INTERVAL AND BAND ESTIMATION FOR CURVES WITH JUMPS

Irène Gijbels[1] **Peter Hall**[1,2] **Aloïs Kneip**[1]

SUMMARY. 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. If the size of the jump in the treatment effect exceeds the range of the error distribution then the design interval within which the jump point lies may be determined concisely. In other cases, however, estimators of the jump point are subject to sampling error. This paper suggests bootstrap methods for quantifying that error, and for constructing a confidence interval for the jump point and a confidence band for the curve. These problems have several unusual features. First, confidence intervals and bands, with asymptotically correct coverage, are possible here only if design points are randomly distributed and are not conditioned upon. (Conditioning is arguably the most common approach in problems of regression.) Secondly, the sampling error of the jump-point estimator has a highly non-Normal distribution, which depends intimately on the distribution of regression errors.

KEY WORDS. Bandwidth, bootstrap, change point, confidence interval, curve estimation, discontinuity, kernel methods, nonparametric regression.

SHORT TITLE. Interval estimation of jump points.

[1] Institut de Statistique, Université Catholique de Louvain, 20 Voie du Roman Pays, B–1348 Louvain-la-Neuve, Belgium.

[2] Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia.

1. INTRODUCTION

In problems where a stochastic system is subject to sudden external influences, such as medical data on an individual who receives a drug injection or other stimulus at an unknown time, an otherwise continuous response to a treatment can suffer a discontinuity. Typically the data are recorded only at discrete design points $X_1 < \ldots < X_n$, rather than in the continuum, so that the position of a jump cannot be known with greater precision than the spacing between adjacent X_i's in the neighbourhood of the jump. Depending on the level of noise relative to the size of the jump, it may not be possible to precisely determine the two design points on either side of the jump point, x_0. In the present paper we consider interval estimation of x_0, and propose ways of combining such estimators with existing methods for constructing confidence bands for smooth functions, so as to produce a 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 describe distributional properties of an optimal estimator $\hat{\imath}_0$ of i_0, and of the jump point estimator $\widehat{x}_0 = \frac{1}{2}(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1})$. Bootstrap methods will be suggested for estimating both distributions. We shall show that if the design points may be regarded as stochastic then an approximate confidence interval for the position of the jump may be constructed from the bootstrap approximation to the distribution of $\hat{\imath}_0$. These ideas will lead us to suggest bootstrap confidence bands for the mean treatment effect.

Our techniques will apply equally to parametric and nonparametric settings. Indeed, in the problem of estimating a jump point, and in the case of increasingly dense design, a parametric context where a polynomial is fitted on either side of the jump is first-order-equivalent (in terms of the distributions of $\hat{\imath}_0$ and $\widehat{x}_0$) to a nonparametric one where the curve is known only up to smoothness constraints.

Our method for estimating the location of a jump is based on a kernel-type diagnostic D and local least-squares fitting. In view of the latter feature, the method

is locally maximum likelihood in the context of Gaussian errors. It is appropriate to a much larger class of error distributions than just the Gaussian, however, and achieves the optimal convergence rate under only moment and continuity conditions. We shall develop theory in the case where the design points X_i become increasingly dense in a fixed interval $\mathcal{I}$, in fact where the X_i's are either regularly spaced or randomly distributed within $\mathcal{I}$. In this context our interval estimator of the jump point is of width $O_p(n^{-1})$. Alternative settings, where spacings between adjacent design points do not shrink as sample size grows, may be treated using similar ideas. Nonparametric curve estimation methods are inappropriate there, however, since local information is not available.

A curious aspect of the problem of estimating jump points is the fact that the very possibility of conducting interval estimation depends critically on the nature of design. If the design points are regularly spaced, say $X_i = i/n$, then even if we are told the value of $i_0 = \sup\{i : i/n \le x_0\}$, we need information outside the data set if we are to describe properties of $X_{i_0} - x_0$ on a finer scale than n^{-1}. Likewise, if we are able only to estimate i_0, and if we take $\widehat{x}_0 = \frac{1}{2}(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1})$ or perhaps $X_{\hat{\imath}_0}$ as our estimator of x_0, then while we can consistently estimate the distributions of $\hat{\imath}_0$ and $\widehat{x}_0$, we are unable to approximate the distribution of $\widehat{x}_0 - x_0$ on a finer scale than n^{-1}. The same conclusion applies if the design points have a stochastic origin and are conditioned upon.

On the other hand, if the design points are stochastic and we are willing to interpret confidence statements unconditionally, then we may make use of the fact that the variables $nf(x_0)\,(x_0 - X_{i_0})$ and $nf(x_0)\,(X_{i_0+1} - x_0)$ are approximately independent and Exponentially distributed with unit mean. (Here, f denotes the design density.) This enables us to consistently estimate the distribution of $n(\widehat{x}_0 - x_0)$. In effect, we are exploiting our ignorance (or more precisely, our knowledge of the distribution of our ignorance) in order to solve the interval estimation problem.

In the case of fixed design, some progress in the direction of interval estima-

tion may be made by randomly perturbing the value of $\widehat{x}_0$ – in effect, adding an additional error of known distribution. That approach increases the variability of the interval estimator, however, and so does not produce an attractive, practical method.

It turns out that the limiting distribution of $n(\widehat{x}_0 - x_0)$ is highly non-Normal, and depends intimately on the entire error distribution, not just a few of its moments. The fact that we are able to successfully use bootstrap methods to solve the distribution estimation problem in such a case, employing resamples of the same size as the original sample, is unusual. There is a "working rule" for bootstrap methods for distribution estimation, which argues that they produce consistent results "if and only if" (roughly speaking) the statistic under investigation is asymptotically Normally distributed. See for example Mammen (1992). When the limit is non-Normal, bootstrap methods based on resamples whose size is an order of magnitude smaller than that of the sample are often necessary to produce accurate distribution estimation. Surprisingly, such techniques are not necessary here, and so the present case is an exception to the rule, although our bootstrap algorithm has several nonstandard features.

We are unaware of other work on confidence procedures and interval estimators for jump points. There is, however, an extensive literature on point estimation, very well surveyed by Wang (1995). Wang's own method, like that of Raimondo (1996), is based wavelets rather than the kernel and least-squares techniques considered in the present paper. It produces an estimator which enjoys a convergence rate of $n^{-1}(\log n)^{1+\delta}$, where $\delta > 0$, for the wavelet estimator, which should be compared to the rate n^{-1} for our estimator based on kernel methods and local least-squares. (A remark in Wang (1995) that the rate $n^{-1}(\log n)^{1+\delta}$ is "best possible" should be interpreted carefully.)

Kernel techniques are considered by Eubank and Speckman (1993, 1994), who (unlike ourselves) use a bandwidth of a size that is appropriate for estimating smooth

parts of the curve. It may be shown that this produces an estimator with a convergence rate no better than $n^{-1+\delta}$, for some $\delta > 0$. Müller (1992) also suggests kernel-type estimators with convergence rates $n^{-1+\delta}$ for $\delta > 0$, and discusses weak limit theory. In cases such as these, with convergence rates an order of magnitude slower than n^{-1}, the limiting distribution of the estimator can be Normal. Gijbels, Hall and Kneip (1995) develop theory for the kernel-based diagnostic, D, that we employ in our paper. Least-squares methods are used as part of the jump-point estimation procedures of McDonald and Owen (1992) and Korostelev and Tsybakov (1993, Chapter 1). In Section 2.4 we shall discuss work on confidence bands.

Section 2 will introduce our methodology and describe its main features. Numerical properties will be discussed in Section 3, which will also address the case of more than one design point. Theoretical results will be summarized in Section 4. Outlines of technical arguments will be deferred to an Appendix.

2. METHODOLOGY

2.1. Model. For the sake of simplicity we consider the problem of estimating the jump point of 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 3. We assume that a sample of n data pairs $\mathcal{X} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ is observed, generated by the model

$$Y_i = g(X_i) + \epsilon_i\,, \quad 1 \le i \le n\,. \tag{2.1}$$

Here, the X_i's 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 $0 \le x_0 \le 1$ and $g(x_0-) \ne g(x_0+)$; and the errors ϵ_i are independent and identically distributed with zero mean and finite variance. The case where the variance of the distribution of ϵ_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 argument and results that we shall

give. In Section 2.3 we discuss modifications of our bootstrap algorithm that are appropriate in this case.

2.2. Point estimators of x_0. If parametric models $g(\cdot|\theta_1)$ and $g(\cdot|\theta_2)$ are available for g in the intervals $[0, x_0]$ and $[x_0, 1]$, respectively, then we may estimate $i_0 = \sup\{i : X_i \le x_0\}$ by minimizing the sum of squares,

$$S(i, \theta_1, \theta_2) = \sum_{j=1}^{i} \{Y_j - g(X_j|\theta_1)\}^2 + \sum_{j=i+1}^{n} \{Y_j - g(X_j|\theta_2)\}^2 ,$$

producing a vector $(\hat{\imath}_0, \hat{\theta}_1, \hat{\theta}_2)$ of parameter estimators.

If g is known only up to smoothness assumptions, for example if it has a bounded derivative on $[0, x_0]$ and on $[x_0, 1]$, then we first use a diagnostic to identify the approximate location of x_0, and then fit constant functions (or, more generally, polynomials) in the vicinity of x_0, again by least squares. Specifically, define D by either

$$D(x, h) = \left(nh^2\right)^{-1} \sum_{i=1}^{n} K'\{(x - X_i)/h\}\, Y_i \quad \text{or} \tag{2.2}$$

$$D(x, h) = \frac{\partial}{\partial x} \left(\frac{\sum_{i=1}^{n} K\{(x - X_i)/h\}\, Y_i}{\sum_{i=1}^{n} K\{(x - X_i)/h\}} \right), \tag{2.3}$$

where K is a compactly supported, differentiable kernel function and h is a bandwidth; define $\widetilde{x}_0$ to be the value of x that maximizes $|D(x, h)|$ in $(vh, 1 - vh)$, where $[-v, v]$ denotes the support of K; let $g(\cdot|\theta_1)$ and $g(\cdot|\theta_2)$ denote two general l'th degree polynomials, for some $l \ge 0$ (thus, θ_1 and θ_2 are both vectors of length $l + 1$); and minimize the local sum of squares,

$$S(i, \theta_1, \theta_2) = \sum_{j \le i}^{(1)} \{Y_j - g(X_j|\theta_1)\}^2 + \sum_{j \ge i+1}^{(2)} \{Y_j - g(X_j|\theta_2)\}^2 , \tag{2.4}$$

where $\sum_{j \le i}^{(1)}$ and $\sum_{j \ge i+1}^{(2)}$ denote summation over j such that $X_j \in [\widetilde{x}_0 - th, X_i]$ and $X_j \in [X_{i+1}, \widetilde{x}_0 + th]$, respectively, and t is a fixed positive number. (Typically we would take t to lie between v and $2v$.) This again produces a vector $(\hat{\imath}_0, \hat{\theta}_1, \hat{\theta}_2)$ of estimators, of which the first component is our estimator of i_0.

Detailed properties of D, and of estimators based solely on the diagnostic (without the least-squares fitting step), are discussed by Gijbels, Hall and Kneip (1995). There it is shown that the diagnostics defined at (2.2) and (2.3) perform almost identically in terms of first-order asymptotics, but that the one at (2.3) is often preferable in that it is more resistant against stochastic fluctuations caused by variations in the sizes of spacings $X_{i+1} - X_i$.

Under quite 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 $\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ be independent and identically distributed as the ϵ_i's in the model at (2.1), and define

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

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

$$\hat{i}_0 - i_0 \to M \tag{2.5}$$

in distribution as $n \to \infty$. Moreover, provided the errors in the regression model are stochastically independent of the design variables $\{X_i\}$, $\hat{i}_0$ is asymptotically independent of the design. The latter property means that if $\mathcal{A}(n)$ and $\mathcal{B}(n)$ denote the σ–fields generated by $(X_1, \ldots, X_n)$ and $\hat{i}_0$, respectively, then

$$\sup_{A \in \mathcal{A}(n),\, B \in \mathcal{B}(n)} \bigl| P\{(X_1, \ldots, X_n) \in A,\, \hat{i}_0 \in B\} - P\{(X_1, \ldots, X_n) \in A\}\, P(\hat{i}_0 \in B) \bigr| \to 0\,. \tag{2.6}$$

(Of course, this statement is vacuous if the design sequence is deterministic.) Formal regularity conditions for these results will be stated in Section 4.

If the design points are the order statistics of independent and identically distributed random variables with a distribution whose density f is continuous and nonzero at x_0, then for large n the spacings $Z_i' = X_{i_0+i+1} - X_{i_0+i}$ (for $i \le -1$), $Z_0' = x_0 - X_{i_0}$, $Z_1' = X_{i_0+1} - x_0$ and $Z_i' = X_{i_0+i} - X_{i_0+i-1}$ (for $i \ge 2$) are approximately independent and identically distributed as Exponential random variables

with mean $\{nf(x_0)\}^{-1}$. Arguing thus we see that if we define $\widehat{x}_0 = \frac{1}{2}(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1})$ then

$$n\,f(x_0)\,(\widehat{x}_0 - x_0) \to T = \begin{cases} -\sum_{i=M-1}^{0} Z_i - \frac{1}{2}Z_M & \text{if } M \le -1 \\ \frac{1}{2}(Z_1 - Z_0) & \text{if } M = 0 \\ \sum_{i=1}^{M} Z_i + \frac{1}{2}Z_{M+1} & \text{if } M \ge 1\,, \end{cases} \tag{2.7}$$

where the convergence is in distribution and the Z_i's 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 $n \to \infty$.)

2.3. Interval estimation of x_0. Bootstrap methods may be used to estimate the distribution of ϵ_i, hence that of T, and hence that of A. (See Section 2.2 for definitions of the latter two random variables.) In this way, and in the case of stochastic design and unconditional inference, an interval estimator of x_0 may be constructed which enjoys asymptotically correct coverage. Such an approach would capture only first-order properties of $\hat{\imath}_0$ and $\widehat{x}_0$, 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 $\hat{\imath}_0$ and $\widehat{x}_0$.

To this end we suggest a bootstrap algorithm based on, first, developing an empirical version of the model at (2.1), and then, resampling from it. We shall confine attention to the nonparametric case, estimate g using local linear regression, and, in the algorithm that produces $\widehat{x}_0$ from $\widetilde{x}_0$, fit a step function to g in a neighbourhood of $\widetilde{x}_0$. To appreciate that our approach does indeed capture second-order features, observe that since computation of $\widehat{x}_0$ involves fitting a step function to g in a neighbourhood of $\widetilde{x}_0$, then second-order properties of $\widehat{x}_0$ are governed in part by g'. If the local linear estimator of g employs a bandwidth that optimizes performance in an L^2 sense, then the derivative of that estimator consistently estimates g', and so second-order properties are preserved in this part of the algorithm. More generally, if we were to fit polynomials of degree l on either side of the jump point when estimating $\widehat{x}_0$ in a neighbourhood of $\widetilde{x}_0$, then in order to capture second-order features we would need to estimate g globally using an $(l+2)$'nd order estimator,

for example by fitting local polynomials of degree $l + 1$.

Our bootstrap algorithm has four parts, which we now describe.

Part 1: Estimation of g, and computation of residuals. Let $\widehat{x}_0 = \frac{1}{2}(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1})$ denote the estimator introduced in Section 2.2. Using local linear regression (see e.g. Hastie and Loader 1993, and Wand and Jones 1995 pp. 3–4, 114–115), construct an estimator $\widehat{g}$ of g on $[0, \widehat{x}_0]$, and another on $[\widehat{x}_0, 1]$. Define residuals $\tilde{\epsilon}_i = Y_i - \widehat{g}(X_i)$, for $1 \le i \le n$; calculate their mean, $\bar{\epsilon}$; and put $\hat{\epsilon}_i = \tilde{\epsilon}_i - \bar{\epsilon}$. (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: Estimation of design density. Using a second-order method such as that based on a symmetric, nonnegative kernel, compute a positive estimator $\widehat{f}$ of the design density f on $\mathcal{I}$.

Part 3: Monte Carlo simulation. Conditional on $\mathcal{X} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, let $\epsilon_1^*, \ldots, \epsilon_n^*$ denote a resample drawn randomly, with replacement, from the set $\{\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n\}$; and independently of this, draw a random sample $X_1', \ldots, X_n'$ from the distribution with density $\widehat{f}$. Write $X_1^* \le \ldots \le X_n^*$ for the order statistics of the sample of X_i''s, and define

$$Y_i^* = \widehat{g}(X_i^*) + \epsilon_i^*\,, \quad 1 \le i \le n\,.$$

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

This approach to bootstrap resampling presupposes that the distribution of the errors ϵ_i is the same for all X_i's, 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 ϵ_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 standardized residuals $\tilde\epsilon_i/\hat\sigma(X_i)$, after centring, and would take ϵ_i^* equal to the resampled quantity multiplied by $\hat\sigma(X_i^*)$.

Part 4: Constructing an interval estimator of x_0. Using the method suggested in Section 2.2 (employing the same bandwidth h in the diagnostic $D(\cdot, h)$, and the same real number t when fitting local polynomials of order $l = 0$ – i.e. step functions – in the neighbourhood of $\widetilde{x}_0$), compute the analogues $\hat\imath_0^*$ and $\widehat{x}_0^* = \frac{1}{2}(X^*_{\hat\imath_0^*} + X^*_{\hat\imath_0^*+1})$ of $\hat\imath_0$ and $\widehat{x}_0$, for the resample $\mathcal{X}^*$ rather than the sample $\mathcal{X}$. Given $\alpha \in (0,1)$, let $\widehat{u}_\alpha$ and $\widehat{v}_\alpha$, defined by

$$P\bigl(\widehat{x}_0^* - \widehat{x}_0 \le \widehat{u}_\alpha \bigm| \mathcal{X}\bigr) = \tfrac{1}{2}\alpha\,, \quad P\bigl(\widehat{x}_0^* - \widehat{x}_0 > \widehat{v}_\alpha \bigm| \mathcal{X}\bigr) = \tfrac{1}{2}\alpha$$

(and chosen so that the identities are satisfied as nearly as possible), be the bootstrap estimators of u_α and v_α, defined by

$$P(\widehat{x}_0 - x_0 \le u_\alpha) = P(\widehat{x}_0 - x_0 > v_\alpha) = \tfrac{1}{2}\alpha\,.$$

Then $(\widehat{x}_0 - \widehat{v}_\alpha, \widehat{x}_0 - \widehat{u}_\alpha)$ is an α-level percentile-method bootstrap confidence interval for x_0, based on the fact that the unconditional distribution of $\widehat{x}_0^* - \widehat{x}_0$ approximates the unconditional distribution of $\widehat{x}_0 - x_0$.

Section 4 will give conditions under which the bootstrap distribution estimator is consistent:

$$\sup_{-\infty<x<\infty} \bigl|P\bigl(\widehat{x}_0^* - \widehat{x}_0 \le x \bigm| \mathcal{X}\bigr) - P\bigl(\widehat{x}_0 - x_0 \le x\bigr)\bigr| \to 0 \tag{2.8}$$

in probability. Under the assumption that the error distribution is continuous, (2.7) and (2.8) imply that $P\{x_0 \in (\widehat{x}_0 - \widehat{v}_\alpha, \widehat{x}_0 - \widehat{u}_\alpha)\} \to 1 - \alpha$. That is, the bootstrap confidence interval has asymptotically correct coverage.

Thus, our bootstrap algorithm is first-order correct. It does not capture all second-order features of the distribution of $\widehat{x}_0 - x_0$, however. This situation cannot be remedied by simply correcting for scale, as in the percentile–t method employed

in more traditional problems. It does not appear to be possible to capture all second-order features without using the iterated or double bootstrap.

2.4. Confidence bands for functions with jumps. A large variety of methods exist in the literature for constructing pointwise confidence bands (e.g. Wahba 1983, Silverman 1985) or simultaneous bands (e.g. Knafl, Sacks and Ylvisaker 1985, Härdle and Bowman 1988, Hall and Titterington 1988, Härdle and Marron 1991, Eubank and Speckman 1993 and 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, to produce confidence bands for functions with jumps.

For the sake of definiteness we shall confine attention to methods for constructing simultaneous bands, in 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 be have a common value such as 0.95. In both the simultaneous and pointwise cases, confidence statements for jump-point confidence intervals will be interpreted unconditionally, and it is desirable that this be true also for confidence bands for smooth parts of the curve. This may mean making minor alterations to the manner of constructing those bands, for example by taking precautions to guard against variance problems (see for example Seifert and Gasser 1996) when constructing bootstrap versions of local linear estimators of curves.

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, $0 < x_1 < \ldots < x_k < 1$. We suggest factorizing the coverage probability $\beta = 1 - \alpha$ into $2k + 1$ parts, say $\beta = \beta_1 \ldots \beta_{2k+1}$, where (a) $0 < \beta_i < 1$ for each i, (b) $\beta_1, \ldots, \beta_k$ will be the

coverages associated with intervals for the respective jump points $x_1, \ldots, x_k$, and (c) $\beta_{k+1}, \ldots, \beta_{2k+1}$ will be the coverages of confidence bands for g in the respective intervals $[0, x_1]$, $[x_1, x_2], \ldots, [x_k, 1]$ between jumps. (The methods suggested in Sections 2.2 and 2.3 are readily generalized to the case of $k \geq 1$ jumps, and may be used to produce estimators $\widehat{x}_j$ of, and confidence intervals $(\widehat{x}_j - \widehat{c}_j, \widehat{x}_j + \widehat{d}_j)$ for, the x_j's. An example will be given in Section 3.)

Extending each confidence interval vertically we obtain a sequence of confidence "pillars", $\{(x, y) : x \in (\widehat{x}_j - \widehat{c}_j, \widehat{x}_j + \widehat{d}_j), -\infty < y < \infty\}$. Using a standard method such as one of those noted in the first paragraph of this section, construct a β_{k+j}-level confidence band for g on the interval $[\widehat{x}_{j-1}, \widehat{x}_j]$, for each $1 \leq j \leq k+1$, where we define $\widehat{x}_0 = 0$ and $\widehat{x}_{k+1} = 1$. Splice the confidence pillars and confidence bands together, to produce a confidence band for g on $\mathcal{I}$. Noting the asymptotic independence of $\hat{\imath}_0$ and the design points $X_1, \ldots, X_n$ (see (2.6)), and the consistency of the bootstrap estimator of the distribution of $\widehat{x}_0$ (see (2.8)), it may be proved that our confidence band has asymptotically correct coverage.

The physical appearance of the band will of course depend on the relative sizes of the β_j's in the factorization of α. Some experimentation is suggested, as illustrated in the next section.

3. NUMERICAL RESULTS

The following two examples illustrate our procedures for constructing confidence intervals and bands for curves with jumps. We used the mean functions

$$g_1(x) = 4x^2 + I(x > 0.5),$$

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

We took the design to be uniform and the errors to be Gaussian with variance $\sigma^2 = 0.10$ or 0.25, and used diagnostic (2.3) with biweight kernel $K(x) = (1 - x^2)^2$ for $|x| \leq 1$. Sample sizes $n = 100$, 250 and 500 were considered, and nominal coverages were 0.90, 0.95 and 0.99.

The function g_1 has a single jump discontinuity of size 1 at 0.5, whereas g_2 has two jump discontinuities, both of size 1 and occurring at 0.35 and 0.65. Thus, the estimation and inference tasks are more awkward in Example 2, and in fact difficulties were experienced keeping track of different local maxima of $|D|$. To overcome them a simple algorithm was introduced: all local maxima of $|D(\cdot, h_i)|$ in $(vh_0, 1-vh_0)$, for a discrete and decreasing set of bandwidths $\{h_i\}$, were monitored until one or more clearly provide a dominant improvement over the case $i = 0$. Then, the locations of these were selected as preliminary estimates of the sites of jumps. More details are provided in Gijbels, Hall and Kneip (1995).

The numerical work reported here employed 1000 simulations for each set of parameter values, and $B = 2000$ bootstrap replicates for all sample sizes except $n = 500$, where we took $B = 500$. Results are reported in Tables 1 and 2 (for Examples 1 and 2, respectively), which compare nominal and actual coverages, and list average lengths and standard deviations of lengths of the 1000 confidence intervals (CI). In the case of Table 2, 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. Coverage accuracy is good, especially given the complexity of Example 2.

Figures 1 and 2 depict a simulated dataset, together with the 95% confidence band, for the respective examples. For the sake of simplicity, pointwise confidence intervals for the smooth part of the curve were constructed using the percentile bootstrap method, as follows. For each fixed x we constructed $\hat{g}^*(x)$ using the bootstrap sample $\mathcal{X}^*$ and local linear smoothing. (To avoid bias problems we undersmoothed, i.e. used a bandwidth slightly smaller than optimal.) Denote by $\hat{g}_1^*(x), \cdots, \hat{g}_B^*(x)$ the estimates resulting from the B bootstrap resamples, and let $c_\ell(x)$ and $c_u(x)$ denote respectively the lower 2.5% and upper 97.5% quantiles of the empirical distribution determined by $\hat{g}_1^*(x), \cdots, \hat{g}_B^*(x)$. Then, $2\hat{g}(x) - c_u(x)$ and $2\hat{g}(x) - c_\ell(x)$ were taken as the quantiles defining a 95% confidence interval for $g(x)$.

The confidence pillars at jump points were constructed as explained in Section

2.3. We joined the confidence pillar at a jump point to an adjacent confidence band by linearly extending the band along its tangent until it had passed through the pillar.

4. SUMMARY OF THEORETICAL PROPERTIES

In this section we state regularity conditions under which the theoretical results in Section 2 are valid. For the sake of definiteness we consider in detail only the case where g, the regression mean, and f, the design density (in the case of stochastic design), are known only up to smoothness conditions. The parametric case will be addressed in remarks at the end of the section.

Of the model at (2.1) we assume that $g(x) = g_1(x) + g_2(x)\, I(x > x_0)$ for all $x \in \mathcal{I} = [0, 1]$, where $0 < x_0 < 1$, the functions g_1 and g_2 both have $d \geq 1$ bounded derivatives on $\mathcal{I}$, and $g(x_0) \neq 0$; that 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 they are the order statistics of a sample of independent and identically distributed random variables with density f, which function has d bounded derivatives and is bounded away from 0 on $\mathcal{I}$, and vanishes outside that interval; and that the errors ϵ_i are independent and identically distributed with zero mean and $E(|\epsilon_i|^\gamma) < \infty$ for some $\gamma > 2$. Of the kernel K and bandwidth h used to construct the diagnostic D defined at (2.2) or (2.3), we suppose that K is supported on a finite interval $[-v, v]$ for some $v > 0$, $K(0) > K(x) \geq 0$ for $x \neq 0$, and K has two bounded derivatives; and that $h = h(n) \to 0$ and $\liminf n^\tau h > 0$ for some $0 < \tau < 1 - (1/\gamma)$. In the step of the algorithm that derives the estimator $\widehat{x}_0$ from $\widetilde{x}_0$, we assume that the polynomials fitted on either side of the jump point are of degree $l \leq d - 1$, and that $t > 0$. Collectively we call these conditions (A_d).

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

Theorem 1. *Assume conditions* (A_d) *for some* $d \geq 1$. *Then* (2.5) *and* (2.6) *hold, and if the design is stochastic then* (2.7) *is also true.*

As a prelude to describing performance of our bootstrap algorithm, assume that the design is stochastic, let K_1 and K_2 be nonnegative, symmetric and Hölder-continuous functions supported on $[-v, v]$ and integrating to 1, and let h_1 and h_2 be bandwidths satisfying $h_i(n) \to 0$ for $i = 1$ and 2, $\liminf n^{\tau_1} h_1(n) > 0$ for some $0 < \tau_1 < 1 - (1/\gamma)$ (where γ is as in (A_d)), and $\liminf n^{\tau_2} h_2(n) > 0$ for some $\tau_2 \in (0, 1)$. Assume that the local linear estimator $\widehat{g}$ is constructed using kernel K_1 and bandwidth h_1; that the kernel density estimator $\widehat{f}$ employs kernel K_2 and bandwidth h_2 to estimate f on $[vh, 1 - vh]$; and that $\widehat{f}$ is taken equal to a constant on $[0, vh) \cup (1 - vh, 1]$. (The constant is uniquely determined from $\widehat{f}$ on $[vh, 1 - vh]$, noting that $\widehat{f}$ must integrate to 1.) Finally, suppose that (A_2) holds with $l = 0$ (i.e. in the definitions of $\widehat{x}_0$ and $\widehat{x}_0^*$, constant functions are fitted on either side of the jump point) and that the distribution of the errors ϵ_i is absolutely continuous. Collectively we call these conditions (B).

Theorem 2. *Under condition* (B), (2.8) *holds.*

It will be seen from the outline proofs of Theorems 1 and 2 in the Appendix that to first order, which is the level explored in those results, the manner in which f and g are estimated in the bootstrap algorithm is largely irrelevant. Indeed, we require only that $\widehat{f} \to f$ in a neighbourhood of x_0, and that $\|\widehat{g} - g\| \to 0$ in the sense of the Skorokhod metric. Therefore, our results are unchanged if the nonparametric estimators of f and g are replaced by consistent parametric ones. That is, to first order it is immaterial whether f of g are determined parametrically or nonparametrically.

Likewise, the degree l of polynomials fitted locally on either side of the jump point is immaterial, to first order. To appreciate why, observe that if f and g have at least $l + 1$ bounded derivatives (outside a neighbourhood of x_0, in the case of g), and if the intervals where the l'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 case $j = 0$ has a first-order effect, leading to the

conclusion that identical limit distributions for $\hat{\imath}_0$ and $\widehat{x}_0$ are obtained using fitted polynomials of degrees $l \geq 0$.

APPENDIX: *Outline of technical arguments*

Proof of Theorem 1. (Here and in the subsequent proof we work only with definition (2.2) of D. The version defined at (2.3) may be treated similarly.) The first step is to show that for each $\eta > 0$,

$$P(|\widetilde{x}_0 - x_0| > \eta h) \to 0\,. \tag{A.1}$$

To this end, observe that $\sup_{vh \leq x \leq 1-vh} |D(x,h) - E\{S(x,h)\}| = o_p(h^{-1})$. (This property may be established using, for example, embedding results from Shorack and Wellner (1988, pp. 60–61 and p. 495). It requires the assumptions, $E|\epsilon_i|^\gamma < \infty$ and $0 < \tau < 1 - (1/\gamma)$, in condition ($\mathrm{A}_d$).) Note too that $E\{D(x,h)\} = h^{-1} f(x_0) g_2(x_0)\, K\{(x - x_0)/h\} + o(h^{-1})$ uniformly in $vh \leq x \leq 1 - vh$, where we take $f \equiv 1$ in the case of regularly spaced design. Result (A.1) follows from these facts and the assumption that $K(0) > K(x) \geq 0$ for each $x \neq 0$.

Now consider the operation of minimizing the quantity $S(i, \theta_1, \theta_2)$ defined at (2.4). There, in the first and second series on the right-hand side, we sum over j's such that $X_j \in [\widetilde{x}_0 - th, X_i]$ and $X_j \in [X_{i+1}, \widetilde{x}_0 + th]$, respectively. Let $S_{t_1}(i, \theta_1, \theta_2)$ denote the quantity that arises if these ranges are changed to $[x_0 - t_1 h, X_i]$ and $[X_{i+1}, x_0 + t_1 h]$, respectively. It may be shown that, defining $a_1 = g(x_0-)$, $a_2 = g(x_0+)$ and

$$\xi(i) = \begin{cases} \sum_{i_0+1 \leq j \leq i} Y_j & \text{if } i \geq i_0 + 1 \\ 0 & \text{if } i = i_0 \\ \sum_{i \leq j \leq i_0 - 1} Y_j & \text{if } i \leq i_0 - 1\,, \end{cases}$$

and writing $S_{t_1}\{i, \hat{\theta}_1(i), \hat{\theta}_2(i)\}$ for the minimum of $S_{t_1}(i, \theta_1, \theta_2)$ over θ_1 and θ_2, we have

$$P\Bigg[\sup_{t_2 \leq t_1 \leq t_3} \sup_{1 \leq |i - i_0| \leq \eta_n n} |i - i_0|^{-1} \big| S_{t_1}\big\{i, \hat{\theta}_1(i), \hat{\theta}_2(i)\big\} - S_{t_1}\big\{i_0, \hat{\theta}_1(i_0), \hat{\theta}(i_0)\big\} - (a_1 - a_2)\,\{2\xi(i) - (i - i_0)\,(a_1 + a_2)\}\big| > \eta \Bigg] \to 0 \tag{A.2}$$

for each $0 < t_2 < t_3 < \infty$, each $\eta > 0$ and all sequences $\eta_n \downarrow 0$. More simply, if $\eta_n \to 0$ so slowly that $n^{1-\delta}\eta_n \to \infty$ for some $\delta > 0$, then for each $\lambda > 0$,

$$P\biggl(\sup_{t_2 \le t_1 \le t_3} \sup_{|i-i_0|>\eta_n n} \bigl[S_{t_1}\bigl\{ i, \hat\theta_1(i), \hat\theta_2(i) \bigr\} - S_{t_1}\bigl\{ i_0, \hat\theta_1(i_0), \hat\theta_2(i_0) \bigr\} \le -\lambda \biggr) \to 1\,.$$

Results (2.5) and (2.6) follow from these properties. (To obtain (2.6), note that in view of (A.2), (2.5) is valid conditional on the design points.) Result (2.7) follows from (2.5), (2.6) and properties of spacings of order statistics.

Proof of Theorem 2. Note that by Theorem 1, $\widehat{x}_0 - x_0 = O_p(n^{-1})$. Condition (B) implies that for each $\eta > 0$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\biggl\{ \sup_{x\in\mathcal{I}:|x-x_0|>n^{-1}\lambda} \bigl|\widehat{g}(x) - g(x)\bigr| > \eta \biggr\} \to 0\,,$$

$$P\biggl\{ \sup_{x\in\mathcal{I}} \bigl|\widehat{g}(x) - g(x)\bigr| \le g_2(x_0) + \eta \biggr\} \to 1\,.$$

Therefore, recalling the definition $i_0 = \sup\{i : X_{i_0} \le x_0\}$,

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\biggl(\sup_{1\le i\le n:|i-i_0|>\lambda} |\tilde\epsilon_i - \epsilon_i| > \eta \biggr) \to 0\,, \tag{A.3}$$

$$P\biggl\{ \sup_{1\le i\le n} |\tilde\epsilon_i - \epsilon_i| \le g_2(x_0) + \eta \biggr\} \to 1\,. \tag{A.4}$$

One consequence of these results is that if $E|\epsilon_i|^\gamma < \infty$ then $n^{-1}\sum_{i\le n} |\hat\epsilon_i|^\gamma \to E|\epsilon_i|^\gamma$ in probability as $n \to \infty$. This condition on boundedness of the moments of the conditional distribution of ϵ_i^*, and results (A.3) and (A.4), enable us to prove (by following the argument leading to Theorem 1) that

$$\sup_{-\infty<x<\infty} \bigl|P\bigl(\widehat{x}_0^* - \widehat{x}_0 \le x\bigl|\mathcal{X}\bigr) - P\bigl(\widehat{x}_0^\dagger - \widehat{x}_0 \le x\bigl|\mathcal{X}\bigr)\bigr| \to 0$$

in probability, where $\widehat{x}_0^\dagger$ denotes the version of $\widehat{x}_0^*$ that arises if, in the bootstrap algorithm, we resample from the collection $\{\epsilon_i - n^{-1}\sum_{j\le n}\epsilon_j,\ 1 \le i \le n\}$ rather than from $\{\hat\epsilon_1, \ldots, \hat\epsilon_n\}$. Using the assumed continuity of the distribution of ϵ_i we may prove similarly that

$$\sup_{-\infty<x<\infty} \bigl|P\bigl(\widehat{x}_0^\dagger - \widehat{x}_0 \le x\bigl|\mathcal{X}\bigr) - P\bigl(\widehat{x}_0 - x_0 \le x\bigr)\bigr| \to 0$$

in probability. Theorem 2 follows from the last two displayed formulae.

Acknowledgement. Gijbels and Kneip were supported by "Projet d'Actions de Recherche Concertées", No. 93/98–164.

REFERENCES

EUBANK, 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, California.

EUBANK, R.L. AND WANG, S. (1994). Confidence regions in non-parametric regression. *Scand. J. Statist.* **21**, 147–157.

GIJBELS, I., HALL, P., AND KNEIP, A. (1995). On the estimation of change points in smooth curves. Institut de Statistique, Université Catholique de Louvain, *Discussion Paper* # 9515.

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.

HALL, P. AND TITTERINGTON, D.M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Mult. Anal.* **27**, 228–254.

HASTIE, T.J. AND LOADER, C. (1993). Local regression: automatic kernel carpentry (with discussion). *Statist. Sci.* **8**, 120–143.

KNAFL, G., SACKS, J., AND YLVISAKER, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80**, 683–691.

KOROSTELEV, A.P. AND TSYBAKOV, A.B. (1993). *Minimax Theory of Image Reconstruction.* Springer Lecture Notes in Statistics **82**. Springer-Verlag, Berlin.

MAMMEN, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations.* 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.

RAIMONDO, M. (1996). *Modèles en Rupture, Situations non Ergodique et Utilisation de Méthode d'Ondelette.* PhD thesis, University of Paris VII.

SEIFERT, B. AND GASSER, T. (1996). Finite-sample variance of local polynomials: analysis and solutions. *J. Amer. Statist. Assoc.* **91**, 267–275.

SHORACK, G.R. AND WELLNER, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

SILVERMAN, B.W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47**, 1–52.

WAHBA, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. 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.

Table 1: *Simulation results for Example 1*

n	σ^2		nominal coverage		
			0.90	0.95	0.99
100	0.10	Coverage frequency CI length: average (stdev.)	90.180 0.024 (0.009)	94.589 0.032 (0.011)	98.798 0.073 (0.036)
100	0.25	Coverage frequency CI length: average (stdev.)	89.590 0.036 (0.028)	93.594 0.054 (0.037)	97.097 0.117 (0.059)
250	0.25	Coverage frequency CI length: average (stdev.)	88.689 0.011 (0.003)	93.994 0.015 (0.004)	98.799 0.027 (0.012)
500	0.25	Coverage frequency CI length: average (stdev.)	91.100 0.005 (0.001)	95.600 0.007 (0.001)	99.300 0.012 (0.003)

Table 2: *Simulation results for Example 2*

n	σ^2	0.35			0.65		
		0.90	0.95	0.99	0.90	0.95	0.99
100	0.10	92.200 0.023 (0.003)	97.200 0.030 (0.004)	99.700 0.051 (0.013)	89.100 0.024 (0.006)	94.200 0.032 (0.010)	98.400 0.085 (0.041)
100	0.25	92.500 0.029 (0.010)	96.900 0.040 (0.014)	99.000 0.079 (0.032)	88.700 0.038 (0.023)	92.700 0.055 (0.032)	96.900 0.112 (0.043)
250	0.25	89.400 0.011 (0.001)	95.700 0.015 (0.002)	99.700 0.045 (0.023)	89.600 0.012 (0.007)	95.500 0.017 (0.009)	99.000 0.052 (0.025)
500	0.25	91.400 0.005 (0.001)	96.900 0.008 (0.001)	99.700 0.027 (0.016)	90.490 0.005 (0.001)	95.495 0.008 (0.001)	99.099 0.026 (0.016)

Figure 1: *A simulated dataset for Example 1, together with the true regression curve (dashed line) and the* 95% *confidence band (solid line). For the smooth part of the curve pointwise confidence intervals were constructed using the percentile bootstrap method.*

Figure 2: *A simulated dataset for Example 2, together with the true regression curve (dashed line) and the* 95% *confidence band (solid line). For the smooth part of the curve pointwise confidence intervals were constructed using the percentile bootstrap method.*