Communications of the Korean Statistical Society Vol. 15, No. 6, 2008, pp. 899–908

Estimation of Jump Points in Nonparametric Regression†

Dongryeon Park[1)]

Abstract

If the regression function has jump points, nonparametric estimation method based on local smoothing is not statistically consistent. Therefore, when we estimate regression function, it is quite important to know whether it is reasonable to assume that regression function is continuous. If the regression function appears to have jump points, then we should estimate first the location of jump points. In this paper, we propose a procedure which can do both the testing hypothesis of discontinuity of regression function and the estimation of the number and the location of jump points simultaneously. The performance of the proposed method is evaluated through a simulation study. We also apply the procedure to real data sets as examples.

Keywords: Discontinuous regression function; jump detection; nonparametric regression.

1. Introduction

Nonparametric regression analysis provides a very powerful tool for estimating regression function from noisy data. Nonparametric regression techniques based on conventional local smoothing procedure assume that the regression function is smooth. In some application, however, it is more appropriate to assume that the regression function has a finite number of jump discontinuities. The December sea-level pressure in Bombay, India was found to have a jump discontinuity around year 1960. See Müller and Stadtmüller (1999) for more background information and many examples of discontinuous regression model.

When the regression function has discontinuity points, the traditional smoothing method such as the local polynomial regression is not statistically consistent at jump points of the true regression function. Therefore, if we know the regression function has a finite number of discontinuity points, we should estimate first the location of the discontinuity points and then apply the traditional smoothing method to each smooth part of regression function separately. Thus the estimation of jump positions is the main issue for the estimation of discontinuous regression function.

The problem of estimating jump positions in nonparametric regression has received much attention over the years. The literature on this topic is quite large. These researches can be classified into two categories based on knowledge of the number of discontinuity

† This work has been supported by Hanshin University Research Grant.

1) Professor, Department of Statistics, Hanshin University, 441 Yangsan-Dong, Osan, Kyunggi-Do 447-791, Korea. Email: drpark@hs.ac.kr.

points. When the number of jumps are known, Müller (1992), Qiu *et al.* (1991) and Wu and Chu (1993) suggested similar jump estimators. These estimators are all based on the difference between two one-sided kernel smoother. One smoother is based on the observations in the right-sided neighborhood of given point while the other one is based on the observations in the left-sided neighborhood of given point. If there is no jump at a given point, the difference would be near zero. Otherwise, the difference would be near the jump magnitude. As a result, the difference includes the information about the jump position and can be used as a jump detection criterion. Qiu *et al.* (1991) called these estimators the difference kernel estimators(DKE). Gijbels and Goderniaux (2004) suggested a modified version of DKE. Their jump detection procedure is based on the derivative of Nadaraya-Watson kernel estimator.

However, it is not always the case in applications that the number of discontinuity points are known. Jump estimation methods when the number of discontinuity points are unknown have been suggested by Hall and Titterington (1992), Wu and Chu (1993), Qiu (1994) and Qiu and Yandell (1998).

Hall and Titterington (1992) suggested the jump detection algorithm based on the differences between three local linear fits. Wu and Chu (1993) also suggested determining the number of discontinuity points by performing a sequence of testing hypotheses $H_0 : p = j$, $H_1 : p > j$ where p is the true number of jumps and j is any positive integer. The testing process stops after the first "fail to reject H_0" is obtained. Qiu (1994) suggested an estimator of the number of jumps based on the idea of DKE. He uses two one-sided kernels for which the domains on which they take non-zero values are apart from the original point for a distance. This procedure is called the difference apart kernel estimation(DAKE). Qiu and Yandell (1998) proposed a jump detection procedure based on local polynomial estimation.

In this paper, we make another attempt to propose a procedure for estimating both the number and the location of discontinuity points. The procedure is based on DKE. In fact, DKE is very efficient and easy to implement, but it can not be used if the number of jumps is not known. In our procedure we estimate first the number of jump points and then estimate the location of jump points. The number of jump points is estimated by performing the test. Therefore, our procedure has some similar aspect with Wu and Chu (1993), but the test proposed in Wu and Chu (1993) is solely based on asymptotic distribution, so its performance at the small sample sizes situation is not good at all. We use the test statistic proposed in Park (2008), which is verified to have an excellent performance even in small sample sizes.

In Section 2, the procedure for determining the number and the location of jump points is described. In Section 3, the simulation results to evaluate the performance of the procedure are presented. The examples of real data set are considered in Section 4.

2. Estimation of Jump Points

Suppose we want to estimate the regression function m using a sample of n data $\{(x_i, Y_i),\ i = 1, \ldots, n\}$ generated from model (2.1).

$$Y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where ϵ_i's are independent and identically distributed with mean 0 and finite variance σ^2. We assume that the regression function m can be expressed by

$$m(x) = g(x) + \sum_{j=1}^{p} d_j I(x > s_j),$$

where g is a continuous function in the entire design interval, p is the number of jump points, $\{s_j,\ j = 1, \ldots, p\}$ are the jump positions and $\{d_j,\ j = 1, \ldots, p\}$ are jump magnitudes. We assume that the distance between any two of s_j is greater than δ. Here δ is an arbitrary small positive constant. Without loss of generality, we assume that m is defined on the interval $[0, 1]$. We also assume that the design points x_i are equally spaced on $[0, 1]$.

For given $x \in [h,\ 1-h]$, the difference kernel estimator, M_{DKE} is defined as

$$M_{DKE}(x) = \hat{m}_+(x) - \hat{m}_-(x),$$

where

$$\hat{m}_+(x) = \sum_{i=1}^{n} Y_i K_1\left(\frac{x_i - x}{h}\right) \Big/ \sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h}\right)$$

and

$$\hat{m}_-(x) = \sum_{i=1}^{n} Y_i K_2\left(\frac{x_i - x}{h}\right) \Big/ \sum_{i=1}^{n} K_2\left(\frac{x_i - x}{h}\right).$$

h is a smoothing parameter and K_1 and K_2 are kernel functions satisfying the following conditions:

1. the support of K_1 is $[0,\ 1]$ and the support of K_2 is $[-1,\ 0)$.
2. $K_j(x) \geq 0,\ j = 1,\ 2.$
3. $\int_{-1}^{1} K_j(x)dx = 1,\ j = 1,\ 2.$

The kernel estimator $\hat{m}_+(x)$ is a weighted average of the observations in the right-sided neighborhood $[x,\ x+h]$, whereas $\hat{m}_-(x)$ is a weighted average of the observations in the left-sided neighborhood $[x-h,\ x)$. Thus if the regression function m is continuous at x, then both $\hat{m}_+(x)$ and $\hat{m}_-(x)$ are good estimator of $m(x)$, so $M_{DKE}(x)$ should be small. However, if m is discontinuous at x, then $\hat{m}_+(x)$ is a good estimator of $m_+(x)$ only and $\hat{m}_-(x)$ is a good estimator of $m_-(x)$ only, so in this case $M_{DKE}(x)$ can be used as an estimator of the jump magnitude $m_+(x) - m_-(x)$ where $m_+(x)$ and $m_-(x)$ denote the right and left limit of $m(x)$, respectively.

Therefore, when it is assumed that there are $p \geq 1$ jumps and p is known, the jump positions $\{s_j,\ j = 1, \ldots, p\}$ and the jump magnitudes $\{d_j,\ j = 1, \ldots, p\}$ can be estimated as follows. Let $\tilde{s}_j$ be the maximizers of $|M_{DKE}(x)|$ over the sets $\tilde{A}_j$, where

$$\tilde{A}_j = [h,\ 1-h] - \bigcup_{k=1}^{j-1} [\tilde{s}_k - h,\ \tilde{s}_k + h],$$

for $j = 1, \ldots, p$. Let

$$\tilde{d}_j = M_{DKE}(\tilde{s}_j), \quad j = 1, \ldots, p.$$

Then we define $\tilde{s}_j$ and $\tilde{d}_j$ as the estimator of s_j and d_j, respectively for $j = 1, \ldots, p$. When we detect several jumps as above, two jumps that are less than h apart can not be distinguished, so δ which is the distance between any two jump points is assumed to be $\delta = h$.

However, it is not assumed that p is known in this paper, so it is not possible to apply the above procedure directly. What we need is an estimator of p. To estimate p, we adopt the statistic proposed by Park (2008). He considered the testing problem

$$\begin{aligned} H_0 &: \ m_+(x) = m_-(x) \quad \forall\, x \in [0,1] \\ H_1 &: \ m_+(x) \neq m_-(x) \quad \exists\, x \in [0,1] \end{aligned} \tag{2.2}$$

and proposed the test statistic

$$T(x) = \frac{M_{DKE}(x) - c_1 \hat{m}'(x)}{\sqrt{c_2 \hat{\sigma}^2}},$$

where

$$c_1 = \frac{\sum_{i=1}^{n} (x_i - x) K_{1h}(x_i - x)}{\sum_{i=1}^{n} K_{1h}(x_i - x)} - \frac{\sum_{i=1}^{n} (x_i - x) K_{2h}(x_i - x)}{\sum_{i=1}^{n} K_{2h}(x_i - x)},$$

$$c_2 = \frac{\sum_{i=1}^{n} K_{1h}^2(x_i - x)}{\left(\sum_{i=1}^{n} K_{1h}(x_i - x)\right)^2} + \frac{\sum_{i=1}^{n} K_{2h}^2(x_i - x)}{\left(\sum_{i=1}^{n} K_{2h}(x_i - x)\right)^2}$$

and

$$K_{jh}(x_i - x) = K_j\left(\frac{x_i - x}{h}\right), \quad j = 1,\, 2.$$

Under H_0, the distribution of T is a standard normal distribution, so if $|T(x)| \geq z_{1-\alpha/2}$, for any $x \in [0,1]$ then we can reject H_0 where $z_{1-\alpha/2}$ is $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

Combing the estimation of the jump points procedure based on M_{DKE} and the testing procedure using $T(x)$, we proposed to take $\hat{s}_j$ as the maximizers of $|T(x)|$ over the sets A_j, where

$$A_j = [h,\, 1-h] - \bigcup_{k=1}^{j-1} [\hat{s}_k - h,\, \hat{s}_k + h],$$

for $j = 1, \ldots, r$ and r is a positive integer which is far less than n. Then we define $\hat{p}$ as the number of $\hat{s}_j$'s such that

$$|T(\hat{s}_j)| \geq z_{1-\frac{\alpha}{2}}, \quad j = 1, \ldots, \hat{p}. \tag{2.3}$$

We propose to use $\hat{p}$ and $\{\hat{s}_j\}_{j=1}^{\hat{p}}$ as the estimator of p and $\{s_j\}_{j=1}^{p}$, respectively.

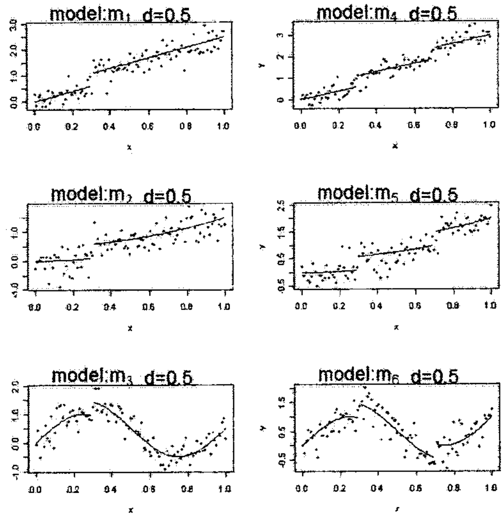

Figure 3.1: The typical data set for each model along with the true regression line

3. Simulation Study

A simulation study was conducted to evaluate the finite sample performance of the proposed estimation procedure. We consider the following discontinuous regression models:

$$\begin{aligned}
m_1(x) &= 2x + d \cdot I(x > s_1),\\
m_2(x) &= x^2 + d \cdot I(x > s_1),\\
m_3(x) &= \sin(2\pi x) + d \cdot I(x > s_1),\\
m_4(x) &= 2x + d \cdot I(x > s_1) + d \cdot I(x > s_2),\\
m_5(x) &= x^2 + d \cdot I(x > s_1) + d \cdot I(x > s_2),\\
m_6(x) &= \sin(2\pi x) + d \cdot I(x > s_1) + d \cdot I(x > s_2),
\end{aligned}$$

where s_1 and s_2 are jump positions and we used $s_1 = 0.3$, $s_2 = 0.7$ and d is a jump magnitude and we consider $d = 0.5$, 1, 2 for each model. The design points were given by $x_i = (i-1)/(n-1)$ for $i = 1, \ldots, 100$. The error terms were generated from $N(0,\, 0.1)$. The typical data sets of each regression model for $d = 0.5$ along with the true regression lines are presented in Figure 3.1.

We chose the kernel function K_1 as $K_1(x) = 1.5(1-x^2)I_{[0,1]}(x)$ and $K_2(x) = K_1(-x)$ for all x. The error variance σ^2 was estimated by $\hat{\sigma}^2 = \sum_{i=2+\nu}^{n-\nu} \xi_i/2(n-1-2\nu)$, where ξ_i denote the rearranged $(Y_i - Y_{i-1})^2$ in ascending order and we chose $\nu = 2$. The first derivative estimator of each regression model was evaluated by the function *glkerns* of the package *lokern* in R. We used the significance level $\alpha = .05$ for the procedure (2.3).

The smoothing parameter plays important role in our procedure. To investigate the effect of the smoothing parameter to the estimation procedure, we used several different smoothing parameters. We used $h = 0.05$, 0.1, 0.15, 0.2, 0.25 for each data set.

The simulation results based on 1000 replications are presented Table 3.1 and 3.2.

Table 3.1: Simulation results based on 1000 replications for one jump point models(The frequencies of $\hat{p}$ and the number of times that $\hat{s}_j$, $j = 1, \ldots, \hat{p}$ fell within ± 0.01 and ± 0.05 from true jump positions)

	d	h	$\hat{p}$				s_1	s_1
			0	1	2	>2	± 0.01	± 0.05
m_1	0.5	0.25	607	388	5	0	254	359
		0.20	710	281	9	0	167	240
		0.15	783	202	15	0	120	151
		0.10	826	164	9	1	57	78
		0.05	871	116	13	0	22	36
	1.0	0.25	21	957	22	0	904	978
		0.20	65	908	27	0	865	932
		0.15	185	780	35	0	737	782
		0.10	426	538	35	1	504	528
		0.05	699	279	21	1	212	233
	2.0	0.25	0	985	15	0	996	1000
		0.20	0	973	27	0	999	1000
		0.15	0	957	41	2	998	1000
		0.10	1	931	64	4	996	999
		0.05	64	869	65	2	924	928
m_2	0.5	0.25	600	391	9	0	266	361
		0.20	730	261	9	0	165	219
		0.15	780	211	9	0	123	151
		0.10	811	170	19	0	59	74
		0.05	851	130	19	0	18	42
	1.0	0.25	24	955	21	0	896	972
		0.20	75	893	32	0	836	921
		0.15	211	744	44	1	724	768
		0.10	401	556	42	1	508	543
		0.05	735	244	19	2	165	192
	2.0	0.25	0	986	14	0	998	1000
		0.20	0	979	21	0	998	1000
		0.15	0	964	36	0	998	1000
		0.10	0	945	54	1	997	1000
		0.05	74	852	70	4	907	915
m_3	0.5	0.25	538	438	24	0	231	323
		0.20	713	268	19	0	172	221
		0.15	782	206	12	0	111	150
		0.10	850	137	13	0	55	80
		0.05	868	122	10	0	27	40
	1.0	0.25	13	932	55	0	910	975
		0.20	49	898	53	0	866	940
		0.15	172	796	31	1	747	805
		0.10	391	566	42	1	537	563
		0.05	726	251	23	0	183	203
	2.0	0.25	0	940	60	0	1000	1000
		0.20	0	952	48	0	998	1000
		0.15	0	958	42	0	1000	1000
		0.10	0	943	55	2	1000	1000
		0.05	49	875	74	2	928	939

In these table, we report the frequencies of $\hat{p}$. We also report the number of times that $\hat{s}_j$, $j = 1, \ldots, \hat{p}$ fell within $s_1 \pm 0.01$ and $s_1 \pm 0.05$ for one jump point model and within $s_j \pm 0.01$ and $s_j \pm 0.05$, $j = 1, 2$ for two jump points model.

Table 3.2: Simulation results based on 1000 replications for two jump points models(The frequencies of $\hat{p}$ and the number of times that $\hat{s}_j$, $j = 1, \ldots, \hat{p}$ fell within ± 0.01 and ± 0.05 from true jump positions)

	d	h	$\hat{p}$				s_1	s_2	s_1	s_2
			0	1	2	>2	± 0.01	± 0.01	± 0.05	± 0.05
m_4	0.5	0.25	434	435	131	0	252	244	336	316
		0.20	555	366	79	0	179	163	247	227
		0.15	673	284	42	1	118	117	147	152
		0.10	781	201	16	2	69	56	88	75
		0.05	860	128	12	0	24	19	33	33
	1.0	0.25	0	38	962	0	903	898	982	976
		0.20	10	143	845	2	850	849	912	911
		0.15	41	341	616	2	720	712	776	763
		0.10	226	463	292	19	489	478	515	519
		0.05	581	345	72	2	201	192	223	223
	2.0	0.25	0	0	1000	0	998	999	1000	1000
		0.20	0	0	999	1	1000	999	1000	1000
		0.15	0	0	991	9	1000	998	1000	1000
		0.10	0	0	974	26	997	999	999	1000
		0.05	6	120	837	37	921	920	929	923
m_5	0.5	0.25	432	450	118	0	229	232	322	328
		0.20	535	399	66	0	184	166	239	240
		0.15	702	260	37	1	109	107	131	139
		0.10	748	224	27	1	68	72	89	97
		0.05	861	130	9	0	21	17	35	32
	1.0	0.25	0	53	947	0	885	896	970	973
		0.20	4	130	861	5	850	852	933	917
		0.15	46	316	628	10	728	727	782	785
		0.10	208	469	309	14	514	486	544	521
		0.05	588	334	77	1	181	192	197	214
	2.0	0.25	0	0	1000	0	1000	1000	1000	1000
		0.20	0	0	1000	0	998	998	1000	1000
		0.15	0	0	991	9	1000	999	1000	1000
		0.10	0	1	975	24	997	996	998	997
		0.05	7	137	824	32	912	907	922	917
m_6	0.5	0.25	397	518	85	0	209	112	286	147
		0.20	564	374	62	0	181	109	244	141
		0.15	679	282	37	2	136	84	166	110
		0.10	787	192	21	0	71	49	97	63
		0.05	853	136	11	0	18	27	30	42
	1.0	0.25	1	78	921	0	911	826	967	900
		0.20	11	163	825	1	849	799	916	857
		0.15	48	355	594	3	732	667	792	718
		0.10	234	478	273	15	469	464	503	498
		0.05	599	340	53	8	192	168	207	189
	2.0	0.25	0	0	1000	0	1000	998	1000	1000
		0.20	0	0	1000	0	999	1000	1000	1000
		0.15	0	0	989	11	999	999	1000	1000
		0.10	0	1	972	27	997	992	1000	997
		0.05	9	138	818	35	902	911	915	920

We can see the effect of the bandwidth to the estimation procedure clearly. As h increases, the performance of the procedure is getting better for all cases. We can observe the same phenomenon in the simulation results of Wu and Chu (1993), Bowman *et al.*

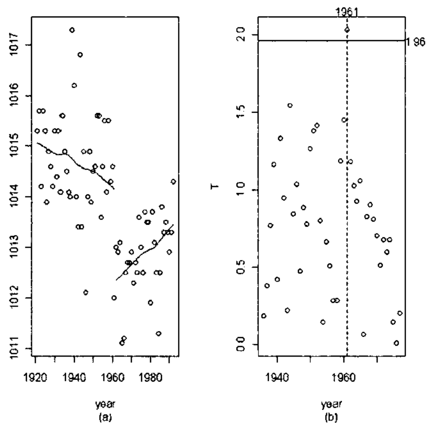

Figure 4.1: The December sea-level pressure from 1921 to 1992 in Bombay, India((a) The small circles represent sea-level pressure and the solid curves represent nonparametric regression estimate, (b) The values of the statistic $|T|$ for each year)

(2006) and Park (2008). These researches are mainly focused on the testing problem of the discontinuous regression function. Since our estimation procedure is based on the test statistics used in Park (2008), it is quite natural that larger bandwidth produces better results. However, we can not detect the discontinuity points if they are in the interval $[0, h]$ and $[1 - h, 1]$, so we should not choose too large bandwidth.

4. Applications to Real Data

4.1. One jump point example

It is well known that the December sea-level pressures during 1921–1992 in Bombay, India has a discontinuity point around 1960 and 1961. The small circles in Figure 4.1(a) represent the sea-level pressures. These data are given in Qiu (2005).

To apply our procedure to real data set, we need to specify the bandwidth. What we know about the bandwidth from the simulation study is that larger bandwidth produces better results but using too large bandwidth is not recommendable. However, we do not have a bandwidth selection method for our estimation procedure yet. Qiu and Yandell (1998) treated this data set and for their own method, they chose $h = 15$ empirically, so we try to choose the same value as h.

Figure 4.1(b) shows the values of the statistic $|T|$ for each year. Note that $|T|$ has a maximum point at 1961 and its value is larger than the critical point of $\alpha = 0.05$. Thus a jump appears to exist at year 1961.

After we estimate the jump position, we divide whole interval into two parts, 1921–1960 and 1961–1992 and then we estimate the discontinuous regression function by applying local linear regression model to each smooth part separately. The fitted curves are presented in Figure 4.1(a).

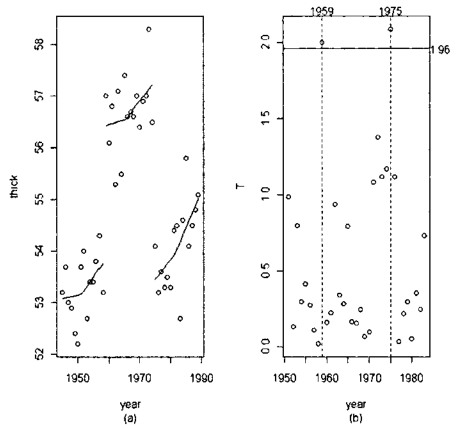

Figure 4.2: Thickness of pennies from 1945 to 1989((a) The small circles represent thickness of pennies and the solid curves represent nonparametric regression estimate, (b) The values of the statistic $|T|$ for each year)

4.2. Two jump points example

The data set for two jump points example is penny thickness data, which is available from Scott (1992). This data set consists of thickness in mils of a sample of 90 U.S. Lincoln pennies dated from 1945 to 1989. Since two pennies were measured for each year, we replace the original data by the mean of each pair per year. It is known that penny thickness was reduced in World War II and restored to its original thickness sometimes around 1960 and reduced again in the 70's. These data are displayed in Figure 4.2(a).

For this data set, Gijbels and Goderniaux (2004) selected the bandwidth by their bootstrap procedure, which is equal to 5. We decide to use the same bandwidth.

Figure 4.2(b) shows the values of $|T|$ for each year. We can see that $\hat{p} = 2$, $\hat{s}_1 = 1959$ and $\hat{s}_2 = 1975$. The discontinuous nonparametric regression function is estimated and displayed in Figure 4.2(a).

5. Conclusion

If the regression function has discontinuity points, the traditional smoothing method is not statistically consistent at jump points. Therefore, when we estimate regression function, we need to know whether it is reasonable to assume that the regression function is continuous. This is about the testing problem of hypothesis in (2.2). After we reject H_0, the next task is naturally to estimate the number and the location of the jump points.

In this paper, we proposed the estimation procedure. Using the procedure, we can actually do both the testing hypothesis of (2.2) and the estimation of the number and the location of the jump points simultaneously. Through the simulation study and real data examples, we have verified that the proposed method works quite well.

However, to apply the proposed estimation procedure in the real situation, we should have the bandwidth selection rule. We need to construct the proper rule and we will leave this further research topic.

Even though the competitiveness of the proposed method is verified empirically, we need to prove $\lim_{n\to\infty} \hat{p} = p$ and $\lim_{n\to\infty} \hat{s}_j = s_j,\ j = 1, \ldots, p$ to provide the theoretical backbone, but the proof of the consistency needs further study.

References

Bowman, A. W., Pope, A. and Ismail, B. (2006). Detecting discontinuities in nonparametric regression curves and surfaces, *Statistics & Computing*, **16**, 377–390.

Gijbels, I. and Goderniaux, A. C. (2004). Bandwidth selection for change point estimation in nonparametric regression, *Technometrics*, **46**, 76–86.

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

Müller, H. G. (1992). Change-points in nonparametric regression analysis, *The Annals of Statistics*, **20**, 737–761.

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

Park, D. (2008). Test for discontinuities in nonparametric regression, To appear in *Communications of the Korean Statistical Society.*

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. (2005). *Image Processing and Jump Regression Analysis*, John Wiley & Sons, New Jersey.

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

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

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley-Interscience, New York.

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

[Received August 2008, Accepted September 2008]