Bandwidth Selection for Changepoint Estimation in Nonparametric Regression

Irène Gijbels and Anne-Cécile Goderniaux

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

Nonparametric estimation of abrupt changes in a regression function involves choosing smoothing (bandwidth) parameters. The performance of estimation procedures depends heavily on this choice. So far, little attention has been paid to the crucial issue of choosing appropriate bandwidth parameters in practice. In this article we propose a bootstrap procedure for selecting the bandwidth parameters in a nonparametric two-step estimation method. This method results in a fully data-driven procedure for estimating a finite (but possibly unknown) number of changepoints in a regression function. We evaluate the performance of the data-driven procedure via a simulation study, which reveals that the fully automatic procedure performs quite well. As an illustration, we apply the procedure to some real data.

KEY WORDS: Bandwidth; Bootstrap; Cross-validation; Discontinuity points; Least squares fitting.

1. INTRODUCTION

There is a vast literature on the estimation of smooth regression functions. In many applications in physical sciences, engineering, medicine, and economics, the regression function is smooth except at a finite number of locations (e.g., jump discontinuities). (See Müller and Stadtmüller 1999 for more background information and many examples of data applications.) Estimation of an unknown regression function with a finite number of discontinuities is usually done in several stages. In the first (and more challenging) stage, the locations of the possible jump discontinuities are estimated. The second stage is the actual estimation of the regression curve, usually by fitting nonparametrically smooth curves to the left and to the right of the estimated locations of jump discontinuities, relying on techniques available for estimating smooth curves. Another approach was described by Kang, Koo, and Park (2000), who estimated the locations of the jump discontinuities as well as the sizes of the jumps. Using these estimations, they then adjusted the data suitably and applied ordinary smoothing techniques to these adjusted data. Thus, estimating regression functions with jump discontinuities involves estimation of the number of jump discontinuities and their locations, the jump sizes, and the regression function itself.

The literature on nonparametric estimation of regression functions that are smooth except at some points is by now quite large. Kernel-based estimation methods have been studied by Canny (1986), Korostelev (1987), Hall and Titterington (1992), Müller (1992), Wu and Chu (1993a, b, c), Chu (1994), Speckman (1994), Eubank and Speckman (1994), Bunt, Koch, and Pope (1995), and Kang et al. (2000), among others. Local polynomial methods were used by McDonald and Owen (1986), Leclerc and Zucker (1987), Loader (1996), Horváth and Kokoszka (1997), Qiu and Yandell (1998), Spokoiny (1998), Hamrouni (1999), and Grégoire and Hamrouni (2002), among others. Work on spline-based methods has been reported by Laurent and Utreras (1986), Girard (1990), and Koo (1997); on wavelet-based methods, by Mallat and Hwang (1992), Potier and Vercken (1994), Wang (1995), Raimondo (1998), Oudshoorn (1998), and Antoniadis and Gijbels (2002). Müller and Song (1997) and Gijbels, Hall, and Kneip (1999) proposed two-step procedures.

Two important issues arise when dealing with nonparametric estimation of a regression curve with jump discontinuities. The first issue is knowledge of the finite number of jump discontinuities. Often it is assumed that this number is known, whereas this is often not the case in applications. The second issue is that any nonparametric estimation method involves the choice of parameters, call them "smoothing parameters," and the performance of the estimation procedures depends heavily on the choice of these parameters. Hence it is very important to address the issue of how to choose these parameters in practice. Some attention has been paid to these issues in the literature. Authors who have explored the first issue include Yin (1988), Wu and Chu (1993a), Bunt et al. (1998), Beunen (1998), Oudshoorn (1998), Müller and Stadtmüller (1999), and Antoniadis and Gijbels (2002). As far as we know, little work has been done on the second issue; Wu and Chu (1993c) and Spokoiny (1998) have provided theoretical contributions.

In this article we focus on estimating the locations of the jump discontinuities, particularly on the practical choice of the smoothing (bandwidth) parameters involved. We develop an estimation procedure with a data-driven choice of the bandwidth parameters and with a built-in estimation of the number of discontinuity points that performs well in practice. We use the two-step estimation method proposed by Gijbels et al. (1999), for which it has been shown that the estimator for the location of a jump discontinuity achieves the optimal rate n^{-1}, where n is the sample size. Other estimation procedures achieve the same optimal rate of convergence, but we choose this method for two reasons. First, this estimation method demonstrated, via extensive simulation good finite-sample performance for various regression functions. Second, as we show later, the method leads to a fully data-driven procedure. The simulation study carried

TECHNOMETRICS, FEBRUARY 2004, VOL. 46, NO. 1
DOI 10.1198/004017004000000130

out for the two-step estimation method by Gijbels et al. (1999) revealed that the choice of the bandwidth parameters is crucial in practice, not only for this estimation method, but also for any other estimation method in this context. In Section 2.3 we illustrate this point by comparing the two-step estimation method with another estimation procedure. The second issue motivates the present article.

We propose a bootstrap procedure for selecting the bandwidth parameters. Gijbels et al. (2004) dealt with interval and band estimation for curves with jumps and suggested this selection procedure. Some of the basic ingredients for the bootstrap procedure used in our bandwidth selection problem are similar to these described in that article, and hence for details we refer to this work. A theoretical justification of the bootstrap bandwidth selection procedure introduced here would rely on theoretical results established in Gijbels et al. (2004). We tested our fully data-driven estimation procedure on various simulation models; all demonstrated very good performance.

The article is organized as follows. In Section 2 we briefly describe the estimation method and discuss the bandwidth parameters it involves. We also compare this method with the estimation method developed by Grégoire and Hamrouni (2002). In Section 3 we introduce the bootstrap bandwidth selection procedure and discuss how to choose the number of discontinuities. We present numerical study of the fully data-driven estimation method, along with an illustration using some real data, in Section 4. We provide some discussion in Section 5.

2. ESTIMATING JUMP DISCONTINUITIES

2.1 Statistical Model

Consider a sample of n observed data pairs $\chi = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ generated from the model

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

The design points X_i are either regularly spaced on $I = [0, 1]$ or are the order statistics of a random sample from a distribution having a density f supported on I. The errors ε_i are assumed to be idd distributed with mean 0 and finite variance σ^2. The unknown function $g(\cdot)$ is smooth except at a finite number of jump discontinuities. For the moment, we assume that the number of jump points is known, and, moreover, for ease of presentation, we explain the estimation and bootstrap selection procedure for the case of a single jump discontinuity in the regression function at the point $x_0 \in]0, 1[$. We address the case of more than one jump discontinuity in Section 3.3 and discuss choosing the number of jump discontinuities in Section 3.4.

2.2 Estimation Procedure

The two-step estimation procedure for estimating x_0 introduced by Gijbels et al. (1999) involves: (1) obtaining a rough estimate of x_0, say $\tilde{x}_0$, and (2) improving $\tilde{x}_0$ using least squares estimation in an interval around $\tilde{x}_0$, say $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$, with $h_2 > 0$ a bandwidth parameter, to obtain an improved estimator $\hat{x}_0$ of x_0. The first step, called the diagnostic step, is based on the derivative of the Nadaraya–Watson estimator of the regression function and involves a bandwidth parameter h_1. Section 5 presents some discussion on various aspects of this estimation method.

2.2.1 *Diagnostic Step.* One way to detect a jump discontinuity is to look at locations with high derivatives. A possible diagnostic function is the derivative of the Nadaraya–Watson estimator (see Nadaraya 1964; Watson 1964), defined as

$$D(x, h_1) = \frac{\partial}{\partial x}\left(\frac{\sum_{i=1}^{n} K\{(x - X_i)/h_1\} Y_i}{\sum_{i=1}^{n} K\{(x - X_i)/h_1\}}\right), \tag{2}$$

where K is a compactly supported (with support $[-v, v]$) differentiable kernel function and $h_1 > 0$ is a bandwidth. A first rough estimator of x_0 is then given by

$$\tilde{x}_0 = \operatorname*{argmax}_{x \in [vh_1, 1 - vh_1]} |D(x, h_1)|.$$

The rate of convergence of this preliminary estimator $\tilde{x}_0$ is $n^{-1}(\log n)^{1/2}$. The second, least squares, step will lead to a n^{-1} estimator (see Gijbels et al. 1999 for further details).

2.2.2 *Least Squares Step.* We construct an interval concentrated around $\tilde{x}_0$, namely $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$, to which x_0 belongs with high probability. Denote by $\{i_1, i_1 + 1, \dots, i_2\}$ the set of integers i for which $X_i \in [\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$. We fit a step function on the interval $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$ using least squares. We estimate that the jump discontinuity occurs between design points X_{i_0} and X_{i_0+1}, where i_0 is chosen to minimize the sum of squares,

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

Denote by $\hat{i}_0$ the minimizer of this sum of squares. The final estimator of the jump discontinuity x_0 is then defined as the midpoint between $X_{\hat{i}_0}$ and $X_{\hat{i}_0+1}$: $\hat{x}_0 = \frac{1}{2}(X_{\hat{i}_0} + X_{\hat{i}_0+1})$.

2.3 Comparison With an Available Method

In this section we present a comparison between the performances of the two-step estimation method explained earlier and the method proposed by Grégoire and Hamrouni (2002) that also achieves the best rate of convergence n^{-1} when using a specific class of kernel functions. The latter method estimates the location of the discontinuity point by that value of t for which $|\hat{g}_+(t) - \hat{g}_-(t)|$ is maximal, where $\hat{g}_+$ and $\hat{g}_-$ are local linear regression estimates of the right and left limits of g. This method requires the choice of a bandwidth parameter h, because it involves local linear estimation. As an illustration, we applied the two-step estimation method and this one-step method to the regression model (1) with regression functions g_1 and g_2 given later (Sec. 4.1) in (4) and (5), respectively, and depicted in Figures 3(a) and (b). These functions show one jump discontinuity at the point .5 of sizes 1 and -2. We simulated 100 samples from model (1) with $N(0; \sigma^2)$ distributed errors and with $n = 50$ and $\sigma^2 = .1$ for g_1 and $n = 200$ and $\sigma^2 = .5$ for g_2. For each fixed bandwidth in the set $h = .03 + .015j$ for $j = 0, 1, \dots, 18$, and for both methods, we obtain the estimate of the unknown location point for the 100 samples. Figure 1 provides a boxplot of these 100 values, for each fixed value of h that we considered (indicated on the horizontal axis). Panels (a) and (c) corre-

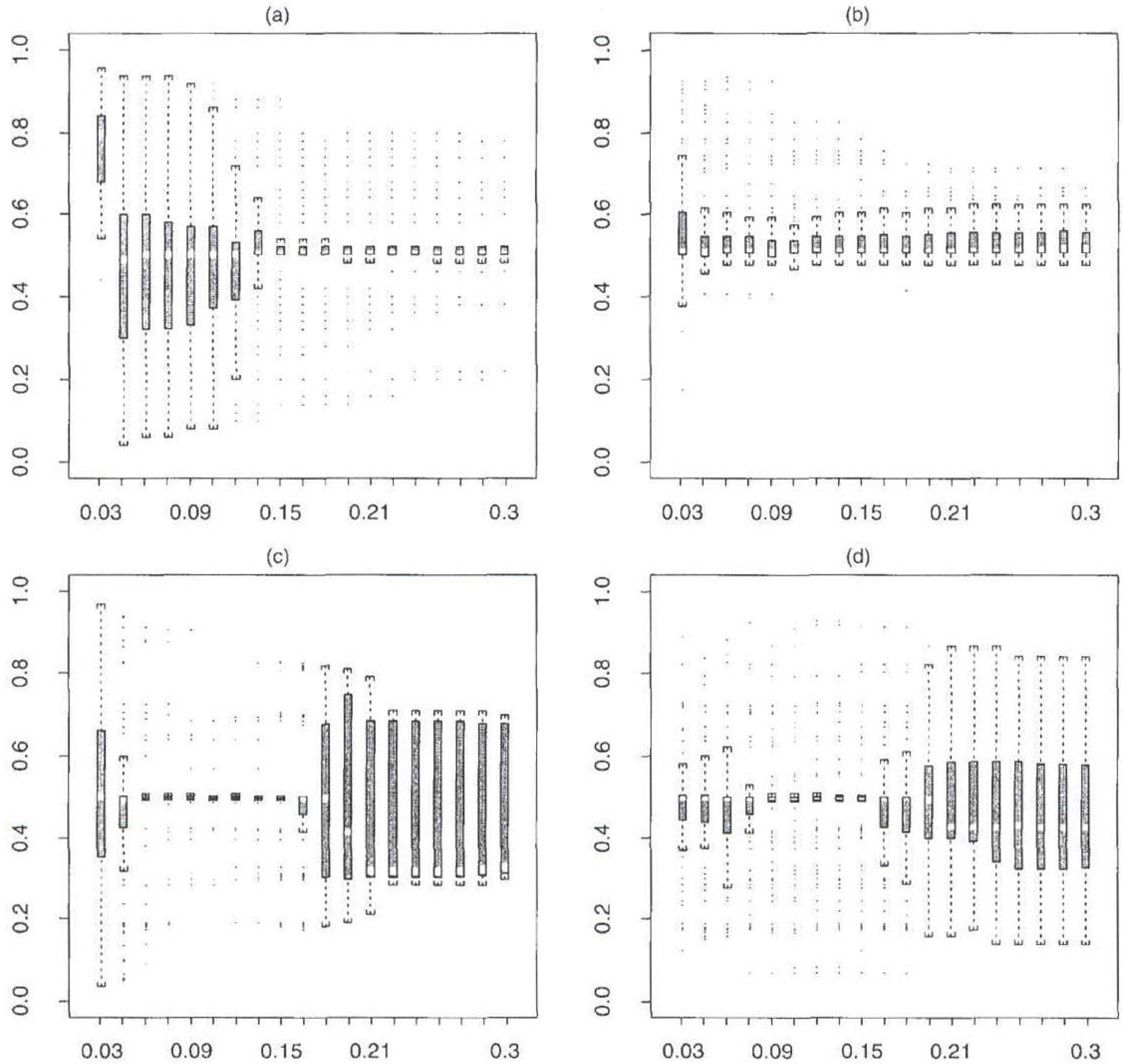

Figure 1. Boxplots of the Estimated Values for the Discontinuity Point x_0 for the Functions g_1 and g_2 Using the Two-Step Estimation Method [(b) and (d)] and the Method of Grégoire and Hamrouni (2002) [(a) and (c)]. (a) and (b) Results for g_1; (c) and (d) results for g_2.

spond to the one-step method, whereas panels (b) and (d) correspond to the two-step procedure. The outer whiskers extend to the extreme values of the data (i.e., the smallest and largest observations) or to the nearest observation not exceeding a distance 1.5 times the interquartile range measured from the quartiles, whichever is smallest. This figure is an illustration of how the methods depend on the choice of the bandwidth parameter, and of how the methods perform for finite samples. Figure 1 shows that the choice of the bandwidth parameter is indeed important. It also illustrates that the quadratic function g_1 [panels (a) and (b)] represents an easier example than the cosinus function g_2 [panels (c) and (d)].

Another remark that can be made is that the method of Grégoire and Hamrouni (2002) tends to be more variable in general, looking at the interquartile range. Figure 1(c), for example, shows that the median values (the little white squares) of the estimate are quite far off from the true value of .5 for a whole range of h values. The two-step estimation method performs better here. This illustrates a feature of the latter method: If at the first step the preliminary estimator of x_0 is rather bad (but not too far off), then the second step gives a chance to "correct" for this.

2.4 Choice of the Bandwidth Parameters

Each step in the two-step estimation procedure involves a bandwidth parameter, h_1 or h_2. Both bandwidth choices are crucial for the performance of the procedure, as has been explored and discussed in detail by Gijbels et al. (1999).

The bandwidth h_1 should be small enough to capture the behavior at the jump discontinuity, but not so small that the diagnostic curve shows artificial peaks. The bandwidth h_2 used in the second step determines the interval in which we carry out the least squares approximation, namely $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$. If this interval is wide and g differs greatly from a step function, then a local constant approximation may be very poor, leading to a bad estimate of i_0 and hence of x_0. Thus the interval should be as small as possible, but still contain sufficient data points for a reasonable least squares fit.

A simplification is obtained by taking the same bandwidth in each step, that is, $h_1 = h_2$. In Section 3 we present the fully data-driven, bootstrap-type procedure for this simplified case. For the more general case of two different bandwidths h_1 and h_2, a fully data-driven procedure has also been developed. Details of this more flexible procedure can be obtained from the authors on request.

2.5 Identification Problem

Gijbels et al. (1999) noted that a simple search for the maximum in the diagnostic function $|D(x, h)|$ might be problematic. As an example, consider the cosine function in (5), depicted in Figure 3(b). This function has many steep declines and inclines. Thus the diagnostic function exhibits many local maxima, and the largest may correspond to a steep gradient instead of to the jump discontinuity. We need to identify the local maximum of the diagnostic function that corresponds to the jump discontinuity. Gijbels et al. (1999) proposed a four-step algorithm for dealing with this identification problem. The algorithm also includes an automatic choice of the bandwidth parameter h_1 in the diagnostic step. Let $h_{1,i} = h_0 r^i$, for $i = 0, 1, 2, \ldots$, be a range of decreasing values of the bandwidth h_1, with $h_0 > 0$ and $0 < r < 1$. The four-step algorithm proceeds as follows:

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

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

Step 3: Termination. Stop the algorithm at iteration $i = \tilde{i}$, when the number of data values in some interval $]x - h_{1,i}, x + h_{1,i}[\subseteq]vh_{1,0}, 1 - vh_{1,0}[$ first falls below a predetermined value. The preliminary estimate $\tilde{x}_0$ of the jump discontinuity x_0 is the value of $\xi_{\tilde{i},j}$ for which $|D|(\xi_{\tilde{i},j}, h_{1,\tilde{i}})| - |D(\xi_{0,j}, h_{1,0})||$ is largest.

Step 4: Least squares. Use local least squares within the interval $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$ to obtain the final estimator $\hat{x}_0$.

The first three steps of the algorithm yield a preliminary estimator for x_0. The two parameters h_0 and r that appear in the automatic choice of the bandwidth h_1 are of little importance, because they represent the starting value of the range of h_1 values and the multiplicative decreasing step in the sequence of h_1 values. The grid of h_1 values should be fine enough and should be of a reasonable range; thus safe choices are h_0 large (e.g., half of the length of the domain of the regression function) and r close to 1.

3. FULLY DATA-DRIVEN ESTIMATION PROCEDURE

The estimator $\hat{i}_0$ obtained as in Section 2.2 is integer-valued and may differ in absolute value from the theoretical (random) i_0 by $0, 1, 2, \ldots$. Ideally, the estimator $\hat{i}_0$ equals i_0 with high probability. We use a bootstrap procedure to estimate the probability $\Pr(\hat{i}_0 - i_0 = 0) \equiv p_0$ and to select the bandwidth for which this estimated probability is maximal. We first describe in detail the bootstrap algorithm for estimating p_0, and then specify the actual bandwidth selection procedure. We provide examples of implementations in Section 4, and discuss some alternative selection criteria in Section 5.

3.1 Bootstrap Algorithm

Step 1: Estimation of g and computation of residuals. Let $\hat{x}_0 = \frac{1}{2}(X_{\hat{i}_0} + X_{\hat{i}_0+1})$ denote the estimator introduced in Section 2.2. Using local linear regression (see, e.g., Fan and Gijbels 1996), we construct $\hat{g}$ on $[0, \hat{x}_0]$ and $]\hat{x}_0, 1]$. We define $\tilde{\varepsilon}_i = Y_i - \hat{g}(X_i)$ for $i = 1, \ldots, n$ and $\bar{\varepsilon}$ the mean of $\tilde{\varepsilon}_i$, and $\hat{\varepsilon}_i = \tilde{\varepsilon}_i - \bar{\varepsilon}$, the centralized estimated residuals.

Step 2: Monte Carlo simulation. Conditional on the observed sample $\chi = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, we consider $\varepsilon_1^*, \ldots, \varepsilon_n^*$, a resample drawn randomly with replacement from the set $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$. We define

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

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

Step 3: Determination of the bootstrap probability. Using the method described in Section 2.2, we compute the analog $\hat{i}_0^*$ and $\hat{x}_0^* = \frac{1}{2}(X_{\hat{i}_0^*} + X_{\hat{i}_0^*+1})$ of $\hat{i}_0$ and $\hat{x}_0$ for the resample χ^* rather than the sample χ. From B bootstrap replications, we obtain B values of $\hat{i}_0^*$, denoted by $\hat{i}_0^{*b}$, $b = 1, 2, \ldots, B$, and we evaluate the discrete probability $\Pr(\hat{i}_0^* - \hat{i}_0 = 0|\chi)$ via

$$\frac{1}{B}\sum_{b=1}^{B} \#\{b : \hat{i}_0^{*b} = \hat{i}_0\}. \tag{3}$$

Using the foregoing procedure, we can evaluate all discrete probabilities $\Pr(\hat{i}_0^* - \hat{i}_0 = k|\chi)$, $k = 0, -1, 1, -2, 2, \ldots$. These have been studied theoretically by Gijbels et al. (2004), who showed that under some regularity conditions and for fixed design $\sup_{k=0,-1,1,\ldots} |\Pr(\hat{i}_0^* - \hat{i}_0 = k|\chi) - \Pr(\hat{i}_0 - i_0 = k)| \to 0$ in probability. The same holds for random design; the only difference is that probabilities must be considered conditionally on the realized design $X_1, \ldots, X_n$.

3.2 Bootstrap Bandwidth Selection Method

With the foregoing bootstrap procedure, the bandwidth selection is simple and reads as follows: For a set of potential bandwidths h, choose that bandwidth for which the bootstrap estimate of $\Pr(\hat{i}_0 - i_0 = 0)$ is maximum. More precisely, denote by h_j, $j = 0, \ldots, H$ the set of potential bandwidths. For each bandwidth h_j, we obtain a preliminary estimator $\tilde{x}(h_j)$, that is, the value of x that maximizes the diagnostic function $|D(x, h_j)|$. Next we construct an interval of length $2h_j$ around $\tilde{x}(h_j)$ and use least squares fitting to obtain the final estimator corresponding to this bandwidth value h_j. Using the bootstrap algorithm described earlier, we then estimate, via (3), the probability $\Pr(\hat{i}_0 - i_0 = 0)$ associated with that fixed bandwidth h_j. Finally, we select the bandwidth that yields the largest bootstrap-estimated probability. We denote this bandwidth by $\hat{h}_{\text{boot}}$. We then use this bandwidth to calculate the final estimator of the jump discontinuity. To be safe, one should consider a sufficiently large set of potential bandwidths.

Functions with possible identification problems, as indicated in Section 2.5, require the (more sophisticated) four-step algorithm. The bandwidth h_1 is chosen automatically by the algorithm, and the bandwidth h_2 is chosen as before. We select a (large) set of potential bandwidths $h_{2,j}$ for $j = 0, \ldots, H$ and estimate the discrete probability $\Pr(\hat{i}_0 - i_0 = 0)$ for each of them, then choose that bandwidth h_2 from the candidate set that maximizes the bootstrap estimate of this probability.

To achieve a fully data-driven bandwidth selection procedure, we use cross-validation (CV) to estimate the (smooth) regression functions. Applying the same bandwidth to estimate g to the left of $\hat{x}_0$ and to the right of $\hat{x}_0$, and denoting the local linear fit on $[0, \hat{x}_0]$ by $\hat{g}_1$ and the local linear fit on $[\hat{x}_0, 1]$ by $\hat{g}_2$ we choose h that minimizes the CV quantity

$$\mathrm{CV}(h) = \sum_{i=1}^{\hat{i}_0} \{\hat{g}_1^{-i}(X_i; h) - Y_i\}^2 + \sum_{i=\hat{i}_0+1}^{n} \{\hat{g}_2^{-i}(X_i; h) - Y_i\}^2,$$

where $\hat{g}_1^{-i}(\cdot)$ and $\hat{g}_2^{-i}(\cdot)$ denote the estimators $\hat{g}_1$ and $\hat{g}_2$ obtained by disgarding the ith data point on the interval $[0, \hat{x}_0]$ and $[\hat{x}_0, 1]$. The CV bandwidth selector is then defined as $\hat{h}_{\mathrm{CV}} = \mathrm{argmin}_h \mathrm{CV}(h)$. In our simulation study presented in Section 4, we incorporate more flexibility by allowing two different bandwidths for estimating g to the left and to the right of $\hat{x}_0$.We discuss alternative data-driven bandwidth selectors for this step in Section 5.

3.3 Generalization for More Than One Discontinuity

The generalization of the data-driven bandwidth selection procedure to the case of more than one jump discontinuity is rather straightforward. For convenience, we use the four-step algorithm of Section 2.5, which generalizes easily to the case of k jump discontinuities $x_1, x_2, \ldots, x_k$ by slight modifications of steps 3 and 4:

Step 3′: Termination (addition). Take the preliminary estimates $\tilde{x}_1, \ldots, \tilde{x}_k$ of the k jump discontinuities $x_1, x_2, \ldots, x_k$ to be those values of $\xi_{\tilde{i},j}$ for which $||D(\xi_{\tilde{i},j}, h_{1,\tilde{i}})| - |D(\xi_{0,j}, h_{1,0})||$ is one of the k largest.

Step 4′: Least squares (addition). Use local least squares within the interval $[\tilde{x}_j - h_2, \tilde{x}_j + h_2]$ for $j = 1, \ldots, k$, and obtain the final estimators $\hat{x}_1, \ldots, \hat{x}_k$ of $x_1, \ldots, x_k$.

The first three steps of the generalized four-step algorithm lead to preliminary estimators for $x_1, \ldots, x_k$, involving an automatic choice of the bandwidth h_1 as in Section 2.5. An illustration of this four-step algorithm is provided in Figure 2, which deals with estimation of the piecewise quadratic function $g(x) = 4x^2 + 1.2I(x > .2) + .8I(x > .5)$, where $I(A)$ denotes the indicator function on a set A. This function has two jump discontinuities of size 1.2 and .8 occurring at .2 and .5. Figure 2 presents four selected iterations in the four-step algorithm with $h_0 = .1$ and $r = .9$, corresponding to $i = 0, 4, 10$, and the final iteration step with $\tilde{i} = 15$. In the initial iteration step ($i = 0$), the diagnostic function shows $M = 3$ local maxima, and hence the four-step algorithm searches in each iteration step for the closest local maxima, indicated by the vertical dotted lines in Figure 2. The preliminary estimates are the two values of $\xi_{\tilde{i},j}$, $j = 1, 2, 3$, for which $||D(\xi_{\tilde{i},j}, h_{1,\tilde{i}})| - |D(\xi_{0,j}, h_{1,0})||$ is one of the two largest. These differences are the differences in magnitude between the maximal values of the diagnostic function achieved in the first and the last iteration step for the three local maxima. The first three steps of the four-step algorithm track back the jump points nicely and lead to the preliminary estimates .203 and .511 of the true values .2 and .5.

For the least squares fitting, we select, via the bootstrap algorithm, a bandwidth h_2 for each jump discontinuity separately. More precisely, we select a (large) set of potential bandwidths $h_{2,j}$ for $j = 0, \ldots, H$ and then estimate, for each jump point x_ℓ, $\ell = 1, \ldots, k$, the discrete probability $\Pr(\hat{i}_{0,\ell} - i_{0,\ell} = 0) \equiv p_{0,\ell}$, where obviously the extra subscript ℓ refers to the jump discontinuity x_ℓ. For each jump discontinuity, we choose that bandwidth h_2 from the candidate set that maximizes the bootstrap estimate of the probability $p_{0,\ell}$.

We need to ensure that the interval around $\tilde{x}_\ell$ in the least squares step contains only one jump discontinuity. Hence the set of potential bandwidths should not contain large bandwidth values.

3.4 Determining the Number of Discontinuities

In most practical examples, the number of discontinuities k ($k \geq 0$) is not known, and hence we use the automatic bootstrap selection method described previously. We compare the quality of the fitted curves for various fixed values of k and select the one with the best fit in terms of minimizing the CV sum of squares,

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

where $g_k^{-i}(X_i)$ is the fit obtained at X_i, assuming k changepoints and excluding the data point(X_i, Y_i) when constructing the fit. This CV approach was proposed by Müller and Stadtmüller (1999). The number of discontinuities k is then estimated by

$$\hat{k} = \operatorname*{argmin}_{k \in \{0,1,\ldots\}} \mathrm{CV}(k),$$

and the estimated curve is then the one associated with this number of discontinuities. With this CV rule for selecting the number of discontinuities, we have a fully data-driven method for estimating curves with possible jump discontinuities.

4. NUMERICAL STUDY

In this section we evaluate, via a simulation study, the fully data-driven estimation procedure developed in Section 3 and provide an illustration with some real data. Emphasis is on the evaluation of the bootstrap bandwidth selection method, but we also briefly demonstrate the performance of the CV method in determining the number of discontinuities, in conjunction with the bootstrap selection method.

4.1 Simulation Study

We consider the regression functions

$$g_1(x) = 4x^2 + I(x > .5) \qquad (4)$$

and

$$g_2(x) = \cos\{8\pi(.5 - x)\} - 2\cos\{8\pi(.5 - x)\}I(x > .5). \qquad (5)$$

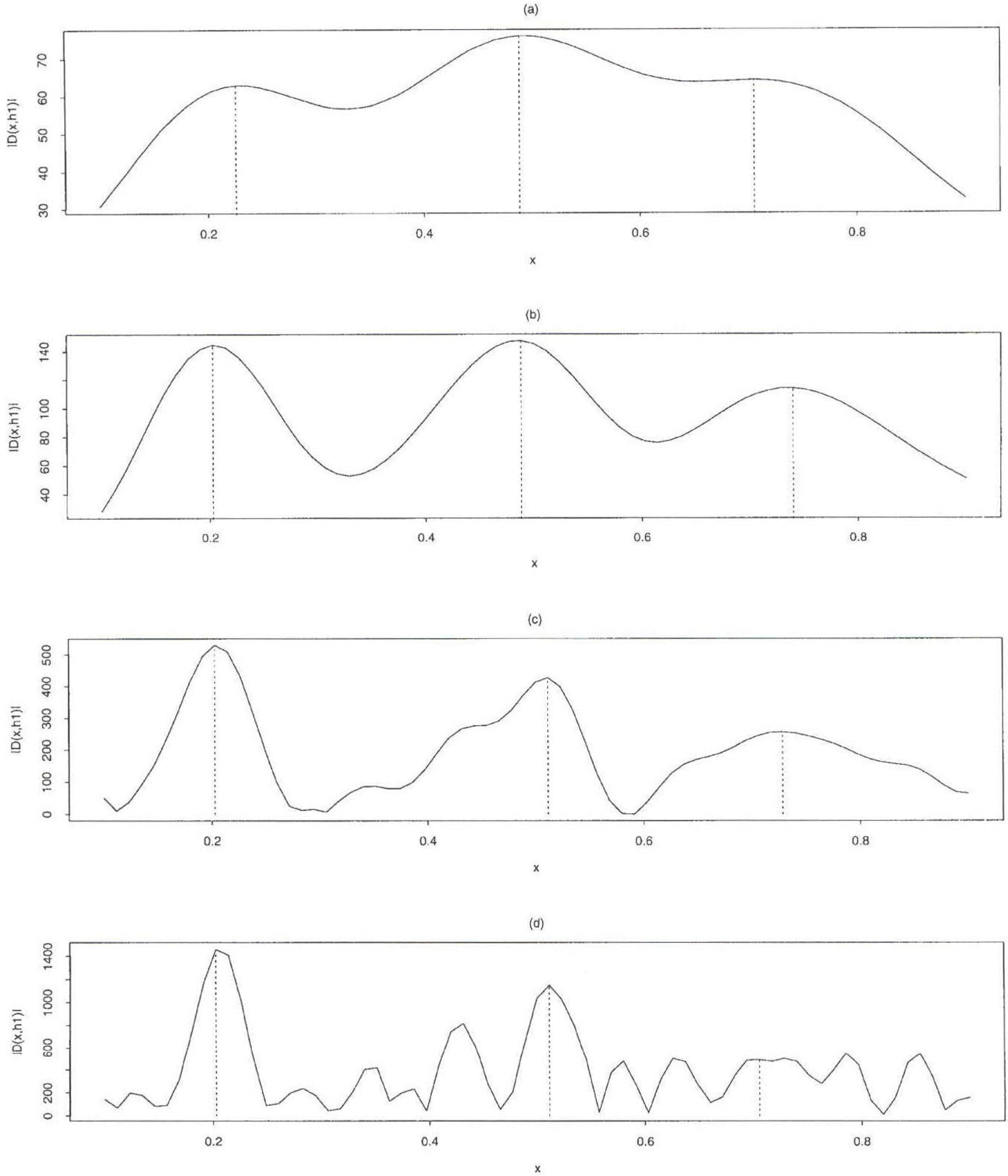

Figure 2. Performance of the Diagnostic Function for the Piecewise Quadratic Function With $n = 50$ and $\sigma^2 = .1$ and Using the Four-Step Algorithm. (a)–(d) Plots of $|D(x, h_{1,i})|$ for $i = 0, 4, 10$, and $i = \tilde{i} = 15$.

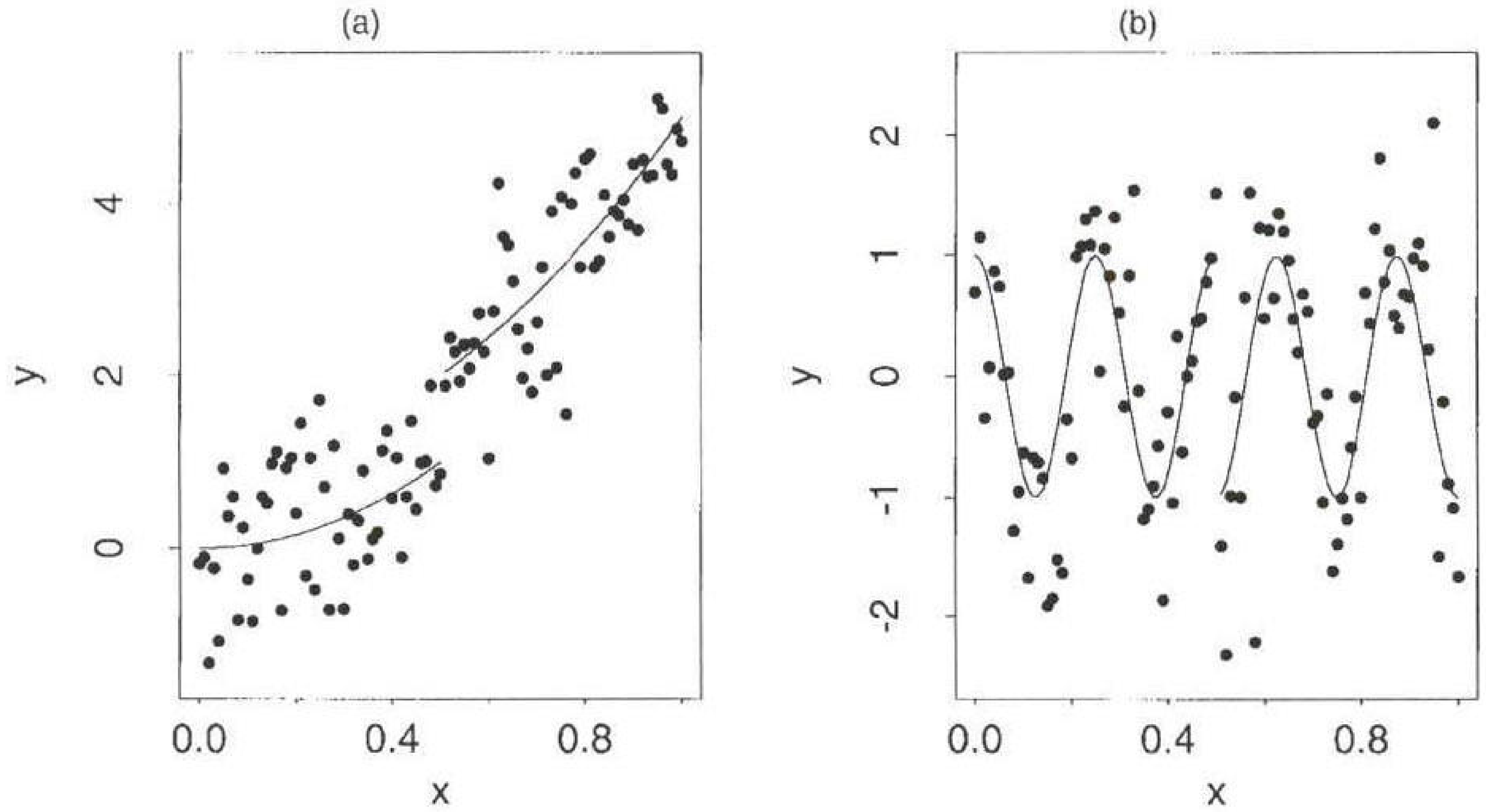

Figure 3. The True Regression Functions (solid curves) With a Typical Simulated Dataset of Size $n = 100$ and Variance $\sigma^2 = .5$ for Regression Functions (a) g_1 and (b) g_2.

We considered a fixed equidistant design, $x_i = i/n$ for $i = 1, \ldots, n$. The errors, ε_i, are Gaussian with variance $\sigma^2 = .1$ or .5. We present simulation results for sample sizes $n = 50$, 100, or 200. Figure 3 presents the true regression functions g_1 and g_2 with typical simulated datasets for sample size $n = 100$ and $\sigma^2 = .5$.

For the diagnostic function in (2), we use a standard Gaussian kernel, 1,000 simulations and $B = 2{,}000$ bootstrap replicates. For each example, we use the CV bandwidths defined in Section 3.2 to choose the smoothing parameters in the first step of the bootstrap algorithm described in Section 3.1.

We first investigate the performance of the bootstrap bandwidth selection method for each of the regression functions. We simulate from model (1) with unknown regression function g_1. We consider the set of potential bandwidths $h_j = .03 + .015j$ for $j = 0, \ldots, 18$. Figure 4(a) depicts a simulated dataset for $n = 50$ and $\sigma^2 = .1$ together with the true function $g_1(\cdot)$. Also presented is a local linear estimator of the regression function, obtained by applying local linear fitting to the left and right of the estimated jump point. Figure 4(b) shows the bootstrap estimates of the probability p_0, associated with the potential bandwidths h_j for $j = 0, \ldots, 18$. For this particular simulation, the bandwidth selected by the bootstrap algorithm is $\hat{h}_{\text{boot}} = .075$, which is the bandwidth with the largest bootstrap estimated probability. With this bandwidth, the final estimator of the discontinuity point is $\hat{x}_0$=.51. To estimate the variability of the bootstrap algorithm for selecting $h = h_1 = h_2$, Figure 4(c) presents a kernel density estimate of the 1,000 bandwidths $\hat{h}_{\text{boot}}$ selected for the 1,000 simulations. This graph shows that for these 1,000 simulations, the bootstrap-selected bandwidth $\hat{h}_{\text{boot}}$

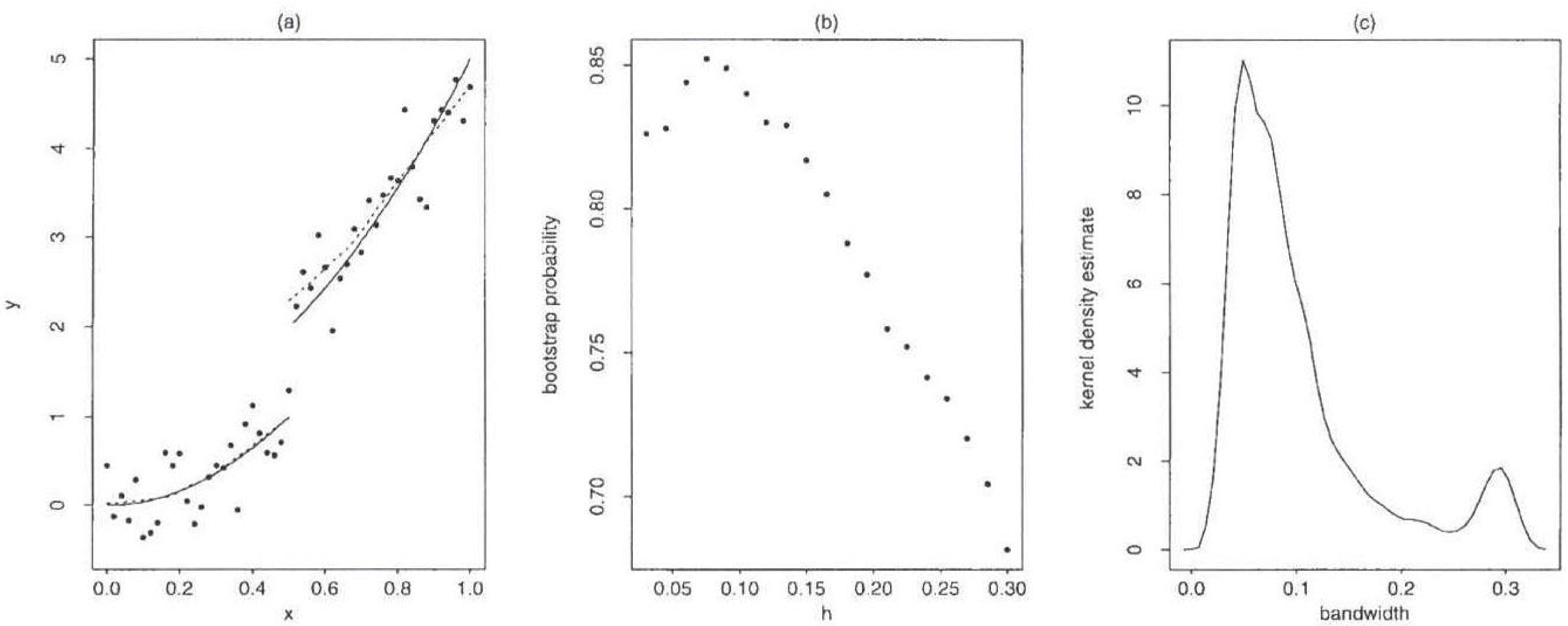

Figure 4. (a) A Simulated Dataset With $n = 50$ and $\sigma^2 = .1$, Together With the True Regression Function g_1 (solid curve) and a Local Linear Estimator, Adapted to the Estimated Changepoint (dashed curve); (b) the Bootstrap Estimate of the Probability p_0 for a Range of Values of the Bandwidth h for That Simulated Dataset; and (c) a Kernel Density Estimate of the Bandwidth $\hat{h}_{boot}$ Selected by the Bootstrap Procedure, Based on 1,000 Simulations.

Table 1. Simulation Results for the Functions g_1 and g_2; Evaluation of the Bootstrap Bandwidth Selection Method

	g_1			g_2		
	$\sigma^2 = .1$		$\sigma^2 = .5$	$\sigma^2 = .1$		$\sigma^2 = .5$
	$n = 50$	$n = 100$	$n = 100$	$n = 100$	$n = 200$	$n = 200$
% in [0, .15[	0	0	.2	.1	.5	1.6
% in [.15, .25[	0	0	.4	1.8	.5	7.1
% in [.25, .35[	0	0	.7	.3	.7	9.2
% in [.35, .45[	.1	.1	2.2	2.5	.7	4.5
% in [.45, .55[	93.9	99.5	78.4	91.2	96.4	58.8
% in [.55, .65[	4.9	.4	9.8	1.6	.4	4.8
% in [.65, .75[	.1	0	3	.9	.4	6.3
% in [.75, .85[	.6	0	3.2	1.6	.4	7.5
% in [.85, 1.0[	.4	0	2.1	0	0	2
Mean of $\hat{x}_0$	.517640	.50685	.533130	.504740	.499390	.491135
SD of $\hat{x}_0$	.035288	.007330	.091399	.06589	.044386	.151566

has a density function concentrated around the value .0488. The little mode around .3 can be explained by the fact that for the function g_1, bandwidths of larger order are also quite appropriate, because the function is overall increasing, and large bandwidths will still allow one to detect the jump discontinuity.

Table 1 summarizes the simulation results for functions g_1 and g_2. Presented are percentages of the estimated values $\hat{x}_0$ falling in the specified intervals. The last two rows list the means and standard deviations of $\hat{x}_0$ across the 1,000 simulations. The function g_2 presents some specific difficulties, because it shows many fluctuations, regions with a steep increase followed by regions with a steep decrease. To identify the "appropriate" local maximum, we use the four-step algorithm as specified in Section 2.5. For the bandwidth h_1 we considered the set of possible values $h_{1,i} = .1 \times .9^i$ for $i = 0, 1, \ldots,$ and we took the set of potential bandwidths h_2 for the least squares step to be $h_{2,j} = .03 + .015j$ for $j = 0, 1, \ldots, 5$. For the function g_2, we used fitting with linear functions instead of constant functions in the least squares step of the estimation procedure. See also Section 5 for some discussion on this issue. Note that the data-driven method performs very well even for these rather small sample sizes.

We now demonstrate the performance of the fully data-driven procedure, including the CV choice of the number of jump points, based on 100 simulations, 1,000 bootstrap replicates, and possible values for k, $k = 0, 1, 2, 3,$ or 4. For each example, we use the CV bandwidth to choose the smoothing parameter that we need to determine g_k^{-i}. Table 2 summarizes the simulation results for the functions g_1 and g_2 and for the smooth function $g_0(x) = x^2$ using different sample sizes and values of σ^2. Presented are the frequencies (out of 100) that the estimated values $\hat{k}$ correspond to the specified values. Clearly, the CV choice of k seems to work nicely.

4.2 Application

As an illustration, we now apply the fully data-driven estimation method to a real dataset. The application concerns 215 average annual temperatures measured in Prague from 1775 to 1989, discussed by Horváth and Kokoszka (1997). This dataset was analyzed by Horváth and Kokoszka (1997) to detect climatic changes occurring over a span of several years or a decade. For this dataset, previous analysis often considered the number of jump points to be either two or three. Antoniadis and Gijbels (2002) also studied these data from 1775 to 1902 and, using a wavelet method, found two jump points occurring at 1787 and 1837.

We apply the fully data-driven procedure to these data and search for changepoints in the data between 1775 and 1989. From the CV criterion, we find $\hat{k} = 3$, and the locations of the jumps are estimated to be 1,786.5, 1,836.5, and 1,942.5. This agrees with previous analysis, in particular with that of Horváth and Kokoszka (1997). Horváth, Kokoszka, and Steinebach (1999) analyzed these data, considering models for dependent observations. They tested for changes in the mean temperature and found that changes occured in the years 1835, 1893, and 1927. Note that the changepoint around 1836 was also an important one in our analysis. Figure 5(a) presents a smooth fit using the method of local linear fitting. Figures 5(b)–(e) depict the fitted curves assuming $k = 1, 2, 3,$ and 4 discontinuities. The large negative slope of the first fitted curve in Figures 5(c)–(e) looks rather strange. A possible explanation is that during this period, some errors appeared in the measurements of the temperatures. Note also that the first discontinuity occurs near the boundary, and hence very little data are used to do the fitting. The plot of the CV function CV(k) is provided in Figure 5(f).

Table 2. Simulation Results for the CV Choice of k

	Function g_0				Function g_1				Function g_2			
$n=$	100		200		100		200		100		200	
k $\quad\sigma^2=$	.1	.5	.1	.5	.1	.5	.1	.5	.1	.5	.1	.5
0	80	72	88	77	1	7	0	1	1	8	3	2
1	7	8	4	7	83	73	87	79	80	71	88	84
2	5	8	4	3	6	6	0	4	7	18	2	3
3	8	11	6	9	8	12	4	15	8	1	4	8
4	0	1	2	4	2	2	9	1	4	2	3	3

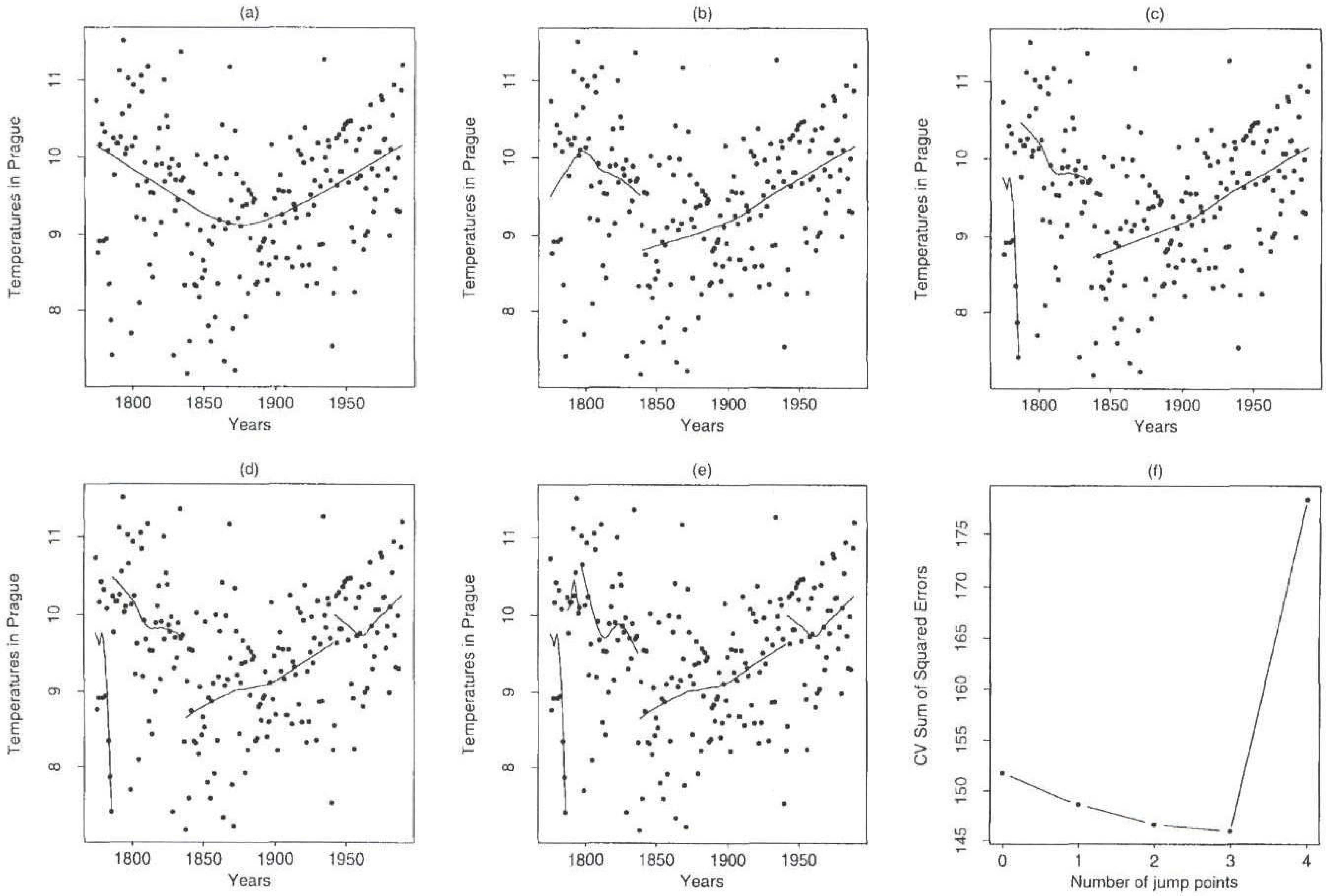

Figure 5. Average Annual Temperatures in Prague With Superimposed Regression Fits. (a) Smooth fit without discontinuity and fit with (b) one discontinuity, (c) two discontinuities, (d) three discontinuities, and (e) four discontinuities; (f) the CV function CV(k).

5. DISCUSSION

5.1 Options for the Two-Step Estimation Method

The two-step estimation method of Gijbels et al. (1999) used here offers several options, including (1) the diagnostic function could be any other consistent estimator of the derivative of the regression function; (2) the least squares step could involve fitting any appropriate parametric function; and (3) the final estimation of the function could be based on any other good nonparametric estimator. In our context of the bandwidth selection problem, we opted for implementing the simplest options: the Nadaraya–Watson estimator as a basis for the diagnostic function, a fitting with constant functions in the least squares step, and local linear estimation for constructing the final estimator. In the final estimation we opted for local linear estimation, because this is known to better handle boundary effects (as opposed to local constant approximation). When using the two-step estimation procedure, one should be aware of the various options and exploit these options according to the context. For example, when dealing with functions for which one suspects significant curvature (such as the cosine function in our simulation study), we recommend using least squares fitting with a polynomial of degree at least 1 to better capture the curvature.

5.2 Alternative Bootstrap-Based Criteria for Selecting the Bandwidth

Our bandwidth selection criterion is based on maximizing the bootstrap estimate of the probability $P(\hat{\imath}_0 - i_0 = 0)$. Here we discuss some alternatives.

A first alternative criterion is to focus on minimizing the mean squared error (MSE) of $\hat{\imath}_0$,

$$\mathrm{MSE}(\hat{\imath}_0) = \sum_{k=-\infty}^{+\infty} k^2 \Pr(\hat{\imath}_0 - i_0 = k) = \{\mathrm{Bias}(\hat{\imath}_0)\}^2 + \mathrm{var}(\hat{\imath}_0).$$

Each probability, $\Pr(\hat{\imath}_0 - i_0 = k)$, can be estimated by $\Pr(\hat{\imath}_0^* - \hat{\imath}_0 = k|\chi)$, and thus this criterion involves taking the bandwidth that minimizes, using (3),

$$\widehat{\mathrm{MSE}}(\hat{\imath}_0) = \sum_{k=-\infty}^{+\infty} k^2 \Pr(\hat{\imath}_0^* - \hat{\imath}_0 = k|\chi).$$

A second alternative criterion is simply to minimize the estimated variance of $\hat{\imath}_0$,

$$\widehat{\mathrm{var}}(\hat{\imath}_0) = \sum_{k=-\infty}^{+\infty} k^2 \Pr(\hat{\imath}_0^* - \hat{\imath}_0 = k|\chi) - \left\{\sum_{k=-\infty}^{+\infty} k \Pr(\hat{\imath}_0^* - \hat{\imath}_0 = k|\chi)\right\}^2.$$

We compared the performances of these alternative criteria with the proposed criterion. The results from 200 simulations for, for example, the quadratic function g_1 (with $n = 50$ and $\sigma^2 = .1$) revealed presentages of estimated values $\hat{x}_0$ falling in the interval $]$.45, .55$[$ to be 97.5%, 87.5%, and 91% for the criterion based on the maximum probability, the minimum MSE, and the minimum variance. For the cosine function g_2 (with $n = 100$ and $\sigma^2 = .1$), these percentages were 90.5%, 91.5%, and 91%. The means and standard deviations of the estimator $\hat{x}_0$ were very close for the three criteria. The data-driven methods based on the different bandwidth selection criteria perform comparably; hence, we opted for the simplest criterion and focused on maximization of $\Pr(\hat{\iota}_0^* - \hat{\iota}_0 = 0|\chi)$.

5.3 Alternatives to the Cross-Validation Bandwidth

For the final estimation of the curve with discontinuities, we use local linear estimation with CV bandwidth selectors. We opted for CV, because it worked well throughout our extensive simulation study. Of course, here the user also has the option to choose his or her favorite data-driven bandwidth selector. To illustrate this point, we implemented more sophisticated bandwidth selectors to replace the CV selectors that we used. We implemented the data-driven constant (global) and variable bandwidth selectors proposed by Fan and Gijbels (1995) in the context of locally weigthed least squares regression. For the 1,000 simulations used to produce Table 1, we found for the quadratic function g_1 (with $n = 50$ and $\sigma^2 = .1$) that the percentages that $\hat{x}_0$ falls in $]$.45, .55$[$ are 93.3% and 93.3% when using the constant and variable bandwidth selector, with means (and standard deviations) of $\hat{x}_0$ of .519 (.0378) and .519 (.0377). These figures should be compared with the corresponding ones in Table 1 (i.e., 93.9%, mean .518, and standard deviation .0353). For the cosine function (with $n = 100$ and $\sigma^2 = .1$), we obtained 92.4%, mean .507 (standard deviation .0588) and 92.4%, mean .509 (standard deviation .0593) for the constant and variable bandwidths. Hence very little difference in these examples results from using the more sophisticated bandwidth selectors. There is a little gain in using the better bandwidth selectors when dealing with (the more difficult) cosine function. Similar conclusions hold true for all other simulations that we carried out (and do not report here). In cases where the functions are less nicely behaved to the right and the left of jump points it might be worthwhile to choose better bandwidth selectors.

5.4 Further Discussion

So far no theoretical results have been established for the data-driven procedure. From the literature, we know that for appropriately chosen fixed bandwidths, the two-step estimator of a jump discontinuity achieves the optimal rate n^{-1}. This would be expected to continue to hold when using the data-driven bandwidths (based on a consistent bootstrap procedure), but theoretical work is needed to establish the rate of convergence of the data-driven estimation method.

Jump discontinuities represent only one type of irregularity that might occur in an otherwise smooth regression function. Other types of irregularities include changes in the derivative functions. The fully data-driven estimation procedure developed in this article can be adapted for detecting jumps in derivative functions. This essentially requires the use of an appropriate diagnostic function and an appropriate family of parametric functions in the least squares step (see Gijbels and Goderniaux 2004b).

One also might be interested in testing whether or not an unknown regression function has jump discontinuities. Such testing problems are rather difficult and are mostly dealt with via asymptotic theory. We have proposed bootstrap testing procedures based on the two-step estimation method that seem to perform very well in comparison with other testing procedures available in the literature. They do not rely on asymptotics and are data-driven, so that the user is not left with a difficult and crucial choice of some smoothing parameter (see Gijbels and Goderniaux 2004a).

ACKNOWLEDGMENTS

Financial support of the "Projet d'Actions de Recherche Concertées," No. 98/03–217 and of the IAP research network No. P5/24 of the Federal Office for Scientific, Technical, and Cultural Affairs, from the Belgian government is gratefully acknowledged. The authors are very grateful to the editor, an associate editor, and two anonymous referees for their excellent review reports, which led to important improvements in the article.

[Received October 2000. Revised July 2002.]

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," unpublished master's thesis, University of Newcastle, Australia.

Bunt, M., Koch, I., and Pope, A. (1995), "Kernel-Based Nonparametric Regression for Discontinuous Functions," Statistics Research Report SRR 012-95, The Australian National University.

——— (1998), "Counting Discontinuities in Nonparametric Regression," Research Report 98-2, University of Newcastle, Australia.

Canny, J. (1986), "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 679–698.

Chu, C. K. (1994), "Estimation of Change-Points in a Nonparametric Regression Function Through Kernel Density Estimation," *Communications in Statistics, Part A—Theory and Methods*, 23, 3037–3062.

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, Hayward, CA: IMS, pp. 130–143.

Fan, J., and Gijbels, I. (1995), "Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation," *Journal of the Royal Statistical Society*, Ser. B, 57, 371–394.

——— (1996), *Local Polynomial Modelling and Its Applications*, New York: Chapman & Hall.

Gijbels, I., and Goderniaux, A.-C. (2004a), "Bootstrap Test for Change Points in Nonparametric Regression," *Journal of Nonparametric Statistics*, to appear.

——— (2004b), "Data-Driven Discontinuity Detection in Derivatives of a Regression Function," *Communications in Statistics—Theory and Methods*, 33, to appear.

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.

——— (2004), "Interval and Band Estimation for Curves With Jumps," *Journal of Applied Probability*, special issue "Stochastic Methods and Their Applications", in honor of Chris Heyde, 41A, 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 (in French).

Grégoire, G., and Hamrouni, Z. (2002), "Change-Point Estimation by Local Linear Smoothing," *Journal of Multivariate Analysis*, 83, 56–83.

Hall, P., and Titterington, D. M. (1992), "Edge Preserving and Peak-Preserving Smoothing," *Technometrics*, 34, 429–440.

Hamrouni, Z. (1999), "Inférence statistique par lissage linéaire local pour une fonction de régression présentant des discontinuités," unpublished doctoral thesis, Université de Joseph Fourier, Grenoble, France (in French).

Horváth, L., and Kokoszka, P. (1997), "Change-Point Detection With Nonparametric Regression," technical report, University of Liverpool.

Horváth, L., Kokoszka, P., and Steinebach, J. (1999), "Testing for Changes in Multivariate Dependent Observations With an Application to Temperature Changes," *Journal of Multivariate Analysis*, 68, 96–119.

Kang, K.-H., Koo, J.-Y., and Park, C.-W. (2000), "Kernel Estimation of Discontinuous Regression Functions," *Statistics & Probability Letters*, 47, 277–285.

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.

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.

Leclerc, Y. C., and Zucker, S. W. (1987), The Local Structure of Image Discontinuities in One Dimension," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9, 341–355.

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 Transactions on 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.

Nadaraya, E. A. (1964), "On Estimating Regression," *Theory of Probability and Applications*, 9, 141–142.

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. P.-J. Laurent, A. Le Mehaute, and L. L. Schumaker, New York: Academic Press, pp. 477–480.

Qiu, P., and Yandell, B. (1998), "A Local Polynomial Jump-Detection Algorithm in Nonparametric Regression," *Technometrics*, 40, 141–152.

Raimondo, M. (1998), "Minimax Estimation of Sharp Change Points," *The Annals of Statistics*, 26, 1379–1397.

Speckman, P. L. (1994), Detection of Change-Points in Nonparametric Regression," unpublished manuscript.

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.

Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385–397.

Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhyā*, Ser. A, 26, 359–372.

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.

——— (1993b), Modification for Boundary Effects and Jump Points in Nonparametric Regression," *Nonparametric Statistics*, 2, 341–354.

——— (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.