\,]}(\eta,\mathscr{M}_\delta,L_2)$ be the bracketing number of the class of functions $\mathscr{M}_\delta$ related to the L_2-norm, i.e. $\|m\|_{L_2}^2=E(|m(\epsilon,X)|^2)$. Let $\tilde{C}$ be a positive generic constant. By Corollary 19.35 in Van der Vaart (1998) it follows that

$$
E_1\leq\tilde{C}\int_0^{\|M_\delta\|_{L_2}}(\max\{1,\ln(N_{[\,]}(\eta,\mathscr{M}_\delta,L_2))\})^{1/2}\mathrm{d}\eta.
$$

Since the density d_X of the distribution X is uniformly bounded on the unit interval, by (8) and $E(\epsilon^2|X)<C$ a.s. it follows for $0\leq\theta_1<\theta_2<1$ that

$$
\begin{aligned}
&E\left(\left|\sup_{\theta_1\leq\theta\leq\theta_2}m_{\theta,q_0}(\epsilon,X)-\inf_{\theta_1\leq\theta\leq\theta_2}m_{\theta,q_0}(\epsilon,X)\right|^2\right)\\
&\leq E(|2\epsilon+2|^2(f_{\theta_1,q_0}(X)-f_{\theta_2,q_0}(X))^2)\\
&\leq\tilde{C}\int_0^1(f_{\theta_1,q_0}(x)-f_{\theta_2,q_0}(x))^2\mathrm{d}x\leq\tilde{C}g_{q_0}^2(\theta_2-\theta_1),
\end{aligned}
$$

where the last inequality follows by Lemma 8. Using standard methods it follows that

$$\begin{aligned}\|M_\delta\|_{L_2} &\le \tilde{C} g_{q_0}(2\bar{\delta}) \qquad (11)\\ N_{[\,]}(\eta, \mathscr{M}_\delta, L_2) &\le [2\bar{\delta}/g_{q_0}^{-1}(\tilde{C}^{-1}\eta)] + 1 \text{ for any} \eta > 0,\end{aligned}$$

where $[x] := \max\{k \in \mathbb{Z} : \ k \le x\}$. Let $\bar{\eta} := g_{q_0}^{-1}(\tilde{C}^{-1}\eta)$, then

$$\begin{aligned}&\int_0^{\|M_\delta\|_{L_2}} (\max\{1, \ln(N_{[\,]}(\eta, \mathscr{M}_\delta, L_2))\})^{1/2} \mathrm{d}\eta\\ &\le \int_0^{\tilde{C} g_{q_0}(2\bar{\delta})} (\max\{1, \ln([2\bar{\delta}/g_{q_0}^{-1}(\tilde{C}^{-1}\eta)] + 1)\})^{1/2} \mathrm{d}\eta\\ &= \tilde{C} \int_0^{2\bar{\delta}} (\max\{1, \ln([2\bar{\delta}/\bar{\eta}] + 1)\})^{1/2}\ g'_{q_0}(\bar{\eta}) \mathrm{d}\bar{\eta}\\ &= \tilde{C} \left(\int_0^{\bar{\delta}} ((\ln([2\bar{\delta}/\bar{\eta}] + 1))^{1/2}\ g'_{q_0}(\bar{\eta}))\ \mathrm{d}\bar{\eta} + g_{q_0}(2\bar{\delta}) - \delta \right) =: A_{q_0}.\end{aligned}$$

For $q_0 \in (0, 1/2)$ it follows

$$A_{q_0} \le \tilde{C} \left(\int_0^{\bar{\delta}} ((2\bar{\delta}/\bar{\eta})^{1/2} \bar{\eta}^{(2q_0-1)/2})\ \mathrm{d}\bar{\eta} + (2\delta^{2/(2q_0+1)})^{(2q_0+1)/2} - \delta \right) \le \tilde{C}\delta.$$

The same conclusion can be drawn for $q_0 > 1/2$. For $q_0 = 1/2$ we get

$$\begin{aligned}A_{1/2} \le \tilde{C} \Bigg(&\int_0^{\bar{\delta}} ((\ln(2\bar{\delta}/\bar{\eta}))^{1/2} (\ln((1-\theta_0)/\bar{\eta}))^{1/2})\ \mathrm{d}\bar{\eta}\\ &+ 2\bar{\delta}(\ln((1-\theta_0)/2\bar{\delta}))^{1/2} - \delta \Bigg) \le \tilde{C}(\bar{\delta} - \bar{\delta}\ln(\bar{\delta}) + \delta) \le \tilde{C}\delta.\end{aligned}$$

□

Proof of Lemma 4 For $\tilde{\delta} > 0$ there exists an n_0, such that, for all $n \ge n_0$, we have that $[-\tilde{\delta}, \tilde{\delta}]^2 \subseteq [-r_n\theta_0, r_n(1-\theta_0)] \times [-\sqrt{n}q_0, \infty)$. For $0 < s \le \tilde{\delta}$ by the mean value theorem and (9) we have that

$$\begin{aligned}E(Z_n(s,t)) &= -nE((f_{\theta_0,q_0}(X) - f_{\theta(s),q(t)}(X))^2)\\ &= -nE\left((f_{\theta_0,q_0}(X) - f_{\theta(s),q_0}(X) - \frac{t}{\sqrt{n}} f_{\theta(s),\tilde{q}(t)}(X) \ln(X - \theta(s)))^2 \right)\\ &= -nE((X-\theta_0)^{2q_0} 1_{(\theta_0,\theta(s)]}(X))\\ &\quad -n \sum_{(k,l)\in\mathbb{N}^2} h_k(q_0) h_l(q_0) (\theta(s) - \theta_0)^{k+l} E((X-\theta_0)^{2q_0-k-l} 1_{(\theta(s),1]}(X))\\ &\quad + \frac{2q_0 nst}{\sqrt{n} r_n} E((X - \tilde{\theta}(s))^{q_0-1} (X - \theta(s))^{\tilde{q}(t)} \ln(X - \theta(s)) 1_{(\theta(s),1]}(X))\end{aligned}$$

$$-t^2 E((X-\theta(s))^{2\tilde{q}(t)}(\ln(X-\theta(s)))^2 1_{(\theta(s),1]}(X))$$
$$=: b_{1,n}(s) + b_{2,n}(s) + b_{3,n}(s,t) + b_{4,n}(s,t),$$

where $\tilde{\theta}(s)$ is a point between θ_0 and $\theta(s)$ and $\tilde{q}(t)$ is a point between q_0 and $q(t)$. Since $f_{\theta(s),q(t)}(X)$ is continuous in s and t and by $\lim_{n\to\infty} \sqrt{n}/r_n = 1_{(1/2,\infty)}(q_0)$ we have that

$$\sup_{(s,t)\in[0,\tilde{\delta}]\times[-\tilde{\delta},\tilde{\delta}]} |b_{3,n}(s,t) - 2a_{12}(q_0)st| \to 0 \quad \text{as } n\to\infty$$
$$\sup_{(s,t)\in[0,\tilde{\delta}]\times[-\tilde{\delta},\tilde{\delta}]} |b_{4,n}(s,t) - a_{22}(q_0)t^2| \to 0 \quad \text{as } n\to\infty.$$

For $b_{1,n}(s) + b_{2,n}(s)$ we have to distinguish three cases:
1. $q_0 > 1/2$: we have $\theta(s) = \theta_0 + s/\sqrt{n}$ and

$$\begin{aligned}
\sup_{0\le s\le\tilde{\delta}} |b_{1,n}(s) + b_{2,n}(s) - a_{11}(q_0)s^2| &\le \sup_{0\le s\le\tilde{\delta}} nE((X-\theta_0)^{2q_0} 1_{(\theta_0,\theta(s)]}(X))\\
&+ \sup_{0\le s\le\tilde{\delta}} E(s^2 q_0^2 (X-\theta_0)^{2q_0-2} 1_{(\theta_0,\theta(s)]}(X))\\
&+ \sup_{0\le s\le\tilde{\delta}} n \sum_{(k,l)\in\mathbb{N}^2\setminus\{(1,1)\}} h_k(q_0)h_l(q_0)(s/\sqrt{n})^{k+l} E((X-\theta_0)^{2q_0-k-l} 1_{(\theta(s),1]}(X))\\
&=: A_n + B_n + C_n.
\end{aligned}$$

Let $\tilde{C}$ be a generic positive constant. Since the density of X is bounded, it follows that

$$\begin{aligned}
A_n &\le nE((X-\theta_0)^{2q_0} 1_{(\theta_0,\theta(\tilde{\delta})]}(X))\\
&\le \tilde{C}n \int_{\theta_0}^{\theta(\tilde{\delta})} (x-\theta_0)^{2q_0}\, \mathrm{d}x = \tilde{C}n(\tilde{\delta}/\sqrt{n})^{2q_0+1} \to 0 \text{ as } n\to\infty. \qquad (12)
\end{aligned}$$

By similar arguments we have that $\lim_{n\to\infty} B_n = 0$.

$$\begin{aligned}
C_n &\le \sup_{0\le s\le\tilde{\delta}} \tilde{C}n \sum_{(k,l)\in\mathbb{N}^2\setminus\{(1,1)\}} h_k(q_0)h_l(q_0)(s/\sqrt{n})^{k+l} \int_{\theta(s)}^1 (x-\theta_0)^{2q_0-k-l}\, \mathrm{d}x\\
&\le \tilde{C}n \Bigg(\sum_{\substack{(k,l)\in\mathbb{N}^2\setminus\{(1,1)\}\\ k+l<2q_0+1}} (h_k(q_0)h_l(q_0)/(2q_0+1-k-l))(\tilde{\delta}/\sqrt{n})^{k+l}\\
&+ \sum_{\substack{(k,l)\in\mathbb{N}^2\setminus\{(1,1)\}\\ k+l>2q_0+1}} (h_k(q_0)h_l(q_0)/(k+l-2q_0-1))(\tilde{\delta}/\sqrt{n})^{2q_0+1}
\end{aligned}$$

$$+ \sum_{\substack{(k,l)\in\mathbb{N}^2\setminus\{(1,1)\}\\ k+l=2q_0+1}} h_k(q_0)h_l(q_0)(\tilde{\delta}/\sqrt{n})^{2q_0+1}\ln(\sqrt{n}/\tilde{\delta})\Big) \to 0 \quad \text{as } n \to \infty. \tag{13}$$

2. $q_0 = 1/2$: we have $\theta(s) = \theta_0 + s/\sqrt{n\ln(n)}$ and

$$\begin{aligned}
&\sup_{0\le s\le\tilde{\delta}} |b_{1,n}(s) + b_{2,n}(s) - a_{11}(1/2)s^2| \le \sup_{0\le s\le\tilde{\delta}} nE((X-\theta_0)1_{(\theta_0,\theta(s)]}(X))\\
&+ \sup_{0\le s\le\tilde{\delta}} |\frac{s^2}{4\ln(n)}E((X-\theta_0)^{-1}1_{(\theta(s),1]}(X)) - \frac{d_X(\theta_0)}{8}s^2|\\
&+ \sup_{0\le s\le\tilde{\delta}} n \sum_{(k,l)\in\mathbb{N}^2\setminus\{(1,1)\}} h_k(q_0)h_l(q_0)(s/\sqrt{n\ln(n)})^{k+l}E((X-\theta_0)^{1-k-l}1_{(\theta(s),1]}(X))\\
&=: D_n + E_n + F_n.
\end{aligned}$$

Analysis similar to that in (12) and (13) shows $\lim_{n\to\infty} D_n = \lim_{n\to\infty} F_n = 0$. Since the function d_X is continuous at θ_0, it follows that

$$\begin{aligned}
E_n \le &(d_X(\theta_0)/4) \sup_{0\le s\le\tilde{\delta}} \left|\frac{s^2}{\ln(n)}\ln\left(\frac{1-\theta_0}{\theta(s)-\theta_0}\right) - \frac{s^2}{2}\right|\\
&+ \sup_{0\le s\le\tilde{\delta}} \left|\frac{s^2}{4\ln(n)}\int_{\theta(s)}^1 \frac{d_X(x)-d_X(\theta_0)}{x-\theta_0}\,\mathrm{d}x\right| \to 0 \quad \text{as } n\to\infty.
\end{aligned}$$

3. $0 < q_0 < 1/2$: for $0 < s \le \tilde{\delta}$ we have by the substitution $y = (x-\theta_0)r_n/s$ that

$$\begin{aligned}
-b_{1,n}(s) - b_{2,n}(s) &= n\Big(\int_{\theta_0}^{\theta(s)} (x-\theta_0)^{2q_0} d_X(x)\,\mathrm{d}x\\
&\quad + \int_{\theta(s)}^1 ((x-\theta_0)^{q_0} - (x-\theta(s))^{q_0})^2 d_X(x)\,\mathrm{d}x\Big)\\
&= s^{2q_0+1}\Big(\int_0^1 y^{2q_0} d_X(\theta_0 + ysr_n^{-1})\,\mathrm{d}y\\
&\quad + \int_1^{(1-\theta_0)r_n/s} (y^{q_0} - (y-1)^{q_0})^2 d_X(\theta_0 + ysr_n^{-1})\,\mathrm{d}y\Big).
\end{aligned}$$

By the dominated convergence theorem it follows, since the function d_X is continuous at θ_0, that

$$\begin{aligned}
&\sup_{0\le s\le\tilde{\delta}} |b_{1,n}(s) + b_{2,n}(s) - a_{11}(q_0)|s|^{2q_0+1}|\\
&\le \sup_{0\le s\le\tilde{\delta}} \left|\int_0^1 y^{2q_0}(d_X(\theta_0) - d_X(\theta_0 + ysr_n^{-1}))\,\mathrm{d}y\right|
\end{aligned}$$

$$+ \sup_{0\le s\le\tilde\delta} \left| \int_1^{(1-\theta_0)r_n/s} (y^{q_0} - (y-1)^{q_0})^2 (d_X(\theta_0) - d_X(\theta_0 + ysr_n^{-1}))\,\mathrm{d}y \right|$$
$$+ \sup_{0\le s\le\tilde\delta} \left| \int_{(1-\theta_0)r_n}^{\infty} (y^{q_0} - (y-1)^{q_0})^2 d_X(\theta_0)\,\mathrm{d}y \right| \to 0 \quad \text{as } n\to\infty.$$

The same conclusion can be drawn for $-\tilde\delta < s < 0$. □

Proof of Lemma 5 The proof is similar to the proof of the previous Lemma 4 and therefore is omitted for the sake of brevity. □

Proof of Lemma 6 We will show that $Z_n - E(Z_n)$ converge to a tight Gaussian process. Similar to (10) we define for $n \in \mathbb{N}$ and for $\tilde\delta > 0$ a class of measurable functions $\mathscr{M}_{n,\tilde\delta}$ by

$$\mathscr{M}_{n,\tilde\delta} := \left\{ \sqrt{n} m_{\theta,q} \,:\, (\theta, q) \in ([\theta(-\tilde\delta), \theta(\tilde\delta)] \cap [0,1]) \times ([q(-\tilde\delta), q(\tilde\delta)] \cap [0,\infty)) \right\}.$$

Observe that the measurable function $M_{n,\tilde\delta} : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$M_{n,\tilde\delta}(\mathfrak{e}, x) := \sqrt{n}|2\mathfrak{e} + 2| \left(f_{\theta(-\tilde\delta),q(-\tilde\delta)}(x) - f_{\theta(\tilde\delta),q(\tilde\delta)}(x) \right)$$

is an envelope function of the class $\mathscr{M}_{n,\tilde\delta}$ and observe that

$$Z_n(s,t) - E(Z_n(s,t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{n} m_{\theta(s),q(t)}(\epsilon_i, X_i) - E\left(\sqrt{n} m_{\theta(s),q(t)}(\epsilon, X)\right).$$

For $\eta > 0$ let $N_{[\,]}(\eta, \mathscr{M}_{n,\tilde\delta}, L_2)$ be the bracketing number of the class of functions $\mathscr{M}_{n,\tilde\delta}$ related to the L_2-norm. Further let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence with $\lim_{n\to\infty} \delta_n = 0$. Let $\tilde C$ be a generic positive constant. Analysis similar to that in the proof of Lemma 3, Eq. (11) gives

$$N_{[\,]}(\eta, \mathscr{M}_{n,\tilde\delta}, L_2) \le \tilde C([2\tilde\delta r_n^{-1}/g_{q_0}^{-1}(n^{-1}\eta)] + 1) \quad \text{for any } \eta > 0,$$
$$\lim_{n\to\infty} \int_0^{\delta_n} (\max\{1, \ln(N_{[\,]}(\eta, \mathscr{M}_{n,\tilde\delta}, L_2))\})^{1/2} \mathrm{d}\eta = 0.$$

Let $(s_1,t_1), (s_2,t_2) \in [-\tilde\delta, \tilde\delta]^2$ with $|s_1 - s_2| + |t_1 - t_2| < \delta_n$. Then we have that $m_{\theta(s_1),q(t_1)} \in \mathscr{M}_{n,\tilde\delta}$ and $m_{\theta(s_2),q(t_2)} \in \mathscr{M}_{n,\tilde\delta}$. Further let $\tilde s_1 = \min\{s_1, s_2\}$, $\tilde s_2 = \max\{s_1, s_2\}$, $\tilde t_1 = \min\{t_1, t_2\}$ and $\tilde t_2 = \max\{t_1, t_2\}$. By (8) and by Lemma 8 it follows that

$$\begin{aligned}
&\lim_{n\to\infty} E((\sqrt{n}m_{\theta(s_1),q(t_1)}(\epsilon, X) - \sqrt{n}m_{\theta(s_2),q(t_2)}(\epsilon, X))^2) \\
&\leq \lim_{n\to\infty} \tilde{C}nE(\epsilon^2)E((f_{\theta(\tilde{s}_1),q(\tilde{t}_1)}(X) - f_{\theta(\tilde{s}_2),q(\tilde{t}_2)}(X))^2) \\
&\leq \lim_{n\to\infty} \tilde{C}nE(\epsilon^2)(E((f_{\theta(\tilde{s}_1),q(\tilde{t}_1)}(X) - f_{\theta(\tilde{s}_1),q(\tilde{t}_2)}(X))^2) \\
&\quad + E((f_{\theta(\tilde{s}_1),q(\tilde{t}_2}(X) - f_{\theta(\tilde{s}_2),q(\tilde{t}_2)}(X))^2)) \\
&\leq \lim_{n\to\infty} \tilde{C}E(\epsilon^2)n(((t_2-t_1)/\sqrt{n})^2 + g^2_{q(\tilde{t}_2)}((s_2-s_1)/r_n)) \\
&\leq \lim_{n\to\infty} \tilde{C}E(\epsilon^2)(\delta_n^2 + ng^2_{q(\tilde{t}_2)}(\delta_n/r_n)) = 0. \qquad (14)
\end{aligned}$$

Next we show that the envelope function $M_{n,\tilde{\delta}}$ satisfies the Lindeberg condition. Similar to (14) we have that $E(M^2_{n,\tilde{\delta}}(\epsilon, X)) = O(1)$ as $n \to \infty$. By (9) we have for $\eta > 0$

$$\begin{aligned}
&\mathbb{1}_{(\eta\sqrt{n},\infty)}(M_{n,\tilde{\delta}}(\epsilon, X)) \\
&= \mathbb{1}_{(\eta,\infty)}(|2\epsilon + 2|(X-\theta(-\tilde{\delta}))^{q(-\tilde{\delta})})\, \mathbb{1}_{(\theta(-\tilde{\delta}),\theta(\tilde{\delta})]}(X) \\
&\quad + \mathbb{1}_{(\eta,\infty)}(|2\epsilon + 2|((X-\theta(-\tilde{\delta}))^{q(-\tilde{\delta})} - (X-\theta(\tilde{\delta}))^{q(\tilde{\delta})}))\, \mathbb{1}_{(\theta(\tilde{\delta}),1]}(X) \\
&\leq \mathbb{1}_{(\eta,\infty)}(|2\epsilon + 2|(2\tilde{\delta}r_n^{-1})^{q(-\tilde{\delta})}) + \mathbb{1}_{(\eta,\infty)}(|2\epsilon + 2|(2\tilde{\delta}r_n^{-1})^{\min\{q(-\tilde{\delta}),1\}}).
\end{aligned}$$

Hence

$$\begin{aligned}
&E(M_{n,\tilde{\delta}}(\epsilon, X)\, \mathbb{1}_{(\eta\sqrt{n},\infty)}(M_{n,\tilde{\delta}}(\epsilon, X))) \\
&\leq \tilde{C}E(\epsilon^2\, \mathbb{1}_{(\eta,\infty)}(|\epsilon|(\tilde{\delta}r_n^{-1})^{\min\{q(-\tilde{\delta}),1\}})) \\
&\quad \cdot E(n(f_{\theta(-\tilde{\delta}),q(-\tilde{\delta})}(\epsilon, X) - f_{\theta(\tilde{\delta}),q(\tilde{\delta})}(\epsilon, X))^2) \ \to 0 \text{ as } n \to \infty.
\end{aligned}$$

The sequence of the covariance functions converges pointwise by Lemma 5. Thus all assumptions of Theorem 19.28 in Van der Vaart (1998) are fulfilled, hence for all $\tilde{\delta} > 0$ we have that $\{Z_n(s,t) - E(Z_n(s,t)) : (s,t) \in [-\tilde{\delta}, \tilde{\delta}]^2\}$ converges to a tight Gaussian process. The trajectories of the limit process are continuous by Lemma 18.15 in Van der Vaart (1998). The assertion follows by Lemma 4 and by the Lemma of Slutsky. □

Acknowledgments The authors would like to thank the Associate Editor and Reviewers for their careful reading and comments. These comments and suggestions have been very helpful for revising and improving the manuscript.

References

Aue, A., Steinebach, J. (2002). A note on estimating the change-point of a gradually changing stochastic process. *Statistics & Probability Letters*, *56*, 177–191.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, *79*, 551–560.

Csörgö, M., Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: Wiley.

Dempfle, A., Stute, W. (2002). Nonparametric estimation of a discontinuity in regression. *Statistica Neerlandica*, *56*, 233–242.

Döring, M., Jensen, U. (2010). Change point estimation in regression models with fixed design. In V. Rykov, M. Nikulin, N. Balakrishnan (Eds.), *Mathematical and Statistical Models and Methods in Reliability* (pp. 207–222). New York: Springer.

Dufner, J., Jensen, U., Schumacher, E. (2004). *Statistik mit SAS (in German)*. Wiesbaden: Teubner.

Feder, P. L. (1975). On asymptotic distribution theory in segmented regression problems. *The Annals of Statistics*, *3*, 49–83.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: Wiley.

Hinkley, D. (1971). Inference in two-phase regression. *Journal of the American Statistical Association*, *66*, 736–743.

Hušková, M. (1999). Gradual changes versus abrupt changes. *Journal of Statistical Planning and Inference*, *76*, 109–125.

Hušková, M. (2001). A note on estimators of gradual changes. In M. de Gunst, C. Klaassen, A. van der Vaart (Eds.), *State of the art in probability and statistics (Leiden, 1999), 36* (pp. 345–358). Institute of Mathematical Statistics Lecture Notes-Monograph Series. Beachwood: Institute of Mathematical Statistics.

Ibragimov, I. A., Has'minskii, R. Z. (1981). *Statistical Estimation—Asymptotic Theory*. New York: Springer.

Kim, J., Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, *18*, 191–219.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

Koul, H. L., Qian, L., Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Process and their Applications*, *103*, 123–154.

Lan, Y., Banerjee, M., Michailidis, G. (2009). Change-point estimation under adaptive sampling. *The Annals of Statistics*, *37*, 1752–1791.

Müller, H. G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics*, *20*, 737–761.

Müller, H. G., Song, K. S. (1997). Two-stage change-point estimators in smooth regression models. *Statistics & Probability Letters*, *34*, 323–335.

Müller, H. G., Stadtmüller, U. (1999). Discontinuous versus smooth regression. *The Annals of Statistics*, *27*, 299–337.

Rukhin, A. L., Vajda, I. (1997). Change-Point Estimation as a Nonlinear Regression Problem. *Statistics*, *30*, 181–200.

Van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Van der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika*, *82*, 385–397.