The Annals of Statistics
1992, Vol. 20, No. 2, 737–761

CHANGE-POINTS IN NONPARAMETRIC REGRESSION ANALYSIS[1]

BY HANS-GEORG MÜLLER

University of California, Davis

Estimators for location and size of a discontinuity or change-point in an otherwise smooth regression model are proposed. The assumptions needed are much weaker than those made in parametric models. The proposed estimators apply as well to the detection of discontinuities in derivatives and therefore to the detection of change-points of slope and of higher order curvature. The proposed estimators are based on a comparison of left and right one-sided kernel smoothers. Weak convergence of a stochastic process in local differences to a Gaussian process is established for properly scaled versions of estimators of the location of a change-point. The continuous mapping theorem can then be invoked to obtain asymptotic distributions and corresponding rates of convergence for change-point estimators. These rates are typically faster than $n^{-1/2}$. Rates of global L^p convergence of curve estimates with appropriate kernel modifications adapting to estimated change-points are derived as a consequence. It is shown that these rates of convergence are the same as if the location of the change-point was known. The methods are illustrated by means of the well known data on the annual flow volume of the Nile river between 1871 and 1970.

1. Introduction. Nonparametric regression methods are usually applied in order to obtain a smooth fit of a regression curve without having to specify a parametric class of regression functions. Sometimes a generally smooth curve might contain an isolated discontinuity or change-point in the curve or in a (possibly higher order) derivative, and in many cases interest focuses on the occurrence of such change-points. In parametric approaches to the regression change-point problem, simple linear regressions before and after a possible change-point are assumed, and then the possibility of a discontinuity in the form of a jump or of a jump in the first derivative, or, equivalently, a slope change, is incorporated into the model; see for instance Hinkley (1969) and Brown, Durbin and Evans (1975).

The analysis of change-points which describe sudden, localized changes typically occurring in economics, medicine and the physical sciences has recently found increasing interest. General smoothness assumptions, allowing for a large class of regression functions to be considered, seem to be more appropriate in a variety of applied problems than parametric modelling. An

Received February 1990; revised September 1991.
[1]Research supported by NSF Grant DMS-90-02423 and AFOSR Grant 89-0386.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62G20.
Key words and phrases. Boundary kernel, change of slope, curve fitting, discontinuity, end-point, functional limit theorem, jump size, kernel estimation, L^p convergence, Nile data, smoothing, weak convergence.

example which will be discussed is the annual Nile river flow data for the years 1871–1970, which have been analysed under a parametric change-point model by Cobb (1978).

If the regression function g is l times continuously differentiable for some $l \geq 0$, $g \in \mathscr{C}^l$, and a kernel smoother with a kernel function of the order k is chosen, that is, a kernel function with exactly $(k - 1)$ vanishing moments, the rate of convergence of mean squared error (MSE) or integrated mean squared error (IMSE) is well known to be $n^{-\alpha}$, where $\alpha = 2\min(k,l)/(2\min(k,l) + 1)$; the rate therefore depends jointly on the smoothness of the curve and the order of the kernel function. Obviously, a change-point in a curve or a derivative will alter this rate of convergence; in fact, if the curve or derivative to be estimated is not continuous at a given point, the ordinary kernel estimator at this point for curve/derivative will not be consistent. It is therefore of interest for purposes of curve estimation itself to adapt for change-points. In addition to this statistical motivation, for many applications it is of intrinsic concern to analyse change-points in curves or derivatives; see Cobb (1978), and the references given in McDonald and Owen (1986); compare also Cline and Hart (1991).

Although the focus in this paper is on fixed design nonparametric regression, the proposed methods can be adapted to other settings of curve estimation. For example, in hazard rate estimation, simple parametric change-point models require already quite sophisticated techniques; see, for example, Matthews, Farewell and Pyke (1985). Smooth approximation of a change-point model by a model which contains a point of most rapid change and the corresponding statistical inference was considered by Müller and Wang (1990) in the context of hazard functions under random censoring.

Other related approaches for detecting a change in the distribution of a sequence of random variables have been developed by Chernoff and Zacks (1964), Bhattacharya and Brockwell (1976) and Siegmund (1986), among many others. Weak convergence of change-point estimators in parametric regression models was investigated by Bhattacharya (1991).

The setting considered here is the fixed design regression model

$$\text{(M1)} \qquad y_{i,n} = g(t_{i,n}) + \varepsilon_{i,n}, \qquad t_{i,n} \in [0,1], 1 \leq i \leq n,$$

where $y_{i,n}$ are noisy measurements of the regression function g taken at points $t_{i,n}$ and $\varepsilon_{i,n}$ are i.i.d. errors with $E(\varepsilon_{i,n}) = 0$, $\operatorname{var}(\varepsilon_{i,n}) = \sigma^2 < \infty$. The design points $t_{i,n}$ are assumed to be equidistant. This assumption could be relaxed by requiring only the existence of a smooth design density, where $\int_0^{t_{i,n}} f(x)\,dx = (i-1)/(n-1)$. With minor modifications, the results then remain valid for asymptotically nonequidistant designs. Fixing the left and right endpoint of the compact support of g at 0 (respectively 1) incurs no loss of generality. In the following, indices n will be omitted whenever feasible.

Let $\nu \geq 0$ be an integer and $k \geq 2$ be an even integer. Assume that a change-point exists for $g^{(\nu)}$ at τ, $0 < \tau < 1$, in the following sense: There exists

$f \in \mathscr{C}^{k+\nu}([0,1])$ such that

$$\text{(M2)} \qquad g^{(\nu)}(t) = f^{(\nu)}(t) + \Delta_\nu 1_{[\tau,1]}(t), \qquad \Delta_\nu > 0, 0 \le t \le 1.$$

The case $\Delta_\nu < 0$ can be treated analogously. Define $g_+^{(\nu)}(\tau) = \lim_{t \downarrow \tau} g^{(\nu)}(t)$, $g_-^{(\nu)}(\tau) = \lim_{t \uparrow \tau} g^{(\nu)}(t)$ and $g^{(\nu)}(\tau) = g_+^{(\nu)}(\tau)$, and observe that

$$\Delta_\nu = g_+^{(\nu)}(\tau) - g_-^{(\nu)}(\tau), \tag{1.1}$$

where Δ_ν is the jump size at the possible change-point τ of the νth derivative. The case $\Delta_\nu = 0$ corresponds to the nonexistence of a change-point at τ.

The main results of this paper concern weak convergence of estimators $\hat{\tau}$ of the location of the change-point τ (Theorem 3.1) and rates of global L^p convergence of kernel estimators adjusted to an estimated change-point (Theorem 4.1). The paper is organized as follows: Section 2 presents a discussion of kernel estimators using kernel functions with one-sided support and their application to change-point estimation, which is based on maximizing the difference between one-sided kernel smoothers. Section 3 is devoted to the study of a functional limit theorem for a local deviation process. The functional mapping theorem is used to obtain the limit distribution for estimated change-points. These results are applied in Section 4 to obtain rates of L^p convergence for kernel estimators with and without change-point adaptation. An application of the methods to the Nile data is discussed in Section 5. The proofs for Section 3 are compiled in Section 6; those for Section 4 in Section 7.

2. One-sided kernels and change-point estimators. Omitting in the following indices n, that is, writing $s_i = s_{i,n}$, $t_i = t_{i,n}$, $y_i = y_{i,n}$ and $\varepsilon_i = \varepsilon_{i,n}$, define $s_i = (t_i + t_{i+1})/2$, $i = 1, \dots, n-1$, $s_0 = 0$, $s_n = 1$ and consider the following kernel estimators $\hat{g}^{(\nu)}(t)$ of $g^{(\nu)}(t)$, $t \in [0,1]$:

$$\hat{g}^{(\nu)}(t) = \frac{1}{b^{\nu+1}} \sum_{i=1}^{n} y_i \int_{s_{i-1}}^{s_i} K^{(\nu)}\left(\frac{t-u}{b}\right) du. \tag{2.1}$$

Here $b = b(n)$ is a sequence of bandwidths which is required to satisfy

$$\text{(B1)} \quad b \to 0, \qquad nb^{2\nu+1} \to \infty \quad \text{as} \quad n \to \infty, \qquad \limsup_{n \to \infty} nb^{2(k+\nu)+1} < \infty,$$

and $K^{(\nu)}$ is the kernel function, which is assumed to be the νth derivative of a function K with compact support $[-1,1]$. Other kernel estimators like the one proposed by Priestley and Chao (1972) could be used as well. Observe that $K^{(\nu)}$ is a two-sided kernel with symmetric support.

For purposes of boundary modification [Rice (1984b); Gasser, Müller and Mammitzsch (1985); Müller (1991)], kernels with asymmetric supports were considered which correspond to smoothing windows which are asymmetric with respect to the point t. For the nonparametric estimation of regression functions when change-points or discontinuities are present, an algorithm which is based on local least squares fits and which also uses asymmetric smoothing windows was proposed by McDonald and Owen (1986); related results on a.s. convergence of change-point estimators, also using one-sided moving averages in a smooth function model with white noise, were obtained by Yin (1988).

Let $K_+^{(\nu)}$ and $K_-^{(\nu)}$ be one-sided kernel functions with support$(K_+^{(\nu)}) = [-1, 0]$ and support$(K_-^{(\nu)}) = [0, 1]$, and define one-sided regression estimates for the νth derivative $g^{(\nu)}(t)$:

$$\hat{g}_{\pm}^{(\nu)}(t) = \frac{1}{b^{\nu+1}} \sum_{i=1}^{n} y_i \int_{s_{i-1}}^{s_i} K_{\pm}^{(\nu)}\left(\frac{t-u}{b}\right) du. \tag{2.2}$$

The idea is to base inference for change-points on differences of right- and left-sided estimates:

$$\hat{\Delta}^{(\nu)}(t) = \hat{g}_+^{(\nu)}(t) - \hat{g}_-^{(\nu)}(t). \tag{2.3}$$

Intuitively, the location of the maximum of these differences will be a reasonable estimator for the location of the change-point. Let $Q \subset (0, 1)$ be a closed interval such that $\tau \in Q$. Define the estimators

$$\hat{\tau} = \inf\left\{\rho \in Q \colon \hat{\Delta}^{(\nu)}(\rho) = \sup_{x \in Q} \hat{\Delta}^{(\nu)}(x)\right\} \tag{2.4}$$

for the location of the change-point τ and

$$\hat{\Delta}^{(\nu)}(\hat{\tau}) = \hat{g}_+^{(\nu)}(\hat{\tau}) - \hat{g}_-^{(\nu)}(\hat{\tau}) \tag{2.5}$$

for the jump size in the νth derivative. Defining $\hat{\tau}$ as maximizer over Q instead of over $[0, 1]$ serves the sole purpose of excluding change-points located arbitrarily close to the boundary.

Assume that for some integer $\mu \geq 0$,

$$\begin{gathered} K \in \mathscr{C}^{\nu+\mu}([-1,1]) \cap \mathscr{K}_{0,k}([-1,1]), \\ K^{(j)}(-1) = K^{(j)}(1) = 0, \qquad 0 \leq j < \nu + \mu, \end{gathered} \tag{K1}$$

where k as before is an even integer $k \geq 2$, $\nu < k$ and

$$\mathscr{K}_{\nu,l}([a_1, a_2]) = \left\{ f \in \mathscr{C}([a_1, a_2]) \colon \text{support}\,(f) = [a_1, a_2], \right.$$

$$\left. \int f(x) x^j \, dx \begin{cases} = (-1)^{\nu} \nu!, & j = \nu, \\ = 0, & 0 \leq j < l,\ j \neq \nu, \\ \neq 0, & j = l, \end{cases} \right\}.$$

It then follows by integration by parts that

$$\begin{gathered} K^{(\nu)} \in \mathscr{C}^{\mu}([-1,1]) \cap \mathscr{K}_{\nu,k+\nu}([-1,1]), \\ K^{(\nu+j)}(-1) = K^{(\nu+j)}(1) = 0, \qquad 0 \leq j < \mu. \end{gathered} \tag{2.6}$$

The integer $\mu \geq 0$ is a measure for the overall smoothness of $K^{(\nu)}$, which is inherited by the curve estimate. According to (2.6), the kernel $K^{(\nu)}$ is $(\mu - 1)$ times differentiable on $\mathbb{R}$ and $K^{(\mu-1)}$ is absolutely continuous.

Analogously, assume for kernels K_+ and K_-,

$$\text{(K2)} \qquad \begin{aligned} &K_+ \in \mathscr{C}^{\nu+\mu}([-1,0]) \cap \mathscr{K}_{0,k}([-1,0]),\\ &K_+^{(j)}(-1) = K_+^{(j)}(0) = 0, \qquad 0 \leq j < \nu + \mu,\\ &K_- \in \mathscr{C}^{\nu+\mu}([0,1]) \cap \mathscr{K}_{0,k}([0,1]),\\ &K_-^{(j)}(0) = K_-^{(j)}(1) = 0, \qquad 0 \leq j < \nu + \mu, \end{aligned}$$

which again implies that

$$\text{(2.7a)} \qquad \begin{aligned} &K_+^{(\nu)} \in \mathscr{C}^{\mu}([-1,0]) \cap \mathscr{K}_{\nu,k+\nu}([-1,0]),\\ &K_+^{(\nu+j)}(-1) = K_+^{(\nu+j)}(0) = 0, \qquad 0 \leq j < \mu, \end{aligned}$$

$$\text{(2.7b)} \qquad \begin{aligned} &K_-^{(\nu)} \in \mathscr{C}^{\mu}([0,1]) \cap \mathscr{K}_{\nu,k+\nu}([0,1]),\\ &K_-^{(\nu+j)}(0) = K_-^{(\nu+j)}(1) = 0, \qquad 0 \leq j < \mu. \end{aligned}$$

Observe that K_+ (respectively K_-) acts on the r.h.s. (respectively l.h.s.) of t according to the convolution property in definition (2.1), so that application of these kernels corresponds to employing smoothing windows $[t, t+b]$ (respectively $[t-b, t]$). For an example of kernels satisfying (2.7b), set ${}_{\mu}K_-(x) = x^{\mu}(1-x)^{\mu}(\alpha_0 + \alpha_1 x)$ on $[0,1]$. From (2.5) and (2.7) one obtains the solutions $\alpha_0 = a_{22}/(a_{11}a_{22} - a_{12}a_{21})$ and $\alpha_1 = -a_{12}/(a_{11}a_{22} - a_{12}a_{21})$, where $a_{11} = (\mu!)^2/(2\mu+1)!$, $a_{12} = a_{21} = \mu!(\mu+1)!/(2\mu+2)!$ and $a_{22} = \mu!(\mu+2)!/(2\mu+3)!$. Special cases are the following kernel functions K_- on the one-sided interval $[0,1]$: $\mu = 0$: ${}_0K_-(x) = 2(2-3x)$, $\mu = 1$: ${}_1K_-(x) = 12x(1-x)(3-5x)$ and $\mu = 2$: ${}_2K_-(x) = 80x^2(1-x)^2(3-5x)$.

Another possibility of interest is to prescribe a zero in the kernel function only at one endpoint of its support to render the estimated function smoother. If, for instance, a zero is required at 1, the approach $K_-(x) = (1-x)(\alpha_0 + \alpha_1 x)$ leads to the quadratic kernel polynomial ${}_{01}K_-(x) = 6(1-x)(1-2x)$.

Observe that it follows from (K2) that if $K_-^{(\nu)}$ satisfies (2.7b), then a kernel $K_+^{(\nu)}$ defined by

$$\text{(K3)} \qquad K_+^{(\nu)}(x) = (-1)^{\nu} K_-^{(\nu)}(-x)$$

satisfies (2.7a). An additional assumption we make is

$$\text{(K4)} \qquad K_-^{(\nu+\mu)} \in \operatorname{Lip}([0,1]),\ K_-^{(\nu+\mu)}(0) > 0,\ (\nu+\mu) \text{ is odd and } \mu \geq 1.$$

Analogous conditions follow for $K_+^{(\nu+\mu)}$, assuming (K3).

These considerations can be extended to cover kernel functions with more general asymmetric supports $[-q, 1]$. Such kernels will be used to estimate near an estimated change-point in Sections 4 and 5. In the right boundary interval $[1-b, 1]$, choose kernels satisfying

$$(2.8)\qquad \begin{aligned} &K_-^{(\nu)}(x,q) \in \mathscr{C}^{\mu}([-q,1]) \cap \mathscr{K}_{\nu,k+\nu}([-q,1]),\\ &K_-^{(\nu+j)}(-q,q) = K_-^{(\nu+j)}(1,q) = 0, \qquad 0 \le j < \mu, 0 \le q \le 1, \end{aligned}$$

where $q = (1-x)/b$. Analogously, for the left boundary interval $[0, b]$,

$$(2.9)\qquad \begin{aligned} &K_+^{(\nu)}(x,q) \in \mathscr{C}^{\mu}([-1,q]) \cap \mathscr{K}_{\nu,k+\nu}([-1,q]),\\ &K_+^{(\nu+j)}(-1,q) = K_+^{(\nu+j)}(q,q) = 0, \qquad 0 \le j < \mu, 0 \le q \le 1, \end{aligned}$$

where $q = x/b$.

These kernels can be constructed analogously as for the special case $q = 0$ above. Special kernels $K_-^{(\nu)}(\cdot, q)$ minimize the functional $\int K_-^{(\nu+\mu)}(x)^2\,dx$, subject to (2.8), for $0 \le q \le 1$; see Müller (1991). These solutions are polynomials of degree $(k + \nu + 2\mu - 1)$ with the following properties:

$$(\text{K5})\qquad \begin{aligned} &\text{support } K_-^{(\nu)}(\cdot, q) = [-q, 1];\\ &K_-^{(\nu+j)}(-q,q) = K_-^{(\nu+j)}(1,q) = 0, \qquad 0 \le j < \mu;\\ &\sup_{x,q} \left| K_-^{(\nu)}(x,q) \right| \le C < \infty;\\ &\sup_{q} \left| K_-^{(\nu)}(x_1,q) - K_-^{(\nu)}(x_2,q) \right| \le C|x_1 - x_2|;\\ &\sup_{x} \left| K_-^{(\nu)}(x,q_1) - K_-^{(\nu)}(x,q_2) \right| \le C|q_1 - q_2|; \end{aligned}$$

and analogously for $K_+^{(\nu)}(\cdot, q)$, where

$$(\text{K6})\qquad K_+^{(\nu)}(x,q) = (-1)^{\nu} K_-^{(\nu)}(x,q).$$

3. Weak convergence of local deviation processes and asymptotic distributions of change-point estimators. In this section a functional limit theorem for a process operating on increments of one-sided function estimates near τ is derived. The functional mapping theorem is then applied to obtain the limit distributions for change-point estimators $\hat{\tau}$. A similar device was used by Eddy (1980, 1982) in the context of estimating the mode of a probability density.

Let

$$\hat{\delta}_{\nu}(y) = \hat{\Delta}^{(\nu)}(\tau + yb) = \hat{g}_+^{(\nu)}(\tau + yb) - \hat{g}_-^{(\nu)}(\tau + yb),$$

and define for some $0 < M < \infty$, $-M \le z \le M$, the sequence of stochastic processes

$$(3.1)\quad \zeta_n(z) = \left(nb^{2\nu+1}\right)^{(\mu+\nu+1)/(2(\mu+\nu))}\left(\hat{\delta}_{\nu}\left(\frac{z}{\left(nb^{2\nu+1}\right)^{1/(2(\mu+\nu))}}\right) - \hat{\delta}_{\nu}(0)\right).$$

The scaling is chosen in such a way that processes ζ_n converge weakly.

Observe that $\zeta_n \in \mathscr{C}([-M, M])$. Denoting weak convergence by $\Rightarrow$, the following functional limit theorem holds.

THEOREM 3.1. *Assume that* (M1), (M2), (B1) *and* (K1)–(K4) *hold. Then*

$$\zeta_n \Rightarrow \zeta \quad \text{on } \mathscr{C}([-M, M]), \tag{3.2}$$

where ζ is a continuous Gaussian process with moment structure

$$E(\zeta(z)) = -\Delta_\nu z^{\mu+\nu+1} K_-^{(\mu+\nu)}(0)/(\mu+\nu+1)!, \tag{3.3}$$

$$\operatorname{cov}(\zeta(z_1), \zeta(z_2)) = 2 z_1 z_2 \sigma^2 \int K_-^{(\nu+1)}(v)^2\, dv. \tag{3.4}$$

Since the Gaussian limit process ζ is determined by its first and second moments, according to (3.3) and (3.4), it can be written equivalently as

$$\zeta(z) = -\Delta_\nu z^{\mu+\nu+1} K_-^{(\mu+\nu)}(0)/(\mu+\nu+1)! + Xz, \tag{3.5}$$

where $X \sim \mathscr{N}(0, 2\sigma^2 \int K_-^{(\nu+1)}(v)^2\, dv)$.

The proof of Theorem 3.1 follows from a sequence of lemmas in Section 6.

Asymptotic distributions of estimated change-points (2.4) can now be obtained as a consequence of this functional limit theorem. Under (K4), the limit process ζ of (3.5) is seen to have a unique maximum at

$$Z^* = \left[X(\mu+\nu)!/\Delta_\nu K_-^{(\mu+\nu)}(0)\right]^{1/(\mu+\nu)}. \tag{3.6}$$

Let Z_n be the location of the maximum of ζ_n. By construction,

$$\hat{\tau} = \tau + Z_n b/(nb^{2\nu+1})^{1/(2(\nu+\mu))}. \tag{3.7}$$

Observing Whitt (1970) and the global properties of ζ, following Eddy (1980), the convergence $\zeta_n \Rightarrow \zeta$ can be extended to $\mathscr{C}(-\infty, \infty)$, equipped with Whitt's metric, and the functional mapping theorem then implies $Z_n \to_{\mathscr{D}} Z^*$, where Z_n is now the global maximizer of ζ_n on $(-\infty, \infty)$.

COROLLARY 3.1. *Under the assumptions of Theorem* 3.1,

$$\begin{aligned} &(nb^{2\nu+1})^{1/2}\left(\frac{\hat{\tau}-\tau}{b}\right)^{\mu+\nu} \\ &\quad \to_{\mathscr{D}} \mathscr{N}\left(0, 2\left(\frac{(\mu+\nu)!}{\Delta_\nu K_-^{(\mu+\nu)}(0)}\right)^2 \sigma^2 \int K_-^{(\nu+1)}(v)^2\, dv\right). \end{aligned} \tag{3.8}$$

Note that the corresponding rate of convergence exceeds $n^{-1/2}$ in most cases. Consider for instance the important case $\mu = 1$, $\nu = 0$, $k = 2$. If the usual bandwidth choice $b = dn^{-1/5}$ is made, $d > 0$, then (3.8) becomes $n^{3/5}(\hat{\tau} - \tau) \to_{\mathscr{D}} \mathscr{N}(0, 2\, d\sigma^2 \int K_-^{(1)}(v)^2\, dv/(\Delta_\nu K_-^{(1)}(0))^2)$.

Another application of the functional mapping theorem shows that $\zeta_n(Z_n) \to_{\mathscr{D}} \zeta(Z^*)$ and therefore

$$(nb^{2\nu+1})^{1/2}\left\{\zeta_n(Z_n)/(nb^{2\nu+1})^{(\mu+\nu+1)/(2(\mu+\nu))}\right\} \to_P 0.$$

This implies $(nb^{2\nu+1})^{1/2}\{\hat{\Delta}^{(\nu)}(\hat{\tau}) - \hat{\Delta}^{(\nu)}(\tau)\} \to_P 0$, where $\hat{\Delta}^{(\nu)}(\cdot)$ is defined in (2.3). According to Lemma 6.6,

$$(nb^{2\nu+1})^{1/2}\left\{\hat{\Delta}^{(\nu)}(\tau) - \Delta_\nu\right\} \to_{\mathscr{D}} \mathscr{N}\left(0, 2\sigma^2 \textstyle\int K_-^{(\nu)}(v)^2\,dv\right),$$

and combining these results one obtains for the jump size estimator $\hat{\Delta}^{(\nu)}(\hat{\tau})$:

COROLLARY 3.2.

$$(nb^{2\nu+1})^{1/2}\left\{\hat{\Delta}^{(\nu)}(\hat{\tau}) - \Delta_\nu\right\} \to_{\mathscr{D}} \mathscr{N}\left(0, 2\sigma^2 \int K_-^{(\nu)}(v)^2\,dv\right). \tag{3.9}$$

If the construction of asymptotic confidence intervals is desired, one needs to substitute consistent estimators $\hat{\sigma}^2$ for σ^2. Such estimators were proposed by Rice (1984a) and Silverman (1985). A general class of quadratic estimators,

$$\hat{\sigma}^2 = \frac{1}{n - (m_1 + m_2)} \sum_{i=m_1+1}^{n-m_2} \left(\sum_{j=-m_1}^{m_2} \omega_j y_{j+i} \right)^2, \tag{3.10}$$

where $m_1, m_2 \geq 0$ are given fixed integers with $m_1 + m_2 \geq 1$, was considered in Müller and Stadtmüller (1987) and more recently in Hall, Kay and Titterington (1990). The law of large numbers implies $\hat{\sigma} \to_P \sigma$, if the regression function g is Lipschitz continuous, and if the weights ω_j satisfy $\sum \omega_j = 0$, $\sum \omega_j^2 = 1$.

A single discontinuity as in change-point model (M2) does not disturb the asymptotic consistency, but for practical purposes it is preferable to exclude a neighborhood around $\hat{\tau}$ when calculating $\hat{\sigma}^2$ (3.10). Calculating $\hat{\sigma}^2$ separately on the left- and on the right-hand side of $\hat{\tau}$ can also serve as a crude check whether σ is subject to change as well at τ. The results remain essentially valid for a model (M1), (M2) with $\operatorname{var}(\varepsilon_{i,n}) = \sigma^2(t_{i,n})$, where $\sigma^2(\cdot)$ is a Lipschitz continuous function with a possible change-point at τ.

Let Φ denote the Gaussian distribution function. With consistent estimators (3.10), one obtains asymptotic $100(1-\alpha)\%$ confidence intervals

$$\begin{aligned}&\hat{\tau} \pm b\left[\Phi^{-1}(1-\alpha/2)(\mu+\nu)!\hat{\sigma}/\hat{\Delta}^{(\nu)}(\hat{\tau})K_-^{(\mu+\nu)}(0)\right]^{1/(\mu+\nu)} \\ &\qquad \times \left[2\int K_-^{(\nu+1)}(v)^2\,dv/(nb^{2\nu+1})\right]^{1/(2\mu+2\nu)}\end{aligned} \tag{3.11}$$

for τ and $100(1-\alpha)\%$ upper/lower confidence bounds

$$\hat{\Delta}^{(\nu)}(\hat{\tau}) \pm \Phi^{-1}(1-\alpha)\left\{\left(2\hat{\sigma}^2\int K_-^{(\nu)}(v)^2\,dv\right)\Big/(nb^{2\nu+1})\right\}^{1/2} \tag{3.12}$$

for the jump size Δ_ν.

A test whether change-points with jump sizes exceeding a given value exist can be based on inverting lower/upper confidence bounds (3.12). If say this lower bound is $\hat{\Delta}_l$, then for any Δ with $0 < \Delta \le \hat{\Delta}_l$, the hypothesis H_0: $0 < \Delta_\nu \le \Delta$ would be rejected at level α. If $\Delta_\nu > \Delta$, this test obviously has asymptotic power 1. Analogously, if $\hat{\Delta}_u$ is the upper bound, for any $\Delta \ge \hat{\Delta}_u$, the hypothesis H_0: $\Delta_\nu \ge \Delta$ would be rejected at level α.

4. Global L^p consistency with unknown change-point and two-step procedures. Consider the problem of uniform consistency of kernel estimators in the L^p sense when a change-point is present. Owing to boundary effects, unmodified kernel estimators (2.1) will not be uniformly consistent,

$$\sup_{t \in [0,1]} \left| \hat{g}^{(\nu)}(t) - g^{(\nu)}(t) \right| = o_p(1), \tag{4.1}$$

even if a change-point is not present. The endpoints at 0 and 1 can be viewed as change-points with known location. Assume now in addition the existence of a change-point with unknown location τ, $0 < \tau < 1$. We investigate kernel estimators which are adjusted to known endpoints and an estimated change-point $\hat{\tau}$ (2.4), employing kernels $K_-^{(\nu)}(\cdot, q)$ and $K_+^{(\nu)}(\cdot, q)$, as defined in (K5) and (K6). Let

$$\hat{g}_\pm^{(\nu)}(t,q) = \frac{1}{b^{\nu+1}} \sum_{i=1}^{n} y_i \int_{s_{i-1}}^{s_i} K_\pm^{(\nu)}\left(\frac{t-u}{b}, q \right) du \tag{4.2}$$

and define

$$\tilde{g}^{(\nu)}(t,\hat{\tau}) = \begin{cases} \hat{g}^{(\nu)}(t), & \text{for } t \ge b,\ t \le 1-b,\ |t-\hat{\tau}| > b \text{ [see (2.1)]}; \\ \hat{g}_+^{(\nu)}(t,q), & \text{for } 0 \le t \le b \text{ with } q = t/b \text{ and} \\ & \text{for } \hat{\tau} < t \le \hat{\tau} + b \text{ with } q = (t-\hat{\tau})/b; \\ \hat{g}_-^{(\nu)}(t,q), & \text{for } 1-b \le t \le 1 \text{ with } q = (1-t)/b \text{ and} \\ & \text{for } \hat{\tau} - b \le t \le \hat{\tau} \text{ with } q = (\hat{\tau}-t)/b. \end{cases} \tag{4.3}$$

Even for adjusted estimators $\tilde{g}^{(\nu)}(\cdot, \hat{\tau})$, sup norm convergence (4.1) does not hold, due to asymptotically nonvanishing biases when estimating at points between τ and $\hat{\tau}$. However, rates of uniform consistency with respect to L_p norms might be of interest since the convergence $\hat{\tau} \to \tau$ is relatively fast for change-point estimators $\hat{\tau}$ (2.4).

To obtain direct global rates of convergence, consider the following additional assumptions on the errors in model (M1) and on the bandwidths:

(M3) There exists $s > 2$ such that $E(|\varepsilon_{i,n}|^s) < \infty$.

(B2)
$$\liminf_{n\to\infty} nb^{\nu+1+\delta} > 0, \qquad (nb^{2\nu+1})/\log n \to \infty,$$
$$(nb^{\nu+1})/(n^{1/(s-\delta)} \log n) \to \infty,$$

as $n \to \infty$, where s is as in (M3) and δ is any positive number $0 < \delta < s$.

LEMMA 4.1. *Under* (M1)–(M3), (B1), (B2) *and* (K1)–(K4), *it holds that*

$$|\hat{\tau} - \tau| = O_p\Big(\big[b^{2\mu-1}/n\big]^{1/2(\mu+\nu)}\Big). \tag{4.4}$$

The proof is in Section 7. For $\nu = 0$, $\mu = 1$, $|\hat{\tau} - \tau| = O_p([b/n]^{1/2})$ and observing (B1) and (B2), assuming (M3) for $s > s_0 \geq 2$ and choosing $b \sim (\log n/n)n^{1/s_0}$, the rate becomes $n^{-1}[n^{1/s_0}\log n]^{1/2}$. If all moments of the errors exist, this rate gets arbitrarily close to $n^{-1}[\log n]^{1/2}$.

THEOREM 4.1. *Assume* (M1)–(M3), (B1) *and* (B2) *for bandwidths* b_1, b.

(a) *Assume the boundary kernels* $K_{\pm}^{(\nu)}(\cdot, q)$ *in* (4.2) *satisfy* (K5) *and* (K6), *and the kernels* $K_{\pm}^{(\nu)}(\cdot, 0) = K_{\pm}^{(\nu)}$ *satisfy* (K1)–(K4). *If* b_1 *is used for the estimation of* τ *by* $\hat{\tau}$ (2.4) *and* b *is used for* $g^{(\nu)}(\cdot, \hat{\tau})$ (4.3), $b_1 \leq b$, *then it holds for* $p \geq 1$:

$$\int_0^1 |\tilde{g}^{(\nu)}(x, \hat{\tau}) - g^{(\nu)}(x)|^p\, dx$$

$$= O_p\left(\left[\frac{b_1^{2\mu-1}}{n}\right]^{1/(2(\mu+\nu))} + \left[\frac{b_1^{2\mu-1}}{nb^{2(\nu+1)(\mu+\nu)}}\right]^{p/(2(\mu+\nu))} + \left[\frac{\log n}{nb^{2\nu+1}}\right]^{p/2}\right) \tag{4.5}$$

$$+ O([b^{kp}]).$$

(b) *If* $\nu = 0$, $\mu = 1$, $1 \leq p < 2 + 1/k$, $s > s_0 = \max\{(k + 1/2)/(2k + 1 - kp), 2\}$ *and the bandwidths are chosen according to* $b_1 \sim (\log n/n)n^{1/s_0}$, $b \sim [\log n/n]^{1/(2k+1)}$, *then*

$$\int_0^1 |\tilde{g}(x, \hat{\tau}) - g(x)|^p\, dx = O_p\big([\log n/n]^{kp/(2k+1)}\big). \tag{4.6}$$

A key step in the proof is:

LEMMA 4.2.

$$\int_0^1 |\tilde{g}^{(\nu)}(x, \hat{\tau}) - g^{(\nu)}(x)|^p\, dx$$

$$= |\hat{\tau} - \tau| O_p(1) + \big[|\hat{\tau} - \tau| O_p(b^{-\nu-1})\big]^p \tag{4.7}$$

$$+ O_p\left(\left[\frac{\log n}{(nb^{2\nu+1})}\right]^{p/2}\right) + O\big([nb^{\nu}]^{-1} + b^{kp}\big).$$

Auxiliary results and proofs for this lemma and for Theorem 4.1 are in Section 7. It is of interest to compare the modified estimator $\tilde{g}(\cdot, \hat{\tau})$ with estimated change-point with the modified estimator $\tilde{g}(\cdot, \tau)$ when the location of the change-point is known. Let $\nu = 0$, $\mu = 1$. If the change-point location is known, minimizing the resulting terms in (4.7) corresponds exactly to the same minimization problem as the one leading to (4.6), so that under the assumptions of Theorem 4.1(b),

$$\int_0^1 |\tilde{g}(x,\tau) - g(x)|^p \, dx = O_p\big([\log n/n]^{kp/(2k+1)}\big); \tag{4.8}$$

that is, the rate is the same as for $\hat{g}(\cdot,\hat{\tau})$.

Another interesting comparison can be made between modified estimators $\tilde{g}(\cdot,\hat{\tau})$, $\tilde{g}(\cdot,\tau)$ (or without change-point) and unmodified estimators $\hat{g}(\cdot)$ (2.1). For the latter, one can show by the same arguments as in the proof of Lemma 4.2, that

$$\int_0^1 |\hat{g}(x) - g(x)|^p \, dx = O_p\big([\log n/nb]^{p/2}\big) + O(b), \tag{4.9}$$

yielding the optimizing bandwidth sequence $b \sim [\log n/n]^{p/(p+2)}$, and under $s > p + 2$,

$$\int_0^1 |\hat{g}(x) - g(x)|^p \, dx = O_p\big([\log n/n]^{p/(p+2)}\big). \tag{4.10}$$

This rate is poor as compared to (4.6) since in (4.9) the usual bias expansion cannot be carried out.

The practical procedure of estimating τ by $\hat{\tau}$ (2.4) with bandwidth b_1 in a first step, followed by applying the accordingly modified kernel estimator $\tilde{g}(x,\hat{\tau})$, with bandwidth b, provides an efficient two-step procedure for the estimation of curves with discontinuities in the sense that the resulting rate of convergence remains the same as when τ is known.

5. Application and discussion. This section serves to illustrate the methods with data on the annual volume of the Nile river from 1871 to 1970. These data and a discussion of various approaches to parametric change-point modelling are given in Cobb (1978). The question is whether and when there occurred an abrupt change in rainfall activity near the turn of the last century. Cobb suggests that a change occurs in the year 1898. The data are displayed in Figure 1.

Assume that the model (M1), (M2) applies to these data. The unmodified kernel estimate $\hat{g}(\cdot)$ (2.1) (choosing $\nu = 0$ and modifications at the boundaries, however not at a possible change-point) is shown in Figure 2. The bandwidth was determined visually and chosen as $b = 10$ years. The kernel used was $K(x) = 3(1 - x^2)/4$, a kernel corresponding to parameters $\nu = 0$, $k = 2$, $\mu = 1$ and $q = 1$ [compare Epanechnikov (1969)].

There is a hint of a relatively strong decline near 1900, but if there was a change-point, it is smoothed out. The evidence gets a little stronger by looking at the first derivative $\hat{g}^{(1)}(\cdot)$, estimated also with unmodified kernels (Figure 3), $b = 10$ and $K^{(1)}(x) = 15(x - x^3)/4$. This derivative estimate seems to corroborate that a change occurs shortly before 1900.

Auxiliary one-sided curve estimates $\hat{g}_+$, $\hat{g}_-$ as defined in (2.2), employing one-sided kernels $K_-(x) = 6(1 - x)(1 - 2x)$, $0 \le x \le 1$, are shown in Figure 4, also with bandwidth $b = 10$.

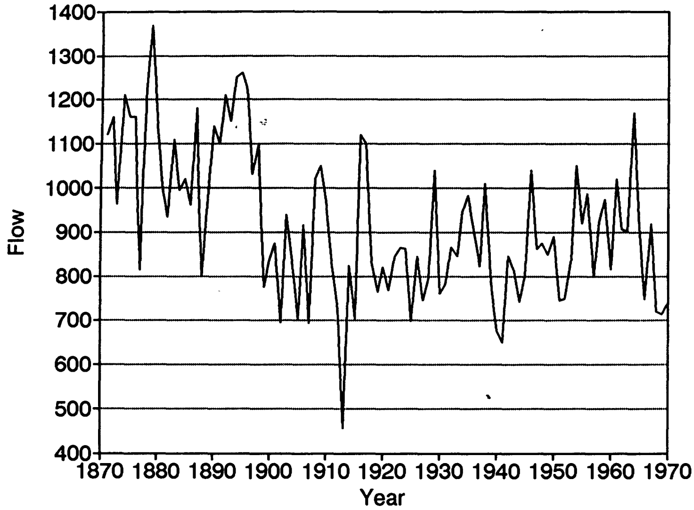

FIG. 1. *Annual volume of the Nile river* (10^8 m^3) *between* 1871 *and* 1970.

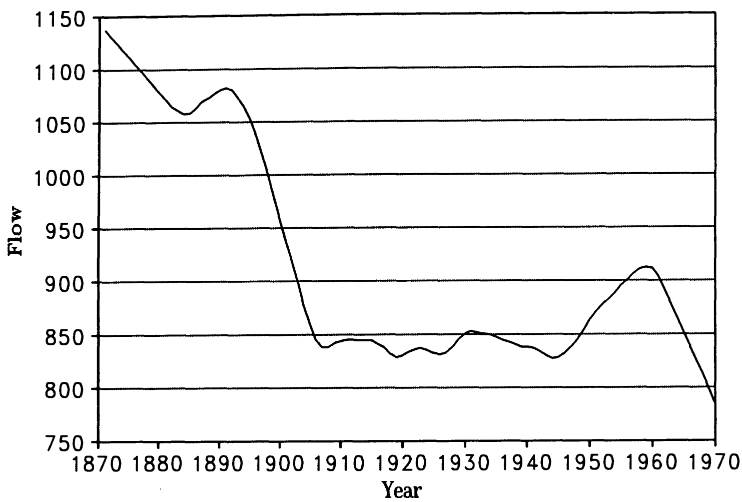

FIG. 2. *Unmodified kernel estimate for the Nile data with bandwidth* $b = 10$.

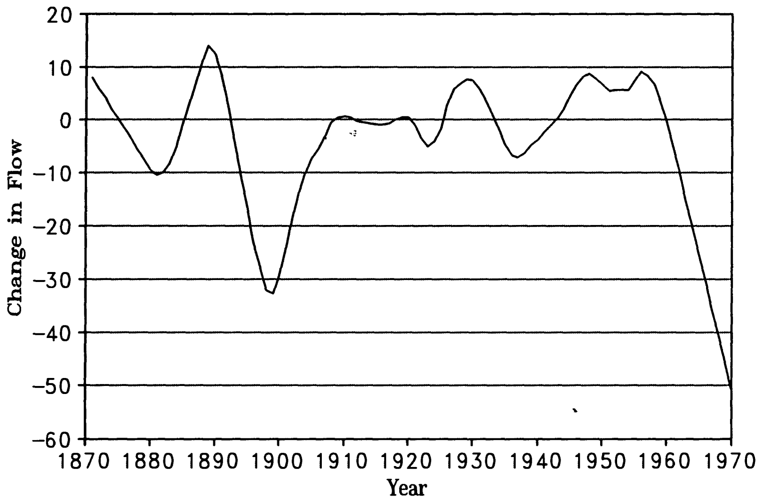

FIG. 3. *Unmodified kernel estimate for the first derivative* ($b = 10$).

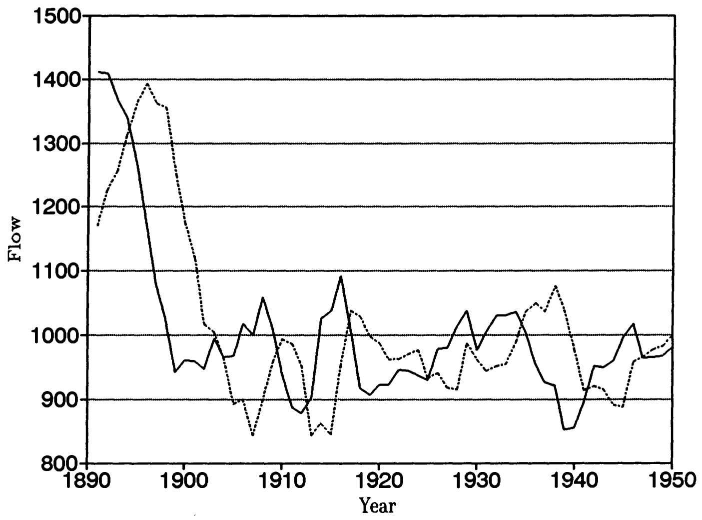

FIG. 4. *One-sided kernel estimates* $\hat{g}_-$ *(dashed) and* $\hat{g}_+$ *(solid)* ($b = 10$).

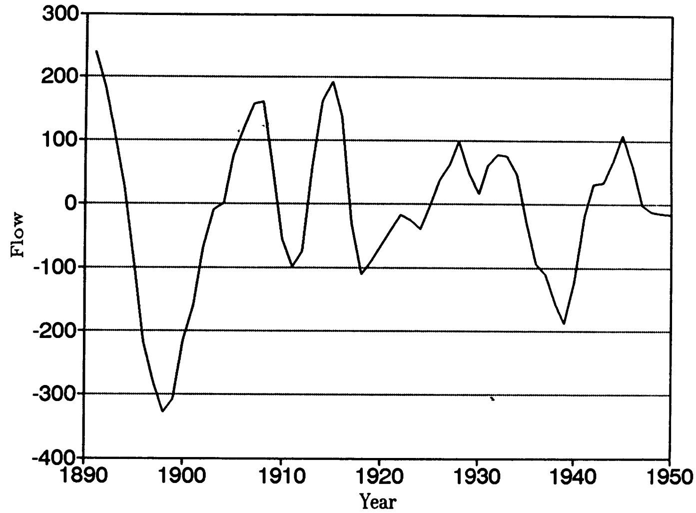

FIG. 5. *The function* $\hat{\Delta}$.

The estimator $\hat{\tau}$ for τ (2.4) is found as the minimizer of the function $\hat{\Delta}(t) = \hat{g}_+(x) - \hat{g}_-(x)$, since in this case the jump is from higher to lower levels. This function is shown in Figure 5. In practice, the set Q introduced in the definition (2.4) of $\hat{\tau}$ can be chosen as the interval from the left endpoint of the data plus bandwidth to the right endpoint of the data minus bandwidth. For the current example, this interval is [1881, 1960].

A clear minimizer stands out at $\hat{\tau} = 1898$, with the associated jump size $\hat{\Delta} = -351$. For $\alpha = 0.05$, one obtains the approximate $100(1 - \alpha)\%$ confidence intervals 1898 ± 1.04, for τ using $\hat{\sigma} = 110.5$, and -351 ± 212 for Δ. Finally, the modified curve estimator $\tilde{g}$ (4.3), which is adapted to the estimated change-point $\hat{\tau}$, is displayed in Figure 6.

A critical issue for applications is the choice of an appropriate bandwidth. While this problem in its bearing on change-point estimators (2.4) is not investigated here, the following points should be considered: Choice of a too small bandwidth will lead to spurious change-point locations, while choice of a too large bandwidth will be conservative in the sense that change-points with relatively small jump sizes will not be picked up. According to Theorem 4.1, it is advisable to choose a relatively small bandwidth for estimating change-point and jump size as compared to the bandwidth chosen for estimating the curve with adapted estimators $\tilde{g}$. One possibility is to adopt for $\nu = 0$, $\mu = 1$, $k = 2$ and $s_0 = \infty$ the ratio $b_1/b \sim (\log n/n)^{4/5}$ according to Theorem 4.1(b) as a guideline.

In order to obtain bandwidths for the "smooth" parts of the curve by global criteria like cross-validation or pilot estimation methods [compare Müller (1988), Chapter 7], one should either cut out the region $[\tilde{\tau} - b, \tilde{\tau} + b]$, where $\tilde{\tau}$ is a preliminary change-point estimator, or adjust to the preliminary change-point $\tilde{\tau}$ according to (4.3), when calculating the global error estimate, the

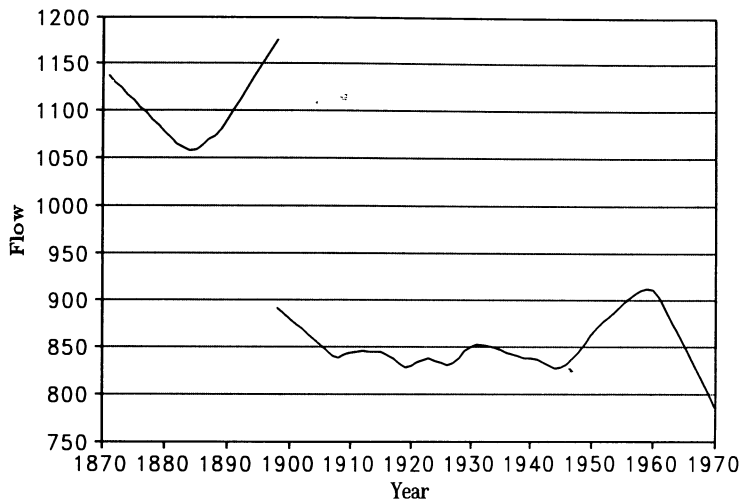

FIG. 6. *Modified kernel estimator $\tilde{g}$, adapted to the estimated change-point.*

minimizer of which is the chosen bandwidth. Failure to carry out these adjustments can lead to severely undersmoothing bandwidth choices as has been demonstrated for the case of known endpoints (Müller, 1991). It is also possible and indeed might be advisable to carry out separate bandwidth choices for the intervals separated by change-points.

There exist various applications of the proposed change-point estimators to nonparametric curve estimation. One is the segmentation of curves according to the locations of change-points in the curve itself or in higher derivatives. Often one might be able to approximate the curve in between change-points by a polynomial. Consider, for instance, polynomials of degree k between change-points of the kth derivative which might be assumed at all points τ where $\hat{\Delta}_k(\tau) > \Delta_0$, Δ_0 being a constant provided by the user. If these polynomials are fitted by least squares, subject to the constraint that they join at the change-point locations for the kth derivative in such a way that $(k - 1)$ continuous derivatives exist, then this provides a method for choosing the number and location of knots for curve fitting by variable knot regression splines.

In many finite sample situations it is not clear whether a change-point model (M2) with a discontinuity (model I) applies or whether the change-point is rather a very rapid but smooth transition; that is, corresponds to a *point of most rapid change* (model II). Consider $\nu = 0$, $\mu = 1$. In model II the change-point corresponds to $\theta = \inf\{x \in [0,1]: |g^{(1)}(x)| = \sup_y |g^{(1)}(y)|\}$ and it can be estimated by $\hat{\theta} = \inf\{x \in [0,1]: |\hat{g}^{(1)}(x)| = \sup_y |\hat{g}^{(1)}(y)|\}$. In analogy to results of Müller and Wang (1990), one can show that $|\hat{\theta} - \theta| = O_p(1/nb^5)$, which is slower than the rate $|\hat{\tau} - \tau| = O_p(b/n)$ obtained for model I. Compare Figures 3 and 5 for these two approaches.

6. Auxiliary results and proofs for Section 3. The following sequence of lemmas leads to the proof of Theorem 3.1.

LEMMA 6.1.

$$E(\zeta_n(z)) = -\Delta_\nu z^{\mu+\nu+1} K_-^{(\mu+\nu)}(0)/(\mu+\nu+1)! + o(1). \tag{6.1}$$

PROOF. Observe that, by approximating sums through integrals, analogous to formula (4.7) in Müller (1988),

$$E(\hat{g}_\pm^{(\nu)}(\tau + yb)) = \frac{1}{b^\nu}\int K_\pm^{(\nu)}(v) g(\tau + yb - vb)\, dv + O([nb^\nu]^{-1}).$$

Therefore, defining

$$\delta_\nu(y) = \frac{1}{b^\nu}\int_{-1}^{1}(K_+^{(\nu)}(v) - K_-^{(\nu)}(v)) g(\tau + yb - vb)\, dv,$$

we obtain

$$E(\hat{\delta}_\nu(y)) = \delta_\nu(y) + O([nb^\nu]^{-1}). \tag{6.2}$$

Observing (M2), (2.7), evenness of k and (K3), employing a Taylor expansion and mean values $\xi_{1n} = \tau + \xi_1(y - v)b$, $\xi_{2n} = \tau + \xi_2(y - v)b$,

$$\begin{aligned}
\delta_\nu(y) &= \frac{1}{b^\nu}\int (K_+^{(\nu)}(v) - K_-^{(\nu)}(v))\left[g(\tau + (y - v)b)(1_{\{v > y\}} + 1_{\{v \le y\}})\right] dv \\
&= \frac{1}{b^\nu}\int (K_+^{(\nu)}(v) - K_-^{(\nu)}(v))\left[\left(\sum_{j=0}^{k+\nu-1} \frac{(y - v)^j}{j!} b^j g_+^{(j)}(\tau)\right.\right. \\
&\qquad \left. + \frac{(y - v)^{k+\nu}}{(k + \nu)!} b^{k+\nu} g_+^{(k+\nu)}(\xi_{1n})\right) 1_{\{v \le y\}} \\
&\qquad \left. + \left(\sum_{j=0}^{k+\nu-1} \frac{(y - v)^j}{j!} b^j g_-^{(j)}(\tau) + \frac{(y - v)^{k+\nu}}{(k + \nu)!} b^{k+\nu} g_-^{(k+\nu)}(\xi_{2n})\right) 1_{\{v > y\}}\right] dv \\
&= \frac{1}{\nu!}\Delta_\nu \int (K_+^{(\nu)}(v) - K_-^{(\nu)}(v))(y - v)^\nu 1_{\{v \le y\}}\, dv + Q_n(y),
\end{aligned}$$

where

$$\begin{aligned}
Q_n(y) = \frac{b^k}{(k + \nu)!}\Bigg[&\int_{-1}^{y} (K_+^{(\nu)}(v) - K_-^{(\nu)}(v))(y - v)^{k+\nu} \\
&\times (g_+^{(k+\nu)}(\xi_{1n}) - g_+^{(k+\nu)}(\tau)) 1_{\{v \le y\}}\, dv \\
&+ \int_{-1}^{y} (K_+^{(\nu)}(v) - K_-^{(\nu)}(v))(y - v)^{k+\nu} \\
&\times (g_-^{(k+\nu)}(\xi_{2n}) - g_-^{(k+\nu)}(\tau)) 1_{\{v > y\}}\, dv\Bigg].
\end{aligned}$$

Observe that

$$R_n(y) = |Q_n(y) - Q_n(0)| = o(b^k y), \tag{6.3}$$

since, for instance, for the difference of the first terms on the r.h.s. of Q_n, for $y > 0$,

$$\int_{-1}^{y} \big(K_+^{(\nu)}(v) - K_-^{(\nu)}(v)\big)(-v)^{k+\nu}\big(g_+^{(k+\nu)}(\tau + \xi_1(-v)b) - g_+^{(k+\nu)}(\tau)\big)\,dv$$

$$= \int_{-1}^{y} \big(K_+^{(\nu)}(v-y) - K_-^{(\nu)}(v-y)\big)(y-v)^{k+\nu} \times \big(g_+^{(k+\nu)}(\xi_{1n}) - g_+^{(k+\nu)}(\tau)\big)\,dv,$$

and analogous calculations for $y < 0$ and for the difference of the second terms on the r.h.s. of Q_n yield (6.3). Observing, for $y \geq 0$, under (K3),

$$\frac{1}{\nu!}\int_0^y K_-^{(\nu)}(v)(y-v)^\nu\,dv = \int_0^y \left[\sum_{i=0}^{\mu-1} \frac{v^i}{i!} K_-^{(\nu+i)}(0) + \frac{v^\mu}{\mu!} K_-^{(\nu+\mu)}(\xi)\right] \frac{(y-v)^\nu}{\nu!}\,dv$$

$$= \frac{y^{\mu+\nu+1}}{(\mu+\nu+1)!}\big(K_-^{(\mu+\nu)}(0) + O(y)\big), \quad \text{as } y \to 0,$$

and analogously, for $y \leq 0$,

$$\frac{1}{\nu!}\int_y^0 K_+^{(\nu)}(v)(y-v)^\nu\,dv = \frac{-y^{\mu+\nu+1}}{(\mu+\nu+1)!}\big(K_+^{(\mu+\nu)}(0) + O(y)\big), \quad \text{as } y \to 0,$$

one obtains, noting that $K_-^{(\mu+\nu)}(0) = (-1)^{\mu+\nu} K_+^{(\mu+\nu)}(0)$ and that $(\mu+\nu)$ is odd,

$$\delta_\nu(y) - \delta_\nu(0) = \frac{-\Delta_\nu K_-^{(\mu+\nu)}(0) y^{\mu+\nu+1}}{(\mu+\nu+1)!}(1 + O(y)) + o(b^k y), \tag{6.4}$$

as $y \to 0$. The result follows. □

Lemma 6.2.

$$\operatorname{cov}(\zeta_n(z_1), \zeta_n(z_2)) = 2 z_1 z_2 \sigma^2 \int K_-^{(\nu+1)}(v)^2\,dv + O\Big(1/(nb^{2\nu+1})^{1/2(\mu+\nu)}\Big). \tag{6.5}$$

Proof. Abbreviate $\alpha = 2\nu + 1$, $\beta = (\mu+\nu+1)/(2(\mu+\nu))$ and $\gamma = 1/(2(\mu+\nu))$. Observe

$$\zeta_n(z) - E(\zeta_n(z))$$

$$= \frac{(nb^\alpha)^\beta}{b^{\nu+1}} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} \left[\left(K_+^{(\nu)}\left(\frac{\tau + (zb)/(nb^\alpha)^\gamma - v}{b}\right) - K_+^{(\nu)}\left(\frac{\tau - v}{b}\right)\right) - \left(K_-^{(\nu)}\left(\frac{\tau + (zb)/(nb^\alpha)^\gamma - v}{b}\right) - K_-^{(\nu)}\left(\frac{\tau - v}{b}\right)\right)\right] dv\, \varepsilon_i. \tag{6.6}$$

This implies

$$\operatorname{cov}(\zeta_n(z_1), \zeta_n(z_2))$$

$$= \frac{(nb^\alpha)^{2\beta}}{b^{2\nu+2}}\sigma^2$$

$$\times \sum_{i=1}^{n}\Bigg[\int_{s_{i-1}}^{s_i}\left[K_+^{(\nu)}\left(\frac{\tau + z_1 b/(nb^\alpha)^\gamma - v}{b}\right) - K_+^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv$$

$$\times \int_{s_{i-1}}^{s_i}\left[K_+^{(\nu)}\left(\frac{\tau + z_2 b/(nb^\alpha)^\gamma - v}{b}\right) - K_+^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv$$

$$- \int_{s_{i-1}}^{s_i}\left[K_+^{(\nu)}\left(\frac{\tau + z_1 b/(nb^\alpha)^\gamma - v}{b}\right) - K_+^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv$$

$$\times \int_{s_{i-1}}^{s_i}\left[K_-^{(\nu)}\left(\frac{\tau + z_2 b/(nb^\alpha)^\gamma - v}{b}\right) - K_-^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv \tag{6.7}$$

$$- \int_{s_{i-1}}^{s_i}\left[K_+^{(\nu)}\left(\frac{\tau + z_2 b/(nb^\alpha)^\gamma - v}{b}\right) - K_+^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv$$

$$\times \int_{s_{i-1}}^{s_i}\left[K_-^{(\nu)}\left(\frac{\tau + z_1 b/(nb^\alpha)^\gamma - v}{b}\right) - K_-^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv$$

$$+ \int_{s_{i-1}}^{s_i}\left[K_-^{(\nu)}\left(\frac{\tau + z_1 b/(nb^\alpha)^\gamma - v}{b}\right) - K_-^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv$$

$$\times \int_{s_{i-1}}^{s_i}\left[K_-^{(\nu)}\left(\frac{\tau + z_2 b/(nb^\alpha)^\gamma - v}{b}\right) - K_-^{(\nu)}\left(\frac{\tau - v}{b}\right)\right]dv\Bigg].$$

By the assumptions, observing the compactness of supports,

$$K_\pm^{(\nu)}\left(\frac{\tau + zb/(nb^\alpha)^\gamma - v}{b}\right) - K_\pm^{(\nu)}\left(\frac{\tau - v}{b}\right)$$

$$= K_\pm^{(\nu+1)}\left(\frac{\tau - v}{b}\right)\frac{z}{(nb^\alpha)^\gamma} \tag{6.8}$$

$$+ O\left(\frac{1}{(nb^\alpha)^{2\gamma}}\right)1_{\{K_\pm^{(\nu+1)}((\tau - v)/b) \neq 0\} \cup \{K_\pm^{(\nu+1)}((\tau + zb/(nb^\alpha)^\gamma - v)/b) \neq 0\}}.$$

Inserting this into (6.7) and observing

$$\sum_{i=1}^{n} 1_{\{K_+^{(\nu+1)}((\tau - v)/b) \neq 0\} \cup \{K_+^{(\nu+1)}((\tau + zb/(nb^\alpha)^\gamma - v)/b) \neq 0\}} = O(nb), \tag{6.9}$$

all the $O(\cdot)$ terms combined result in a summary $O(\cdot)$ term of $O(1/(nb^\alpha)^\gamma)$. Observing $2\beta - 2\gamma = 1$ and combining

$$\sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_\pm^{(\nu+1)}\left(\frac{\tau-u}{b}\right) du \int_{s_{i-1}}^{s_i} K_\pm^{(\nu+1)}\left(\frac{\tau-u}{b}\right) du = \frac{b}{n}\int K_\pm^{(\nu+1)}(v)^2\,dv + O\left(\frac{1}{n^2}\right),$$

$$(6.10)\qquad \sum_{i=1}^{n} \int_{s_{i-1}\pm}^{s_i} {}^{(\nu+1)}\left(\frac{\tau-u}{b}\right) du \int_{s_{i-1}}^{s_i} K_\mp^{(\nu+1)}\left(\frac{\tau-u}{b}\right) du = \frac{b}{n}\int K_+^{(\nu+1)}(v)K_-^{(\nu+1)}(v)\,dv + O\left(\frac{1}{n^2}\right) = O\left(\frac{1}{n^2}\right)$$

with (6.7), where the differences are substituted by the leading terms of (6.8), completes the proof. □

LEMMA 6.3. *For fixed z, $z \in [-M, M]$,*

$$\zeta_n(z) - E(\zeta_n(z)) \to_{\mathscr{D}} \mathscr{N}\left(0, 2z^2\sigma^2\int K_-^{(\nu+1)}(v)^2\,dv\right).$$

PROOF. Since $\zeta_n(z) - E(\zeta_n(z)) = \sum_{i=1}^n W_{in}\varepsilon_i$, with

$$W_{in} = \frac{(nb^\alpha)^\beta}{b^{\nu+1}}\int_{s_{i-1}}^{s_i}\left[\left(K_+^{(\nu)}\left(\frac{\tau+(zb)/(nb^\alpha)^\gamma - v}{b}\right) - K_+^{(\nu)}\left(\frac{\tau-v}{b}\right)\right) - \left(K_-^{(\nu)}\left(\frac{\tau+(zb)/(nb^\alpha)^\gamma - v}{b}\right) - K_-^{(\nu)}\left(\frac{\tau-v}{b}\right)\right)\right]dv,$$

it follows from the Lindeberg condition for the central limit theorem that asymptotic normality is implied by $\max_{1\le i\le n}|W_{in}|/(\sum W_{in}^2)^{1/2} \to 0$ as $n\to\infty$. The result follows by combining (6.5), (6.6) and (6.8). □

LEMMA 6.4. *For fixed $z_1, z_2, \dots, z_l$, $z_i \in [-M, M]$,*

$$(6.11)\qquad (\zeta_n(z_1) - E(\zeta_n(z_1)), \dots, \zeta_n(z_l) - E(\zeta_n(z_l))) \to_{\mathscr{D}} \mathscr{N}(0, A),$$

where $A = (a_{ij})_{1\le i,j\le l}$ and $a_{ij} = 2z_iz_j\sigma^2\int K_-^{(\nu+1)}(v)^2\,dv$.

LEMMA 6.5. *The sequence $\bar\zeta_n(\cdot) = \zeta_n(\cdot) - E(\zeta_n(\cdot))$ is tight.*

PROOF. We show that there exists a constant $c > 0$ such that

$$(6.12)\qquad E\left(\bar\zeta_n(z_1) - \bar\zeta_n(z_2)\right)^2 \le c(z_1 - z_2)^2$$

for n sufficiently large. According to Billingsley (1968), the moment condition

(6.12) implies tightness of $\bar{\zeta}_n$. Using the same notation as in the proof of Lemma 6.2 and defining

$$A_{\pm}(z) = \left\{ u \in [0,1]: K_{\pm}^{(\nu)}\left(\frac{\tau + zb/(nb^{\alpha})^{\gamma} - u}{b} \right) \neq 0 \right\},$$

the Lipschitz continuity of $K^{(\nu)}$ implies

$$\begin{aligned} &E\big(\bar{\zeta}_n(z_1) - \bar{\zeta}_n(z_2)\big)^2 \\ &\quad \le \frac{(nb^{\alpha})^{2\beta}}{b^{2\nu+2}} \sigma^2 \sum_{i=1}^{n} \left[\int_{s_{i-1}}^{s_i} \left[\frac{|z_1 - z_2|}{(nb^{\alpha})^{\gamma}} \big(1_{A_+(z_1) \cup A_+(z_2)} + 1_{A_-(z_1) \cup A_-(z_2)}\big)\, du \right]^2 \right. \\ &\quad \le c|z_1 - z_2|^2 \end{aligned}$$

since $2\beta - 2\gamma = 1$ and

$$\sum_{i=1}^{n} \left[\int_{s_{i-1}}^{s_i} \big(1_{A_+(z_1) \cup A_+(z_2) \cup A_-(z_1) \cup A_-(z_2)}\big)\, du \right]^2 = O\left(\frac{b}{n}\right). \qquad \square$$

PROOF OF THEOREM 3.1. Weak convergence of processes $\bar{\zeta}_n$ follows now from Theorems 8.1 and 12.3 of Billingsley, applying Lemmas 6.4 and 6.5. The moment structure of the limit process ζ is a consequence of Lemmas 6.1 and 6.2. □

The following lemma is used in the proof of Corollary 3.2.

LEMMA 6.6.

$$(nb^{2\nu+1})^{1/2}\big(\hat{\Delta}^{(\nu)}(\tau) - \Delta_{\nu}\big) \to_{\mathscr{D}} \mathscr{N}\left(0, 2\sigma^2 \int K_{-}^{(\nu)}(v)^2\, dv\right).$$

PROOF. From (4.15) and (4.16) in Müller (1988),

$$(nb^{2\nu+1})^{1/2}\big(\hat{g}_{\pm}^{(\nu)}(\tau) - E(\hat{g}_{\pm}^{(\nu)}(\tau))\big) \to_{\mathscr{D}} \mathscr{N}\left(0, \sigma^2 \int K_{\pm}^{(\nu)}(v)^2\, dv\right),$$

and from (6.1) it follows by Taylor expansions that

$$(nb^{2\nu+1})^{1/2}\big(E(\hat{g}_{+}^{(\nu)}(\tau)) - E(\hat{g}_{-}^{(\nu)}(\tau))\big) \to_P 0.$$

Therefore, the independence of $\hat{g}_{+}^{(\nu)}(\tau)$, $\hat{g}_{-}^{(\nu)}(\tau)$ implies the result. □

7. Auxiliary results and proofs for Section 4.

PROOF OF LEMMA 4.1. The proof uses similar arguments as Parzen (1962) and Müller (1985) for the estimation of modes (respectively peaks). The assumptions imply, according to Lemma 5.2, Theorem 5.1 and Remark (ii) of Müller and Stadtmüller (1987) [respectively, Theorem 11.2 and Corollary 11.2

in Müller (1988)] that

$$\sup_{Q} \left| \hat{\Delta}^{(\nu)}(x) - E\big(\hat{\Delta}^{(\nu)}(x)\big) \right| = o_p(1),$$

where Q is defined in (2.4). Further,

$$\sup_{Q \setminus [\tau - b, \tau + b]} \left| E\big(\hat{\Delta}^{(\nu)}(x)\big) \right| = o(1)$$

and

$$\sup_{[\tau - b, \tau + b]} \left| E\big(\hat{\Delta}^{(\nu)}(x)\big) \right| \to \Delta_\nu > 0$$

imply that

$$|\hat{\tau} - \tau| = O_p(b). \tag{7.1}$$

It is therefore sufficient to consider estimators $\hat{\tau} = \tau + \hat{y} b$, with

$$\hat{y} = \arg\sup_{-1 \le y \le 1} \left| \hat{\delta}_\nu(y) \right|, \tag{7.2}$$

where $\hat{\delta}_\nu$, δ_ν are defined in Section 3.

Analogous to Lemma 6.6, observing (6.2), one can show

$$\left(nb^{2\nu+1}\right)^{1/2}\left(\hat{\delta}_\nu^{(j)}(0) - \delta_\nu^{(j)}(0)\right) = O_p(1), \tag{7.3}$$

for $0 \le j \le \mu + \nu + 1$. Observe

$$\begin{aligned} \delta_\nu^{(1)}(0) - \hat{\delta}_\nu^{(1)}(0) &= \hat{\delta}_\nu^{(1)}(\hat{y}) - \hat{\delta}_\nu^{(1)}(0) \\ &= \sum_{j=1}^{\mu+\nu-1} \left[\hat{\delta}_\nu^{(j+1)}(0) - \delta_\nu^{(j+1)}(0)\right] \hat{y}^j / j! \\ &\quad + \sum_{j=1}^{\mu+\nu-1} \delta_\nu^{(j+1)}(0) \hat{y}^j / j! + \hat{\delta}_\nu^{(\mu+\nu+1)}(\xi) \hat{y}^{\mu+\nu} / (\mu+\nu)!. \end{aligned}$$

Observe now $\delta_\nu^{(j+1)}(0) \to 0$ according to (6.4), $0 \le j \le \mu + \nu - 1$, and $\hat{\delta}_\nu^{(\mu+\nu+1)}(\xi) \to c \neq 0$, which follows from $\delta_\nu^{(\mu+\nu+1)}(0) \to c \neq 0$ and the uniform convergences

$$\sup_{|y| \le 1} \left| \hat{\delta}_\nu^{(\mu+\nu+1)}(y) - E\big(\hat{\delta}_\nu^{(\mu+\nu+1)}(y)\big) \right| = o_p(1)$$

and

$$\sup_{|y| \le 1} \left| E\big(\hat{\delta}_\nu^{(\mu+\nu+1)}(y)\big) - \delta_\nu^{(\mu+\nu+1)}(y) \right| = o(1),$$

where the former follows from the above mentioned results in Müller and Stadtmüller (1987) with some minor modifications. Combining these considerations, one obtains $(nb^{2\nu+1})^{1/2} \hat{y}^{\mu+\nu} = O_p(1)$, whence the result follows. □

Considering now the proofs of Theorem 4.1 and of Lemma 4.2, let $A = [0, \min(\tau, \hat{\tau})] \cup [\max(\tau, \hat{\tau}), 1]$, let b_1, b be as in Theorem 4.1 and Lemma 4.2

and define weight functions W_i implicitly by

$$\tilde{g}^{(\nu)}(t,\tau) = \sum_{i=1}^{n} y_i W_i(t,\tau), \qquad \tilde{g}^{(\nu)}(t,\hat{\tau}) = \sum_{i=1}^{n} y_i W_i(t,\hat{\tau}).$$

Decompose the deviation between $\tilde{g}^{(\nu)}$ and $g^{(\nu)}$ as follows:

$$\begin{aligned}
&\sup_{t\in A} |\tilde{g}^{(\nu)}(t,\hat{\tau}) - g^{(\nu)}(t)| \\
&\quad \le \sup_{t\in A} |\Sigma W_i(t,\tau)\varepsilon_i| + \sup_{t\in A} |\Sigma (W_i(t,\hat{\tau}) - W_i(t,\tau))\varepsilon_i| \\
&\qquad + \sup_{t\in A} |\Sigma (W_i(t,\hat{\tau}) - W_i(t,\tau)) g(t_i)| \\
&\qquad + \sup_{t\in A} |\Sigma (W_i(t,\tau)) g(t_i) - g^{(\nu)}(t)| \\
&\quad = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.
\end{aligned} \tag{7.4}$$

The following sequence of lemmas provides bounds for these terms.

LEMMA 7.1.

$$\sup_{t\in A} |W_i(t,\hat{\tau}) - W_i(t,\tau)| = |\hat{\tau} - \tau| O\big([nb^{\nu+2}]^{-1}\big). \tag{7.5}$$

PROOF. Consider the case $0 \le t \le \min(\tau,\hat{\tau})$. The other case is analogous. Observing (K5) and that for $\min|\hat{\tau} - \tau| \le b$, $t \le \min(\tau,\hat{\tau})$, $t + b \ge \min(\tau,\hat{\tau})$, the employed kernel for both $W_i(t,\tau)$ and $W_i(t,\hat{\tau})$ can be represented as $K_-^{(\nu)}(\cdot, q)$ for possibly different $q(t,\tau)$, $q(t,\hat{\tau})$, $0 \le q \le 1$, one obtains

$$\begin{aligned}
&\sup_{0\le t\le \min(\tau,\hat{\tau})} |W_i(t,\hat{\tau}) - W_i(t,\tau)| \\
&\quad \le \sup_{0\le t\le \min(\tau,\hat{\tau})} |W_i(t,\hat{\tau}) - W_i(t,\tau)| 1_{\{|\hat{\tau}-\tau|>b\}} \\
&\qquad + \sup_{0\le t\le \min(\tau,\hat{\tau})} |W_i(t,\hat{\tau}) - W_i(t,\tau)| 1_{\{|\hat{\tau}-\tau|\le b\}} \\
&\quad \le O\big([nb^{\nu+1}]^{-1}\big) \left|\frac{\hat{\tau}-\tau}{b}\right| \left(1_{\{|(\hat{\tau}-\tau)/b|>1\}} \Big/ \frac{|\hat{\tau}-\tau|}{b} \right) \\
&\qquad + \frac{1}{b^{\nu+1}} \sup_{0\le t\le \min(\tau,\hat{\tau})} \int_{s_{i-1}}^{s_i} \left| K_-^{(\nu)}\left(\frac{t-u}{b}, \min\left(\frac{\hat{\tau}-t}{b}, 1\right)\right) \right. \\
&\qquad\qquad \left. - K_-^{(\nu)}\left(\frac{t-u}{b}, \min\left(\frac{\tau-t}{b}, 1\right)\right) \right| du,
\end{aligned}$$

whence (7.5) follows. □

LEMMA 7.2.

$$\sup_{0\le t\le 1} \left| \sum_{i=1}^{n} W_i(t,\tau) g(t_i) - g^{(\nu)}(t) \right| = O\big(b^k + [nb^\nu]^{-1}\big). \tag{7.6}$$

PROOF. First obtain corresponding pointwise results for fixed t by means of the usual integral approximation and Taylor expansion. These involve bounds on the Lipschitz continuity of $K_{\pm}^{(\nu)}(\cdot, q)$, q fixed, and on $|K_{\pm}^{(\nu)}(x, q)|$, $|\int K_{\pm}^{(\nu)}(x, q)x^k\,dx|$. According to (K5), these bounds are uniform in t. Since $g^{(k+\nu)}(\cdot)$ is also uniformly bounded, (7.6) follows. □

LEMMA 7.3.

$$\sup_{0\le\rho\le1}\ \sup_{0\le t\le1}\left|\sum_{i=1}^{n} W_i(t,\rho)\varepsilon_i\right| = O_p\Big(\big[\log n/nb^{2\nu+1}\big]^{1/2}\Big). \tag{7.7}$$

PROOF. One checks that the assumptions of appropriate extensions of Theorem 5.1A in Müller and Stadtmüller (1987) for derivatives are satisfied. Following through the proof, one finds that bounds on $\sum W_i^2(t,\tau)$, $|W_i(t,\tau)|$ are used which remain unaltered due to (K5). The bound on the r.h.s. is uniform in ρ, again due to (K5), which implies (7.7). □

PROOF OF LEMMA 4.2. Observe

$$\sup_{0\le t\le1}|\tilde{g}^{(\nu)}(t,\hat\tau)| \le O(1) + \sup_{0\le\rho\le1}\ \sup_{0\le t\le1}|\Sigma W_i(t,\rho)\varepsilon_i|,$$

and therefore, by Lemma 7.3, $\sup_{0\le t\le1}|\tilde{g}^{(\nu)}(t,\hat\tau)| = O_p(1)$. This implies

$$\int_{\min(\tau,\hat\tau)}^{\max(\tau,\hat\tau)}|\tilde{g}^{(\nu)}(x,\hat\tau) - g^{(\nu)}(x)|^p\,dx = O_p(1)|\hat\tau-\tau|. \tag{7.8}$$

Let $H(\tau,\hat\tau) = \{t_i,\ 1\le i\le n\colon\ |t_i-\tau|\le b \text{ or } |t_i-\hat\tau|\le b\}$. Observe that the cardinality of $H(\tau,\hat\tau)$ is $O(nb)$, uniformly in $\tau,\hat\tau$. Then according to Lemma 7.1,

$$\begin{aligned}
&\sup_{t\in A}\Sigma|W_i(t,\hat\tau) - W_i(t,\tau)||g(t_i)|\\
&\quad\le|\hat\tau-\tau|O\Big(\big[nb^{\nu+2}\big]^{-1}\Big)\Sigma|g(t_i)|1_{\{t_i\in H(\tau,\hat\tau)\}}\\
&\quad=|\hat\tau-\tau|O\Big(\big[b^{\nu+1}\big]^{-1}\Big).
\end{aligned}\tag{7.9}$$

Observing

$$\sup_{t\in A}|\Sigma(W_i(t,\hat\tau) - W_i(t,\tau))\varepsilon_i| \le 2\sup_{0\le\rho\le1}\ \sup_{0\le t\le1}|\Sigma W_i(t,\rho)\varepsilon_i|,$$

the result (4.7) follows from (7.8), the c_p inequality, (7.4), (7.9), (7.7) and (7.6). □

PROOF OF THEOREM 4.1. (a) Observing $[nb^{\nu}]^{-1} = O([\log n/nb^{2\nu+1}]^{1/2})$ and combining Lemmas 4.2 and 4.5, we obtain (4.5).

(b) is a direct consequence. □

Acknowledgments. I wish to thank an Associate Editor and three referees for very helpful comments which led to substantial improvements of previous versions.

REFERENCES

BHATTACHARYA, P. K. (1991). Weak convergence of the log likelihood process in the two-phase linear regression problem. In *Proceedings of R. C. Bose Symposium*. Wiley, New York. To appear.

BHATTACHARYA, P. K. and 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*. Wiley, New York.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. Roy. Statist. Soc. Ser. B* **37** 149–192.

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subject to change in time. *Ann. Math. Statist.* **35** 999–1018.

CLINE, D. B. H. and HART, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* **22** 69–84.

COBB, G. W. (1978). The problem of the Nile: Conditional solution to a change-point problem. *Biometrika* **65** 243–251.

EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882.

EDDY, W. F. (1982). The asymptotic distributions of kernel estimators of the mode. *Z. Wahrsch. Verw. Gebiete* **59** 279–290.

EPANECHNIKOV, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Theory Probab. Appl.* **14** 153–158.

GASSER, TH., MÜLLER, H. G. and MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.

HINKLEY, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56** 495–504.

MATTHEWS, D. E., FAREWELL, V. T. and PYKE, R. (1985). Asymptotic score statistic processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13** 583–591.

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

MÜLLER, H. G. (1985). Kernel estimators of zeros and of location and size of extrema of regression functions. *Scand. J. Statist.* **12** 221–232.

MÜLLER, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer, New York.

MÜLLER, H. G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78** 521–530.

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

MÜLLER, H. G. and 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.

PARZEN, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.* **33** 1065–1076.

PRIESTLEY, M. B. and CHAO, M. T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 384–392.

RICE, J. (1984a). Bandwidth choice for nonparametric kernel regression. *Ann. Statist.* **12** 1215–1230.

RICE, J. (1984b). Boundary modification for kernel regression. *Commun. Statist. Theory Methods* **13** 893–900.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14** 361–404.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47** 1–41.

WHITT, W. (1970). Weak convergence of probability measures on the function space $\mathscr{C}([0,\infty])$. *Ann. Math. Statist.* **41** 939–944.

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

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616