ELSEVIER

Statistics & Probability Letters 34 (1997) 323–335

STATISTICS & PROBABILITY LETTERS

Two-stage change-point estimators in smooth regression models

Hans-Georg Müller [a,*], Kai-Sheng Song [b]

[a] *Division of Statistics, University of California, Davis, CA 95616, USA*
[b] *Department of Statistics, Florida State University, Tallahassee, FL 32306, USA*

Received July 1992; revised July 1996

Abstract

We consider a fixed design regression model where the regression function is assumed to be smooth, i.e., Lipschitz continuous, except for a point where it has only one-sided limits and a local discontinuity occurs. We propose a two-step estimator for the location of this change point and study its asymptotic convergence properties. In a first step, initial pilot estimates of the change point and associated asymptotically shrinking intervals which contain the true change point with probability converging to 1 are obtained. In the second step, a weighted mean difference depending on the assumed location of the change point is maximized within these intervals and the maximizing argument is then the final change point estimator. It is shown that this estimator attains the rate $O_p(n^{-1})$ in the fixed jump case. In the contiguous case, the estimator attains the rate $O_p(n^{-1}\Delta_n^{-2})$, where Δ_n is the sequence of jump sizes which in this case is assumed to converge to 0. For the contiguous case an invariance principle is established. A sequence of appropriately scaled deviation processes is shown to converge to a two-sided Brownian motion with triangular drift.

keyword: Asymptotics; Brownian motion; Discontinuity; Functional limit theorem; Nonparametric regression; Rate of convergence; Triangular drift; Weak convergence

1. Introduction

The problem of locating a discontinuity or change point in an otherwise smooth curve is highly relevant for the statistical description and analysis of discontinuous phenomena; compare, for instance, McDonald and Owen (1986) for various interesting applications. Estimates for change points in nonparametric regression functions have been proposed in Müller (1992) and Wu and Chu (1993) based on kernel estimates and in Müller (1993) based on locally weighted least-squares estimators. The locally weighted least-squares approach for change-point detection can be easily extended to cover generalized linear and quasi-likelihood models, using similar methods as in Fan et al. (1995). A somewhat different approach based on semi-parametric modelling was developed in Eubank and Speckman (1994).

Expressed in terms of "equivalent" kernels, all these estimators are different variants of essentially the same idea: Take differences of estimates of right- and left-sided limits of the unknown regression function and then use the maximizing argument as change-point estimate. The auxiliary estimates of one-sided limits of

* Corresponding author.

PII S0167-7152(96)00197-6

functions used in these procedures are constructed by using boundary kernels (or, in case of locally weighted least squares, equivalent boundary kernels) which have one-sided supports and center to a point on the edge of the support in terms of their moments. For this type of procedure, using smooth kernels, respectively, weight functions, it was shown in Müller (1992) that the O_p-rate of convergence is $n^{-1+\gamma}$ for arbitrary small $\gamma > 0$. The limit process of an appropriately scaled deviation process was found to be a Gaussian process with parabolic drift. A corresponding invariance principle was obtained for both fixed and contiguous cases.

We show here that the asymptotic rate of convergence can be improved by adding a second step to the difference-based procedure. We thus effectively achieve an O_p-rate of convergence of n^{-1} in the fixed jump case as well as a limit process. The limit process corresponds to a two-sided Brownian motion with drift in the "contiguous" case.

More specifically, we consider the following model: Assume n measurements $y_{1,n}, y_{2,n}, \ldots, y_{n,n}$ are made, following the fixed design regression model

$$y_{i,n} = g(x_{i,n}) + \varepsilon_{i,n}, \qquad x_{i,n} = i/n. \tag{1.1}$$

We assume that the errors $\varepsilon_{i,n}$ are i.i.d. with $E\varepsilon_{i,n} = 0$ and $E\varepsilon_{i,n}^2 = \sigma^2$, and that the unknown regression function g can be decomposed into a smooth part h and a discontinuous part:

$$g(x) = h(x) + \Delta_n 1_{[\tau,1]}(x), \quad 0 \leqslant x \leqslant 1, \quad |\Delta_n| = \gamma_n^{-1}, \tag{1.2}$$

where $1_A(\cdot)$ is the indicator function of a set A, τ is the unknown change point to be estimated, and Δ_n is the jump size which possibly depends on n; h is a "smooth" function, assumed to be Lipschitz continuous, but not necessarily differentiable, and γ_n is a positive sequence of real numbers. The "contiguous" jump size case (Bhattacharya and Brockwell, 1976), for which we will derive a functional limit theorem, is characterized by $\gamma_n \to \infty$ as $n \to \infty$, whereas in the fixed jump case, $\gamma_n \equiv$ const.

We note that no specific assumptions are necessary regarding the shape of the "smooth" part h of the regression function g, beyond basic smoothness properties. This implies that the relevant measurements for estimating the change point τ are only those in a (small) neighborhood around τ; if this neighborhood is sufficiently small, the actual shape of h will not matter. It will become evident that for asymptotic results one may approximate the regression function in a small neighborhood of τ by a step function

$$g \equiv c_1 + c_2 1_{[\tau,1]} \quad \text{with constants } c_1, c_2.$$

Assumptions and the proposed new estimator $\hat{\tau}$ for the change point τ are introduced in Section 2, which also contains the main results (Theorems 2.1–2.3). We show in Theorem 2.1 that under reasonable assumptions, one obtains for the rate of convergence $|\hat{\tau} - \tau| = O_p(n^{-1}\gamma_n^2)$. This implies $|\hat{\tau} - \tau| = O_p(n^{-1})$ for the fixed jump case.

In Theorem 2.2 below, a sequence of stochastic processes is introduced which have a properly scaled deviation between $\hat{\tau}$ and τ as argument. The weak convergence of this sequence of processes to a two-sided Brownian motion with triangular drift is established. Theorems 2.1 and 2.2 then imply the limit result (Theorem 2.3)

$$n\gamma_n^{-2}(\hat{\tau} - \tau) \xrightarrow{\mathscr{L}} \sigma^2 \operatorname*{arg\,max}_{t \in \Re} \left\{ W(t) - \frac{|t|}{2} \right\}$$

for the contiguous case, where $\gamma_n \to \infty$. Here $W(\cdot)$ stands for two-sided Brownian motion. These results indicate that the asymptotic convergence behavior of the new estimator is superior as compared to asymptotic convergence properties established previously for kernel-based estimators or other estimators of change points in a smooth regression setting under minimal restrictions on the errors $\varepsilon_{i,n}$.

In Section 3, it is demonstrated that a kernel-based pilot estimator exists which has the required properties. The proofs for the convergence rates are in Section 4, while those for the weak convergence and invariance principle are in Section 5.

2. Proposed estimators and main results

In the following we omit indices n whenever feasible. We assume $\tau \in (0,1)$ for the change point τ defined in (1.2) and make the following assumptions:

(A1) The errors $\varepsilon_{i,n}$ occurring in the model (1.1) form a triangular array of i.i.d. random variables with $E\varepsilon_{i,n} = 0$ and $\operatorname{var}(\varepsilon_{i,n}) = \sigma^2 < \infty$.

(A2) The "smooth" part h of the regression function in (1.2) is Lipschitz continuous on $[0,1]$.

(A3) The jump sizes Δ_n in (1.2) satisfy $|\Delta_n| = \gamma_n^{-1}$, where $\gamma_n \to \infty$, $n\gamma_n^{-2} \to \infty$ ("contiguous" jump case). Or alternatively,

(A3′) The jump size is fixed at $\Delta_n = \Delta$.

(A4) A pilot estimator $\tilde{\tau}$ of τ is available which satisfies

$$P(|\tilde{\tau} - \tau| \leqslant a_n) \to 1 \quad \text{as } n \to \infty,$$

for a sequence a_n satisfying

$$\gamma_n a_n \to 0, \quad \gamma_n^2 (na_n)^{-1} \to 0, \quad na_n^2\gamma_n^{-1} \to 0.$$

Note that this implies $a_n \to 0$ and $na_n \to \infty$ as $n \to \infty$.

A specific example for a pilot estimator $\tilde{\tau}$ satisfying (A4) is provided in Section 3. Consider now the random intervals

$$[r_0, r_1] = [\tilde{\tau} - a_n, \tilde{\tau} + a_n] \quad \text{and} \quad [s_0, s_1] = [\tilde{\tau} - 2a_n, \tilde{\tau} + 2a_n].$$

Let λ_n be the counting measure on the points $x_{i,n} = i/n$, i.e., $\lambda_n(A)$ is the count of those $x_{i,n}$ satisfying $x_{i,n} \in A$ for a set $A \subset \Re$. The statistic on which we base our proposed change point estimator $\hat{\tau}$ is

$$T_n(t) = [(t - s_0)(s_1 - t)]^{1/2} \left\{ \frac{1}{\lambda_n\{[t, s_1]\}} \sum_{x_i \in [t, s_1]} y_i - \frac{1}{\lambda_n\{[s_0, t)\}} \sum_{x_i \in [s_0, t)} y_i \right\}, \tag{2.1}$$

and we define

$$\hat{\tau} = \underset{t \in [r_0, r_1]}{\arg\max} |T_n(t)|. \tag{2.2}$$

Note that $T_n(t)$ is a weighted difference of the means of the observations falling into the subintervals defined by the partition of $[s_0, s_1]$ induced by t.

Theorem 2.1. *Assume* (A1), (A2) *and* (A4).

(A) *If* (A3′) *is satisfied* (*fixed jump case*), *then*

$$|\hat{\tau} - \tau| = O_p(n^{-1}). \tag{2.3}$$

(B) *If* (A3) *is satisfied* (*contiguous jump case*), *then*

$$|\hat{\tau} - \tau| = O_p(n^{-1}\gamma_n^2). \tag{2.4}$$

The proofs are in Section 4. This result shows that in the nonparametric regression setting the convergence rate for the proposed change-point estimate is n^{-1}, respectively $n^{-1}\gamma_n^2$, and therefore the same as the usual parametric rate for change points; compare Bhattacharya and Brockwell (1976) and Dümbgen (1991) for these results when a sequence of i.i.d. random variables is subject to a change in distribution at the change point. For this problem, nonparametric methods for change-point detection were also investigated by Darkhovski (1976) and Carlstein (1988). In these works, "nonparametric" refers to the unknown distribution of the sequence of i.i.d. random variables, whereas in the setting considered here, "nonparametric" refers to the unknown regression function which is only supposed to be "smooth", but is not required to belong to a class of finitely parametrized functions.

Our main result concerns the weak convergence of a suitably standardized deviation process to a two-sided Brownian motion with triangular drift. Let $\Rightarrow$ denote weak convergence in the space $\mathscr{D}([-M,M])$; cf. Billingsley (1968).

Theorem 2.2. *Let $M>0$ be an arbitrary constant and assume that* (A1)–(A4) *hold, with $\gamma_n \to \infty$ (contiguous jump case). Then*

$$Z_n(t) \Rightarrow Z(t) \quad \text{on } \mathscr{D}([-M,M]), \tag{2.5}$$

where

$$Z_n(t) = (n\gamma_n^{-1})\{[(\tau - s_0)(s_1 - \tau)]^{1/2}/(s_1 - s_0)\}\{|T_n(\tau + n^{-1}\gamma_n^2 t)| - |T_n(\tau)|\}, \tag{2.6}$$

$$Z(t) = \sigma W(t) - \tfrac{1}{2}|t|, \tag{2.7}$$

and $W(\cdot)$ is a two-sided Brownian motion.

In the fixed jump case, one obtains a random walk in the limit. The weak convergence result Theorem 2.2 together with the consistency result of Theorem 2.1 implies

Theorem 2.3. *Under* (A1)–(A4), *with $\gamma_n \to \infty$ (contiguous jump case),*

$$n\gamma_n^{-2}(\hat{\tau} - \tau) \xrightarrow{\mathscr{L}} \sigma^2 \operatorname*{arg\,max}_{t \in \Re} \left\{ W(t) - \frac{|t|}{2} \right\}. \tag{2.8}$$

The existence of a unique maximizer of the r.h.s. of (2.8) with probability 1 was demonstrated in Bhattacharya and Brockwell (1976).

3. A kernel-based pilot estimator

For the results in Section 2, a pilot estimator of τ is required which has property (A4). In this section, we provide examples for such pilot estimators which are based on differences of one-sided kernel estimates. These are aimed at estimating one-sided limits $g_+(x) = \lim_{y \downarrow x} g(y)$, $g_-(x) = \lim_{y \uparrow x} g(y)$, and are constructed with one-sided kernel functions K_+, K_-. As an example, one may use the fixed design nonparametric regression estimators

$$\hat{g}_{\pm}(x) = \frac{1}{b_n} \sum_{i=1}^{n} y_i \int_{(x_{i-1}+x_i)/2}^{(x_i+x_{i+1})/2} K_{\pm}\left(\frac{x-u}{b_n}\right) \mathrm{d}u, \tag{3.1}$$

which target $cg_{\pm}(x)$. Here $b = b_n$ is a sequence of bandwidths, c is a constant depending on the kernel (see (K1) below), and we set $x_0 = 0$, $x_{n+1} = 1$. The precise form of the estimator is not important, however, and one might as well choose other one-sided smoothers, like locally weighted least-squares estimators, where one would fit local lines in one-sided windows.

For the kernel functions K_+, K_- in (3.1) we require

(K1) K_- has support $[0,1]$, is bounded and Lipschitz continuous with the possible exception of finitely many points where only one-sided limits exist; K_- satisfies $c = \int_0^1 K_-(x)\,dx > 0$, $\int_0^1 K_-(x)x\,dx = 0$ and $\int_0^1 K_-(x)x^2\,dx < \infty$.

Examples are $K_-(x) = 12x(1-x)(3-5x)1_{[0,1]}$ or $K_-(x) = 4(1-\frac{3}{2}x)1_{[0,1]}$, both with $c = 1$. It is natural to relate K_+, K_- by

(K2) $K_+(x) = K_-(-x)$.

For the errors ε_i we make the additional assumption

(K3) $E|\varepsilon_i|^s < \infty$ for some $s > 4$.

Finally, for the sequence of bandwidths b_n, we require

(K4) $b_n \to 0$, $nb_n/\log n \to \infty$, $\gamma_n^2(nb_n)^{-1} \to 0$, $nb_n^2\gamma_n^{-1} \to 0$, $\liminf_{n\to\infty} n^{-2/r}(nb_n\log n)^{1/2} > 0$ for an r with $2 < r < s$, where s is as in (K3), and

(K5) $\gamma_n[b_n + (\log n/nb_n)^{1/2}] \to 0$,

where γ_n is the rate of convergence of the jump size as defined in (A3). Requirements (K4), (K5) on the bandwidth sequence are consistent and impose a lower bound on the rate of convergence $\Delta_n \to 0$ for the contiguous change-point case.

Defining the preliminary or pilot kernel estimate $\tilde{\tau}$ of τ by

$$\tilde{\tau} = \arg\max_x |\hat{g}_+(x) - \hat{g}_-(x)|, \tag{3.2}$$

the desired result is

Theorem 3.1. *Under* (K1)–(K5) *and* (A1)–(A3), *where* (A3) *can be replaced by* (A3′), *it holds that*

$$P(|\tilde{\tau} - \tau| \leqslant b_n) \to 1 \quad \text{as } n \to \infty. \tag{3.3}$$

It is obvious that (3.3) implies (A4) for pilot estimators $\tilde{\tau}$ (3.2), choosing $a_n = b_n$. The method of one-sided kernels may thus be employed as a first step to obtain appropriate initial estimators $\tilde{\tau}$, from which we obtain in a second step the final estimator $\hat{\tau}$ (2.2).

Proof of Theorem 3.1. We note that the assumptions on the errors (K3) and on Kernels (K1), (K2) and bandwidths (K4) allow us to infer, by slightly modifying the arguments in Müller and Stadtmüller (1987), that

$$\sup_{x\in I} |\hat{g}_\pm(x) - cg_\pm(x)| = O_p\left(b + \left(\frac{\log n}{nb}\right)^{1/2}\right), \tag{3.4}$$

where the interval I is such that all points in the b-neighborhood of I,

$$\mathcal{N}_b(I) = \{z \in [0,1] : |z - x| \leqslant b, x \in I\},$$

are continuity points of g, i.e., $g_+(y) = g_-(y)$ for all $y \in \mathcal{N}_b(I)$.

If we can ignore boundary effects, for instance by assuming that the range of abscissae where data are available extends somewhat beyond $[0,1]$, then (3.4) can be applied on $\bar{I}_n = [0, \tau - b] \cup [\tau + b, 1]$,

$$\sup_{x\subset \bar{I}_n} |\hat{g}_\pm(x) - cg_\pm(x)| = O_p\left(b + \left(\frac{\log n}{nb}\right)^{1/2}\right). \tag{3.5}$$

By definition of τ, this implies

$$\Delta_n^{-1}|\hat{g}_+(\tau) - \hat{g}_-(\tau)| \to c \quad \text{in probability as } n \to \infty. \tag{3.6}$$

It follows from (3.5) and (3.6) by a geometric consideration that $\tilde{\tau} = \arg\max |\hat{g}_+(x) - \hat{g}_-(x)|$ then satisfies

$$P(|\tilde{\tau} - \tau| > b_n) \leqslant P\left(\sup_{x \in \tilde{I}_n} |\hat{g}_\pm(x) - cg_\pm(x)| > \tfrac{1}{10} c\Delta_n \right) \to 0 \quad \text{as } n \to \infty,$$

according to (3.5) and (K5). □

We note that the familiar bias expansion can be carried out for $(\hat{g}_+(x) - \hat{g}_-(x)) - c(g_+(x) - g_-(x))$ without the assumption $\int K_-(z) z \, dz = 0$, if the smooth part h of the regression function is differentiable at x. Thus, if h is differentiable everywhere, except possibly at τ, then the result holds with condition (K1) modified by removing this additional moment condition.

4. Proofs: Consistency

Throughout this section, we use the functions

$$\rho(t) = |(t - s_0)(s_1 - t)|^{1/2},$$

$$R(t) = \rho(t) \left\{ (\lambda_n\{[t, s_1]\})^{-1} \sum_{x_i \in [t, s_1]} \varepsilon_i - (\lambda_n\{[s_0, t)\})^{-1} \sum_{x_i \in [s_0, t)} \varepsilon_i \right\}, \quad s_0 < t < s_1,$$

and $h(t)$ as defined in (1.2). Maxima and minima are denoted by $\vee$ and $\wedge$. We also define random variables

$$V_n = 2L(|\tau - s_0| \vee |\tau - s_1|),$$

where L is the Lipschitz constant of h. We assume that (A1)–(A4) hold, with (A3) possibly replaced by (A3′), and define the event

$$E_n = \{\tau \in [r_0, r_1]\}.$$

According to (A4), $P(E_n) \to 1$ as $n \to \infty$, and, denoting by E^c the complement of any event E, this implies for any sequence ζ_n and any $M > 0$:

$$P(\zeta_n |\hat{\tau} - \tau| > M) \leqslant P(\{\zeta_n |\hat{\tau} - \tau| > M\} \cap E_n) + P(E_n^c),$$

so that it suffices for the consistency proof to consider what happens on events E_n. The following arguments are therefore on events E_n, which is not always explicitly stated.

The proof of Theorem 2.1 is by contradiction. Suppose that $|\hat{\tau} - \tau| = O_p(n^{-1}\gamma_n^2)$ does not hold. Then there exists a $\delta > 0$ and $M_n \to \infty$ such that for events

$$F_n = \{n\gamma_n^{-2} |\hat{\tau} - \tau| > M_n\}$$

one has $\limsup_{n \to \infty} P(F_n) > \delta > 0$. One can therefore find subsequences of the n's where $P(F_n) > \delta/2$ and

$$P(E_n \cap F_n) > \delta/4 \quad \text{for } n \text{ large enough.} \tag{4.1}$$

The proof of Theorem 2.1 follows if

$$P(|T_n(\tau)| > |T_n(\hat{\tau})|) > 0, \tag{4.2}$$

as this leads to a contradiction with the fact that by definition,

$$\hat{\tau} = \arg \max_{t \in [r_0, r_1]} |T_n(t)| \quad \text{for all } n.$$

To show (4.2), representations of $T_n(t)$ are necessary which relate differences in the values of T_n to differences in the arguments.

Lemma 4.1. *For $t \in (s_0, s_1)$, on E_n, one has the following representation for T_n:*

$$T_n(t) = \rho(t)\left\{ \frac{\lambda_n\{[\tau, s_1]\}}{\lambda_n\{[t, s_1]\}} \wedge \frac{\lambda_n\{[s_0, \tau)\}}{\lambda_n\{[s_0, t)\}} \right\} \Delta_n + \rho(t)V_n(t) + R(t),$$

where $\sup_{t \in (s_0, s_1)} |V_n(t)| \leqslant V_n = O_p(a_n)$.

Proof. Assuming $t < \tau$, we find

$$T_n(t) = \rho(t)\left\{ (\lambda_n\{[t, s_1]\})^{-1} \sum_{x_i \in [t,\tau)} h(x_i) + (\lambda_n\{[t, s_1]\})^{-1} \sum_{x_i \in [t, s_1]} (h(x_i) + \Delta_n) \right.$$

$$\left. - (\lambda_n\{[s_0, t)\})^{-1} \sum_{x_i \in [s_0, t)} h(x_i) \right\} + R(t)$$

$$= \rho(t)\frac{\lambda_n\{[\tau, s_1]\}}{\lambda_n\{[t, s_1]\}}\Delta_n + \rho(t)\left\{ (\lambda_n\{[t, s_1]\})^{-1} \sum_{x_i \in [t, s_1)} h(x_i) - (\lambda_n\{[s_0, t)\})^{-1} \sum_{x_i \in [s_0, t)} h(x_i) \right\} + R(t).$$

Let L be the Lipschitz constant of h. Then

$$\left| (\lambda_n\{[t, s_1]\})^{-1} \sum_{x_i \in [t, s_1]} h(x_i) - (\lambda_n\{[s_0, t)\})^{-1} \sum_{x_i \in [s_0, t)} h(x_i) \right|$$

$$= \left| (\lambda_n\{[t, s_1]\})^{-1} \sum_{x_i \in [t, s_1]} (h(x_i) - h(\tau)) - (\lambda_n\{[s_0, t)\})^{-1} \sum_{x_i \in [s_0, t)} (h(x_i) - h(\tau)) \right|$$

$$\leqslant 2L(|\tau - s_0| \vee |\tau - s_1|) = V_n = O_p(a_n)$$

according to (A4). Thus,

$$T_n(t) = \rho(t)\frac{\lambda_n\{[\tau, s_1]\}}{\lambda_n\{[t, s_1]\}}\Delta_n + \rho(t)V_n + R(t). \tag{4.3}$$

Similar calculations yield in the case $t \geqslant \tau$:

$$T_n(t) = \rho(t)\frac{\lambda_n\{[s_0, \tau)\}}{\lambda_n\{[s_0, t)\}}\Delta_n + \rho(t)V_n + R(t), \tag{4.4}$$

and (4.3) and (4.4) together imply the result. $\square$

Lemma 4.2. *On E_n, it holds that*

$$|T_n(\tau)| - |T_n(\hat{\tau})| = \left[\rho(\tau) - \rho(\hat{\tau})\left(\frac{\lambda_n\{[\tau, s_1]\}}{\lambda_n\{[\hat{\tau}, s_1]\}} \wedge \frac{\lambda_n\{[s_0, \tau)\}}{\lambda_n\{[s_0, \hat{\tau})\}} \right)\right] |\Delta_n|$$

$$+ \{\operatorname{sign} \Delta_n\} \{[\rho(\tau) + \rho(\hat{\tau})]\, V_n + R(\tau) - R(\hat{\tau})\}. \tag{4.5}$$

Proof. Observe that by (A4),

$$|\Delta_n^{-1}|\, V_n = o_p(1), \tag{4.6}$$

and note that

$$\begin{aligned}
&|\Delta_n^{-1} R(\hat\tau)(\rho(\hat\tau))^{-1}| \\
&= \left|\Delta_n^{-1}\left\{ (\lambda_n\{[\hat\tau, s_1]\})^{-1} \sum_{x_i \in [\hat\tau, s_1]} \varepsilon_i - (\lambda_n\{[s_0, \hat\tau]\})^{-1} \sum_{x_i \in [s_0, \hat\tau]} \varepsilon_i \right\}\right| \\
&\leqslant \gamma_n (na_n)^{-1} \left\{ \left| \sum_{x_i \in [\hat\tau, s_1]} \varepsilon_i \right| + \left| \sum_{x_i \in [s_0, \hat\tau]} \varepsilon_i \right| \right\} \\
&\leqslant 3\gamma_n (na_n)^{-1} \left\{ \max_{\tau - 2a_n \leqslant t \leqslant \tau} \left| \sum_{x_i \in [t, \tau]} \varepsilon_i \right| + \max_{\tau \leqslant t \leqslant \tau + 2a_n} \left| \sum_{x_i \in [\tau, t]} \varepsilon_i \right| \right\}.
\end{aligned} \tag{4.7}$$

Since by Kolmogorov's inequality

$$\max_{\tau - 2a_n \leqslant t \leqslant \tau} \left| \sum_{x_i \in [t, \tau]} \varepsilon_i \right| = O_p([na_n]^{1/2}) \quad \text{and} \quad \max_{\tau \leqslant t \leqslant \tau + 2a_n} \left| \sum_{x_i \in [\tau, t]} \varepsilon_i \right| = O_p([na_n]^{1/2}),$$

it follows from (4.7) and (A4) that

$$|\Delta_n^{-1} R(\hat\tau)(\rho(\hat\tau))^{-1}| = O_p(\gamma_n [na_n]^{-1/2}) = o_p(1). \tag{4.8}$$

Now applying Lemma 4.1 for $t = \hat\tau$, we obtain from (4.6), (4.8)

$$|T_n(\hat\tau)| = \rho(\hat\tau) \left\{ \frac{\lambda_n\{[\tau, s_1]\}}{\lambda_n\{[\hat\tau, s_1]\}} \wedge \frac{\lambda_n\{[s_0, \tau]\}}{\lambda_n\{[s_0, \hat\tau]\}} \right\} |\Delta_n| + \{\operatorname{sign} \Delta_n\} \{\rho(\hat\tau) V_n + R(\hat\tau)\},$$

and analogously

$$|T_n(\tau)| = \rho(\tau) |\Delta_n| + \{\operatorname{sign} \Delta_n\} \{\rho(\tau) V_n + R(\tau)\},$$

whence the result follows. $\square$

Lemma 4.3. *Let*

$$Q_0 = \rho(\tau) - \rho(\hat\tau) \left(\frac{\lambda_n\{[\tau, s_1]\}}{\lambda_n\{[\hat\tau, s_1]\}} \wedge \frac{\lambda_n\{[s_0, \tau)\}}{\lambda_n\{[s_0, \hat\tau)\}} \right).$$

Then, on E_n,

$$Q_0 \geqslant |\hat\tau - \tau|/32.$$

Proof. Let $\hat\tau < \tau$. Then

$$\begin{aligned}
Q_0 &= \rho(\tau) - \rho(\hat\tau) \left(\frac{(s_1 - \tau)}{(s_1 - \hat\tau)} \wedge \frac{(\tau - s_0)}{(\hat\tau - s_0)} \right) \\
&= \rho(\tau) - \rho(\hat\tau) \frac{(s_1 - \tau)}{(s_1 - \hat\tau)} = \frac{(\tau - \hat\tau)(s_1 - \tau)(\hat\tau - s_0)(s_1 - s_0)}{\rho(\hat\tau)[\rho(\hat\tau)\rho(\tau) + (\hat\tau - s_0)(s_1 - \tau)]}.
\end{aligned}$$

Observe that $(\hat{\tau}-s_0)\vee(\tau-s_0)\vee(s_1-\hat{\tau})\vee(s_1-\tau)\leqslant s_1-s_0=4a_n$. Therefore,

$$\rho(\hat{\tau})[\rho(\hat{\tau})\rho(\tau)+(\hat{\tau}-s_0)(s_1-\tau)]\leqslant(4a_n)[(4a_n)^2+(4a_n)^2]=2(4a_n)^3.$$

Note also that $(s_1-\tau)\wedge(\hat{\tau}-s_0)\geqslant a_n$. We find

$$Q_0\geqslant\frac{(\tau-\hat{\tau})(s_1-\tau)(\hat{\tau}-s_0)(s_1-s_0)}{2(4a_n)^3}=\frac{(\tau-\hat{\tau})(s_1-\tau)(\hat{\tau}-s_0)}{2(4a_n)^2}\geqslant\frac{(\tau-\hat{\tau})a_n^2}{2(4a_n)^2}=\frac{(\tau-\hat{\tau})}{32}.$$

Analogously, for the case $\hat{\tau}\geqslant\tau$,

$$Q_0=\frac{(\hat{\tau}-\tau)(s_1-\hat{\tau})(\tau-s_0)(s_1-s_0)}{\rho(\hat{\tau})[\rho(\hat{\tau})\rho(\tau)+(\tau-s_0)(s_1-\hat{\tau})]}\geqslant\frac{(\hat{\tau}-\tau)}{32}. \qquad \square$$

Proof of Theorem 2.1. By Lemma 4.2,

$$|T_n(\tau)|-|T_n(\hat{\tau})|=Q_0|\Delta_n|\left\{1+(\operatorname{sign}\Delta_n)\left[\frac{(\rho(\tau)+\rho(\hat{\tau}))V_n}{Q_0|\Delta_n|}+\frac{R(\tau)-R(\hat{\tau})}{Q_0|\Delta_n|}\right]\right\}. \tag{4.9}$$

Now combine Lemma 4.3 with the fact that $|\hat{\tau}-\tau|>n^{-1}\gamma_n^2M_n$ on $E_n\cap F_n$. One finds that for arbitrarily small $\xi_1>0$, it holds for all sufficiently large n in the subsequence of n's which satisfy (4.1), that

$$P\left[\left\{\frac{|\rho(\tau)+\rho(\hat{\tau})|V_n}{Q_0|\Delta_n|}\leqslant\xi_1\right\}\cap E_n\cap F_n\right]\geqslant P\left[\left\{\frac{(\rho(\tau)+\rho(\hat{\tau}))\gamma_nV_n}{\hat{\tau}-\tau|}\leqslant\frac{\xi_1}{32}\right\}\cap E_n\cap F_n\right]$$

$$\geqslant P\left[\left\{\left|\frac{\rho(\tau)+\rho(\hat{\tau})}{a_n}\right|\frac{V_n}{a_n}\leqslant\frac{\xi_1\gamma_nM_n}{32na_n^2}\right\}\cap E_n\cap F_n\right].$$

Since $V_n/a_n=O_p(1)$, and $\gamma_nM_n/(na_n^2)\to\infty$ by (A4), this implies that for any arbitrary $\xi_2>0$, for sufficiently large n,

$$\left|P\left[\left\{\frac{|\rho(\tau)+\rho(\hat{\tau})|V_n}{Q_0|\Delta_n|}\leqslant\xi_1\right\}\cap E_n\cap F_n\right]-P(E_n\cap F_n)\right|<\xi_2. \tag{4.10}$$

Some simple algebra shows that

$$\begin{aligned}R(\tau)-R(\hat{\tau})&=[\rho(\tau)(\lambda_n\{[\tau,s_1]\})^{-1}-\rho(\hat{\tau})(\lambda_n\{[\hat{\tau},s_1]\})^{-1}]\sum_{x_i\in[\tau,s_1]}\varepsilon_i\\&+[\rho(\hat{\tau})(\lambda_n\{[s_0,\hat{\tau})\})^{-1}-\rho(\tau)(\lambda_n\{[s_0,\tau)\})^{-1}]\sum_{x_i\in[s_0,\tau)}\varepsilon_i\\&+\operatorname{sign}(\hat{\tau}-\tau)[\rho(\hat{\tau})(\lambda_n\{[\hat{\tau},s_1]\})^{-1}-\rho(\hat{\tau})(\lambda_n\{[s_0,\hat{\tau})\})^{-1}]\sum_{x_i\in[\tau\wedge\hat{\tau},\tau\vee\hat{\tau}]}\varepsilon_i\\&=:Q_1-Q_2+Q_3.\end{aligned} \tag{4.11}$$

Next, we note that the function $\varphi(t)=\rho(t)(\lambda_n\{[t,s_1]\})^{-1}$ satisfies

$$|\varphi(t_1)-\varphi(t_2)|\leqslant\frac{D}{na_n}|t_1-t_2|$$

for $r_0 \leqslant t_1$, $t_2 \leqslant r_1$ and a sufficiently large constant D. Observe that

$$\left|\sum_{x_i \in [\tau, s_1]} \varepsilon_i\right| \leqslant \max_{\tau \leqslant t \leqslant \tau + 2a_n} \left|\sum_{x_i \in [\tau, t]} \varepsilon_i\right| = O_p((na_n)^{1/2})$$

by Kolmogorov's inequality. Therefore, by (A4),

$$Q_1 / |\Delta_n(\hat\tau - \tau)| = O_p(\gamma_n [na_n]^{-1}) = o_p(1), \tag{4.12}$$

and analogously,

$$Q_2 / |\Delta_n(\hat\tau - \tau)| = o_p(1). \tag{4.13}$$

Let $\tau < \hat\tau$ without loss of generality. Then on $E_n \cap F_n$,

$$\frac{1}{n|\hat\tau - \tau|}\left|\sum_{x_i \in [\tau \wedge \hat\tau, \tau \vee \hat\tau]} \varepsilon_i\right| \leqslant \max_{M_n \gamma_n^2 \leqslant j \leqslant 2na_n} \left\{\frac{1}{j}\left|\sum_{i=1}^{j} \varepsilon_i\right|\right\}.$$

From this we obtain by the Hájek–Rényi inequality for constants $C_1, C_2, \lambda > 0$,

$$\begin{aligned}
P\left\{\left|\frac{Q_3}{\Delta_n(\hat\tau - \tau)}\right| > \lambda\right\} &\leqslant P\left\{\frac{\gamma_n}{n(\hat\tau - \tau)}\left|\sum_{x_i \in [\tau, \hat\tau)} \varepsilon_i\right| > \lambda / C_1\right\} \\
&\leqslant P\left(\max_{M_n \gamma_n^2 \leqslant j \leqslant 2na_n} \left\{\frac{1}{j}\left|\sum_{i=1}^{j} \varepsilon_i\right|\right\} > \gamma_n^{-1} \lambda / C_1\right) \leqslant \left(\frac{\sigma C_1 \gamma_n}{\lambda}\right)^2 \sum_{j=M_n \gamma_n^2}^{2na_n} \frac{1}{j^2} \\
&\leqslant \left(\frac{\sigma C_2 \gamma_n}{\lambda}\right)^2 \left(\frac{1}{M_n \gamma_n^2} - \frac{1}{2na_n}\right) \to 0,
\end{aligned} \tag{4.14}$$

as $n \to \infty$ by (A4) and since $M_n \to \infty$. Now combing (4.11)–(4.14) and Lemma 4.3 we obtain for any $\beta_1 > 0$, $\beta_2 > 0$, and for sufficiently large n, in analogy to (4.10),

$$\left|P\left[\left\{\frac{|R(\tau) - R(\hat\tau)|}{Q_0 |\Delta_n|} \leqslant \beta_1\right\} \cap E_n \cap F_n\right] - P(E_n \cap F_n)\right| < \beta_2. \tag{4.15}$$

Since $Q_0 < 0$, according to Lemma 4.3, it follows from (4.9) that

$$\begin{aligned}
&P[\{|T_n(\tau)| - |T_n(\hat\tau)| > 0\} \cap E_n \cap F_n] \\
&\quad \geqslant P\left[\left\{\left|\frac{(\rho(\tau) + \rho(\hat\tau)) V_n}{Q_0 \Delta_n}\right| < \frac{1}{4}\right\} \cap \left\{\left|\frac{R(\tau) - R(\hat\tau)}{Q_0 \Delta_n}\right| < \frac{1}{4}\right\} \cap E_n \cap F_n\right].
\end{aligned}$$

Now (4.10) and (4.15) imply that

$$|P(\{|T_n(\tau)| - |T_n(\hat\tau)| > 0\} \cap E_n \cap F_n) - P(E_n \cap F_n)| \leqslant \beta_2 + \xi_2 < \tfrac{1}{8}\delta,$$

by choosing appropriate β_2, ξ_2. It follows from (4.1) that $P(|T_n(\tau)| > |T_n(\hat\tau)|) > \frac{1}{8}\delta$, which implies (4.2). □

5. Proofs: Weak convergence

We note that by (A4),

$$Z_n(t)1_{[r_0,r_1]}(\tau) = Z_n(t) + o_p(1)$$

for processes Z_n (2.6), where the o_p-term is uniform in t. It is therefore sufficient to consider in the following only what happens on the events $E_n = \{r_0 \leqslant \tau \leqslant r_1\}$, as in Section 4. Consider sequences

$$u_n(t) = \tau + n^{-1}\gamma_n^2 t \quad \text{for } t \in \Re,$$

define

$$\alpha(t) = [(t-s_0)/(s_1-t)]^{1/2},$$

and consider processes

$$X_n(t) = n\Delta_n \rho(\tau)(s_1-s_0)^{-1}\,[T_n(u_n(t)) - T_n(\tau)].$$

We introduce the notations

$$S_n(v,z) = \sum_{x_i \in [v,z]} \varepsilon_i \quad \text{for } v \leqslant z,$$

defining $S_n(z,v) = -S_n(v,z)$ in case that $v > z$,

$$\tilde{X}_n(t) = n\Delta_n\rho(\tau)(s_1-s_0)^{-1}\left\{\rho(u_n(t))\left\{\frac{\lambda_n\{[\tau,s_1]\}}{\lambda_n\{[u_n(t),s_1]\}} \wedge \frac{\lambda_n\{[s_0,\tau)\}}{\lambda_n\{[s_0,u_n(t))\}}\right\}\Delta_n - \rho(\tau)\Delta_n\right\},$$

and

$$\bar{X}_n(t) = \Delta_n S_n(u_n(t),\tau) + \tilde{X}_n(t). \tag{5.1}$$

Lemma 5.1. *It holds that*

$$X_n(t) = \bar{X}_n(t) + o_p(1), \tag{5.2}$$

where the remainder term is uniform in $t \in [-M,M]$.

Proof. From Lemma 4.1,

$$\begin{aligned}
X_n(t) &= n\Delta_n\rho(\tau)(s_1-s_0)^{-1}\{[R(u_n(t)) - R(\tau)] + [(\rho(u_n(t)) - \rho(\tau))O_p(a_n)]\} + \tilde{X}_n(t)\\
&= n\Delta_n\rho(\tau)(s_1-s_0)^{-1}\{[[\alpha(u_n(t)) - \alpha(\tau)]n^{-1}S_n(\tau,s_1)]\\
&\quad + [([\alpha(\tau)]^{-1} - [\alpha(u_n(t))]^{-1})n^{-1}S_n(s_0,\tau)]\\
&\quad + [([\rho(u_n(t))]^{-1} - [\rho(\tau)]^{-1})(s_1-s_0)n^{-1}S_n(u_n(t),\tau)]\\
&\quad + [(\rho(u_n(t)) - \rho(\tau))O_p(a_n)] + (s_1-s_0)[\rho(\tau)]^{-1}n^{-1}S_n(u_n(t),\tau)\} + \tilde{X}_n(t)\\
&= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \Delta_n S_n(u_n(t),\tau) + \tilde{X}_n(t).
\end{aligned}$$

Next we evaluate I–IV. By a Taylor expansion for α around τ and applying (A4), we obtain

$$|\mathrm{I}| = O_p(n\gamma_n^{-1}a_n^{-1}|u_n(t)-\tau|n^{-1}[na_n]^{1/2}) = O_p\left(\left[\frac{\gamma_n^2}{na_n}\right]^{1/2}\right) = o_p(1)$$

$$|\mathrm{II}| = O_p([[na_n]^{-1}\gamma_n^2]^{1/2}) = o_p(1),$$

$$|\mathrm{III}| = O_p(n\gamma_n^{-1}a_n^{-1}|u_n(t)-\tau|n^{-1}|n(u_n(t)-\tau)|^{1/2}) = O_p([na_n]^{-1}\gamma_n^2) = o_p(1)$$

$$|\mathrm{IV}| = O_p(n\gamma_n^{-1}a_n^2) = o_p(1),$$

where the remainder terms are found to be uniform in $t \in [-M,M]$. □

Lemma 5.2. *It holds that*

$$E\bar{X}_n(t) = -\tfrac{1}{2}|t| + o(1). \tag{5.3}$$

$$\bar{X}_n(t) - E\bar{X}_n(t) = \Delta_n S_n(u_n(t),\tau) + o(1), \tag{5.4}$$

where the remainder terms are uniform in $t \in [-M,M]$.

Proof. As $ES_n(u_n(t),\tau) = 0$, we find from (5.1) that

$$E\bar{X}_n(t) = E\tilde{X}_n(t). \tag{5.5}$$

Note that

$$\tilde{X}_n(t) = n\Delta_n\rho(\tau)(s_1-s_0)^{-1}\{[\alpha(u_n(t))-\alpha(\tau)](s_1-\tau)\Delta_n 1_{\{u_n(t)<\tau\}}$$
$$+([\alpha(u_n(t))]^{-1}-[\alpha(\tau)]^{-1})(\tau-s_0)\Delta_n 1_{\{u_n(t)\geq\tau\}}\}.$$

Applying a Taylor expansion for α around τ, we find that

$$\begin{aligned}\tilde{X}_n(t) = {} & n\Delta_n\rho(\tau)(s_1-s_0)^{-1}\left\{\left[\frac{1}{2}\rho^{-1}(\tau)\frac{s_1-s_0}{s_1-\tau}(u_n(t)-\tau)\right.\right.\\ & \left.+O(a_n^{-2})(u_n(t)-\tau)^2\right](s_1-\tau)\Delta_n 1_{\{u_n(t)<\tau\}}\\ & \left.+\left[-\frac{1}{2}\rho^{-1}(\tau)\frac{s_1-s_0}{\tau-s_0}(u_n(t)-\tau)+O(a_n^{-2})(u_n(t)-\tau)^2\right](\tau-s_0)\Delta_n 1_{\{u_n(t)\geq\tau\}}\right\}\\ = {} & -\frac{1}{2}n\Delta_n^2|u_n(t)-\tau| + O(n\gamma_n^{-2}/a_n)(u_n(t)-\tau)^2,\end{aligned}$$

where the O-terms are uniform in t. Therefore,

$$E\tilde{X}_n(t) = -\tfrac{1}{2}|t| + o(1)$$

by (A4), which implies (5.3). Moreover,

$$\tilde{X}_n(t) = E\tilde{X}_n(t) + o(1),$$

which together with (5.1) and (5.5) implies (5.4). □

Proof of Theorem 2.2. By (5.4) and Donsker's theorem,

$$\bar{X}_n(t) - E\bar{X}_n(t) = \Delta_n S_n(u_n(t),\tau) + o(1) \Rightarrow \sigma W(t), \quad t \in [-M,M].$$

Since by (5.3), $X_n(t) \Rightarrow \sigma W(t) - \frac{1}{2}|t|$, and as X_n and $\bar{X}_n$ have the same limits according to (5.2), this implies

$$X_n(t) \Rightarrow \sigma W(t) - \tfrac{1}{2}|t|, \quad t \in [-M, M].$$

Furthermore, we have the following facts, regarding processes $Z_n(\cdot)$ (2.6):

$$Z_n(t) = \left|X_n(t) + n\Delta_n\rho(\tau)(s_1 - s_0)^{-1}T_n(\tau)\right| - \left|n\Delta_n\rho(\tau)(s_1 - s_0)^{-1}T_n(\tau)\right|, \tag{5.6}$$

$$n\Delta_n\rho(\tau)(s_1 - s_0)^{-1}T_n(\tau) = n\Delta_n^2\rho^2(\tau)(s_1 - s_0)^{-1}(1 + o_p(1)) \xrightarrow{\mathrm{P}} +\infty, \tag{5.7}$$

$X_n(t)$ is stochastically bounded. (5.8)

The result then follows from (5.6)–(5.8) and (5.3). □

Acknowledgements

This research was supported in part by NSA grant MDA 904-96-1-0026 and NSF grants DMS-93-05484 and DMS-96-25984. We thank Uli Stadtmüller, other colleagues and a referee for helpful and constructive comments.

References

Bhattacharya, P.K. and P.J. Brockwell (1976), The minimum of an additive process with applications to signal estimation and storage theory, *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **37**, 51–75.

Billingsley, P. (1968), *Convergence of Probability Measures* (Wiley, New York).

Carlstein, E. (1988), Nonparametric estimation of a change-point, *Ann. Statist.* **16**, 188–197.

Darkhovski, B.S. (1976), A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables, *Theory Probab. Appl.* **21**, 178–183.

Dümbgen, L. (1991), The asymptotic behavior of some nonparametric change-point estimators, *Ann. Statist.* **19**, 1471–1495.

Eubank, R.L. and P. Speckman (1994), Nonparametric estimation of functions with jump discontinuities, in: E. Carlstein, H.G. Müller and D. Siegmund, eds., *Change-point Problems*. IMS Lecture Notes and Monograph Series, Vol. 23, pp. 130–144.

Fan, J.Q., N. Heckman and M. Wand (1995), Local polynomial kernel regression for generalized linear models and quasi-likelihood functions, *J. Amer. Statist. Assoc.* **90**, 141–150.

McDonald, J.A. and A.B. Owen (1986), Smoothing with split linear fits, *Technometrics* **28**, 195–208.

Müller, H.G. (1992), Change-points in nonparametric regression analysis, *Ann. Statist.* **20**, 737–761.

Müller, H.G. (1993), Discussion, *Statist. Sci.* **8**, 134–139.

Müller, H.G. and U. Stadtmüller (1987), Estimation of heteroscedasticity in regression analysis, *Ann. Statist.* **15**, 610–625.

Wu, J.S. and C.K. Chu (1993), Kernel-type estimators of jump points and values of a regression function, *Ann. Statist.* **21**, 1545–1566.