On Jump Structure Consideration in One-Dimensional Nonparametric Regression

Technical Report #627

School of Statistics
University of Minnesota

Peihua Qiu

ON JUMP STRUCTURE CONSIDERATION IN ONE-DIMENSIONAL NONPARAMETRIC REGRESSION

Peihua Qiu
School of Statistics
University of Minnesota
270 Vincent Hall
206 Church Street SE
Minneapolis, MN 55455

Abstract

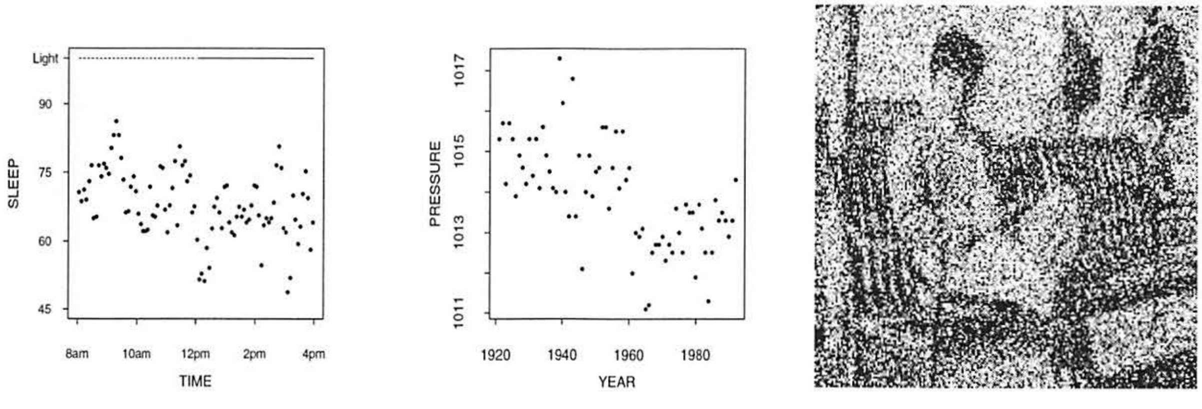
This article introduces some recent local smoothing methods in fitting one dimensional jump regression models. Their strengths and limitations are discussed from several directions including: (1) their ability to get rid of the effect of slope or curvature of the regression curve on jump detection, (2) their ability to diminish the effect of noise, and (3) their ability to detect jumps in both the regression function itself and its derivatives.

Key Words: Difference kernel estimator; Jump detection; Jump-preserving curve fitting; Least squares line; Local polynomial kernel smoothing.

1 Introduction

Regression analysis provides a tool to build functional relationship between response variables and independent variables. For a long time, the nonparametric regression function is assumed to be continuous. This assumption is challenged by some statisticians recently, supported in part by many reallife applications in which jump regression models appear to be more appropriate.

Figure 1.1(a) shows a rat sleep dataset. Several psychiatrists at The University of Wisconsin-Madison were interested in statistical modeling of the percentage of time in each five-minute interval that a Lewis rat was in sleep (Qiu *et al.* 1997). The rat was exposed to light before 12:00pm and then the light was turned off. Rats are noctunal animals. They are expected to have more sleep under light. So 12:00pm should be a jump point of the response variable although it is not obvious



(a) (b) (c)

Figure 1.1: (a) The percentage of time a Lewis rat is in sleep in each five-minute interval. The dotted line at top indicates period of light. The solid line represents dark period. (b) The December sea-level pressures during 1921-1992 in Bombay, India. (c) A noisy picture.

in the plot. Figure 1.1(b) gives another data set about the December sea-level pressures during 1921-1992 observed by the Bombay weather station in India. Meteorologists noticed a possible jump around year 1960 (Shea *et al.* 1994) and it was confirmed by us (Qiu and Yandell 1998). Figure 1.1(c) presents a noisy picture. Its intensity function (representing the brightness at each pixel) has step edges at the outlines of the objects. Since much of the information in a picture is conveyed by the edges and our eye-brain system has evolved to extract edges by preprocessing that begins right at the retina (Chapter 5, Bracewell 1995), edge detection and edge-preserving image reconstruction are important research topics in image processing. This third example is related to two dimensional jump surface fitting which is beyond the scope of this article.

Jump regression model fitting is currently under rapid development. In one dimensional (1-D) case, most jump-preserving curve fitting methods in the literature detect possible jump points first and then fit the regression curve as usual in design subintervals separated by the detected jump points. It is therefore essential to detect jumps by various criteria in fitting jump regression models. Suppose that the regression model concerned is

$$Y_i = f(x_i) + \epsilon_i,\ i = 1, 2, \cdots, n, \tag{1.1}$$

where $0 \leq x_1 < x_2 < \cdots < x_n \leq 1$ are design points, and $\{\epsilon_i\}$ are i.i.d. random errors with mean zero and unknown variance σ^2. The regression function $f(x)$ has jumps at positions $\{s_i, i =$

$1, 2, \cdots, p\}$ and it can be expressed by

$$f(x) = g(x) + \sum_{i=1}^{p} d_i I_{[s_i, s_{i+1})}(x), \tag{1.2}$$

where $g(x)$ is a continuous function, p is the number of jumps, $\{|d_i - d_{i-1}|, i = 1, 2, \cdots, p\}$ are jump magnitudes, $d_0 = 0$ and $s_{p+1} = 1$.

The kernel-type methods (e.g., Müller 1992; Qiu 1994; Qiu *et al.* 1991; Wu and Chu 1993a, b) detect jumps based on criterion $JDC_{DKE}(x)$ which is defined by

$$JDC_{DKE}(x) := \frac{1}{nh_n} \sum_{i=1}^{n} Y_i \left[K_2\left(\frac{x_i - x}{h_n}\right) - K_1\left(\frac{x_i - x}{h_n}\right) \right], \text{ for } h_n \leq x \leq 1 - h_n, \tag{1.3}$$

where $h_n > 0$ is a bandwidth parameter, $K_1(x)$ and $K_2(x)$ are two density kernel functions (non-negative functions with unit integrations) satisfying (i) $K_1(x) \equiv K_2(-x)$ and (ii) $K_2(x)$ has support $[0, 1]$. Intuitively, $JDC_{DKE}(x)$ is a difference of two weighted averages of the observations in $[x - h_n, x)$ and $(x, x + h_n]$, respectively, as illustrated by Figure 1.1(a). If x is a continuous point ($x = x1$ in the plot), then $|JDC_{DKE}(x)|$ is relatively small. Otherwise, it is approximately equal to the jump magnitude (the case when $x = x2$ in the plot). Qiu *et al.* (1991) called these kernel methods *the difference kernel estimation (DKE)* methods.

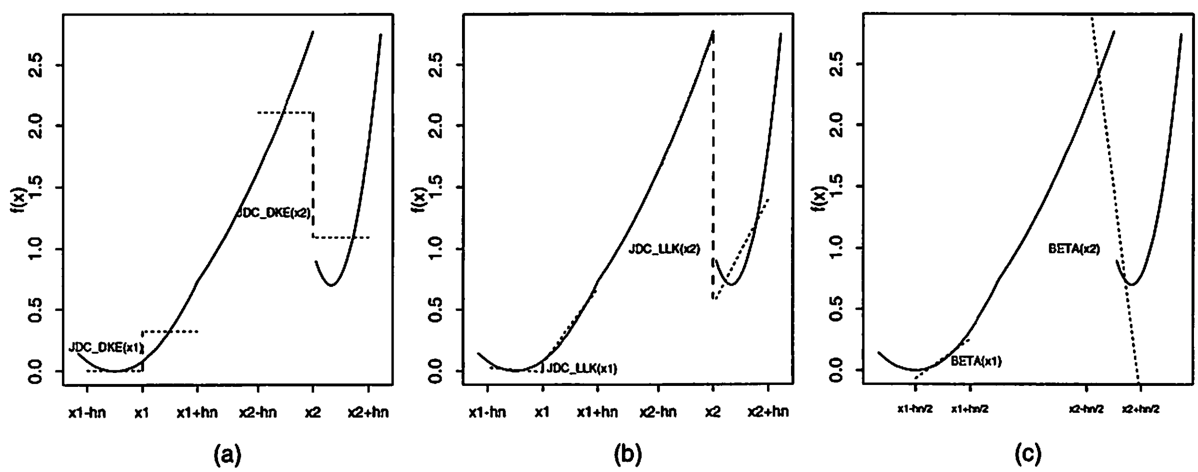


Figure 1.2: (a) Jump detection criterion of the DKE method. (b) Jump detection criterion of the LLK method. (c) Fitted local LS lines at $x1$ and $x2$ (the jump detection criterion of the LLS method is base on the slopes of the fitted local LS lines). The regression function (plotted by the solid curves) is continuous at $x1$ and it has a jump at $x2$.

Recently the local polynomial kernel smoothing method has been demonstrated to have some preferable properties in fitting regression curves (Fan and Gijbels 1996; Hastie and Loader 1993).

Loader (1996) showed that the detected jumps had some favorable properties too if the kernel estimators were replaced by the local linear kernel (LLK) estimators in the construction of $JDC_{DKE}(x)$. The resulting jump detection criterion is defined by

$$JDC_{LLK}(x) := \frac{\sum_{i=1}^{n} Y_i \left[K_2(\frac{x_i-x}{h_n})\,(r_2-(x_i-x)r_1) - K_1(\frac{x_i-x}{h_n})\,(r_2+(x_i-x)r_1)\right]}{r_0r_2 - r_1^2}, \tag{1.4}$$

where $h_n \leq x \leq 1-h_n$, and $r_j = \sum_{i=1}^{n} K_2(\frac{x_i-x}{h_n})(x_i-x)^j$ for $j = 0, 1$ and 2. The construction of $JDC_{LLK}(x)$ is also demonstrated by Figure 1.2(b).

Qiu and Yandell (1998) suggested detecting jumps based on estimated coefficients of the local least squares (LLS) estimation. If we are interested in detecting jumps in the m-th derivative $f^{(m)}(x)$, then a local polynomial function of order $(m+1)$ is fitted by the LS procedure in neighborhood $(x-h_n/2, x+h_n/2)$. The fitted polynomial can be expressed by

$$\hat{Y}^{(i)}(t) = \hat{\beta}_0(x) + \hat{\beta}_1(x)t + \cdots + \hat{\beta}_{m+1}(x)t^{m+1}, \text{ for } t \in (x-h_n/2, x+h_n/2),\ \ x \in [h_n/2, 1-h_n/2].$$

Then $\{\hat{\beta}_{m+1}(x)\}$ could be used to detect jumps in $f^{(m)}(x)$ based on the intuitivity that $\hat{\beta}_{m+1}(x)$ approximates to $f^{(m)}(x)$ if x is a continuous point and it has an abrupt change around x otherwise (Figure 1.2(c) demonstrates a case when $m=0$). In order to exclude the effect of the continuous part of $f(x)$ (which is $g(x)$ in (1.2)) on jump detection, Qiu and Yandell (1998) suggested applying a difference operator on $\{\hat{\beta}_{m+1}(x)\}$. When $m=0$, the resulting jump detection criterion is

$$JDC_{LLS}(x) := \begin{cases} \hat{\beta}_1(x) - \hat{\beta}_1(x-h_n/2), & \text{if } |\hat{\beta}_1(x) - \hat{\beta}_1(x-h_n/2)| \leq |\hat{\beta}_1(x) - \hat{\beta}_1(x+h_n/2)| \\ \hat{\beta}_1(x) - \hat{\beta}_1(x+h_n/2), & \text{if } |\hat{\beta}_1(x) - \hat{\beta}_1(x-h_n/2)| > |\hat{\beta}_1(x) - \hat{\beta}_1(x+h_n/2)| \end{cases}. \tag{1.5}$$

Besides the three jump detection methods mentioned above, there exist many other jump detection and jump-preserving curve fitting methods in the literature which include the "split linear smoother" algorithm (McDonald and Owen 1986) and its simplified version (Hall and Titterington 1992); the semiparametric method (Eubank and Speckman 1994); the wavelet transformation method (Wang 1995); and the smoothing spline method (Koo 1997, Shiau *et al.* 1986).

In this article, we discuss the strengths and limitations of several jump detectors. This effort should be helpful for users in choosing an appropriate method for a specific application problem. To keep the presentation simple, our discussion will mainly focus on three jump detectors: DKE, LLK and LLS, which might be good representatives of local smoothing methods for jump detection.

The three jump detectors are investigated from several directions including (1) their ability to get rid of the effect of the regression function derivatives on jump detection in the regression function itself (Section 2); (2) their ability to diminish the effect of noise (Section 3); and (3) their ability to detect jumps in both the regression function itself and its derivatives (Section 4). Most jump detection methods in the literature assume that the number of jumps is known, which is hard to be satisfied in many applications. In Section 5, we introduce some jump detection procedures which could work well without this assumption. Finally, several remarks conclude the article in Section 6.

2 Effect of $f'(x)$ on Jump Detection in $f(x)$

The impact of $f'(x)$ on jump detection in $f(x)$ has not been well discussed yet in the literature. This impact might be negligible in large sample theory. But it could play an important role in finite sample situations.

Theorem 2.1 In model (1.1), suppose that $f(x)$ has continuous second order right and left derivatives in a neighborhood of $x \in (0,1)$. Then

$$E(JDC_{DKE}(x)) = [f_+(x) - f_-(x)] + [f'_+(x) + f'_-(x)]\frac{a_1}{a_0}h_n + [f''_+(x) - f''_-(x)]\frac{a_2}{2a_0}h_n^2 + o(h_n^2) \quad (2.1)$$

$$E(JDC_{LLK}(x)) = [f_+(x) - f_-(x)] + [f''_+(x) - f''_-(x)]\frac{a_2^2 - a_1a_3}{2(a_0a_2 - a_1^2)}h_n^2 + o(h_n^2) \quad (2.2)$$

$$E(\hat{\beta}_1(x)) = [f_+(x) - f_-(x)]\frac{3}{2}h_n^{-1} + \frac{f'_+(x) + f'_-(x)}{2} + [f''_+(x) - f''_-(x)]\frac{3}{16}h_n + o(h_n) \quad (2.3)$$

where $a_i = \int_0^1 x^i K_2(x)\,dx$ for $i = 0, 1, 2$ and 3.

In Theorem 2.1, if we use the Epanechnikov kernel function $K_2(x) = \frac{3}{2}(1 - x^2)I_{[0,1]}(x)$ (which is optimal in minimizing MSE of the conventional LLK estimator, see e.g., Section 3.2.6, Fan and Gijbels 1996), then

$$E(JDC_{DKE}(x)) = [f_+(x) - f_-(x)] + \frac{3}{8}[f'_+(x) + f'_-(x)]h_n + \frac{1}{10}[f''_+(x) - f''_-(x)]h_n^2 + o(h_n^2)$$

and

$$E(JDC_{LLK}(x)) = [f_+(x) - f_-(x)] + \frac{11}{190}[f''_+(x) - f''_-(x)]h_n^2 + o(h_n^2).$$

The above equations indicate that the first order derivatives affect the mean of $JDC_{DKE}(x)$ in order of h_n. But they do not affect the mean of $JDC_{LLK}(x)$. The impact of the second or higher order derivatives is relatively negligible and will not be discussed here. It is possible to modify $JDC_{DKE}(x)$ such that the resulting criterion can get rid of the effect of the first order derivatives to some degree. For example, $JDC_{DKE}(x)-[\hat{f}'_+(x)+\hat{f}'_-(x)]\frac{a_1}{a_0}h_n$ is one possibility, where $\hat{f}'_+(x)$ and $\hat{f}'_-(x)$ are the conventional consistent estimators of $f'_+(x)$ and $f'_-(x)$, respectively (see e.g., Section 2.3.2, Fan and Gijbels 1996). But we may not want to pursue this idea since the LLK procedure can accomplish this automatically as indicated by (2.2) and it has some other good properties as well (see the related discussion in Hastie and Loader (1993)).

Both the DKE and LLK procedures detect jumps based on consistent estimators of the jump magnitude. The LLS criterion makes use of the property of $\hat{\beta}_1(x)$ that it tends to infinity if x is a jump point and approximates to the slope of the regression curve otherwise. To diminish the effect of the first order derivative, it is constructed by applying a difference operator defined by (1.5) on $\{\hat{\beta}_1(x)\}$. From (2.3), it is not hard to check that when x is a jump point,

$$\begin{aligned} E(JDC_{LLS}(x)) &= [f_+(x)-f_-(x)]\frac{3}{2}h_n^{-1} + P\left[\frac{f'_+(x)+f'_-(x)}{2} - \frac{f'_+(x-h_n/2)+f'_-(x-h_n/2)}{2}\right] \\ &\quad + (1-P)\left[\frac{f'_+(x)+f'_-(x)}{2} - \frac{f'_+(x+h_n/2)+f'_-(x+h_n/2)}{2}\right] + O(h_n), \qquad (2.4) \end{aligned}$$

where $P = Pr(|\hat{\beta}_1(x)-\hat{\beta}_1(x-h_n/2)| \leq |\hat{\beta}_1(x)-\hat{\beta}_1(x+h_n/2)|)$. If $f(x)$ is continuous in a neighborhood of x, (2.4) is still true except that the first term on the right hand side of the equation disappears. As indicated by (2.4), $E(JDC_{LLS}(x))$ does not depend on the first order derivative if the derivative is continuous in the design space. When the first order derivative has jumps itself, it does affect $E(JDC_{LLS}(x))$.

We would like to point out the difference between the ways the first order derivative affecting the DKE and LLS procedures. In the DKE procedure, the derivative affects $E(JDC_{DKE}(x))$ through $f'_+(x)+f'_-(x)$. It could cause false jump detection even if $f'(x)$ is continuous in the entire design space but large at some places. For $E(JDC_{LLS}(x))$, the derivative plays the role through its jump magnitudes. In other words, its impact on jump detection should be taken into account only when its jump magnitudes are large. Furthermore the detected false jumps because of this reason are most probably the jumps of the first order derivative itself, which will be discussed in some detail in Section 4.

<u>**Example 2.1**</u> Consider a regression function $f(x) = cx - I_{(.5,1]}(x)$. It has a jump at $x = .5$ with magnitude 1 and slope c at the continuous points. If $n = 256$, $h_n = 21/256$, $c = 10$, the

Epanechnikov kernel function is used, and the noise is ignored, then the jump detection criteria $JDC_{LLS}(x)$, $JDC_{LLK}(x)$ and $JDC_{DKE}(x)$ are presented in Figures 2.1(b), 2.1(c) and 2.1(d), respectively. The regression function itself is shown in Figure 2.1(a). As indicated by the plots, the value of c affects $JDC_{DKE}(x)$ dramatically. The values of $JDC_{DKE}(x)$ at most continuous points are even larger than its value at the real jump point. Both criteria $JDC_{LLS}(x)$ and $JDC_{LLK}(x)$, however, depend little on c. Their values are large around the true jump point and zero otherwise.

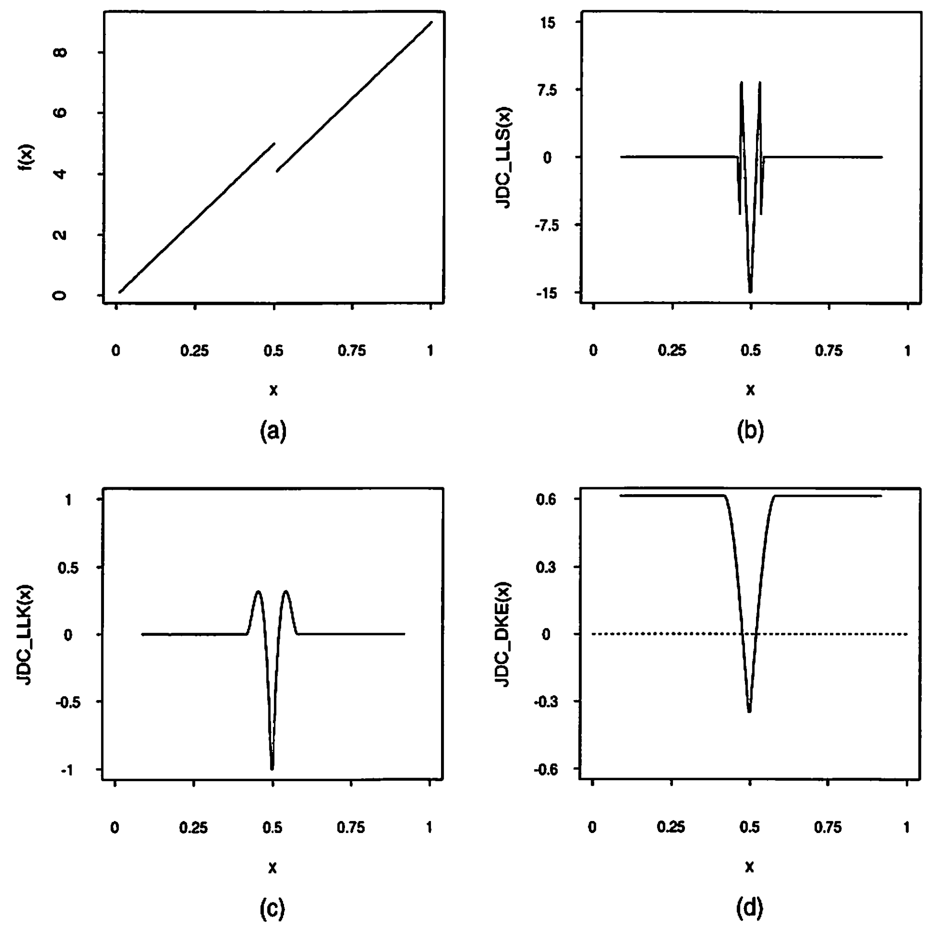


Figure 2.1: (a) The regression function $f(x) = 10x - I_{(.5,1]}(x)$; (b) $JDC_{LLS}(x)$; (c) $JDC_{LLK}(x)$; (d) $JDC_{DKE}(x)$. The dotted line in plot (d) indicates $y = 0$.

We next add i.i.d. noise with $\sigma = .25$ to the data and let c change from 1 to 10 as well. It is assumed that the number of jumps is known beforehand, which is 1 in this case. The detected jumps are defined by the maximizers of $|JDC_{DKE}(x)|$, $|JDC_{LLK}(x)|$, and $|JDC_{LLS}(x)|$, respectively. (The case when the number of jumps is unknown will be discussed in Section 5.) For each c value, the simulation is repeated 1000 times. A jump detection is flaged as "correct" if the distance between the detected jump point and the true jump point is less than h_n. The number of correct jump detections by each method is shown in Figure 2.2. We can see that the LLK and LLS methods perform stably with respect to c. Results of the DKE procedure get worse when c increases.

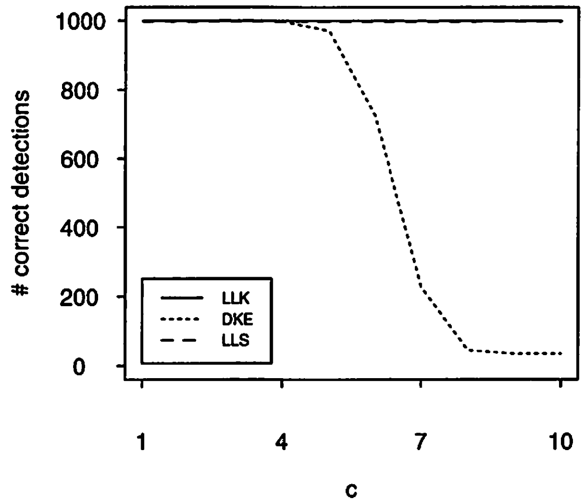


Figure 2.2: Numbers of correct jump detections of the three jump detection procedures out of 1000 replications when the slope of the regression function changes from 1 to 10.

3 Effect of Noise on Jump Detection

In the previous section, we investigated the three jump detection criteria by analysing their means. It could be concluded that the DKE procedure is most sensitive to the slope of the regression curve among the three methods. In this section, the three criteria are studied by their variances. We will explain that the DKE procedure is most capable to smooth away the noise.

Theorem 3.1 Under the conditions stated in Theorem 2.1, we have

$$Var(JDC_{DKE}(x)) = \frac{2\sigma^2 \int_0^1 K^2(x)\, dx}{a_0^2 n h_n} \tag{3.1}$$

$$Var(JDC_{LLK}(x)) = \frac{2\sigma^2 \int_0^1 K^2(x)(a_2 - a_1 x)^2\, dx}{(a_0 a_2 - a_1^2)^2 n h_n} \tag{3.2}$$

$$Var(JDC_{LLS}(x)) \leq \frac{18\sigma^2}{n h_n^3} \tag{3.3}$$

where $\{a_i\}_{i=0}^2$ are defined in Theorem 2.1.

If the Epanechnikov kernel function is used in the DKE and LLK procedures, then

$$Var(JDC_{DKE}(x)) = \frac{2.4\sigma^2}{n h_n}, \quad Var(JDC_{LLK}(x)) = \frac{4.49\sigma^2}{n h_n}.$$

The variance of $JDC_{LLK}(x)$ is about 2 times the variance of $JDC_{DKE}(x)$. Therefore in regions where $f(x)$ is quite flat such that its derivatives do not affect the jump detection much, the DKE procedure could outperform the LLK method by detecting some jumps of small magnitudes, which will be further explained by Example 3.1 below.

The variance of $JDC_{LLS}(x)$ is of higher order than the variances of the other two criteria. If we look at their standard deviations and means (which are expressed by (2.1)-(2.4)) simultaneously, then we can notice that both the mean and standard deviation of $JDC_{LLS}(x)$ are of h_n^{-1} higher order than the means and standard deviations of the other two criteria. In other words, the LLS criterion makes its mean tend to infinity at the jump points by increasing its standard deviation. Although the ratio of its mean and standard deviation is of the same order as the ratios of the other two methods, the difference between its mean and standard deviation tends to infinity at the jump points while the corresponding differences of the other two methods tend to the jump magnitudes. This amplification property might be helpful to visualize the jump structure and to select the threshold values (which will be formally defined in Section 5) of the jump detection criteria as well.

Example 3.1 For the regression function used in Example 2.1, let us assume that $c = 1$. As indicated by Figure 2.2, the impact of slope on jump detection is limited in such case for all three jump detectors. When $\sigma = .25$, the three jump detection criteria are shown in Figure 3.1 along with the true regression function.

We then let σ change from .1 to 1. The numbers of correct jump detections of the three methods out of 1000 replications are presented in Figure 3.2. As indicated by the plot, all three methods detect the jump well when σ is small. When σ gets larger, it becomes more obvious that the DKE method outperforms the other two methods.

4 Detect Jumps in Both $f(x)$ and $f'(x)$

The criteria $JDC_{DKE}(x)$ and $JDC_{LLK}(x)$ defined by (1.3) and (1.4) are designed to detect jumps in $f(x)$. If jump detection in $f'(x)$ is also our concern, then estimators of $f_-(x)$ and $f_+(x)$ in the definitions of $JDC_{DKE}(x)$ and $JDC_{LLK}(x)$ need to be replaced by the corresponding estimators of $f'_-(x)$ and $f'_+(x)$, respectively. The local slope estimators $\{\hat{\beta}_1^{(i)}\}$ on which $JDC_{DKE}(x)$ is based,

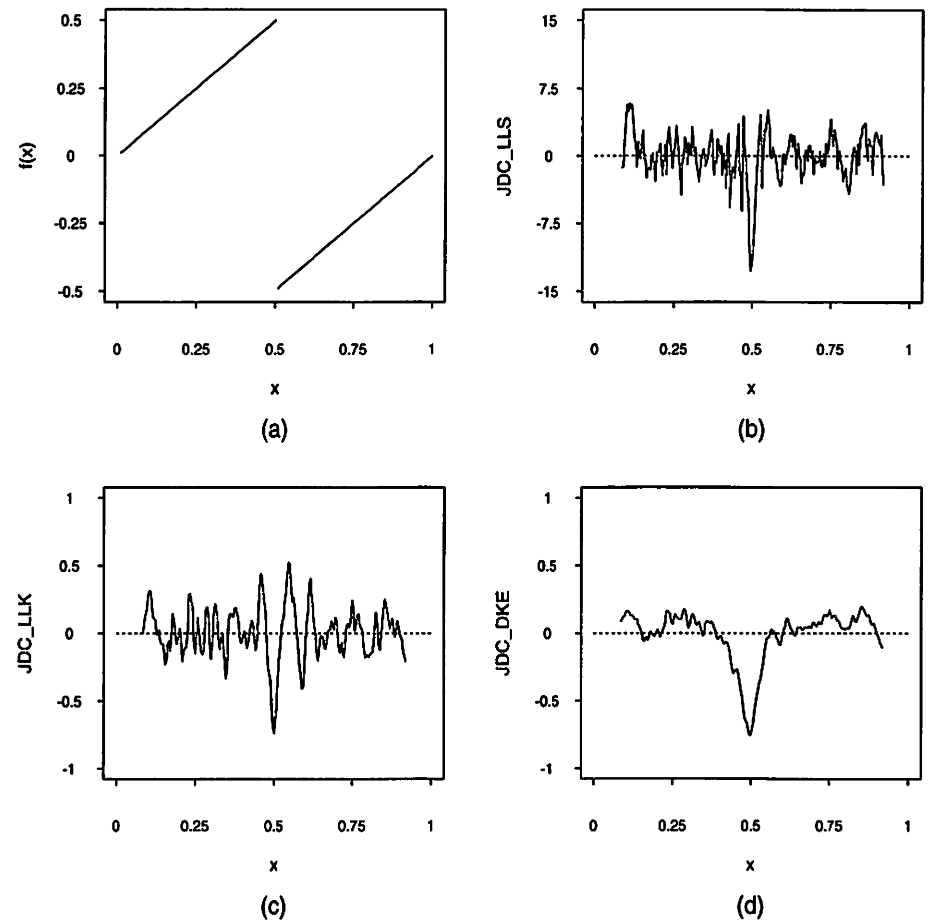


Figure 3.1: (a) The regression function $f(x) = x - I_{(.5,1]}(x)$; (b) $JDC_{LLS}(x)$; (c) $JDC_{LLK}(x)$; (d) $JDC_{DKE}(x)$. The error standard deviation $\sigma = .25$. The dotted lines in plots (b)-(d) indicate $y = 0$.

however, could be used to detect jumps in both $f(x)$ and $f'(x)$, which will be demonstrated in some detail in this section.

As we pointed out in Section 2, after the difference operator (1.5) being applied to $\{\hat{\beta}_1^{(i)}\}$, the resulting criterion $JDC_{LLS}(x)$ could detect jumps in $f(x)$ and some jumps with large magnitudes in $f'(x)$ as well. If we apply the following difference operator to $\{\hat{\beta}_1^{(i)}\}$,

$$JDC^*_{LLS}(x) = \hat{\beta}_1(x + h_n/2) - \hat{\beta}_1(x - h_n/2), \tag{4.1}$$

then $JDC^*_{LLS}(x)$ could be used to detect jumps in $f'(x)$ because $\hat{\beta}_1(x + h_n/2)$ and $\hat{\beta}_1(x - h_n/2)$ are good estimators of $f'_+(x)$ and $f'_-(x)$, respectively.

In the local linear kernel regression, we know that slope of the fitted local line in a neighborhood of a given point could also be used as an estimator of the value of the first order derivative at that

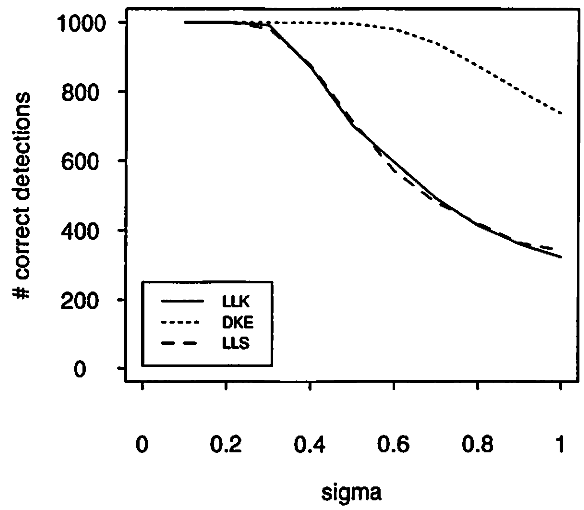


Figure 3.2: Numbers of correct jump detections out of 1000 replications when σ changes from .1 to 1.

point (Chapter 3, Fan and Gijbels 1996). It does not need much extra computation to calculate this slope (see related formulas given by Fan and Gijbels 1996). By this idea, the slopes of the fitted one-sided local lines, which are obtained at the time when we construct $JDC_{LLK}(x)$ by (1.4), could be used as estimators of $f'_-(x)$ and $f'_+(x)$. Then a criterion to detect jumps in $f'(x)$ can be constructed as follows:

$$JDC^*_{LLK}(x) = \frac{\sum_{i=1}^n Y_i \left[K_2(\frac{x_i - x}{h_n})\left(-r_1 + (x_i - x)r_0\right) - K_1(\frac{x_i - x}{h_n})\left(r_1 + (x_i - x)r_0\right)\right]}{r_0 r_2 - r_1^2}, \qquad (4.2)$$

for $h_n \leq x \leq 1 - h_n$. If the kernel functions used in (4.2) are the uniform functions, namely $K_2(x) = K_1(-x) = I_{[0,1]}(x)$, then the slopes of the fitted LLK lines are the same as the slopes of the fitted LLS lines. Consequently, $JDC^*_{LLK}(x)$ and $JDC^*_{LLS}(x)$ are equivalent to each other.

Example 4.1 Consider a regression function $f(x) = 3x$ when $x \in [0, .5)$; and $f(x) = c(x - .5) + 1.5$ when $x \in [.5, 1]$. Then $f'(x)$ has a jump at $x = .5$ with magnitude $c - 3$. If the noise is ignored (by setting $\sigma = 0$) and $c = 4$ (the jump magnitude is 1), then the criteria $JDC_{DKE}(x)$, $JDC_{LLK}(x)$, $JDC^*_{LLK}(x)$, $JDC_{LLS}(x)$, and $JDC^*_{LLS}(x)$ are shown in Figure 4.1 along with the true regression function.

Next we let the jump magnitude of $f'(x)$ vary from 0 to 20 by changing c from 3 to 23. We also add noise with $\sigma = .25$ to the data. The numbers of correct jump detections out of 1000 replications by the related criteria are shown in Figure 4.2. As indicated by the plot, the criteria $JDC_{DKE}(x)$ and $JDC_{LLK}(x)$ could hardly detect the jump in $f'(x)$; $JDC_{LLS}(x)$ detects the jump in $f'(x)$ well only when the jump magnitude is large; the criteria $JDC^*_{LLK}(x)$ and $JDC^*_{LLS}(x)$ perform much

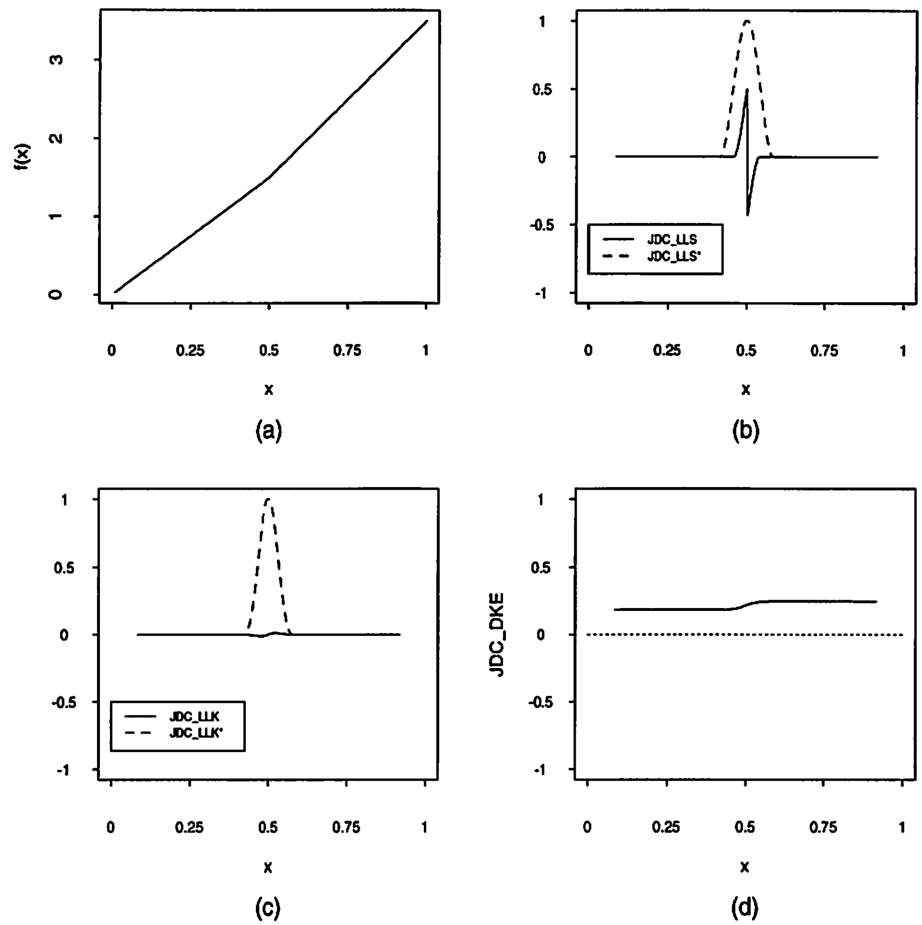


Figure 4.1: (a) The true regression function $f(x)$ ($f'(x)$ has a jump with magnitude 1 at $x = .5$); (b) $JDC_{LLS}(x)$ and $JDC^*_{LLS}(x)$; (c) $JDC_{LLK}(x)$ and $JDC^*_{LLK}(x)$; (d) $JDC_{DKE}(x)$. The dotted line in plot (d) indicates $y = 0$.

better than the other three.

The LLS method is based on the fitted LLS slopes $\{\hat{\beta}_1(x)\}$, which can be easily computed by most statistical softwares. Hence its computation is simple. Users do not need any nonparametric regression knowledgement to apply this procedure. The LLS slope $\hat{\beta}_1(x)$ is a good estimator of $f'(x)$. So the LLS procedure detects jumps in $f(x)$ based on local estimation of $f'(x)$. This idea is intuitively appealing since estimators of $f^{(m+1)}(x)$ could often be used in detecting jumps in $f^{(m)}(x)$ for any non-negative integer m (Qiu and Yandell 1998). In this section, we have shown that jump detectors can be obtained for both $f(x)$ and $f'(x)$ by applying different difference operators to $\{\hat{\beta}_1(x)\}$. The LLS procedure is therefore flexible in detecting jumps in $f(x)$ and its derivatives.

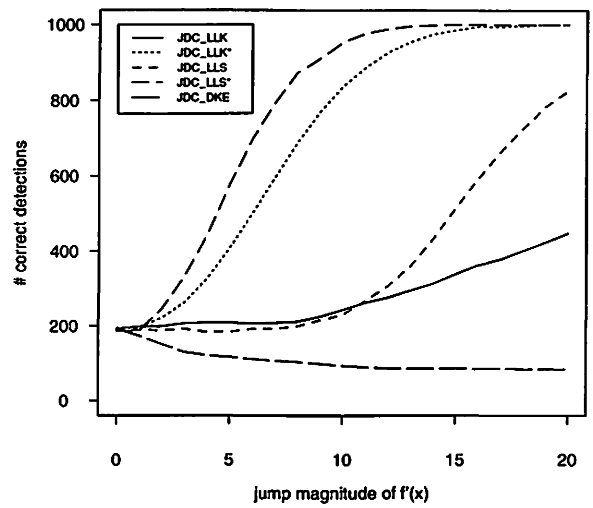


Figure 4.2: The numbers of correct jump detections out of 1000 replications when the jump magnitude of $f'(x)$ changes from 0 to 20.

5 When the Number of Jumps Is Unknown

In the previous sections, we assume that the number of jumps is known beforehand. In many applications, however, this kind of prior information is not available. In the Bombay sea-level pressure example (c.f. Figure 1.1(b)), there is no convincing scientific evidence that a jump exists around year 1960, as indicated by the plot. In the rat sleep example (c.f. Figure 1.1(a)), psychiatrists expect a jump at 12:00pm. But it is not obvious in the plot. Therefore visual perception is not always dependable to know the jump structure of the related model.

The nonparametric regression analysis can be regarded as a generalization of the linear regression analysis because the former can be applied to fitting both linear and nonparametric regression models, although we should use the linear regression analysis if we know beforehand that the true regression model is indeed linear. Similarly it might not be appropriate to say that the jump regression methods generalize the conventional nonparametric regression methods if the former can only deal with the situation in which the number of jumps is known. To handle more applications and to give the word "generalization" some real meaning, it is important to suggest some jump regression methods which do not require any prior information about the number of jumps.

Wu and Chu (1993a) proposed an algorithm to detect jumps when the number of jumps was unknown. Their method was based on $JDC_{DKE}(x)$ and function $S(x)$ defined by:

$$S(x) = \hat{m}_3(x) - \hat{m}_4(x), \tag{5.1}$$

where $\hat{m}_3(x)$ and $\hat{m}_4(x)$ were two kernel estimators of $f(x)$ defined similarly to the two kernel estimators used in (1.3) except that a new bandwidth g_n and two new kernel functions $K_3(x)$ and $K_4(x)$ were used in defining $\hat{m}_3(x)$ and $\hat{m}_4(x)$. They also provided some guidelines for choosing h_n, g_n, and the kernel functions.

Their procedure consisted of several steps. For $j \geq 0$, let $\hat{d}^*_{j+1}$ be the supremum of $|S(x)|$ in $[\delta, 1-\delta] - \bigcup_{k=1}^{j}[\hat{s}_k - 2h_n, \hat{s}_k + 2h_n]$, where $\delta > 0$ was an arbitrarily small number and $\hat{s}_k$ was the k-th maximizer of $JDC_{DKE}(x)$ (see Wu and Chu (1993a) for definition). They then derived an asymptotic distribution for $\hat{d}^*_{j+1}$ under the assumption that $p \leq j$ where p was the true number of jumps. With this distribution, a series of hypothesis tests were performed for $H_0 : p = j$ *vs* $H_a : p > j$, for $j \geq 0$, until an acceptance, from which $\hat{p}$ (an estimator of p) could be defined. Then $\hat{p}$ maximizers $\{\hat{s}_j\}_{j=1}^{\hat{p}}$ of $JDC_{DKE}(x)$ were defined as estimators of the jump positions. Finally, they used rescaled $\{S(\hat{s}_j)\}_{j=1}^{\hat{p}}$ to estimate the jump magnitudes.

Qiu (1994) suggested an alternative procedure by using a threshold value and a modification procedure to estimate the number of jumps and the jump positions. If x is not a jump point, then it is not hard to check that $JDC_{DKE}(x)$ is asymptotically normally distributed with mean 0 and variance $2\sigma^2 \int_0^1 K^2(x)\,dx/(a_0^2 n h_n)$. A natural threshold value for $JDC_{DKE}(x)$ is then

$$u_n = \frac{\hat{\sigma} Z_{\alpha_n/2}}{a_0}\sqrt{\frac{2\int_0^1 K_2^2(x)\,dx}{nh_n}}, \tag{5.2}$$

where $Z_{\alpha_n/2}$ is the $1 - \alpha_n/2$ quantile of the standard normal distribution and $\hat{\sigma}$ is some consistent estimate of σ. Design points $\{x_{i_j} : |JDC_{DKE}(x_{i_j})| > u_n, j = 1, 2, \cdots, n_1\}$ can be flagged as candidate jump positions. But if x_{i_j} is flagged, its neighboring design points will be flagged with high probability. Qiu (1994) defined *tie* sets of the flagged candidates and suggested using the middle point of each tie set to replace the entire set as a new jump candidate. After this modification procedure, the current jump candidates are assumed to be $b_1 < b_2 < \cdots < b_q$. Then q and $\{b_i\}_{i=1}^{q}$ are used as estimators of the number of jumps and the jump positions, respectively. Qiu and Yandell (1998) applied this idea to the LLS procedure to estimate the number of jumps. We think that it could be applied to most existing jump detectors including $JDC_{LLK}(x)$ to get rid of the required prior information about the number of jumps.

Example 5.1 Consider the regression function used in Example 2.1 and let $n = 256$ and $h_n = 21/256$. The Wu and Chu (W-C) (1993a) and Qiu and Yandell (Q-Y) (1998) procedures are used to estimate the number of jumps (the true number is 1). The W-C procedure gives a correct estimation if it rejects H_0 for $H_0 : p = 0$ vs $H_a : p > 0$ and accepts H_0 for $H_0 : p = 1$ vs $H_a : p > 1$ as well. We first let c change from 2 to 3 and fix σ at .25. The numbers of correct estimations by the two methods out of 1000 replications are presented in Figure 5.1(a). We then fix c at 2 and change σ from .1 to 1. The corresponding simulation results are given in Figure 5.1(b). In the W-C procedure, g_n is chosen $2h_n$ and the related kernel functions are selected by the Remark 3 in Wu and Chu (1993a).

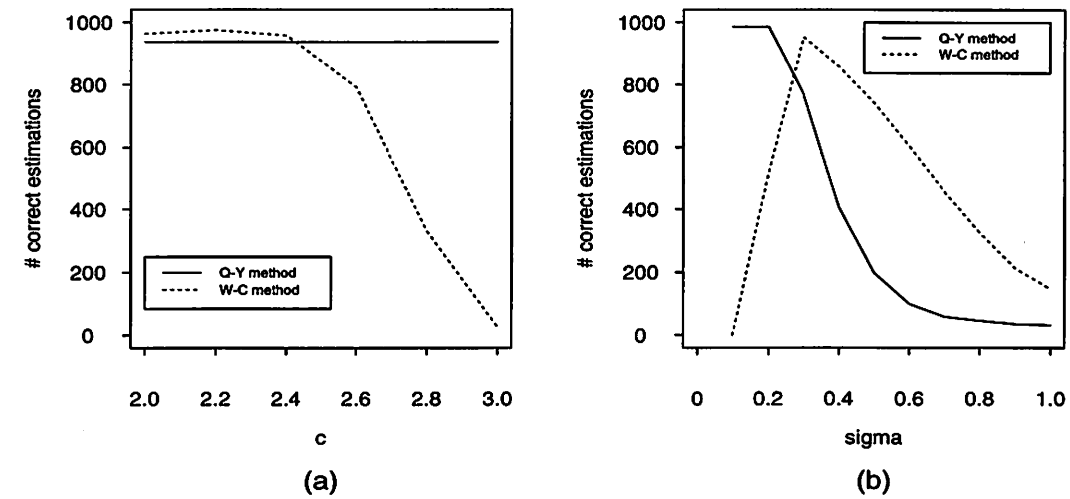


Figure 5.1: (a) The numbers of correct estimations of the number of jumps out of 1000 replications when c changes from 2 to 3 and $\sigma = .25$. (b) The corresponding results when σ changes from .1 to 1 and $c = 2$.

As indicated by Figure 5.1(a), the W-C procedure is sensitive to the slope of the regression curve while the Q-Y procedure performs stablely when c changes, which is consistent with what we found in Figure 2.2 about the DKE and LLS criteria on which the W-C and Q-Y procedures are based. From Figure 5.1(b), the W-C procedure performs better than the Q-Y procedure when σ is larger than a certain number, which might be explained by the fact that the LLS criterion is noisier than the DKE criterion as we found in Figure 3.2. The performance of the W-C procedure is not good when σ is small. That may be related to the effect of $f'(x)$ on jump detection in $f(x)$ which is relatively large when σ is small. Qiu and Yandell (1998) proved that the estimated number of jumps by the Q-Y procedure was almost surely consistent. The Q-Y procedure was generalized to 2-D case by Qiu and Yandell (1997).

<u>**Example 5.2**</u> In Example 5.1, let $c = 2$ and $\sigma = .25$. After the number of jumps and the jump positions are estimated by the Q-Y procedure, we fit $f(x)$ by the conventional local linear kernel smoothing procedure in each design subinterval separated by the estimated jump positions. The

true jump magnitude d changes from 0 to 2 (the regression function is $f(x) = 2x - dI_{(.5,1]}(x)$). For each d value, the averaged MSE value of the fitted $f(x)$ out of 1000 replications is presented in Figure 5.2(a) by the solid curve. As a comparison, we also fit $f(x)$ by the conventional local linear kernel method without considering the jump structure. Its averaged MSE values are plotted in Figure 5.2(a) by the dotted curve. Figure 5.2(b) shows a noisy version of $f(x)$ when $d = 1$ and the LLK estimators of $f(x)$ with and without considering the jump structure.

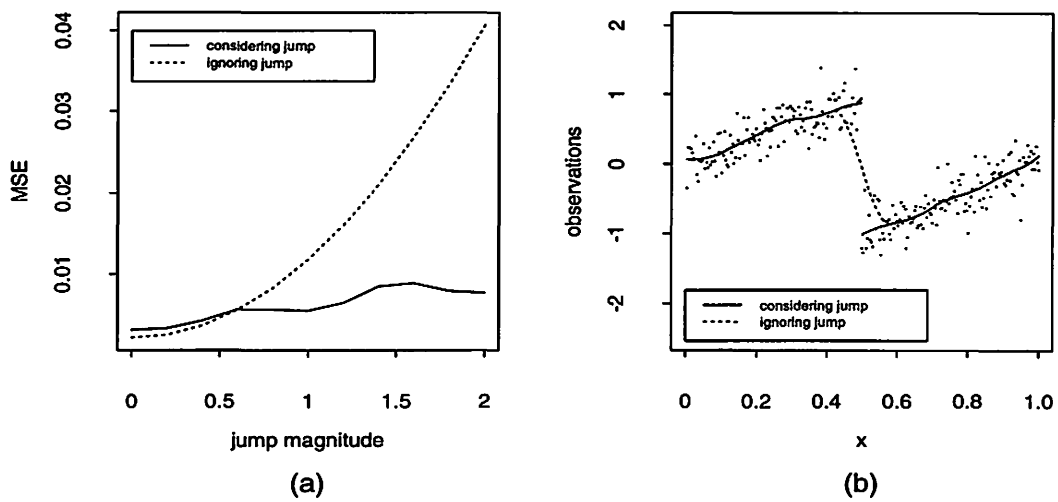


Figure 5.2: (a) Averaged MSE values out of 1000 replications of the LLK estimators of $f(x)$ with and without considering the jump structure of the model when the true jump magnitude changes from 0 to 2. (b) A noisy version of $f(x)$ and its LLK estimators with and without considering the jump structure.

Figure 5.2(a) shows that the jump-preserving procedure does not loss much accuracy when the regression function is continuous ($d = 0$) or when d is too small to be detected. When d gets larger, it becomes more obvious that the jump-preserving procedure outperforms the conventional procedure. As indicated by Figure 5.2(b), the jump structure is smoothed away by the conventional smoothing procedure. It is not hard to check that the conventional smoothing estimator of $f(x)$ is not statistically consistent at the jump position.

6 Concluding Remarks

We have discussed several jump detection and jump-preserving curve fitting methods. Generally speaking, the LLK and LLS procedures perform better than the DKE procedure at places where the regression curve is steep but continuous. On the other hand, the DKE procedure is more capable to smooth away the noise in data. Regarding the computational complexity, the LLS procedure is the simplest one and the LLK procedure requires the most extensive computation.

Many problems are still open in this area. For example, in some applications it is hard to know in which derivative the regression function has jumps. It can also happen that the regression function itself and some of its derivatives have jumps simultaneously. We have not seen much discussion in the literature about these issues. The bandwidth selection is always a problem for local smoothing procedures. The variable bandwidth idea is natural for jump curve fitting. It needs further theoretical research and simulation studies for us to give some practical guidelines for practitioners to choose bandwidths in finite sample situations. (Some interesting discussions can be found in Wu and Chu (1993b) about fixed bandwidth selection by the cross validation procedure.) Another problem that has not been well discussed in the literature is how to detect jumps in border regions of the design space. In large sample case, this might not be a problem since most jump detectors can detect jumps a small distance away from the boundary of the design space and this distance tends to 0 when the sample size increases. In finite sample case, however, the border regions could be relatively large and some methods need to be developed to detect jumps efficiently in those regions.

Two-dimensional problems are much complicated and many important issues have not been well addressed yet in the literature. Some existing methods in jump location curves estimation include the "maximin" procedure (Korostelev and Tsybakov 1993; Müller and Song 1994), the "contrast statistic" algorithm based on smoothing spline (O'Sullivan and Qian 1994), the "rotational difference kernel estimation" proposal (Qiu 1997), the local least squares estimation algorithm (Qiu and Yandell 1997), and the "change curve estimation via wavelets" method (Wang 1998), among many others. Hall and Raimondo (1997, 1998) studied special features of the case where design variables were on a regular grid. For jump-preserving surface fitting methods, see Chu *et al.* (1998), Qiu (1998), and the references cited there.

References

Bracewell, R.N. (1995), *Two-Dimensional Imaging,* Prentice-Hall, Inc.: New Jersey.

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

Eubank, R.L., and Speckman, P.L. (1994), "Nonparametric estimation of functions with jump discontinuities," IMS Lecture Notes, vol. 23, *Change-Point Problems* (E. Carlstein, H.G. Müller and D. Siegmund eds.), 130-144.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications,* Chapman & Hall: London.

Hall, P., and Raimondo, M. (1997), "Approximating a line thrown at random onto a grid," *The Annals of Applied Probability 7,* 648-665.

Hall, P., and Raimondo, M. (1998), "On global performance of approximations to smooth curves using gridded data," *The Annals of Statistics 26,* (December Issue).

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

Hastie, T., and Loader, C. (1993), "Local regression: automatic kernel carpentry," *Statistical Science 8*, 120-143.

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

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

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

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. (1994), "Maximin estimation of multidimensional boundaries," *Journal of the Multivariate Analysis 50*, 265-281.

O'Sullivan, F., and Qian, M. (1994), "A regularized contrast statistic for object boundary estimation – implementation and statistical evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 561-570.

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 (A) 59*, 268-294.

Qiu, P. (1998), "Discontinuous regression surfaces fitting," *The Annals of Statistics 26,* (December Issue).

Qiu, P., Asano, Chi., and Li, X.(1991), "Estimation of jump regression functions," *Bulletin of Informatics and Cybernetics 24,* 197-212.

Qiu, P., Chappell, R., Obermeyer, W., and Benca, R. (1997), "Modelling daily and subdaily cycles in rat sleep data," *Biometrics* (second revision).

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. (1998), "A local polynomial jump detection algorithm in nonparametric regression," *Technometrics 40,* 141-152.

Shea, D.J., Worley, S.J., Stern, I.R., and Hoar, T.J. (1994),"An introduction to atmospheric and oceanographic data," NCAR/TN-404+IA, Climate and Global Dynamics Division, National Center For Atmospheric Research, Boulder, Colorado.

Shiau, J.H., Wahba, G., and Johnson, D.R. (1986), "Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two- and three-dimensional objective analysis," *Journal of Atmospheric and Oceanic Technology 3*, 714-725.

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.

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

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