



Interval and Band Estimation for Curves with Jumps

Author(s): Irène Gijbels, Peter Hall, Aloïs Kneip

Source: *Journal of Applied Probability*, Vol. 41, Stochastic Methods and Their Applications, (2004), pp. 65-79

Published by: Applied Probability Trust

Stable URL: <http://www.jstor.org/stable/3215969>

Accessed: 04/04/2008 17:11

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=apt>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We enable the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# INTERVAL AND BAND ESTIMATION FOR CURVES WITH JUMPS

IRÈNE GIJBELS  
PETER HALL  
ALOÏS KNEIP

## Abstract

Jump points in curves arise when the conditions under which data are generated change suddenly, for example because of an unplanned change in a treatment. This paper suggests bootstrap methods for quantifying the error in estimates of jump points, and for constructing confidence intervals for jump points and confidence bands for the curve. These problems have the unusual feature that the sampling error of the jump-point estimator often has a highly non-normal distribution, which depends intimately on the distribution of regression errors. The methods are illustrated by a simulation study as well as by an application to data on the annual flow volume of the Nile river.

*Keywords:* Bandwidth; bootstrap; change point; confidence interval; curve estimation; discontinuity; kernel methods; nonparametric regression

2000 Mathematics Subject Classification: Primary 62G15  
Secondary 62G09

## 1. Introduction

In problems where a stochastic system is subject to sudden external influences, an otherwise continuous response to a treatment can suffer a discontinuity. In this context, we consider a nonparametric regression setting where data are recorded at discrete design points  $X_1 < \dots < X_n$ , rather than in the continuum. The position of a jump in a regression function,  $g$ , cannot be known with greater precision than the spacing between adjacent points  $X_i$  in the neighbourhood of the jump. Moreover, if the signal-to-noise ratio is low, then it will not be known which interval between successive design points contains the jump point,  $x_0$  say. In the present paper we suggest nonstandard bootstrap methods for interval estimation of  $x_0$ . We propose ways of combining such interval estimators with existing methods for constructing confidence bands for smooth functions, so as to produce a confidence band for a piecewise-continuous regression mean.

Let  $i_0$  denote the index of the largest  $X_i$  that does not exceed the jump point. We shall construct an estimator  $\hat{i}_0$  of  $i_0$ , and then an estimator  $\hat{x}_0 = \frac{1}{2}(X_{\hat{i}_0} + X_{\hat{i}_0+1})$  of  $x_0$ . In order that our method attains a high level of statistical performance we shall base it on a two-step argument. In the first step we shall use a kernel-type method to identify a small interval to which  $x_0$  belongs with high probability. Then we shall estimate  $i_0$  by least squares, from a local parametric model. Thus, the method will be 'locally maximum likelihood' in the context of Gaussian errors. However, it can be shown to work under much more general assumptions, and to produce estimators with convergence rate of order  $n^{-1}$ . A detailed study of this point estimation procedure is provided by Gijbels *et al.* (1999).

Having estimated jump points, bootstrap methods will be applied to approximate the distribution of  $\hat{i}_0$ . Interval estimators of the form  $[\hat{i}_1, \hat{i}_2]$ ,  $\hat{i}_1 < \hat{i}_2$ , of  $i_0$  will then lead to interval estimators  $[X_{\hat{i}_1}, X_{\hat{i}_2+1}]$  of  $x_0$ . It will turn out that the bootstrap procedure is consistent under very general assumptions on the error distribution. Indeed, in asymptotic terms it is asked of the design points  $X_i$  only that they become increasingly dense in a fixed interval  $\mathcal{I}$  as sample size increases. They may be either regularly spaced or randomly distributed within  $\mathcal{I}$ .

Importantly, the asymptotic distribution of  $\hat{i}_0 - i_0$  is highly non-normal, and depends intimately on the entire error distribution, not just a few of its moments. Indeed, the random variable  $\hat{i}_0$  has a discrete distribution which might be concentrated at very few integer values. Since empirical approximations to distributions of estimators are essential to constructing accurate confidence bands, the bootstrap method that we propose must be able to produce good performance in highly non-normal settings. This goal is indeed achieved by our procedure, as is evidenced by its theoretical properties and by a simulation study. The bootstrap even captures some second-order features related to the length of the interval used in the second step. Hence, our bootstrap methods are an exception to the usual ‘working rule’ for the bootstrap for distribution estimation, which argues that it produces consistent results ‘if and only if’ the statistic under investigation is asymptotically normally distributed. See, for example, Mammen (1992).

The only other work which addresses confidence procedures for change points in nonparametric regression seems to be that of Loader (1996). Loader proposes a jump-point estimator based on the maximal difference between left and right fits of local polynomials at each design point. The method requires the assumption of Gaussian errors, however, and that restriction limits applicability.

There is an extensive literature on jump-point estimation. The work most closely related to ours is that of Müller and Song (1997). They propose a two-stage change-point estimator which also attains the  $n^{-1}$  rate of convergence for regularly spaced design. A good literature survey was given by Wang (1995). See also McDonald and Owen (1986), Müller (1992), Eubank and Speckman (1993), (1994) and Raimondo (1996). The convergence rates of the majority of estimators are  $n^{-1}(\log n)^{1+\delta}$  or  $n^{-1+\delta}$ , where  $\delta > 0$ . See also Gijbels *et al.* (1999) and Müller and Stadtmüller (1999).

Section 2 briefly describes the methodology used for point estimation of jump points. Section 3 discusses our bootstrap procedure for interval estimation. The method’s numerical properties are presented in Section 4, which also addresses the case of more than one jump point and treats an application to data on the annual flow volume of the Nile river. Theoretical results are summarised in Section 5.

## 2. Point estimation

### 2.1. Model

For the sake of simplicity we shall consider the problem of estimating a function  $g$  which has only one discontinuity, at  $x_0$  say. The case of more than one jump is similar in most respects, and will be addressed in Section 4. We shall assume that a sample of  $n$  data pairs  $\mathcal{X} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  is observed, generated by the model

$$Y_i = g(X_i) + \varepsilon_i, \quad 1 \leq i \leq n. \quad (1)$$

Here, the  $X_i$  are either the order statistics of a random sample from a distribution having density  $f$ , supported on  $\mathcal{I} = [0, 1]$ , or are given by  $X_i = (i + c)/n$  for a constant  $c \in [0, 1]$ ;  $g$  is continuous on  $[0, x_0]$  and on  $[x_0, 1]$ , where  $0 \leq x_0 \leq 1$  and  $g(x_0-) \neq g(x_0+)$ ; and the errors  $\varepsilon_i$

are independent and identically distributed with zero mean and finite variance. The case where the variance of the distribution of  $\varepsilon_i$  is a function of  $X_i$ , in particular where it is different on either side of  $x_0$ , may be addressed with only minor changes to the method and results that we shall give.

## 2.2. Point estimation of $x_0$

We first discuss our method in general terms, with the aim of giving the intuition behind it rather than a mathematically rigorous account. Suppose that we have determined a small interval, say  $[z_0, z_1]$ , with  $x_0 \in [z_0, z_1]$ . If this interval is so small that  $\sup_{z_0 \leq x < x_0} |g(x) - g(x_0-)|$  and  $\sup_{x_0 < x \leq z_1} |g(x) - g(x_0+)|$  are both considerably less than the error standard deviation, then we might reasonably consider the following local approximation to (1):

$$Y_i \approx \begin{cases} \theta_1 + \varepsilon_i & \text{if } X_i \in [z_0, x_0], \\ \theta_2 + \varepsilon_i & \text{if } X_i \in [x_0, z_1], \end{cases} \quad (2)$$

where  $\theta_1 \approx g(x_0-)$  and  $\theta_2 \approx g(x_0+)$ . We are then (approximately) in the parametric setting of estimating the change point of a piecewise-constant function, which has been extensively studied in the literature (see, for example, Korostelev and Tsybakov (1993, Chapter 1)). Under (2) we may estimate  $i_0 = \max\{i : X_i \leq x_0\}$  by minimising the sum of squares,

$$S(i, \theta_1, \theta_2) = \sum_{X_j \in [z_0, X_i]} (Y_j - \theta_1)^2 + \sum_{X_j \in [X_{i+1}, z_1]} (Y_j - \theta_2)^2, \quad (3)$$

producing a vector  $(\hat{i}_0, \hat{\theta}_1, \hat{\theta}_2)$  of parameter estimators. Clearly, the least-squares estimator,  $\hat{i}_0$ , of  $i_0$  is also an (approximate) maximum-likelihood estimator in the case of Gaussian errors. We then estimate  $x_0$  by simply taking the midpoint of the estimated interval  $[X_{\hat{i}_0}, X_{\hat{i}_0+1}]$ :

$$\hat{x}_0 = \frac{1}{2}(X_{\hat{i}_0} + X_{\hat{i}_0+1}). \quad (4)$$

In the context of a valid parametric model, (2) and normal errors, we may define estimators which are slightly more efficient (in the sense of quadratic loss) than the maximum-likelihood estimator  $\hat{i}_0$ . However, in our context the important point is that (3) will still provide a reasonable estimator,  $\hat{i}_0$ , if the model is only approximately true, simply because the values taken by the function  $g$  possess different levels before and after the jump. (See also Section 2.4.)

How should we determine, as a first step, an appropriate interval  $[z_0, z_1]$ ? We shall obtain first a preliminary estimator  $\tilde{x}_0$  of  $x_0$ , and then define  $[z_0, z_1]$  as the interval concentrated around  $\tilde{x}_0$ . The preliminary estimator  $\tilde{x}_0$  is constructed by using a diagnostic to identify the approximate location of  $x_0$ . Specifically, define  $D$  by

$$D(x, h) = \frac{\partial}{\partial x} \left( \frac{\sum_{i=1}^n K\{(x - X_i)/h\} Y_i}{\sum_{i=1}^n K\{(x - X_i)/h\}} \right), \quad (5)$$

where  $K$  is a compactly supported, differentiable kernel function and  $h$  is a bandwidth; and define  $\tilde{x}_0$  to be the value of  $x$  that maximises  $|D(x, h)|$  in  $(vh, 1 - vh)$ , where  $[-v, v]$  denotes the support of  $K$ . The interval  $[z_0, z_1]$  is then defined as  $[\tilde{x}_0 - th, \tilde{x}_0 + th]$ , with  $t$  typically between  $v$  and  $2v$ .

The diagnostic,  $D$ , was studied by Gijbels *et al.* (1999); it was shown that, with probability tending to 1 as  $n \rightarrow \infty$ ,  $\tilde{x}_0 \in [x_0 - vh, x_0 + vh]$  and, hence,  $x_0 \in [\tilde{x}_0 - vh, \tilde{x}_0 + vh]$ , provided that  $h \rightarrow 0$  and  $n^{1-\delta}h \rightarrow \infty$  (where  $\delta > 0$ ). Note that  $\max_{x \in [x_0 - vh, x_0 + vh]} |D(x, h)| \rightarrow \infty$ .

In summary, our two-step procedure for estimating  $x_0$  is as follows:

**Step 1.** Locate  $\tilde{x}_0$ , the global maximum of  $|D(x, h)|$  on  $(vh, 1 - vh)$ .

**Step 2.** Put  $[z_0, z_1] = [\tilde{x}_0 - th, \tilde{x}_0 + th]$ , and determine the least-squares estimate  $\hat{i}_0$  of  $i_0$  by minimising (3). Then use (4) to determine the final estimator  $\hat{x}_0$  of  $x_0$ .

The diagnostic at (5) might be replaced by any other diagnostic which allows identification of small intervals which contain  $x_0$  with high probability. Likewise, our basic results remain unchanged if a least-squares approximation by higher-order local polynomials is used instead of (2) and (3).

### 2.3. Properties of $\hat{i}_0$ and $\hat{x}_0$

Under general conditions, the asymptotic distribution of the estimator  $\hat{i}_0$  is identical in parametric and nonparametric cases. The distribution may be defined as follows. Put  $a_1 = g(x_0-)$  and  $a_2 = g(x_0+)$ , let  $\dots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \dots$  be independent and identically distributed as the  $\varepsilon_i$  in the model at (1), and define

$$\xi_m = \begin{cases} a_1 + |m|^{-1} \sum_{i=m+1}^0 \varepsilon_i & \text{if } m \leq -1, \\ 0 & \text{if } m = 0, \\ a_2 + m^{-1} \sum_{i=1}^m \varepsilon_i & \text{if } m \geq 1, \end{cases}$$

and  $S(m) = m(a_1 - a_2)\{2\xi_m - (a_1 + a_2)\}$ . Let the random integer  $M$  be chosen to maximise  $S(m)$ . Then,

$$\hat{i}_0 - i_0 \rightarrow M \tag{6}$$

in distribution as  $n \rightarrow \infty$ . In the case of stochastic design (that is, where the  $X_i$  are random variables, rather than regularly spaced points), (6) should be interpreted conditionally on the design points. However, the distribution of  $\hat{i}_0 - i_0$  is asymptotically independent of the design.

Since  $X_{i+1} - X_i = O(n^{-1})$ , or  $O_P(n^{-1})$  for random design, then (6) and the definition of  $\hat{x}_0$  imply that  $\hat{x}_0 - x_0 = O_P(n^{-1})$ . In Section 3 we shall make use of the fact that, for all integers  $m_1$  and  $m_2$  with  $m_1 > m_2$ , the asymptotic distribution at (6) allows us to approximate the discrete probabilities

$$P(x_0 \in [X_{\hat{i}_0 - m_1}, X_{\hat{i}_0 - m_2 + 1}]) = P(\hat{i}_0 - m_1 \leq i_0 \leq \hat{i}_0 - m_2),$$

where, in the case of random design, the probabilities should be interpreted conditionally on the realised design points. For regularly spaced design, nothing more can be said about the asymptotic distribution of  $n(\hat{x}_0 - x_0)$ , since then no information is available about where  $x_0$  might lie between design points. The same conclusion applies if design points have a stochastic origin but are conditioned upon at the outset (as distinct from being conditioned upon in the context of the probabilities above).

However, for random design an *unconditional* asymptotic distribution can be identified. If the design density  $f$  is continuous and nonzero at  $x_0$ , then for large  $n$  the spacings

$$Z'_i = \begin{cases} X_{i_0+i+1} - X_{i_0+i} & \text{for } i \leq -1, \\ x_0 - X_{i_0} & \text{for } i = 0, \\ X_{i_0+1} - x_0 & \text{for } i = 1, \\ X_{i_0+i} - X_{i_0+i-1} & \text{for } i \geq 2, \end{cases}$$

are approximately independent and identically distributed as exponential random variables with mean  $\{nf(x_0)\}^{-1}$ . Arguing thus, we see that

$$nf(x_0)(\hat{x}_0 - x_0) \rightarrow T = \begin{cases} -\sum_{i=M-1}^0 Z_i - \frac{1}{2}Z_M & \text{if } M \leq -1, \\ \frac{1}{2}(Z_1 - Z_0) & \text{if } M = 0, \\ \sum_{i=1}^M Z_i + \frac{1}{2}Z_{M+1} & \text{if } M \geq 1, \end{cases} \quad (7)$$

where the convergence is in distribution and the  $Z_i$  are independent and exponentially distributed with unit mean, independent also of  $M$ . We may interpret  $Z_i$  as the weak limit of  $nf(x_0)Z'_i$  as  $n \rightarrow \infty$ .

#### 2.4. Choice of $h$ and $t$

The arguments above show that the behaviour of our method is asymptotically first order, independent of choice of the bandwidth  $h$  and of the value of  $t$ . There are, however, second-order effects that may play a significant role in practice.

**2.4.1. The identification problem.** Even for finite samples,  $|D(x, h)|$  will usually possess a large local maximum in  $[x_0 - vh, x_0 + vh]$  if a reasonable bandwidth is applied. However, for some realisations, the random error, together with a complicated structure of  $g$ , may lead to a still larger maximum of  $|D(x, h)|$  at a point far from  $x_0$ . We may then find that  $x_0 \notin [\tilde{x}_0 - th, \tilde{x}_0 + th]$ , which will result in the breakdown of our estimation procedure. The probability of such an event will not be negligible if a very small bandwidth is used or if the signal-to-noise ratio is very low. There is thus good reason to avoid particularly small bandwidths. This problem is not a specific feature of our diagnostics  $D$ , and in fact we encounter the same difficulty if we apply a diagnostic based on the maximal difference between left and right smooths, as proposed by Müller (1992) or Loader (1996).

Nevertheless, the probability of correctly specifying the interval to be used in Step 2 can be increased by using a more sophisticated identification procedure in Step 1. In practice, prior knowledge of the location of a jump will often help to determine a region,  $S$ , for which  $x_0 \in S \subset [0, 1]$ . In Step 1 one will then only look for the maximum,  $\tilde{x}_0$ , of  $|D(x, h)|$  for  $x \in S$ . Another possibility consists in not simply using the maximum of  $|D(x, h)|$ , but matching the local maxima of  $|D(x, h)|$  with those of the functions  $|D(\cdot, h_i)|$  obtained for a set  $\{h_i\}$  of different bandwidths. A preliminary estimator,  $\tilde{x}_0$ , might then be obtained under the additional constraint that, for all bandwidths  $h_i$ , a large local maximum existed in the interval  $[\tilde{x}_0 - (h_i + h), \tilde{x}_0 + (h_i + h)]$ . Further details, as well as an alternative identification procedure, are provided by Gijbels *et al.* (1999).

**2.4.2. The problem of interval length.** For finite samples, the local model (2) will usually present only an approximation, since  $g$  will show some fluctuations on  $[\tilde{x}_0 - th, \tilde{x}_0 + th]$ . The finite-sample distribution of  $\hat{i}_0$  is then better approximated when replacing, in the definition of  $M$ , the random term  $\varepsilon_i$  by  $g(X_{i_0-i}) - a_1 + \varepsilon_i$  for  $m \leq -1$  and by  $g(X_{i_0+i}) - a_2 + \varepsilon_i$  for  $m \geq 1$ . A still more precise approximation is obtained if  $a_1$  and  $a_2$  are redefined to denote the average of the  $g(X_i)$  over all  $X_i \in [z_0, x_0]$  and all  $X_i \in [x_0, z_1]$  respectively. If both  $\sup_{z_0 \leq x < x_0} |g(x) - a_1|$  and  $\sup_{x_0 < x \leq z_1} |g(x) - a_2|$  are considerably smaller than  $|a_1 - a_2|$ , the distribution of  $\hat{i}_0$  will still be concentrated around  $i_0$ . Unfortunately, this is not guaranteed if (a) the interval is very

large and (b) a local constant approximation is poor. It is easy to construct examples where least-squares estimation on a large interval  $[z_0, z_1]$  leads to heavily biased estimates of  $i_0$ . There is thus interest in choosing the interval as small as possible.

Concerning the choice of  $h$ , there is thus reason to avoid very small and very large bandwidths. A reasonable choice might be based on a cross-validation procedure, as discussed by Müller (1992). We might then choose a slightly undersmoothing bandwidth, possibly combined with a more sophisticated identification procedure as explained above. Also, the bootstrap method for interval estimation of  $i_0$  and  $x_0$ , presented in Section 3, can be used to select a reasonable bandwidth. See Section 4.

Choice of  $t$  constitutes a somewhat simpler problem. With high probability,  $x_0 \in [\tilde{x}_0 - th, \tilde{x}_0 + th]$  for  $t = v$ . The least-squares step will, of course, not work if  $x_0$  is located exactly at the boundary of this interval. One thus has to choose  $t > v$ . On the other hand, following the discussion above, using a large interval does not make sense. Hence, a reasonable value of  $t$  should certainly be no larger than  $2v$ . The numerical results in Section 4 will show that  $t = 1.5v$  may be considered a reasonable compromise.

### 3. Interval estimation

#### 3.1. Interval estimation for $x_0$

Bootstrap methods may be used to estimate the distribution of  $\varepsilon_i$  and therefore that of  $M$ . In this way, an interval estimator of  $i_0$  and, hence, of  $x_0$  could be constructed. Such an approach would capture only first-order properties of  $\hat{i}_0$  and  $\hat{x}_0$ , however. It is more appropriate to use an algorithm that captures second-order features to at least some extent, by mimicking the methods used to construct  $\hat{i}_0$  and  $\hat{x}_0$ . The following bootstrap algorithm, which has three parts, achieves that end. Examples of implementation will be given in Section 4.

**Part 1.** (*Estimation of  $g$ , and computation of residuals.*) Let  $\hat{x}_0 = \frac{1}{2}(X_{\hat{i}_0} + X_{\hat{i}_0+1})$  denote the estimator introduced in Section 2.2. Using local linear regression (see e.g. Hastie and Loader (1993), Wand and Jones (1995, pp. 3–4, 114–115), Fan and Gijbels (1996)), construct an estimator,  $\hat{g}$ , of  $g$  on  $[0, \hat{x}_0]$ , and another on  $[\hat{x}_0, 1]$ . Define residuals  $\tilde{\varepsilon}_i = Y_i - \hat{g}(X_i)$  for  $1 \leq i \leq n$ ; calculate their mean,  $\bar{\varepsilon}$ ; and put  $\hat{\varepsilon}_i = \tilde{\varepsilon}_i - \bar{\varepsilon}$ . (If desired we could also rescale these quantities, so that their mean-squared error coincided with the value of an estimator of the error variance.)

**Part 2.** (*Monte Carlo simulation.*) Conditional on  $\mathcal{X} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ , let  $\varepsilon_1^*, \dots, \varepsilon_n^*$  denote a resample drawn randomly, with replacement, from the set  $\{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n\}$ , and define

$$Y_i^* = \hat{g}(X_i) + \varepsilon_i^*, \quad 1 \leq i \leq n.$$

Then  $\mathcal{X}^* = \{(X_1, Y_1^*), \dots, (X_n, Y_n^*)\}$  is the bootstrap version of  $\mathcal{X}$ .

This approach to bootstrap resampling presupposes that the distribution of the errors,  $\varepsilon_i$ , is the same for all  $X_i$ , and in particular is the same when  $X_i$  lies on either side of the jump point. If empirical evidence sheds doubt on this assumption, then the resampling algorithm should be modified in an appropriate way. For example, if the error variance assumes different values on either side of  $x_0$  then resampling on a given side should be only from residuals computed there, which should be centred (and rescaled, if desired) independently of those on the other side. If the variance of  $\varepsilon_i$ , conditional on  $X_i$ , is a function of  $X_i$  then it may be modelled and estimated, by  $\hat{\sigma}(X_i)^2$ , say. Then, resampling would be from the set of standardised residuals  $\tilde{\varepsilon}_i/\hat{\sigma}(X_i)$ , after centring, and would take  $\varepsilon_i^*$  equal to the resampled quantity multiplied by  $\hat{\sigma}(X_i)$ .

**Part 3.** (Constructing an interval estimator of  $x_0$ .) Using the method suggested in Section 2 (employing the same values of  $h$  and  $t$ ), compute the analogues  $\hat{i}_0^*$  and  $\hat{x}_0^* = \frac{1}{2}(X_{\hat{i}_0^*} + X_{\hat{i}_0^*+1})$  of  $\hat{i}_0$  and  $\hat{x}_0$ , for the resample  $\mathcal{X}^*$  rather than the sample  $\mathcal{X}$ . For  $m = 0, -1, 1, -2, 2, -3, 3, \dots$ , determine the bootstrap probabilities

$$p_m = \mathbb{P}(\hat{i}_0^* - \hat{i}_0 = m \mid \mathcal{X}). \quad (8)$$

Given  $\alpha \in (0, 1)$ , determine integers  $m_1^\alpha$  and  $m_2^\alpha$  with  $m_1^\alpha < m_2^\alpha$ , and with minimal distance  $m_2^\alpha - m_1^\alpha$ , such that, for some  $\beta \leq \alpha$ ,

$$\sum_{m=m_1^\alpha}^{m_2^\alpha} p_m = \mathbb{P}(m_1^\alpha \leq \hat{i}_0^* - \hat{i}_0 \leq m_2^\alpha \mid \mathcal{X}) = 1 - \beta \geq 1 - \alpha.$$

Then  $[\hat{i}_0 - m_2^\alpha, \hat{i}_0 - m_1^\alpha]$  is a bootstrap confidence interval for  $i_0$  with nominal level  $\beta \leq \alpha$ . The corresponding bootstrap confidence interval for  $x_0$  is given by  $[X_{\hat{i}_0 - m_2^\alpha}, X_{\hat{i}_0 - m_1^\alpha + 1}]$ .

Note that we rely on only discrete probabilities for the different integer values of  $\hat{i}_0^* - \hat{i}_0$ . It will thus not always be possible to construct confidence intervals with nominal level exactly equal to  $\alpha$ .

Section 5 will give conditions under which the bootstrap distribution estimator is consistent: in the case of fixed design,

$$\sup_{m=0, \pm 1, \dots} |\mathbb{P}(\hat{i}_0^* - \hat{i}_0 = m \mid \mathcal{X}) - \mathbb{P}(\hat{i}_0 - i_0 = m)| \rightarrow 0 \quad (9)$$

in probability. Under the assumption that the error distribution is continuous, (9) implies that for the actual level,  $\beta \leq \alpha$ , obtained from the bootstrap procedure,

$$\mathbb{P}(x_0 \in [X_{\hat{i}_0 - m_2^\alpha}, X_{\hat{i}_0 - m_1^\alpha + 1}]) = \mathbb{P}(\hat{i}_0 - m_2^\alpha \leq i_0 \leq \hat{i}_0 - m_1^\alpha) \rightarrow 1 - \beta \geq 1 - \alpha.$$

That is, the bootstrap confidence intervals have asymptotically correct coverage. The same holds for random design. The only difference is that probabilities have to be considered conditionally on the realised design  $X_1, \dots, X_n$ , and that (9) changes to

$$\sup_{m=0, \pm 1, \dots} |\mathbb{P}(\hat{i}_0^* - \hat{i}_0 = m \mid \mathcal{X}) - \mathbb{P}(\hat{i}_0 - i_0 = m \mid X_1, \dots, X_n)| \rightarrow 0. \quad (10)$$

Thus, our bootstrap algorithm is first-order correct. To appreciate that it captures some second-order features, observe that the procedure incorporates estimates  $\hat{g}$ ,  $\hat{g}(\hat{x}_0 -)$  and  $\hat{g}(\hat{x}_0 +)$  of  $g$ ,  $a_1 = g(x_0 -)$  and  $a_2 = g(x_0 +)$  respectively, as well as the approximations  $\hat{\varepsilon}_i$  to  $\varepsilon_i$  for  $i = 1, \dots, n$ , which are asymptotically consistent if reasonable bandwidths (for example, of size  $n^{-1/5}$ ) are used. Following the discussion of interval length in Section 2.4.2, it is thus immediately clear that second-order properties are preserved in this part of the algorithm.

Our method does not capture all second-order features of the distribution of  $\hat{x}_0 - x_0$ , however. This situation cannot be remedied by simply correcting for scale, as in the percentile- $t$  method employed in more traditional problems. One approach to capturing second-order features would be to use the iterated or double bootstrap.

If the design is stochastic, then it is possible to use the bootstrap to approximate the unconditional distribution of  $\hat{x}_0 - x_0$  (see (7)). Only a slight modification of Part 2 of our bootstrap procedure is necessary, as follows. Draw resamples  $X_1^* < \dots < X_n^*$  of the design points by relying on a nonparametric estimator  $\hat{f}$  of the design density, and set  $Y_i^* = \hat{g}(X_i^*) + \varepsilon_i^*$ .

Let  $\hat{x}_0^* = \frac{1}{2}(X_{i_0^*}^* + X_{i_0^*+1}^*)$ , and use the distribution of  $\hat{x}_0^* - \hat{x}_0$ , given the data, to approximate the unconditional distribution of  $\hat{x}_0 - x_0$ .

### 3.2. Confidence bands for functions with jumps

A large variety of methods exists in the literature for constructing pointwise confidence bands (see e.g. Wahba (1983), Silverman (1985)) or simultaneous bands (e.g. Knafl *et al.* (1985), Härdle and Bowman (1988), Hall and Titterton (1988), Härdle and Marron (1991), Eubank and Speckman (1993), Eubank and Wang (1994)) for smooth regression means known only up to bounds on derivatives. Additionally, there is a large classical literature on confidence bands in parametric regression. In the present section we suggest a technique for combining a general confidence band method with our technique for constructing confidence intervals for jump points, so as to produce confidence bands for functions with jumps.

For the sake of definiteness, we shall confine attention to methods for constructing simultaneous bands, for which  $g$  is estimated on the interval between two jumps using only data pairs  $(X_i, Y_i)$  whose  $X_i$  coordinate lies in that interval. The case of pointwise bands is similar; there, we suggest taking the pointwise coverage probability for the bands and the coverage probability of confidence intervals for jump points to have a common value such as 0.95.

Suppose, as before, that the design points are distributed increasingly densely on the interval  $\mathcal{I} = [0, 1]$ , and that there are just  $k$  jump points,  $0 < x_1 < \dots < x_k < 1$ . We suggest factorising the coverage probability  $\beta = 1 - \alpha$  into  $2k + 1$  parts, say  $\beta = \beta_1 \cdots \beta_{2k+1}$ , where (a)  $0 < \beta_i < 1$  for each  $i$ , (b)  $\beta_1, \dots, \beta_k$  will be the coverages associated with intervals for the respective jump points  $x_1, \dots, x_k$ , and (c)  $\beta_{k+1}, \dots, \beta_{2k+1}$  will be the coverages of confidence bands for  $g$  in the respective intervals  $[0, x_1], [x_1, x_2], \dots, [x_k, 1]$  between jumps. (The methods suggested in Sections 2 and 3.1 are readily generalised to the case of  $k \geq 1$  jumps, and may be used to produce estimators  $\hat{x}_j$  of, and confidence intervals of the form  $[X_{\hat{i}_j - m_{2,j}^\alpha}, X_{\hat{i}_j - m_{1,j}^\alpha + 1}]$  for, the  $x_j$ .)

Extending each confidence interval vertically, we obtain a sequence of confidence ‘pillars’,  $\{(x, y) : x \in [X_{\hat{i}_j - m_{2,j}^\alpha}, X_{\hat{i}_j - m_{1,j}^\alpha + 1}], -\infty < y < \infty\}$ . Using a standard method such as one of those noted in the first paragraph of this section, construct a  $\beta_{k+j}$ -level confidence band for  $g$  on the interval  $[\hat{x}_{j-1}, \hat{x}_j]$  for each  $j \in \{1, \dots, k+1\}$ , where we define  $\hat{x}_0 = 0$  and  $\hat{x}_{k+1} = 1$ . Splice the confidence pillars and confidence bands together, to produce a confidence band for  $g$  on  $\mathcal{I}$ . Noting the consistency of our bootstrap estimators of the distributions of  $\hat{i}_j - i_j$  (see (8) and (9)), it may be proved that our confidence band has asymptotically correct coverage.

## 4. Numerical results

### 4.1. Simulation study

The following two examples illustrate our procedures for constructing confidence intervals and bands for curves with jumps. We take  $g$  to be one of the following two functions:

$$g_1(x) = 4x^2 + \mathbf{1}(x > 0.5)$$

and

$$g_2(x) = \begin{cases} \exp\{-2(x - 0.35)\} - 1 & \text{if } x \in [0, 0.35), \\ \exp\{-2(x - 0.35)\} & \text{if } x \in [0.35, 0.65), \\ \exp\{2(x - 0.65)\} + \exp(-0.6) - 2 & \text{if } x \in [0.65, 1]. \end{cases}$$

We treat fixed, equally spaced design as well as uniform, random design. Errors are taken to be Gaussian with standard deviations  $\sigma = 0.1, 0.3$  or  $0.6$ , and the biweight kernel  $K(x) = (1 - x^2)^2$

TABLE 1: Bootstrap estimated probabilities for the function  $g_1$ .

$n$	$\sigma$	$h$	Relative frequency ( $\times 100$ ) of values of $ \hat{i}_0 - i_0 $ Bootstrap mean (and standard deviation)				
			0	$\leq 1$	$\leq 2$	$\leq 3$	$\leq 10$
50	0.1	0.1	100.0	100.0	100.0	100.0	100.0
			100.0 (0.0)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)
50	0.3	0.05	83.3	93.0	94.0	94.6	97.9
			78.8 (19.4)	88.2 (15.2)	90.8 (13.4)	92.0 (12.3)	96.9 (6.7)
50	0.3	0.1	87.5	97.2	98.1	98.6	99.7
			82.3 (15.6)	92.8 (9.3)	95.8 (6.4)	97.0 (4.8)	99.2 (1.6)
50	0.3	0.15	84.9	96.5	98.2	99.0	99.7
			79.9 (16.6)	91.6 (10.3)	95.4 (6.9)	97.0 (5.0)	99.2 (1.5)
50	0.3	0.2	80.3	92.6	96.2	97.5	99.2
			74.9 (18.7)	88.1 (12.8)	93.0 (9.1)	95.4 (6.8)	98.9 (2.5)
100	0.3	0.1	87.9	98.6	99.7	99.9	100.0
			84.7 (11.6)	95.4 (5.6)	98.0 (3.4)	99.0 (2.1)	99.8 (0.5)
100	0.6	0.1	49.2	72.5	81.5	85.7	95.5
			52.4 (18.4)	71.7 (16.7)	80.1 (14.5)	84.7 (12.7)	93.5 (8.1)

TABLE 2: 95% confidence intervals for the function  $g_1$ .

$n$	$\sigma$	Actual (and nominal) coverage frequencies of bandwidths Average length of confidence interval		
		0.05	0.10	0.15
50	0.10	100.0 (100.0)	100.0 (100.0)	100.0 (99.9)
		0.020	0.020	0.020
50	0.3	92.7 (96.1)	95.0 (96.6)	95.4 (96.6)
		0.152	0.085	0.083
100	0.3	96.6 (96.4)	96.2 (96.7)	96.4 (96.7)
		0.043	0.029	0.032
100	0.6	82.1 (95.1)	89.5 (95.2)	90.9 (95.4)
		0.331	0.173	0.162
500	0.6	94.1 (95.5)	94.6 (95.6)	95.3 (95.6)
		0.023	0.021	0.023
100*	0.3	93.7 (95.1)	94.5 (96.4)	94.0 (96.3)
		0.147	0.060	0.064

for  $|x| \leq 1$  is used throughout. Sample sizes  $n = 50, 100$  and  $500$  are considered, and the numerical work employs 1000 simulations for each set of parameter values, as well as  $B = 2000$  bootstrap replicates.

The function  $g_1$  has a single jump discontinuity of size 1 at 0.5. The procedure described above, with  $t = 1.5$ , was used to estimate the location of the jump and to construct bootstrap confidence intervals. Simulations with  $t \in [1.5, 2]$  did not lead to any important differences, the choice  $t = 1.5$  being slightly better. Based on the 1000 simulations, Table 1 shows the resulting relative frequencies of some values of  $|\hat{i}_0 - i_0|$  in different situations, together with means and standard deviations of the corresponding bootstrap probability estimates. Results on the construction of 95% confidence intervals are reported in Table 2. It provides actual and

TABLE 3: 95% confidence intervals for the function  $g_2$ .

$n$	$\sigma$	0.35			0.65		
		0.05	0.10	0.15	0.05	0.10	0.15
50	0.1	100.0 (100.0)	100.0 (99.9)	100.0 (99.9)	100.0 (100.0)	100.0 (100.0)	99.8 (99.6)
		0.020	0.020	0.021	0.020	0.020	0.021
50	0.3	91.6 (95.8)	94.7 (96.3)	94.8 (96.2)	93.1 (96.1)	94.5 (96.7)	94.4 (96.7)
		0.142	0.081	0.116	0.124	0.065	0.070
100	0.3	95.4 (96.1)	96.5 (96.5)	96.1 (96.5)	97.0 (96.2)	96.6 (96.5)	96.0 (96.4)
		0.040	0.030	0.036	0.039	0.033	0.038
100	0.6	79.8 (95.2)	88.0 (95.3)	84.0 (95.2)	81.9 (95.2)	88.1 (95.4)	87.5 (95.4)
		0.225	0.209	0.297	0.205	0.131	0.128
500	0.6	94.4 (95.3)	94.3 (95.4)	94.2 (95.3)	94.7 (95.4)	94.8 (95.4)	94.6 (95.3)
		0.028	0.020	0.031	0.023	0.020	0.025
100*	0.3	94.0 (95.9)	95.4 (96.5)	96.1 (96.3)	94.2 (96.1)	96.2 (96.4)	95.3 (96.4)
		0.062	0.035	0.052	0.044	0.033	0.038

nominal coverages, as well as the average lengths of the 1000 confidence intervals. (Recall that, due to the discrete nature of our procedure, it is usually not possible to obtain nominal coverages of exactly 95%.) All results in these two tables refer to regularly spaced design, except for the last situation considered in Table 2, marked by ‘\*’, which relies on uniform random design.

Given the small sample sizes treated, the results are quite satisfactory. Coverage frequencies are poor only in certain situations, for the very small bandwidth  $h = 0.05$  or for  $(n, \sigma) = (100, 0.6)$ . The latter represents an example with a very low signal-to-noise ratio. Table 1 demonstrates the existence of an identification problem in these situations, as discussed in Section 2.4.1. There is a non-negligible probability that  $|\hat{i}_0 - i_0| > 10$ , which for the sample sizes essentially corresponds to selecting an invalid interval for least-squares estimation. Interestingly, the bootstrap probabilities given in Table 1 reflect this fact quite well, whereas actual and nominal coverages in Table 2 differ by a corresponding amount. One reason seems to be the following. Even if  $x_0 \notin [\tilde{x}_0 - th, \tilde{x}_0 + th]$ , the estimated curve  $\hat{g}$  may show a large increase (or decrease) near  $x_0$ , and many bootstrap replicates will provide estimates  $\hat{x}_0^*$  and  $\hat{i}_0^*$  close to  $x_0$  and  $i_0$  respectively. But then the sign of  $\hat{i}_0^* - \hat{i}_0$  will be the opposite of that of  $\hat{i}_0 - i_0$ . This is confirmed by the fact that, when focusing on symmetric confidence intervals (such as when we take  $m_1^\alpha = -m_2^\alpha$  in Part 3 of the bootstrap procedure), actual coverage increases to 98.7% and nominal coverage to 95.7% in the case  $(h, n, \sigma) = (0.1, 100, 0.6)$ .

The function  $g_2$  has two jump discontinuities, both of size 1 and occurring at 0.35 and 0.65. The existence of two jumps requires a slight modification of the identification procedure in Step 1 of our estimation method. The simplest way would be to take the two largest maxima of  $|D(x, h)|$ . Though this method will work asymptotically, it would multiply the identification problem discussed above. A more sophisticated identification procedure, as indicated in Section 2.4.1, might be used instead. However, for simplicity we include the prior knowledge that the first and second jump occur before and after  $x = 0.5$ , and to rely on the respective maxima of  $|D(x, h)|$  in  $[0, 0.5]$  and  $[0.5, 1]$ . Results are given in Table 3, where the first and second blocks of the table (headed ‘0.35’ and ‘0.65’ respectively) refer to the first and second jumps of  $g_2$ .

Figures 1 and 2 depict a simulated dataset, together with the 95% confidence band for the functions  $g_1$  and  $g_2$  respectively. For the sake of simplicity, pointwise confidence intervals for the smooth part of the curve were constructed using the percentile bootstrap method, as

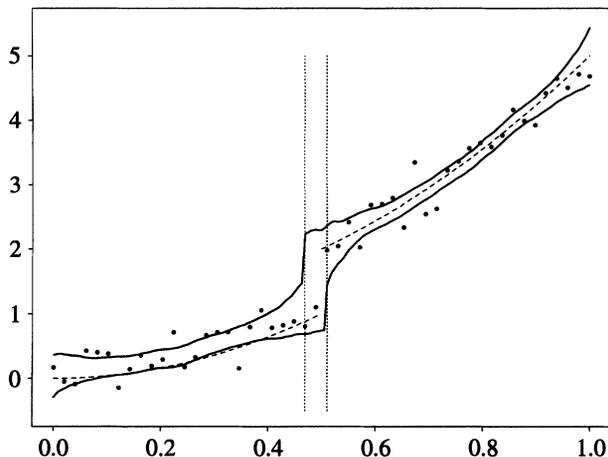


FIGURE 1: Simulated dataset for the function  $g_1$  ( $n = 50$ ,  $h = 0.1$ ,  $\sigma = 0.3$ ), together with the true regression curve (dashed line) and a 95% confidence band (solid lines). For the smooth part of the curve, pointwise confidence intervals were constructed using the percentile bootstrap method.

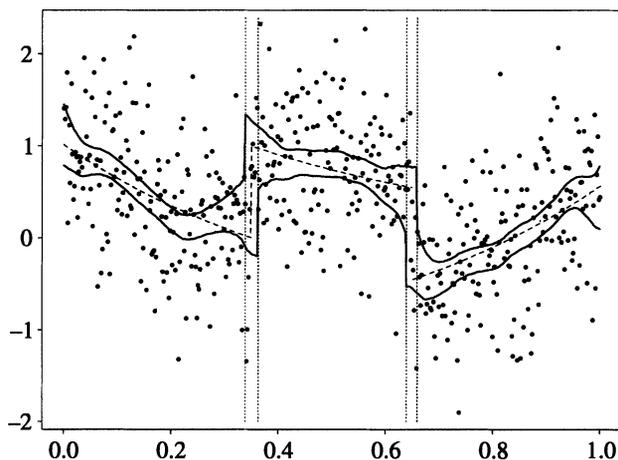


FIGURE 2: Simulated dataset for the function  $g_2$  ( $n = 500$ ,  $h = 0.1$ ,  $\sigma = 0.6$ ), together with the true regression curve (dashed line) and a 95% confidence band (solid lines). For the smooth part of the curve, pointwise confidence intervals were constructed using the percentile bootstrap method.

follows. For each fixed  $x$   $\hat{g}^*(x)$  was constructed using the bootstrap sample  $\mathcal{X}^*$  and local linear smoothing. (To avoid bias problems we undersmooth, i.e. use a bandwidth slightly smaller than the optimal one obtained by cross-validating the local linear fits of  $g$  between estimated jump-points.) Denote by  $\hat{g}_1^*(x), \dots, \hat{g}_B^*(x)$  the estimates resulting from the  $B$  bootstrap resamples, and let  $c_{\text{low}}(x)$  and  $c_{\text{up}}(x)$  denote, respectively, the lower 2.5% and upper 97.5% quantiles of the empirical distribution determined by  $\hat{g}_1^*(x), \dots, \hat{g}_B^*(x)$ . Then  $2\hat{g}(x) - c_{\text{up}}(x)$  and  $2\hat{g}(x) - c_{\text{low}}(x)$  can be taken as the quantiles defining a 95% confidence interval for  $g(x)$ .

The confidence pillars at jump points can be constructed as explained in Section 3.2. Join the confidence pillar at a jump point to an adjacent confidence band by linearly extending the band along its tangent until it passes through the pillar.

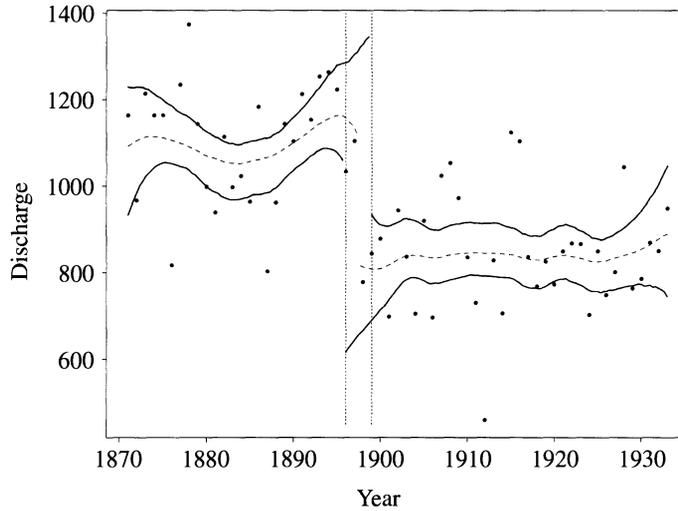


FIGURE 3: Annual flow volume of the Nile river from 1871 to 1934. Shown are the data, together with a nonparametric estimate of the regression function (dashed line) and a 95% confidence band (solid lines). For the smooth part of the curve, pointwise confidence intervals were constructed using the percentile bootstrap method.

#### 4.2. An application

A further illustration of the methods is given by an application to the well-known data on the annual flow volume of the Nile river from 1871 to 1934. In the statistical literature, these data were first presented and discussed by Cobb (1978).

Visual inspection of the data, as well as the results of Cobb (1978) or Müller (1992), suggest a change point around the year 1898. Figure 3 shows the data, as well as the resulting confidence bands obtained with our bootstrap procedure. Confidence bands are determined in the same way as for Figures 1 and 2. As in Müller (1992), a bandwidth of 10 years was chosen in our method for constructing a confidence interval for  $x_0$ . It should be noted that bandwidth was not crucial in this application, and a number of alternative bandwidths led to the same interval. Despite the rather high variability of the data, no evidence of a significant identification problem was found.

The bootstrap confidence interval for the change-point is [1896, 1899]. This corresponds to the early years of the construction of the famous barrage at the Aswan dam. The length of this interval is not unreasonable. Given the high variability of the data, we cannot exclude the possibility that the observations in 1897 are particularly large and those in 1898 particularly small, which might lead to an erroneous location of the jump.

### 5. Summary of theoretical properties

In the model (1) we can impose the following conditions.

**Condition 1.** (i) For all  $x \in \mathcal{I} = [0, 1]$ ,  $g(x) = g_1(x) + g_2(x) \mathbf{1}(x > x_0)$ , where  $0 < x_0 < 1$ , the functions  $g_1$  and  $g_2$  both have  $d \geq 1$  bounded derivatives on  $\mathcal{I}$ , and  $g(x_0) \neq 0$ .

(ii) The design points  $X_i$  are either regularly spaced on  $\mathcal{I}$ , in which case they equal  $(i + c)/n$  for a constant  $c \in [-1, 0]$ , or are the order statistics of a sample of independent and identically

distributed random variables with density  $f$ , which function has a bounded derivative and is bounded away from 0 on  $\mathcal{I}$ , and vanishes outside that interval.

(iii) The errors  $\varepsilon_i$  are independent of the  $X_i$  and are identically distributed with zero mean and  $E(|\varepsilon_i|^\gamma) < \infty$  for some  $\gamma > 2$ .

We can also impose the following conditions on the kernel  $K$  and bandwidth  $h$  used to construct the diagnostic  $D$  defined at (5).

**Condition 2.** The kernel  $K$  is supported on a finite interval  $[-v, v]$  for some  $v > 0$ ,  $K(0) > K(x) \geq 0$  for all  $x \neq 0$ , and  $K$  has two bounded derivatives.

**Condition 3.** As  $n \rightarrow \infty$ ,  $h = h(n) \rightarrow 0$  and  $\liminf n^\tau h > 0$  for some  $\tau$  with  $0 < \tau < 1 - 1/\gamma$ .

In the algorithm for estimating  $x_0$  of Section 2.2, we can make the following assumption.

**Condition 4.** In Step 2 of the algorithm, where the estimator  $\hat{x}_0$  is derived from  $\tilde{x}_0$ , we assume that  $t > 0$ .

Note that the assumption of bounded derivatives in Condition 1(i) does not imply that the values of those bounds need be known in order to apply our methods.

First we describe properties of the point estimators  $\hat{i}_0$  and  $\hat{x}_0$ .

**Theorem 1.** Assume that Conditions 1–4 hold. Then (6) holds and, if the design is stochastic, then (7) is also true.

As a prelude to describing performance of our bootstrap algorithm we impose the following conditions.

**Condition 5.** Let  $K_1$  be a nonnegative, symmetric and Hölder-continuous function supported on  $[-v, v]$  and integrating to 1, and let  $h_1$  be a bandwidth satisfying  $h_1(n) \rightarrow 0$  and  $\liminf n^{\tau_1} h_1(n) > 0$  for some  $\tau_1$  with  $0 < \tau_1 < 1 - 1/\gamma$ , where  $\gamma$  is as in Condition 1(iii). Assume that the local linear estimator,  $\hat{g}_i$ , is constructed using kernel  $K_1$  and bandwidth  $h_1$ .

**Condition 6.** Suppose that Conditions 1–4 hold and that the distribution of the errors  $\varepsilon_i$  is absolutely continuous.

**Theorem 2.** Under Conditions 5 and 6, (9) holds for fixed design, whereas (10) is true for random design.

To first order, which is the level explored in these results, the degree,  $\ell$ , of polynomials fitted locally on either side of the jump point is largely irrelevant. To appreciate why, note that, if  $g$  has at least  $\ell + 1$  bounded derivatives (outside a neighbourhood of  $x_0$ ), and if the intervals where the  $\ell$ th degree polynomials are fitted are of width  $O(h)$ , then the contribution made to the fitted polynomials by terms of degree  $j$  is of order  $O(h^j)$ . Therefore, only the local constant has a first-order effect, leading to the conclusion that fitting polynomials of degrees  $\ell \geq 0$  does not alter the limit distribution of  $\hat{i}_0$ .

## Appendix A. Outlines of technical arguments

### A.1. Proof of Theorem 1

Let  $S(i, \theta_1, \theta_2)$  be defined by (3) with  $[z_0, z_1] = [\tilde{x}_0 - th, \tilde{x}_0 + th]$ . For given  $i$ , write  $S(i)$  for the minimum of  $S(i, \theta_1, \theta_2)$  with respect to the parameters  $\theta_1$  and  $\theta_2$ . It follows from

results of Gijbels *et al.* (1999) that  $P(|\bar{x}_0 - x_0| > \eta h) \rightarrow 0$  for each  $\eta > 0$ . Moreover,  $\sup_{1 \leq i - i_0 \leq nh} |Y_i - a_2 - \varepsilon_i| \rightarrow 0$  and  $\sup_{1 \leq i_0 - i \leq nh} |Y_i - a_1 - \varepsilon_i| \rightarrow 0$ . When analysing the structure of  $S(i)$  it can then be shown that, for all  $\delta$  with  $0 < \delta < 1$  and each  $\eta > 0$ ,

$$P \left[ \sup_{1 \leq |i - i_0| \leq (nh)^\delta} |i - i_0|^{-1} |S(i) - S(i_0) + (i - i_0)(a_1 - a_2)\{2\xi(i) - (a_1 + a_2)\}| > \eta \right] \rightarrow 0,$$

where

$$|i - i_0|\xi(i) = \begin{cases} \sum_{i_0+1 \leq j \leq i} (a_2 + \varepsilon_j) & \text{if } i > i_0, \\ \sum_{i \leq j \leq i_0-1} (a_1 + \varepsilon_j) & \text{if } i < i_0. \end{cases}$$

More simply,  $P[\inf_{|i - i_0| \geq (nh)^\delta} \{S(i) - S(i_0)\} < \eta] \rightarrow 0$ . The result (6) follows from these properties. Since, for stochastic design, (6) is valid conditional on the design points, (7) follows from (6) and properties of spacings of order statistics.

## A.2. Proof of Theorem 2

Conditions 5 and 6 imply that, for each  $\eta > 0$ ,

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \sup_{1 \leq i \leq n: |i - i_0| > \lambda} |\hat{g}(X_i) - g(X_i)| > \eta \right\} \rightarrow 0, \quad (11)$$

$$P \left\{ \sup_{1 \leq i \leq n} |\hat{g}(X_i) - g(X_i)| \leq |g_2(x_0) - g_1(x_0)| + \eta \right\} \rightarrow 1. \quad (12)$$

Moreover,  $|\hat{g}(\hat{x}_0-) - g(x_0-)| \rightarrow 0$  and  $|\hat{g}(\hat{x}_0+) - g(x_0+)| \rightarrow 0$  in probability. We may conclude that (11) and (12) still hold when  $\hat{g}(X_i)$  and  $g(X_i)$  there are replaced by  $\hat{\varepsilon}_i$  and  $\varepsilon_i$ , or by  $\hat{\varepsilon}_i$  and  $\varepsilon_i$ , respectively. Following the argument leading to Theorem 1, we can then show that  $\sup_m |P(\hat{i}_0^* - \hat{i}_0 = m | \mathcal{X}) - P(M^* = m | \mathcal{X})| \rightarrow 0$ , where  $M^*$  denotes the conditional version of  $M$  that arises if, in the definition of  $M$ , the random variables  $\varepsilon_i$  are replaced by resamples obtained from the collection  $\{\varepsilon_i - n^{-1} \sum_{j \leq n} \varepsilon_j, 1 \leq i \leq n\}$  of realised errors. The assumed continuity of the distribution of  $\varepsilon_i$  implies that  $M^* \rightarrow M$  in distribution. Theorem 2 follows.

## Acknowledgements

The first and third authors were supported by 'Projet d'Actions de Recherche Concertées' grants 93/98-164 and 98/03-217. The first author was supported by FNRS grant 1.5.001.95 of the National Science Foundation, Belgium. This research was carried out when the second author was visiting the Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

## References

- COBB, G. W. (1978). The problem of the Nile: conditional solution to a change-point problem. *Biometrika* **65**, 243–251.
- EUBANK, R. L. AND SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287–1301.
- EUBANK, R. L. AND SPECKMAN, P. L. (1994). Nonparametric estimation of functions with jump discontinuities. In *Change-Point Problems*, eds. E. Carlstein, H.-G. Müller and D. Siegmund, Institute of Mathematical Statistics, Hayward, CA.
- EUBANK, R. L. AND WANG, S. (1994). Confidence regions in non-parametric regression. *Scand. J. Statist.* **21**, 147–157.
- FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

- GJIBELS, I., HALL, P. AND KNEIP, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51**, 231–251.
- HALL, P. AND TITTERINGTON, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27**, 228–254.
- HÄRDLE, W. AND BOWMAN, A. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83**, 102–110.
- HÄRDLE, W. AND MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–796.
- HASTIE, T. J. AND LOADER, C. (1993). Local regression: automatic kernel carpentry (with discussion). *Statist. Sci.* **8**, 120–143.
- KNAFL, G., SACKS, J. AND YLVIKAKER, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80**, 683–691.
- KOROSTEL'EV, A. P. AND TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction* (Lecture Notes Statist. **82**). Springer, New York.
- LOADER, C. (1996). Change point estimation using nonparametric regression. *Ann. Statist.* **24**, 1667–1678.
- MAMMEN, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations* (Lecture Notes Statist. **77**). Springer, New York.
- MCDONALD, J. A. AND OWEN, A. B. (1986). Smoothing with split linear fits. *Technometrics* **28**, 195–208.
- MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 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 STADTMÜLLER, U. (1999). Discontinuous versus smooth regression. *Ann. Statist.* **27**, 299–337.
- RAIMONDO, M. (1996). *Modèles en Rupture, Situations non Ergodique et Utilisation de Méthode d'Ondelette*. Doctoral Thesis, Université de Paris VII.
- SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.
- SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. R. Statist. Soc. B* **47**, 1–52.
- WAHBA, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *J. R. Statist. Soc. B* **45**, 133–150.
- WAND, M. P. AND JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385–397.