Aust. N. Z. J. Stat. **46**(3), 2004, 425–441

DETECTION OF A CHANGE POINT WITH LOCAL POLYNOMIAL FITS FOR THE RANDOM DESIGN CASE

J. HUH[1] AND B.U. PARK[2*]

Duksung Women's University and Seoul National University

Summary

Regression functions may have a change or discontinuity point in the νth derivative function at an unknown location. This paper considers a method of estimating the location and the jump size of the change point based on the local polynomial fits with one-sided kernels when the design points are random. It shows that the estimator of the location of the change point achieves the rate $n^{-1/(2\nu+1)}$ when ν is even. On the other hand, when ν is odd, it converges faster than the rate $n^{-1/(2\nu+1)}$ due to a property of one-sided kernels. Computer simulation demonstrates the improved performance of the method over the existing ones.

Key words: discontinuity point; non-parametric regression; one-sided kernel; rate of convergence; two-sided Brownian motion; weak convergence.

1. Introduction

Most non-parametric regression techniques have been developed to estimate a smooth regression function without the need for parametric assumptions on the regression function. The estimators are usually smooth as well, and their rates of convergence depend on the smoothness of the underlying regression function. In practice, however, we are often interested in estimating a regression function which has some change points in itself or in its derivatives. The usual non-parametric approaches suffer from poor practical and theoretical performance in such situations.

In recent work with the kernel-based approach, Müller (1992) gave weakly consistent estimators for the location and the corresponding jump size of a change point in the νth derivative of the regression function, and provided the rate of the global L^p convergence for a kernel regression estimator adjusted at the location of the change point. For the case $\nu = 0$, Loader (1996) proposed a change point estimator, based on the local polynomial fits, that attains the n^{-1} rate. It is assumed in her paper that the errors are normal. Müller & Song (1997) and Gijbels, Hall & Kneip (1999) also suggested two-step kernel type estimators achieving the rate n^{-1}. Gijbels & Goderniaux (2004) generalized the procedure of Gijbels *et al.* for discontinuous νth derivatives and proposed a method of bandwidth selection. Horváth & Kokoszka (2002) gave a test statistic, based on local polynomial fits, for testing existence of a change point in the νth derivative in the fixed design case. Grégoire & Hamrouni (2002) established the n^{-1} rate of convergence for the change point estimator of a regression function based on local linear fits in the random design case. They conjectured that the rate $n^{-1/(2\nu+1)}$

Received March 2003; revised July 2003; accepted August 2003.
*Author to whom correspondence should be addressed.
[1] Dept of Statistics, Duksung Women's University, Seoul 132-714, Korea.
[2] Dept of Statistics, Seoul National University, Seoul 151-747, Korea. e-mail: bupark@stats.snu.ac.kr
Acknowledgments. This work was supported by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University.

would be achieved for a change point in the νth derivative. All these investigations, except Grégoire & Hamrouni (2002), focused on regression models with fixed designs or the case where $\nu = 0$.

In this paper, we extend Loader's work to the case where a change point occurs in the νth ($\nu \geq 0$) derivative, where the design points are random, and errors have an arbitrary distribution. We consider an estimator of the change point in the νth derivative using the local polynomial methods. As in Loader (1996), we use a one-sided kernel which is supported on the positive half-line and has a non-zero value at the left end of the support. The latter property of the kernel leads to a faster rate of convergence in comparison with Müller (1992). In fact, the estimator for the location of the change point achieves the rate $n^{-1/(2\nu+1)}$ when ν is even, and a different, faster, rate when ν is odd. These rates of convergence are superior to those of Müller, and we show that they are achieved under a weaker assumption of smoothness on the regression function. In the simulation study, we observe that the proposed estimator outperforms Müller's estimator in small sample cases too.

There are some related works in the change point detection problem. Raimondo (1998) provided a minimax optimal rate for a class of regression functions. Wang (1995) and Raimondo (1998) followed wavelet coefficient approaches for detecting change points. Koo (1997) used the linear splines to estimate discontinuous regression functions. Yin (1988), Wu & Chu (1993), Qiu (1994) and Braun & Müller (1998) considered multiple change points detection problems. Especially, Wu & Chu (1993) suggested using a series of tests to determine the number of change points. Qiu (1994) proposed an almost surely consistent estimator of the number of change points. McDonald & Owen (1986), Hall & Titterington (1992), and Qiu & Yandell (1998) introduced smoothing algorithms to detect change points and calculate the regression estimates. Other related results can be found in Müller & Wang (1990), Müller (1993), Carlstein, Müller & Siegmund (1994), Eubank & Speckman (1994) and Jose & Ismail (1999).

This paper is organized as follows. Section 2 defines the estimators for the location of the change point and for the corresponding jump size. Section 3 describes their asymptotic properties. Section 4 investigates the finite sample performances of the methods through several simulated examples. Section 5 gives the proofs of the asymptotic results.

2. Change point model and the estimators

The random design regression model arises when we observe a bivariate sample $\{(X_i, Y_i): i = 1, 2, \dots, n\}$ of (X, Y). Let $m(x) = \mathrm{E}(Y \mid X = x)$ denote the regression function. We denote by $v(x)$ the conditional variance of Y given $X = x$, and by f the design density of X. In this case, the regression model can be written as

$$Y_i = m(X_i) + v(X_i)^{1/2}\varepsilon_i \qquad (i = 1, \dots, n), \tag{1}$$

where, conditional on $X_1, \dots, X_n$, the ε_i are independent random variables with mean 0 and variance 1.

We assume that a change point exists for the νth derivative of m, denoted by $m^{(\nu)}$, at some point τ in the interior of the support of f, as given in Assumption A1.

Assumption A1. *There exists a constant L such that*

$$|m^{(\nu)}(x) - m^{(\nu)}(y)| \leq L|x - y| \quad \textit{whenever } (x - \tau)(y - \tau) > 0,$$

i.e. $m^{(\nu)}$ *satisfies the Lipschitz condition of order 1 over* $[0, \tau)$ *and* $(\tau, 1]$. *The jump size at the change point* τ *in the* ν*th derivative of* m *is given by* $\Delta_\nu = m_+^{(\nu)}(\tau) - m_-^{(\nu)}(\tau)$ *where* $m_+^{(\nu)}(\tau) = \lim_{x\to\tau+} m^{(\nu)}(x)$, $m_-^{(\nu)}(\tau) = \lim_{x\to\tau-} m^{(\nu)}(x)$. *We assume* $0 < |\Delta_\nu| < \infty$.

Let us assume the following additional conditions on the design density f and the conditional variance function v.

Assumption A2. *The function* f *is supported on* $[0, 1]$, *and satisfies the Lipschitz condition of order 1 over* $[0, 1]$, *and* $\inf_{x\in[0,1]} f(x) > 0$.

Assumption A3. *The function* v *satisfies the Lipschitz condition of order 1 over* $[0, 1]$.

Let p be a positive integer satisfying $p \geq \nu$. Define $\hat{m}_+^{(\nu)}(x) = \nu!\hat{\alpha}_\nu^+$ as the right-side estimator for $m^{(\nu)}(x)$, where the $(p+1)\times 1$ column vector $\hat{\boldsymbol{\alpha}}^+ = (\hat{\alpha}_0^+, \hat{\alpha}_1^+, \ldots, \hat{\alpha}_p^+)$ minimizes the right-side kernel weighted local least squares:

$$\sum_{j=1}^{n} \Big(Y_j - \textstyle\sum_{\ell=0}^{p}\alpha_\ell (X_j - x)^\ell\Big)^2 K\Big(\frac{X_j - x}{h}\Big) . \tag{2}$$

Likewise, define the left-side estimator $\hat{m}_-^{(\nu)}(x)$ by $\nu!\hat{\alpha}_\nu^-$, where the $(p+1)\times 1$ vector $\hat{\boldsymbol{\alpha}}^- = (\hat{\alpha}_0^-, \hat{\alpha}_1^-, \ldots, \hat{\alpha}_p^-)$ minimizes (2) with $K(h^{-1}(X_j - x))$ being replaced by $K(h^{-1}(x - X_j))$. Here, K is a kernel function with support $[0, 1]$ and $h = h_n$ is a sequence of bandwidths, which satisfy the following assumptions.

Assumption A4. *The function* K *satisfies* $\int_0^1 K(u)\,du = 1$, *and* $K(0) > 0, K(u) \geq 0$ *for* $0 < u \leq 1$, *and its first derivative* K' *satisfies the Lipschitz condition of order 1.*

Assumption A5. $h \to 0$, $nh^{2\nu+1}/\log n \to \infty$, *and* $nh^{2\nu+3} \to 0$, *as* $n \to \infty$.

The estimators $\hat{m}_+^{(\nu)}(x)$ and $\hat{m}_-^{(\nu)}(x)$ are based on the one-sided data at the right and the left of x, respectively. We estimate the jump size at a point x by taking the differences of these two estimators: $\hat{\Delta}_\nu(x) = \hat{m}_+^{(\nu)}(x) - \hat{m}_-^{(\nu)}(x)$. A reasonable estimator $\hat{\tau}$ of τ is the value of x that maximizes $|\hat{\Delta}_\nu(x)|$. Let $Q \subset (0, 1)$ denote a closed interval such that $\tau \in Q$. Define

$$\hat{\tau} = \inf\Big\{z \in Q\colon \big|\hat{\Delta}_\nu(z)\big| = \sup_{x\in Q}\big|\hat{\Delta}_\nu(x)\big|\Big\}$$

for the location of the change point τ. We call this the local polynomial change point (LPCP) estimator. An estimator of the jump size Δ_ν can be obtained by

$$\hat{\Delta}_\nu(\hat{\tau}) = \hat{m}_+^{(\nu)}(\hat{\tau}) - \hat{m}_-^{(\nu)}(\hat{\tau}) . \tag{3}$$

3. Asymptotic properties

Let $\boldsymbol{X}$ denote an $n \times (p+1)$ matrix with (i, j)th element equal to $((X_i - x)/h)^{j-1}$, and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$. The first column vector of $\boldsymbol{X}$ defined above is $(1, 1, \ldots, 1)$. Let $\boldsymbol{W}^+$ be an $n \times n$ diagonal matrix having $K((X_i - x)/h)/h$ as its diagonal elements. Define $\boldsymbol{W}^-$ likewise. Then the weighted least squares function (2) can be written as

$$(\boldsymbol{Y} - \boldsymbol{X\beta})^{\mathsf{T}}\boldsymbol{W}^+(\boldsymbol{Y} - \boldsymbol{X\beta}) , \tag{4}$$

where $\boldsymbol{\beta} = (\alpha_0, h\alpha_1, \ldots, h^p\alpha_p)$. Define the minimizer of (4) by $\hat{\boldsymbol{\beta}}^+ = (\hat{\beta}_0^+, \ldots, \hat{\beta}_p^+)$. Assuming invertibility of $\boldsymbol{X}^\top \boldsymbol{W}^+ \boldsymbol{X}$, the standard weighted least squares theory leads to the solution $\hat{\boldsymbol{\beta}}^+ = (\boldsymbol{X}^\top \boldsymbol{W}^+ \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^+ \boldsymbol{Y}$. Analogously, $\hat{\boldsymbol{\beta}}^- = (\boldsymbol{X}^\top \boldsymbol{W}^- \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^- \boldsymbol{Y}$. Then, the estimators $\hat{\beta}_\ell^+$ and $\hat{\beta}_\ell^-$ can be written as

$$\hat{\beta}_\ell^+ = h^\ell \hat{\alpha}_\ell^+, \qquad \hat{\beta}_\ell^- = h^\ell \hat{\alpha}_\ell^- \qquad (\ell = 0, \ldots, p).$$

Let $\boldsymbol{S}_n = n^{-1} \boldsymbol{X}^\top \boldsymbol{W}^+ \boldsymbol{X}$ and $\boldsymbol{T}_n = n^{-1} \boldsymbol{X}^\top \boldsymbol{W}^- \boldsymbol{X}$. The (j, ℓ)th entries of $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$ are $\sum_{i=1}^n ((X_i - x)/h)^{j+\ell-2} K((X_i - x)/h)/(nh)$, $\sum_{i=1}^n ((X_i - x)/h)^{j+\ell-2} K((x - X_i)/h)/(nh)$, respectively. Define $\boldsymbol{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 appearing at the ith position, and

$$W_\nu^+(x, u) = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}_n^{-1}(1, u, \ldots, u^p) K(u), \quad W_\nu^-(x, u) = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{T}_n^{-1}(1, -u, \ldots, (-u)^p) K(u).$$

Here, we write $W_\nu^\pm(x, u)$ rather than just $W_\nu^\pm(u)$ to stress their dependence on x through $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$. Then, the estimators $\hat{m}_+^{(\nu)}(x)$ and $\hat{m}_-^{(\nu)}(x)$ can be written as

$$\hat{m}_+^{(\nu)}(x) = \frac{\nu!}{nh^{\nu+1}} \sum_{j=1}^n W_\nu^+\Big(x, \frac{X_j - x}{h}\Big) Y_j \quad \text{and} \quad \hat{m}_-^{(\nu)}(x) = \frac{\nu!}{nh^{\nu+1}} \sum_{j=1}^n W_\nu^-\Big(x, \frac{x - X_j}{h}\Big) Y_j.$$

To derive the asymptotic expressions for $\hat{m}_+^{(\nu)}(x)$ and $\hat{m}_-^{(\nu)}(x)$, let $\boldsymbol{S}$ and $\boldsymbol{T}$ denote $(p+1) \times (p+1)$ matrices having their (i, j)th entries equal to $\int_0^1 K(u) u^{i+j-2}\, du$ and $\int_0^1 K(u)(-u)^{i+j-2}\, du$, respectively. Note that

$$(\boldsymbol{S}_n)_{ij} = \big(\mathrm{E}(\boldsymbol{S}_n)\big)_{ij} + O_p\Big(\Big(\frac{\log(1/h)}{nh}\Big)^{1/2}\Big) = f(x)(\boldsymbol{S})_{ij} + O(h) + O_p\Big(\Big(\frac{\log(1/h)}{nh}\Big)^{1/2}\Big)$$

uniformly in $x \in Q$ by Assumption A2, from which we obtain $\boldsymbol{S}_n = f(x)\boldsymbol{S}(1 + o_p(1))$ uniformly in $x \in Q$ by Assumption A5. Let $K_\nu^+(u) = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}^{-1}(1, u, \ldots, u^p) K(u)$. Then

$$W_\nu^+(x, u) = \frac{1}{f(x)} K_\nu^+(u)\big(1 + o_p(1)\big). \tag{5}$$

Similarly, letting $K_\nu^-(u) = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{T}^{-1}(1, -u, \ldots, (-u)^p) K(u)$, we have

$$W_\nu^-(x, u) = \frac{1}{f(x)} K_\nu^-(u)\big(1 + o_p(1)\big). \tag{6}$$

In (5) and (6), the $o_p(1)$ terms are uniform in $x \in Q$ and in $u \in [0, 1]$. The functions K_ν^+ and K_ν^- are the so-called equivalent kernels discussed in Müller (1987). They satisfy the following moment conditions (see e.g. Fan & Gijbels, 1996 p. 103):

$$\int u^\ell K_\nu^+(u) = \delta_{\nu\ell} \quad \text{and} \quad \int (-u)^\ell K_\nu^-(u) = \delta_{\nu\ell} \qquad (0 \le \nu, \ell \le p),$$

where δ is the Kronecker delta function. Assume that the equivalent kernel K_ν^- satisfies Assumption A6.

Assumption A6. $K_\nu^-(0) > 0$.

First, in Theorem 1 below we describe weak convergence of the sequence of the process $\{\varphi_{n\nu}(z)\colon -M < z < M\}$ where

$$\varphi_{n\nu}(z) = \begin{cases} (nh^{2\nu+1})^{(\nu+1)/(2\nu+1)}\Big(\hat{\Delta}_\nu\Big(\tau + \dfrac{h}{(nh^{2\nu+1})^{1/(2\nu+1)}}z\Big) - \hat{\Delta}_\nu(\tau)\Big) & \text{when } \nu \text{ is even,} \\ (nh^{2\nu+1})^{(\nu+1)/2\nu}\Big(\hat{\Delta}_\nu\Big(\tau + \dfrac{h}{(nh^{2\nu+1})^{1/2\nu}}z\Big) - \hat{\Delta}_\nu(\tau)\Big) & \text{when } \nu \text{ is odd,} \end{cases} \tag{7}$$

and $M < \infty$. Existence of the unique maximizer (minimizer) of the limit of the process $\varphi_{n\nu}$ when $\Delta_\nu > 0$ ($\Delta_\nu < 0$) is discussed later on. The process $\varphi_{n\nu}$ lies in the space, denoted $\mathcal{D}([-M, M])$, of functions defined on $[-M, M]$ having, at most, finitely many discontinuities. To obtain the theorem, consider the following additional assumption.

Assumption A7. $\mathrm{E}(|Y|^{2+\zeta} \mid X = x) < \infty$, *for all* x *and some positive* ζ.

Let $\stackrel{\mathrm{w}}{\to}$ denote weak convergence in the space $\mathcal{D}([-M, M])$, and denote the first-order derivatives of $K_\nu^{\pm}$ by $K_\nu'^{\pm}$.

Theorem 1. *Suppose that Assumptions (A1)–(A7) are satisfied.*

(i) *If* ν *is even, then*

$$\varphi_{n\nu}(z) \stackrel{\mathrm{w}}{\to} \varphi_\nu(z) = -\frac{\Delta_\nu}{\nu+1}K_\nu^-(0)|z|^{\nu+1} + \sigma_1 W(z)\,, \tag{8}$$

where $W(z)$ *is the two-sided Brownian motion defined in Bhattacharya & Brockwell (1976), and*

$$\sigma_1 = \sqrt{\frac{4(\nu!)^2 v(\tau) + \Delta_0^2\,\mathrm{I}\,(\nu = 0)}{f(\tau)}}\,K_\nu^-(0)\,. \tag{9}$$

(ii) *If* ν *is odd, then*

$$\varphi_{n\nu}(z) \stackrel{\mathrm{w}}{\to} \varphi_\nu(z) = -\frac{\Delta_\nu}{\nu+1}K_\nu^-(0)|z|^{\nu+1} + zU\,, \tag{10}$$

where $U \stackrel{\mathrm{d}}{=} \mathrm{N}(0, \sigma_2^2)$ *and*

$$\sigma_2^2 = 2(\nu!)^2\frac{v(\tau)}{f(\tau)}\int_0^1 K_\nu'^-(u)^2\,du\,. \tag{11}$$

Remark 1. When the conditional variance function v depends on m, it is likely that v also has a change point at τ. In that case, the asymptotic variance parts of the limit processes φ_ν in (8) and (10) are slightly changed. Define $v_+(\tau) = \lim_{x\to\tau+} v(x)$ and $v_-(\tau) = \lim_{x\to\tau-} v(x)$. Then, $v(\tau)$ in (9) is replaced by $v_+(\tau)$ when $z \geq 0$, and by $v_-(\tau)$ when $z < 0$. On the other hand, $2v(\tau)$ in (11) is replaced by $v_+(\tau) + v_-(\tau)$.

Remark 2. The asymptotic variance of $\varphi_{n\nu}$ given at (9) and (11) depends on the design density f as well as the variance function v. When $\nu = 0$, it involves the extra term Δ_0^2 too. We note that the asymptotic variances of Müller (1992) and Loader (1996) depend only on v (see Müller, 1992 Theorem 3.1 and Loader, 1996 Theorem 1) since they are for the fixed uniform design case.

Remark 3. Since $T = ESE$ where $E = \mathrm{diag}((-1)^0, (-1)^1, \ldots, (-1)^p)$, the equivalent kernel K_ν^- can be rewritten as $K_\nu^-(u) = (-1)^\nu e_{\nu+1}^{\mathrm{T}} S^{-1}(1, u, \ldots, u^p)K(u)$. It follows that

$$K_\nu^+(u) = (-1)^\nu K_\nu^-(u) \qquad (0 \le u \le 1)\,. \tag{12}$$

In fact, this relation makes it possible to state the theorems in terms of K_ν^- and $K_\nu'^-$ only, and has an important consequence as we discuss below.

Next, we describe the asymptotic distribution of $\hat\tau$. For this, we observe that the maximizer of the limit process φ_ν exists and is unique with probability one. In the case where $\Delta_\nu > 0$ and ν is even, this follows directly from Bhattacharya & Brockwell (1976 Remark 5.3), where it is argued that the maximizer of the two-sided Brownian motion with an additional drift is unique with probability 1. Now, when $\Delta_\nu > 0$ and ν is odd, the Gaussian process φ_ν in (10) has a unique maximum at

$$Z_\nu = \Big(\frac{U}{\Delta_\nu K_\nu^-(0)}\Big)^{1/\nu}\,. \tag{13}$$

The other case where $\Delta_\nu < 0$ is analogous.

Theorem 2 describes the asymptotic distribution of $\hat\tau$.

Theorem 2. *Suppose that the assumptions in Theorem 1 are satisfied.*

(i) *If ν is even, then*

$$n^{1/(2\nu+1)}(\hat\tau - \tau) \xrightarrow{\mathrm{d}} \begin{cases} \operatorname*{argmax}_{z\in(-\infty,\infty)} \varphi_\nu(z) & \text{when } \Delta_\nu > 0, \\ \operatorname*{argmin}_{z\in(-\infty,\infty)} \varphi_\nu(z) & \text{when } \Delta_\nu < 0. \end{cases}$$

(ii) *If ν is odd, then*

$$\sqrt{nh}(\hat\tau - \tau)^\nu \xrightarrow{\mathrm{d}} \mathrm{N}\Big(0, \frac{\sigma_2^2}{\Delta_\nu^2 K_\nu^-(0)^2}\Big)\,.$$

According to Theorem 2, the rate of convergence of $\hat\tau$ decays rapidly as ν increases. The rate differs for ν even and ν odd. When ν is even, the LPCP estimator achieves the rate $n^{-1/(2\nu+1)}$. When ν is odd, it is $(1/nh)^{1/2\nu}$. If we choose the bandwidth $h = O(n^{-1/(2\nu+2)})$, which satisfies Assumption A5, $\hat\tau$ achieves the rate $n^{-(2\nu+1)/(2\nu(2\nu+2))}$. For instance, a change point of the first derivative can be estimated at the rate $O_p(n^{-3/8})$ rather than $O_p(n^{-1/3})$. This faster rate $n^{-(2\nu+1)/(2\nu(2\nu+2))}$ for odd ν comes from the property (12) of the equivalent kernels K_ν^+ and K_ν^-. In the following paragraph we elaborate this phenomenon further.

Consider a process $\chi_{n\nu}(\cdot, b_{n\nu}, c_{n\nu}) = c_{n\nu}(\hat\Delta_\nu(\tau + b_{n\nu}\cdot) - \hat\Delta_\nu(\tau))$ for some sequences $b_{n\nu}$ and $c_{n\nu}$ converging to 0 and infinity, respectively, as n tends to infinity. Note that $\chi_{n\nu}$ equals the process $\varphi_{n\nu}$ defined at (7) with specific choices of $b_{n\nu}$ and $c_{n\nu}$. The rate of convergence of the change point estimator $\hat\tau$ is determined by $b_{n\nu}$ for which the process $\chi_{n\nu}$ has a proper limit. We see later, following the proof of Lemma 1 in Section 5, that

$$\mathrm{E}\big(\chi_{n\nu}(z, b_{n\nu}, c_{n\nu})\big) = c_1(z) b_{n\nu}^{\nu+1} c_{n\nu} h^{-\nu-1}\,.$$

Also, we show, following the proof of Lemma 2 in Section 5, that the asymptotic covariance of $\chi_{n\nu}(z_1, b_{n\nu}, c_{n\nu})$ and $\chi_{n\nu}(z_2, b_{n\nu}, c_{n\nu})$ for different z_1 and z_2 equals

$$c_2(z_1, z_2) b_{n\nu} c_{n\nu}^2 n^{-1} h^{-2\nu-2}\big(K_\nu^+(0) + K_\nu^-(0)\big)^2 + O(b_{n\nu}^2 c_{n\nu}^2 n^{-1} h^{-2\nu-3})\,. \tag{14}$$

In the case where ν is odd, the first leading term on the right-hand side of (14) vanishes since $K_\nu^+(0) + K_\nu^-(0) = 0$ by (12). For $\chi_{n\nu}$ to have a stable limit process, $b_{n\nu}$ needs to satisfy $b_{n\nu}^{\nu+1} c_{n\nu} h^{-\nu-1} \sim 1$ and $b_{n\nu}^2 c_{n\nu}^2 n^{-1} h^{-2\nu-3} \sim 1$. These considerations yield the rate $b_{n\nu} = (1/nh)^{1/2\nu}$ when ν is odd. In the case where ν is even, the first term on the right-hand side of (14) does not vanish, and thus the relatively slower rate $b_{n\nu} = n^{-1/(2\nu+1)}$ is obtained from $b_{n\nu}^{\nu+1} c_{n\nu} h^{-\nu-1} \sim 1$ and $b_{n\nu} c_{n\nu}^2 n^{-1} h^{-2\nu-2} \sim 1$.

Remark 4. The asymptotic variance of $\hat{\tau}$ depends on $f(\tau)$. Theorem 2 shows that the change point estimator gets more stable as the density at the change point increases.

Remark 5. For all ν, the rates for the LPCP estimators are faster than those for the Müller (1992) estimators under the weaker smoothness assumption on the regression function as given at Assumption A1. As in Loader (1996), this property of faster convergence is due to Assumption A6 on the one-sided kernel function. Assumptions (M2) and (K1) of Müller (1992) require the existence of a function g which is at least $2\nu+1$ times continuously differentiable and that $m^{(\nu)}(x) = g^{(\nu)}(x) + \Delta_\nu \, \mathrm{I}(\tau \le x \le 1)$, $0 \le x \le 1$. For example, Müller (1992 Corollary 3.1) shows that his estimator may achieve the rates $(b/n)^{1/2}$ and $(b^3/n)^{1/6}$ for $\nu = 0$ and 1, respectively, when a bandwidth b is chosen to satisfy the assumption (B1) in that paper. In particular, if the bandwidth $b \sim (\log n/n) n^{1/(2+\zeta)}$ for ζ which appears in Assumption A7, the achieved rate for $\nu = 0$ equals $n^{-1}(\log n)^{1/2} n^{1/(2(2+\zeta))}$, which gets arbitrarily close to $n^{-1}(\log n)^{1/2}$ as ζ tends to infinity.

As another consequence of Theorem 1, Corollary 1 describes the asymptotic distribution of the estimator $\hat{\Delta}_\nu(\hat{\tau})$ for the jump size defined at (3).

Corollary 1. *Under the assumptions of Theorem 1,*

$$\sqrt{nh^{2\nu+1}}\,\big(\hat{\Delta}_\nu(\hat{\tau}) - \Delta_\nu\big) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0,\, 2(\nu!)^2 \frac{v(\tau)}{f(\tau)} \int_0^1 K_\nu^-(u)^2\, du\right).$$

If v has a change point at τ, the asymptotic variance in Corollary 1 is replaced by $(\nu!)^2 f(\tau)^{-1}(v_+(\tau) + v_-(\tau)) \int_0^1 K_\nu^-(u)^2\, du$.

4. Numerical experiments

We compare the LPCP estimator $\hat{\tau}$ with the one given by Müller (1992), the latter being based on the Gasser–Müller-type one-sided regression estimators

$$\tilde{m}_\pm^{(\nu)}(x) = \frac{1}{h^{\nu+1}} \sum_{j=1}^n Y_{[j]} \int_{s_{j-1}}^{s_j} M_\nu^\pm\Big(\frac{x-u}{h}\Big)\, du\,,$$

where M_ν^+ and M_ν^- are one-sided kernel functions with supports $[-1, 0]$ and $[0, 1]$ respectively, which satisfy $M_\nu^+(x) = (-1)^\nu M_\nu^-(-x)$ and $M_\nu^-(0) = 0$. Here $s_j = \frac{1}{2}(X_{(j)} + X_{(j+1)})$, $s_0 = 0$, $s_n = 1$ and $(X_{(j)}, Y_{[j]})$, $j = 1, \ldots, n$, denote the (X_j, Y_j) ordered with respect to the X_j values.

We consider the uniform fixed (UF) design, and the three random design densities of X:

$$f_1(x) = \mathrm{I}(0 \le x \le 1)\,,$$
$$f_2(x) = p_1 \varphi\Big(\frac{x-0.5}{0.40}\Big)\, \mathrm{I}(0 \le x \le 1)\,,$$
$$f_3(x) = p_2\Big(\varphi\Big(\frac{x-0.25}{0.15}\Big) + \varphi\Big(\frac{x-0.75}{0.15}\Big)\Big)\, \mathrm{I}(0 \le x \le 1)\,,$$

where φ denotes the standard normal density function, and the p_i are the normalizing constants to make f_i proper densities. The densities f_2 and f_3 are unimodal and bimodal, respectively. The errors ε_i in (1) are assumed to be normally distributed.

For the regression function, we consider the function m_1 in Nason & Silverman (1994) with a change point at $\tau = 0.5$. We take the jump size $\Delta_0 = -0.5$. Thus, m_1 is given by

$$m_1(x) = \begin{cases} 4x^2(3-4x) & 0 \le x < \frac{1}{2}, \\ \frac{4}{3}x(4x^2 - 10x + 7) - \frac{3}{2} & \frac{1}{2} \le x < \frac{3}{4}, \\ \frac{16}{3}x(x-1)^2 & \frac{3}{4} \le x \le 1. \end{cases}$$

The conditional variance $v(x)$ is 0.25 for all x.

Our second example concerns a regression function which has a change point in the first derivative. The regression function m_2 is given by

$$m_2(x) = \begin{cases} \frac{1}{2}\sin(4\pi x) + \frac{1}{2} & 0 \le x < \frac{1}{2}, \\ -\frac{1}{2}\sin(4\pi x) + \frac{1}{2} & \frac{1}{2} \le x \le 1. \end{cases}$$

We choose $v(x) = 0.10$ for all x, and the jump size of the first derivative is $\Delta_1 = -4\pi$ at the location of the change point $\tau = 0.5$.

We consider the one-sided kernel function $K(x) = \frac{3}{2}(1-x^2)\,\mathrm{I}\,(0 \le x \le 1)$ and the degree of the local polynomial $p = 1$ for the LPCP estimators. The kernel functions $M_0^-(x) = 12x(1-x)(3-5x)\,\mathrm{I}\,(0 \le x \le 1)$ and $M_1^-(x) = 60x(1-x)(1-2x)\,\mathrm{I}\,(0 \le x \le 1)$ in Müller (1992) are chosen for the Müller estimator in the models m_1 and m_2, respectively. These kernels vanish at the left end of the support.

Table 1 contains the results of the simulations based on 1000 pseudo samples of size 1000. To estimate the location of the change point, we first compute the jump sizes at $x_k = k/500$, $k = 1, \ldots, 500$, and then choose a point which maximizes the absolute value of the calculated jump sizes over the interval Q. As suggested in Müller (1992), we take $Q = [h, 1-h]$ for our simulation settings. We compute the Monte Carlo estimates of the root mean squared errors (RMSE) of the LPCP estimator and the one by Müller (1992), the latter being denoted by M, for various values of bandwidth h. We report here the minimum RMSEs of $\hat{\tau}$ over h and the minimizing bandwidths $h_{\hat{\tau}}$. The table also includes the averages of $\hat{\tau}$ and $\hat{\Delta}$, and the RMSEs of $\hat{\Delta}$ when $h_{\hat{\tau}}$ are used. Standard errors of these Monte Carlo estimates are given in the parentheses. The standard errors of the RMSEs are obtained by the formula

$$(\text{SE of RMSE}) = \frac{1}{2}\left(\sqrt{\text{RMSE}^2 + (\text{SE of RMSE}^2)} - \sqrt{\text{RMSE}^2 - (\text{SE of RMSE}^2)}\,\right),$$

where the standard error of RMSE^2 is calculated in the usual way; e.g. the standard error of RMSE^2 of $\hat{\tau}$ based on 1000 Monte Carlo samples equals $\frac{1}{1000}\sqrt{\sum_{i=1}^{1000}((\hat{\tau}_i - \tau)^2 - \text{RMSE}^2)^2}$. The table shows that, in every case, the LPCP estimator has less RMSE than the estimator M in terms of estimating τ. The LPCP outperforms M with the ratios of the RMSEs roughly between $\frac{1}{4}$ and $\frac{1}{2}$ for m_1 and for every design density. For m_2, these ratios are much smaller than for m_1. The ratios lie between $\frac{1}{6}$ and $\frac{1}{3}$ approximately.

As mentioned in Remark 4, the density at the change point τ has a crucial influence on the performance of the LPCP estimators. In our eight simulation models, the RMSEs of $\hat{\tau}$ for the

TABLE 1

The minimum RMSEs of $\hat{\tau}$ over h and the minimizing bandwidths $h_{\hat{\tau}}$. Also included are the averages of $\hat{\tau}$ and $\hat{\Delta}$, and RMSEs of $\hat{\Delta}$ when $h_{\hat{\tau}}$ are used, with the standard errors given in parentheses.

Case	$h_{\hat{\tau}}$	Estimator	Average of $\hat{\tau}$	RMSE of $\hat{\tau}$	Average of $\hat{\Delta}$	RMSE of $\hat{\Delta}$
(m_1, UF)	0.15	LPCP	0.500406	0.001586	–0.500454	0.059716
			(0.000048)	(0.000084)	(0.001888)	(0.001281)
	0.10	M	0.500660	0.003237	–0.522328	0.086267
			(0.000100)	(0.000089)	(0.002635)	(0.001913)
(m_1, f_1)	0.11	LPCP	0.500230	0.001847	–0.508728	0.068673
			(0.000058)	(0.000105)	(0.002154)	(0.001500)
	0.14	M	0.500264	0.004646	–0.522517	0.084894
			(0.000147)	(0.000136)	(0.002588)	(0.001820)
(m_1, f_2)	0.12	LPCP	0.500088	0.001039	–0.506832	0.049970
			(0.000033)	(0.000070)	(0.001565)	(0.001131)
	0.16	M	0.500520	0.003722	–0.526550	0.066566
			(0.000117)	(0.000090)	(0.001930)	(0.001473)
(m_1, f_3)	0.12	LPCP	0.500464	0.002800	–0.513091	0.077498
			(0.000087)	(0.000140)	(0.002415)	(0.001691)
	0.16	M	0.500494	0.005603	–0.528391	0.091346
			(0.000176)	(0.000158)	(0.002746)	(0.001993)
(m_2, UF)	0.12	LPCP	0.500034	0.002592	–9.758507	2.846361
			(0.000082)	(0.000061)	(0.014754)	(0.014642)
	0.70	M	0.499846	0.007305	–11.645859	1.318820
			(0.000231)	(0.000123)	(0.029865)	(0.024385)
(m_2, f_1)	0.12	LPCP	0.499988	0.002583	–9.660561	2.939372
			(0.000082)	(0.000067)	(0.014006)	(0.014047)
	0.08	M	0.500552	0.010571	–11.144008	1.923343
			(0.000334)	(0.001419)	(0.040941)	(0.174435)
(m_2, f_2)	0.12	LPCP	0.500108	0.002165	–9.671326	2.921531
			(0.000068)	(0.000050)	(0.012412)	(0.012358)
	0.07	M	0.500402	0.009428	–11.601712	1.655966
			(0.000298)	(0.001052)	(0.042564)	(0.201045)
(m_2, f_3)	0.12	LPCP	0.500054	0.003224	–9.526879	3.079188
			(0.000102)	(0.000090)	(0.015585)	(0.015707)
	0.09	M	0.500014	0.018184	–10.608519	2.566069
			(0.000575)	(0.004697)	(0.052455)	(0.205019)

density f_2 are smaller than those for UF and the density f_1 because the value of the density f_2 at τ is larger than the values of UF and f_1. By the same reasoning, the RMSEs of $\hat{\tau}$ for UF and the density f_1 are superior to those for the density f_3. Comparing the simulation results for UF and f_1, we find that the RMSEs of M for UF are relatively smaller than those for f_1, while the RMSEs of LPCP for UF are similar to those for f_1. This can be expected because (see Jones, Davies & Park, 1994) for the Gasser–Müller regression estimator, the asymptotic variance in the random design case is 1.5 times larger than the asymptotic variance in the corresponding fixed design case. These two are identical for the local polynomial regression estimator.

In the case of m_2, the averages and the RMSEs of $\hat{\Delta}$ of the estimator M seem to be superior to those of LPCP, but the values reported here are for bandwidths which minimize the RMSEs of $\hat{\tau}$. We find that the minimum RMSEs of the LPCP $\hat{\Delta}$ are usually smaller than those of M. Table 2 reports the minimum RMSEs of $\hat{\Delta}$ over h with the minimizing bandwidths $h_{\hat{\Delta}}$. It also shows the averages of $\hat{\tau}$ and $\hat{\Delta}$, and RMSEs of $\hat{\tau}$ when $h_{\hat{\Delta}}$ are used. Although the RMSEs of $\hat{\tau}$ in Table 2 are for the bandwidths which minimize the RMSEs of $\hat{\Delta}$, those for LPCP are rather smaller than those for M.

TABLE 2

The minimum RMSEs of $\hat{\Delta}$ over h and the minimizing bandwidths $h_{\hat{\Delta}}$. Also included are the averages of $\hat{\tau}$ and $\hat{\Delta}$, and RMSEs of $\hat{\tau}$ when $h_{\hat{\Delta}}$ are used, with the standard errors given in parentheses.

Case	$h_{\hat{\Delta}}$	Estimator	Average of $\hat{\tau}$	RMSE of $\hat{\tau}$	Average of $\hat{\Delta}$	RMSE of $\hat{\Delta}$
(m_1, UF)	0.25	LPCP	0.500496	0.001625	–0.519848	0.051166
			(0.000049)	(0.000082)	(0.001491)	(0.001153)
	0.20	M	0.501952	0.004738	–0.538073	0.069735
			(0.000137)	(0.000119)	(0.001847)	(0.001463)
(m_1, f_1)	0.23	LPCP	0.500612	0.002147	–0.525066	0.052336
			(0.000065)	(0.000124)	(0.001453)	(0.001185)
	0.21	M	0.501472	0.005415	–0.538529	0.075928
			(0.000165)	(0.000133)	(0.002069)	(0.001653)
(m_1, f_2)	0.20	LPCP	0.500222	0.001137	–0.518813	0.044045
			(0.000035)	(0.000077)	(0.001259)	(0.001011)
	0.18	M	0.500748	0.003930	–0.530928	0.065780
			(0.000122)	(0.000107)	(0.001836)	(0.001437)
(m_1, f_3)	0.24	LPCP	0.501250	0.003247	–0.531248	0.059228
			(0.000095)	(0.000173)	(0.001591)	(0.001326)
	0.26	M	0.503438	0.006920	–0.549967	0.080100
			(0.000190)	(0.000170)	(0.001980)	(0.001661)
(m_2, UF)	0.07	LPCP	0.500194	0.003275	–11.788341	1.267834
			(0.000103)	(0.000082)	(0.031656)	(0.026466)
	0.06	M	0.500052	0.007472	–12.176618	1.210304
			(0.000236)	(0.000127)	(0.036234)	(0.025968)
(m_2, f_1)	0.06	LPCP	0.500080	0.003773	–11.955366	1.342218
			(0.000119)	(0.000096)	(0.037792)	(0.029243)
	0.06	M	0.499132	0.018497	–12.304319	1.539952
			(0.000584)	(0.005378)	(0.047987)	(0.158480)
(m_2, f_2)	0.06	LPCP	0.500050	0.003395	–11.916279	1.249040
			(0.000107)	(0.000091)	(0.033727)	(0.027588)
	0.06	M	0.499698	0.017854	–12.181929	1.460146
			(0.000565)	(0.005113)	(0.044545)	(0.169504)
(m_2, f_3)	0.07	LPCP	0.500032	0.004241	–11.660030	1.437004
			(0.000134)	(0.000102)	(0.035264)	(0.028807)
	0.07	M	0.499414	0.019632	–11.831970	1.723373
			(0.000621)	(0.005013)	(0.049302)	(0.201107)

5. Proofs

See http://stats.snu.ac.kr/~brain03/download/papers/park/cplp.ps for a more detailed proof than the one we sketch here.

Let $\bar{m}_+(x,u) = \sum_{\ell=0}^{\nu} m_+^{(\ell)}(x)(u-x)^\ell/\ell!$, $\bar{m}_-(x,u) = \sum_{\ell=0}^{\nu} m_-^{(\ell)}(x)(u-x)^\ell/\ell!$ and $\tilde{\Delta}_\nu(x) = \tilde{m}_+^{(\nu)}(x) - \tilde{m}_-^{(\nu)}(x)$ where

$$\tilde{m}_\pm^{(\nu)}(x) = \frac{\nu!}{nh^{\nu+1}} \sum_{j=1}^{n} W_\nu^\pm\Big(x, \pm\frac{X_j - x}{h}\Big)\big(Y_j - \bar{m}_\pm(x, X_j)\big)\,.$$

It follows that

$$\hat{\Delta}_\nu(x) = \begin{cases} \tilde{\Delta}_\nu(x) & x \neq \tau\,, \\ \tilde{\Delta}_\nu(\tau) + \Delta_\nu & x = \tau\,, \end{cases}$$

since the weights $W_\nu^\pm$ satisfy the following discrete moment conditions (see Fan & Gijbels, 1996 p. 103):

$$\frac{1}{nh^{\nu+1}}\sum_{j=1}^{n} W_\nu^{\pm}\Big(x,\ \pm\frac{X_j-x}{h}\Big)(X_j-x)^{\ell}=\delta_{\nu\ell}\qquad (0\le\nu,\ell\le p)\,. \tag{15}$$

This implies that

$$\varphi_{n\nu}(z)=a_{n\nu}^{\nu+1}\Big(\tilde\Delta_\nu\Big(\tau+\frac{h}{a_{n\nu}}z\Big)-\tilde\Delta_\nu(\tau)-\Delta_\nu\Big)\qquad \text{for all } z\neq 0\,,$$

where

$$a_{n\nu}=\begin{cases}(nh^{2\nu+1})^{1/(2\nu+1)} & \text{when } \nu \text{ is even,}\\ (nh^{2\nu+1})^{1/2\nu} & \text{when } \nu \text{ is odd.}\end{cases}$$

Define $z_{n\nu}=zh/a_{n\nu}$. By (15), we have the following identity:

$$\Delta_\nu=\frac{1}{nh^{\nu+1}}\sum_{j=1}^{n} W_\nu^{\pm}\Big(\tau+z_{n\nu},\ \pm\frac{X_j-\tau-z_{n\nu}}{h}\Big)(X_j-\tau-z_{n\nu})^{\nu}\Delta_\nu\,. \tag{16}$$

Let

$$\begin{aligned}
C_{n\nu}^{+}(w,u,z)&=\frac{1}{f(\tau+z_{n\nu})}K_\nu^{+}\Big(\frac{w-\tau-z_{n\nu}}{h}\Big)\Big(u-\bar m_{+}(\tau+z_{n\nu},w)\\
&\qquad-(w-\tau-z_{n\nu})^{\nu}\frac{\Delta_\nu}{\nu!}\,\mathrm{I}\,(z<0)\Big)-\frac{1}{f(\tau)}K_\nu^{+}\Big(\frac{w-\tau}{h}\Big)\big(u-\bar m_{+}(\tau,w)\big)\,,\\
C_{n\nu}^{-}(w,u,z)&=\frac{1}{f(\tau+z_{n\nu})}K_\nu^{-}\Big(\frac{\tau+z_{n\nu}-w}{h}\Big)\Big(u-\bar m_{-}(\tau+z_{n\nu},w)\\
&\qquad+(w-\tau-z_{n\nu})^{\nu}\frac{\Delta_\nu}{\nu!}\,\mathrm{I}\,(z>0)\Big)-\frac{1}{f(\tau)}K_\nu^{-}\Big(\frac{\tau-w}{h}\Big)\big(u-\bar m_{-}(\tau,w)\big)\,,\\
\phi_{n\nu}(z)&=a_{n\nu}^{\nu+1}\frac{\nu!}{nh^{\nu+1}}\sum_{j=1}^{n}\big(C_{n\nu}^{+}(X_j,Y_j,z)-C_{n\nu}^{-}(X_j,Y_j,z)\big)
\end{aligned}$$

for $z\neq 0$, and $\phi_{n\nu}(0)=0$. It follows from (5), (6) and (16) that $\varphi_{n\nu}(z)=\phi_{n\nu}(z)(1+o_p(1))$ uniformly in $z\in[-M,M]$. We need the following four lemmas to prove Theorem 1.

Lemma 1. *Suppose that Assumptions A1, A2, A4, A5 and A6 are satisfied. Then,*

$$\mathrm{E}\big(\phi_{n\nu}(z)\big)=-\frac{\Delta_\nu}{\nu+1}K_\nu^{-}(0)|z|^{\nu+1}+o(1)$$

uniformly in $z\in[-M,M]$.

Proof. We prove the lemma for $z>0$, as the other case can be dealt with similarly. By Assumptions A1, A2 and A4, it can be shown that

$$\mathrm{E}\big(C_{n\nu}^{+}(X_1,Y_1,z)\big)=O\Big(\frac{h^{\nu+2}}{a_{n\nu}}\Big) \tag{17}$$

uniformly in z. Now, by Assumption A2,

$$\begin{aligned}\mathrm{E}\big(C_{n\nu}^{-}(X_1,Y_1,z)\big) &= h\int K_\nu^{-}(u)\Big(m(\tau+z_{n\nu}-hu)-\bar{m}_{-}(\tau+z_{n\nu},\tau+z_{n\nu}-hu)\\ &\quad+(-hu)^\nu\frac{\Delta_\nu}{\nu!}-m(\tau-hu)+\bar{m}_{-}(\tau,\tau-hu)\Big)\,du\big(1+O(h)\big)\\ &\quad+h\int K_\nu^{-}(u)\Big(\frac{f(\tau+z_{n\nu}-hu)}{f(\tau+z_{n\nu})}-\frac{f(\tau-hu)}{f(\tau)}\Big)\big(m(\tau-hu)-\bar{m}_{-}(\tau,\tau-hu)\big)\,du\,,\end{aligned} \tag{18}$$

where the $O(h)$ term is uniform in $z\in[-M,M]$. The second term of (18) is $O(h^{\nu+3}/a_{n\nu})$ uniformly in z by (A1) and (A2). For the first term, we divide the interval of integration into two parts since the change point τ lies between $\tau+z_{n\nu}-h$ and $\tau+z_{n\nu}$. For $0<u\le z/a_{n\nu}$,

$$m(\tau+z_{n\nu}-hu)-\bar{m}_{-}(\tau+z_{n\nu},\tau+z_{n\nu}-hu)=\frac{m_{-}^{(\nu)}(\tau_{-}^{*})-m_{-}^{(\nu)}(\tau+z_{n\nu})}{\nu!}(-hu)^\nu\,, \tag{19}$$

where τ_{-}^{*} lies between $\tau+z_{n\nu}$ and $\tau+z_{n\nu}-hu$. However, for $z/a_{n\nu}<u<1$,

$$\begin{aligned}&m(\tau+z_{n\nu}-hu)-\bar{m}_{-}(\tau+z_{n\nu},\tau+z_{n\nu}-hu)\\ &\quad=-\frac{\Delta_\nu}{\nu!}(z_{n\nu}-hu)^\nu+\Big(\frac{m_{-}^{(\nu)}(\tau_{-}')-m_{-}^{(\nu)}(\tau)}{\nu!}-\frac{m_{+}^{(\nu)}(\tau_{+}'')-m_{+}^{(\nu)}(\tau)}{\nu!}\Big)(z_{n\nu}-hu)^\nu\,,\end{aligned} \tag{20}$$

where τ_{-}' lies between τ and $\tau+z_{n\nu}-hu$ and τ_{+}'' lies between τ and $\tau+z_{n\nu}$.

By (19) and (20), the integral of the first term in (18) equals

$$\int_0^{z/a_{n\nu}}K_\nu^{-}(u)\frac{\Delta_\nu}{\nu!}(z_{n\nu}-hu)^\nu\,du+O\Big(\frac{h^{\nu+1}}{a_{n\nu}}\Big) \tag{21}$$

uniformly in z. Combining (17) and (21), we have that

$$\mathrm{E}\big(\phi_{n\nu}(z)\big)=-a_{n\nu}^{\nu+1}\Delta_\nu\int_0^{z/a_{n\nu}}K_\nu^{-}(u)\Big(\frac{z}{a_{n\nu}}-u\Big)^\nu\,du+O(a_{n\nu}^\nu h)$$

uniformly in $z\in[0,M]$. Since $K_\nu^{-}(u)=K_\nu^{-}(0)(1+o(1))$ uniformly for $u\in[0,M/a_{n\nu}]$, the result follows.

Lemma 2. *Suppose that Assumptions A1–A6 are satisfied.*

(i) *If ν is even, then*

$$\operatorname{cov}\big(\phi_{n\nu}(z_1),\phi_{n\nu}(z_2)\big)=\begin{cases}\dfrac{4(\nu!)^2v(\tau)+\Delta_0^2\,\mathrm{I}(\nu=0)}{f(\tau)}\min(|z_1|,|z_2|)K_\nu^{-}(0)^2+o(1) & z_1z_2\ge 0,\\ o(1) & \textit{elsewhere,}\end{cases}$$

uniformly in $z_1,z_2\in[-M,M]$.

(ii) *If ν is odd, then*

$$\operatorname{cov}\big(\phi_{n\nu}(z_1),\phi_{n\nu}(z_2)\big)=2(\nu!)^2z_1z_2\frac{v(\tau)}{f(\tau)}\int_0^1K_\nu'^{-}(u)^2\,du+o(1)$$

uniformly in $z_1,z_2\in[-M,M]$.

Proof. We prove the lemma for $z_1, z_2 > 0$ first. By Lemma 1,

$$\operatorname{cov}\big(\phi_{n\nu}(z_1),\phi_{n\nu}(z_2)\big) = n(\nu!)^2 \frac{a_{n\nu}^{2(\nu+1)}}{(nh^{\nu+1})^2} \mathrm{E}\big(C^+_{n\nu}(X_1,Y_1,z_1)C^+_{n\nu}(X_1,Y_1,z_2)$$
$$- C^+_{n\nu}(X_1,Y_1,z_1)C^-_{n\nu}(X_1,Y_1,z_2) - C^-_{n\nu}(X_1,Y_1,z_1)C^+_{n\nu}(X_1,Y_1,z_2)$$
$$+ C^-_{n\nu}(X_1,Y_1,z_1)C^-_{n\nu}(X_1,Y_1,z_2)\big) + O(n^{-1})\,. \tag{22}$$

Now, define $z_{\min} = \min(z_1, z_2)$ and $z_{\max} = \max(z_1, z_2)$ and $\tau^{\min}_{n\nu} = \tau + z_{\min}h/a_{n\nu}$ and $\tau^{\max}_{n\nu} = \tau + z_{\max}h/a_{n\nu}$. Let

$$D^+_{n\nu}(w,z) = \frac{1}{f(\tau+z_{n\nu})} K^+_\nu\Big(\frac{w-\tau-z_{n\nu}}{h}\Big) - \frac{1}{f(\tau)} K^+_\nu\Big(\frac{w-\tau}{h}\Big),$$
$$D^-_{n\nu}(w,z) = \frac{1}{f(\tau+z_{n\nu})} K^-_\nu\Big(\frac{\tau+z_{n\nu}-w}{h}\Big) - \frac{1}{f(\tau)} K^-_\nu\Big(\frac{\tau-w}{h}\Big).$$

Consider the first term in the brackets in (22) first. By the definition of $\bar m_+$, we have

$$\sup_{u\in[x,x+h]} |m(u) - \bar m_+(x,u)| \le (\mathrm{const.})h^{\nu+1} \tag{23}$$

for $x = \tau$ or $\tau + z_{n\nu}$. By Assumptions A2–A5 and (23),

$$\mathrm{E}\big(C^+_{n\nu}(X_1,Y_1,z_1)C^+_{n\nu}(X_1,Y_1,z_2)\big)$$
$$= \mathrm{E}\big(D^+_{n\nu}(X_1,z_{\min})D^+_{n\nu}(X_1,z_{\max})\big(Y_1 - m(X_1)\big)^2\big) + O(h^{2\nu+3})$$
$$= h\frac{v(\tau)}{f(\tau)}\Big(K^+_\nu(0)^2\frac{z_{\min}}{a_{n\nu}} + K^+_\nu(0)K'^+_\nu(0)\frac{z_{\min}z_{\max}}{a^2_{n\nu}}\big(1+o(1)\big)\Big)\big(1+O(h)\big)$$
$$+ \int_{\tau^{\max}_{n\nu}}^{\tau^{\max}_{n\nu}+h} D^+_{n\nu}(u,z_{\min})D^+_{n\nu}(u,z_{\max})v(u)\,f(u)\,du + O(h^{2\nu+3}) \tag{24}$$

uniformly in z_1 and z_2.

Next, consider the second term in the brackets in (22) for the case $z_{\min} = z_1$. The other cases can be dealt with in a similar way. We note that $C^+_{n\nu}(w,u,z) = 0$ for $w < \tau$, and that

$$\sup_{u\in[\tau,\tau+z_{n\nu}]} |m(u) - \bar m_\pm(\tau+z_{n\nu},u)| \le (\mathrm{const.})\Big(\frac{h}{a_{n\nu}}\Big)^{\nu+1};$$

$$\sup_{u\in[\tau,\tau+z_{n\nu}]} |m(u) - \bar m_+(\tau,u)| \le (\mathrm{const.})\Big(\frac{h}{a_{n\nu}}\Big)^{\nu+1};$$

$$\sup_{u\in[\tau,\tau+z_{n\nu}]} \Big|m(u) - \bar m_-(\tau,u) - \frac{\Delta_\nu}{\nu!}(u-\tau)^\nu\Big| \le (\mathrm{const.})\Big(\frac{h}{a_{n\nu}}\Big)^{\nu+1}. \tag{25}$$

By (23) and (25),

$$\mathrm{E}\big(C^+_{n\nu}(X_1,Y_1,z_1)C^-_{n\nu}(X_1,Y_1,z_2)\big)$$
$$= -h\frac{v(\tau)}{f(\tau)}\Big(K^+_\nu(0)K^-_\nu(0)\frac{z_{\min}}{a_{n\nu}} + \tfrac12\big(K^+_\nu(0)K'^-_\nu(0) + K'^+_\nu(0)K^-_\nu(0)\big) \tag{26}$$
$$\times\Big(2\frac{z_{\min}z_{\max}}{a^2_{n\nu}} - \Big(\frac{z_{\min}}{a_{n\nu}}\Big)^2\Big)\big(1+o(1)\big)\Big)\big(1+O(h)\big) + O\Big(\Big(\frac{h}{a_{n\nu}}\Big)^{2\nu+2} + h^{2\nu+3}\Big)$$

uniformly in z_1 and z_2. Analogously,

$$\begin{aligned}
&\mathrm{E}\big(C^-_{n\nu}(X_1, Y_1, z_1)C^+_{n\nu}(X_1, Y_1, z_2)\big) \\
&\quad = -h\frac{v(\tau)}{f(\tau)}\Big(K^-_\nu(0)K^+_\nu(0)\frac{z_{\min}}{a_{n\nu}} + \tfrac{1}{2}\big(K^-_\nu(0)K'^+_\nu(0) + K'^-_\nu(0)K^+_\nu(0)\big) \\
&\qquad \times \Big(\frac{z_{\min}}{a_{n\nu}}\Big)^2\big(1+o(1)\big)\Big)\big(1+O(h)\big) + O\Big(\Big(\frac{h}{a_{n\nu}}\Big)^{2\nu+2} + h^{2\nu+3}\Big) \qquad (27)
\end{aligned}$$

uniformly in z_1 and z_2. Now we consider the last term in the brackets in (22). We have

$$\sup_{u\in[\tau-h,\tau)} |m(u) - \bar{m}_-(\tau, u)| \le (\text{const.})h^{\nu+1}\,. \qquad (28)$$

On the other hand, by (20),

$$\sup_{u\in[\tau+z_{n\nu}-h,\tau)} \Big|m(u) - \bar{m}_-(\tau+z_{n\nu}, u) + \frac{\Delta_\nu}{\nu!}(u-\tau-z_{n\nu})^\nu\Big| \le (\text{const.})h^{\nu+1}\,. \qquad (29)$$

By (28) and (29) with the definition of $C^-_{n\nu}$, the last term at (22) equals

$$\begin{aligned}
&\mathrm{E}\big(C^-_{n\nu}(X_1, Y_1, z_1)C^-_{n\nu}(X_1, Y_1, z_2)\big) = \frac{h}{f(\tau)}\Big(\big(v(\tau) + \Delta_0^2\,\mathrm{I}(\nu=0)\big)K^-_\nu(0)^2\frac{z_{\min}}{a_{n\nu}} \\
&\quad + v(\tau)K^-_\nu(0)K'^-_\nu(0)\frac{z_{\min}z_{\max}}{a^2_{n\nu}}\big(1+o(1)\big)\Big)\big(1+O(h)\big) \\
&\quad + \int_{\tau-h}^{\tau} D^-_{n\nu}(u, z_{\min})D^-_{n\nu}(u, z_{\max})v(u)f(u)\,du + O\Big(\Big(\frac{h}{a_{n\nu}}\Big)^{2\nu+1}\mathrm{I}(\nu\ne 0) + h^{2\nu+3}\Big) \\
&\qquad (30)
\end{aligned}$$

uniformly in z_1 and z_2. Combining the first leading terms in (24), (26), (27) and (30) concludes the proof of Lemma 2(i) for the case $z_1, z_2 > 0$.

Next, we prove the second part of the lemma for the case $z_1, z_2 > 0$. If ν is odd, all terms in the brackets in (24), (26), (27) and (30) are cancelled due to the relation (12). Note that $D^{\pm}_{n\nu}(w, z) = O(a^{-1}_{n\nu})$ uniformly in w and z by Assumptions A2 and A4. Hence, the integral terms in (24) and (30) are $O(h/a^2_{n\nu})$. Since $h^2/a_{n\nu} = o(h/a^2_{n\nu})$ for odd ν by the assumption $nh^{2\nu+3} \to 0$ in Assumption A5, the $O(h^2/a_{n\nu})$ terms in (24), (26), (27) and (30) are negligible. Thus, it is enough to consider the integral terms in (24) and (30). By Assumptions A2 and A4, the integral term in (24) can be written as

$$h\frac{z_{\min}z_{\max}}{a^2_{n\nu}}\frac{v(\tau)}{f(\tau)}\int_0^1 K'^+_\nu(u)^2\,du\big(1+o(1)\big) \qquad (31)$$

uniformly in z_1 and z_2. Similarly, we can show that the integral term in (30) equals

$$h\frac{z_{\min}z_{\max}}{a^2_{n\nu}}\frac{v(\tau)}{f(\tau)}\int_0^1 K'^-_\nu(u)^2\,du\big(1+o(1)\big) \qquad (32)$$

uniformly in z_1 and z_2. From (12), (31) and (32), Lemma 2(ii) follows for the case $z_1, z_2 > 0$. It can be shown in a similar way that the lemma follows for the case $z_1, z_2 < 0$ too.

Following the lines in the proof for the case $z_1z_2 > 0$, we obtain the results for the case $z_1z_2 < 0$. This concludes the proof.

Lemma 3. *Suppose that the assumptions in Theorem 1 are satisfied. For each $z \in [-M, M]$, $\phi_{n\nu}(z)$ satisfies Lyapounov's condition.*

Proof. By Lemma 2, $\mathrm{var}(\phi_{n\nu}(z)) = O(1)$. It can be shown that, for some positive ζ,

$$\left(\frac{a_{n\nu}^{\nu+1}}{nh^{\nu+1}}\right)^{2+\zeta} \sum_{j=1}^{n} \mathrm{E}\left(\left|C_{n\nu}^{+}(X_j, Y_j, z) - C_{n\nu}^{-}(X_j, Y_j, z)\right|^{2+\zeta}\right) \longrightarrow 0 \qquad \text{as } n \to \infty\,.$$

This implies the result.

Lemma 4. *Suppose that the assumptions in Theorem 1 are satisfied. Then, the sequence of the process $\psi_{n\nu}(\cdot) = \phi_{n\nu}(\cdot) - \mathrm{E}(\phi_{n\nu}(\cdot))$ is tight.*

Proof. By Lemma 2, there exists a positive constant C_3 such that

$$\begin{aligned}\mathrm{E}\left(\left(\psi_{n\nu}(z_1) - \psi_{n\nu}(z_2)\right)^2\right) &= \mathrm{var}\left(\phi_{n\nu}(z_1)\right) + \mathrm{var}\left(\phi_{n\nu}(z_2)\right) - 2\,\mathrm{cov}\left(\phi_{n\nu}(z_1), \phi_{n\nu}(z_2)\right) \\ &\le \begin{cases} C_3|z_2 - z_1| & \text{when } \nu \text{ is even,} \\ C_3(z_2 - z_1)^2 & \text{when } \nu \text{ is odd,} \end{cases}\end{aligned}$$

for sufficiently large n. According to Billingsley (1968 Theorem 12.3), this concludes the proof of Lemma 4.

Proof of Theorem 1. Lemma 3 implies that $\psi_{n\nu}(z)$, for fixed $z \in [-M, M]$, converges weakly to a normal distribution. Furthermore, by the Cramér–Wold device we can show that for fixed $z_1, \ldots, z_\ell, z_i \in [-M, M]$,

$$\left(\psi_{n\nu}(z_1), \ldots, \psi_{n\nu}(z_\ell)\right) \xrightarrow{\mathrm{d}} \mathrm{N}_\ell(0, \Sigma)\,,$$

where N_ℓ denotes the ℓ-variate normal distribution and Σ is the asymptotic covariance described in Lemma 2. This concludes the proof. See Billingsley (1968 Theorems 8.1 and 12.3).

Proof of Theorem 2. Consider the case $\Delta_\nu > 0$. The other case $\Delta_\nu < 0$ can be treated in the same way. According to Billingsley (1968 Theorem 5.1), we have

$$\operatorname*{argmax}_{z\in[-M,M]} \varphi_{n\nu}(z) \xrightarrow{\mathrm{d}} \operatorname*{argmax}_{z\in[-M,M]} \varphi_\nu(z) \tag{33}$$

for any $M > 0$. If we prove

$$\sup_{x\in Q, |x-\tau|\ge(h/a_{n\nu})M} \hat{\Delta}_\nu(x) = o_p(1)$$

for any $M > 0$, the result (33) can be extended to the whole real line $(-\infty, \infty)$. The uniform convergence can be proved by a standard technique; see e.g. Mack & Silverman (1982). By (13), the asymptotic distribution of $\hat{\tau}$ in Theorem 2(ii) is obtained explicitly for odd ν.

Proof of Corollary 1. Theorem 1 shows that $\sqrt{nh^{2\nu+1}}(\hat{\Delta}_\nu(\hat{\tau}) - \hat{\Delta}_\nu(\tau)) \xrightarrow{\mathrm{p}} 0$. Now,

$$\sqrt{nh^{2\nu+1}}\left(\hat{\Delta}_\nu(\hat{\tau}) - \Delta_\nu\right) = \sqrt{nh^{2\nu+1}}\left(\hat{\Delta}_\nu(\hat{\tau}) - \hat{\Delta}_\nu(\tau)\right) + \sqrt{nh^{2\nu+1}}\left(\hat{\Delta}_\nu(\tau) - \Delta_\nu\right). \tag{34}$$

Note that $\hat{\Delta}_\nu(\tau) = \tilde{\Delta}_\nu(\tau) + \Delta_\nu$. Define

$$\tilde{\lambda}_\nu(\tau) = \frac{\nu!}{nh^{\nu+1} f(\tau)} \sum_{j=1}^{n} \Big(K_\nu^+\Big(\frac{X_j - \tau}{h}\Big)\big(Y_j - \bar{m}_+(\tau, X_j)\big) - K_\nu^-\Big(\frac{\tau - X_j}{h}\Big)\big(Y_j - \bar{m}_-(\tau, X_j)\big)\Big).$$

By (5) and (6), $\sqrt{nh^{2\nu+1}}(\tilde{\Delta}_\nu(\tau) - \tilde{\lambda}_\nu(\tau)) = o_p(1)$. According to Assumptions A1, A2 and A4, $\sqrt{nh^{2\nu+1}}\,\mathrm{E}(\tilde{\lambda}_\nu(\tau)) = \sqrt{nh^{2\nu+1}}O(h)$ which is $o(1)$ by the assumption $nh^{2\nu+3} \to 0$ in Assumption A5. Now, since the support of K is $[0, 1]$,

$$nh^{2\nu+1}\mathrm{var}\big(\tilde{\lambda}_\nu(\tau)\big) = (\nu!)^2 \frac{v(\tau)}{f(\tau)} \Big(\int K_\nu^+(u)^2 + \int K_\nu^-(u)^2 \Big) du \big(1 + O(h)\big) + O(h^{2\nu+1}).$$

Using Assumption A7, the Lyapounov's condition for $\sqrt{nh^{2\nu+1}}\,\tilde{\lambda}_\nu(\tau)$ is satisfied. These together with (34) imply Corollary 1.

References

BHATTACHARYA, P.K. & BROCKWELL, P.J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrsch. Verw. Gebiete* **37**, 51–75.

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

BRAUN, J.V. & MÜLLER, H.G. (1998). Statistical methods for DNA sequence segmentation. *Statist. Sci.* **13**, 142–162.

CARLSTEIN, E., MÜLLER, H.G. & SIEGMUND, D. (Eds) (1994). *Change-Point Problems*. Institute of Mathematical Statistics — Monograph Series, **23**. Hayward, CA: Institute of Mathematical Statistics.

EUBANK, R.L. & SPECKMAN, P.L. (1994). Nonparametric estimation of functions with jump discontinuities. In *Change-Point Problems*, eds E. Carlstein, H.G. Müller & D. Siegmund, pp. 130–144. Institute of Mathematical Statistics — Monograph Series, **23**. Hayward, CA: Institute of Mathematical Statistics.

FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and Its Application*. London: Chapman & Hall.

GIJBELS, I. & GODERNIAUX, A.C. (2004). Bandwidth selection for change point estimation in non-parametric regression. *Technometrics* **46**, 76–86.

GIJBELS, I., HALL, P. & KNEIP, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51**, 231–251.

GRÉGOIRE, G. & HAMROUNI, Z. (2002). Change point estimation by local linear smoothing. *J. Multivariate Anal.* **83**, 56–83.

HALL, P. & TITTERINGTON, D.M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34**, 429–440.

HORVÁTH, L. & KOKOSZKA, P. (2002). Change-point detection with non-parametric regression. *Statistics* **36**, 9–31.

JONES, M.C., DAVIES, S.J. & PARK, B.U. (1994). Versions of kernel-type regression estimators. *J. Amer. Statist. Assoc.* **89**, 825–832.

JOSE, C.T. & ISMAIL, B. (1999). Change points in non-parametric regression functions. *Comm. Statist. Theory Methods* **28**, 1883–1902.

KOO, J.Y. (1997). Spline estimation of discontinuous regression functions. *J. Comput. Graph. Statist.* **6**, 266–284.

LOADER, C.R. (1996). Change point estimation using non-parametric regression. *Ann. Statist.* **24**, 1667–1678.

MACK, Y.P. & SILVERMAN, B.W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61**, 405–415.

MCDONALD, J.A. & OWEN, A.B. (1986). Smoothing with split linear fits. *Technometrics* **28**, 195–208.

MÜLLER, H.G. (1987). Weighted local regression and kernel methods for non-parametric curve fitting. *J. Amer. Statist. Assoc.* **82**, 231–238.

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

MÜLLER, H.G. (1993). Comment on 'Local regression: automatic kernel carpentry', by T. Hastie & C. Loader. *Statist. Sci.* **8**, 134–139.

MÜLLER, H.G. & SONG, K.S. (1997). Two-stage change-point estimators in smooth regression models. *Statist. Probab. Lett.* **34**, 323–335.

MÜLLER, H.G. & WANG, J.L. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika* **77**, 305–314.

NASON, G.P. & SILVERMAN, B.W. (1994). The discrete wavelet transform in S. *J. Comp. Graph. Statist.* **3**, 163–191.

QIU, P. (1994). Estimation of the number of jumps of the jump regression functions. *Comm. Statist. Theory Methods* **23**, 2141–2155.

QIU, P. & YANDELL, B. (1998). A local polynomial jump detection algorithm in non-parametric regression. *Technometrics* **40**, 142–152.

RAIMONDO, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26**, 1379–1397.

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

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

YIN, Q. (1988). Detection of the number, locations and magnitudes of jumps. *Comm. Statist. Stochastic Models* **4**, 445–455.