

ESTIMATION OF THE NUMBER OF JUMPS OF THE JUMP REGRESSION
FUNCTIONS

Peihua Qiu

Department of Statistics
University of Wisconsin - Madison
1210 West Dayton Street
Madison, WI 53706

Key Words and Phrases: Jump regression functions; Kernel smoothers; a.s. consistent; Rate of convergence

ABSTRACT

This paper suggests an estimator of the number of jumps of the jump regression functions. The estimator is based on the difference between right and left one-sided kernel smoothers. It is proved to be a.s. consistent. Some results about its rate of convergence are also provided.

1 Introduction

Regression analysis is one of the most mature branches in statistics. For a long time, however, its main theory is about the continuous regression functions (c. f.

Draper and Smith (1981), Härdle (1990), etc.). Recently, discontinuous regression functions have gotten more and more attention from statisticians all over the world. This, we think, is mainly due to their great application background (e. g. Wahba (1986) used the discontinuous regression model to explore the equi-temperature surfaces of the high sky and the deep ocean).

By now, we have found that jump regression functions are discussed in two statistical fields. One is the change-point field, in which statisticians discuss the jump regression functions by means of the change-point methods (see Chen (1988), Medvedev and Kazachenok (1986), etc.). The other is the partial spline method, which merges the jump regression functions into the so-called partial linear regression models and deals with them as a whole (see Wahba (1986), Shiau (1987), etc.). There is a common point in these two methods that they both merge jump regression models into other more general models. Doing so inevitably loses much of the special features of the jump regression functions.

McDonald and Owen (1986) may be the first to analyse the discontinuous regression functions directly from their own features. In that paper, they provided a split linear smoothing algorithm which is based on three smoothed estimates of the regression function. These three estimates were taken to be linear fits corresponding to the observations on the right, on the left, and on both sides of the point in question . Then they got another smooth fit by the weighted average of these three estimates. The weights were determined by the goodness-of-fit values of the estimates. If there was a jump at the given point, then some of these three estimates were good while the others were not. Because of the weighted average nature, the new smooth fit consisted mainly of the good estimates. So the new smooth fit had much flexibility when discontinuities existed. In their algorithm, they also used the different window sizes in each linear fit.

Hall and Titterington (1992) proposed an alternative method to that of McDonald and Owen (1986). They found and used some features of the differences

between three linear fits which were similar to those of McDonald and Owen (1986) and then diagnosed the existence of the discontinuities directly. Yin (1988) suggested an algorithm to estimate the number, locations and magnitudes of jumps. He averaged the observations on the right and left side of a given point respectively, and then made a difference between them. This difference was the basic criterion in his paper.

Almost at the same time, Wu and Chu (1992), Müller (1992), Qiu (1991) and Qiu et al (1991) suggested similar estimators of the jump regression functions. These estimators were all based on the difference between two one-sided kernel smoothers. One such smoother used the right-sided kernel while the other one used the left-sided kernel. Because of the weighted average nature of the kernel smoothers, the difference would be near to zero when there was no jump at a given point. Otherwise, the difference would be near to the jump magnitude. So this difference was a good criterion for us to judge the existence of the jumps. Qiu et al (1991) called these estimators *The Difference Kernel Estimators* (DKE). DKE connected the jump regression models with the kernel regression techniques. Doing so has at least two benefits. One is that it is possible for us to explore the properties of the DKE by the abundant kernel regression theory. Actually, Wu and Chu (1992) proved the asymptotic normality of DKE, Müller (1992) got the global L^p consistency and Qiu et al (1991) proved the a. s. and L^2 consistencies of DKE in various cases. The other benefit is that we can make use of the existing kernel regression programs in SAS or other statistical software to do our computations. So these estimators are not too difficult to use. Based on these points, we can say that DKE is a good tool for the analysis of the jump regression models.

Up to now, most of the jump regression theory is under the assumption that the number of jumps is known. The estimation of the number of jumps has been bothersome for a long time. However, this is a very important problem in applications.

We can find some related work about the estimation of the number of jumps in the literature. Chen (1988) gave a test method for the hypothesis that there are no jumps in the step regression function. Wu and Chu (1992) suggested a method to test the hypotheses $H_0 : p = j$ vs $H_1 : p > j$ where p is the true number of jumps and j is any positive integer. Hall and Titterton (1992) developed an algorithm to detect the jumps at every design point. Finally, as we mentioned above, Yin (1988) gave a strongly consistent estimator of the number of jumps. But its computation is complicated and also it is not easy for us to obtain its further properties.

In this paper, we make another attempt. An estimator of the number of jumps based on the idea of DKE is suggested. We also use two one-sided kernels for which the domains on which they take non-zero values are apart from the original point for a distance. Our method detects jumps at m points instead of at every design point like Hall and Titterton (1992) and Yin (1988). The quantity m is far less than the number of design points (c. f. section 2 for its definition); Hence our computation is simpler. Section 2 will describe the construction of this estimator. Its a. s. consistency and some results about its rate of convergence are provided in section 3. We do some simulation studies in section 4.

2 The construction of the estimator of the number of jumps

Suppose the regression model concerned is

$$Y_i = f(t_i) + \varepsilon_i, i = 1, 2, \dots, n \tag{1}$$

where $0 \leq t_1 < t_2 < \dots < t_n \leq 1$ are the design points, $E\varepsilon_i\varepsilon_j = \sigma^2\delta_{ij}$, $i, j = 1, 2, \dots, n$. The regression function $f(t)$ has the following form

$$f(t) = g(t) + \sum_{i=1}^p C_i I_{[s_i, s_{i+1})}(t), t \in [0, 1] \tag{2}$$

where $g(t)$ is the continuous part of $f(t)$, p is the number of jumps, $\{s_i\}_1^p$ are the jump locations, $\{d_i \equiv C_i - C_{i-1}\}_{i=1}^p$ are the jump magnitudes, $C_0 = 0$ and $s_{p+1} = 1$. We will suggest an estimator of p in this section which will be proved to be a. s. consistent in the next section.

First of all, suppose $\min_{1 \leq i \leq p} |d_i| \geq B_0 > 0$, where B_0 is a finite positive number. Namely, the absolute values of the jump magnitudes have a common positive lower bound. This assumption will be removed at the end of this section.

Let $K_1(x)$ and $K_2(x)$ be two kernels which satisfy the following conditions:

- (i) $K_1(x) = 0$, when $x \notin [-2, -1]$; $K_2(x) = 0$, when $x \notin [1, 2]$.
- (ii) $K_i(x) \geq 0, i = 1, 2$.
- (iii) $\int_{-2}^2 K_i(x) dx = 1, i = 1, 2$.

Conditions (ii) and (iii) guarantee the kernels to be probability densities which are commonly used in kernel regression techniques (c. f. Härdle (1990)). Condition (i) is crucial to our method. It requires (1) both of these kernels are one-sided ($K_1(x)$ is left-sided, $K_2(x)$ is right-sided), (2) both of them take non-zero values in special intervals. These intervals are apart from the original point for a distance which is equal to the lengths of the intervals.

Define

$$M_n(t) = \frac{1}{nh_n} \sum_{i=1}^n Y_i \left[K_2 \left(\frac{t_i - t}{h_n} \right) - K_1 \left(\frac{t_i - t}{h_n} \right) \right], 2h_n \leq t \leq 1 - 2h_n$$

where the window size h_n satisfies $h_n > 0, \lim_{n \rightarrow \infty} h_n = 0$ and $\lim_{n \rightarrow \infty} nh_n = \infty$.

$M_n(t)$ is the difference of two kernel regression estimators of $f(t)$. As we pointed out in section 1, it is a good criterion to detect the jumps. In the above condition (i), $K_1(x)$ and $K_2(x)$ are required to vanish on $(-1, 0]$ and $[0, 1)$ respectively. As a result, $M_n(s), \forall s \in (t - h_n, t + h_n)$, is actually the difference between the weighted average of the observations on a right-side neighbour of t and that of the observations on a left-side neighbour of t . So all $M_n(s), s \in (t - h_n, t + h_n)$, instead of $M_n(t)$ only, keep useful information about

the observations near the jump at t . In other words, the criterion function $M_n(\cdot)$ is more “robust” for jump detection in this case than that in the case when we only use one-sided kernels as Müller (1992), Wu and Chu (1992), Qiu (1991) and Qiu et al (1991) did. This property makes the jump detection easier. If there is no jump at t , then $\|M_n(s)\|_{(t-h_n, t+h_n)}$ will be close to zero, especially when n is large enough. Otherwise, $\|M_n(s)\|_{(t-h_n, t+h_n)}$ is close to the jump magnitude at t . Here, $\|M_n(s)\|_{(t-h_n, t+h_n)}$ means $\max_{t-h_n < s < t+h_n} |M_n(s)|$.

Now, denote

$$a_i = \frac{i}{m}, i = 1, 2, \dots, m$$

$$m = \left[\frac{1}{2h_n} \right] + 1$$

where $[x]$ is the integer part of x . Then m will tend to infinity when n increases. As a result, $\{a_i, i = 1, 2, \dots, m\}$ will be dense on $[0, 1]$. Instead of detecting jumps at every design point as Hall and Titterton (1992) do, we concentrate our attention only on $\{a_i, i = 1, 2, \dots, m\}$.

Suppose $a_{i_1} < a_{i_2} < \dots < a_{i_r}$ are the points in $\{a_i\}_1^m$ which satisfy

$$|M_n(a_{i_j})| \geq \frac{B_0}{2}, j = 1, 2, \dots, r. \quad (3)$$

Then $\{a_{i_j}, j = 1, 2, \dots, r\}$ can be regarded as the candidates of the jump points. But if a_{i_j} is judged as a jump point, then some of its neighbouring points in $\{a_i\}_1^m$ will probably also be regarded as jump points even if they are actually not. So we need to cancel some of the candidates in $\{a_{i_j}\}_{j=1}^r$. We do this by the following procedure. If there are $r_1 < r_2$ such that

$$\begin{cases} a_{i_{j+1}} - a_{i_j} = \frac{1}{m}, j = r_1, r_1 + 1, \dots, r_2 - 1 \\ a_{i_{r_1}} - a_{i_{r_1-1}} > \frac{1}{m} \\ a_{i_{r_2+1}} - a_{i_{r_2}} > \frac{1}{m} \end{cases} \quad (4)$$

then we cancel all of the candidates $\{a_{i_j}\}_{j=r_1}^{r_2}$ and use the point $\frac{a_{i_{r_2}} + a_{i_{r_1}}}{2}$ as a new candidate. If we call the phenomenon described by (4) a *tie* in $\{a_{i_j}\}_{j=1}^r$, then we

just use the middle point of a tie to replace all of the tie points. After this modification, the present candidates consist of two types of points. One type of such points are those in $\{a_{i_j}\}_{j=1}^r$ which do not belong to any ties. The other type consists of the middle points of all of the ties of $\{a_{i_j}\}_{j=1}^r$. We denote the present candidates as $b_1 < b_2 < \dots < b_q$ which obviously satisfy $b_{i+1} - b_i > \frac{1}{m}, i = 1, 2, \dots, (q - 1)$. Then we use $q, \{b_i\}_1^q$ and $\{M_n(b_i)\}_{i=1}^q$ as the estimates of $p, \{s_i\}_1^p$ and $\{d_i\}_1^p$ respectively and call them *The Difference Apart Kernel Estimates (DAKE)*.

Remark 2.1. The name DAKE comes from its construction. DAKE depends on the difference of two kernel regression smoothers. Each kernel takes non-zero values on an interval which is apart from the original point.

The construction of DAKE and its consistency are based on two properties of $M_n(\cdot)$ (c. f. the proof of theorem 3.1 in section 3).

- (1) $\lim_{n \rightarrow \infty} \|M_n(t)\|_{\bigcup_{i=0}^p [s_i+2h_n, s_{i+1}-2h_n]} = 0, a.s.$ where $s_0 = 0$.
- (2) $\lim_{n \rightarrow \infty} \|M_n(t) - d_i\|_{(s_i-h_n, s_i+h_n)} = 0, a.s., \forall i = 1, 2, \dots, p$.

The first property says that $M_n(t)$ will vanish uniformly beyond neighbourhoods with widths $4h_n$ of the jump points when n is large enough. The second is just the “robust” property of $M_n(t)$ as we mentioned before. If the kernels $K_1(x)$ and $K_2(x)$ are required to vanish on $(-\delta, 0]$ and $[0, \delta)$ respectively for some positive number $\delta \neq 1$ instead of on $(-1, 0]$ and $[0, 1)$, then the above property (1) will be violated when $\delta < 1$ and the property (2) will be violated when $\delta > 1$. That is the reason why we choose $\delta = 1$ in the construction of DAKE.

In most applications, we do not know the lower bound of the absolute jump magnitudes. In such cases, we modify the construction of the DAKE as follows. Let $\{B_n\}$ be a series of positive numbers which converge to zero (theorem 3.2 gives the concrete condition on $\{B_n\}$). $\{a_{i_j}^*\}_{j=1}^{r^*}$ are all points in $\{a_i\}_1^m$ which satisfy $|M_n(a_{i_j}^*)| \geq B_n, j = 1, 2, \dots, r^*$, instead of (3). After the similar modification procedure to that described by (4) (namely, we use the middle points of the ties

of $\{a_{i_j}^*\}_{j=1}^{r^*}$ to replace all of the tie points), we can obtain the final estimators q^* , $\{b_i^*\}_1^{q^*}$ and $\{M_n(b_i^*)\}_{i=1}^{q^*}$ of p , $\{s_i\}_1^p$ and $\{d_i\}_1^p$ respectively. Theorem 3.2 shows that these estimators have the same strong consistency as that of q , $\{b_i\}_1^q$ and $\{M_n(b_i)\}_{i=1}^q$.

At the end of this section, we want to point out a fact that some quantities in DAKE (e. g. h_n and B_n) are still not well-defined for practitioners although we give some large sample conditions on them in section 3. In this paper, we do not consider the boundary problem either. All of these need much further research.

3 Strong consistency of the DAKE

In this section, we will prove the strong consistency of the DAKE. First of all, we give a result in Lemma 3.1 about the strong consistency of the kernel regression estimator of the smooth regression function. This result is similar to the Theorem 3 in Cheng and Lin (1981). Cheng and Lin (1981) used their own kernel regression estimator form. We use the Priestley-Chao kernel regression estimator form here. This is the only difference between our Lemma 3.1 and the Theorem 3 in Cheng and Lin (1981).

Lemma 3.1. Let ν be a positive number and β_n be a series of numbers which satisfy $\lim_{n \rightarrow \infty} \beta_n = \infty$. Suppose $f(x) \in Lip(\alpha)$, $\alpha > 0$, $E|\varepsilon_1|^\rho < \infty$, $\rho \geq 2$, $\max_{1 \leq i \leq n} |\frac{1}{n} - (t_i - t_{i-1})| = O(n^{-1-\lambda})$, $\lambda > 0$. $K(x)$ is a non-negative bounded kernel function which satisfies (a) $K(x) = 0$, when $x \notin [-L, L]$, where L is a positive constant (b) $\int_{-L}^L K(x) dx = 1$ (c) $K(x) \in Lip(\beta)$, $\beta > 0$. h_n is the positive window size which satisfies $\lim_{n \rightarrow \infty} h_n = 0$, $\lim_{n \rightarrow \infty} nh_n = \infty$ and the following additional conditions (when n is large enough) (1) $\frac{n^\nu}{\beta_n \log n} [h_n^\alpha + \frac{1}{n^\lambda h_n} + \frac{1}{n^\beta h_n^{\beta+1}}] = o(1)$; (2) $\frac{n^{2\nu}}{n\beta_n h_n} = O(1)$; (3) $\frac{n^{\nu+\frac{1}{p}-1}}{h_n \beta_n \log n} = o(1)$. Then we have, $\forall 0 < a \leq b < 1$,

$$\lim_{n \rightarrow \infty} \frac{n^\nu}{\beta_n \log n} \|f_n(t) - f(t)\|_{[a,b]} = 0, \text{ a.s.}$$

where $f_n(t) = \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{t_i-t}{h_n}\right)$.

(The proof is similar to that of the Theorem 3 in Cheng and Lin (1981).)

Corollary 3.1. In the above lemma, if $\rho \geq 2\beta + 1$, $h_n = O(n^{-\varepsilon})$, $\varepsilon \geq \frac{2\beta-1}{2\beta+1}$, $\alpha \geq \frac{1}{2\beta-1} > 0$, $\lambda \geq \frac{2\beta}{2\beta+1}$, $\nu = \frac{1}{2\beta+1}$, then the conditions (1)-(3) are satisfied and hence under the other conditions, we have

$$\|f_n(t) - f(t)\|_{[a,b]} = o(n^{-\frac{1}{2\beta+1}} \log n);$$

if $2 \leq \rho < 2\beta + 1$, $\alpha \geq \frac{1}{2\beta-1}$, $\nu = \frac{2}{2\beta+1} - \frac{1}{\rho} > 0$, $h_n = O(n^{-\varepsilon})$, $\varepsilon \geq \frac{2\beta-1}{2\beta+1}$, $\lambda \geq 1 - \frac{1}{\rho}$, then the conditions (1)-(3) are also satisfied and hence under the other conditions, we have

$$\|f_n(t) - f(t)\|_{[a,b]} = o(n^{\frac{1}{\rho} - \frac{2}{2\beta+1}} \log n).$$

Theorem 3.1. Under the regression model (1) and (2), suppose $g(t) \in Lip(\alpha)$, $\alpha > 0$, $E|\varepsilon_1|^\rho < \infty$, $\rho \geq 2$, $\max_{1 \leq i \leq n} |\frac{1}{n} - (t_i - t_{i-1})| = O(n^{-1-\lambda})$, $\lambda > 0$, and $\min_{1 \leq i \leq p} |d_i \equiv C_i - C_{i-1}| \geq B_0 > 0$, where B_0 is a constant. Kernel functions $K_1(x)$, $K_2(x)$ are bounded and satisfy conditions (i)-(iii) (in section 2) and (iv) $K_i(x) \in Lip(\beta)$, $\beta > 0$, $i = 1, 2$. ν is a positive number. $\{\beta_n\}$ are a sequence of numbers which satisfy $\lim_{n \rightarrow \infty} \beta_n = \infty$. h_n is the positive window size which satisfies $\lim_{n \rightarrow \infty} h_n = 0$, $\lim_{n \rightarrow \infty} nh_n = \infty$ and the conditions (1)-(3) in Lemma 3.1 when n is large enough. Then we have

- (1) $\lim_{n \rightarrow \infty} q = p$, *a.s.*; $\lim_{n \rightarrow \infty} b_i = s_i$, *a.s.*; $\lim_{n \rightarrow \infty} M_n(b_i) = d_i$, *a.s.*, $i = 1, 2, \dots, p$.
- (2) The rates of convergence of $\lim_{n \rightarrow \infty} q = p$, *a.s.*, and $\lim_{n \rightarrow \infty} M_n(b_i) = d_i$, $i = 1, 2, \dots, p$, *a.s.*, are $o(\frac{\beta_n \log n}{n^\nu})$. The rates of convergence of $\lim_{n \rightarrow \infty} b_i = s_i$, $i = 1, 2, \dots, p$, *a.s.*, are faster than h_n .

Proof. By lemma 3.1, we have

$$\lim_{n \rightarrow \infty} \|M_n(t)\|_{\bigcup_{i=0}^p [s_i + 2h_n, s_{i+1} - 2h_n]} = 0, \text{ a.s.}$$

So, almost surely, $a_{i_1} < a_{i_2} < \dots < a_{i_r}$ are all in $\bigcup_{i=1}^p (s_i - 2h_n, s_i + 2h_n)$ when n is large enough. For $\forall 1 \leq i \leq p$, $M_n(t) - d_i$ can be regarded as the difference of two kernel smoothers of the continuous regression function $f(t) - d_i I_{t > s_i}$ when $t \in [s_i - h_n, s_i + h_n]$. The above fact is guaranteed by the one-sided natures of the two kernels (c. f. condition (i) on two kernels in section 2). So

$$\lim_{n \rightarrow \infty} \|M_n(t) - d_i\|_{(s_i - h_n, s_i + h_n)} = 0, a.s. \quad (5)$$

By the definition of $\{a_j\}_1^m$ and the relationship between m and h_n (namely, $m = \lfloor \frac{1}{2h_n} \rfloor + 1$), we know that, $\forall 1 \leq i \leq p$, one of the points $\{a_j\}_1^m$, without loss of generality suppose it is a_{j_i} , is in $(s_i - h_n, s_i + h_n)$. So by (5),

$$\lim_{n \rightarrow \infty} M_n(a_{j_i}) = d_i, i = 1, 2, \dots, p, a.s..$$

Hence there must be one point of $\{a_{i_j}\}_{j=1}^r$ (c. f. their definitions in section 2) to be in $(s_i - h_n, s_i + h_n)$ when n is large enough. On the other hand, $\forall 1 \leq i \leq p$, it is obvious that $(s_i - 2h_n, s_i + 2h_n)$ covers at most three points of $\{a_{i_j}\}_{j=1}^r$. According to the construction of the DAKE, almost surely, there is one and only one point of $\{b_i\}_1^q$ to be in $(s_i - 2h_n, s_i + 2h_n)$ when n is large enough and this point must be in $(s_i - h_n, s_i + h_n)$. So

$$\lim_{n \rightarrow \infty} q = p, a.s.$$

and

$$|b_i - s_i| < h_n, i = 1, 2, \dots, p, a.s. \quad (6)$$

when n is large enough.

According to (5) and (6), we have

$$\lim_{n \rightarrow \infty} b_i = s_i, i = 1, 2, \dots, p, a.s.$$

$$\lim_{n \rightarrow \infty} M_n(b_i) = d_i, i = 1, 2, \dots, p, a.s..$$

It is easy to see, according to Lemma 3.1, that the rate of convergence of (5) reaches $o((\frac{n^\nu}{\beta_n \log n})^{-1})$. So do the rates of convergence of $\lim_{n \rightarrow \infty} q = p, a.s.$, and

of $\lim_{n \rightarrow \infty} M_n(b_i) = d_i, i = 1, 2, \dots, p, a.s.$ (6) says that the rates of convergence of $\lim_{n \rightarrow \infty} b_i = s_i, i = 1, 2, \dots, p, a.s.$, are faster than h_n . We have now proved the theorem.

Corollary 3.2. In theorem 3.1, if $\rho \geq 2\beta + 1, h_n = O(n^{-\varepsilon}), \varepsilon \geq \frac{2\beta-1}{2\beta+1}, \alpha \geq \frac{1}{2\beta-1} > 0, \lambda \geq \frac{2\beta}{2\beta+1}, \nu = \frac{1}{2\beta+1}$ and other conditions remain, then the rates of convergence of $\lim_{n \rightarrow \infty} q = p, a.s.$, and $\lim_{n \rightarrow \infty} M_n(b_i) = d_i, i = 1, 2, \dots, p, a.s.$, reach $o(n^{-\frac{1}{2\beta+1}} \log n)$; if $2 \leq \rho \leq 2\beta + 1, \alpha \geq \frac{1}{2\beta-1}, \nu = \frac{2}{2\beta+1} - \frac{1}{\rho} > 0, h_n = O(n^{-\varepsilon}), \varepsilon \geq \frac{2\beta-1}{2\beta+1}, \lambda \geq 1 - \frac{1}{\rho}$ and other conditions remain, then the rates of convergence of $\lim_{n \rightarrow \infty} q = p, a.s.$, and $\lim_{n \rightarrow \infty} M_n(b_i) = d_i, i = 1, 2, \dots, p, a.s.$, reach $o(n^{\frac{1}{\rho} - \frac{2}{2\beta+1}} \log n)$.

Remark 3.1. From Corollary 3.2 and Theorem 3.1, we know that the faster the rates of convergence of $\lim_{n \rightarrow \infty} q = p, a.s.$, and $\lim_{n \rightarrow \infty} M_n(b_i) = d_i, i = 1, 2, \dots, p, a.s.$, the slower the rates of convergence of $\lim_{n \rightarrow \infty} b_i = s_i, i = 1, 2, \dots, p, a.s.$. The inverse relation is also true. When $\beta = 1$ and $\rho \geq 3$, these rates of convergence are all equivalent to $O(n^{-1/3})$.

When the lower bound of the absolute jump magnitudes is unknown, we establish the following theorem about the consistencies of the DAKE $q^*, \{b_i^*\}_1^{q^*}$ and $\{M_n(b_i^*)\}_{i=1}^{q^*}$.

Theorem 3.2. Let $\{B_n\}$ be a series of positive numbers which satisfy $B_n = O(n^{-\nu} \beta_n \log n)$. All of the conditions in theorem 3.1 are satisfied except that “ $\min_{1 \leq i \leq p} |d_i| \geq B_0 > 0$ ”. Then

- (i) $\lim_{n \rightarrow \infty} q^* = p, a.s.; \lim_{n \rightarrow \infty} b_i^* = s_i, a.s.; \lim_{n \rightarrow \infty} M_n(b_i^*) = d_i, a.s., i = 1, 2, \dots, p,$
- (ii) The rates of convergence of $\lim_{n \rightarrow \infty} q^* = p, a.s.$, and $\lim_{n \rightarrow \infty} M_n(b_i^*) = d_i, i = 1, 2, \dots, p, a.s.$, reach $o(\frac{\beta_n \log n}{n^\nu})$. The rates of convergence of $\lim_{n \rightarrow \infty} b_i^* = s_i, i = 1, 2, \dots, p, a.s.$, are faster than h_n .

Proof. By lemma 3.1, we know that the rate of convergence of

$$\lim_{n \rightarrow \infty} \|M_n(t)\|_{\bigcup_{i=0}^p [s_i + 2h_n, s_{i+1} - 2h_n]} = 0, a.s.$$

is $o(B_n)$. So all of the points $\{a_{i_j}^*\}_{j=1}^{r^*}$ are in $\bigcup_{i=1}^p (s_i - 2h_n, s_i + 2h_n)$ almost surely when n is large enough. The remaining proof is similar to that of Theorem 3.1.

4 Simulation studies

In this section, we do some computer simulations about the DAKE when B_0 is unknown. First of all, we use $f(x) = x^4 + I_{[0.3,1]}(x)$, $x \in [0, 1]$, which was also used by Wu and Chu (1992). $f(x)$ has one jump point at 0.3. The sample size n equals 100, $\varepsilon_1 \sim N(0, \sigma^2)$, $\sigma^2 = 0.25^2$, $K_2(x) = -\frac{\pi}{2} \sin(\pi x) I_{[1,2]}(x)$, $K_1(x) = K_2(-x)$, and $h_n = \frac{1}{5} n^{-1/3}$. The factor $n^{-1/3}$ in the expression for h_n coincides with the Remark 3.1. The factor 1/5 is mainly due to the following two considerations. One is to guarantee the fact that there are enough observations to be actually used in the kernel regression at any point. In this case, $M_n(t)$ actually uses 8 observations. The second consideration is from the value of $m = \lceil \frac{1}{2h_n} \rceil + 1$ (c. f. section 2); m should not be very small. In the present case, $m = 12$. In $B_n = 2n^{-1/4}$, $n^{-1/4}$ comes from Theorem 3.2. The factor 2 is about 5 times the mean of $\{M_n(a_i)\}_{i=1}^m$. This is from our experience. Of course, these constant factors are not very important when the sample size n is large enough. One thousand replications were done, and the results are summarized in TABLE I.

TABLE I

| q^* | frequencies | $b_i^*, i = 1, 2, \dots, q^*$ | frequencies |
|-------|-------------|-------------------------------|-------------|
| 1 | 950 | $\in (0.299, 0.301)$ | 983 |
| 2 | 50 | $\in (0.25, 0.35)$ | 1000 |
| > 2 | 0 | | |

We also apply our method to the data on the annual volume of the Nile river from 1871 to 1970 in Cobb (1978). According to Cobb (1978) and Müller (1992), there is an abrupt change in this set of data in the year 1898. The data are displayed in Fig.4.1.

Fig.4.1

By our program, we know that the mean of $\{M_n(a_i)\}_{i=1}^m$ is 88.3694. So we use $B_n = 400n^{-1/4}$ where the constant 400 is about 5 times the mean of $\{M_n(a_i)\}_{i=1}^m$; $h_n, K_1(x)$ and $K_2(x)$ are the same as those in the previous numerical example. The results are $q^* = 1, b^*(1)$ is close to the year 1899, and $M_n(b^*(1)) = -389.369$. These results coincide with those in Cobb (1978) and Müller (1992).

ACKNOWLEDGEMENTS

The author is grateful to a referee for his helpful comments.

BIBLIOGRAPHY

- Chen, X.R.(1988), "Hypothesis testing and interval estimation about the model which has only one change-point," *Scientia Sinica 8A*, 1-13.
- Cheng, K.F., and Lin, P.E.(1981), "Nonparametric estimation of a regression function," *Z. Wahrscheinlichkeitstheorie verw. Gebiete 57*, 223-233.
- Cobb, G.W.(1978), "The problem of the Nile: conditional solution to a changepoint problem," *Biometrika 65*, 243-251.
- Draper, N.R., and Smith, H.(1981), *Applied Regression Analysis*, New York: John-

Wiley & Sons, 2nd edition.

- Hall, P., and Titterton, M. (1992), "Edge-preserving and peak-preserving smoothing," *Technometrics* 34, 429-440.
- Härdle, W. (1990), *Applied Nonparametric regression*, Oxford University Press.
- McDonald, J.A., and Owen, A.B. (1986), "Smoothing with split linear fits," *Technometrics* 28, 195-208.
- Medvedev, G., and Kazachenok, V. (1986), "Estimation of the discontinuous regression function," In *detection of Changes in Random Processes* (Laimutis Telksnys eds.) New York: Optimal Software Inc..
- Müller, H.G. (1992), "Change-points in nonparametric regression analysis," *The Annals of Statistics* 20, 737-761.
- Qiu, P. (1991), "Estimation of a kind of jump regression functions," *Systems Science and Mathematical Sciences* 4, 1-13.
- Qiu, P., Asano, Chi., and Li, X. (1991), "Estimation of jump regression function," *Bulletin of Informatics and Cybernetics* 24, 197-212.
- Shiau, J.H. (1987), "A note on MSE coverage intervals in a partial spline model," *Communications in Statistics-Theory and Methods* 16, 1851-1866.
- Wahba, G. (1986), "Partial spline modelling of the tropopause and other discontinuities," *Function Estimate, Contemporary Mathematics, Vol. 59* (J.S. Marron eds.), 125-135.
- Wu, J.S., and Chu, C.K. (1992), "Kernel type estimators of jump points and values of a regression function," to appear in *The Annals of statistics*.

Yin, Y.Q.(1988), "Detecting of the number, locations and magnitudes of jumps,"
Communications in Statistics-Stochastic Models 4, 445-455.

FIG 4.1 Annual Volume Of The Nile River From 1871 To 1970