Journal of Multivariate Analysis **83**, 56–83 (2002)
doi:10.1006/jmva.2001.2038

Change Point Estimation by Local Linear Smoothing

Gérard Grégoire[1] and Zouhir Hamrouni

Université Joseph Fourier, Grenoble, France
E-mail: Gerard.Gregoire@imag.fr

Published online December 19, 2001

We consider the problem of estimating jump points in smooth curves. Observations (X_i, Y_i) $i=1, \dots, n$ from a random design regression function are given. We focus essentially on the basic situation where a unique change point is present in the regression function. Based on local linear regression, a jump estimate process $t \to \hat{\gamma}(t)$ is constructed. Our main result is the convergence to a compound Poisson process with drift, of a local dilated-rescaled version of $\hat{\gamma}(t)$, under a positivity condition regarding the asymmetric kernel involved. This result enables us to prove that our estimate of the jump location converges with exact rate n^{-1} without any particular assumption regarding the bandwidth h_n. Other consequences such as asymptotic normality are investigated and some proposals are provided for an extension of this work to more general situations. Finally we present Monte-Carlo simulations which give evidence for good numerical performance of our procedure.

AMS 2000 subject classifications: 62G08; 62G05; 62G20; 62G55.

Key words and phrases: nonparametric regression; local linear regression; change points; compound Poisson processes.

1. INTRODUCTION

In this paper we are interested in change point problems in regression estimation. In recent years a number of authors were concerned with these problems in a nonparametric setting. Fields of applications include econometrics, biostatistics, reliability, and signal processing. To give only one example, in theoretical economics it is well known that when shocks are produced by sudden decisions of government, variables such as investment, wages, consumption, and prices may exhibit abrupt changes. Statistical inference for change point problems in such contexts are readily of interest. Change point inference procedures are also involved on a more theoretical ground as they are related to edge effects. Usually, estimating a regression function without corrections in the neighborhood of the boundaries of the interval result in an increased bias. Note that the same remark holds when we estimate a discontinuous function by a continuous estimate. A reasonable strategy for dealing with edge effects is to consider the frontiers of the interval as discontinuity points.

[1] Research supported by the IDOPT project (CNRS-INRIA-UJF-INPG).

0047-259X/01 $35.00

A great deal of literature focuses on the parametric problem. References can be founded in the monograph edited by Müller *et al.* [28]. In this paper, our setting is a nonparametric one. We observe independent and identically distributed (X_i, Y_i), $i = 1, \ldots, n$ such that

$$Y_i = m(X_i) + \sigma(X_i)\,\epsilon_i, \qquad i = 1, \ldots, n,$$

where ϵ_i are i.i.d. variables, independent of the X_i with means zero and variances unity. The regression function $m(\cdot)$ is smooth but for some points where jumps in the function itself or in one of its derivatives are allowed.

Several smoothing approaches were proposed for dealing with discontinuities in this framework. First a semi-parametric method is based on writing the regression function in the following form

$$m(x) = m_0(x) + \sum_{i=1}^{q} \gamma_i\, p_{\tau_i}(x)$$

with m_0 a smooth function and $p_{\tau_i}(x) = p_{\tau_i}^{(k_i)}(x) = (x - \tau_i)_+^{k_i}$ is an order-k_i discontinuity function. Then, to create inference about the function m_0 and parameters (τ_i, γ_i), $i = 1, \ldots, q$ spline methods (Girard [16], Laurent and Utreras [27], Shiau [35]) have been used as well as kernel methods (Eubank and Speckman [10], Antoniadis and Grégoire [1] in the context of hazard rate functions). Another strategy (Gijbels *et al.* [21, 22]) is a two-step procedure: first the observations are smoothed as if the function m were a continuous one, getting in this way a pilot estimate $\hat{m}$ and searching for points maximizing the slope of $\hat{m}$. The second step consists of a refinement of the localization by a method based on differences between left and right averaging. Finally, the most popular approach is probably the one based on differences between left and right estimates. At any point τ, the right limit $m_+(\tau)$ is estimated using observations located at the right of point τ, and similarly $m_-(\tau)$ is estimated using data located at the left. Accordingly, an estimate for a possible jump $\gamma(\tau) = m_+(\tau) - m_-(\tau)$ follows. Papers of Müller [32], Loader [26], and Wu and Chu [39], to cite only a few, resort to this latter approach. Müller uses kernel smoothing with left kernel K_+ with support in $[-1, 0]$ and K_- defined by $K_-(x) = K_+(-x)$. The achieved convergence rate for the estimate of a change point is $n^{-(1+\varepsilon)}$ for some $\varepsilon > 0$. Wu and Chu give improvements with modified kernels and using two different bandwidths for estimating location and size parameters.

Our paper is in line with this approach of left and right smoothing. But we appeal to linear local regression instead of kernel smoothing. An essential motivation for focusing on this method is the absence of the edge effects observed with other methods such as the kernel one. In contrast with kernel method, the rate of convergence for the bias is the same near the boundaries as inside the interval. Thus, at least to the extent that we are concerned with asymptotic properties, linear local regression is likely to provide better results for change point inference than other smoothing methods. Note that in his paper, Loader takes advantage of the same idea but in an essentially different setting. Loader uses a fixed regular design and supposes that the noise is gaussian with constant variance, while in contrast we work in an unconstrained setting. More precisely we use a random design; no particular assumption is imposed on the noise distribution and the variance is allowed to depend on the location.

Several problems are of interest when abrupt changes are possibly present in data: detection of (tests for) discontinuities, estimation for the number of jumps, estimation of the locations and sizes of jumps, and reconstruction of the function. In this paper we are concerned with the estimation of the locations and sizes of jumps. This problem is well identified in the literature (see Müller [32], Loader [26], Wu and Chu [39], Wang [37], to cite only a few). We focus on the case where one jump is known to be present in the function. The method can be extended to the case of more than one jump. Also, although it would require some substantial work, we think it is possible to carry over the method to the case where the change point is a derivative (see Section 4 for proposals concerning both sorts of extension).

Our estimator $\hat{\tau}$ is defined as a point where $\hat{\gamma}(\cdot)=\hat{m}_+(\cdot)-\hat{m}_-(\cdot)$ is maximum (for the sake of simplicity we assume in this Introduction that $\gamma(\tau)>0$). Here $\hat{m}_+(t)$ and $\hat{m}_-(t)$ are the estimates of the left and right limits of $m(\cdot)$ at point t, obtained by linear local regression with a positive kernel $K_+(\cdot)$ supported on $[-1, 0]$ and K_- defined by $K_-(x)=K_+(-x)$. A basic assumption is that $K_+(0)>0$. This makes the samples of the process $\hat{\gamma}(t)$ discontinuous, but we show that it allows us to estimate the jump location τ with rate n^{-1}, while using a kernel such that $K_+(0)=K_+(-1)=0$ would result only in a rate $n^{-1+\delta}$ for some $\delta>0$ (see [10, 32, 26, 39]). Furthermore the rate n^{-1} is achieved for any rate of convergence of the bandwidth h_n such that $nh_n\to\infty$, without any other restriction. Our conclusion agrees with the result given by Loader in his particular setting. The basic tool for deriving our results is the deviation process $\mathscr{Z}_n(z)=\alpha(n, h_n)(\hat{\gamma}(\tau+(h_n/(\beta(n, h_n)))\,z)-\hat{\gamma}(\tau))$. Our location estimate $\hat{\tau}$ can be seen to satisfy $\hat{\tau}=\arg\sup \mathscr{Z}_n(z)$ when z lies in $[-M, M]$ for M large enough. We prove that, when the rescaling and dilating parameters $\alpha(n, h_n)$ and $\beta(n, h_n)$ are chosen in a convenient way, the process $\mathscr{Z}_n(z)$ converges to a compound Poisson process with an additional drift, which implies that $\hat{\tau}$ is also consistent.

The structure of the paper is as follows. Notation, recalls for local linear regression, and definitions of right and left estimates are given in Section 2. We define in Section 3 our procedure for estimating τ and provide our main results together with numerical experiments. Section 4 is devoted to numerical experiments and to provide some proposals for possible extensions of our results. Section 5 contains the proof of our central theorem; we show that the deviation process $\mathscr{Z}_n(z)$ is asymptotically equivalent to a process $\mathscr{M}_n(z)$ which converges to a compound Poisson process and derive from this the central result of the paper.

2. THE MODEL AND SOME PRELIMINARIES

This section introduces the model and notations and presents the assumptions we will deal with in the remainder of the paper.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a set of independent and identically distributed vectors. Assume X_i to be on the interval $[0, 1]$ and that

$$Y_i = m(X_i) + \sigma(X_i)\, \varepsilon_i, \tag{1}$$

where the function $m(\cdot)$ is smooth except for τ; that is

$$m(x) = m_0(x) + \gamma I_{[\tau, 1]}(x),$$

where $\gamma > 0$, $m_0(\cdot)$ is smooth (more precisely see A2 below), and $I_{[a, b]}(x) = 1$ when $x \in [a, b]$ and 0 otherwise. The ε_i are independent and identically distributed random variables with zero mean and unit variance and are independent of the X_i's.

The functions $m(\cdot)$ and $\sigma(\cdot)$ are unknown. A flexible estimation method does not make any assumption on the forms of these functions. These forms should be determined completely by the data. In other words, a nonparametric approach is preferable.

We need the following conditions

(A1) The marginal density f of covariate X is continuous and bounded away from zero.

(A2) The function $m_0(x)$ has both continuous second derivatives. It follows that the regression function $m(\cdot)$ and its first and second derivatives have left and right limits at τ.

(A3) The conditional variance $\sigma^2(x) = Var(Y \mid X = x)$ is continuous.

(A4) The jump point τ of the regression function is supposed to be in the interval (0, 1).

Assumptions (A1) and (A3) are in fact unnecessary and could be relaxed. Our results are stated under (A1) and (A3) only for the sake of the proofs' simplicity. Modifications to take into account a possible jump in $f(\cdot)$ and/or in $\sigma(\cdot)$ are rather straightforward.

In the following, for $v=0, 1, 2$, we always denote $m_-^{(v)}(\tau)=\lim_{t\nearrow\tau} m^{(v)}(t)$, $m_+^{(v)}(\tau)=\lim_{t\searrow\tau} m^{(v)}(t)$, and $\gamma^{(v)}(\tau)=m_+^{(v)}(\tau)-m_-^{(v)}(\tau)$.

$\gamma(\tau)=m_+(\tau)-m_-(\tau)$ is the jump size at the change point τ of the regression function. In this paper we study the case $\gamma(\tau)>0$. The case $\gamma(\tau)<0$ can be treated analogously. Instead of looking for the maximum of some quantity, we look for its minimum. So all the results of this paper can be proved in the case where the jump size $\gamma(\tau)$ is nonpositive.

The local linear regression is known to share the simplicity and consistency of the kernel estimators as Nadaraya–Watson or Gasser–Müller estimators but overcomes the main problems of those estimators. Especially, the local linear estimator avoids boundary effects, at least when convergence rates are concerned. The local linear estimator is based on local least squares fitting using kernel weights and can be written as a weighted sum of the response Y_i,

$$\hat{m}(x)=\frac{\sum_{i=1}^n w_i(x)\, Y_i}{\sum_{i=1}^n w_i(x)}, \tag{2}$$

where

$$w_i(x)=K\left(\frac{x-X_i}{h_n}\right)(S_2(x)-(x-X_i)\, S_1(x)), \qquad i=1, ..., n, \tag{3}$$

$$S_l(x)=\sum_{i=1}^n (x-X_i)^l\, K\left(\frac{x-X_i}{h_n}\right) \qquad l=0, 1, 2. \tag{4}$$

For further motivation and study of the local linear estimator, see Fan and Gijbels [12], Ruppert and Wand [33], Stone [36], and Cleveland [6]. In those aforementioned papers, authors showed that the local linear estimator has attractive mathematical properties. Particularly, this estimator is known to have an optimal rate of convergence; see Stone [36]. The asymptotic bias at boundary points is shown to be of the same order of magnitude as that of the interior points, whereas the Nadaraya–Watson and Gasser–Müller estimators both have an asymptotic bias on the order of $\mathcal{O}(h_n)$ instead of $\mathcal{O}(h_n^2)$ at boundary points. The fact that the estimator has no boundary effects is a very appealing property when we deal with the localization of change points, since those points can be assimilated to boundary points. Fan [11] shows that local linear estimator has an important asymptotic minimax property.

Let $K_+(\cdot)$ be a continuous kernel function with support in $[-1, 0]$ and $K_-(x)=K_+(-x)$. We will use the following notation $K_l^+=\int_{-1}^0 x^l K_+(x)\, dx$, $B_+=(K_2^+)^2-K_3^+K_1^+$, $L_l^+=\int_{-1}^0 x^l K_+^2(x)\, dx$, and $V_+=\int_{-1}^0 (K_2^+-xK_1^+)^2\, K_+^2(x)\, dx$.

Analogous formulas for K_i^-, B_-, L_l^-, and V_- follow under changing from "+" to "−".

We impose a kind of normalization to $K_+(\cdot)$ and $K_-(\cdot)$:

$$|K_2 K_0 - (K_1)^2| = 1. \tag{5}$$

In fact if $K(\cdot)$ is any kernel function, $K(\cdot)/\sqrt{|K_2 K_0 - (K_1)^2|}$ satisfies the constraint (5). For instance, in the experiment shown later on, we use the kernel given by $K_+(x) = 12\sqrt{5/19}\,(1-x^2)\,1_{[-1,0]}(x)$.

We define right-sided and left-sided regression estimates for $m(t)$ by using K_+ and K_-, respectively, instead of K in (3). Thus $\hat{m}_+(t)$ and $\hat{m}_-(t)$ are estimators of $m_+(t)$ and $m_-(t)$, respectively, and $\hat{\gamma}(t) = \hat{m}_+(t) - \hat{m}_-(t)$ estimates $\gamma(t)$.

Statements in the following will often make use of kernel M_+ defined as

$$M_+(x) = (K_2^+ - xK_1^+)\,K_+(x).$$

The similar definition for M_- implies that $M_-(x) = M_+(-x)$ and we set $M(x) = M_+(x) - M_-(x)$. The kernel M_+ is called the equivalent kernel. It was first introduced by Lejeune [25]. M_+ and M_- arise in a natural way in our developments. In the Monte-Carlo simulations we report later on, we use the normalized kernel defined by

$$K_+(x) = 12\sqrt{5/19}\,(1-x^2)\,\mathbb{I}_{[-1,0]}(x),$$

for which the equivalent kernel M_+ is given by

$$M_+(x) = \tfrac{4}{19}\,(24+45x)(1-x^2)\,\mathbb{I}_{[-1,0]}(x).$$

3. THE MAIN RESULTS

As was said in the Introduction, our procedure is based on the behaviour of the jump estimate process $t \to \hat{\gamma}(t) = \hat{m}_+(t) - \hat{m}_-(t)$. It is very natural to define the estimate of τ as a value of t that maximizes $\hat{\gamma}(t)$ over $\kappa = [h_n, 1-h_n]$:

$$\hat{\tau} = \inf\{t \in \kappa;\ \hat{\gamma}(t) = \sup_{x \in \kappa} \hat{\gamma}(x)\}. \tag{6}$$

We exclude right and left edges of the interval since 0 and 1 are themselves discontinuous points. To investigate the asymptotic behaviour of $\hat{\tau}$ we use a rescaled-dilated version of the process $\hat{\gamma}(t)$ around τ:

$$\mathscr{Z}_n(z) = \alpha(n, h_n)\left(\hat{\gamma}\left(\tau + \frac{h_n}{\beta(n, h_n)}\,z\right) - \hat{\gamma}(\tau)\right), \qquad z \in [-M, M].$$

For M large enough, we readily have $\hat{\tau} = \tau + (h_n/(\beta(n, h_n))\,\hat{z}$, with

$$\hat{z} = \arg \sup_{z \in [-M, M]} \mathscr{Z}_n(z).$$

This idea is quite similar in spirit to the one used by Eddy [9] to estimate the mode of a distribution.

We list here assumptions needed in the remainder of the paper:

H1. $K_+(0) > 0$ and $K_+(-1) = 0$.
H2. $0 < \lim_{n \to \infty} \frac{\alpha(n, h_n)}{\beta(n, h_n)} = L_4 < \infty$.
H3. $0 < \lim_{n \to \infty} \frac{\alpha(n, h_n)}{nh_n} = L_5 < \infty$.

For asymptotic results, throughout the rest of the paper we always assume that $h_n \to 0$ and $nh_n \to \infty$. We always suppose also that assumption H1 is satisfied. Note that in fact the assumption $K_+(-1) = 0$ is unnecessary. This condition is only set for the sake of simplicity. Using a positive value for $K_+(-1)$ would lead to similar results with more tedious calculations. Moreover we also denote the following: (1) $\lambda_1 = 2L_5 M_+(0)$, (2) $\lambda_2 = (L_4/L_5)\, f_X(\tau)$, (3) $\lambda_3 = L_4 M_+(0)\, f_X(\tau) = \lambda_1 \lambda_2 / 2$.

The key result for our approach is given now.

THEOREM 3.1. *Assume that $h_n \to 0$, $nh_n \to \infty$, and conditions* H1–H3 *are satisfied. Then we have*

$$\mathscr{Z}_n \Rightarrow \mathscr{Z}, \qquad on\ \mathscr{D}[-M, M], \tag{7}$$

where

$$\mathscr{Z}(z) = -\frac{\lambda_3}{f_X(\tau)}\, \gamma(\tau)\, |z| + \frac{\lambda_1}{f_X(\tau)}\, \mathscr{N}(z), \tag{8}$$

with $\mathscr{N}(z)$ defined by

$$\mathscr{N}(z) = \begin{cases} \sum_{i=1}^{N_z^+} (-\gamma(\tau) - 2\sigma(\tau)\, \varepsilon_i^+) & \text{if } z \geqslant 0, \\ \sum_{i=1}^{N_z^-} (-\gamma(\tau) + 2\sigma(\tau)\, \varepsilon_i^-) & \text{if } z < 0. \end{cases} \tag{9}$$

The sequences (ε_i^+) and (ε_i^-) are independent and built with i.i.d. variables distributed as the model error variable ε. N_z^+ and N_z^- are independent *homogeneous Poisson processes with λ_2 as parameter and are independent of the sequences (ε_i^+) and (ε_i^-).*

Thus the limit process is a bilateral compound Poisson process with an additional drift. It is interesting to notice that the process is asymmetric. One can think it rather surprising. In fact this asymmetric feature is due to the jump sign $\gamma(\tau) > 0$. The alternative assumption $\gamma(\tau) < 0$ would have changed the sign before $2\sigma(\tau)\, \varepsilon_i^+$ in both cases $z \geqslant 0$ and $z < 0$.

To get an insight into the behaviour of the process $\mathscr{Z}_n(z)$, we performed simulations for the model defined by the regression function

$$m(x) = 4\sin(5x) + 3x + I_{[0.7,\,1]}(x)$$

with random design points uniformly distributed on $[0, 1]$ and gaussian noise with standard deviation $\sigma = 0.5$. We have done 20 simulations, each with $n = 1000$ observations: the curves at left in Fig. 1 are the samples of $z \to \mathscr{Z}_n(z)$ locally around $\tau = 0.7$. The right part shows the mean of the 20 samples. The bandwidth is $h_n = 0.06$, $M = 10$, and the zoom parameters are set to $\alpha(n, h_n) = \beta(n, h_n) = nh_n$, so that $(h_n / \beta(n, h_n))\, z$ lies in $[-0.01, 0.01]$. This experimentation provides an interesting light on the theorem. The jumps and linear drift between jumps are clearly shown and the empirical mean given is coherent with what is obtained from (8), namely $E(\mathscr{Z}(z)) = -3 f_X(\tau)\, \lambda_3 \gamma(\tau)\, |z|$.

To prove Theorem 3.1, we show the same result for $\mathscr{M}_n(z)$ which is asymptotically equivalent to $\mathscr{Z}_n(z)$. The proof is rather technical and lengthy. So we postpone it to Section 5. It relies on the fact that we can write $\mathscr{M}_n(z)$ as the sum of the terms of a row in a triangular array.

This representation allows us to use standard arguments for convergence to infinitely divisible distributions.

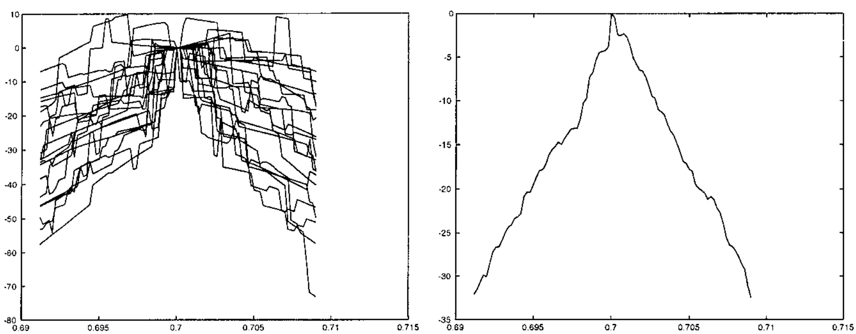

FIG. 1. Twenty samples of the process $\mathscr{Z}_n$ and their average.

COROLLARY 3.1. *Assume that* H1 *is satisfied and that* $h_n \to 0$, $nh_n \to \infty$, $\beta(n, h_n) \to \infty$, *and* $\lim_{n\to\infty} \beta(n, h_n)/(nh_n) = L$ *with* $L < \infty$. *Then*

$$\frac{\beta(n, h_n)}{h_n}(\hat{\tau}-\tau) \xrightarrow{\mathscr{D}} \mathscr{T}, \tag{10}$$

where $\mathscr{T}$ *is a* $\mathbb{R}$*-valued variable defined as*

$$\mathscr{T} = \arg\sup_z \left\{ -\frac{\lambda_3}{f_X(\tau)}\gamma(\tau)\,|z| + \frac{\lambda_1}{f_X(\tau)}\mathscr{N}(z) \right\}. \tag{11}$$

Proof of Corollary 3.1. In order for Theorem 3.1 to apply, it is enough to set $\alpha(n, h_n) = \beta(n, h_n)$. Then conditions H2 and H3 are satisfied. Hence we get the convergence of $\mathscr{Z}_n$ to $\mathscr{Z}$ on $\mathscr{D}([-M, M])$. By use of Whitt's result [38] the convergence can be extended to $\mathscr{D}(\mathbb{R})$. So we have

$$\sup_z \mathscr{Z}_n(z) \xrightarrow{\mathscr{D}} \sup_z \mathscr{Z}(z) \qquad \text{and} \qquad \arg\sup_z \mathscr{Z}_n(z) \xrightarrow{\mathscr{D}} \arg\sup_z \mathscr{Z}(z).$$

Denote by $\hat{z}(n)$ the point where the process $\mathscr{Z}_n$ reaches its maximum. By construction, we have

$$\hat{\tau} = \tau + \frac{h_n}{\beta(n, h_n)}\hat{z}(n). \tag{12}$$

We deduce that

$$\frac{\beta(n, h_n)}{h_n}(\hat{\tau}-\tau) \xrightarrow{\mathscr{D}} \hat{z} = \arg\sup_z \left\{ -\frac{\lambda_3}{f_X(\tau)}\gamma(\tau)\,|z| + \frac{\lambda_1}{f_X(\tau)}\mathscr{N}(z) \right\}.$$

For the proof to be complete, it remains to show that a.s. the values of $\mathscr{T}$ are finite. This is stated by the following lemma.

LEMMA 3.1. *The process* $(\mathscr{Z}(z))$ *satisfies:*

(a) $\mathbb{P}[\sup_{z\in\mathbb{R}} \mathscr{Z}(z) < \infty] = 1$.

(b) $\mathbb{P}[\arg\sup_{z\in\mathbb{R}} \mathscr{Z}(z) < \infty] = 1$.

Proof of Lemma 3.1.

(a) Since $\sup_{z\in\mathbb{R}} \mathscr{Z}(z) = \max(\sup_{z\leqslant 0} \mathscr{Z}(z), \sup_{z\geqslant 0} \mathscr{Z}(z))$, we can concentrate on the domain $z \geqslant 0$.

For any $z \geqslant 0$, we have

$$\mathscr{Z}(z) = -\frac{\lambda_3}{f_X(\tau)}\gamma(\tau)\,z + \frac{\lambda_1}{f_X(\tau)}\sum_{i=1}^{\mathscr{N}_z^+}(-\gamma(\tau) - 2\sigma(\tau)\,\varepsilon_i^+).$$

Hence, we get

$$\sup_{z \geqslant 0} \mathscr{X}(z) = \frac{\lambda_1}{f_X(\tau)} \sup_{z \geqslant 0} \left\{ -\frac{\lambda_2}{2} z - \sum_{i=1}^{\mathcal{N}_z^+} \left(1 + \frac{2\sigma(\tau)}{\gamma(\tau)} \varepsilon_i^+ \right) \right\}. \tag{13}$$

Denote by $T_1^+, T_2^+, \ldots$ the dates of occurrence of the Poisson process $\mathcal{N}_z^+$, $T_0^+ = 0$, and $E_i^+ = T_{i+1}^+ - T_i^+$.

From (13), we see that $\mathscr{X}(z)$ is decreasing between consecutive points of the process $\mathscr{X}(z)$. Consequently the supremum of $\mathscr{X}(z)$ is reached at the left edge of a segment $[T_i, T_{i+1})$. Hence

$$\begin{aligned} \sup_{z \geqslant 0} \mathscr{X}(z) &= \sup_{j \geqslant 0} \{ \mathscr{X}(T_j^+) \}, \\ &= \frac{\lambda_1}{f_X(\tau)} \sup_{j \geqslant 0} \left\{ \sum_{i=0}^{j} \Theta_i^+ \right\}, \end{aligned}$$

where $\Theta_0^+ = 0$, and for $i \geqslant 1$, $\Theta_i^+ = -(\lambda_2/2)\, E_i^+ - (1 + (2\sigma(\tau)/\gamma(\tau))\, \varepsilon_i^+)$.

Consider the random walk (W_j^+) defined by $W_j^+ = \sum_{i=0}^{j} \Theta_i^+$. Since $E\Theta_i^+ = -3$, it turns out that $\sup_i W_i^+$ is a.s. finite.

(b) Let $\mathscr{T} = \arg\sup_{z \in \overline{\mathbb{R}}_+} \mathscr{X}(z)$, where $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$. We have

$$\begin{aligned} \mathbb{P}[\mathscr{T} = +\infty] &= \mathbb{P}[\forall u \geqslant 0;\ \sup_{z \in [0,u]} \mathscr{X}(z) \leqslant \sup_{z > u} \mathscr{X}(z)], \\ &= \mathbb{P}[\forall u \geqslant 0;\ \sup_{z \in [0,u]} \mathscr{X}(z) - \mathscr{X}(u) \leqslant \sup_{z > u} \mathscr{X}(z) - \mathscr{X}(u)], \\ &= \mathbb{P}[\forall u \geqslant 0;\ \sup_{z \in [0,u]} -\mathscr{X}_1(z) \leqslant \sup_{z > 0} \mathscr{X}_2(z)], \\ &= \mathbb{P}[-\inf_{z \geqslant 0} \mathscr{X}_1(z) \leqslant \sup_{z > 0} \mathscr{X}_2(z)], \end{aligned}$$

where $\mathscr{X}_1(z)$ and $\mathscr{X}_2(z)$ are two independent copies of $\mathscr{X}(z)$.

Now (see, e.g., Feller [14, Chap. XII]), $\mathscr{X}_1(z) \to -\infty$ as $z \to +\infty$. Hence $\inf_{z \geqslant 0} \mathscr{X}_1(z) = -\infty$. It turns out from (a) that

$$\mathbb{P}[\mathscr{T} = +\infty] = \mathbb{P}[\sup_{z > 0} \mathscr{X}_2(z) \geqslant +\infty] = 0. \quad \blacksquare \tag{14}$$

An important by-product of Corollary 3.1 is obtained when plugging $\beta(n, h_n) = nh_n$ in Corollary 3.1. It says that the rate n^{-1} can be achieved for any sequence of bandwidths h_n such that $nh_n \to \infty$.

THEOREM 3.2. *Assume that* H1 *is satisfied and that* $h_n \to 0$, $nh_n \to \infty$; *then*

$$n(\hat{\tau}-\tau) = \mathcal{O}_P(1).$$

We now derive asymptotic normality for $\hat{\tau}$ and $\gamma(\hat{\tau})$.

COROLLARY 3.2. *Assume that* $h_n \to 0$, $nh_n \to \infty$, *and* $nh_n^5 \to 0$. *Assume also that* $E(|\varepsilon|^{2+\delta}) < \infty$ *for some* $\delta > 0$ *and condition* H1 *is satisfied. Then*

$$\sqrt{nh_n}\,(\hat{\gamma}(\hat{\tau})-\gamma(\tau)) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \frac{2\sigma^2(\tau)}{f_X(\tau)} V_+^2\right).$$

Proof of Corollary 3.2. We have

$$\begin{aligned}\sqrt{nh_n}\,(\hat{\gamma}(\hat{\tau})-\gamma(\tau)) &= \sqrt{nh_n}\,(\hat{\gamma}(\hat{\tau})-\hat{\gamma}(\tau))+\sqrt{nh_n}\,(\hat{\gamma}(\tau)-\gamma(\tau)),\\ &= \frac{\sqrt{nh_n}}{\alpha(n, h_n)}\,\mathscr{Z}_n(\hat{z}(n))+\sqrt{nh_n}\,(\hat{\gamma}(\tau)-\gamma(\tau)).\end{aligned}$$

Hence, choosing the dilating parameter $\alpha(n, h_n)$ in such a way that $\sqrt{nh_n}/\alpha(n, h_n) \to 0$, we get

$$\sqrt{nh_n}\,(\hat{\gamma}(\hat{\tau})-\gamma(\tau)) = o_P(1)+\sqrt{nh_n}\,(\hat{\gamma}(\tau)-\gamma(\tau)) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \frac{2\sigma^2(\tau)}{f_X(\tau)} V_+^2\right),$$

where the normality follows from standard arguments concerning the usual smooth case. ∎

When using consistent estimators for $f_X(\tau)$ and $\sigma^2(\tau)$, we obtain the following corollary

COROLLARY 3.3. *Let* $\hat{f}_X(\cdot)$ *and* $\hat{\sigma}^2(\cdot)$ *be consistent estimators of* $f(\cdot)$ *and* $\sigma^2(\cdot)$ *such as* $\sup_{t\in(0,1)} |\hat{f}_X(t)-f_X(t)| = o_p(1)$ *and* $\sup_{t\in(0,1)} |\hat{\sigma}^2(t)-\sigma^2(t)| = o_p(1)$.

The estimator $\hat{f}_X(\cdot)$ *is supposed to be nonnegative. Under the same assumptions as in Corollary* 3.2 *we have*

$$\sqrt{nh_n}\,\frac{\sqrt{\hat{f}_X(\hat{\tau})}\,(\hat{\gamma}(\hat{\tau})-\gamma(\tau))}{(2\hat{\sigma}^2(\hat{\tau}))^{1/2}} \xrightarrow{\mathscr{L}} \mathcal{N}(0, V_+).$$

Note that, for instance, estimation of the density $f_X(\cdot)$ can be done by the usual kernel method, and for the variance function $\sigma^2(\cdot)$ the estimators proposed by Müller and Stadmüller [30] are convenient. Notice also that the convergence conditions laid down on $\hat{f}_X(\cdot)$ and $\sigma^2(\cdot)$ above could be weakened by making use of the high rate of convergence of $\hat{\tau}$.

4. SIMULATIONS: EXTENSIONS AND RELATED ISSUES

Simulations. Tables I and II provide results of simulations performed for the following example:

$$m(x) = 4\sin(5x) + 3x + \mathbb{I}_{[0.7,\,1]}(x).$$

Samples of 200 points (x_i, y_i) were generated for Table I, while we used 500 points for Table II. In both cases, for the estimation of MSE the experiment is replicated 200 times, thus leading to 200 observations of $\hat{\tau}$ and $\hat{\gamma}(\hat{\tau})$. In both cases the design points are random and uniformly distributed on $[0, 1]$, and the standard error for the noise ε is $\sigma = 0.5$. In Tables I and II, B^2 stands for the squared bias and V for the variance. Notice that, as expected, $\hat{\tau}$ has a better convergence rate than $\hat{\gamma}(\hat{\tau})$. Mean squared error is dominated by the variance part. Comparisons of results given in Tables I and II seem to indicate a rate as fast as n^{-2} for the convergence of $\hat{\tau}'$ MSE. It appears that for most h_n-values this rate is almost the same. Figure 2 shows two histograms for the 200 values of $\hat{\tau}_n$ we obtained in the case of experiments with 200 points (x_i, y_i). Note that these distributions were obtained for bandwidths 0.15 and 0.17 and look

TABLE I

$n = 200$

	$\hat{\tau}$		$\hat{\gamma}(\hat{\tau})$	
h_n	B^2	V	B^2	V
0.07	0.0283	0.0528	0.3496	0.3468
0.09	0.0167	0.0435	0.1018	0.1389
0.11	0.0076	0.0329	0.0342	0.0670
0.13	0.0040	0.0287	0.0144	0.0538
0.15	0.0021	0.0223	0.0041	0.0461
0.17	0.0015	0.0223	0.0004	0.0413
0.19	0.0018	0.0249	0.0003	0.0364
0.21	0.0033	0.0294	0.0291	0.0311
0.23	0.0066	0.0337	0.0063	0.0266
0.25	0.0174	0.0389	0.0077	0.0222

TABLE II

$n = 500$

	$\hat{\tau}$		$\hat{\gamma}(\hat{\tau})$	
h_n	B^2	V	B^2	V
0.07	0.0011	0.0150	0.0033	0.0439
0.09	0.0002	0.0053	0.0003	0.0392
0.11	0.0001	0.0037	0.0000	0.0338
0.13	0.0000	0.0020	0.0008	0.0312
0.15	0.0000	0.0013	0.0029	0.0279
0.17	0.0000	0.0011	0.0070	0.0249
0.19	0.0000	0.0012	0.0143	0.0221
0.21	0.0001	0.0030	0.0259	0.0193
0.23	0.0009	0.0078	0.0402	0.0160
0.25	0.0066	0.0142	0.0489	0.0127

somewhat similar. The histograms for $h = 0.13$ and 0.19 are also not very different from those we show.

Extensions and related issues. The method developed in this paper could be generalized in a fairly natural way to νth derivatives of regression functions. This can be done by using local polynomial regression rather than L.L.R. More precisely we could adjust locally a polynomial with degree $q \geqslant \nu$ by minimizing

$$\sum [Y_i - (a_0 + a_1(x - X_i) + \cdots + a_q(x - X_i)^q)]^2 \, K((x - X_i)/h).$$

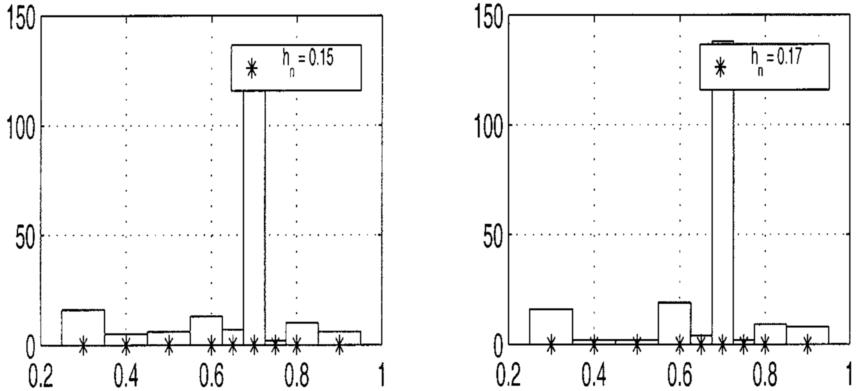

FIG. 2. Histograms for values of the estimate $\hat{\tau}_n$ with bandwidths 0.15 (left) and 0.17 (right).

Let $(\hat{a}_0, \ldots, \hat{a}_q)$ be the minimizer; then $\nu!\hat{a}_\nu$ is a natural estimate of the derivative $m^{(\nu)}(\cdot)$. Thus we can mimic the method developed for the regression function, that is estimate the jump $\gamma^{(\nu)}(x) = m_+^{(\nu)}(x) - m_-^{(\nu)}(x)$ by $\hat{\gamma}^{(\nu)}(x) = \nu!\,(\hat{a}_{\nu,+}(x) - \hat{a}_{\nu,-}(x))$. We can then define the dilated-rescaled version of the process $\hat{\gamma}^{(\nu)}(t)$,

$$\mathscr{Z}_n^{(\nu)}(z) = \alpha(n, h_n)\left(\hat{\gamma}^{(\nu)}\left(\tau + \frac{h_n}{\beta(n, h_n)}\, z\right) - \hat{\gamma}^{(\nu)}(\tau)\right),$$

and proceed along the lines of this paper. We think that with *ad hoc* modifications our arguments will work and we conjecture that the rate $n^{-1/(1+2\nu)}$ will be achieved for the location estimate. However, although the principle is clear, the proof requires tedious developments and is very complicated. This could be the subject of another work.

Note that our procedure makes sense when a change point actually exists, that is there is a τ with $\gamma(\tau) \neq 0$. When there is no change point, our procedure still provides an estimate $\hat{\tau}$, but $\hat{\gamma}(\hat{\tau})$ is not significant and $\hat{\tau}$ has no convergence property. In fact, when the existence of a jump is questionable, we have first to apply a detection or test procedure. A fairly natural test can be based on Corollary (4.2) or (4.3). See also [19] for two procedures for testing for a jump at τ or in a neighborhood of τ. A global test for the existence of a jump in $[0, 1]$ can be based on the asymptotic distribution, under the null hypothesis, of the sup of the process $t \to \hat{\gamma}(t)$ with an accurate normalization (see [24]). See also [29] for a related work when the design is an equally spaced fixed one and the noise variance is constant.

We can also be concerned with the case where there are $p > 1$ change points. When p is known, we only have to look for the locations $(\hat{\tau}_1, \ldots, \hat{\tau}_p)$ of the p highest values of $\hat{\gamma}(t)$. The first one is searched for in $A_1 = [h, 1-h]$, the second one in $A_2 = A_1 \setminus [\hat{\tau}_1 - h, \hat{\tau}_1 + h]$, and so on. Note that, for n large enough, since $h_n/\beta(n, h_n) = o(1)$, $\mathscr{Z}_n(z)$ involves only observations in a neighborhood of τ. This implies that asymptotically the processes $\mathscr{Z}_n^{(i)}(z_i) = \alpha(n, h_n)(\hat{\gamma}(\tau_i + (h_n/(\beta(n, h_n)))\, z_i) - \hat{\gamma}(\tau_i))$ will be independent and that the results of the previous section apply to each discontinuity.

When the number p is unknown, the key point is to have an a.s. consistent estimate of p. If we have such an estimate, say $\hat{p}_n$, then there exists a.s. an n_0 for which $\hat{p}_n = p$, $n \geqslant n_0$. Thus we are led to the case "p known." To get such an estimate $\hat{p}_n$, we adapt a procedure initially proposed by Yin [42]. See Wu and Chu [39] and Qiu [34] for this type of approach in the setting of an equally spaced fixed design.

To estimate the number p, we write the interval $[0, 1]$ as a set of elementary intervals whose lengths are approximately $1/2h_n$, namely $[0, 1] = \bigcup [g_j, g_{j+1})$ with

$$G = \left[\frac{1}{2h_n}\right] + 1, \qquad g_j = \frac{j}{G}, \qquad j = 1, 2, ..., G-1,$$

where $[x]$ is the largest integer less than or equal to x.

Let $S_0 > 0$ be a threshold value such that $|\gamma(\tau_i)| > S_0$, $i = 1, ..., p$. The set J is defined by

$$J = \{g_j, j = 1, ..., G-2, \|\hat{\gamma}(\cdot)\|_{[g_j, g_{j+1})} > S_0\}, \tag{15}$$

where $\|h(\cdot)\|_A = \sup_{t \in A} |h(t)|$. Now we define

$$\mathscr{C} = \{C = \{g_j\}_{j=r, ..., r+k} \mid g_j \in J, j = r, ..., r+k, g_{r-1} \notin J \text{ and } g_{r+k+1} \notin J\}.$$

Clearly $\mathscr{C}$ defines a set of areas $\hat{\mathscr{E}}_i$, each of which is obtained by putting together elementary contiguous intervals where $|\hat{\gamma}(\cdot)|$ exceeds the threshold value:

$$\hat{\mathscr{E}}_i = \left[\min C_i, \max C_i + \frac{1}{G}\right], \qquad C_i \in \mathscr{C}.$$

Finally our estimates for p and τ_i are given by

$$\hat{p} = \#\mathscr{C},$$

and

$$\hat{\tau} = \arg\max_{t \in \hat{\mathscr{E}}_i} |\hat{\gamma}(t)|, \qquad i = 1, ..., \hat{p}$$

We denote $\mathscr{D}_n = \bigcup_{i=1}^{p} [\tau_i - h_n, \tau_i + h_n]$ (discontinuity set) and list below the assumptions needed to get a.s. consistency of $\hat{p}$.

K1. $K_+(\cdot)$ is a continuous function on $[-1, 0]$.

M1. There exists $s > 2$ such that $E|\varepsilon_1|^s < \infty$.

M2. $f_X(\cdot)$ is continuous and strictly positive on $[0, 1]$.

M3. The function $m(\cdot)$ is continuous and bounded on each interval $[\tau_i, \tau_{i+1})$, $i = 1, ..., p$.

We then have the following results.

THEOREM 4.1. *Suppose that* K1, M1–M3 *hold true. Suppose that the bandwidth* h_n *satisfies* $n^\beta h_n \to \infty$ *for some* $0 < \beta < 1 - s^{-1}$ *and that* $\lambda > 0$

exists such that $\sum_n h_n^{\lambda} < \infty$. *Suppose also that* $S_0 < \min_{i=1,\dots,p} |\gamma(\tau_i)|$. *We then have*

(a) $\|\hat{\gamma}(\cdot)\|_{\mathscr{D}_n} = o(1)$, *a.s.*

(b) $\inf_{1 \leqslant i \leqslant p} \liminf_n \|\hat{\gamma}(\cdot)\|_{]\tau_i - h_n, \tau_i + h_n[} > S_0$, *a.s.*

As a consequence of the above result we get the a.s. consistency of $\hat{p}$ and $\hat{\tau}_i$, $i = 1, \dots, p$.

THEOREM 4.2. *Under the assumptions of Theorem* 4.1 *we have*

(a) $\hat{p} \to p$, *a.s.*

(b) $\forall 1 \leqslant i \leqslant p$, $\hat{\tau}_i \to \tau_i$, *a.s.*

Thus, under the assumptions K1, M1–M3, and with $S_0 < \min_{i=1,\dots,p} |\gamma(\tau_i)|$ we are ensured that a.s. there exists an n_0 such that $\hat{p}_n = p$, $n \geqslant n_0$. See [24] or [20] for details and proofs of the above results, which will appear elsewhere.

Bandwidth selection. A relevant point of practical importance is how to choose the bandwidth. The issue is crucial as the result is highly dependent on an accurate smoothing parameter. As far as we know, few authors have tackled this matter.

To start with, we suggest a naive way to proceed, which can be seen as a rough approach. Usually, near the optimal bandwidth the estimates and the error are fairly stable. This is the case for the simulations we show below and generally for other experiments we have not reported here. So a sensible strategy could be to carry out the estimation procedures with an extended grid of h_n values. A relative stability for the estimates can be interpreted as the fact that the bandwidth used is a reasonable one.

A strategy more satisfactory from a theoretical point of view could be based on the accuracy of the estimation of the regression function $m(\cdot)$ on the whole interval $[0, 1]$. We suppose that we have selected an error measure (ISE, MISE, ASE, MASE, etc.) for the accuracy of any estimate $\hat{m}(\cdot)$ on $[0, 1]$ and that an estimate I of this measure, e.g., based on cross-validation or bootstrap, is available. For each value h of a grid M, we compute an estimate $\hat{\tau}(h)$, derive $\hat{m}_h$ the resulting estimate for m on $[0, \hat{\tau})$ and $[\hat{\tau}, 1]$, and finally get the estimate $I(h)$ of the error measure. Then we chose $\hat{h}$ as the bandwidth minimizing $I(h)$ when h runs over M. Basically the idea is that an accurate estimate of τ is necessary to get an accurate estimate of m on the whole interval $[0, 1]$. The estimate $\hat{\tau}$ is likely to be close to τ when $\hat{m}$ is close to m on $[0, 1]$. This procedure has been suggested by several authors in the past ten years. To the best of our knowledge only Wu and Chu [40] provide a precise implementation of the procedure and present theoretical results. In their work, they use the kernel method and are concerned with the case where the number of discontinuities is unknown. Although we cannot provide precise theoretical arguments, we think it likely that the method will work in our setting of L.L.R. smoothing with random design. As a last point, note that in this outline, since we are using L.L.R. we do not take care of boundary effects at points 0 and 1. Nevertheless, to improve the procedure we could also generate pseudo-data as suggested by Wu and Chu [41] and Hall and Wehrly [23].

Finally we briefly describe the bootstrap method proposed by I. Gijbels and A. C. Goderniaux [15]. Once again we consider the case where it is known that m is regular except for one jump located at τ. The method is designed for the two-step estimate defined in [21, 22]; nevertheless we think that it can be adapted to our setting. Given observations (X_i, Y_i), $i=1, ..., n$, where $0 \leqslant X_1 \leqslant X_2 \leqslant \cdots \leqslant X_i \leqslant X_{i+1} \leqslant \cdots \leqslant X_n \leqslant 1$, let i_0 be defined by $X_{i_0} \leqslant \tau < X_{i_0+1}$ and $\hat{i}_0(h)$ the estimate of i_0 using bandwidth h, $\hat{\tau} = (X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1})/2$. Then the unknown function m is estimated on both the intervals $[0, \hat{\tau})$ and $[\hat{\tau}, 1]$ using an independent data driven procedure. Next the authors get residuals $\hat{\varepsilon} = Y_i - \hat{m}(X_i)$, $i=1, ..., n$, which they center: $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \bar{\varepsilon}$, $\bar{\varepsilon} = (1/n) \sum \hat{\varepsilon}_i$. The next step consists in bootstrapping the centered residuals to get bootstrapped observations Y_i^* from which an $\hat{i}_0^*$ is derived. Repeating this step a large number of times will provide an estimate of $P(\hat{i}_0 = i_0)$. The bandwidth $\hat{h}$ is chosen as the one which gives the best value for this estimate. The authors provide extensive simulations which show the procedure works quite well. ∎

5. PROOF OF THEOREM 3.1

We first give some auxiliary lemma that will prove to be useful through the rest of the section. The arguments for the proof of this lemma can be seen as extensions of those leading to the asymptotic expansions of $S_l(x)$ and $\sum_{i=1}^n w_i(x)$. Recall [13] that

$$S_l(x) = nh_n^{l+1} f(x)(K_l + o_P(1)), \tag{16}$$

$$\sum_{i=1}^n w_i(x) = n^2 h_n^4 f^2(x)(K_2 K_0 - (K_1)^2)(1 + o_P(1)), \tag{17}$$

where the functions $o_P(1)$ can be defined in a way independent from x. Denote

$$D_l^+(x) = \sum_{i=1}^n (x - X_i)^l \, K_+\left(\frac{x - X_i}{h_n}\right)(y_i - m_+(\tau)).$$

LEMMA 5.1.

(a) *Assume that $h_n \to 0$ and $nh_n \to \infty$; then*

$$D_l^+(\tau) = \mathcal{O}(nh_n^{l+2}) + \mathcal{O}_P\left(\sqrt{nh_n^{2l+1}}\right).$$

(b) *Assume that $h_n \to 0$, $nh_n \to \infty$, and $h_n/\beta(n, h_n) \to 0$ is satisfied; then*

$$S_l^+(\tau + h_n y) = nh_n^{l+1} f_X(\tau)(K_l^+ + o_p(1)),$$

$$\sum_{i=1}^{n} w_i^+(\tau + h_n y) = n^2 h_n^4 f_X^2(\tau)(1 + o_p(1)),$$

where $y = z/\beta(n, h_n)$.

(c) *Under* H1 *and the conditions given in* (b)

$$S_l^+(\tau + h_n y) - S_l^+(\tau) = \mathcal{O}\left(\frac{nh_n^{l+1}}{\beta(n, h_n)}\right) + \mathcal{O}_P\left(\sqrt{\frac{nh_n^{2l+1}}{\beta(n, h_n)}}\right).$$

Proof of Lemma 5.1. We only sketch the proof. For more precise arguments see [18].

Writing down $E[D_l^+(\tau)]$ and $Var[D_l^+(\tau)]$ we easily get (a). The proof of (b) follows from a fairly straightforward modification of the one of (16) and (17). The proof of (c) makes use of the following asymptotic behaviours:

(1) $\int_{-1-y}^{-1} (x+y)^l K_+(x+y) f_X(\tau - h_n x)\, dx = \mathcal{O}(1/\beta^2(n, h_n))$,
(2) $\int_{-y}^{0} x^l K_+(x) f_X(\tau - h_n x)\, dx = \mathcal{O}(1/\beta(n, h_n))$,
(3) $\int_{-1}^{-y} [(x+y)^l K_+(x+y) - x^l K_+(x)] f_X(\tau - h_n x)\, dx = \mathcal{O}(1/\beta(n, h_n))$.

∎

5.1. *An Approximation Result for the Deviation Process*

Recall that $\mathcal{Z}_n(z)$, the rescaled-dilated version of our process $\hat{\gamma}(t)$ around τ, is defined by:

$$\mathcal{Z}_n(z) = \alpha(n, h_n)\left(\hat{\gamma}\left(\tau + \frac{h_n}{\beta(n, h_n)} z\right) - \hat{\gamma}(\tau)\right), \qquad z \in [-M, M].$$

The aim of this section is to show that $\mathcal{Z}_n(z)$ is asymptotically equivalent to the sum of the row in a triangular array.

THEOREM 5.1. *Under the conditions* H1, H2, *and* H3, *and with* $h_n \to 0$, $nh_n \to \infty$

$$\mathscr{Z}_n(z) = \frac{(1+o_P(1))}{f_X(\tau)} (\mathscr{M}_n^+(z) - \mathscr{M}_n^-(z)) + o_P(1),$$
$$= \frac{(1+o_P(1))}{f_X(\tau)} \mathscr{M}_n(z) + o_P(1),$$

where $\mathscr{M}_n(z) = \mathscr{M}_n^+(z) - \mathscr{M}_n^-(z)$ *and*

$$\mathscr{M}_n^{\pm}(z) = \frac{\alpha(n, h_n)}{nh_n} \sum_{i=1}^{n} \left(M_{\pm} \left(\frac{\tau + h_n z/\beta(n, h_n) - X_i}{h_n} \right) - M_{\pm} \left(\frac{\tau - X_i}{h_n} \right) \right)$$
$$\times (Y_i - m_{\pm}(\tau)).$$

We prove later on in this section that $\mathscr{M}_n(z)$ converges to a compound Poisson process with drift, and of course the same result holds true for $\mathscr{Z}_n(z)$.

We first give a basic decomposition for $\mathscr{Z}_n(z)$ from which the asymptotic equivalence of $\mathscr{Z}_n(z)$ and $\mathscr{M}_n(z)$ will be derived.

We have

$$\mathscr{Z}_n(z) = \mathscr{Z}_n^+(z) - \mathscr{Z}_n^-(z),$$

where

$$\mathscr{Z}_n^+(z) = \alpha(n, h_n)(\hat{m}_+(\tau + h_n z/\beta(n, h_n)) - \hat{m}_+(\tau))$$

and $\mathscr{Z}_n^-$ is defined analogously using $\hat{m}_-$.

For convenience, throughout the remainder of the paper we set $y = z/\beta(n, h_n)$.

$$\mathscr{Z}_n^+(z) = \alpha(n, h_n) \left[\frac{\sum_{i=1}^n w_i^+(\tau + h_n y)\, y_i}{\sum_{i=1}^n w_i^+(\tau + h_n y)} - \frac{\sum_{i=1}^n w_i^+(\tau)\, y_i}{\sum_{i=1}^n w_i^+(\tau)} \right],$$
$$= \alpha(n, h_n) \frac{\sum_{i=1}^n (w_i^+(\tau + h_n y) - w_i^+(\tau))(y_i - m_+(\tau))}{\sum_{i=1}^n w_i^+(\tau + h_n y)}$$
$$- \alpha(n, h_n) \left(\frac{\sum_{i=1}^n w_i^+(\tau)(y_i - m_+(\tau))}{\sum_{i=1}^n w_i^+(\tau)} \right)$$
$$\times \left(\frac{\sum_{i=1}^n w_i^+(\tau + h_n y) - w_i^+(\tau)}{\sum_{i=1}^n w_i^+(\tau + h_n y)} \right),$$
$$= A_n - B_n \times C_n \quad \text{(say)}. \tag{18}$$

Now we have

$$w_i^+(\tau+h_n y)-w_i^+(\tau)=[S_2^+(\tau+h_n y)-S_2^+(\tau)]\,K_+\left(\frac{\tau-X_i}{h_n}\right)$$
$$-[S_1^+(\tau+h_n y)-S_1^+(\tau)](\tau-X_i)\,K_+\left(\frac{\tau-X_i}{h_n}\right)$$
$$+S_2^+(\tau+h_n y)\left[K_+\left(\frac{\tau+h_n y-X_i}{h_n}\right)-K_+\left(\frac{\tau-X_i}{h_n}\right)\right]$$
$$-S_1^+(\tau+h_n y)\left[(\tau+h_n y-X_i)\,K_+\left(\frac{\tau+h_n y-X_i}{h_n}\right)\right.$$
$$\left.-(\tau-X_i)\,K_+\left(\frac{\tau-X_i}{h_n}\right)\right].$$

and consequently, the numerators of A_n and C_n can be written as follows:

$$\sum_{i=1}^{n} w_i^+(\tau+h_n y)-w_i^+(\tau)$$
$$=[S_0^+(\tau+h_n y)-S_2^+(\tau)]\,S_0^+(\tau)-[S_1^+(\tau+h_n y)-S_1^+(\tau)]\,S_1^+(\tau)$$
$$+[S_2^+(\tau+h_n y)-S_0^+(\tau)]\,S_2^+(\tau+h_y)-[S_1^+(\tau+h_n y)-S_1^+(\tau)]$$
$$\times S_1^+(\tau+h_n y),\tag{19}$$

and

$$\sum_{i=1}^{n} (w_i^+(\tau+h_n y)-w_i^+(\tau))(y_i-m_+(\tau))$$
$$=[S_2^+(\tau+h_n y)-S_2^+(\tau)]\,D_0^+(\tau)-[S_1^+(\tau+h_n y)-S_1^+(\tau)]\,D_1^+(\tau)$$
$$+[D_0^+(\tau+h_n y)-D_0^+(\tau)]\,S_2^+(\tau+h_n y)$$
$$-[D_1^+(\tau+h_n y)-D_1^+(\tau)]\,S_1^+(\tau+h_n y),\tag{20}$$

Thus, A_n, B_n, and C_n are written using the quantities $S_l^+(\tau+h_n y)$, $S_l^+(\tau+h_n y)-S_l^+(\tau)$, and $D_l^+(\tau)$ for which we gave asymptotic behaviour in Lemma 5.1.

The following lemma gives the behaviours of A_n, B_n, and C_n defined in (18).

Lemma 5.2. *Suppose that $h_n\to 0$, $nh_n\to\infty$, and $h_n/\beta(n,h_n)\to 0$ when $n\to\infty$, and* H1 *is satisfied; then*

(a)

$$\frac{\sum_{i=1}^{n} w_i^+(\tau)(y_i-m_+(\tau))}{\sum_{i=1}^{n} w_i^+(\tau)}=\mathcal{O}_P\left(h_n^2+\frac{1}{\sqrt{nh_n}}\right).$$

(b)

$$\frac{\sum_{i=1}^{n} w_i^+(\tau+h_n y)-w_i^+(\tau)}{\sum_{i=1}^{n} w_i^+(\tau+h_n y)} = \mathcal{O}_P\left(\frac{1}{\beta(n,h_n)}+\frac{1}{\sqrt{nh_n\beta(n,h_n)}}\right).$$

(c)

$$\alpha(n,h_n)\frac{\sum_{i=1}^{n}(w_i^+(\tau+h_n y)-w_i^+(\tau))(y_i-m_+(\tau))}{\sum_{i=1}^{n} w_i^+(\tau+h_n y)}$$
$$=\frac{1+o_P(1)}{f_X(\tau)}\mathcal{M}_n^+(z)+\mathcal{O}_P\left(\left[\frac{\alpha(n,h_n)}{\beta(n,h_n)}+\frac{\alpha(n,h_n)}{\sqrt{nh_n\beta(n,h_n)}}\right]\left[h_n+\frac{1}{\sqrt{nh_n}}\right]\right),$$

with $\mathcal{M}_n^+(z)=\frac{\alpha(n,h_n)}{nh_n}\sum_{i=1}^{n}(M_+(\frac{\tau+h_n y-X_i}{h_n})-M_+(\frac{\tau-X_i}{h_n}))(Y_i-m_+(\tau))$ *and we use* y *for the rescaled* z *variable; i.e.,* $y=z/\beta(n,h_n)$.

Proof of Lemma 5.2. Assertion (a) follows from standard arguments similar to those needed for the proof of the standard result giving the asymptotic behaviour of the RLL estimate conditional MSE in the smooth case. Part (b) follows from (19) and the second assertion in (b) of Lemma 5.1. The proof of part (c) needs to notice the following equality

$$[D_0^+(\tau+h_n y)-D_0^+(\tau)]\,K_2^+-\frac{1}{h_n}[D_1^+(\tau+h_n y)-D_1^+(\tau)]\,K_1^+=\mathcal{M}_n^+(z).$$

See [18] for more details. ∎

5.2. *Proof of Theorem* 3.1

We first show that $\mathcal{M}_n(z)$ converges weakly to a bilateral compound Poisson process. To do that, we proceed as follows: (1) we prove the 2-dimensional convergence of the process $\mathcal{M}_n(z)$; (2) we show that the sequence of processes $(\mathcal{M}_n(z))$ is tight. To prove the 2-dimensional convergence, according to the Cramer–Wold device, we need to show that for any (a,b) and for each pair $(z_1,z_2)\in[-M,M]\times[-M,M]$, $a\mathcal{M}_n(z_1)+b\mathcal{M}_n(z_2)$ have an asymptotic distribution.

As a preliminary remark, observe that $\mathcal{M}_n(z)$ is the cumulative sum of the terms of the same row of a triangular array. Indeed

$$\mathcal{M}_n(z)=\mathcal{M}_n^+(z)-\mathcal{M}_n^-(z),$$

where

$$\mathcal{M}_n^{\pm}(z)=\frac{\alpha(n,h_n)}{nh_n}\sum_{i=1}^{n}\left(M_{\pm}\left(\frac{\tau+h_n y-X_i}{h_n}\right)-M_{\pm}\left(\frac{\tau-X_i}{h_n}\right)\right)(Y_i-m_{\pm}(\tau))$$
$$=\sum_{i=1}^{n}M_i^{\pm}(z). \tag{21}$$

Consequently

$$\mathscr{M}_n(z) = \sum_{i=1}^{n} (M_i^+(z) - M_i^-(z)) = \sum_{i=1}^{n} M_i(z), \tag{22}$$

$$a\mathscr{M}_n(z_1) + b\mathscr{M}_n(z_2) = \sum_{i=1}^{n} (aM_i(z_1) + bM_i(z_2)), \tag{23}$$

where we have denoted $M_i(z) = M_i^+(z) - M_i^-(z)$. In fact the variables $M_i(z)$, $M_i^+(z)$, and $M_i^-(z)$ depend on n, but to simplify notation we omit the index n. Let

$$T_i^n = aM_i(z_1) + bM_i(z_2).$$

The proof of the convergence of $\sum_{i=1}^{n} T_i^n$ relies upon the result of convergence of Gnedenko [17] to infinitely divisible distributions which we recall below:

THEOREM 5.2 (Gnedenko [17]). *Let $\Gamma_n(u) = \sum_{i=1}^{n} \int_{-\infty}^{u} x^2 \, d\tilde{F}_{ni}(x)$ where $\tilde{F}_{ni}(\cdot)$ is the cdf of $\tilde{T}_i^n = T_i^n - E(T_i^n)$. Assume that the array (T_i^n) satisfies the following condition:*

$$\forall \epsilon > 0, \qquad \sup_{1 \leqslant i \leqslant n} \mathbb{P}(|\tilde{T}_i^n| > \epsilon) \to 0 \qquad (n \to \infty).$$

In order that the distribution laws of the sums $\sum_{i=1}^{n} \tilde{T}_i^n$ converge to a limit, and the variance of these sums converges to the variance of the limit law, it is necessary and sufficient that there exists a nondecreasing function $\Gamma(\cdot)$ such that $\Gamma_n(u)$ converges weakly to $\Gamma(u)$. The logarithm of the characteristic function of the limit law, $\phi(\cdot)$, is given by Kolmogorov's formula

$$\log \phi(t) = it\gamma + \int (\exp(itu) - 1 - itu) \frac{1}{u^2} \Gamma(du),$$

where γ is chosen according to the formula

$$\gamma - E\left(\sum_{i=1}^{n} T_i^n\right) = o(1).$$

Let us introduce the following notation:

(1) $g_{\pm}(x, z) = M_{\pm}(x + y) - M_{\pm}(x) = M_{\pm}\left(x + \frac{z}{\beta(n, h_n)}\right) - M_{\pm}(x)$,

(2) $g_1^{\pm}(x, z) = g_{\pm}(x, z)(m(\tau - xh_n) - m_{\pm}(\tau))$, $g_1(x, z) = g_1^+(x, z) - g_1^-(x, z)$,

(3) $g_2^{\pm}(x, z) = g_{\pm}(x, z)(m(\tau - xh_n) - m_{\pm}(\tau) + \sigma(\tau - xh_n)\,\varepsilon)$, $g_2(x, z) = g_2^{+}(x, z) - g_2^{-}(x, z)$,

(4) $A_n = \sum_{i=1}^{n} E[T_i^n]$,

(5) $\Gamma_n(u) = \sum_{i=1}^{n} E\,[(T_i^n - E(T_i^n))^2\, \mathbb{I}_{\{T_i^n - E(T_i^n) \leqslant u\}}]$,

where $\mathbb{I}_A(\cdot)$ is the indicator function of the set A.

Elementary calculations provide

$$A_n = \alpha(n, h_n) \int (ag_1(x, z_1) + bg_1(x, z_2))\, f_X(\tau - xh_n)\, dx,$$

$$\Gamma_n(u) = \frac{\alpha(n, h_n)}{nh_n} \int E\left[\left(ag_2(x, z_1) + bg_2(x, z_2) - \frac{h_n}{\alpha(n, h_n)} A_n\right)^2 \times \mathbb{I}_{\{(\alpha(n, h_n))/(nh_n)(ag_2(x, z_1) + bg_2(x, z_2)) - 1/n\, A_n \leqslant u\}}\right] f_X(\tau - xh_n)\, dx.$$

Under the conditions H1–H3 and $h_n \to 0$, we can show:

(a) If $x < -y < 0$ or $x < 0 < -y$ or $-y < 0 < x$ or $0 < -y < x$,

(1) $g_1(x, z)\, f_X(\tau - xh_n) = \mathcal{O}\left(\frac{h_n}{\beta(n, h_n)}\right)$,

(2) $g_2(x, z) - \frac{h_n}{\alpha(n, h_n)} A_n = \mathcal{O}_P\left(\frac{1}{\beta(n, h_n)}\right)$.

(b) If $-y < x < 0$ or $0 < x < -y$,

(1) $g_1(x, z)\, f_X(\tau - xh_n) = -\gamma(\tau)\, f_X(\tau)\, M_+(0) + o(1)$,

(2) $g_2(x, z) - \frac{h_n}{\alpha(n, h_n)} A_n = (-\gamma(\tau) - 2\,\mathrm{sign}(z)\, \sigma(\tau)\, \varepsilon)\, M_+(0) + o_P(1)$.

Therefore

$$A_n = -\frac{\alpha(n, h_n)}{\beta(n, h_n)}\, \gamma(\tau)\, f_X(\tau)\, M_+(0)(a\,|z_1| + b\,|z_2|) + o\left(\frac{\alpha(n, h_n)}{\beta(n, h_n)}\right).$$

We deduce

LEMMA 5.3. *Under the conditions* H1–H3, $h_n \to 0$ *and* $nh_n \to \infty$, *we have*

(a) $A_n \to A = -L_4 \gamma(\tau)\, f_X(\tau)\, M_+(0)(a\,|z_1| + b\,|z_2|)$,

(b) $\Gamma_n(u) \to \Gamma(u)$,

where

(1) *If* $0 \leqslant z_1 \leqslant z_2$ *then* $\Gamma(u) = 4L_4 L_5 M_+^2(0)\, f_X(\tau)[z_1(a+b)^2\, v_{a+b}^{+}(u) + (z_2 - z_1)\, b^2 v_b^{+}(u)]$,

(2) *if* $z_2 \leqslant z_1 \leqslant 0$ *then* $\Gamma(u) = 4L_4 L_5 M_+^2(0)\, f_X(\tau)[-z_1(a+b)^2\, v_{a+b}^{-}(u) + (z_1 - z_2)\, b^2 v_b^{-}(u)]$,

(3) *if* $z_1 \leqslant 0 \leqslant z_2$ *then* $\Gamma(u) = 4L_4L_5M_+^2(0)\, f_X(\tau)[-z_1 a^2 v_a^-(u) + z_2 b^2 v_b^+(u)]$,

with

$$v_a^{\pm}(u) = E[(-\gamma(\tau) \mp 2\sigma(\tau)\,\varepsilon)^2\, \mathbb{I}_{\{2L_5 a M_+(0)(-\gamma(\tau)+\sigma(\tau)\,\varepsilon) \leqslant u\}}].$$

We also have

$$\begin{aligned}
&\sup_{1\leqslant i\leqslant n} \mathbb{P}[|aM_i(z_1)+bM_i(z_2)-E[aM_i(z_1)+bM_i(z_2)]| \geqslant \varepsilon] \\
&\quad = \mathbb{P}[|aM_1(z_1)+bM_1(z_2)-E[aM_1(z_1)+bM_1(z_2)]| \geqslant \varepsilon], \\
&\quad \leqslant \frac{\mathbb{E}[aM_1(z_1)+bM_1(z_2)-E[aM_1(z_1)+bM_1(z_2)])^2]}{\epsilon^2}, \\
&\quad \leqslant \frac{\gamma_n(\infty)}{n\epsilon^2} \to 0 \qquad (n\to\infty).
\end{aligned}$$

Hence this yields:

Lemma 5.4. *Under the conditions* H1, H2, *and* H3 *and for any* $\epsilon > 0$, *we have*

$$\sup_{1\leqslant i\leqslant n} \mathbb{P}[|aM_i(z_1)+bM_i(z_2)-E[aM_i(z_1)+bM_i(z_2)]| \geqslant \epsilon] \to 0,$$

as $n \to \infty$.

Denote $\xi^{\pm} = -\gamma(\tau) \mp 2\sigma(\tau)\,\varepsilon$, and recall that $\lambda_1 = 2L_5M_+(0)$, $\lambda_2 = (L_4/L_5)\, f_X(\tau)$, and $\lambda_3 = L_4M_+(0)\, f_X(\tau) = \lambda_1\lambda_2/2$.

We have

(a) If $0 \leqslant z_1 \leqslant z_2$, then $d\Gamma(u)/u^2 = \lambda_2[z_1 dF_{\lambda_1(a+b)\,\xi^+}(u) + (z_2 - z_1)\, dF_{\lambda_1 b\xi^+}(u)]$.

(b) If $z_2 \leqslant z_1 \leqslant 0$, then $d\Gamma(u)/u^2 = \lambda_2[z_1 dF_{\lambda_1(a+b)\,\xi^-}(u) + (z_1 - z_2)\, dF_{\lambda_1 b\xi^-}(u)]$.

(c) If $z_1 \leqslant 0 \leqslant z_2$, then $d\Gamma(u)/u^2 = \lambda_2[z_1 dF_{\lambda_1 a\xi^-}(u) + z_2 dF_{\lambda_1 b\xi^+}(u)]$, where $F_X(\cdot)$ is the *cdf* of X.

According to Gnedenko [17], Lemmas 5.3 and 5.4 prove that $a\mathcal{M}_n(z_1) + b\mathcal{M}_n(z_2)$ has an asymptotic distribution. The logarithm of the characteristic function of the limit law is given by Kolmogorov's formula of (see Theorem 5.2) with constant equals to A and the function $\Gamma(u)$ is defined as in Lemma 5.3.

Note $\phi(\cdot)$ the characteristic function of the limit law and $\phi_X(\cdot)$ X's one. From Lemma 5.3, it follows:

(a) If $0 \leqslant z_1 \leqslant z_2$, then $\phi(t) = \exp[-i\lambda_3 \gamma(\tau)(a\,|z_1| + b\,|z_2|)\,t] \times \exp[\lambda_2 z_1 (\phi_{\lambda_1(a+b)\,\xi^+}(t) - 1)] \times \exp[\lambda_2(z_2 - z_1)(\phi_{\lambda_1 b\xi^+}(t) - 1)]$.

(b) If $z_2 \leqslant z_1 \leqslant 0$, then $\phi(t) = \exp[-i\lambda_3 \gamma(\tau)(a\,|z_1| + b\,|z_2|)\,t] \times \exp[-\lambda_2 z_1 (\phi_{\lambda_1(a+b)\,\xi^-}(t) - 1)] \times \exp[\lambda_2(z_1 - z_2)(\phi_{\lambda_1 b\xi^-}(t) - 1)]$.

(c) If $z_1 \leqslant 0 \leqslant z_2$, then $\phi(t) = \exp[-i\lambda_3 \gamma(\tau)(a\,|z_1| + b\,|z_2|)\,t] \times \exp[-\lambda_2 z_1 (\phi_{\lambda_1 a\xi^-}(t) - 1)] \times \exp[\lambda_2 z_2 (\phi_{\lambda_1 b\xi^+}(t) - 1)]$.

Hence we have proved the finite-dimensional convergence, under the conditions H1, H2, and H3, of the process $\mathscr{M}_n$ to the bilateral compound Poisson process $\mathscr{M}$ defined as

$$\mathscr{M}(z) = -\lambda_3 \gamma(\tau)\,|z| + \lambda_1 \mathscr{N}(z),$$

where

$$\mathscr{N}(z) = \begin{cases} 0 & \text{if} \quad z = 0, \\ \sum_{i=1}^{N_z^+} -\gamma(\tau) - 2\sigma(\tau)\,\varepsilon_i^+ & \text{if} \quad z > 0, \\ \sum_{i=1}^{N_z^-} -\gamma(\tau) + 2\sigma(\tau)\,\varepsilon_i^- & \text{if} \quad z < 0. \end{cases} \tag{24}$$

The sequences (ε_i^+) and (ε_i^-) are independent and built with i.i.d. variables distributed as the model error variable ε. N_z^+ and N_z^- are independent homogeneous Poisson processes with λ_2 as parameter and are independent of the sequences (ε_i^+) and (ε_i^-).

We verify easily

(i) $E[\mathscr{M}(z)] = -3\lambda_3 \gamma(\tau)\,|z|$,

(ii) $Cov(\mathscr{M}(z_1), \mathscr{M}(z_2)) = \lambda_1^2 \lambda_2 (\gamma^2(\tau) + \sigma^2(\tau))(\min(|z_1|, |z_2|)\,\mathbb{I}_{\{z_1 z_2 > 0\}})$.

Since

$$\mathscr{Z}_n(z) = \frac{1 + o_P(1)}{f_X(\tau)}\,\mathscr{M}_n(z),$$

then

$$\mathscr{Z}_n(z) \Rightarrow \mathscr{Z}(z) = -\frac{\lambda_3}{f_X(\tau)}\,\gamma(\tau)\,|z| + \frac{\lambda_1}{f_X(\tau)}\,\mathscr{N}(z).$$

LEMMA 5.5. *Under the conditions* H1, H2, *and* H3, *the sequence of the process* $(\mathscr{M}_n)$ *is tight.*

Proof of Lemma 5.5. We only sketch the proof. The reader is refered to [18] or [24] for more precise arguments. Under assumptions H1, H2, and H3, the process $\mathscr{Z}_n(z)$ is an element of $\mathscr{D}[-\mathscr{M}, \mathscr{M}]$. To show that sequence of the process $(\mathscr{M}_n(z))$ is tight, it suffices (see Billingsley [3]) to check that for any $z_1 \leqslant z \leqslant z_2$:

$$E[|\mathscr{M}_n(z_2)-\mathscr{M}_n(z)|\,|\mathscr{M}_n(z_1)-\mathscr{M}_n(z)|] \leqslant \lambda(z_2-z_1)^2.$$

Basically, the result will follow from the three following inequalities.

$$\begin{aligned} &E[|\mathscr{M}_n^{\pm}(z_2)-\mathscr{M}_n^{\pm}(z)|\,|\mathscr{M}_n^{\pm}(z_1)-\mathscr{M}_n^{\pm}(z)|] \\ &\quad\leqslant nE[|M_1^{\pm}(z_2)-M_1^{\pm}(z)|\,|M_1^{\pm}(z_1)-M_1^{\pm}(z)|] \\ &\qquad+n(n-1)\,E[|M_1^{\pm}(z_2)-M_1^{\pm}(z)|]\times E[|M_1^{\pm}(z_1)-M_1^{\pm}(z)|]. \end{aligned}$$

$$E[|M_1^{\pm}(z_2)-M_1^{\pm}(z)|\,|M_1^{\pm}(z_1)-M_1^{\pm}(z)|] \leqslant c_2\,\frac{\alpha_2(n, h_n)}{n^2 h_n \beta^2(n, h_n)}\,|z_2-z|\,|z_1-z|,$$

$$E[|M_1^{\pm}(z_1)-M_1^{\pm}(z)|] \leqslant c_3\,\frac{\alpha(n, h_n)}{n\beta(n, h_n)}\,|z_1-z|. \quad \blacksquare$$

Finally, by using the finite-dimensional convergence and the tightness of the process $\mathscr{M}_n$, we establish the weak convergence of the process $\mathscr{M}_n$ to a limit process $\mathscr{M}$.

ACKNOWLEDGMENTS

We are grateful to the editor and a reviewer for valuable remarks and comments.

REFERENCES

1. A. Antoniadis and G. Grégoire, Nonparametric estimation in change-point hazard rate models for censored data: a counting process approach, *Nonparametric Statist.* **3** (1993), 135–154.
2. A. Antoniadis and I. Gijbels, "Inference in Change-Point Problems via Wavelet Decompositions," Technical report, Institut de Statistique, Université Catholique de Louvain, 1996.
3. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.
4. M. Basseville and I. V. Nikiforov, "Detection of Abrupt Changes: Theory and Application," Prentice–Hall Information and System Sciences Series, Prentice–Hall, Englewood Cliffs, NJ, 1993.
5. W. Csörgö and L. Horváth, Nonparametric methods for changepoint problems, *Handbook Statist.* **7** (1998), 403–425.

6. W. S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Amer. Statist. Assoc.* **74** (368) (1979), 829–836.
7. G. W. Cobb, The problem of the Nile: Conditional solution to a changepoint problem, *Biometrika* **65** (1978), 243–251.
8. J. Durbin, R. L. Brown, and J. M. Evans, Techniques for testing the constancy of regression relationships over time, *J. Roy. Statist. Soc. B* **37** (1975), 149–163.
9. W. F. Eddy, The asymptotic distributions of the mode, *Ann. Statist.* **59** (1982), 279–290.
10. L. Eubank and P. L. Speckman, Nonparametric estimation of functions with jump discontinuities, "Change-Point Problems" (H.-G. Müller, E. Carlstein, and D. Siegmund, Eds.), Institute of Mathematical Statistics, Hayward, California, 1994.
11. J. Fan, Local linear regression smoothers and their minimax efficiencies, *Ann. Statist.* **21** (1993), 196–216.
12. J. Fan and I. Gijbels, Variable bandwidth and local linear regression smoothers, *Ann. Statist.* **20** (1992), 2008–2036.
13. J. Fan and L. Gijbels, "Local Polynomial Modelling and Its Applications," Chapman and Hall, London, 1996.
14. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol. II, Wiley, New York, 1996.
15. I. Gijbels and A. C. Goderniaux, "Bandwidth Selection for Change Point Estimation in Nonparametric Regression," discussion paper, Institut de Statistique, Université Catholique de Louvain, available at http://www.stat.ucl.ac.be, 2000.
16. D. Girard, From template matching to optimal approximation by piecewise smooth curves, "SPIE Proceedings on Curves and Surfaces in Computer Vision and Graphics, Santa Clara, CA, Feb. 11–16, 1990."
17. B. V. Gnedenko and A. N. Kolmogorov, "Limit Distributions for Sum of Independent Random Variables," Addison–Wesley Series in Statistics, Addison–Wesley, Reading, MA, 1968.
18. G. Grégoire and Z. Hamrouni, "Local Linear Smoothing for Change Point Estimation," Technical report, Laboratoire LMC/IMAG, Université Joseph Fourier, 1999.
19. G. Grégoire and Z. Hamrouni, Two non-parametric tests for change-point problems, *Nonparametric Statist.*, to appear.
20. G. Grégoire and Z. Hamrouni, Exponential rate for estimating the number of discontinuities, in preparation.
21. P. Hall, I. Gijbels, and A. Kneip, "Interval and Band Estimation for Curves with Jumps," Discussion paper 9609, Institut de Statistique, Université Catholique de Louvain, available at http://www.stat.ucl.ac.be, 1996.
22. P. Hall, I. Gijbels, and A. Kneip, On the estimation of jump points in smooth curves, *Ann. Inst. Statist. Math.* **51** (1999), 231–251.
23. P. Hall and T. Wehrly, A geometrical method for removing edge effects from kernel-type nonparametric regression estimator, *J. Amer. Statist. Assoc.* **86** (415) (1991), 665–672.
24. Z. Hamrouni, "Inférence statistique par lissage linéaire local pour une fonction de régression présentant des discontinuité," Unpublished thesis, Laboratoire LMC/IMAG, Université Joseph Fourier, 1999.
25. M. Lejeune, Estimation non-paramétrique par noyaux: Régression polynomiale mobile, *Rev. Statist. Appl.* **33** (1985), 43–64.
26. C. R. Loader, Change point estimation using nonparametric regression, *Ann. Statist.* **24** (1996), 1667–1678.
27. P. J. Laurent and F. Utreras, Optimal smoothing of noisy broken data with spline functions. *J. Approx. Theory Appl.* **2** (1986), 71–94.
28. H. G. Müller, E. Carlstein, and D. Siegmund, "Change-Point Problems," Institute of Mathematical Statistics Lecture Notes Series, Vol. 23, Institute of Mathematical Studies, 1994.

29. H. G. Müller and U. Stadtmüller, Discontinuous versus smooth regression, *Ann. Statist.* **274** (1999), 299–337.
30. H. G. Müller and U. Stadtmüller, Estimation of heteroscedasticity in regression analysis, *Ann. Statist.* **15** (1987), 610–625.
31. H. G. Müller and K. S. Song, Two-stage change-point estimators in smooth regression models, *Statist. Probab. Lett.* **34** (1997), 323–335.
32. H. G. Müller, Change-points in nonparametric regression analysis, *Ann. Statist.* **20** (1992), 737–761.
33. D. Ruppert and M. P. Wand, Multivariate locally weighted least squares regression, *Ann. Statist.* **22** (1994), 1346–1370.
34. P. Qiu, Estimation of the number of jumps of the jump regression functions, *Statist. Theory Method* **23** (1994), 2141–2155.
35. J. H. Shiau, "Smoothing Spline Estimation of Functions with Discontinuities," Ph.D. thesis, University of Wisconsin, Madison, Wisconsin, 1985.
36. C. J. Stone, Optimal global rates of convergence for nonparametric regression, *Ann. Statist.* **10** (1982), 1040–1053.
37. Y. Wang, Jump and sharp cusp detection by wavelets, *Biometrika* **82** (1995), 385–397.
38. W. Whitt, Some useful functions for functional limit theorems, *Math. Oper. Res.* **5** (1980), 67–84.
39. J. S. Wu and C. K. Chu, Kernel-type estimators of jump points and values of a regression function, *Ann. Statist.* **21** (3) (1993), 1545–1566.
40. J. S. Wu and C. K. Chu, Nonparametric function estimation and bandwidth selection for discontinuous regression functions, *Statist. Sinica* **3** (1993), 557–576.
41. J. S. Wu and C. K. Chu, Modification for boundary effects and jump points in nonparametric regression, *Nonparametric Statist.* **2** (1993), 341–354.
42. Y. Q. Yin, Detection of the number, locations and magnitudes of jumps, *Commun. Statist. Stochastic Models* **4** (1988), 445–455.