Nonparametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring.

Anestis Antoniadis,
IMAG - LMC, University Joseph Fourier, Grenoble,
Irène Gijbels,
Institute of Statistics, U. C. L., Louvain-La-Neuve
and
Brenda MacGibbon,
University of Quebec at Montreal.

September 16, 1998

Abstract

A nonparametric wavelet based estimator is proposed for the location of a change-point in an otherwise smooth hazard function under noninformative random right censoring. The proposed estimator is based on wavelet coefficients differences via an appropriate parametrization of the time-frequency plane. The study of the estimator is facilitated by the strong representation theorem for the Kaplan-Meier estimator established by Lo and Singh (1986). The performance of the estimator is checked via simulations and two real examples conclude the paper.

1 Introduction

An important practical issue of many research papers in survival analysis is the study of the risk patterns of disease or other events in time. Although most display results with plots of the survival function (Kaplan and Meier (1958)) or the cumulative hazard function (Nelson (1972)), a careful study of the hazard function itself would be beneficial in many cases as argued by Efron (1980). The hazard function of time-related events is often time-varying in a structured fashion, which can sometimes be decomposed into additive phases. Each phase is in general shaped by a different generic hazard function and the transition or jump from one phase to the other (the change-point) is in general difficult to locate. One way to conduct inference in such hazard rate change-point models is to use nonparametric estimation procedures whose nature allows great flexibility in the possible form of the generic functions modeling each phase.

Müller and Wang (1990) estimate the point of the most rapid change of a hazard function via the location of an extremum of a nonparametric estimator of the derivative of the hazard function. Müller and Wang (1994) considered the estimation of the location of a discontinuity in a hazard function that is supposed to be left and right continuous at an unknown jump point. Their analysis is based on differences of two one-sided kernel type estimators of the hazard function and their change-point estimator is shown to be weakly consistent with a rate of convergence that is faster than $n^{-1/2}$ where n denotes the sample size. The procedure they propose is adopted from the one Müller (1992) has developed for "standard" nonparametric regression. An enhancement of their procedure for the regression setting has been recently proposed by Loader (1996) and could possibly be adapted to the case of hazard function estimation.

In the regression setting, the theory of wavelets has been recently used to detect jumps or cusps in an otherwise smooth regression function. Using the fact that the oscillating property as well as the multiresolution structure allows a precise characterization of local regularity of a signal in terms of its wavelet coefficients, several authors (Mallat and Hwang (1992), Wang (1995), Raimondo (1996), Antoniadis and

Gijbels (1997)) have proposed jump-point detection procedures to estimate singularities in signals observed with noise. In all the regression setting papers mentioned above, the models have been formulated by means of additive i.i.d. Gaussian noise. The estimation of a smooth hazard function by wavelet methods has been addressed by Antoniadis, Grégoire and McKeague (1994), and more recently by Antoniadis, Grégoire and Nason (1999), but to our knowledge there are no papers (yet) dealing with wavelet detection of change-points in hazard functions with censored observations, which is the primary goal of this paper.

For the sake of clarity, we consider here the simplest case of a single jump or a sharp cusp in an otherwise smooth hazard function which presents some regularity before and after the location of the change-point. It is, however, clear that our method can be easily extended to the case where a jump or a sharp cusp occurs in higher derivatives of the hazard function.

The organization of the paper is as follows. In Section 2 we introduce the notation and the random censoring setup, and present an i.i.d strong representation of the Nelson estimator of the cumulative hazard function which will be needed later for the study of the wavelet estimator proposed. Section 3 recalls some basic facts from Raimondo (1996) on the behavior of wavelet coefficients differences for functions presenting local singularities. These results have been adapted to the case of a function that presents an abrupt discontinuity or a sharp cusp. The main results are discussed in Section 4. The final section is devoted to the presentation of some simulations and to the analysis of some real examples. Proofs are collected in the appendix.

2 Some preliminaries on hazard rates and change-points

Let $T_1, \ldots, T_n$ be i.i.d. lifetimes with cumulative distribution function (c.d.f) F, and let $C_1, \ldots, C_n$ be i.i.d. censoring times with c.d.f G. Assume also that the T_i's and the C_i's are independent. The observed data is a realization of $\{(X_i, \delta_i), i = 1, \ldots, n\}$ with

$$X_i = \min(T_i, C_i), \qquad \delta_i = \mathbf{I}_{\{X_i = T_i\}} = \mathbf{I}_{\{T_i \leq C_i\}},$$

where $\mathbf{I}_A$ denotes the indicator function of a set A. We will also assume that F is absolutely continuous and we will denote by f its density with respect to Lebesgue measure. Denote by Λ the cumulative hazard function defined by

$$\Lambda(x) = -\log(1 - F(x)),$$

and let λ be the associated hazard rate function defined by

$$\lambda(x) = \frac{d}{dx}(\Lambda(x)) = -\frac{f(x)}{\bar{F}(x)}, \quad \text{for } x \text{ such that } F(x) < 1,$$

where $\bar{F}(x) = 1 - F(x)$.

A change-point in a hazard function refers here to a localized jump or a sharp cusp in an otherwise smooth hazard function. The change-point model that we are going to consider may be defined as follows.

Let K, L, M, s, α, $0 \leq \alpha < s \leq 1$ be positive constants and denote by I an interval of the real line $\mathbb{R}$. The class of real functions presenting a single jump or a sharp cusp in the interior of I, say $\overset{\circ}{I}$, will be denoted by $\mathcal{F}_{s,\alpha}(K, L, M, I)$. More precisely, a function f belongs to the class $\mathcal{F}_{s,\alpha}(K, L, M, I)$ if and only if, there exists a unique $\theta_f \in \overset{\circ}{I}$ such that:

(1) there exists $h_0 > 0$ such that for all h with $|h| \leq h_0$ we have
$|f(\theta_f - h) - f(\theta_f + h)| \geq K|2h|^\alpha$;

(2) for all $(x, y) \in I \times I$ such that $\theta_f \notin [x, y]$ we have $|f(x) - f(y)| \leq L|x - y|^s$;

(3) $\|f\|_\infty = \sup_{x \in I} |f(x)| \leq M$.

Remark 2.1 Note that we assume that the regularity condition of the function with a jump or a cusp is the same before and after the location of the singularity. This assumption can be easily relaxed: if the Hölder exponents of a function before and after the change-point are s_1 and s_2 respectively everything remains valid by taking $s = \min(s_1, s_2)$ in the above definition of the class $\mathcal{F}_{s,\alpha}(K, L, M, I)$.

The constants K, L and M will be assumed to be fixed and, without loss of generality the interval I will be assumed equal to $[0, 1]$. For ease of notation the class $\mathcal{F}_{s,\alpha}(K, L, M, I)$ will be denoted $\mathcal{F}_{s,\alpha}$ hereafter.

We will assume henceforth that $\lambda \in \mathcal{F}_{s,\alpha}(K, L, M, I)$ where $I = [0, T]$ with $T < \inf\{x \geq 0; L(x) = 1\}$ equal to the right end of the support of the distribution function L of the X_i's. To simplify the notation we will assume without any loss of generality that $T = 1$. We will also denote by

$$L_1(x) = \mathrm{P}(X_i \leq x, \delta_i = 1)$$

the subdistribution function of the uncensored observations.

The assumption that λ belongs to the class $\mathcal{F}_{s,\alpha}$ leads naturally to a wavelet detection of jumps based on the results of the next section. The basic idea for estimating the location of the jump or the cusp consists in looking at the wavelet coefficient differences for an appropriate estimator of λ. Since these differences are of a smaller order of amplitude, except at the change-point, it can be expected that the maximum of these differences occurs near the location of the change-point. This reasoning will lead to the change-point estimator $\hat{\theta}_\lambda$ to be defined and studied in Section 4.

A basic tool for studying the behavior of the estimates of the wavelet coefficient differences for λ will be an asymptotic representation of the product-limit (Nelson) estimator of the cumulative hazard rate function established by Lo and Singh (1986) (see also the paper by Lo, Mack and Wang (1989)). Using the same notation as in these papers, let

$$g(x) = \int_0^x \frac{dL_1(s)}{\bar{L}(s)^2}, \tag{2.1}$$

where $\bar{L}(x) = \mathrm{P}(X_i > x) = \bar{F}(x)\bar{G}(x) = \mathrm{P}(T_i > x, C_i > x)$.

For positive real z and x, and δ taking values 0 or 1, let

$$\xi(z, \delta, x) = g(\min(z, x)) + [\bar{L}(z)]^{-1}\mathbf{I}_{\{z \leq x, \delta = 1\}}, \tag{2.2}$$

and set $\xi_i(x) = \xi(X_i, \delta_i, x)$. Note that the random variables ξ_i are i.i.d and uniformly bounded on $[0, 1]$. Moreover, it is easy to see that $\mathbb{E}(\xi_i(x)) = 0$ and $\mathrm{cov}(\xi_i(x), \xi_i(y)) = g(\min(x, y))$. In the sequel we will use the improved version of Theorem 1 of Lo and Singh (1986), as it is proved in Lemma 2.1 of Lo, Mack and Wang (1989), which we state here for the sake of completeness.

Let $\hat{\Lambda}_n$ denote the Nelson (1972) estimator of the cumulative hazard function Λ, given by

$$\hat{\Lambda}_n(x) = \sum_{X_{(i)} \leq x} \frac{\delta_{(i)}}{n-i+1},$$

where $X_{(i)}$ denotes the i-th order statistic of $X_1, \ldots, X_n$ and $\delta_{(i)}$ the corresponding censoring indicator variable. We then have, for all $x \in [0,1]$:

$$\hat{\Lambda}_n(x) = \Lambda(x) + \frac{1}{n}\sum_{i=1}^{n} \xi_i(x) + r_n(x), \tag{2.3}$$

where

$$\sup_{x \in [0,1]} |r_n(x)| = \mathcal{O}\left(\frac{\log n}{n}\right), \quad a.s. \tag{2.4}$$

and for any $\beta \geq 1$,

$$\sup_{x \in [0,1]} \mathbb{E}(|r_n(x)|^\beta) = \mathcal{O}\left(\left[\frac{\log n}{n}\right]^\beta\right).$$

To end this preliminary results section we state the integration by parts lemma of Földes, Rëjto and Winter (1981) which by now is a standard device in the hazard function estimation literature:
If $\mathcal{G}$ and $\mathcal{H}$ are nondecreasing mappings from $\mathbb{R}$ into $\mathbb{R}$ and if $-\infty < c < d < \infty$, then

$$\int_{[c,d]} \mathcal{G}(x^+) d\mathcal{H}(x) + \int_{[c,d]} \mathcal{H}(x^-) d\mathcal{G}(x) = \mathcal{G}(d^+)\mathcal{H}(d^+) - \mathcal{G}(d^-)\mathcal{H}(d^-), \tag{2.5}$$

where $\mathcal{G}(x^+)$ (resp. $\mathcal{G}(x^-)$) denotes the right- (resp. left-) hand limit of $\mathcal{G}$ at the point x.

3 Differences in wavelet coefficients

We first recall some results on wavelet coefficients differences characterizing jumps or sharp cusps of a function in $\mathcal{F}_{s,\alpha}$. In this study of hazard rate change-points, we will follow closely the presentation of Raimondo (1996), adapting it to our case. Note that only the oscillating property and the multiresolution structure of the compactly supported spline scaling function are used; the orthogonality of the associated wavelet basis is not used at all.

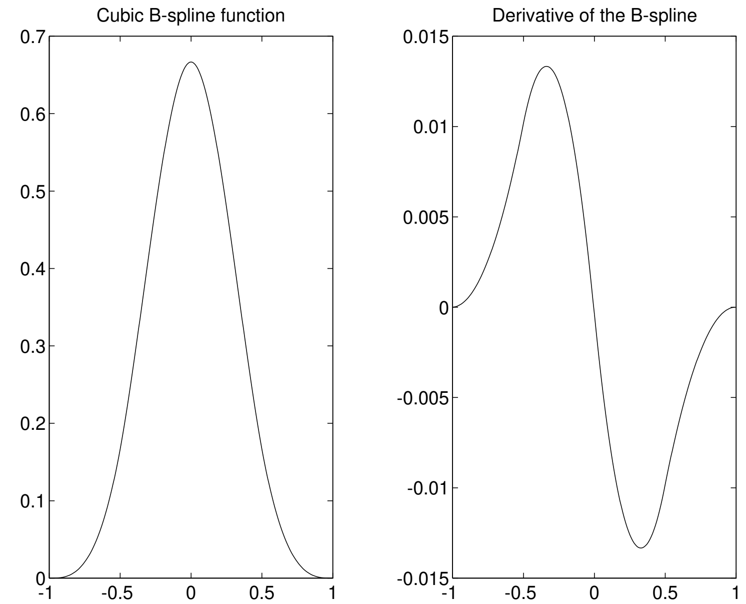

Figure 1. The scaling function ϕ and the associated wavelet ψ.

We consider a smooth enough compactly supported scaling function whose first order derivative generates an admissible zero-moment wavelet. To be precise, the scaling function ϕ used to derive our results will be a cubic B-spline (De Boor (1978)) whose support is the interval $[-1, 1]$ and whose integral is equal to 1. Mallat and Hwang (1992) have also adopted such a cubic B-spline for extrema detection. The first order derivative $\psi = d\phi/dx$ of such a function generates a compactly supported admissible wavelet. Figure 1 displays the scaling function ϕ and the associated wavelet ψ. Since a cubic B-spline is positive, we have $\int \phi(x)dx = \int |\phi(x)|dx = 1$. Moreover, the function ϕ is smooth in the sense that it admits derivatives of any order.

By translating and changing its scale we generate from ϕ a doubly-indexed family of scaling functions and corresponding wavelets as follows:

$$\begin{aligned} \phi_{j,k}(x) &= 2^{j/2}\phi(2^j x - k), \qquad j \geq 0,\ k \in \mathbb{Z}, \\ \psi_{j,k}(x) &= 2^{j/2}\psi(2^j x - k), \qquad j \geq 0,\ k \in \mathbb{Z}, \end{aligned}$$

Note that since ϕ and ψ are supported on $[-1, 1]$, we have

$$\text{Support}\,(\phi_{j,k}) = \text{Support}\,(\psi_{j,k}) = \left[\frac{-1+k}{2^j}, \frac{1+k}{2^j}\right].$$

The scaling and wavelet coefficients of a square-integrable function f are respectively defined by:

$$s_{j,k} = \int f(x)\phi_{j,k}(x)\,dx \qquad \text{and} \qquad d_{j,k} = \int f(x)\psi_{j,k}(x)\,dx.$$

To characterize functions of $\mathcal{F}_{s,\alpha}$, defined in Section 2, via their scaling or wavelet coefficients, we will use the parametrization of time-scale space presented in Raimondo (1996). For $k \in \mathbb{Z}$, let $\tau(k) = 2k+1$ and consider the following coefficient differences:

$$\begin{aligned} \Delta_{j,k}(f) &= s_{t,\tau(k)} - s_{t,\tau(k+2)} \\ &= \int f(x)\phi_{j,\tau(k)}(x)dx - \int f(x)\phi_{j,\tau(k+2)}(x)dx = \int f(x)\Phi_{j,k}(x)dx, \end{aligned} \quad (3.1)$$

where $\Phi_{j,k} = \phi_{j,\tau(k)} - \phi_{j,\tau(k+2)}$. Note that

$$\text{Support}\,(\Phi_{j,k}) = \left[\frac{2k}{2^j}, \frac{2k+2}{2^j}\right] \cup \left[\frac{2k+4}{2^j}, \frac{2k+6}{2^j}\right].$$

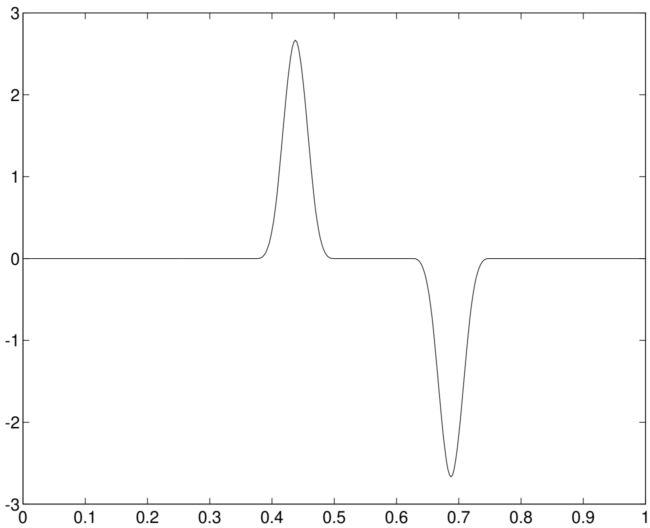

Figure 2. The wavelet difference function Φ used in detecting a jump or a cusp.

Figure 2 displays a typical graph of the function $\Phi_{j,k}$. It is the gap between the intervals that allows the characterization of sharp cusps via the wavelet coefficient differences $\Delta_{j,k}(f)$ defined in (3.1). Indeed, for any positive resolution j and for any $f \in \mathcal{F}_{s,\alpha}$ let us denote by k_1 the (unique) integer such that $\theta_f \in \left[\frac{\tau(k_1)+1}{2^j}, \frac{\tau(k_1)+3}{2^j}\right]$. Then there exist two positive constants C_1 and C_2 such that:

(i) For all $j > 0$ and $0 \leq k \leq \frac{2^j - 6}{2}$ we have:

$$\max_{|k-k_1|\geq 2} |\Delta_{j,k}(f)| \leq C_1 2^{-j(s+1/2)}.$$

(ii) If j is large enough and $0 \leq k \leq \frac{2^j - 6}{2}$

$$\max_{|k-k_1|<2} |\Delta_{j,k}(f)| \geq C_2 2^{-j(\alpha+1/2)}.$$

Moreover the constants C_1 and C_2 do not depend on j or k_1. Note that $0 \leq k \leq \frac{2^j-6}{2}$ ensures that the support of $\Phi_{j,k}$ overlaps with the interval $[0, 1]$. To see why the above assertions hold, see the brief sketch of the proof in the appendix.

Remark 3.1 All the above concerns a jump or a sharp cusp in the function f itself. It is not difficult to see that the above characterization may be extended to a more general class of functions $\mathcal{B}_{s,\alpha}$ with $1 \leq [\alpha] \leq \alpha < s \leq [\alpha] + 1$, where $[\alpha]$ denotes the integer part of α, with

$$f \in \mathcal{B}_{s,\alpha} \Leftrightarrow f^{([\alpha])} \in \mathcal{F}_{s-[\alpha],\alpha-[\alpha]}.$$

In such a case one seeks for a jump or a sharp cusp in the $[\alpha]$th derivative of f. In order to extend the previous results to this apparently more general case using Mallat and Hwang's (1992) local regularity characterization for $f^{([\alpha])}$ we would use instead of the B-spline ϕ, a wavelet ϕ whose support is $[-[\alpha], [\alpha]]$, with $[\alpha] - 1$ vanishing moments and one should define $\tau(k) = [\alpha] + 2k[\alpha]$. The results are then exactly analogous to those of this section, but in order to avoid heavy notation we will not pursue this issue further here.

4 Estimating the change-point

Our aim is to first define an empirical version $\hat{\Delta}_{j,k}(\lambda)$ of the wavelet transform differences $\Delta_{j,k}(\lambda)$ in the hazard function setting of Section 2. From the properties of $\Delta_{j,k}(\lambda)$ at a sufficiently large resolution level j_1 we then derive an estimation procedure for the location of the change-point of λ. The level j_1 should be chosen in such a way that the global energy of λ at the level j_1 is of the same order as the stochastic component of $\hat{\Delta}_{j,k}(\lambda)$, and thus the local energy of λ will be significantly larger around the location of the change-point. The procedure is similar in spirit to the one used by Wang (1995) for detecting cusps in a classical nonparametric regression setting. See also Antoniadis and Gijbels (1997).

Using the notation of Section 3, we may write:

$$\Delta_{j,k}(\lambda) = \langle \lambda, \Phi_{j,k} \rangle = \int \lambda(x)\Phi_{j,k}(x)dx = \int \Phi_{j,k}(x)d\Lambda(x).$$

The support of $\Phi_{j,k}$ is included within the interval $\left[\frac{\tau(k)-1}{2^j}, \frac{\tau(k+2)+1}{2^j}\right]$. Using the integration by parts formula (2.5) we obtain:

$$\begin{aligned}
\Delta_{j,k}(\lambda) &= \int_{\left[\frac{\tau(k)-1}{2^j}, \frac{\tau(k+2)+1}{2^j}\right]} \Phi_{j,k}(x)d\Lambda(x) = -\int_{\left[\frac{\tau(k)-1}{2^j}, \frac{\tau(k+2)+1}{2^j}\right]} \Lambda(x^-)d\Phi_{j,k}(x) \\
&\quad + \Phi_{j,k}\left(\frac{\tau(k+2)+1}{2^j}\right)\Lambda\left(\frac{\tau(k+2)+1}{2^j}\right) - \Phi_{j,k}\left(\frac{\tau(k)-1}{2^j}\right)\Lambda\left(\frac{\tau(k)-1}{2^j}\right) \\
&= -\int \Lambda(x)d\Phi_{j,k}(x).
\end{aligned}$$

This alternative form of $\Delta_{j,k}(\lambda)$ allows us now to define an appropriate version of the empirical coefficient differences, namely:

$$\hat{\Delta}_{j,k}(\lambda) = \hat{\Delta}_{j,k} = -\int \hat{\Lambda}_n(u^-)d\Phi_{j,k}(u) = \int \Phi_{j,k}(x)d\hat{\Lambda}_n(u),$$

with $\hat{\Lambda}_n$ the Nelson estimator of Λ as defined in Section 2. For practical use it is more convenient to rewrite $\hat{\Delta}_{j,k}$, as follows:

$$\hat{\Delta}_{j,k} = \sum_{i=1}^{n} \Phi_{j,k}(X_{(i)})\frac{\delta_{(i)}}{n-i+1}. \tag{4.1}$$

We will use the notation $\bar{\xi}_n(x)$ for $n^{-1}\sum_{i=1}^{n}\xi_i(x)$ with $\xi_i(x)$ the random variables defined in (2.2). Using the strong representation result (2.3) for $\hat{\Lambda}_n$ we may write:

$$\begin{aligned}
\hat{\Delta}_{j,k} &= -\int \Lambda(u)d\Phi_{j,k}(u) - \int \bar{\xi}_n(u^-)d\Phi_{j,k}(u) - \int r_n(u^-)d\Phi_{j,k}(u) \\
&= \Delta_{j,k}(\lambda) - \int \bar{\xi}_n(u^-)d\Phi_{j,k}(u) - \int r_n(u^-)d\Phi_{j,k}(u),
\end{aligned}$$

which splits the empirical coefficient differences into a deterministic term and a stochastic term. We already know that the deterministic term has an exponential decay in j for any Hölder continuous function. What remains to be seen is the behavior of the stochastic term in order to select the appropriate level j_1 for which

$$\max_{k=0,\ldots,(2^{j_1}-6)/2} |\Delta_{j_1,k}(\lambda)|$$

is of the same order as

$$\max_{k=0,\ldots,(2^{j_1}-6)/2} |V_{j_1,k}|,$$

where

$$V_{j,k} = -\int \bar{\xi}_n(u^-)d\Phi_{j,k}(u) - \int r_n(u^-)d\Phi_{j,k}(u).$$

Note that, for any level j the coefficients are only computed for $k = 0, \ldots, (2^j - 6)/2$, since these are the only indices for which the support of $\Phi_{j,k}$ overlaps with the interval $[0,1]$.

To study the random variable $V_{j,k}$ we will separate it into two parts $U_{j,k}$ and $W_{j,k}$ defined by:

$$\begin{aligned} U_{j,k} &= -\int \bar{\xi}_n(u^-)d\Phi_{j,k}(u) \\ W_{j,k} &= -\int r_n(u^-)d\Phi_{j,k}(u). \end{aligned}$$

Since the stochastic process $\bar{\xi}_n$ is centered, the random variables $U_{j,k}$ have zero-mean. Let k and ℓ be two indices between 0 and $(2^j - 6)/2$ and let us compute for the same resolution level j, the covariance of $U_{j,k}$ and $U_{j,\ell}$, $\mathrm{cov}(U_{j,k}, U_{j,\ell}) = I\!\!E(U_{j,k}U_{j,\ell})$. Our first task is to show that the covariance of the $U_{j,k}$ and $U_{j,\ell}$ ($k \neq \ell$) is zero. (See proof in the Appendix).

When $k = \ell$, using the zero moment property of ψ and Fubini's formula we obtain:

$$\begin{aligned} \mathrm{var}(U_{j,k}) &= 2\frac{2^j}{n}\int_{2k}^{2k+2} g(v/2^j)\psi(v-2k-1)(\phi(1)-\phi(v-2k-1))dv \\ &\quad +2\frac{2^j}{n}\int_{2k+4}^{2k+6} g(v/2^j)\psi(v-2k-5)(\phi(1)-\phi(v-2k-5))dv \\ &= -2\frac{2^j}{n}\int_{-1}^{1}\left[g(2^{-j}(v+2k+5)) + g(2^{-j}(v+2k+1))\right]\psi(v)\phi(v)dv. \end{aligned}$$

By definition the function g is differentiable. Using Taylor's formula with integral remainder we obtain the following bound for $\mathrm{var}(U_{j,k})$:

$$\begin{aligned} \mathrm{var}(U_{j,k}) &\leq 2\frac{2^j}{n}\int_{-1}^{1}\frac{1}{2^j}\left| v\int_0^1 g'(2^{-j}(2k+1+vw)dw \right. \\ &\quad \left. +(v+4)\int_0^1 g'(2^{-j}(2k+1+(v+4)w)dw\psi(v)\right|\phi(v)dv \leq \frac{C}{n}, \end{aligned}$$

where

$$
\begin{aligned}
C &= 2 \max_{j;k=0,\dots,(2^j-6)/2} \int_{-1}^{1} \Big| v \int_0^1 g'(2^{-j}(2k+1+vw)dw \\
&\quad +(v+4)\int_0^1 g'(2^{-j}(2k+1+(v+4)w)dw\psi(v)\Big| \phi(v)dv,
\end{aligned}
\tag{4.2}
$$

which is finite since g' is bounded on $[0,1]$.

By asymptotic normality of the sequence of stochastic processes $(\bar{\xi}_n(x); x \in [0,1])$, it follows that for any resolution level j, the sequence $(U_{j,k})_{k=0,\dots,(2^j-6)/2}$ is asymptotically (in n) a sequence of Gaussian independent random variables. Using the extremal results for independent normal variables of Leadbetter, Lindgren and Rootzén (1983), and letting j tend to infinity at a $\log_2 n$ rate we have:

$$
\lim_{n\to\infty} \mathrm{P}\left(\max |U_{j,k}| < \frac{(2C\log n)^{1/2}}{n^{1/2}}\right) = 1. \tag{4.3}
$$

To end the study of the stochastic part of $\hat{\Delta}_{j,k}$, note that

$$
\begin{aligned}
|W_{j,k}| &= |\int r_n(u^-)d\Phi_{j,k}(u)| \\
&\leq \sup |r_n(u)| \int |d\Phi_{j,k}(u)| \leq 2^{j/2} \sup |r_n(u)| \int |\psi(x)|dx.
\end{aligned}
$$

Now by taking $2^{-j} \simeq \left(\frac{\log n}{n}\right)^{1/(1+2\alpha)}$ and recalling that by Lo *et al.* (1989), $\|r_n\|_\infty = \mathcal{O}\left(\frac{\log n}{n}\right)$ a.s., it is easy to see that the above integral behaves as $o((\log n/n)^{1/2})$. By grouping the upper bounds of the stochastic terms $U_{j,k}$ and $W_{j,k}$ and taking

$$
j_1(n) = \frac{\log_2(n/\log n)}{2\alpha+1}
$$

we conclude that:

$$
\lim_{n\to\infty} \mathrm{P}\left(\max_{k=0,\dots,(2^{j_1}-6)/2} |V_{j_1,k}| \leq \mathcal{O}\left(\frac{(\log n)^{1/2}}{n^{1/2}}\right)\right) = 1. \tag{4.4}
$$

We may now state the following Proposition whose proof is given in the Appendix:

Proposition 4.1 *Let $\lambda \in \mathcal{F}_{s,\alpha}(K,L,M,[0,1])$ and $j_1(\alpha) = \frac{\log_2(n/\log n)}{2\alpha+1}$. Let k_1 be the unique integer such that $\theta_\lambda \in [2^{-j_1}(\tau(k_1)+1), 2^{-j_1}(\tau(k_1+2)+1)]$. Then there exist two positive constants C_1' and C_2' such that*

(i)

$$\max_{k\in\{0,\ldots,(2^{j_1}-6)/2\};|k-k_1|\geq 2} \left|\Delta_{j_1,k}(\lambda)\right| \leq C_1'\left(\frac{\log n}{n}\right)^{1/2}$$

(ii)

$$\sup_{k\in\{0,\ldots,(2^{j_1}-6)/2\};|k-k_1|<2} \left|\Delta_{j_1,k}(\lambda)\right| \geq C_2'\left(\frac{\log n}{n}\right)^{1/2}.$$

Our change-point location estimator is now defined as follows. Let $j_1(\alpha, n)$ be the resolution level given in Proposition 4.1. We set

$$\hat{k}_n = \arg\max_{k\in\{0,\ldots,(2^{j_1}-6)/2\}} \left|\hat{\Delta}_{j_1,k}\right|, \tag{4.5}$$

and

$$\hat{\theta}_\lambda = \frac{\tau(\hat{k}_n) + \tau(\hat{k}_n + 2) + 2}{2^{j_1+1}}. \tag{4.6}$$

Our next theorem proves that $\hat{\theta}_\lambda$ as defined by (4.6) is consistent with the rate $(\log n/n)^{1/(2\alpha+1)}$ which is a little bit faster than the rate obtained by Müller and Wang (1994).

Theorem 4.1 *Using the notation of Proposition 4.1, and assuming further that the constants K and L are such that $C_2' \geq C_1' + 2\sqrt{2C}$, we have:*

$$\lim_{n\to\infty} \sup_{\lambda\in\mathcal{F}_{s,\alpha}} P\left(v_n^{-1}|\hat{\theta}_n - \theta_\lambda| \leq 6\right) = 1,$$

with $v_n = (\log n/n)^{1/(1+2\alpha)}$.

Recalling the definition of the constants C_1', C_2' and the expression of the variance upper bound C defined in (4.2), the estimator will be consistent if the size of the jump at the jump point is relatively large with respect to the Hölder constant L, since obviously C tends to 0 as n goes to infinity.

The proof of Theorem 4.1 relies on all the previous bounds on the coefficient differences and is quite standard. For details see the Appendix.

The previous result can be further sharpened in order to get rid of the extra $(\log n)^{1/(2\alpha+1)}$ term in the rate of convergence of $\hat{\theta}_n$. This can be done following the same device as the one used in Korostelev (1987), Raimondo (1996), Müller and Song (1997) or Gijbels, Hall and Kneip (1998) in the nonparametric regression setting. To achieve this task we must introduce a narrower grid in the neighborhood of $\hat{\theta}_n$ and update the position of the change-point on this new grid.

Let $j_0(\alpha) = \log_2(n)/(2\alpha+1)$. Note that by definition of $j_1(\alpha)$, we have $j_0(\alpha) > j_1(\alpha)$. Let $\hat{I}_n$ be the interval defined by $[\hat{\theta}_n - 6\,(2^{-j_1})\,, \hat{\theta}_n + 6\,(2^{-j_1}])$ and let us denote by $S_0, \ldots, S_{m_0}$ the points that partition $\hat{I}_n$ into intervals of length $2/2^{j_0}$. From Theorem 4.1 it is known that the probability that $\hat{I}_n$ contains the point θ_λ tends to 1 as n goes to infinity. There exists therefore an index k_0, $0 \leq k_0 \leq m_0 - 1$ such that

$$\mathrm{P}(\theta_\lambda \in [S_{k_0}, S_{k_0+1}[) \rightarrow_{n\rightarrow\infty} 1.$$

To estimate θ_λ it is therefore necessary to determine as accurately as possible this index k_0, i.e. in such a way that

$$\mathrm{P}(|\hat{k}_0 - k_0| < c) \rightarrow_{n\rightarrow\infty} 1,$$

with c an arbitrary positive number. The updated estimate of θ_λ will be $\hat{\theta}_n^* = S_{\hat{k}_0}$.

Computations that are similar to those used by the above cited authors, once the behavior of the stochastic part of the $\hat{\Delta}_{j,k}$ is controlled, reduces the problem of estimating k_0 to the problem of estimating a jump in a sequence of independent Bernoulli random variables. To construct the estimator $\hat{\theta}_n^*$ we may proceed as follows:

a) We first translate the center of the transform toward the point $S_0 = \hat{\theta}_n - 6(\log n/n)^{1/(2\alpha+1)}$, i.e.

$$\phi^c_{j,\tau(k)}(x) = 2^{j/2}\phi((2^j(x - S_0) - \tau(k)).$$

b) We construct independent Bernoulli random variables at level $j_0(\alpha)$ by

$$\eta_k = \mathbf{I}_{\{\sqrt{n}|\hat{\Delta}^*_{j_0,k}| > C^*\}}$$

with C^* an appropriate constant and where

$$\hat{\Delta}^*_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\left(\phi^c_{j,\tau(k)}(X_{(i)}) - \phi^c_{j,-\tau(k)}(X_{(i)})\right)\frac{\delta_{(i)}}{n-i+1}.$$

c) We take

$$\hat{k}_0 = \arg \min_{\ell \in \{0,\dots,m_0-1\}} \left(\sum_{k=0}^{\ell} \eta_k + \sum_{k=\ell+1}^{m_0-1} (1-\eta_k) \right),$$

and put

$$\hat{\theta}_n^* = S_{\hat{k}_0}.$$

With the above notation we then have

Proposition 4.2

$$\lim_{C^* \to \infty} \limsup_{n \to \infty} \sup_{\lambda \in \mathcal{F}_{s,\alpha}} P_\lambda(n^{1/(2\alpha+1)} |\hat{\theta}_n^* - \theta_\lambda| > C^*) = 0$$

The proof of this theorem follows step by step the one given for example by Raimondo (1996) in the nonparametric regression case and is therefore omitted.

5 Simulations and some real examples

It follows from the proof of Theorem 4.1 that knowledge about the constant C is really important for the practical application of the detection algorithm outlined in the previous sections.

This constant C (up to a factor $1/n$) appears as an upper bound on the variances of the variables $U_{j,k}$. Since the error term $W_{j,k}$ is uniformly bounded above by $\log n/n$, this upper bound on the variances of the $U_{j,k}$'s does not differ much from the upper bound on the variances of the $V_{j,k}$, which denotes the stochastic part of the estimates $\hat{\Delta}_{j,k}$. One way to derive this upper bound for practical purposes is to bootstrap the distribution of the $\hat{\Delta}_{j,k}$, since the stochastic part of these variables is unaffected by the presence of a jump or a cusp. We have seen that most of our results rely upon the almost sure representation of the Nelson estimator of the cumulative hazard function. In order to compute the variances of the $\hat{\Delta}_{j,k}$ it is sufficient to approximate the distribution of $\hat{\Lambda}_n$ by an appropriate consistent bootstrap approximation that also admits an almost sure representation similar to the one used in Section 2.

The bootstrap procedure can be described as follows: we resample with replacement pairs (X_i^*, δ_i^*) from (X_1, δ_1), ..., (X_n, δ_n). Based on the bootstrap sample the analogue $\hat{\Lambda}_n^*$ of the Nelson estimator is easily computed. In case of ties we adopt the usual convention that uncensored observations are considered to occur just before censored observations. Most of the conditions of Theorem 4.1 of Van Keilegom and Veraverbeke (1997) hold, even in the presence of a discontinuity and therefore, asymptotically in n, the strong representation of the bootstrapped cumulative hazard function holds with bounds similar to those of the true sample. A good estimate of C is therefore obtained by computing, for the relevant resolution j_1 all the associated k's the variances of the bootstrapped samples $\hat{\Delta}_{j_1,k}^{*b}, b = 1, \ldots, B$.

6 Simulation examples

A small simulation study was carried out to explore the properties of our change-point detection procedure. The following case was considered for the hazard function:

$$\lambda(x) = \lambda_0 + \theta \mathbf{I}_{[\tau,\infty[}(x), \tag{6.1}$$

where $\lambda_0 > 0$ and $\lambda_0 + \theta > 0$, and τ is a change-point parameter. This is a typical model when testing for a constant hazard rate against alternatives with hazard rates involving a single change-point. In the simulation, we generated 100 random samples for each of the following parameter settings:

(a) $\lambda_0 = 4$, $\theta = -3$ and $\tau = 0.4$

(b) $\lambda_0 = 1$ and $\theta = 0$.

The parameter values in (a) might be typical of a clinical trial in which we would expect a high initial rate and a lower hazard rate after the treatment has been in place for some time. Model (b) with no change-point was designed to test the sensitivity of our detection procedure. Independent random censoring times were generated from an exponential distribution with a parameter $\lambda = 0.69$ selected to give an expected censoring proportion of 25 %. For each Monte Carlo run we generated samples of size $n = 500$. The number of bootstrap samples used to compute the variances of the estimated wavelet coefficients was set to 150. All computations were done in Matlab.

The simulation results for the two cases are summarized below. For model (a), the change-point could not be located six times among the 100 simulations (the standardized wavelet difference coefficients were below the threshold). For the remaining 94 runs, the results are given in Table 1 and graphically summarized with the boxplot of Figure 3.

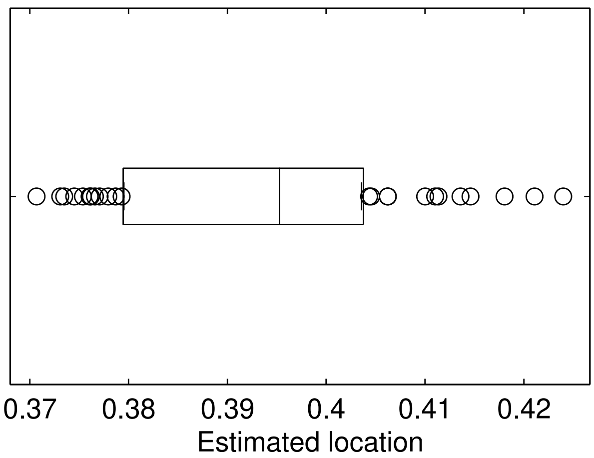

Figure 3. Boxplot of the estimated location of the change-point for the 94 successful Monte Carlo runs.

Mean	Std	Median	5-percentile	95-percentile
0.39398224	0.01429710	0.39528437	0.37345025	0.41818359

Table 1: Summary statistics for the estimation of change-point location by our wavelet procedure over 94 out of the 100 Monte Carlo runs under the simulation setting (a).

For case (b), designed to see how often our procedure detects change-points in models without any change-point, among the 100 Monte Carlo runs, an estimated discontinuity was detected 3 times, every time at the right end of the time interval containing the uncensored observations and suggesting that much larger bootstrap samples are needed to properly estimate the variance of wavelet difference coefficients when the data are heavily censored.

The above results indicate that the change-point detection procedure in the presence of censoring yields good results. One difficultly, however, occurs in the estimation of the exact location of the discontinuity. This is due to the discrete nature of our algorithm, which detects the possible discontinuities within a grid defined by the resolution $j_1 = \log_2(n/(log(n))$, which, for moderate sample sizes, can be quite large. The larger the sample size, the narrower will be the grid and a more precise estimate of the location will then be obtained. One way to achieve this is to introduce a narrower grid in the neighborhood of the estimated location of the change-point and update its position on this new grid, as suggested at the end of Section 4. However, this was not implemented here because it substantially increases the number of numerical computations.

7 Real data examples

To illustrate the method developed in this paper we apply our procedure to two real datasets.

First we consider a version of the Stanford heart transplant data from Loader (1991), which is a subset of the full dataset given in Cox and Oakes (1984) and originally analyzed by Miller and Halpern (1982). It concerns 184 patients who received a heart transplant. Of these, 119 died during the follow-up period, and 65 were censored. This dataset was previously analyzed by Loader (1991), using likelihood ratio type tests and large deviation approximations in change-point models.

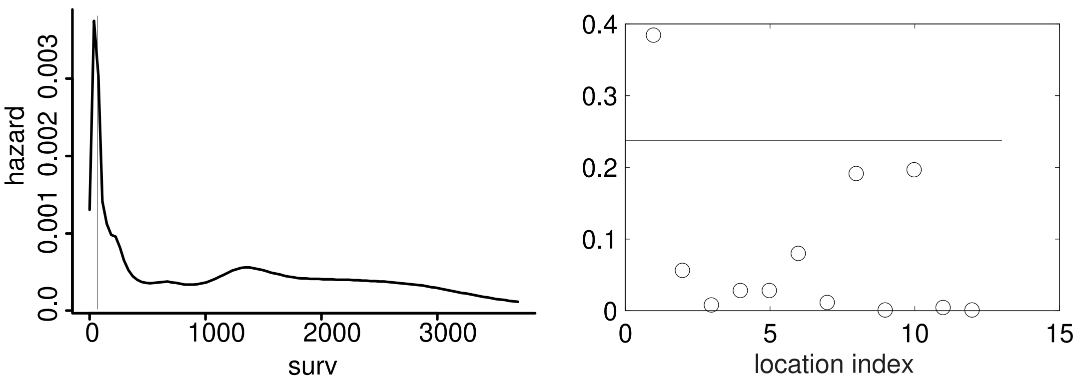

Figure 4. Left plot: local likelihood hazard estimation applied to heart transplant data. Right plot: absolute values of the standardized wavelet coefficient differences. The horizontal line indicates the threshold $\sqrt{2\log(n)/n}$ above which a coefficient is significantly detecting a discontinuity. The vertical line in the left plot indicates the estimated location.

In the left panel of Figure 4 we report a plot of the estimated hazard rate using the local likelihood method of Loader (1997), as it is implemented in the Splus library Locfit. The absolute values of the wavelet coefficient differences, standardized by our bootstrap procedure, are plotted on the right panel. The horizontal line indicates the threshold $\sqrt{2\log(n)/n}$ above which a coefficient difference is judged as significantly detecting a discontinuity. For this particular dataset, the estimated location $\hat{\tau}$ of the discontinuity is 71.75 days. Given the moderate sample size, which implies a wide detection grid, this result is in a good agreement with the results of Loader (1991) who showed a precipitous drop in the hazard rate function around 68 days. In fact, the value 71.75 is well within the approximate 95% confidence interval for τ determined by Loader in the above mentioned paper.

The second example used to illustrate our method concerns a clinical trial in which nonleukaemic causes of death after the bone marrow transplant were studied (see Brochstein *et. al.*, (1987)). A subset of this dataset was analyzed by Müller and Wang (1990) to illustrate a kernel-based nonparametric method for detecting the point the most rapid change of a hazard rate. We first applied our method to the same data subset used by Müller and Wang, concerning 53 patients, aged 20 years or above, with acute lymphocytic leukaemia after bone marrow transplantation. The lifetime is the time in months from bone marrow transplantation to death due to nonleukaemic causes. Of these 53 patients, 41 died of nonleukaemic causes by the end of the study and 12 were thus censored. A hazard rate plot obtained by Locfit is given on the left panel of Figure 5. The right panel plot displays the absolute values of the standardized wavelet coefficient differences and our procedures indicates that a change-point occurs at 2.9 months. Considering that no observations occur between the observed times 2.79 and 4.17, and given that the transition from one hazard to the other is quite smooth, our estimation is in close agreement with the one found by Müller and Wang at around 3.5 months. It should be noted also that the wavelet procedure does not support the existence of a relatively minor second change-point at approximately 10.4 months as suggested by Müller and Wang.

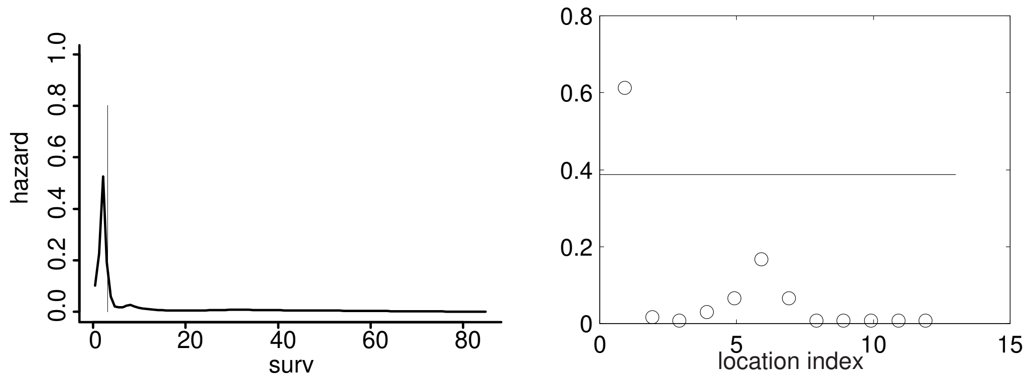

Figure 5. Left plot: local likelihood hazard estimation applied to the leukaemia data. Right plot: absolute values of the standardized wavelet coefficient differences. The horizontal line indicates the threshold $\sqrt{2\log(n)/n}$ above which a coefficient is significantly detecting a discontinuity. The vertical line in the left plot indicates the estimated location of the change-point.

We have also applied our procedure to the subset of the 73 "young" patients, aged less than 19.9 years. Of these 73 "young" patients 22 died from nonleukaemic causes. This is therefore a dataset with heavy censoring. Our procedure did not detect any change-point in the hazard rate for this dataset. The maximum of the absolute values of the standardized by bootstrap wavelet differences coefficients was found to be 0.119 which is well below the threshold 0.3429. Concatenating the two subsets, "old" and "young" into one set and applying our procedure leads to the same discontinuity position as for the “old” subset.

In conclusion, the method presented here seems to be well suited as an exploratory tool for detecting possible change-points in hazard rates from censored data.

Acknowledgements. Anestis Antoniadis would like to thank the Institute of Statistics at Louvain-la-Neuve for its warm hospitality and financial support during the completion of this work. Financial support by the contract “Projet d'Actions de Recherches Concertées”, No. 93/98-164 of the Belgian Government is gratefully acknowledged. The third author wishes to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada. All three authors are grateful to the Sloan Kettering Memorial Cancer Center, to Dr R. J. O'Reilly, and to Dr. Susan Groshen of the University of Southern California, Department of Preventive Medicine for making available to us the bone marrow transplant data (see Brochstein *et. al.*, (1987)) that we used to illustrate our method here.

A Proof of assertions *(i)* and *(ii)* about the wavelet coefficients differences

Note that by a change of variables we have:

$$\Delta_{j,k}(f) = 2^{-j/2}\int_{-1}^{1} f\left(\frac{x+\tau(k)}{2^j}\right)\phi(x)dx - 2^{-j/2}\int_{-1}^{1} f\left(\frac{x+\tau(k+2)}{2^j}\right)\phi(x)dx.$$

We can now take advantage of the Hölder continuity of f away from its jump point or cusp. By adding and subtracting $f\left(\frac{\tau(k)}{2^j}\right)$ and $f\left(\frac{\tau(k+2)}{2^j}\right)$ inside the integrals we obtain:

$$\begin{aligned}\Delta_{j,k}(f) &= 2^{-j/2}\left[f\left(\frac{\tau(k)}{2^j}\right) - f\left(\frac{\tau(k+2)}{2^j}\right)\right] \\ &+2^{-j/2}\int_{-1}^{1}\left[f\left(\frac{v+\tau(k)}{2^j}\right) - f\left(\frac{\tau(k)}{2^j}\right)\right]\phi(v)dv \\ &-2^{-j/2}\int_{-1}^{1}\left[f\left(\frac{v+\tau(k+2)}{2^j}\right) - f\left(\frac{\tau(k+2)}{2^j}\right)\right]\phi(v)dv,\end{aligned}$$

since $\int\phi(x)dx = 1$ according to our assumptions on ϕ.

Let

$$D_{j,k}(f) = 2^{-j/2}\left[f\left(\frac{\tau(k)}{2^j}\right) - f\left(\frac{\tau(k+2)}{2^j}\right)\right], \tag{A.1}$$

and

$$R_{j,k}(f) = 2^{-j/2}\int_{-1}^{1}\left[f\left(\frac{v+\tau(k)}{2^j}\right) - f\left(\frac{\tau(k)}{2^j}\right)\right]\phi(v)dv. \tag{A.2}$$

Using expressions (A.1) and (A.2) we can therefore write:

$$\Delta_{j,k}(f) = D_{j,k}(f) + R_{j,k}(f) - R_{j,k+2}(f). \tag{A.3}$$

For any index k, $0 \le k \le (2^j-6)/2$, such that $k < k_1 - 1$ it is easy to see that, for any $v \in [-1,+1]$ we have:

$$\frac{\tau(k+2)+v}{2^j} \le \frac{\tau(k_1)+1}{2^j}.$$

Similarly, for $k > k_1+1$, we have

$$\frac{\tau(k)+v}{2^j} \ge \frac{\tau(k_1)+3}{2^j}.$$

It follows that for any k such that $|k-k_1| \ge 2$ and using the facts that $f \in \mathcal{F}_{s,\alpha}$ and $\theta_f \in \left[\frac{\tau(k_1)+1}{2^j}, \frac{\tau(k_1)+3}{2^j}\right]$, the function f is Hölder continuous with exponent s within the intervals of integration defining $R_{j,k}(f)$ and $R_{j,k+2}(f)$. We therefore have

$$|R_{j,k}(f)| \leq 2^{-j/2} L \int_{-1}^{1} \phi(v) \frac{|v|^s}{2^{js}} dv = 2^{-j(s+1/2)} L \int_{-1}^{1} \phi(v)|v|^s dv,$$

and

$$|R_{j,k+2}(f)| \leq 2^{-j(s+1/2)} L \int_{-1}^{1} \phi(v)|v|^s dv.$$

By the triangular inequality we therefore have:

$$|R_{j,k}(f) - R_{j,k+2}(f)| \leq 2^{-j(s+1/2)} 2L \int_{-1}^{1} \phi(v)|v|^s dv.$$

We also have

$$|D_{j,k}(f)| \leq 2^{-j/2} L | \frac{\tau(k) - \tau(k+2)}{2^j} |^s = 2^{-j(s+1/2)} 4^s L.$$

Combining all these upper bounds we obtain for indices k such that $|k - k_1| \geq 2$:

$$|\Delta_{j,k}(f)| \leq 2^{-j(s+1/2)} (4^s L + 2L \int_{-1}^{1} \phi(v)|v|^s dv). \qquad \text{(A.4)}$$

The constant C_1 is therefore given by $C_1 = (4^s L + 2L \int_{-1}^{1} \phi(v)|v|^s dv)$.

For $0 \leq k \leq (2^j - 6)/2$ with $k = k_1$, note first that, for all $v \in [-1, +1]$:

$$\frac{\tau(k_1) + v}{2^j} \leq \frac{\tau(k_1) + 1}{2^j} \qquad \text{and} \qquad \frac{\tau(k_1 + 2) + v}{2^j} \geq \frac{\tau(k_1) + 3}{2^j}.$$

Now, since

$$\theta_f \in \left[\frac{\tau(k_1) + 1}{2^j}, \ \frac{\tau(k_1) + 3}{2^j} \right] \subset \left[\frac{\tau(k_1)}{2^j}, \ \frac{\tau(k_1 + 2)}{2^j} \right],$$

we have

$$|D_{j,k}(f)| \geq 2^{-j/2} K | \frac{\tau(k_1)}{2^j} - \frac{\tau(k_1 + 2)}{2^j} |^\alpha = 2^{-j(\alpha+1/2)} 4^\alpha K.$$

By the triangular inequality, for $k = k_1$ and for j large enough we have:

$$|\Delta_{j,k}(f)| \geq 2^{-j(\alpha+1/2)} 4^\alpha K - 2^{-j(s+1/2)} 2L \int_{-1}^{1} \phi(v)|v|^s dv.$$

By our assumptions $s > \alpha$. Hence,

$$\begin{aligned} |\Delta_{j,k}(f)| &\geq 2^{-j(\alpha+1/2)} 4^\alpha K - 2^{-j(s-\alpha+\alpha+1/2)} L \int_{-1}^{1} \phi(v)|v|^s dv \\ &= 2^{-j(\alpha+1/2)} \left[4^\alpha K - 2^{-j(s-\alpha)} L \int_{-1}^{1} \phi(v)|v|^s dv \right]. \end{aligned}$$

For j large enough $4^\alpha K - 2^{-j(s-\alpha)} L \int_{-1}^{1} \phi(v)|v|^s dv \geq K/2$. Since

$$\max_{|k-k_1|<2} |\Delta_{j,k}(f)| \geq |\Delta_{j,k_1}(f)|,$$

this finishes the proof of (ii) with $C_2 = K/2$.

B Proof of $\mathrm{cov}(U_{j,k}, U_{j,\ell}) = 0$ $(k \neq \ell)$

Recalling the covariance structure of the ξ's and their independence, we obviously have:

$$\begin{aligned}
\mathrm{cov}(U_{j,k}, U_{j,\ell}) &= \frac{1}{n}\int\int g(\min(u,v))d\Phi_{j,k}(u)d\Phi_{j,k}(v) \\
&= \frac{1}{n}\int\int g(\min(u,v)) \\
&\quad \Big[d\phi_{j,\tau(k+2)}(u)d\phi_{j,\tau(\ell+2)}(v) - d\phi_{j,\tau(k)}(u)d\phi_{j,\tau(\ell+2)}(v) \\
&\quad -d\phi_{j,\tau(k+2)}(u)d\phi_{j,\tau(\ell)}(v) + d\phi_{j,\tau(k)}(u)d\phi_{j,\tau(\ell)}(v)\Big].
\end{aligned}$$

Now,

$$d\phi_{j,k}(u) = d\left(2^{j/2}\phi(2^j u - k)\right) = 2^j\psi_{j,k}(u).$$

Therefore, with a change of variables, we have:

$$\begin{aligned}
\mathrm{cov}(U_{j,k}, U_{j,\ell}) &= \frac{2^j}{n}\int\int g(\min(u/2^j, v/2^j)) \\
&\quad [\psi(u-2k-5)\psi(v-2\ell-5) - \psi(u-2k-5)\psi(v-2\ell-1) \\
&\quad -\psi(u-2k-1)\psi(v-2\ell-5) + \psi(u-2k-1)\psi(v-2\ell-1)]\,du\,dv.
\end{aligned}$$

The support of ψ is the interval $[-1,+1]$ and induces a decomposition of the domain of integration into $([2k,2k+2]\cup[2k+4,2k+6])\times([2\ell,2\ell+2]\cup[2\ell+4,2\ell+6])$. It follows that:

$$\begin{aligned}
\mathrm{cov}(U_{j,k}, U_{j,\ell}) &= \frac{2^j}{n}\int_{2k}^{2k+2}\int_{2\ell}^{2\ell+2} g(\min(u/2^j, v/2^j))\psi(u-2k-1)\psi(v-2\ell-1)dudv \\
&\quad -\frac{2^j}{n}\int_{2k}^{2k+2}\int_{2\ell+4}^{2\ell+6} g(\min(u/2^j, v/2^j))\psi(u-2k-1)\psi(v-2\ell-5)dudv \\
&\quad -\frac{2^j}{n}\int_{2k+4}^{2k+6}\int_{2\ell}^{2\ell+2} g(\min(u/2^j, v/2^j))\psi(u-2k-5)\psi(v-2\ell-1)dudv \\
&\quad +\frac{2^j}{n}\int_{2k+4}^{2k+6}\int_{2\ell+4}^{2\ell+6} g(\min(u/2^j, v/2^j))\psi(u-2k-5)\psi(v-2\ell-5)dudv.
\end{aligned}$$

When $k \neq \ell$ (say $\ell > k$ without any loss of generality), and since $2k+2 \leq 2\ell$, we have

$$\begin{aligned}
&\frac{2^j}{n}\int_{2k}^{2k+2}\int_{2\ell}^{2\ell+2} g(\min(u/2^j, v/2^j))\psi(u-2k-1)\psi(v-2\ell-1)dudv \\
&\quad = \frac{2^j}{n}\int_{2k}^{2k+2} g(u/2^j)\psi(u-2k-1)du\int_{2\ell}^{2\ell+2}\psi(v-2\ell-1)dv = 0,
\end{aligned}$$

since

$$\int_{2\ell}^{2\ell+2} \psi(v-2\ell-1)dv = \int_{-1}^{1} \psi(v)dv$$

and ψ is a wavelet. Similar arguments on the other integrals involved in the $\mathrm{cov}(U_{j,k}, U_{j,\ell})$ lead finally to

$$\mathrm{cov}(U_{j,k}, U_{j,\ell}) = 0, \qquad \text{if } k \neq \ell.$$

C Proof of Proposition 4.1

Using the inequalities of Section 3, we have:

$$\sup_{k\in\{0,\ldots,(2^{j_1}-6)/2\};|k-k_1|\geq 2} \left|\Delta_{j_1,k}(\lambda)\right| \leq C_1 2^{-j_1(s+1/2)}$$

and since $j_1(\alpha) \to \infty$ as n tends to ∞,

$$\sup_{k\in\{0,\ldots,(2^{j_1}-6)/2\};|k-k_1|<2} \left|\Delta_{j_1,k}(\lambda)\right| \geq C_1 2^{-j_1(\alpha+1/2)}.$$

Now at $j_1 = j_1(\alpha)$, for any $s > \alpha$ we have

$$2^{-j_1(s+1/2)} < 2^{-j_1(\alpha+1/2)} = \left(\frac{\log n}{n}\right)^{1/2},$$

and the result follows. □

D Proof of Theorem 4.1

Let

$$\mathcal{A}_n = \left\{\omega\,;\, \max_{k\in\{0,\ldots,(2^{j_1}-6)/2\}} |V_{j_1,k}(\omega)| \leq \sqrt{C}\left(\frac{2\log n}{n}\right)^{1/2}\right\}.$$

Using Proposition 4.1 and the triangular inequality on

$$\hat{\Delta}_{j_1,k}(\omega) = \Delta_{j_1,k}(\lambda) + V_{j_1,k}(\omega)$$

we have, for any $\omega \in \mathcal{A}_n$:

$$\max_{k\in\{0,\ldots,(2^{j_1}-6)/2\};|k-k_1|\geq 2} \left|\hat{\Delta}_{j_1,k}(\omega)\right| \leq (C_1' + \sqrt{2C})\left(\frac{\log n}{n}\right)^{1/2}$$

and

$$\max_{k\in\{0,\ldots,(2^{j_1}-6)/2\};|k-k_1|<2} |\hat{\Delta}_{j_1,k}(\lambda)| \geq (C_2' - \sqrt{2C})\left(\frac{\log n}{n}\right)^{1/2}.$$

The maximum of $|\hat{\Delta}_{j_1,k}(\lambda)|$ will therefore be achieved for some $\hat{k}_n(\omega)$ such that $|\hat{k}_n(\omega) - k_1| < 2$ as soon as $C_2' \geq C_1' + 2\sqrt{2C}$. It therefore follows that

$$\begin{aligned}\mathcal{A}_n \;&\subset\; \left\{\omega\,;\,|\frac{\tau(\hat{k}_n)+\tau(\hat{k}_n+2)+2}{2^{j_1+1}} - \frac{\tau(k_1)+\tau(k_1+2)+2}{2^{j_1+1}}| \leq \frac{4}{2^{j_1}}\right\}\\ &\subset\; \left\{\omega\,;\,|\hat{\theta}_n - \frac{\tau(k_1)+\tau(k_1+2)+2}{2^{j_1+1}}| \leq \frac{4}{2^{j_1}}\right\} \subset \left\{\omega\,;\,|\hat{\theta}_n - \theta_\lambda| \leq \frac{6}{2^{j_1}}\right\}\end{aligned}$$

by definition of k_1. The result now follows by Proposition 4.1 □

REFERENCES

Antoniadis, A. and Gijbels, I. (1997). Detecting abrupt changes by wavelet methods. Discussion paper 9716, Institute of Statistics, Louvain-la-Neuve.

Antoniadis, A., Grégoire, G. and McKeague, I. (1994). Wavelet methods for curve estimation. *Jour. Amer. Statist. Assoc.*, **89**, 1340–1353.

Antoniadis, A., Grégoire, G. and Nason, G. (1999). Density and hazard rate estimation for right censored data using wavelet methods. *Jour. Royal Statist. Society*, Ser. B, **61**, to appear.

Brochstein, J.A., Kernan, N.A., Groshen, S., Cirrincione, C., Shank, B., Emanual, D., Laver, J. and O'Reilly, R.J. (1987). Allogenic bone marrow transplantation after hyperfractionated total body irradiation and cyclophosphamide in children with acute leukaemia. *New England Journal of Medicine*, **317**, 1618–1624.

De Boor, C. (1978). *A practical guide to splines.* Springer-Verlag, New-York.

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

Efron, B. (1980). Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association*, **83**, 414–425.

Földes, A., Rejtö, L. and Winter, B.B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data, II: estimation of density and failure rate. *Periodica Mathematica Hungarica*, Vol 12 (1), 15–29.

Gijbels, I., Hall, P. and Kneip, A. (1998). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.*, to appear.

Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Korostelev, A.P. (1987). On minimax estimation of a discontinuous signal. *Theory of Prob. and Its Applic.*, **32**, 727–730.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes.* Springer-Verlag, New York.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Prob. Th. Rel. Fields*, **71**, 455–465.

Lo, S.-H., Mack, Y.P. and Wang, J.-L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Prob. Th. Rel. Fields*, **80**, 461–473.

Loader, C.R. (1991). Inference for a hazard rate change point. *Biometrika*, **78**, No. 4, 749–757.

Loader, C.R. (1996). Change-point estimation using nonparametric regression. *Ann. Statist.*, **24**, No. 4, 1667–1678.

Loader, C.R. (1997). Locfit: An introduction. *Statistical Computing and Graphics*, **8**, No. 1, 11–17.

Mallat, S. and Hwang, W.L. (1992). Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theory*, **2**, 617–643.

Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **69**, 521–531.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.*, **20**, No. 2, 737–761.

Müller, H.-G. and Song, K.-S. (1997). Two-stage change-point estimators in smooth regression models. *Statist. & Prob. Lett.*, **34**, 323–335.

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.

Müller, H.-G. and Wang, J.-L. (1994). Change-point models for hazard functions. in *Change-point problems*, IMS Lecture Notes, Vol 23, 224–241.

Nelson, W. (1972). Theory and applications of hazard plotting for censored data. *Technometrics*, **14**, 27–52.

Raimondo, M. (1996). Modèles de ruptures: situations non ergodiques et utilisation de méthodes d'ondelettes. Doct. Dissertation, Université Paris VII, France.

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467–491.

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