T E C H N I C A L
R E P O R T

0317

INFERENCE FOR NONSMOOTH REGRESSION CURVES AND SURFACES USING KERNEL-BASED METHODS

I. GIJBELS

IAP STATISTICS
NETWORK

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Inference for nonsmooth regression curves and surfaces using kernel-based methods

I. Gijbels

Institut de Statistique, Université catholique de Louvain,
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.*

In this paper we review kernel-based methods for detecting discontinuities in an otherwise smooth regression function or surface. In case of a possible discontinuous curve the interest might be in detecting the discontinuities, their jump sizes and finally to estimate the discontinuous curve. Alternatively, one might be uniquely interested in estimating directly the discontinuous curve preserving the jumps. A brief discussion on available kernel-based methods for testing for a continuous versus a discontinuous regression function, and for detecting discontinuities in regression surfaces is also provided.

1. INTRODUCTION

When estimating a regression curve it is often assumed that this function is smooth, and standard smoothing techniques will result in an estimator that smooths away discontinuities. As a consequence these estimators are not consistent at points of discontinuity of the domain. To obtain consistent estimators over the whole domain, one needs to adapt the standard techniques. Apart from discontinuities a curve may show other irregularities, such as for example cusps, discontinuities in a higher order derivative, etc. In this (non-exhaustive) survey we mainly focus on simple jump discontinuities in the regression function itself.

Several approaches have been proposed to estimate such discontinuous curves. A first approach is to estimate first the locations of the discontinuities, as well as their jump sizes. The estimated jump locations divide the domain into several segments and within these segments the underlying regression curve is continuous. The estimation of the regression function can then be done by simply using standard smoothing techniques on each of the segments. A second approach focuses on the estimation of the regression function directly and aims not at estimating separately the jump locations nor their sizes. We will refer to these as direct estimation methods.

Standard smoothing techniques include kernel-based methods, local polynomial fitting, wavelet-based methods, splines etc. It would take us too far to discuss the adaptations to estimating discontinuous curves for all these techniques. We will therefore restrict to a survey of kernel-based methods and local polynomial fitting methods for estimating

*Financial support from the contract 'Projet d'Actions de Recherche Concertes' nr 98/03–217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

curves with possible jump discontinuities. Among the papers dealing with spline-based methods are Laurent and Utreras (1986), Girard (1990), Potier and Vercken (1994), Shiau (1988), Koo (1997) and Lee (2002). Semiparametric methods for change-point analysis have been dealt with by Speckman (1995) and Cline, Eubank and Speckman (1995). Wavelet methods for change-point detection have been studied by Mallat and Hwang (1992), Wang (1995, 1998, 1999), Raimondo (1998), Oudshoorn (1998) and Antoniadis and Gijbels (2002), among others.

The paper is organized as follows. In Section 2 we recall a standard jump detection method. Section 3 then provides a brief survey of available methods all based in some way on the standard method. In this section we also discuss some identification problem when detecting jump discontinuities, and discuss the case of a finite (unknown) number of jump points. Section 4 focuses on estimation of a discontinuous regression curve and Section 5 provides some references for tests for a continuous versus a discontinuous curve. We end by giving a few basic references for generalizations of the methods to the two-dimensional case.

2. A STANDARD JUMP DETECTION METHOD

Suppose that we have a sample of n data pairs $(X_1, Y_1), \cdots, (X_n, Y_n)$ generated from the model

$$Y_i = g(X_i) + \sigma(X_i)\varepsilon_i \,, \qquad 1 \le i \le n \,, \tag{1}$$

where ε_i are i.i.d. variables, independent of the X_i, with mean zero and variance one. The quantity $\sigma^2(\cdot)$ is the conditional variance function. Without loss of generality we assume that the design points are in the interval $[0, 1]$. The regression function g is smooth except at a finite number k of points where the function is discontinuous. For simplicity we will first restrict attention to the case of one discontinuity point (i.e. $k = 1$), located at $\tau \in]0, 1[$. The case of more than one discontinuity point will be discussed in Section 3.4.

We can write

$$g(x) = g_0(x) + \gamma \mathbf{1}_{[\tau,1]}(x) = \begin{cases} g_0(x) & \text{if} \quad x \in [0, \tau[\\ g_1(x) \equiv g_0(x) + \gamma & \text{if} \quad x \in [\tau, 1] \,, \end{cases} \tag{2}$$

where the function g_0 is a continuous function and where γ denotes the size of the jump of the regression function g at location τ. Clearly, at the point τ the left-hand limit of the function g (denoted by g_-; $g_-(x) = \lim\limits_{\substack{t \to x \\ <}} g(t)$) differs (for $\gamma \neq 0$) from the right-hand limit (denoted by g_+; $g_+(x) = \lim\limits_{\substack{t \to x \\ >}} g(t)$), whereas at all other points they coincide. Therefore, a reasonable estimator of the unknown point τ is provided by the point at which the difference between estimates of the right-hand limit and the left-hand limit of g is maximal in absolute value. An estimate $\widehat{g}_+(x)$ of $g_+(x)$ is obtained via a nonparametric kernel estimate using all observations located to the right of x. Similarly, an estimate $\widehat{g}_-(x)$ of $g_-(x)$ is based on all observations located to the left of x. An estimate of τ is then

$$\widehat{\tau} = \arg\max_x |\widehat{g}_+(x) - \widehat{g}_-(x)| \ . \tag{3}$$

For the jump size of g at the point τ we have

$$\gamma = \gamma(\tau) = g_+(\tau) - g_-(\tau) \ , \tag{4}$$

and an estimate of this quantity is

$$\widehat{\gamma} = \widehat{\gamma}(\widehat{\tau}) = \widehat{g}_+(\widehat{\tau}) - \widehat{g}_-(\widehat{\tau}) \ . \tag{5}$$

An estimate of the unknown regression function g is obtained by using a standard smoothing technique on all observations falling into the interval $[0, \widehat{\tau}[$ and obtaining as such $\widehat{g}_0$, as well as on all observations falling into the interval $[\widehat{\tau}, 1]$ for obtaining $\widehat{g}_1$. In correspondence with (2) an estimate for the regression curve g is then

$$\widehat{g}(x) = \begin{cases} \widehat{g}_0(x) & \text{if} \quad x \in [0, \widehat{\tau}[\\ \widehat{g}_1(x) & \text{if} \quad x \in [\widehat{\tau}, 1] \ . \end{cases} \tag{6}$$

This approach of estimating τ, γ and finally the regression curve g has been followed in many papers. A first difference between the various kernel-based methods available in the literature is related to the use of different smoothers for estimating g_+, g_- and the final estimate of g, such as, for example, the Nadaraya-Watson kernel estimator, the Gasser-Müller estimator, a local linear estimator or more generally a local polynomial estimator. (See for example Wand and Jones (1995) and Fan and Gijbels (1996).) A second difference is in the choice of the kernel functions. Thirdly, some procedures involve several steps in the estimation in order to achieve better rates of convergence as well as a better performance for the resulting estimators.

The minimax optimal rate of convergence for detecting τ is $O(n^{-1})$, for either regularly-spaced or stochastic design. Indeed, the 'average' spacings between design points are of order n^{-1}, and so jump discontinuities cannot be determined with a greater accuracy. See also Korostelov and Tsybakov (1993) for optimal rates mainly in parametric settings.

In the next section we briefly discuss the different methods, and indicate how to deal with the case of more than one discontinuity point.

The generalization of this approach for detecting discontinuities in the νth ($\nu \geq 0$) derivative of a regression function g is straightforward. One then looks at differences of estimates of the νth left-hand and right-hand derivative of the regression function, denoted by $\widehat{g}_+^{(\nu)}(\cdot)$ and $\widehat{g}_-^{(\nu)}(\cdot)$ respectively. The location of a discontinuity point τ of the νth derivative of the function g is then estimated by

$$\widehat{\tau}_\nu = \arg\max_x \left| \widehat{g}_+^{(\nu)}(x) - \widehat{g}_-^{(\nu)}(x) \right| \ ,$$

and an estimate of the jump size γ of the derivative function $g^{(\nu)}$ at the point τ is

$$\widehat{\gamma}_\nu = \widehat{\gamma}_\nu(\widehat{\tau}) = \widehat{g}_+^{(\nu)}(\widehat{\tau}) - \widehat{g}_-^{(\nu)}(\widehat{\tau}) \ .$$

3. DETECTION OF DISCONTINUITIES IN REGRESSION CURVES

3.1. One-step procedures

We first discuss the one-step procedures and then provide a brief survey on the more-step procedures.

Among the first papers using kernel-based smoothing techniques, there is the work by Canny (1986).

Müller (1992) considered a fixed-design regression model with homoscedastic errors and studied estimation of a single discontinuity point in the νth derivative of a regression function. The approach is as the one explained in Section 2. As a smoother he used the Gasser-Müller estimator with one-sided kernels. A one-sided kernel K_+ supported on $[-1, 0]$ is used to estimate the right-hand limit and a one-sided kernel K_- defined via $K_-(x) = K_+(-x)$, supported on $[0, 1]$, is used for estimating the left-hand limit. Furthermore it is assumed that $K_+(-1) = K_+(0) = 0$. The resulting estimator for τ can achieve the rate $O(n^{-1+\delta})$ for some $\delta > 0$. See also Eubank and Speckman (1994) for similar results obtained via a least-squares type of estimator.

Wu and Chu (1993a) consider a diagnostic function

$$J(x) = \widehat{g}_2(x) - \widehat{g}_3(x) , \tag{7}$$

where $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$ are Gasser-Müller estimators, using the same bandwidth h, and different kernel functions K_2 and K_3 respectively, with K_2 and K_3 compactly supported on $[-1, 1]$ and satisfying $K_2(x) = K_3(-x)$ and $\int_0^1 K_2 \neq \int_0^1 K_3$. Discontinuity points are then detected by searching for locations where $|J(x)|$ is large. Indeed, for points x where the regression function g is continuous both estimates $\widehat{g}_2(x)$ and $\widehat{g}_3(x)$ are consistent estimates for $g(x)$ and hence $J(x)$ will be close to zero; one can show that $E[J(x)] = O(h)$. For a point x which is a discontinuity point of g though, it can be shown that the expectation of $J(x)$ is given by

$$E\left[J(x)\right] = \gamma \left[\int_0^1 K_3(u)du - \int_0^1 K_2(u)du \right] + O\left(h + (nh)^{-1}\right) , \tag{8}$$

where γ is the jump size at the discontinuity point x. An estimate for τ in model (2) is then the point where $|J(x)|$ is maximal. To estimate the jump size γ Wu and Chu (1993a) consider the quantity

$$S(x) = \widehat{g}_4(x) - \widehat{g}_5(x) ,$$

where $\widehat{g}_4(\cdot)$ and $\widehat{g}_5(\cdot)$ are Gasser-Müller estimators, using a same bandwidth g, and different kernel functions K_4 and K_5, satisfying conditions similar to these imposed on K_2 and K_3. An expression similar to (8) can be proved for $S(x)$. Therefore, an appropriate estimate for γ is given by

$$\widehat{\gamma} = S(\widehat{\tau}) \left[\int_0^1 K_5(u)du - \int_0^1 K_4(u)du \right]^{-1} .$$

The bandwidth g is taken to be larger than the bandwidth h, such that in the neighbourhood of a discontinuity point the quantity $S(x)$ is expected to be of order $O(g)$ and hence could be of a better performance then an estimate based on $J(\widehat{\tau})$ for $\widehat{\tau}$ in a neighbourhood of the true τ.

Modifications of kernel regression estimates for points near the boundary of a support and discontinuity points are proposed in Wu and Chu (1993b). The proposed modifications built further on the methodology of Hall and Wehrly (1991) proposed for regression estimation at boundary points.

Loader (1996) considers regression model (1) with g as in (2), but restricts to equispaced fixed design points and to Gaussian errors. The estimator for τ is as in (3) where the estimates $\widehat{g}_+(\cdot)$ and $\widehat{g}_-(\cdot)$ are obtained by local polynomial fitting of degree p. Essential assumptions on the kernel functions K_+ and K_- are that $K_+(x) = K_-(-x)$ and K_- is defined on $[0, 1]$ with $K_-(1) = 0$ and $K_-(0) > 0$, and integrates to one. Loader (1996) shows that the estimate of τ achieves rate of convergence $O(n^{-1})$ and also derives the asymptotic law for $\widehat{\tau}$ in this restrictive setup of normal errors. The assumption that $K_-(0) > 0$ is essential for obtaining the $O(n^{-1})$ rate of convergence of the estimator.

Grégoire and Hamrouni (2002a) consider the more general random regression model (1) and use local linear smoothing with (3) for detecting jump points. They use a kernel $K_+(\cdot)$ that is supported on $[-1, 0]$ with $K_-(x) = K_+(-x)$ and make the basic assumption that $K_+(0) > 0$. They show that the rate of convergence $O(n^{-1})$ is obtained for all choices of the bandwidth parameter $h = h_n$ satisfying the usual conditions $h_n \to 0$ and $nh_n \to \infty$. This result is along the same lines as the result in Loader (1996) obtained under the particular setup. The use of local linear smoothing is motivated by the fact that local linear smoothers have the nice property of adapting automatically to the boundary of the support. Note that estimation of the left-hand and right-hand limits g_+ and g_- is indeed assimilated to estimation at 'boundary' points. Grégoire and Hamrouni (2002a) study the jump estimate $\widehat{\gamma}(x)$ as a process in x, and establish convergence of the process. From this they deduce the asymptotic normality of the estimate $\widehat{\tau}$ as well as of the estimate $\widehat{\gamma}(\widehat{\tau})$ of the jump size γ.

Extensions of these methods to the case of a jump discontinuity in the νth derivative of the regresion function is rather straightforward, by relying on local polynomial fitting of degree p with $p \geq \nu$.

A key issue in these estimation methods is the choice of the bandwidth parameter h. This choice is rather crucial as was pointed out by many authors. Among the proposals available in the literature there are those of Grégoire and Hamrouni (2002a) who suggest to carry out the estimation procedure for an extended grid of h-values. If the estimates show a certain stability than the bandwidth used is a reasonable one. A more involved approach would be to obtain $\widehat{\tau} = \widehat{\tau}(h)$, for a fixed value of h, and from this derive the regression estimate $\widehat{g}$ as defined in (6), using the same bandwidth. Consider then a measure of performance of the regression estimate, such as for example, an Integrated Squared Error, or an Average Squared Error, and get an estimate $I(h)$ of this measure of error. The idea is then to choose the bandwidth, among a grid of bandwidth values, for which this estimated error is minimal. See also Wu and Chu (1993c) for such a procedure. Note that such a method focuses on final estimation of the regression function g. An important question here is whether the bandwidth for estimating the location of the jump point on one hand and the estimation for the regression curve on the other hand should be taken the same, and whether an appropriate error criterion is one that focuses on the estimation of the regression function.

Horváth and Kokoszka (2002) consider regression model (1) with fixed equidistant design $x_i = i/n$, $1 \leq i \leq n$, and fit locally polynomials of degree ν to the left and to the right of points x of the domain of definition of the regression function. Consistency of the resulting estimates for the change point in the νth derivative is shown under different conditions on the kernel function. The issue of choice of the bandwidth parameters is not addressed.

The above estimation procedures are commony-based on looking at differences of estimates of left-hand and right-hand limits (see (3) and (5)) or on a diagnostic function which is the difference between two kernel type estimates of the regression function as in (7).

3.2. Two-step procedures

Other kernel-based procedures have been proposed. Among these are several two-step procedures. A first two-step procedure was introduced in Korostelev (1987) who considered estimation of jump points in a Gaussian white noise model.

Müller and Song (1997) consider a fixed design regression model and propose a two-step estimator for the case of a single jump discontinuity. Instead of (2) they consider

$$g(x) = g_0(x) + \gamma_n \mathbf{1}_{[\tau,1]}(x) ,$$

with $|\gamma_n| = a_n^{-1}$, where a_n is a positive sequence of real numbers. The 'contiguous' jump size case is characterized by $a_n \to \infty$ as $n \to \infty$ whereas the fixed jump case, described above, corresponds to $a_n =$ constant. In this contiguous jump size case Müller and Song (1997) propose a two-step estimator for τ and establish its rate of convergence as well as its limiting distribution. In step one of the procedure a pilot estimator $\widetilde{\tau}$ of τ is obtained via (3) using a bandwidth h_n, and kernels K_+ and K_- as indicated before, without requiring that $K_+(0) > 0$. Müller and Song (1997) showed that, under suitable conditions, $P(|\widetilde{\tau} - \tau| \leq h_n) \to 1$ as $n \to \infty$. In a second step of the procedure random intervals

$$[r_0, r_1] = [\widetilde{\tau} - h_n, \widetilde{\tau} + h_n] \qquad \text{and} \qquad [s_0, s_1] = [\widetilde{\tau} - 2h_n, \widetilde{\tau} + 2h_n] ,$$

are considered, where $h_n > 0$ should satisfy $a_n h_n \to 0$, $a_n^2 (nh_n)^{-1} \to 0$ and $nh_n^2 a_n^{-1} \to 0$. Note that this implies that $h_n \to 0$ and $nh_n \to \infty$. Note also that the interval $[r_0, r_1]$ contains the true unknown location point τ with probability tending to one, as shown in the theoretical property mentioned above. In this second step one calculates for each point x in $[r_0, r_1]$, a weighted difference between averages of Y-observations to the left and to the right, contained in the bigger interval $[s_0, s_1]$. The final estimate is then the point x for which this weighted difference is maximal. More precisely, the final estimation of τ

is based on the diagnostic function

$$T_n(x) = [(x - s_0)(s_1 - x)]^{1/2} \left\{ \frac{1}{\lambda_n([x, s_1])} \sum_{i: x_i \in [x, s_1]} Y_i - \frac{1}{\lambda_n([s_0, x\,))} \sum_{i: x_i \in [s_0, x\,)} Y_i \right\} ,$$

where $\lambda_n(A)$, with A a set, is the counting measure counting the number of design points $x_i = i/n$ in the set A. The final estimator of τ is defined as

$$\widehat{\tau} = \arg \max_{x \in [r_0, r_1]} |T_n(x)| .$$

Müller and Song (1997) prove that this final estimator has rate of convergence $O_P(n^{-1} a_n^2)$ in the contiguous jump case, and rate $O_P(n^{-1})$ in the fixed jump case (i.e. a_n is a constant). Practical choices of the bandwidth parameter are not discussed.

Gijbels, Hall and Kneip (1999) introduced another two-step procedure under the regression model (1). The proposed estimator has desired rate n^{-1}. In addition practical choices of the bandwidth parameters involved have been proposed, resulting into a fully data-driven estimation procedure.

The two-step procedure of Gijbels, Hall and Kneip (1999), consists of obtaining a rough estimate of τ, via maximization of an appropriate diagnostic function, and then to do a refined search by fitting a piecewise constant function via least squares in an interval surrounding the rough estimate. The first step involves a diagnostic function, which is taken to be an estimate of the 'derivative' of the unknown regression function. The motivation behind this is simple: the estimated 'continuous' regression curve will typically show a large slope at the location of the discontinuity point of g. Therefore a good pilot estimator is obtained by taking the location where an estimate of the first derivative is large. Examples of estimates of derivative curves are: the derivative of the Nadaraya-Watson or Gasser-Müller estimate, or the direct estimator of the derivative obtained via local linear (or polynomial) fitting. For simplicity of presentation we take the derivative of the Nadaraya-Watson estimator, and define the diagnostic function D by

$$D(x, h_1) = \frac{\partial}{\partial x} \left(\frac{\sum_{i=1}^{n} K\{(x - X_i)/h_1\} Y_i}{\sum_{i=1}^{n} K\{(x - X_i)/h_1\}} \right) , \tag{9}$$

where K is a kernel function, supported on $[-1, 1]$, with two Hölder continuous derivatives on $[0, 1]$, $K(0) \geq K(x) > 0$ for all $x \neq 0$ and $K''(0) > 0$. The parameter $h_1 = h_{1,n} > 0$ is a bandwidth parameter. A first rough estimator for τ is

$$\widetilde{\tau} = \arg \max_{x \in]h_1, 1-h_1[} |D(x, h_1)| .$$

In the second step one considers an interval around $\widetilde{\tau}$ that contains the unknown point τ with high probability (justified by theoretical results). Consider the interval $[\widetilde{\tau} - h_2, \widetilde{\tau} + h_2]$, with $h_2 = h_{2,n}$ bandwidth parameter (possibly different from h_1). Denote by $\{i_1, i_1, \cdots, i_2\}$ the set of integers i for which $X_i \in [\widetilde{\tau} - h_2, \widetilde{\tau} + h_2]$. Then fit a step function on the interval

$[\widetilde{\tau} - h_2, \widetilde{\tau} + h_2]$ using least-squares. The jump point is estimated to occur between design points X_{i_0} and X_{i_0+1}, where i_0 is chosen to minimize the (residual) sum of squares

$$\sum_{i=i_1}^{i_0} \left\{ Y_i - (i_0 - i_1 + 1)^{-1} \sum_{j=i_1}^{i_0} Y_j \right\}^2 + \sum_{i=i_0+1}^{i_2} \left\{ Y_i - (i_2 - i_0)^{-1} \sum_{j=i_0+1}^{i_2} Y_j \right\}^2 .$$

Denoting by $\widehat{\imath}_0$ the minimizer of this sum of squares, the final estimator of τ is given by the midpoint between $X_{\hat{\imath}_0}$ and $X_{\hat{\imath}_0+1}$, i.e.

$$\widehat{\tau} = \frac{1}{2} \left(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1} \right) .$$

Gijbels, Hall and Kneip (1999) proved that this estimator achieves the optimal n^{-1} rate of convergence. Moreover, their extensive simulation study shows that the introduction of a second step really improves the estimation: if the pilot estimation is bad, the second step permits an improvement of it.

The performance of the estimator $\widehat{\tau}$ heavily depends on the choice of the bandwidth parameters h_1 and h_2. Data-driven choices for h_1 and h_2, based on a bootstrap procedure relying on an appropriate error criterion for the estimation of τ, have been worked out by Gijbels and Goderniaux (2002a). The fully data-driven procedure has been tested extensively via simulation studies and shows a very good performance of the method. Its performance has also been compared, via simulations with other n^{-1}-rate estimation procedures, such as the one-step procedure of Grégoire and Hamrouni (2002a). See Goderniaux (2001) and Gijbels and Goderniaux (2002a,c). One of the findings is that the data-driven two-step method tends to be less variable than the one-step procedure of Grégoire and Hamrouni (2002a), although both achieve, theoretically, the optimal rate of convergence.

Several remarks can be made here. Recalling the definition of a derivative one can see that the diagnostic functions in (7) and (9) are similar in spirit. The diagnostic function (7) is focusing on the numerator in the definition of the derivative. Note also that in the case of fixed equally spaced design, one could, instead of relying on (9), simply look at the derivative of the numerator of the Nadaraya-Watson estimator, as this is a consistent estimator of $g' \cdot f$, with f the design density. The estimation can as such be based on $(nh_1)^{-1} \sum_{i=1}^{n} K'((x - x_i)/h_1) Y_i$. Bias and variance of the corresponding estimator $\widehat{\tau} = \arg\max_x (nh_1)^{-1} \sum_{i=1}^{n} K'((x - x_i)/h_1) Y_i$ of τ are derived in Koch and Pope (1997).

Further, the two-step procedures by Müller and Song (1997) and Gijbels, Hall and Kneip (1999) are somewhat similar in spirit, but a major difference is in the error criterion used in the second step, and in the fact that the latter is a fully data-driven estimation procedure.

Another kernel-based algorithm for detecting jumps in the υth derivative of a regression function is provided in Qiu and Yandell (1998a). For detecting a jump in the regression function g itself, a local linear fit is carried out in a neighbourhood of design points, and the estimates of the derivatives of the regression function at these design points are used to detect discontinuity points. This basically relies on ideas similar to the justification of (9). For determining whether a point is a jump point or not, Qiu and Yandell (1998a) rely on asymptotic theory for local linear (polynomial) smoothers.

3.3. Identification problem

A possible problem that might appear in all of the above methods is the distinction between a discontinuity point of the regression function g and a location where g shows a high derivative in absolute value. For convenience, we explain this problem in the context of the procedure of Gijbels, Hall and Kneip (1999). A simple search for the maximum in the diagnostic function $|D(x, h_1)|$, defined in (9), might give some problems, as was already pointed out in Gijbels, Hall and Kneip (1999). As an example, consider the cosinus function

$$g(x) = \cos\{8\pi(0.5 - x)\} - 2\cos\{8\pi(0.5 - x)\}\mathbf{1}(x > 0.5) \;, \tag{10}$$

which has a discontinuity point at 0.5 of size -2, and shows many steep decreasing and increasing parts. As a consequence the diagnostic function exhibits many local maxima and the largest among these may correspond to a steep gradient instead of to the jump discontinuity. This is illustrated in Figure 1 (a) where we have represented, for a simulated dataset of size $n = 100$ from (1) with $\sigma^2 = \text{constant} = 0.1$, $\varepsilon \sim \mathrm{N}(0; 1)$ and equidistant design, the diagnostic function for $h_1 = h_{1,0} = 0.1$. We can see that the diagnostic function shows 8 local maxima and the largest of them does not correspond to the location of the jump discontinuity, 0.5. So the problem that is highlighted with this example is one of identifying that local maximum of the diagnostic function that corresponds to the jump discontinuity. This identification problem can be dealt with in a fairely easy way, via a three-step algorithm proposed by Gijbels, Hall and Kneip (1999). The algorithm also includes an automatic choice of the bandwidth h_1. Let $h_{1,i} = h_0 r^i$, for $i = 0, 1, 2, \cdots$, be a range of decreasing values of the bandwidth h_1 (starting with $h_{1,0} = h_0$), with $h_0 > 0$ and $0 < r < 1$. The three-step algorithm reads as follows:

Step 1: Initialization.
Let M denote the number of local maxima of $|D(., h_{1,0})|$ on $]h_{1,0}, 1 - h_{1,0}[$. Let $\{\xi_{0,j}, 1 \leq j \leq M\}$ be the set of points at which the maxima are achieved.

Step 2: Iteration.
Given a set $\{\xi_{i,j}, 1 \leq j \leq M\}$ of local maxima of $|D(., h_{1,i})|$ on $]h_{1,0}, 1 - h_{1,0}[$, let $\xi_{i+1,j}$ denote the local maximum of $|D(., h_{1,i+1})|$ that is nearest to $\xi_{i,j}$.

Step 3: Termination.
Stop the algorithm at iteration $i = \tilde{\imath}$, when the number of data values in some interval $]x - h_{1,i}, x + h_{1,i}[\subseteq]h_{1,0}, 1 - h_{1,0}[$ first falls below a predetermined value.
Take the preliminary estimate $\tilde{\tau}$ of the jump discontinuity x_0 to be this value of $\xi_{\tilde{\imath},j}$ for which $||D(\xi_{\tilde{\imath},j}, h_{1,\tilde{\imath}})| - |D(\xi_{0,j}, h_{1,0})||$ is the largest.

The reason for choosing the preliminary estimator of τ by maximizing $||D(\xi_{\tilde{\imath},j}, h_{1,\tilde{\imath}})| - |D(\xi_{0,j}, h_{1,0})||$ is the following: when the bandwidth $h_{1,i}$ decreases, the slope of the Nadaraya-Watson estimator increases around the jump, i.e. the value taken by the diagnostic function increases. Therefore, it is expected that the most noticable differences in the diagnostic function, when decreasing h, are encountered around jump points. Figure 1 illustrates the three steps of the algorithm for four selected iterations corresponding to $i = 0, 4, 8$ and $\tilde{\imath} = 11$. Figure 1 (a) represents the initial iteration ($i = 0$) and the diagnostic function shows $M = 8$ local maxima at that iteration. For each iteration i the vertical dotted lines present the 8 local maxima, $\xi_{i,j}$ for $j = 1, ..., 8$ that are closest to those obtained in the previous iteration, $\xi_{i-1,j}$ $j = 1, ..., 8$. In the last iteration we then have 8 (considered) local maxima, $\xi_{\tilde{\imath},j}$ for $j = 1, ..., 8$ and we must choose among them by maximizing the difference in magnitude between the maximal values of the diagnostic function achieved in the first and in the last iteration for the 8 local maxima, i.e. $||D(\xi_{\tilde{\imath},j}, h_{1,\tilde{\imath}})| - |D(\xi_{0,j}, h_{1,0})||$, $j = 1, 2, ..., 8$. These values are reported in Table 1 with the values of $\xi_{\tilde{\imath},j}$, $j = 1, ..., 8$. This leads to the conclusion that the preliminary estimator of the jump point is 0.50.

Table 1
Tabulated $||D(\xi_{\tilde{\imath},j}, h_{1,\tilde{\imath}})| - |D(\xi_{0,j}, h_{1,0})||$ and $\xi_{\tilde{\imath},j}$ values for a simulated data set for (10).

j	$\|\|D(\xi_{\tilde{\imath},j}, h_{1,\tilde{\imath}})\| - \|D(\xi_{0,j}, h_{1,0})\|\|$	$\xi_{\tilde{\imath},j}$
1	395.94	0.134286
2	379.04	0.305714
3	683.21	0.420000
4	**887.88**	**0.500000**
5	677.06	0.580000
6	546.55	0.694286
7	418.44	0.785714
8	314.07	0.854285

Note that two parameters h_0 and r appear in the automatic choice of the bandwidth h_1. These choices however are of little importance. They represent respectively the starting value of the range of h_1-values and the multiplicative decreasing step in the sequence of h_1 values. The only precaution is to choose h_0 and r not too small. In other words, the grid of h_1 values should be fine enough and should be of a reasonable range. Safe choices are h_0 large (for example half of the length of the domain of the regression function) and r close to 1.

3.4. More than one discontinuity

In case of several discontinuity points $\tau_1, \tau_2, \cdots, \tau_k$ one has, instead of (2),

$$g(x) = g_0(x) + \sum_{j=1}^{k} \gamma_j \mathbf{1}_{[\tau_j,1]}(x) ,$$

where γ_j denotes the size of the jump of the regression function g at location τ_j and where it is assumed, for simplicity of presentation, that $|\gamma_j| > |\gamma_{j+1}|$ for $j = 1, \cdots, k$ with $\gamma_{k+1} = 0$. Furthermore one needs to assume that the discontinuity points τ_j are such that $\tau_j \in [\delta, 1-\delta]$ for some $\delta > 0$, i.e. are located in the interior of $[0, 1]$. Furthermore, it is assumed that the jumps points are at a least a distance δ located from each other.

For k known, Wu and Chu (1993a) suggest the following procedure for detecting all jump points.

STEP 1: Estimate τ_1 by $\arg\max_x |J(x)|$;

STEP 2: For $j = 2, \cdots, k$, estimate τ_j by searching for the location of the maximum over the set

$$A_j = [\delta, 1-\delta] - \bigcup_{k=1}^{j-1} [\widehat{\tau}_k - 2h, \widehat{\tau}_k + 2h] \; .$$

In other words, for each estimated discontinuity point, remove a symmetric interval of length $4h$, centered at the estimated point, from the domain and search for the next discontuity point in the remaining set.

Yin (1988) proposed a method to estimate the number of discontinuity points, by introducing a (pre-specified) threshold value. Wu and Chu (1993a) propose tests to test for the number of jump points. See also Qiu (1994). Other procedures for determining the number of change-points using kernel-based estimation methods can be found in Beunen (1998) and Bunt, Koch and Pope (1998).

As a second example, we describe the generalization of the two-step procedure of Gijbels, Hall and Kneip (1999) to the case of general k. Suppose first that k is known. One can then proceed in the same way: identify the k positions using a kernel method and improve the rate of convergence by a least-squares step on an interval around each pilot estimator. It is convenient to use the three-step algorithm of Section 3.3 which generalizes easily to the case of k jump discontinuities $\tau_1, \tau_2, \cdots, \tau_k$ by a slight modification of step 3.

Step 3: Termination (addition):
Take the preliminary estimates $\widetilde{\tau}_1, \ldots, \widetilde{\tau}_k$ of the k jump discontinuities $\tau_1, \tau_2, \cdots, \tau_k$ to be those values of $\xi_{\tilde{\imath},j}$ for which $||D(\xi_{\tilde{\imath},j}, h_{1,\tilde{\imath}})| - |D(\xi_{0,j}, h_{1,0})||$ is one of the k largest.

For refining these preliminary estimators, we then fit via least-squares piecewise constant functions within the intervals $[\widetilde{\tau}_j - h_2, \widetilde{\tau}_j + h_2]$ for $j = 1, \ldots, k$. We then obtain the final estimators $\widehat{\tau}_1, \ldots, \widehat{\tau}_k$ of $\tau_1, \ldots, \tau_k$.
Also here data-driven choices of the bandwidth parameters involved have been proposed. See Gijbels and Goderniaux (2002a).

When the number of discontinuity points is unknown it can be estimated simply, by for example using a cross-validation method. For a fixed value of k ($k = 0, 1, 2, 3, \cdots$) locate the k discontinuity points, and estimate the regression curve g by using a standard smoothing method with a good data-driven bandwidth selector on the segments, as indicated in (6) in the case of one jump point. Denote this estimate by $\widehat{g}_k(\cdot)$. Consider the cross-validation sum of squares:

$$\mathrm{CV}(k) = \sum_{i=0}^{n} (Y_i - g_k^{-i}(X_i))^2 \; ,$$

where $g_k^{-i}(X_i)$ is the fit obtained at X_i, assuming k discontinuity points and excluding the data point (X_i, Y_i) when constructing the fit. This approach has been proposed by

Müller and Stadtmüller (1999). The number of discontinuities k is estimated by

$$\widehat{k} = \arg\min_{k\in\{0,1,\ldots\}} \mathrm{CV}(k) ,$$

and the estimated curve is then the one associated with this number of discontinuities.

4. ESTIMATION OF DISCONTINUOUS REGRESSION CURVES

4.1. Indirect methods

As explained in the introduction, once estimates for the locations of the jump discontinuity point are obtained, say $\widehat{\tau}_1, \cdots, \widehat{\tau}_k$, the interval of estimation is divided into $k+1$ intervals $[0, \widehat{\tau}_1[$, $[\widehat{\tau}_1, \widehat{\tau}_2[$, $\cdots$, $[\widehat{\tau}_k, 1]$ and on each segment the regression function is continuous and estimated using standard smoothing techniques (such as for example a local linear smoother with cross-validation bandwidth).

An alternative indirect method has been proposed by Kang, Koo and Park (2000) in the fixed design regression case. They suggest to estimate the location τ and γ via (3) and (4) respectively, to obtain suitably adjusted data

$$Z_i = Y_i - \widehat{\gamma}\mathbf{1}_{[\widehat{\tau},1]}(x_i) , \qquad i = 1, \cdots, n$$

and then to use standard kernel regression methods to estimate the function g_0 in (2).

4.2. Direct methods

Several estimation methods have been proposed for estimating directly the unsmooth regression curve, without estimating first the locations and sizes of the jumps separately.

McDonald and Owen (1986) suggest to obtain for any given point three smoothed estimates of the regression function at that point, based on data to the right, to the left and on both sides of the given point. They used least squares fits to obtain, locally, the three linear fits. These three local linear fits are then used to construct a 'split linear fit' as a weighted average of these three estimates with weights determined by the goodness-of-fit values of the estimates.

Hall and Titterington (1992) proposed an alternative to the smoothed linear fitting method of McDonald and Owen (1986). Let $m \in \{0, 1, 2, \cdots\}$ be a window size. At all design points $x_i = i/n$ three smooths, the central smooth, the right and the left smooth, are computed using a linear conbination of the $(2m+1)$st observations in the neighbourhood of the fixed design point. The central smooth is based on the observation at the design point itself and on those at its m nearest neighbours to the left and to the right. The left (right) smooth is based on the observation at the design point itself and on those at the $2m$ nearest design points to the left (right) of the given design point. More precisely, the central, left and right smooths at the design point $x_i = i/n$ are defined as

$$\widehat{g}_c(x_i) = \sum_{j=-m}^{m} c_j Y_{i+j} \qquad \widehat{g}_l(x_i) = \sum_{j=-2m}^{0} r_{-j} Y_{i+j} \qquad \widehat{g}_r(x_i) = \sum_{j=0}^{2m} r_j Y_{i+j} ,$$

where the constants c_j and r_j are chosen such that the leading terms in Taylor expansions of the expected values of $\widehat{g}_c(x_i)$ and $\widehat{g}_r(x_i)$ coincide. The method of Hall and Titterington (1992) is in fact an indirect method, since it searches for discontinuity points and then constructs the estimate for the unsmooth curve. To find the discontinuity points the central, left and right smooths are compared with one another. If there is no discontinuity and the regression function is locally linear then the three smooths should be very similar. At discontinuity points the three smooths behave quite different, and various diagnostics can be used to detect discontinuity points based on comparisons of the three smooths. These diagnostics involve the choice of a threshold value to determine whether a point is a discontinuity point or not. Values of a threshold are determined by referring to a linear regression function and Gaussian errors.

Recently, Qui (2003) proposed a jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. For each x in the interval $[h_n/2, 1-h_n/2]$, with $h_n > 0$ a bandwidth, obtain a local linear fit to the left and to the right, using respectively a kernel function K_+, defined on $[-1/2, 0)$ and K_-, defined on $[0, 1/2]$. These fits estimate the left and right-hand limits of the function g at the given point x. Clearly, if x is a point of continuity of g, then both estimates are very close. If x is not a jump point but a jump point is near to it, then one of the two estimates will be a good estimate of g. The idea is now to choose for each $x \in [h_n/2, 1 - h_n/2]$ between one of these estimates, according to the residual sum of squares of the two fits. Denote the left-hand and right-hand estimates by $\widehat{g}_+(x)$ and $\widehat{g}_-(x)$ respectively. Further let $\mathrm{RSS}_+(x)$ and $\mathrm{RSS}_-(x)$ denote the residual sum of squares of these fits. Then the proposed estimate for g at the given point x is

$$\widehat{g}(x) = \begin{cases} \widehat{g}_-(x) & \text{if} \quad \mathrm{RSS}_+(x) > \mathrm{RSS}_-(x) \\ \frac{1}{2}\left(\widehat{g}_+(x) + \widehat{g}_-(x)\right) & \text{if} \quad \mathrm{RSS}_+(x) = \mathrm{RSS}_-(x) \\ \widehat{g}_+(x) & \text{if} \quad \mathrm{RSS}_+(x) < \mathrm{RSS}_-(x) \; . \end{cases}$$

An adaptive estimation procedure that simultaneously adapts to the smoothness of the curve and is sensitive to discontinuities of the curve has been proposed by Spokoiny (1998). The basic idea behind the method is to try to find by data the maximal interval such that g within that interval can be well approximated by a polynomial. The fits are evaluated via least-squares analysis of the residuals. The intervals are allowed to be asymmetric. Intuitively: when x is a point near a jump discontinuity τ, the automatic procedure will come up with an asymmetric window excluding the near jump discontinuity.

Other direct estimation methods involve M-smoothers. Chu, Glad, Godtliebsen and Marron (1998) estimate g in (1) by minimizing

$$S(a; x_i) = (-1) \sum_{j=1}^{n} K_h(x_i - x_j) L_g(Y_j - a) \; ,$$

with respect to a, and choosing among the local minima the one that is closest to Y_i. Here K_h and L_g are rescaled versions of kernels K and L. For example $K_h(\cdot) = h^{-1} K(\cdot/h)$. This results into a so-called local constant M-smoother.

An improvement of this is a local linear M-smoother defined as the local minimizer of

$$S(a,b;x_i) = (-1)\sum_{j=1}^{n} K_h(x_i - x_j)L_g(Y_j - a - b(x_i - x_j)) \ ,$$

with respect to a and b, and choosing among the local minima $(\widehat{a}, \widehat{b})$ the one for which $\widehat{a}$ is closest to Y_i. This estimator is introduced and studied by Rue, Chu, Godtliebsen and Marron (2002). Both the constant and linear M-smoothers have been designed for signals showing abrupt changes.

5. TESTING PROBLEMS AND CONFIDENCE INTERVALS AND BANDS

Several testing problems have been dealt with in the literature. Denote by x_0 a point in the interval $]0,1[$, and by $\kappa = \sum_{j=1}^{k} \gamma_j^2$ the sum of squared jump sizes of all jump points. Among the testing problems of interest are:

$$\begin{cases} H_0 : g & \text{is continuous in the point } x_0 \\ H_1 : g & \text{has a jump discontinuity at the point } x_0 \end{cases} \tag{11}$$

$$\begin{cases} H_0 : g & \text{is continuous on the interval }]0,1[\\ H_1 : g & \text{is discontinuous in at least one point on the interval }]0,1[\end{cases} \tag{12}$$

$$\begin{cases} H_0 : \kappa = 0 \\ H_1 : \kappa > 0 \end{cases} \tag{13}$$

In testing problem (11), called a local testing problem, one focuses on testing whether the regression function g has a jump discontinuity at a certain fixed (pre-specified) point x_0. In contrast, testing problem (12), called a global testing problem, aims at finding out whether the regression function is continuous or not. Similarly for problem (13).

Grégoire and Hamrouni (2002b) studied tests for testing problems (11) and (12) based on asymptotic theory for the estimates developed in Grégoire and Hamrouni (2002a). The tests involve the choice of a bandwidth parameter. Gijbels and Goderniaux (2002b) proposed data-driven bootstrap based tests for the same problems, as well as for similar testing problems regarding the νth derivative of a regression function. Horváth and Kokoszka (2002) also study problem (11) but in general for the νth derivative function $g^{(\nu)}$ in the case of fixed regression.

Müller and Stadtmüller (1999) and Dubowik and Stadtmüller (2000) consider testing problem (13) and propose tests based on asymptotic theory. An important practical issue here is also the choice of a span-parameter.

Apart from testing it is also interesting to derive confidence bands (and intervals) for curves with jumps. This has been dealt with in Gijbels, Hall and Kneip (2003) using the estimation procedure introduced by Gijbels, Hall and Kneip (1999) and relying on non-standard bootstrap procedures.

6. DISCONTINUOUS BIVARIATE REGRESSION FUNCTIONS

Given space limitation we will only provide a few basic references for kernel-based methods for detecting jump curves in a regression surface, that generalize in some way the univariate methods discussed in Sections 2 and 3.

Detection of jump curves in regression surfaces by looking at differences of kernel estimates is exploited in Qiu (1997, 2002). Methods relying on partial derivatives of the regression surface can be found in Qiu (1998, 2000). A related approach using local least squares fitting of a plane is used in Qiu and Yandell (1996, 1997).

A generalization to the two-dimensional case of the estimation method of Gijbels, Hall and Kneip (1999) has been worked out by Goderniaux (2001), including automatic selection of the bandwidth parameters involved using bootstrap methods.

References

Antoniadis, A. and Gijbels, I. (2002). Detecting abrupt changes by wavelet methods. *Journal of Nonparametric Statistics*, **14**, 7–29.

Beunen, J.A. (1998). Implementation of a method for detection and location of discontinuities. *Thesis* for Master of Scientific Studies (Statistics). University of Newcastle, Australia.

Bunt, M., Koch, I. and Pope, A. (1998). Counting discontinuities in nonparametric regression. *Research Report 98-2*, University of Newcastle, Australia.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 679–698.

Chu, C.-K., Glad, I.K., Godtliebsen, F. and Marron, J.S. (1998). Edge preserving smoothers for image processing (with discussion). *Journal of the American Statistical Association*, **93**, 526–556.

Cline, D.B.H., Eubank, R.L. and Speckman, P.L. (1995). Nonparametric estimation of regression curves with discontinuous derivatives. *Journal of Statistical Research*, **29**, 17–30.

Dubowik, Ch. and Stadtmüller, U. (2000). Detecting jumps in nonparametric regression. In *Asymptotics in Statistics and Probability*, M.L. Puri (editors), pp. 171–184.

Eubank, R.L. and Speckman, P.L. (1994). Nonparametric estimation of functions with jump discontinuities. In *Change-point Problems*, IMS–Lecture Notes, Eds. E. Carlstein, H.-G. Müller and D. Siegmund, **23**, 130–143.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall, New York.

Gijbels, I. and Goderniaux, A.-C. (2002a). Bandwidth selection for change point estimation in nonparametric regression. Institut de Statistique, Université catholique de Louvain, *Discussion Paper* # 0024. *Revised version.*

Gijbels, I, and Goderniaux, A.-C. (2002b). Bootstrap test for detecting change points in nonparametric regression. Institut de Statistique, Université catholique de Louvain, *Discussion Paper* # 0144. http://www.stat.ucl.ac.be

Gijbels, I, and Goderniaux, A.-C. (2002c). Discontinuity detection in derivatives of a regression function. Institut de Statistique, Université catholique de Louvain, *Discussion Paper* # 0130. http://www.stat.ucl.ac.be

Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *The Annals of the Institute of Statistical Mathematics*, **51**, 231–251.

Gijbels, I., Hall, P. and Kneip, A. (2003). Interval and band estimation for curves with jumps. *Journal of Applied Probability*, Special Issue in honour of Chris Heyde, to appear.

Girard, D. (1990). Détection de discontinuités dans un signal par inf-convolution spline et validation croisée. *Technical Report RR 702-I-M*, University of Grenoble, France.

Goderniaux, A.-C. (2001). Automatic detection of change-points in nonparametric regression. *Doctoral dissertation.* Institute of Statistics, Catholic University of Louvain, Louvain-la-Neuve, Belgium.

Grégoire, G. and Hamrouni, Z. (2002a). Change-point estimation by local linear smoothing. *Journal of Multivariate Analysis*, **83**, 56–83.

Grégoire, G. and Hamrouni, Z. (2002b). Two nonparametric tests for change-point problem. *Journal of Nonparametric Statistics*, **14**, 87–112.

Hall, P. and Titterington, D.M. (1992). Edge preserving and peak-preserving smoothing. *Technometrics*, **34**, 429–440.

Hall, P. and Wehrly, T.E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association*, **86**, 665–672.

Horváth, L. and Kokoszka, P. (2002). Change-point detection with non-parametric regression. *Statistics*, **36**, 9–31.

Kang, K.-H., Koo, J.-Y. and Park, C.-W. (2000). Kernel estimation of discontinuous regression functions. *Statistics & Probability Letters*, **47**, 277–285.

Koch, I. and Pope, A. (1997). Asymptotic bias and variance of a kernel-based estimator for the location of a discontinuity. *Journal of Nonparametric Statistics*, **8**, 45–64.

Koo, J.-Y. (1997). Spline estimation of discontinuous regression function. *Journal of Computational and Graphical Statistics*, **6**, 266–284.

Korostelev, A.P. (1987). On minimax estimation of a discontinuous signal. *Theory of Probability and its Applications*, **32**, 727–730.

Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics, **82**. Springer-Verlag, New York.

Laurent, P.J. and Utreras, F. (1986). Optimal smoothing of noisy broken data with spline functions. *Journal of Approximation Theory and its Applications*, **2**, 71–94.

Lee, Thomas C.M. (2002). Automatic smoothing for discontinuous regression functions. *Statistica Sinica*, **12**, 823–842.

Loader, C.R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics*, **24**, 1667–1678.

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

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. *The Annals of Statistics*, **20**, 737–761.

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

Müller, H.-G. and Stadtmüller, U. (1999). Discontinuous versus smooth regression. *The Annals of Statistics*, **27**, 299–337.

Oudshoorn, C.G.M. (1998). Asymptotically minimax estimation of a function with jumps. *Bernoulli*, **4**, 15–33.

Potier, C. and Vercken, C. (1994). Spline fitting numerous noisy data with discontinuities. In *Curves and Surfaces*, Eds Laurent, P.J. et al., 477–480, Academic Press.

Qiu, P. (1994). Estimation of the number of jumps of the jump regression functions. *Communications in Statistics, Theory and Methods*, **23**, 2141–2155.

Qiu, P. (1997). Nonparametric estimation of jump surface. *Sankhyā, Series A*, **59**, 268–294.

Qiu, P. (1998). Discontinuous regression surfaces fitting. *The Annals of Statistics*, **26**, 2218–2245.

Qiu, P. (2000). Some recent developments on edge detection and image reconstruction based on local smoothing and nonparametric regression. In *Recent Research Developments in Pattern Recognition*, **1**, 41–49.

Qiu, P. (2002). A nonparametric procedure to detect jumps in regression surfaces. *Journal of Computational and Graphical Statistics*, **11**, 799–822.

Qiu, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *Journal of Nonparametric Statistics*, to appear.

Qiu, P. and Yandell, B. (1996). Discontinuity detection in regression surfaces. In *ASA Proc. Statist. Comput. Sect.*, 85–90.

Qiu, P. and Yandell, B. (1997). Jump detection in regression surfaces. *Journal of Computational and Graphical Statistics*, **6**, 332–354.

Qiu, P. and Yandell, B. (1998a). A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics*, **40**, 141–152.

Qiu, P. and Yandell, B. (1998b). Discontinuous regression surfaces fitting. *The Annals of Statistics*, **26**, 2218–2245.

Raimondo, M. (1998). Minimax estimation of sharp change points. *The Annals of Statistics*, **26**, 1379–1397.

Rue, H., Chu, C.-K., Godtliebsen, F. and Marron, J.S. (2002). M-smoother with local linear fit. *Journal of Nonparametric Statistics*, **14**, 155–168.

Shiau, Jyh-Jen Horng (1988). Efficient algorithms for smoothing spline estimation of functions with or without discontinuities. In *Comput. Sci. Statist.*: Proc. 20th Symp. Interface, Wegman, Edward J., ed., 260–265.

Speckman, P.L. (1995). Fitting curves with features: semiparametric change-point methods. *Computing Science and Statistics*, **26**, 257–264.

Spokoiny, V.G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26**, 1356–1378.

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.

Wang, Y. (1998). Change curve estimation via wavelets. *Journal of the American Statistical Association*, **93**, 163–172.

Wang, Y. (1999). Change-point analysis via wavelets for indirect data. *Statitica Sinica*, **9**, 103–118.

Wu, J.S. and Chu, C.K. (1993a). Kernel type estimation of jump points and values of regression function. *The Annals of Statistics*, **21**, 1545–1566.

Wu, J.S. and Chu, C.K. (1993b). Modification for boundary effects and jump points in nonparametric regression. *Nonparametric Statistics*, **2**, 341 – 354.

Wu, J.S. and Chu, C.K. (1993c). Nonparametric function estimation and bandwidth selection for discontinuous regression functions. *Statistica Sinica*, **3**, 557 – 576.

Yin, Y.Q. (1988). Detection of the number, locations and magnitudes of jumps. *Communications in Statistics, Stochastic Models*, **4**, 445–455.

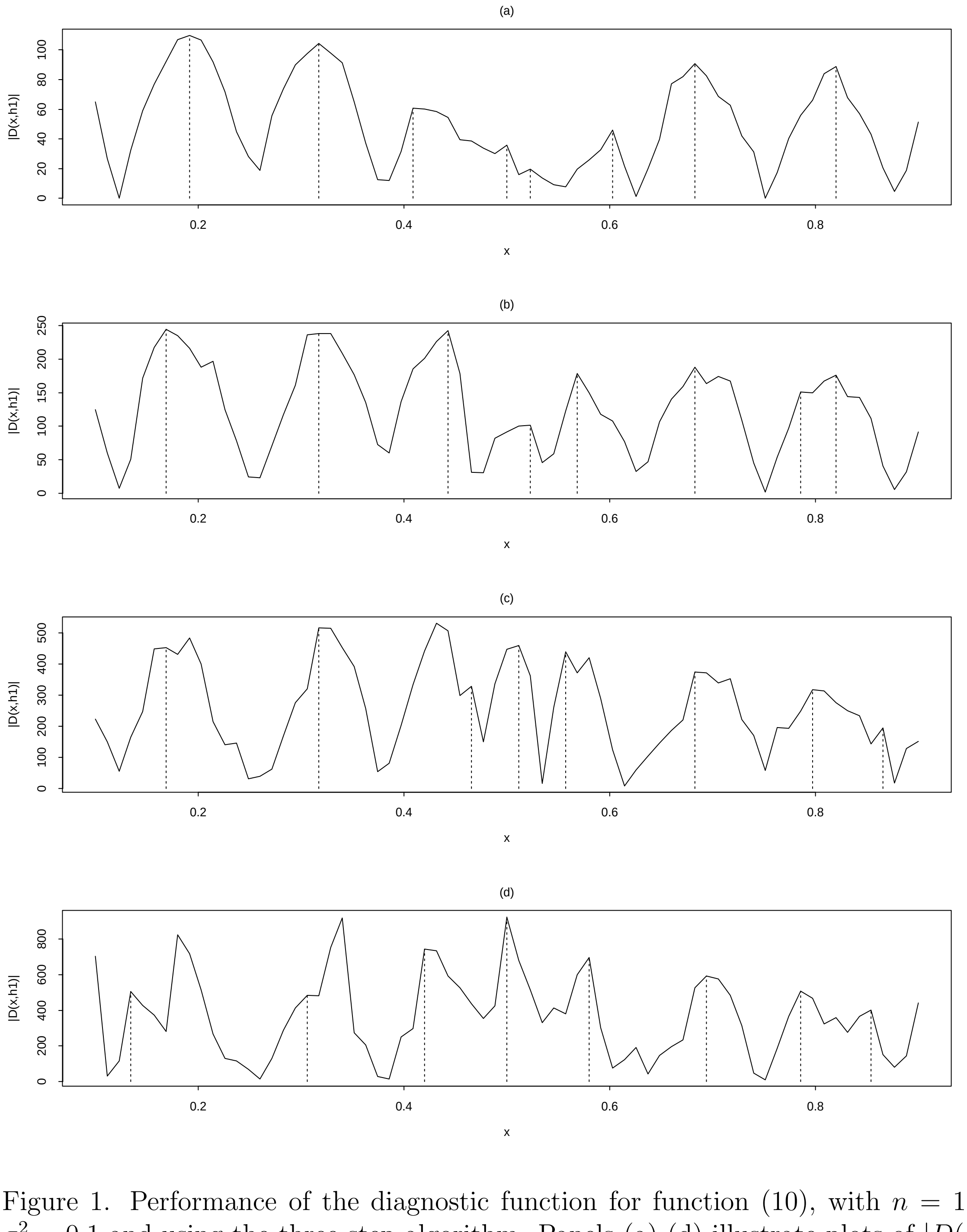

Figure 1. Performance of the diagnostic function for function (10), with $n = 100$ and $\sigma^2 = 0.1$ and using the three-step algorithm. Panels (a)-(d) illustrate plots of $|D(x, h_{1,i})|$ for $i = 0, 4, 8$ and $i = \tilde{\imath} = 11$ respectively.