Communications of the Korean Statistical Society
2009, Vol. 16, No. 3, 519–528

Comparison of Jump-Preserving Smoothing and Smoothing Based on Jump Detector

Dongryeon Park[1,a]

[a]Department of Statistics, Hanshin University

Abstract

This paper deals with nonparametric estimation of discontinuous regression curve. Quite number of researches about this topic have been done. These researches are classified into two categories, the *indirect* approach and *direct* approach. The major goal of the indirect approach is to obtain good estimates of jump locations, whereas the major goal of the direct approach is to obtain overall good estimate of the regression curve. Thus it seems that two approaches are quite different in nature, so people say that the comparison of two approaches does not make much sense. Therefore, a thorough comparison of them is lacking. However, even though the main issue of the indirect approach is the estimation of jump locations, it is too obvious that we have an estimate of regression curve as the subsidiary result. The point is whether the subsidiary result of the indirect approach is as good as the main result of the direct approach. The performance of two approaches is compared through a simulation study and it turns out that the indirect approach is a very competitive tool for estimating discontinuous regression curve itself.

Keywords: Difference Kernel estimators, discontinuous regression function, jump detector, jump-preserving smoothing, local constant M-smoother.

1. Introduction

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 (1.1).

$$Y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

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

$$m(x) = f(x) + \sum_{j=1}^{p} d_j I(x > s_j), \tag{1.2}$$

where f 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.

Since the regression function of (1.2) has discontinuity points, the traditional smoothing such as the local polynomial regression is not statistically consistent, so we need quite different smoothing

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. E-mail: drpark@hs.ac.kr.

methods. Quite a number of researches about estimating discontinuous regression function have been done. These researches are classified into two categories. The first approach, which is called the *indirect* approach by Gijbels *et al.* (2007), estimates the locations of the jump points first using one of various jump detection procedure and then applies the traditional smoothing technique to each smooth parts of regression function separately. The literature on indirect estimation method includes Müller (1992), Hall and Titterington (1992), Wu and Chu (1993a), Qiu and Yandell (1998), Park (2008), among others. The second approach estimates the regression curve directly without detecting the jumps explicitly, so it is called the *direct* approach or jump-preserving smoothing. The local constant M-smoother by Chu *et al.* (1998) and the local linear M-smoother by Rue *et al.* (2002) are widely used direct methods. Adaptive weights smoothing proposed by Polzehl and Spokoiny (2000) and the jump-preserving curve fitting procedures proposed by Qiu (2003) and Gijbels *et al.* (2007) are also the direct estimation methods.

The indirect approach is based on the jump detection procedure, so the major issue is to obtain good estimates of jump locations. Once we have correct estimates of jump locations, we have a good chace of getting accurate estimate of regression function. Thus estimation of regression function itself is often secondary in the indirect approach. On the other hand, the major goal of the direct approach is to obtain an overall good estimation of the regression function. In fact, the direct approach does not even require correct estimates of jump location explicitly. It just treat every single data point as a potential discontinuity point.

Apparently, two approaches have quite different major goals, so they look very different in nature. In this point of view, people say that the comparison of two approaches does not make much sense and it is even difficult to choose a proper criterion for comparison because of different major goals. Therefore the research about the comparison of two approaches has not started yet.

However, even though the main issue of the indirect approach is the estimation of jump locations and the estimation of regression function is just subsidiary, it is too obvious that we have the estimate of the regression function as the subsidiary result. The point is whether the subsidiary result of the indirect approach is as good as the main result of the direct approach. Gijbels *et al.* (2007) noted that a good overall estimate of the regression function may not necessarily reveal jump locations and size well. They actually presumed that the direct approach performs better than the indirect approach for obtaining a good overall estimate of discontinuous regression function, but they did not provide any evidence for their presumption.

In this paper, we want to compare the performance of two above-mentioned approaches for estimating discontinuous regression function. Since this kind of comparison has never been tried before, there is no relevant reference for this problem. Therefore, this paper may produce a premature result, but will be a good starting point for this problem.

As the direct estimation method, we choose the local constant M-smoother proposed by Chu *et al.* (1998), which is one of the most widely used direct method. This method is based on the local constant fit, so to make the fair comparison the corresponding indirect estimation method should be based on the local constant fit too. Thus we choose the difference kernel estimators(DKE) as a jump detector and the local constant regression as a smoother. However, the original DKE can be used only if the number of jump is known, which is not the case in this paper. Park (2008) proposed the jump detection procedure. His method is based on DKE and can be used when the number of jump is unknown, so we choose his method as the jump detector for the indirect method.

The paper is organized as follows. In Section 2, we briefly describe the estimating procedures of both approaches. A simulation study investigating the performances of two approaches in Section 3. In Section 4, we provide some concluding remarks and further research topics.

2. Estimation Procedures

In this section, we briefly describe both the local constant M-smoother by Chu *et al.* (1998) and the modified DKE procedure by Park (2008).

2.1. Local constant M-smoother

Throughout the paper, we assume that $\{(x_i, Y_i), i = 1, \ldots, n\}$ are generated from model (1.1), where the design points x_i are equally spaced, and the regression function m is expressed by (1.2). We assume that the number and the location of jumps are unknown, and the distance between any two of s_j of (1.2) 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].

For each x_i, Chu *et al.* (1998) proposed to take the M-smoother $\hat{m}_M(x_i)$ as the local minimizer with respect to θ of

$$S(\theta, x_i) = \sum_{j=1}^{n} K_{h_M}^{M}(x_i - x_j) L_{g_M}^{M}(Y_j - \theta) \tag{2.1}$$

that is closest to Y_i. Here K^M and L^M are kernel functions, h_M and g_M are two bandwidths, $K_{h_M}^{M}(\cdot) = h_M^{-1} K^M(\cdot/h_M)$ and $L_{g_M}^{M}(\cdot) = g_M^{-1} L^M(\cdot/g_M)$. For both kernel function K^M and L^M, they recommended to use the Gaussian density function.

Like any other smoother, the performance of M-smoother heavily depends on the choice of the bandwidth h_M and g_M. The bandwidth h_M works more like a traditional smoothing parameter, and g_M acts like a tuning constant in usual M-estimation. They also provided some recommendation for choice of both parameters, and we will look at them at Section 3.

In many cases, the local constant M-smoother gives remarkably good results. Especially, when the true regression function is constant with abrupt changes at several locations, it shows an excellent performance, preserving edges and spikes. Chu *et al.* (1998) derived the asymptotic bias and variance of the local constant M-smoother and showed that it suffers from the boundary effect.

2.2. Modified DKE procedure

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

$$M_{DKE}(x) = \hat{m}_+(x) - \hat{m}_-(x), \tag{2.2}$$

where

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

and

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

g is a smoothing parameter and K_1 and K_2 are kernel functions. Here the support of K_1 is [0,1] and the support of K_2 is $[-1, 0)$.

The modified DKE procedure proposed by Park (2008) is based on the following testing problem

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

He considered the test statistic

$$T(x) = \frac{M_{DKE}(x) - c_1\hat{m}'(x)}{\sqrt{c_2\hat{\sigma}^2}}, \tag{2.3}$$

where

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

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

and

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

Under H_0, the asymptotic 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 locations procedure based on M_{DKE} and the testing procedure using $T(x)$, $\hat{s}_j$ is proposed to take the maximizers of $|T(x)|$ over the sets A_j, where

$$A_j = [g, 1 - g] - \bigcup_{k=1}^{j-1} [\hat{s}_k - g, \hat{s}_k + g], \tag{2.4}$$

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

$$|T(\hat{s}_j)| \geq z_{1-\frac{\alpha}{2}}, \quad j = 1, \ldots, \hat{p}. \tag{2.5}$$

After jump locations are estimated, the estimator of modified DKE procedure, $\hat{m}_{DKE}(x)$ is now obtained by applying the local constant regression to each B_j separately, where

$$B_j = [\hat{s}_{j-1}, \hat{s}_j], \quad j = 1, \ldots, \hat{p} + 1. \tag{2.6}$$

Here $\hat{s}_0 = 0$ and $\hat{s}_{\hat{p}+1} = 1$. We assume that we use the same bandwidth h when we apply local constant regression estimator to each B_j.

Any theoretical properties of $\hat{m}_{DKE}(x)$ is not derived yet, and it looks even very challenging. Therefore, only empirical comparison of $\hat{m}_M$ and $\hat{m}_{DKE}$ is possible in this paper. In Section 3, we compare the finite sample properties of these two estimators using simulation study.

3. Simulation Study

In this section, we investigate the performance of two estimation procedures described in Section 2 by a simulation study.

We first introduce the simulation settings. We consider the following regression models:

$$m_1(x) = 2 + 2I(x \geq s_1) - 3I(x \geq s_2),$$

$$m_2(x) = 2x + 1I(x \geq s_1) + 0.5I(x \geq s_2),$$

$$m_3(x) = \begin{cases} -3x + 2, & x < s_1, \\ -3x + 3 - \sin\left((x - 0.3)\frac{\pi}{0.2}\right), & s_1 \leq x < s_2, \\ \frac{x}{2} + 1.55, & x \geq s_2, \end{cases}$$

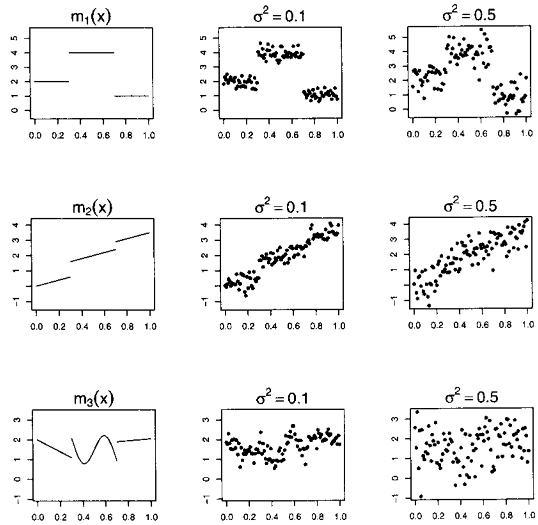

Figure 1: *Left panel: graphs of the true regression functions m_1, m_2 and m_3. Center panel: simulated data with $n = 100$ and $\epsilon \sim N(0, 0.1)$. Right panel: simulated data with $n = 100$ and $\epsilon \sim N(0, 0.5)$*

where $s_1 = 0.3$ and $s_2 = 0.7$ for all models. The design points are given by $x_i = (i-1)/(n-1)$ for $i = 1, \ldots, n$. The sample sizes considered are $n = 20, 50, 100, 200$. The error terms are generated from $N(0, \sigma^2)$ with $\sigma^2 = 0.1, 0.5$. The typical data sets of each regression model for $n = 100$ along with the true regression curves are presented in Figure 1.

$S(\theta, x_i)$ of (2.1) is too complicated to find a closed form for their local minimum. Chu *et al.* (1998) provide sophisticated algorithm based on Newton's method to find the local minimum closest to Y_i, but it is complicated to program. Simpson *et al.* (1998) suggest an iterative weighted least squares(IWLS) algorithm, which is straightforward to program. Given an initial estimate $\theta_i^{(0)}$, IWLS updating with step t is computed as follows:

$$\theta_i^{(t+1)} = \frac{\sum_{j=1}^{n} K_{h_M}^{M}\left(x_i - x_j\right) L_{g_M}^{M}\left(Y_j - \theta_i^{(t)}\right) Y_j}{\sum_{k=1}^{n} K_{h_M}^{M}\left(x_i - x_k\right) L_{g_M}^{M}\left(Y_k - \theta_i^{(t)}\right)}.$$

For both kernel function K^M and L^M, we choose the Gaussian density function. As a stopping rule, if $|\theta_i^{(t)} - \theta_i^{(t+1)}| < 10^{-5}$, then we take $\theta_i^{(t)}$ as $\hat{m}_M(x_i)$.

Although the IWLS algorithm can potentially find the wrong local minimum, Simpson *et al.* (1998) argue that if one uses the data as starting values and reasonable bandwidths g_M and h_M, there is no problem for getting correct convergence. To use the data as starting values, we simply take $\theta_i^{(0)} = Y_i$, $i = 1, \ldots, n$. The bandwidth selection methods for both g_M and h_M, however, are not well established. For g_M, we have two different recommendation. Chu *et al.* (1998) recommend the range of $g \in [2\sigma/3, 3\sigma/2]$ based on their empirical experience, but Burt (2000) recommends to take $g_M = 2.11\sigma$. The number 2.11 is determined by the Asymptotic Relative Efficiency(ARE) calculation. We have examined two recommendation in the simulation, and it turns out that $g_M = 2.11\hat{\sigma}$ produces better results, so we report the simulation results only for the $g_M = 2.11\hat{\sigma}$ case.

For the estimator of σ^2, we use the trimmed mean version used in Wu and Chu (1993a), which is defined as $\hat{\sigma}^2 = \sum_{i=2+\nu}^{n-\nu} \xi_i / 2(n - 1 - 2\nu)$ where ξ_i denote the rearranged $(Y_i - Y_{i-1})^2$ in ascending order,

Table 1: The smallest $\widehat{\text{MISE}}$ values along with their bandwidths based on 1000 replications, and $\widehat{\text{MISE}}_{s=0.3} + \widehat{\text{MISE}}_{s=0.7}$ values for given bandwidths. Ratio (last column) is the ratio of $\widehat{\text{MISE}}$(DKE) to $\widehat{\text{MISE}}$(M-smoother).

	σ^2	n	Modified DKE			M-smoother			Ratio
			h	$\widehat{\text{MISE}}$	$\sum\widehat{\text{MISE}}_s$	h_M	$\widehat{\text{MISE}}$	$\sum\widehat{\text{MISE}}_s$	
m_1	0.1	20	0.09	0.0526	0.0137	0.25	0.0159	0.0025	3.30
		50	0.25	0.0073	0.0022	0.21	0.0088	0.0021	0.82
		100	0.25	0.0034	0.0010	0.25	0.0079	0.0021	0.43
		200	0.25	0.0018	0.0006	0.21	0.0049	0.0014	0.36
	0.5	20	0.13	0.2129	0.0614	0.09	0.1669	0.0465	1.27
		50	0.13	0.1060	0.0588	0.05	0.1037	0.0552	1.02
		100	0.23	0.0416	0.0254	0.05	0.0682	0.0450	0.60
		200	0.25	0.0149	0.0089	0.03	0.0462	0.0284	0.32
m_2	0.1	20	0.15	0.0330	0.0089	0.09	0.0558	0.0096	0.59
		50	0.17	0.0210	0.0119	0.09	0.0367	0.0125	0.57
		100	0.13	0.0110	0.0067	0.07	0.0263	0.0117	0.41
		200	0.13	0.0063	0.0042	0.05	0.0168	0.0084	0.37
	0.5	20	0.25	0.0852	0.0176	0.13	0.1243	0.0240	0.68
		50	0.23	0.0529	0.0221	0.09	0.0608	0.0240	0.87
		100	0.19	0.0387	0.0210	0.07	0.0389	0.0196	0.99
		200	0.13	0.0267	0.0161	0.05	0.0268	0.0151	0.99
m_3	0.1	20	0.09	0.0482	0.0092	0.09	0.0774	0.0162	0.62
		50	0.07	0.0357	0.0164	0.05	0.0569	0.0221	0.62
		100	0.05	0.0256	0.0121	0.03	0.0455	0.0170	0.56
		200	0.03	0.0175	0.0079	0.03	0.0333	0.0154	0.52
	0.5	20	0.25	0.1276	0.0242	0.25	0.1402	0.0183	0.91
		50	0.13	0.0845	0.0330	0.07	0.0924	0.0366	0.91
		100	0.11	0.0600	0.0265	0.05	0.0627	0.0302	0.95
		200	0.09	0.0448	0.0208	0.03	0.0411	0.0207	1.09

and we choose $\nu = 2$.

The bandwidth h_M works more like a traditional smoothing parameter, so the conventional data adaptive bandwidth selection method, like cross-validation or plug-in rule can be used, but performance of these methods for choosing h_M have not been thoroughly investigated yet. Thus instead of using data adaptive method, we use several different values of h_M in the simulation.

For the modified DKE procedure, the bandwidth g and h are also very important factors. The bandwidth g is used for the jump detection procedure of (2.4), and h is used for the local constant regression which is applied to each B_j of (2.6). On the issue about the bandwidth g, however, little work has been done. Gijbels and Goderniaux (2004) proposed the data-driven bandwidth selection method using bootstrap procedure, but their method is complicated, so it is difficult to implement in practice. Park (2008) investigated the effect of the bandwidth g to the modified DKE procedure and found that the larger bandwidth produced the better results. We can observe the same empirical evidence in Wu and Chu (1993a) and Bowman *et al.* (2006). Thus we may need to choose the value of g as large as possible, but we can detect the discontinuity points only in the interval $[g, 1-g]$, so too large value of g should be avoided. We take $g = 0.2$ for the simulation, but this choice is rather arbitrary.

For the bandwidth h, we may use the cross-validation procedure proposed in Wu and Chu (1993b). However, for the fair comparison, we use several different values of h just like the case of h_M.

For the kernel function of M_{DKE} in (2.2), we choose $K_1(x) = 1.5(1 - x^2)I_{[0,1]}(x)$ and $K_2(x) = K_1(-x)$ for all x. The first derivative estimate of regression function, $\hat{m}'(x)$ in $T(x)$ of (2.3) is evaluated by the function *glkerns* of the package *lokern* in R. The computation of the local constant regression

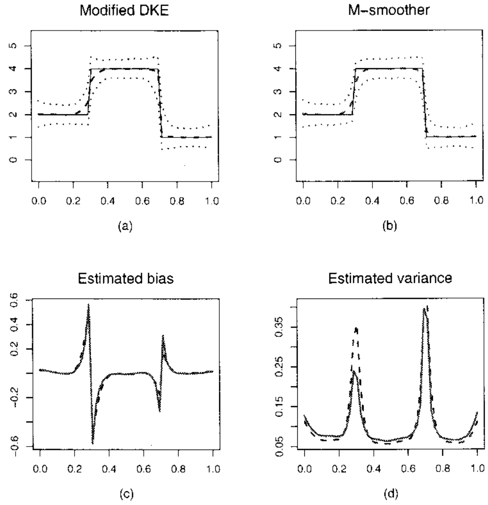

Figure 2: *Plots of two estimates for* m_1, $n = 50$, $\sigma^2 = 0.5$, $h = 0.13$ *and* $h_M = 0.05$*((a), (b) The true regression curve is denoted by the solid line. The average of 1000 replicated fits is denoted by the dashed line. The corresponding* 5^{th} *percentile and the* 95^{th} *percentile are denoted by dotted lines. (c), (d) The solid line is for M-smoother and the dashed line is for modified DKE)*

estimates for each B_j of (2.6) is done by R function *ksmooth*. The significance level for the procedure (2.5) is set to $\alpha = 0.05$.

We want to reemphasize that the main objective of two procedures is to obtain good overall estimation of regression function m. Therefore a natural criterion for comparison of two procedure is the Mean Integrated Squared Error(MISE). In order to avoid boundary effects the MISE is calculated for the slightly smaller interval [0.1,0.9]. Since we estimate a jump regression curve, the curve estimates also need to be jump-preserving. To measure jump-preserving around a given jump point s, Gijbels *et al.* (2007) proposed to use the following local MISE:

$$\text{MISE}_s = E\left[\int_{s-0.05}^{s+0.05} \{\hat{m}(x) - m(x)\}^2\, dx\right].$$

The number 0.05, the half-width of interval, is subjectively selected.

For each model, we estimate the MISE of two procedures based on 1000 replications for various h_M and h values, and report the smallest $\widehat{\text{MISE}}$ of each procedure along with the corresponding bandwidth in Table 1. We also report the estimated local MISE, $\widehat{\text{MISE}}_{s=0.3} + \widehat{\text{MISE}}_{s=0.7}$ based on the bandwidth which produces the smallest $\widehat{\text{MISE}}$.

As we can see in Table 1. $\widehat{\text{MISE}}$ and $\sum \widehat{\text{MISE}}_s$ show almost identical pattern. Moreover, during the simulation, we have observed that the small values of $\sum \widehat{\text{MISE}}_s$ do not necessarily imply excellent performance of jump-preserving. Maybe we need other criterion for measuring jump-preserving property.

The regression model m_1 is the case that the underlying true regression function is constant with abrupt changes at serval locations, and in this case, it is well known that the local constant M-smoother gives remarkably good results, but it turns out that modified DKE procedure is competitive and even better than M-smoother at large sample case. The M-smoother performs better than modified DKE

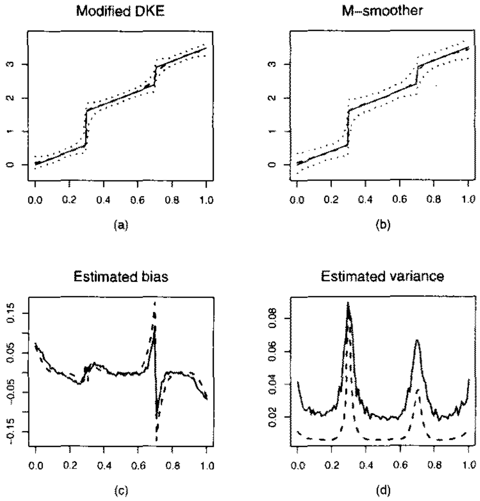

Figure 3: *Plots of two estimates for m_2, $n = 100$, $\sigma^2 = 0.1$, $h = 0.13$ and $h_M = 0.07$((a), (b) The true regression curve is denoted by the solid line. The average of 1000 replicated fits is denoted by the dashed line. The corresponding 5^{th} percentile and the 95^{th} percentile are denoted by dotted lines. (c), (d) The solid line is for M-smoother and the dashed line is for modified DKE)*

only for small sample case. For m_2 and m_3 with $\sigma^2 = 0.1$, modified DKE is superior to M-smoother for all sample sizes, and in fact, the relative performance of modified DKE, which is measured by the ratio of $\widehat{\text{MISE}}$(DKE) to $\widehat{\text{MISE}}$(M-smoother), is getting better as n increases. However, for m_2 and m_3 with $\sigma^2 = 0.5$, we have different pattern. The modified DKE performs slightly better than M-smoother only for $n = 20$ and 50 cases. Moreover, the relative performance of modified DKE is getting worse as n increases. Eventually, for m_3 case, M-smoother has slightly smaller $\widehat{\text{MISE}}$ at $n = 200$.

Figure 2 to 4 give the graphical display of the performance of two procedures. In (a) and (b) of each plot, the dashed line represents the average of 1000 replicated fits, so the difference between the solid and the dashed line can be considered as the estimated bias, and the difference between the upper and lower dotted line is a kind of measure of variation about 1000 replicated fits. In Figure 2, we can see that two procedure have almost the same performance, except that modified DKE has slightly larger variance at $x = 0.3$. In Figure 3 and Figure 4, we can see that modified DKE has slightly larger bias at jump points, but M-smoother has much larger variance, which results in larger $\widehat{\text{MISE}}$ of M-smoother.

4. Conclusion

Even though the indirect approach and the direct approach have quite different major goals, the final results of both approaches are the estimates of the regression curve. Surprisingly, however, the comparison of two approaches as the estimator of the discontinuous regression function has never been done before. In this paper, we compared the performance of two approaches using the simulation study and it turns out that the indirect approach is a very competitive tool for estimating regression curve even though the major goal of the indirect approach is not the estimation of the regression curve itself.

We only compared the local constant version of two approaches, but we know that the local linear

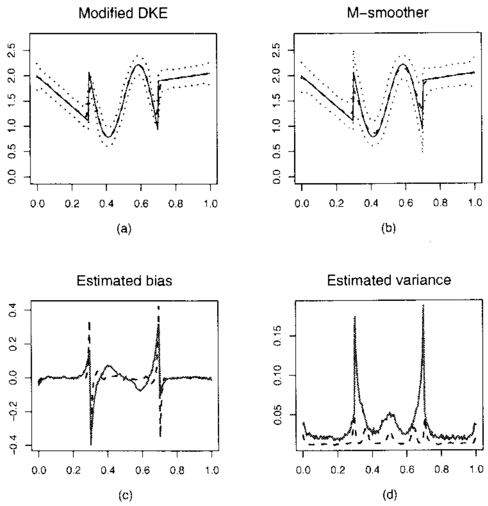

Figure 4: *Plots of two estimates for* m_3, $n = 200$, $\sigma^2 = 0.1$, $h = 0.03$ *and* $h_M = 0.03$*((a), (b) The true regression curve is denoted by the solid line. The average of 1000 replicated fits is denoted by the dashed line. The corresponding* 5^{th} *percentile and the* 95^{th} *percentile are denoted by dotted lines. (c), (d) The solid line is for M-smoother and the dashed line is for modified DKE)*

version is more appropriate in situations where the true curve is far from being constant but still contains abrupt changes. We will extend the comparison to the local linear version of two approaches and then we may have more concrete conclusion about which approach provides better overall estimate of discontinuous regression function.

An important question which we have not addressed is how to choose smoothing parameters data-adaptively. We do not have clear answer about how to choose g_M, h_M and g yet. The optimal selection methods for these bandwidths need further study.

References

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

Burt, D. A. (2000). *Bandwidth selection concerns for jump point discontinuity preservation in the regression setting using M-smoothers and the extension to hypothesis testing*, Ph. D. dissertation, Virginia Polytechnic Institute and State University, Department of Statistics.

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

Gijbels, I., Lambert, A. and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: A compromise, *Annals of the Institute of Statistical Mathematics*, **59**, 235–272.

Gijbels, I. and Goderniaux, A. C. (2004). Bandwidth selection for change point estimation in nonparametric regression, *Technometics*, **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.

Park, D. (2008). Estimation of jump points in nonparametric regression, *Communications of the Korean Statistical Society*, **15**, 899–908.

Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration, *Journal of the Royal Statistical Society: Series B*, **62**, 335–354.

Qiu, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation, *Journal of Nonparametric Statistics*, **15**, 437–453.

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

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

Simpson, D. G., He, X. and Liu, Y. (1998). Comment on: Edge-preserving smoothers for image, In *Processing Journal of the American Statistical Association*, **93**, 544–548.

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.

Received March 2009; Accepted March 2009