

ESTIMATION OF BOUNDARY AND DISCONTINUITY POINTS IN DECONVOLUTION PROBLEMS

A. Delaigle and I. Gijbels

University of California at San Diego and Katholieke Universiteit of Leuven

Abstract: We consider estimation of the boundary of the support of a density function when only a contaminated sample from the density is available. This estimation problem is a necessary step when estimating a density with unknown support, different from the whole real line, since then modifications of the usual kernel type estimators are needed for consistent estimation of the density at the endpoints of its support. Boundary estimation is also of interest on its own, since it is the basic problem in, for example, frontier estimation in efficiency analysis in econometrics. The method proposed in this paper can also be used for estimating locations of discontinuity points of a density in the same deconvolution context. We establish the limiting distribution of the proposed estimator as well as approximate expressions for its mean squared error and deduce rates of convergence of the estimator. The finite sample performance of the procedure is investigated via simulation.

Key words and phrases: Asymptotic distribution, boundary points, deconvolution problem, density estimation, diagnostic function, discontinuity points, endpoints, rates of convergence.

1. Introduction

We consider kernel estimation of boundary points and discontinuity points of a density from a contaminated sample of that density, i.e., from a sample that contains measurement errors. The contamination problem, often referred to as a deconvolution problem, has applications in such different fields as chemistry and public health. A so-called deconvolution kernel estimator of the density has been proposed in the literature. This estimator, however, is not consistent at a discontinuity point or at a finite left/right endpoint where the density is discontinuous, and has to be modified by taking these points into account. See for example Zhang and Karunamuni (2000) for the modifications to apply in the case of boundary points. It is necessary to provide good estimators of these boundary points or, more generally, discontinuity points when they are unknown.

Boundary estimation also arises when investigating efficiencies of firms like banks, or public services. These investigations involve estimation of quantities such as the maximum level of output that can be produced for a given level of input, often referred to as an economic frontier estimation problem, but can be

seen as a problem of estimation of the boundary of a density. Many different methods have been proposed to estimate a frontier in the case where the observations do not contain any measurement error, but these methods do generally not provide consistent estimators in the more realistic setting in which frontiers (or boundary points) are to be estimated from data that are contaminated by noise.

Boundary estimation from contaminated data has been studied by Kneip and Simar (1996), Neumann (1997b) and Hall and Simar (2002), for example. These papers, however, focus on very specific contexts, or propose methods that are difficult to implement in practice. Our goal is to provide a method that works in general contexts, and to provide a way to implement the method in practice. The idea is to estimate the boundary point by the maximiser of a certain diagnostic function, and this is related to procedures used in the error-free case to estimate discontinuity points of a density or a regression function.

As with density estimation in the deconvolution context, the behaviour of the proposed estimator depends strongly on the type of error that contaminates the data. See for example Fan (1991c), who considers two classes of error densities called ordinary smooth and supersmooth. We established the rates of convergence of the estimator for the two types of error, but for brevity we only present detailed results for the ordinary smooth case. A detailed treatment of the supersmooth case can be found in a longer version of the paper that is available on the web (Delaigle and Gijbels (2003)).

This paper is organized as follows. In Section 2 we present the problem of boundary point estimation and introduce the estimation procedure. In Section 3 we establish the asymptotic distribution for the estimator and deduce approximate expressions for its bias and variance. In Section 4, the finite sample performance of the procedure is illustrated on simulated examples. The proofs of results are given in Section 5.

2. The estimation method

Suppose we are interested in a density f_X , but we observe an i.i.d. sample Y_1, \dots, Y_n from the density f_Y , where $Y_i = X_i + Z_i$, $i = 1, \dots, n$, and where for all i , Z_i is a r.v. independent of X_i , of known density f_Z , representing the error in the data, and X_i is a r.v. of density f_X . The case where f_Z is totally or partially unknown can also be handled if further information, for example a sample from f_Z itself, is available. See Barry and Diggle (1995), Neumann (1997a) and Li and Vuong (1998) in this connection.

When f_X is continuous, a so-called deconvolving kernel density estimator of f_X has been proposed. Consider a kernel function K and a smoothing parameter

$h = h_n > 0$, depending on n , called the bandwidth. The deconvolving kernel estimator of f_X is

$$\hat{f}_X(x; h) = \frac{1}{nh} \sum_{j=1}^n K_Z\left(\frac{x - Y_j}{h}\right), \tag{2.1}$$

where $K_Z(u) = (2\pi)^{-1} \int e^{-itu} \varphi_K(t) / \varphi_Z(t/h) dt$, with φ_L the Fourier transform (resp. characteristic function) of a function (resp. random variable) L . See Carroll and Hall (1988) and Stefanski and Carroll (1990) for an introduction to this estimator. Throughout this paper, we assume that, for all $t \in \mathbb{R}$, $\varphi_Z(t) \neq 0$. In order to guarantee that the integral in (2.1) exists, we choose K to be a real, continuous and symmetric function, such that φ_K has a compact support $[-B_K, B_K]$, with $0 < B_K < \infty$. For simplicity, we assume that f_Z is symmetric, which ensures that K_Z is real and symmetric. After slight modifications, our results apply to the non-symmetric case as well. See Remark 1 in Section 3.

Here, we consider the case where f_X has one or two finite boundary points and f_X is not continuous in these points. When the data of interest are observed without error, a simple and consistent approach is to estimate the left endpoint τ_1 of the support by the smallest observation $X_{(1)}$, and the right endpoint τ_2 by the largest observation $X_{(n)}$. In the case of measurement error however, the simple estimators $Y_{(1)}$ and $Y_{(n)}$ are not consistent estimators of the boundary points of f_X but rather those of f_Y . If f_Z is supported on a finite interval $[a_Z, b_Z]$, then τ_1 and τ_2 can simply be estimated by $Y_{(1)} - a_Z$ and $Y_{(n)} - b_Z$. In the general case however, we need a more elaborate procedure.

The method we propose uses the fact that a boundary point is a particular discontinuity point of the density. The idea is to use methods that exist in the error-free case to detect a discontinuity point, and adapt them to boundary point estimation with contaminated data. We focus on kernel methods. In the error-free case, several methods have been proposed to detect a discontinuity point. They all estimate a discontinuity point by the maximizer of an appropriate diagnostic function. Chu and Cheng (1996) choose as diagnostic function the difference of two kernel density estimators; Couallier (1999, 2000) uses the derivative of a kernel density estimator. See Müller (1992), Wu and Chu (1993), Gijbels, Hall and Kneip (1999), Goderniaux (2001) and Gijbels and Goderniaux (2004), among others, for similar methods in the regression context.

We propose a diagnostic function based on derivative estimation. For a density f_X with a single boundary point τ , estimate τ by

$$\hat{\tau} = \underset{x}{\operatorname{argmax}} |\hat{J}(x)|, \tag{2.2}$$

where the diagnostic function $\hat{J}(x)$ is proportional to the derivative of the deconvolution kernel density estimator of f_X : $\hat{J}(x) = (nh)^{-1} \sum_{i=1}^n K'_Z((x - Y_i)/h)$. Unlike f_X , the kernel estimate \hat{f}_X is a smooth function, even at τ . In the current context, it is to be expected that this estimate will be continuous, but with large derivatives when approaching the endpoints (large positive derivative for a left endpoint and large negative derivative for a right endpoint).

3. Asymptotic Distribution of the Estimator

Suppose f_X has a single boundary or discontinuity point τ . We show that $\hat{\tau}$ is a consistent estimator of τ and establish its asymptotic law. The proof uses techniques and conditions somewhat similar to those used for proving the consistency of a discontinuity point estimator in the error-free case. See for example Müller (1992), Chu and Cheng (1996) and Couallier (1999). In particular, we assume that τ lies in a compact interval $[A, B]$, and thus our estimator is $\hat{\tau} = \operatorname{argmax}_{x \in [A, B]} |\hat{J}(x)|$.

The asymptotic properties of our estimator depend strongly on the error distribution, which dictates the behaviour of K_Z . In deconvolution problems, it is common to consider two types of error distributions: the ordinary smooth error distributions of order β have a characteristic function $\varphi_Z(t)$ satisfying $d_0|t|^{-\beta} \leq |\varphi_Z(t)| \leq d_1|t|^{-\beta}$ as $|t| \rightarrow \infty$, for some positive constants d_0, d_1 and β ; supersmooth error distributions have characteristic functions decreasing exponentially fast in the tails. See Fan (1991c). For simplicity, we restrict our presentation to the ordinary smooth error case.

For any set $D \subset \mathbb{R}$ and positive integer m , let $\mathcal{C}_m(D)$ denote the set of functions m times continuously differentiable on D and $\mathcal{D}_m(D) = \{f \in \mathcal{C}_m(D) : \sup_{0 \leq j \leq m} \sup_{x \in D} |f^{(j)}(x)| < \infty\}$. For a square integrable function f , let $R(f)$ denote $\int f^2$. Finally, let $d = f_X(\tau^+) - f_X(\tau^-)$ denote the size of the discontinuity of f_X at τ where, for any function g and point $a \in \mathbb{R}$, we denote $g(a^+) = \lim_{x \rightarrow a^+} g(x)$ and $g(a^-) = \lim_{x \rightarrow a^-} g(x)$. Then, we take $r_X = f_X - dI_{[\tau, +\infty[}$, continuous on \mathbb{R} and, in particular, at τ since $r_X(\tau^+) = r_X(\tau^-) = f_X(\tau^-)$.

Condition A.

- (A1) $K \in \mathcal{C}_3(\mathbb{R})$ is a symmetric, k th order kernel ($k \geq 2$ even), such that $\|K\|_\infty = K(0) > \max_{x \neq 0} |K(x)|$, $K''(0) < 0$, and $\int |uK^{(r)}(u)| du < \infty$ for $r = 0, 1, 2, 3$;
- (A2) r_X is Lipschitz continuous with Lipschitz constant L ;
- (A3) f_Y is differentiable on $\mathbb{R} \setminus \{\tau\}$ such that $\sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y^{(\ell)}(x)| < \infty$ for $\ell = 0, 1$;
- (A4) $K_Z \in \mathcal{C}_4(\mathbb{R})$ is such that $\int |K_Z''(u)| du = O(h^{-\beta})$, $\int |u| \cdot |K_Z''(u)|^2 du = O(h^{-2\beta})$ and, for $r = 0, \dots, 4$, $\|K_Z^{(r)}\|_\infty = O(h^{-\beta})$ and $R(K_Z^{(r)}) \sim h^{-2\beta}$;

(A5) $h \rightarrow 0$ as $n \rightarrow \infty$ so that, for some $0 < \delta < 1/2$, $p \in \mathbb{N}_0 = \mathbb{N} \setminus \{0\}$ and $q \in \mathbb{N}_0$, $nh^{(2+2\delta+\beta)/(1+q)} \rightarrow \infty$ and $\sum_{n=1}^{\infty} n^{1+q-p} h^{-4\delta p-p-2\beta p} < \infty$.

A discussion on these and related conditions is given in Delaigle (2003). There it is shown that the conditions can be expressed in a rather simple form if one sacrifices generality of the functions f_Z and K . In particular, Condition (A1) is satisfied by most kernels commonly used in deconvolution problems. See also Fan (1991a,b) and Delaigle and Gijbels (2002, 2004a,b).

The asymptotic distribution of the estimator is described in the next theorem. See Section 5 for a proof.

Theorem 3.1. *Suppose the error is ordinary smooth of order β and Conditions (A1)–(A5) hold. Assume $r_X \in \mathcal{C}_l(\mathbb{R}) \cap \mathcal{D}_3(\mathbb{R} \setminus \{\tau\})$ with $l \geq 0$, $\int |u|^3 |K''(u)| du < \infty$, and let $k_2 = 0$ if $l = 0$ and 1 otherwise. Then, for $h = o(n^{-1/(2\beta+2k_2+5)})$, we have*

$$\sqrt{\frac{n}{h}} \left[\frac{\tau - \hat{\tau}}{\sqrt{R(K''_Z)}} - \frac{h^{k_2+2} D_\tau}{dK''(0) \sqrt{R(K''_Z)}} \right] \xrightarrow{L} N\left(0; \frac{B_\tau}{d^2\{K''(0)\}^2}\right), \tag{3.1}$$

where $D_\tau = [(-1)^{k_2+1}/(k_2 + 1)!][r_X^{(k_2+1)}(\tau^+) + (-1)^{k_2+1} r_X^{(k_2+1)}(\tau^-)] \int_{-\infty}^0 u^{k_2+1} K''(u) du$ and $B_\tau = [f_Y(\tau^+) + f_Y(\tau^-)]/2$.

The indirect definition of the estimator makes it quite hard to derive the asymptotic bias and variance of $\hat{\tau}$, and thus the asymptotic mean squared error (AMSE) of $\hat{\tau}$. An approximation of the mean squared error (ApMSE) can be obtained from the first two moments of the asymptotic distribution. Although these are not necessarily the asymptotic moments of the estimator, they can be used as a first approximation to assess some asymptotic properties of the estimator.

Corollary 3.1. *Under the conditions of Theorem 3.1, we have,*

$$ApMSE[\hat{\tau}] = \frac{h^{2k_2+4} D_\tau^2}{d^2\{K''(0)\}^2} + \frac{hR(K''_Z)B_\tau}{nd^2\{K''(0)\}^2}. \tag{3.2}$$

When $r_X \in \mathcal{C}_l(\mathbb{R})$, with $l \geq 1$, we obtain the same asymptotic expression whatever the value of l . If $D_\tau = 0$, one has to go one or several steps further in the Taylor expansions used in the proofs, until finding a non-zero leading term.

The above results show that the larger the discontinuity, the easier the estimation. This is easy to understand intuitively, as a large discontinuity is more likely to produce large derivatives of \hat{f}_X , and thus easily detectable maxima of the diagnostic function \hat{J} .

From Theorem 3.1, we deduce that

$$\tau - \hat{\tau} = O_{\mathbb{P}}\left(\sqrt{\frac{hR(K_Z'')}{n}}\right) + O(h^{k_2+2}). \quad (3.3)$$

Under the conditions of the theorem, we know that $hR(K_Z'')$ is of order $h^{1-2\beta}$, and thus the rates of convergence depend on the sign of $1 - 2\beta$. If $0 \leq \beta \leq 1/2$, minimization of (3.3) with respect to h leads to choosing h as small as possible. Under the conditions of the theorem, we can take $h \sim n^{-(2\beta+1)^{-1}+\eta}$ with $\eta > 0$ which provides a rate of convergence slightly slower than $n^{-1/(2\beta+1)}$, more precisely, $\tau - \hat{\tau} = O_{\mathbb{P}}(n^{-(2\beta+1)^{-1}+\epsilon})$ with $\epsilon > 0$. For $\beta > 1/2$, we see that the optimal bandwidth is the balancing bandwidth (i.e., the bandwidth which makes the two terms of (3.3) of the same order). If $D_{\tau} \neq 0$, this bandwidth satisfies $h \sim n^{-(2\beta+2k_2+3)^{-1}}$. From (3.3), we then conclude that $\tau - \hat{\tau} = O_{\mathbb{P}}(n^{-(k_2+2)/(2\beta+2k_2+3)})$.

Remark 1. It is easy to see that the estimator can be applied for non-symmetric error densities as well, but in this case, K_Z is not necessarily symmetric. Although the rates of convergence do not differ from the symmetric case, the asymptotic limiting distribution has to be modified to read $\sqrt{n/h}[(\tau - \hat{\tau})/\sqrt{G_{Z,h}} - h^{k_2+2}D_{\tau}/(dK''(0)\sqrt{G_{Z,h}})] \xrightarrow{L} N(0; d^{-2}\{K''(0)\}^{-2})$, where $G_{Z,h} = f_Y(\tau^+) \int_{-\infty}^0 \{K_Z''(x)\}^2 dx + f_Y(\tau^-) \int_0^{\infty} \{K_Z''(x)\}^2 dx$. The ApmSE expression and Lemmas 5.4 and 5.6 of Section 5 have to be modified in a similar way.

Remark 2. Similar results for the supersmooth error case, of order β , can be found in the long version of this paper, where we show that $\hat{\tau} - \tau = O_{\mathbb{P}}((\ln n)^{-(k_2+2)/\beta})$. Although this slow rate seems rather discouraging, simulations carried out with a Gaussian error indicate that the estimator works well in that case too, even for samples of size $n = 50$ or 100 .

4. Simulations

In this section, we illustrate the finite sample performance of the procedure on a few examples, in the case of a Laplace error. In practice, our estimator can be calculated as $\hat{\tau} = \operatorname{argmax}_x \hat{J}(x)$, (respectively $\hat{\tau} = \operatorname{argmin}_x \hat{J}(x)$), for a left (respectively right) boundary point τ . In the case of two boundary points τ_1 and τ_2 , we take the estimators $\hat{\tau}_1 = \operatorname{argmax}_x \hat{J}(x)$ and $\hat{\tau}_2 = \operatorname{argmin}_x \hat{J}(x)$. For estimating a discontinuity point, one has no information on the sign of the discontinuity and has to stick to (2.2).

The typical shape of the diagnostic function \hat{J} is illustrated in Figure 4.1 for increasing bandwidths (this for a sample of size $n = 100$ from Density #3 below, contaminated by a Laplace error), with the actual endpoints (here -3

and 3) indicated by vertical lines. We see that the diagnostic function is indeed maximized at points close to -3 and minimized at points close to 3 .

We select the bandwidth by estimating the asymptotic MISE optimal bandwidth for estimating the derivative of a density developed in the case of continuous and differentiable densities, i.e., for a second order kernel K , we estimate the bandwidth that minimizes $\text{AMISE}\{\hat{f}'_X(x; h)\} = R(K'_Z)/(nh^3) + h^4\mu_{K,2}^2\theta_3/4$, where $\mu_{K,2}$ denotes the second moment of the kernel K , $\theta_3 = R(f_X^{(3)})$ is estimated by the normal reference or plug-in estimation methods described in Delaigle and Gijbels (2002, 2004b) and, in the symmetric error case, $R(K'_Z)$ can be calculated by $(2\pi)^{-1} \int t^2|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2} dt$.

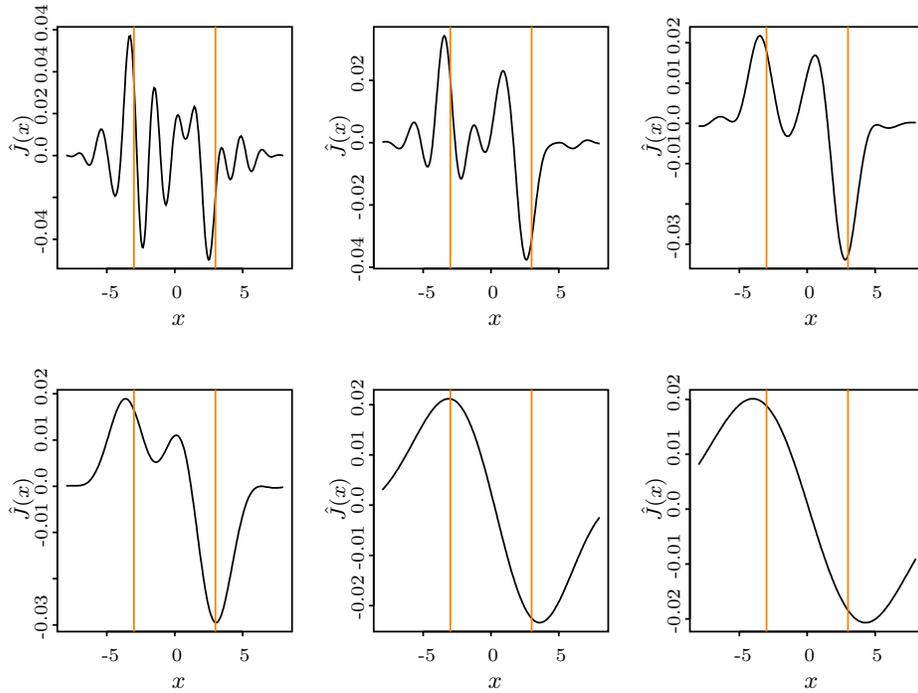


Figure 4.1. Typical shape of $\hat{J}(x)$ for a sample of size $n = 100$ from Density #3 contaminated with a Laplace error, for increasing bandwidths (from the left to the right and from the top to the bottom).

We use the second order kernel K corresponding to $\varphi_K(t) = (1 - t^2)^3 \cdot 1_{[-1,1]}(t)$, commonly used in deconvolution problems. We consider four densities with compact support: density #1 is $f_X(x) = 1/3 \cdot 1_{[0,3]}(x)$; density #2 is $f_X(x) = 3/175(-x^2 + 6x + 5) \cdot 1_{[0,5]}(x)$; density #3 is $f_X(x) = (\sin^2(x/2) + 2)/(15 - \sin 3) \cdot 1_{[-3,3]}(x)$; and density #4 is $f_X(x) = (\sin x + 1.1)/(28.5 - \cos 25) \cdot 1_{[0,25]}(x)$.

For each of the above densities, we generated 100 samples of size $n = 100$ and 250, contaminated by a Laplace error with a noise-to-signal ratio $\text{Var } Z/\text{Var } X = 10\%$. We only present the results for densities #2 and #4. The results for the other densities were similar, although slightly better. Figure 4.2 shows scatterplots of the 100 replicated estimators of the left and right endpoints of densities #2 and #4, indicated by $+$ and \diamond , respectively. The true endpoints are indicated by horizontal lines. We see that the method performs quite well, even in these rather difficult cases, and the results improve as the sample size increases. As one could have expected, the left endpoint of Density #2 is more difficult to estimate than the right endpoint, because it has a smaller jump size, yet we see that the bias decreases as the sample size increases. See Delaigle and Gijbels (2006) for more detailed results on this type of problem and other more challenging ones.

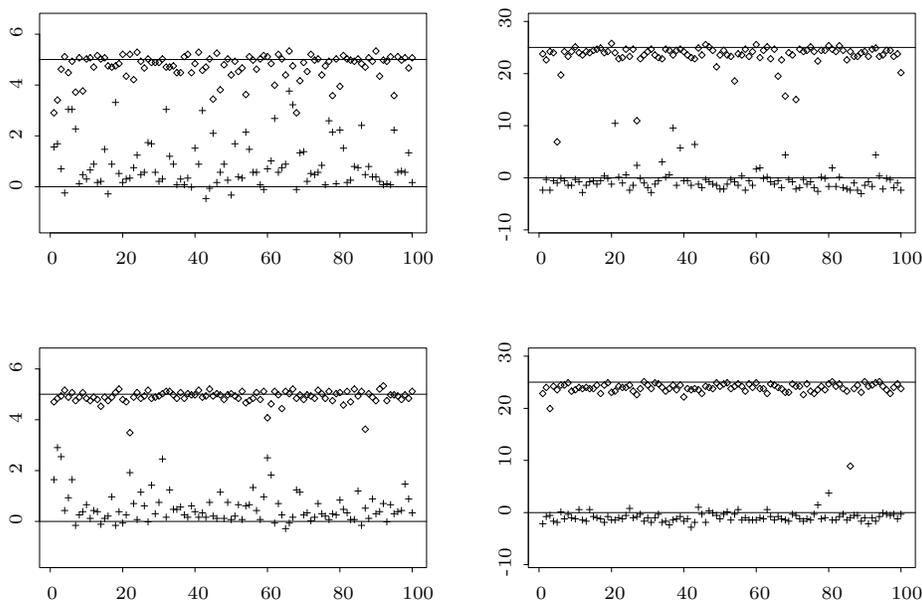


Figure 4.2. Scatterplots of 100 replicated estimators for samples of size $n = 100$ (left panels), or 250 (right panels), from density #2 (top panels) and #4 (bottom panels) contaminated by a Laplace error with a noise-to-signal ratio $\text{Var } Z/\text{Var } X = 10\%$. Estimates of the right (respectively left) endpoint are indicated by the character \diamond (respectively $+$).

Finally, we mention that we implemented our estimator on other types of error densities such as the supersmooth Gaussian density and the non-symmetric and discontinuous exponential density. The method seems to work well in these

cases too. See also Delaigle and Gijbels (2003).

5. Proofs of the Results

For $u < 0$ and $r_X \in \mathcal{C}_{\ell+1}(\mathbb{R} \setminus \{\tau\})$, the ℓ th order Taylor expansion of r_X around τ may be written as $r_X(\tau+u) = r_X(\tau) + ur'_X(\tau^-) + \dots + (u^\ell/\ell!)r_X^{(\ell)}(\tau^-) + R_\ell(\tau)$, where $R_\ell(\tau) = (u^{\ell+1}/(\ell+1!))r_X^{(\ell+1)}(\tau + \theta u)$, with $0 < \theta < 1$. We obtain a similar expansion for $u > 0$, with τ^+ instead of τ^- .

We partition the interval $[A, B]$ in n^{1+q} intervals of equal size, and define E_n as the set of endpoints of the partition, i.e., $E_n = \{z_0, \dots, z_{n^{1+q}}\}$, where $z_0 = A < z_1 < \dots < z_{n^{1+q}} = B$, and $z_{j+1} - z_j = (B - A)/n^{1+q}$ for $j = 0, \dots, n^{1+q} - 1$, where q satisfies Condition (A5).

5.1. Auxiliary results

The following sequence of lemmas lead to the proof of Theorem 3.1. Throughout, K is a k th order symmetric kernel with $k \geq 2$. The following condition will be useful.

Condition C. (C_1^m) $K \in \mathcal{C}_m(\mathbb{R})$ is such that $\lim_{|x| \rightarrow \infty} K^{(m-1)}(x) = 0$; (C_2^m) $K_Z \in \mathcal{C}_m(\mathbb{R})$; (C_3^m) $\int |u| \cdot |K_Z^{(m)}(u)|^2 du = O(h^{-2\beta})$; (C_4^m) $\int |K_Z^{(m)}(u)| du = O(h^{-\beta})$; (C_5^m) $\|K_Z^{(m)}\|_\infty = O(h^{-\beta})$; (C_6^m) $R(K_Z^{(m)}) \sim h^{-2\beta}$; (C_7^m) $\int |uK^{(m)}(u)| du < \infty$.

The next lemma is a generalization of a result of Stefanski and Carroll (1990). See Delaigle and Gijbels (2002) for a proof.

Lemma 5.1. *Let $r \geq 0$. If $K \in \mathcal{C}_r(\mathbb{R})$, we have $E[K_Z^{(r)}((x - Y)/h)] = E[K^{(r)}((x - X)/h)]$.*

Lemma 5.2. *Assume (C_1^2) and (C_2^2) , and $r_X \in \mathcal{C}_l(\mathbb{R}) \cap \mathcal{D}_3(\mathbb{R} \setminus \{\tau\})$ with $l \geq 0$. Let $k_2 = 0$ if $l = 0$ and 1 otherwise. Then, if $K'(0) = 0$ and $\int |u|^3 |K''(u)| du < \infty$,*

$$E \left[\frac{1}{nh} \sum_{i=1}^n K_Z'' \left(\frac{\tau - Y_i}{h} \right) \right] = h^{k_2+1} D_\tau + O(h^{k_2+2}). \tag{5.1}$$

Proof. From Lemma 5.1 and the condition $K'(0) = 0$, we can write

$$E \left[\frac{1}{nh} \sum_{i=1}^n K_Z'' \left(\frac{\tau - Y_i}{h} \right) \right] = \int_{-\infty}^0 K''(u) r_X(\tau - hu) du + \int_0^{+\infty} K''(u) r_X(\tau - hu) du .$$

A Taylor expansion of r_X of order 2 around τ^- (resp. τ^+) for $u > 0$ (resp. $u < 0$), combined with the fact that $\int u^j K''(u) du = 0$ for $j = 0, 1$, provides the result.

Lemma 5.3. *Let $r \geq 0$. Under (A2), (C_1^{r+1}) , (C_2^{r+1}) and (C_7^{r+1}) , we have $E[h^r \hat{J}^{(r)}(x)] = dK^{(r)}((x - \tau)/h) + O(h)$, uniformly in x .*

One requires Lemma 5.1 and Lipschitz continuity of r_X for this.

Lemma 5.4. *Let $r \geq 0$. Under (A2), (A3), (C_1^{r+1}) , (C_2^{r+1}) , (C_3^r) , (C_7^{r+1}) , and if $K_Z^{(r)}$ is symmetric, we have*

$$\text{Var} \left[h^{-1} K_Z^{(r)} \left(\frac{\tau - Y}{h} \right) \right] = \frac{f_Y(\tau^+) + f_Y(\tau^-)}{2h} \int \{K_Z^{(r)}(u)\}^2 du + O(h^{-2\beta}).$$

Proof. From Lemma 5.1, a first order Taylor expansion of f_Y around τ^+ or τ^- , and the symmetry of $K_Z^{(r)}$, we have

$$\begin{aligned} \text{Var} \left[h^{-1} K_Z^{(r)} \left(\frac{\tau - Y}{h} \right) \right] &= h^{-1} \int \{K_Z^{(r)}(u)\}^2 f_Y(\tau - hu) du + O(1) \\ &= \frac{f_Y(\tau^+) + f_Y(\tau^-)}{2h} \int \{K_Z^{(r)}(u)\}^2 du + R_2 + O(1), \end{aligned}$$

where $|R_2| \leq \sup_{x \in \mathbb{R} \setminus \{\tau\}} |f'_Y(x)| \int |u| \cdot |K_Z^{(r)}(u)|^2 du = O(h^{-2\beta})$.

The next lemma generalizes a result of Fan (1991a) to the case where the density f_X is not continuous. The proof is similar to the proof of the result in Fan (1991a). See Delaigle and Gijbels (2003).

Lemma 5.5. *Let $r \geq 0$. Under (A2), (A3), (C_1^{r+1}) , (C_2^{r+1}) , (C_3^r) , (C_4^r) , (C_5^r) , (C_6^r) , (C_7^{r+1}) , and if $K_Z^{(r)}$ is symmetric and $nh \rightarrow \infty$ as $n \rightarrow \infty$, we have*

$$\frac{h^{r-1} \hat{J}^{(r-1)}(\tau) - E[h^{r-1} \hat{J}^{(r-1)}(\tau)]}{\sqrt{\text{Var}[h^{r-1} \hat{J}^{(r-1)}(\tau)]}} \xrightarrow{L} N(0; 1). \tag{5.2}$$

Lemma 5.6. *Let $r \geq 0$. Under (A2), (C_1^{r+1}) , (C_2^{r+1}) , (C_5^{r+1}) , (C_6^{r+1}) , (C_7^{r+1}) , if $\sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y(x)| < \infty$ and if $nh \rightarrow \infty$ as $n \rightarrow \infty$, we have, for all $p \in \mathbb{N}_0$ and for n large enough,*

$$\begin{aligned} &E[h^r \hat{J}^{(r)}(x) - h^r E \hat{J}^{(r)}(x)]^{2p} \\ &\leq 2n^{-p} h^{-p-2\beta p} \cdot \left\{ \frac{2}{\pi} \int |t|^{2r+2+2\beta} |\varphi_K(t)| dt \cdot \sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y(x)| \left(\frac{2}{d_0} \right)^2 \right\}^p. \end{aligned}$$

Proof. Let T_j denote $K_Z^{(r+1)}((x - Y_j)/h)$. Then $E[h^r \hat{J}^{(r)}(x) - h^r E \hat{J}^{(r)}(x)]^{2p}$ can be written as

$$\begin{aligned} & \frac{1}{(nh)^{2p}} \sum_{i=1}^n E [T_i - E(T_1)]^{2p} \\ & + \frac{1}{(nh)^{2p}} \sum_{i=1}^n \sum_{j \neq i} \sum_{l_i=2}^{2p-2} E [T_i - E(T_1)]^{l_i} \cdot E [T_j - E(T_1)]^{2p-l_i} \\ & + \dots + \frac{1}{(nh)^{2p}} \sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{j=i_1}^{i_p} E [T_j - E(T_1)]^2, \end{aligned} \tag{5.3}$$

where we used $E[T_i - E(T_i)] = 0$. By Lemma 5.3, we have $ET_i = O(h)$ and, for all $j \geq 2$, using arguments similar to the proof of Lemma 5.5, we have $E[T_i^j] = O(h^{1-j\beta})$. We deduce that, for all $l \geq 2$, $E [T_i - E(T_1)]^l = O(h^{1-l\beta})$. Finally, we get $E[h^r \hat{J}^{(r)}(x) - h^r E \hat{J}^{(r)}(x)]^{2p} = O(n^{-p}h^{-p-2\beta p})$, since $nh \rightarrow \infty$ as $n \rightarrow \infty$. From the above calculations, we also see that we can write

$$\begin{aligned} & E[h^r \hat{J}^{(r)}(x) - h^r E \hat{J}^{(r)}(x)]^{2p} \\ & = \frac{1}{n^p h^{2p}} \left\{ E \left[K_Z^{(r+1)} \left(\frac{x - Y_1}{h} \right) \right]^2 \right\}^p \cdot (1 + o(1)) \\ & \leq \frac{1}{n^p h^{2p}} \left\{ 2h \sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y(x)| \cdot R(K_Z^{(r+1)}) \right\}^p \cdot (1 + o(1)) \\ & \leq n^{-p} h^{-p-2\beta p} \cdot \left\{ 2c_\beta(K) \sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y(x)| \cdot \left(\frac{2}{d_0}\right)^2 \right\}^p \cdot (1 + o(1)), \end{aligned}$$

where $c_\beta(K) = \pi^{-1} \int |t|^{2r+2+2\beta} |\varphi_K(t)|^2 dt$. Details of the last inequality can be found in Delaigle and Gijbels (2003).

Lemma 5.7. *Let $r \geq 0$. Assume (A2), (A5), (C_1^{r+1}) , (C_2^{r+2}) , (C_5^{r+1}) , (C_5^{r+2}) , (C_6^{r+1}) and (C_7^{r+1}) . Further assume that $\|K^{(r+1)}\|_\infty < \infty$ and $\sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y(x)| < \infty$. Then if $nh \rightarrow \infty$ as $n \rightarrow \infty$, we have for all $\epsilon > 0$,*

$$\sum_{n=1}^\infty P \left(h^{-2\delta} \sup_{x \in [A,B]} |h^r \hat{J}^{(r)}(x) - h^r E \hat{J}^{(r)}(x)| > \epsilon \right) < \infty, \tag{5.4}$$

with $\delta > 0$ as in (A5).

Proof. Recall the definition of E_n . For each x in $[A, B]$ there exists at least one point $z(x)$ in E_n such that $|x - z(x)| \leq (B - A)n^{-(1+q)}$. For all $\omega \in \Omega$, the sample space, we have

$$\sup_{x \in [A,B]} |h^r \hat{J}^{(r)}(x) - h^r E \hat{J}^{(r)}(x)| \leq S_{1,n} + S_{2,n} + S_{3,n}, \tag{5.5}$$

where $S_{1,n} \equiv \sup_{x \in [A,B]} |h^r \hat{J}^{(r)}(x) - h^r \hat{J}^{(r)}(z(x))|$, $S_{2,n} \equiv \sup_{x \in [A,B]} |h^r \hat{J}^{(r)}(z(x)) - h^r \mathbf{E} \hat{J}^{(r)}(z(x))|$ and $S_{3,n} \equiv \sup_{x \in [A,B]} |h^r \mathbf{E} \hat{J}^{(r)}(z(x)) - h^r \mathbf{E} \hat{J}^{(r)}(x)|$. To simplify notation, we do not indicate the dependence of the random variables on ω . We treat the three terms separately. For the first, note that for all $\omega \in \Omega$ and for all $x \in [A, B]$ we have (by the Mean-Value Theorem)

$$|h^r \hat{J}^{(r)}(x) - h^r \hat{J}^{(r)}(z(x))| \leq h^r |\hat{J}^{(r+1)}(\xi)| \cdot |x - z(x)| \leq h^{-2} \|K_Z^{(r+2)}\|_\infty \cdot |x - z(x)|,$$

where ξ lies between x and $z(x)$ and $\|K_Z^{(r+2)}\|_\infty \leq c_1 h^{-\beta}$, with c_1 a positive constant independent of ω and n . We conclude that

$$h^{-2\delta} S_{1,n} = h^{-2\delta} \sup_{x \in [A,B]} |h^r \hat{J}^{(r)}(x) - h^r \hat{J}^{(r)}(z(x))| \leq c_1 \cdot (B - A) h^{-2-2\delta-\beta} n^{-1-q}.$$

For handling the second term in (5.5), note that we have, for all $\omega \in \Omega$,

$$\sup_{x \in [A,B]} |h^r \hat{J}^{(r)}(z(x)) - h^r \mathbf{E} \hat{J}^{(r)}(z(x))| \leq \sup_{z \in E_n} |h^r \hat{J}^{(r)}(z) - h^r \mathbf{E} \hat{J}^{(r)}(z)|.$$

Hence for all $\epsilon > 0$ and $\ell \geq 1$, we can write

$$\sum_{n=1}^{\ell} P\left(h^{-2\delta} S_{2,n} > \epsilon\right) \leq \sum_{n=1}^{\ell} \sum_{z \in E_n} \left[P\left(h^{-2\delta} |h^r \hat{J}^{(r)}(z) - h^r \mathbf{E} \hat{J}^{(r)}(z)| > \epsilon\right) \right].$$

Now by Chebychev's Inequality and Lemma 5.6, we have that, for any $z \in \mathbb{R}$ and n large enough (say $n \geq M$),

$$\begin{aligned} P\left(h^{-2\delta} |h^r \hat{J}^{(r)}(z) - h^r \mathbf{E} \hat{J}^{(r)}(z)| > \epsilon\right) &\leq \frac{\mathbf{E} [h^r \hat{J}^{(r)}(z) - h^r \mathbf{E} \hat{J}^{(r)}(z)]^{2p}}{\frac{\epsilon^{2p}}{h^{-4\delta p}}} \\ &\leq c_2 n^{-p} h^{-p-2\beta p} h^{-4\delta p}, \end{aligned}$$

where c_2 is independent of n . Taking the limit as $\ell \rightarrow \infty$, we deduce that

$$\begin{aligned} &\sum_{n=1}^{\infty} P\left(h^{-2\delta} S_{2,n} > \epsilon\right) \\ &\leq M - 1 + c_2 \sum_{n=M}^{\infty} n^{1+q-p} h^{-4\delta p-p-2\beta p} + c_2 \sum_{n=M}^{\infty} n^{-p} h^{-4\delta p-p-2\beta p} < \infty, \end{aligned}$$

by Condition (A5). For the third term in (5.5), using Lipschitz continuity of r_X and the bound $|z(x) - x| \leq (B - A) n^{-1-q}$, we get

$$h^{-2\delta} S_{3,n} = h^{-2\delta} \sup_{x \in [A,B]} |h^r \mathbf{E} \hat{J}^{(r)}(z(x)) - h^r \mathbf{E} \hat{J}^{(r)}(x)| \leq c_3 n^{-1-q} h^{-1-2\delta},$$

with c_3 a positive constant independent of n (and of ω). Let ϵ be any positive real number. Since we have shown that, for all $\omega \in \Omega$, $h^{-2\delta}S_{1,n} + h^{-2\delta}S_{3,n} \leq c_1 (B - A)h^{-2-2\delta-\beta}n^{-1-q} + c_3n^{-1-q}h^{-1-2\delta}$ (which, under Condition (A5), tends to zero as $n \rightarrow \infty$), we have for n large enough (say $n \geq M$), $h^{-2\delta}S_{1,n} + h^{-2\delta}S_{3,n} \leq \epsilon/2$. Thus we can write

$$\sum_{n=1}^{\infty} P\left(h^{-2\delta}S_{1,n} + h^{-2\delta}S_{2,n} + h^{-2\delta}S_{3,n} > \epsilon\right) \leq (M - 1) + \sum_{n=M}^{\infty} P\left(h^{-2\delta}S_{2,n} > \frac{\epsilon}{2}\right) < \infty.$$

Lemma 5.8. *Suppose that $\hat{\tau} = \tau + O(h^{1+\eta})$ a.s., with $\eta > 0$. Assume (A2), (A5), (C_1^3) , (C_2^4) , (C_3^3) , (C_4^4) , (C_5^3) , (C_6^3) , and suppose that $\|K^{(3)}\|_{\infty} < \infty$ and $\sup_{x \in \mathbb{R} \setminus \{\tau\}} |f_Y(x)| < \infty$. Then if $nh \rightarrow \infty$ as $n \rightarrow \infty$, we have $h^2 \hat{J}''(\xi) \xrightarrow{a.s.} dK''(0)$ for any ξ between τ and $\hat{\tau}$.*

Proof. We have

$$\begin{aligned} |h^2 \hat{J}''(\xi) - dK''(0)| &\leq |h^2 \hat{J}''(\xi) - h^2 E \hat{J}''(\tau)| + |h^2 E \hat{J}''(\tau) - dK''(0)| \\ &\leq T_{1,n} + T_{2,n} + T_{3,n}, \end{aligned}$$

with $T_{1,n} = \sup_{x \in [A, B]} |h^2 \hat{J}''(x) - h^2 E \hat{J}''(x)|$, $T_{2,n} = \sup_{x \in [\tau \wedge \hat{\tau}, \tau \vee \hat{\tau}]} |h^2 E \hat{J}''(x) - h^2 E \hat{J}''(\tau)|$, and $T_{3,n} = |E h^2 \hat{J}''(\tau) - dK''(0)|$. By Lemmas 5.7 and 5.3, $T_{1,n} \xrightarrow{a.s.} 0$ and $T_{3,n} \rightarrow 0$.

Now, by Lemma 5.3 and a Taylor expansion of K'' around 0, we have for all $x \in [A, B]$, $h^2 E \hat{J}''(x) - h^2 E \hat{J}''(\tau) = d[(x - \tau)/h]K^{(3)}(\theta) + O(h)$, with θ between 0 and $(x - \tau)/h$, and where the remainder term $O(h)$ is uniform in x . In particular, since $[\tau \wedge \hat{\tau}, \tau \vee \hat{\tau}] \subset [A, B]$, we can write $T_{2,n} \leq h^{-1}|d| \cdot \|K^{(3)}\|_{\infty} \cdot |\hat{\tau} - \tau| + O(h) = O(h^\eta) + O(h)$, where the last equality holds almost surely. This proves that $T_{2,n} \xrightarrow{a.s.} 0$.

Proposition 5.2 below shows that under the conditions of the theorem, the condition $\hat{\tau} - \tau = O(h^{1+\eta})$ a.s. of Lemma 5.8 is satisfied, with η corresponding to δ in (A5). Its proof requires Proposition 5.1, the proof of which can be found in Delaigle and Gijbels (2003). See also Couallier (2000). Let $I_n = \{x \in [A, B] : |x - \tau| > h^{1+\delta}\}$, with $\delta > 0$.

Proposition 5.1. *Under (A1) to (A5), we have $\sum_{n=1}^{\infty} P(\sup_{x \in I_n} |\hat{J}(x)| \geq |\hat{J}(\tau)|) < \infty$.*

Proposition 5.2. *Under (A1) to (A5), we have $\hat{\tau} - \tau = O(h^{1+\delta})$ a.s., with $\delta > 0$ as in (A5).*

Proof. By definition of $\hat{\tau}$, we have $P(\hat{\tau} \in I_n) \leq P(\sup_{x \in I_n} |\hat{J}(x)| \geq |\hat{J}(\tau)|)$, and

$$\sum_{n=1}^{\infty} P(\sup_{x \in I_n} |\hat{J}(x)| \geq |\hat{J}(\tau)|) < \infty \implies P(A) = 1,$$

by the Borel-Cantelli lemma, if we define $A = \{w : \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|\hat{\tau}_m - \tau|/h_m^{1+\delta} \leq 1\}\}$.

We have that $\forall \omega \in A : \limsup_{n \rightarrow \infty} |\hat{\tau}_n - \tau|/h_n^{1+\delta} \leq 1$, and thus $\hat{\tau} - \tau = O(h^{1+\delta})$ almost surely.

5.2. Proof of Theorem 3.1.

By applying a Taylor expansion of \hat{J}' around τ , we can write $0 = \hat{J}'(\hat{\tau}) = \hat{J}'(\tau) + (\hat{\tau} - \tau)\hat{J}''(\xi)$ where ξ lies between τ and $\hat{\tau}$. Thus we have

$$\tau - \hat{\tau} = \hat{J}'(\tau)/\hat{J}''(\xi), \quad (5.6)$$

where by Lemma 5.8, $\hat{J}''(\xi)$ is almost surely different from zero as $n \rightarrow \infty$, since $K''(0) < 0$. Under the conditions of the theorem, we have $\hat{\tau} - \tau = O(h^{1+\delta})$ a.s. (see Proposition 5.2). Hence the conditions of Lemma 5.8 are satisfied. The asymptotic law of $\tau - \hat{\tau}$ follows from the asymptotic law of $\hat{J}'(\tau)/\hat{J}''(\xi)$. From Lemma 5.5 for $r = 2$, we know that

$$\frac{h\hat{J}'(\tau) - E[h\hat{J}'(\tau)]}{\sqrt{\text{Var}(h\hat{J}'(\tau))}} \xrightarrow{L} N(0; 1),$$

where $\text{Var}[h\hat{J}'(\tau)]$ follows from Lemma 5.4. From Lemma 5.2, we know that $E[h\hat{J}'(\tau)] = h^{k_2+1}D_\tau + O(h^{k_2+2})$, and we deduce that

$$\frac{h\hat{J}'(\tau) - h^{k_2+1}D_\tau}{\sqrt{\frac{B_\tau}{nh} R(K_Z'')}} \xrightarrow{L} N(0; 1).$$

The conclusion follows from (5.6) and Lemma 5.8.

Acknowledgement

This research was supported in part by a Belgian American and Educational Foundation Post-Doctoral Fellowship. Financial support from the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy) and from a grant of the National Science Foundation, Belgium, is gratefully acknowledged. The authors are grateful to an associate editor and two referees for their very valuable comments and suggestions which led to an improvement of the paper.

References

- Barry, J. and Diggle, P. (1995). Choosing the smoothing parameter in a Fourier approach to nonparametric deconvolution of a density function. *J. Nonparametr. Statist.* **4**, 223-232.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.
- Chu, C. K. and Cheng, P. E. (1996). Estimation of jump points and jump values of a density function. *Statist. Sinica* **6**, 79-95.
- Couallier, V. (1999). Estimation non paramétrique d'une discontinuité dans une densité. *Comptes Rendus de l'Académie des Sciences, I t.* **329**, 633-636.
- Couallier, V. (2000). Inférence statistique pour des estimateurs de discontinuités dans un cadre non paramétrique. Ph.D. thesis, Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse III, France.
- Delaigle, A. (2003). Kernel estimation in deconvolution problems. Ph.D. dissertation. Institut de Statistique, Université Catholique de Louvain, Belgium.
- Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *J. Roy. Statist. Soc. Ser. B* **64**, 869-886.
- Delaigle, A. and Gijbels, I. (2003). Boundary estimation and estimation of discontinuity points in deconvolution problems. Available at <http://www.math.ucsd.edu/~adelaigl>
- Delaigle, A. and Gijbels, I. (2004a). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* **56**, 19-47.
- Delaigle, A. and Gijbels, I. (2004b). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45**, 249-267.
- Delaigle, A. and Gijbels, I. (2006). Data-driven boundary estimation in deconvolution problems. *Comput. Statist. Data Anal.* **50**, 1965-1994.
- Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53**, 97-110.
- Fan, J. (1991b). Global behaviour of deconvolution kernel estimates. *Statist. Sinica* **1**, 541-551.
- Fan, J. (1991c). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.
- Gijbels, I. and Goderniaux, A.-C. (2004). Bandwidth selection for change point estimation in nonparametric regression. *Technometrics* **46**, 76-86.
- Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51**, 231-251.
- Goderniaux, A.-C. (2001). Automatic detection of change-points in nonparametric regression. Ph.D. thesis, Institut de Statistique, Université Catholique de Louvain, Belgium.
- Hall, P. and Simar, L. (2002). Estimating a changepoint, boundary or frontier in the presence of observation error. *J. Amer. Statist. Assoc.* **97**, 523-534.
- Kneip, A. and Simar, L. (1996). A general framework for frontier estimation with panel data. *J. Productivity Anal.* **7**, 187-212.
- Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.* **65**, 139-165.
- Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737-761.
- Neumann, M. H. (1997a). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* **7**, 307-330.

- Neumann, M. H. (1997b). Optimal change-point estimation in inverse problems. *Scand. J. Statist.* **24**, 503-521.
- Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **2**, 169-184.
- Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and jump values of a regression function. *Ann. Statist.* **21**, 1545-1566.
- Zhang, S. and Karunamuni, R. (2000). Boundary bias correction for nonparametric deconvolution. *Ann. Inst. Statist. Math.* **52**, 612-629.

Department of Mathematics, University of California at San Diego, CA 92093, U.S.A.

E-mail: delaigle@math.ucsd.edu

Department of Mathematics, Katholieke Universiteit of Leuven, W. de Croylaan 54, B-3001 Heverlee, Belgium.

E-mail: Irene.Gijbels@wis.kuleuven.be

(Received December 2004; accepted April 2005)