Comparison of Nonparametric Function Estimation Methods for Discontinuous Regression Functions

Dongryeon Park[1]

[1]Department of Statistics, Hanshin University

Abstract

There are two main approaches for estimating the discontinuous regression function nonparametrically. One is the *direct* approach, the other is the *indirect* approach. The major goal of the two approaches are different. The direct approach focuses on the overall good estimation of the regression function itself, whereas the indirect approach focuses on the good estimation of jump locations. Apparently, the two approaches are quite different in nature. Gijbels *et al.* (2007) argue that the comparison of two approaches does not make much sense and that it is even difficult to choose an appropriate criterion for comparisons. However, it is obvious that the indirect approach also has the regression curve estimate as the subsidiary result. Therefore it is necessary to verify the appropriateness of the indirect approach as the estimator of the discontinuous regression function itself. Park (2009a) compared the performance of two approaches through a simulation study. In this paper, we consider a more general case and draw some useful conclusions.

Keywords: Discontinuous regression function, jump detector, jump-preserving smoothing, local M-smoother.

1. Introduction

Suppose we want to estimate a discontinuous regression function nonparametrically. In this case, the traditional smoothing technique such as the local polynomial regression is not statistically consistent and we need quite different smoothing methods. Numerous research on the nonparametric estimation of discontinuous regression functions exist that are classified into two categories. The first approach is called the *indirect* approach and estimates the locations of the jump points first using one of the various jump detection procedures and then estimates each smooth parts of the regression function by separately applying the traditional smoothing technique. The second approach is called the *direct* approach or jump-preserving smoothing and estimates the regression function directly without detecting the jumps explicitly. Refer to Gijbels *et al.* (2007) and Park (2009a) for the literature on both direct and indirect estimation methods.

[1]Professor, Department of Statistics, Hanshin University, 441 Yangsan-dong, Osan, Gyeonggi 447-791, Korea. E-mail: drpark@hs.ac.kr

The main issue of the indirect approach is to obtain good estimates of jump locations. The estimation of regression function itself is just a subsidiary result in the indirect approach. On the other hand, the main issue of the direct approach is to obtain an overall good estimate of the regression function. The two approaches are quite different in nature and Gijbels *et al.* (2007) argue that the comparison of two approaches does not make sense and that it is even difficult to choose an appropriate criterion for comparisons. However, even though the estimation of the regression function is not the main issue in the indirect approach, it is too obvious that the indirect approach also produces the estimate of the regression function in the final step. Therefore, it is quite natural to verify the appropriateness of the indirect approach as the estimator of the discontinuous regression function.

Park (2009a) compared the performance of the two above-mentioned approaches as the estimator of the discontinuous regression function itself through a simulation study. He chose the local constant M-smoother proposed by Chu *et al.* (1998) as the direct approach and the modified difference kernel estimator(DKE) proposed by Park (2008) as the indirect approach. Note that both local constant M-smoother and modified DKE are based on the local constant regression.

In this paper, we extend the simulation study of Park (2009a). We include the procedures based on the local linear regression for both the direct and indirect approach. It is well known that the local linear regression has several good properties over the local constant regression. One of them is that the local linear regression does not suffer from the boundary problem. Since the local smoothing jump detection procedures use the one-sided kernel, every single design points can be considered as a 'boundary point'. In this respect, the jump detection procedure based on the local linear regression is expected to have a considerable advantage. In addition, the local linear M-smoother is expected to be superior to the local constant M-smoother according to the numerical results of Rue *et al.* (2002).

The paper is organized as follows. In Section 2, we briefly describe the estimating procedures considered in the simulation study. The procedures for the simulation study and the numerical results are given in Section 3. Section 4 provides the concluding remarks.

2. Discontinuous Regression Function Estimation Procedures

There are many estimation procedures in both the direct and indirect approach. For a fair comparison, the estimation procedures are selected from each approach such that the corresponding estimation procedures are based on the same smoothing technique. As the direct approach, the local constant M-smoother by Chu *et al.* (1998) and the local linear M-smoother by Rue *et al.* (2002) are selected, and as the indirect approach, the modified DKE by Park (2008) and the modified **d**ifference of the two one-sided local **l**inear **k**ernel estimator(DLK) by Park (2009b) are selected. Both the local constant M-smoother and modified DKE are based on the local constant regression, and both the local linear M-smoother and modified DLK are based on the local linear regression. In this section, we briefly describe the above-mentioned four estimation procedures.

Throughout the paper, we assume that $\{(x_i, Y_i), i = 1, \ldots, n\}$ are generated from the 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, and the design points x_i are equally spaced. We also assume that the regression function m can be expressed

by

$$m(x) = f(x) + \sum_{j=1}^{p} d_j I(x > s_j), \tag{2.2}$$

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. The number and the location of jumps are assumed to be unknown, and without a loss of generality we assume that m is defined on the interval $[0, 1]$.

2.1. The direct approach

For each x_i, the local constant M-smoother estimate $\hat{m}_{M0}(x_i)$ is found by taking the local minimizer with respect to α of

$$S(\alpha, x_i) = \sum_{j=1}^{n} K_{h_M}^{M}(x_i - x_j) L_{g_M}^{M}(Y_j - \alpha) \tag{2.3}$$

that is closest to Y_i. Refer to Chu *et al.* (1998) for further details. As a trivial extension, the local linear M-smoother estimate $\hat{m}_{M1}(x_i)$ is found by taking the local minimizer $(\hat{\alpha}, \hat{\beta})$ of

$$S(\alpha, \beta, x_i) = \sum_{j=1}^{n} K_{h_M}^{M}(x_i - x_j) L_{g_M}^{M}(Y_j - \alpha - \beta(x_i - x_j)) \tag{2.4}$$

over (α, β) such that $\hat{\alpha}$ is closest to Y_i. Refer to Rue *et al.* (2002) for further details. 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 $L_{g_M}^{M}(\cdot) = g_M^{-1} L^M(\cdot/g_M)$.

2.2. The indirect approach

The original version of both DKE and DLK can be used to detect the jump locations only when the number of jump points are known. Refer to Qiu (2005) for further details. For the case of unknown number of jump points, Park (2008) and Park (2009b) proposed the modified version of DKE and DLK, respectively.

The modified DKE and DLK procedures are based on the following testing problem

$$H_0 : m_+(x) = m_-(x), \quad \forall \, x \in [0, 1],$$
$$H_1 : m_+(x) \neq m_-(x), \quad \exists \, x \in [0, 1].$$

To state $T_0(x)$, the test statistic for the modified DKE and $T_1(x)$, the test statistic for the modified DLK, the following notation will be used. Set

$$w_{j_1 j_2}^{(l)} = \sum_{i=1}^{n} (x_i - x)^{j_1} K_l^{j_2} \left(\frac{x_i - x}{g} \right), \quad \text{for } j_1 = 0, 1, 2 \text{ and } j_2, \ l = 1, 2.$$

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

Now $T_0(x)$ is defined as follows.

$$T_0(x) = \frac{M_{DKE}(x) - c_1 \hat{m}'(x)}{\sqrt{c_2 \hat{\sigma}^2}}, \tag{2.5}$$

where

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

and

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

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

and

$$c_1 = \frac{w_{11}^{(1)}}{w_{01}^{(1)}} - \frac{w_{11}^{(2)}}{w_{01}^{(2)}},$$

$$c_2 = \frac{w_{02}^{(1)}}{\left(w_{01}^{(1)} \right)^2} + \frac{w_{02}^{(2)}}{\left(w_{01}^{(2)} \right)^2}.$$

$T_1(x)$ is also defined as follows.

$$T_1(x) = \frac{\hat{m}_+(x) - \hat{m}_-(x)}{\sqrt{\mathrm{Var}\left(\hat{m}_+(x) - \hat{m}_-(x) \right)}}, \tag{2.6}$$

where

$$\hat{m}_+(x) = \sum_{i=1}^{n} \frac{w_{21}^{(1)} - w_{11}^{(1)}(x_i - x)}{w_{01}^{(1)} w_{21}^{(1)} - \left(w_{11}^{(1)} \right)^2} Y_i K_1 \left(\frac{x_i - x}{g} \right),$$

$$\hat{m}_-(x) = \sum_{i=1}^{n} \frac{w_{21}^{(2)} - w_{11}^{(2)}(x_i - x)}{w_{01}^{(2)} w_{21}^{(2)} - \left(w_{11}^{(2)} \right)^2} Y_i K_2 \left(\frac{x_i - x}{g} \right)$$

and

$$\mathrm{Var}\left(\hat{m}_+(x) - \hat{m}_-(x) \right) = V_1 + V_2,$$

where

$$V_i = \frac{\sigma^2 \left[\left(w_{21}^{(i)} \right)^2 w_{02}^{(i)} - 2 w_{21}^{(i)} w_{11}^{(i)} w_{12}^{(i)} + \left(w_{11}^{(i)} \right)^2 w_{22}^{(i)} \right]}{\left(w_{01}^{(i)} w_{21}^{(i)} - \left(w_{11}^{(i)} \right)^2 \right)^2}, \quad i = 1, 2.$$

Under H_0, the asymptotic distribution of both T_0 and T_1 is the standard normal distribution, so if $|T_i(x)| \geq z_{1-\alpha/2}, i = 1, 2$, for any $x \in [0, 1]$ then we can reject H_0 where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

The jump detection procedures for the two above-mentioned methods are just identical. The candidates of jump locations are found by taking $\hat{s}_j$ as the maximizer of $|T_\bullet(x)|$ over the set A_j where $T_\bullet$ represents T_0 or T_1, and

$$A_j = [g, 1-g] - \bigcup_{k=1}^{j-1} [\hat{s}_k - g, \hat{s}_k + g], \tag{2.7}$$

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

$$|T_\bullet(\hat{s}_j)| \geq z_{1-\frac{\alpha}{2}}, \quad j = 1, \ldots, \hat{p}, \tag{2.8}$$

and then the estimated jump locations are given by $\hat{s}_j, j = 1, \ldots, \hat{p}$.

After the jump locations are estimated, traditional smoothing procedures with the bandwidth h are applied to each B_j separately to obtain the discontinuous regression function estimate where

$$B_j = [\hat{s}_{j-1}, \hat{s}_j], \quad j = 1, \ldots, \hat{p}+1. \tag{2.9}$$

Here $\hat{s}_0 = 0$ and $\hat{s}_{\hat{p}+1} = 1$. For the modified DKE estimate, $\hat{m}_{DKE}(x)$, local constant regression is used, and local linear regression is used for the modified DLK estimate, $\hat{m}_{DLK}(x)$.

3. Simulation

A simulation study was conducted to evaluate the finite sample properties of four estimation procedures described in Section 2. We considered the following discontinuous regression models:

$$m_1(x) = \cos(8\pi(0.5 - x))I(x \leq 0.5) - \cos(8\pi(0.5 - x))I(x > 0.5),$$

$$m_2(x) = \begin{cases} -3x + 1, & x \leq 0.3, \\ -3x + 3 - \sin\dfrac{(x - 0.3)\pi}{0.2}, & 0.3 < x \leq 0.7, \\ \dfrac{x}{2} + 2.55, & x > 0.7, \end{cases}$$

$$m_3(x) = \sin(5.5\pi) + \sum_{j=1}^{7} d_j I(x > s_j),$$

where $\{s_j\}_{j=1}^{7} = \{0.2, 0.275, 0.55, 0.6, 0.7, 0.71, 0.825\}$ and $\{d_j\}_{j=1}^{7} = \{4, -2, -1.75, 2, -3, 3, -1.75\}$. The design points, x_i's were evenly spaced from 0 to 1 with a sample size of $n = 200$, and Y_i's were generated from model (2.1) with $\epsilon_i \sim N(0, 0.2^2)$.

The left panels of Figure 3.1 depict the three discontinuous regression functions. Note that m_1 has one jump point at the position 0.5 with jump size 2, and m_2 has two jump points at the positions 0.3 and 0.7 with the same jump magnitude 2. The function m_1 and m_2 was considered in the simulation study of Gijbels *et al.* (2007). The function m_3 is a sine function broken by 7 jump points at the positions 0.2, 0.275, 0.55, 0.6, 0.7, 0.71 and 0.825 with jump magnitudes 4, -2, -1.75, 2, -3, 3 and -1.75 respectively. This function was considered in Rue *et al.* (2002). The right panels of Figure 3.1 present the typical data sets of each regression model along with the true regression curves.

$S(\alpha, x_i)$ of (2.3) and $S(\alpha, \beta, x_i)$ of (2.4) are too complicated to find the closed form for their local minimum, so numerical algorithm must be implemented. For the local constant M-smoother, Chu
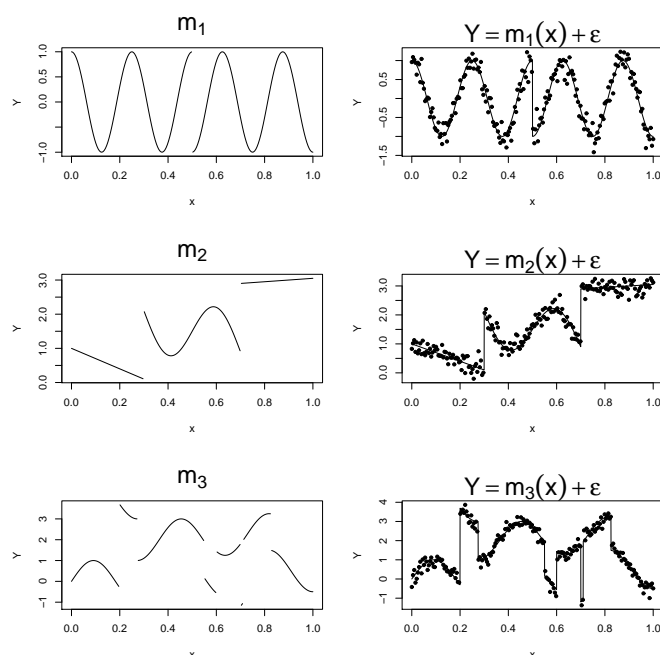
Figure 3.1. The true discontinuous regression functions and the typical data set of each model

et al. (1998) provide numerical algorithm based on Newton's method to find the local minimum closest to Y_i. For the local linear M-smoother, Rue *et al.* (2002) suggest numerical algorithm that considers all possible roots and then chooses the correct one based on Newton's iteration algorithm to find local minimizer $\hat{\alpha}$ of $S(\alpha, \beta, x_i)$ that is closest to Y_i. However, these methods are complicated to program. Simpson *et al.* (1998) propose the iterative weighted least squares (IWLS) algorithm for the local constant M-smoother, and Burt (2000) suggest the IWLS algorithm for the local linear M-smoother. Although there is no guarantee that the IWLS algorithm will converge to the correct local minimum, Simpson *et al.* (1998) and Burt (2000) found that using reasonable bandwidth and choosing Y_i as the starting value for the intercept parameter helps IWLS algorithm converge to the correct root. Thus, we adopted the IWLS algorithm for both local constant and local linear M-smoother.

For both kernel function K^M and L^M of (2.3) and (2.4), we chose the Gaussian density function. For the bandwidth g_M, Burt (2000) recommends to take $g_M = 2.11\sigma$. The number 2.11 is determined by Asymptotic Relative Efficiency (ARE) calculation. For the estimator of σ^2, we used 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, and we choose $\nu = 2$. The bandwidth h_M works more like a traditional smoothing parameter, so the data adaptive bandwidth selection methods like cross-validation or a plug-in rule can be used; however, the performance of these methods for choosing h_M has not been thoroughly investigated yet. Thus, we used several different values of h_M in the simulation.

For both the modified DKE and DLK procedure, the bandwidth g and h are also very important factors. The bandwidth g is used for the jump detection procedure of (2.7), but limited study has

Table 3.1. The smallest $\widehat{\text{MISE}}$ values along with their bandwidths based on 1000 replications

| | Direct Approach | | | | Indirect Approach | | | | |
| | M0 | | M1 | | M-DKE | | M-DLK | | Ratio |
	h_M	$\widehat{\text{MISE}}$	h_M	$\widehat{\text{MISE}}$	h	$\widehat{\text{MISE}}$	h	$\widehat{\text{MISE}}$	
m_1	0.01	0.0287	0.035	0.0232	0.02	0.0075	0.035	0.0062	3.71
m_2	0.045	0.0226	0.06	0.0195	0.035	0.0049	0.055	0.0038	5.02
m_3	0.015	0.0215	0.025	0.0185	0.01	0.0322	0.01	0.0297	0.62

been done about the selection method of g. Gijbels and Goderniaux (2004) propose the bootstrap based data-driven bandwidth selection method; however, their method is complicated and it would not be a good idea to use their method in practice. Thus, the choice of the proper bandwidth g is problematic. However, according to the numerical investigation of Park (2008) and Park (2009a), the larger bandwidth seems to produce the better results for both $\hat{m}_{DKE}(x)$ and $\hat{m}_{DLK}(x)$. We can observe the same empirical evidence in both Wu and Chu (1993a) and Bowman and Pope (2006). Without proper theoretical support, it is dangerous to generalize the empirical evidence, but we cannot help selecting the value of g as large as possible since we do not have alternatives. By the jump detection procedure of (2.7), the discontinuity points only in the interval $[g, 1-g]$ can be detected, so too large value of g should be avoided. We set $g = 0.15$ for both $\hat{m}_{DKE}(x)$ and $\hat{m}_{DLK}(x)$ in the simulation; however, this choice is rather arbitrary.

The bandwidth h is used for local constant regression and local linear regression that are applied to each B_j of (2.9). We could use the cross-validation procedure proposed by Wu and Chu (1993b); however, we used several different values of h just like the case of h_M.

For the kernel function of $T_0(x)$ and $T_1(x)$, we chose $K_1(x) = 1.5(1 - x^2)I_{[0,1]}(x)$ and $K_2(x) = K_1(-x)$ for all x. The first derivative estimate of regression function $\hat{m}'(x)$ in $T_0(x)$ of (2.5) was evaluated by the function *glkerns* of the package *lokern* in R. The computation of the local constant regression estimate and the local linear regression estimate for each B_j of (2.9) was done by R function *locfit*. The significance level for the procedure (2.8) was set to $\alpha = 0.05$.

Since our main objective is to obtain good overall estimation of regression function m, the natural choice of the criterion for the comparison is the Mean Integrated Squared Error, MISE $= E[\int_{0.1}^{0.9} (\hat{m}(x) - m(x))^2 dx]$. In order to avoid boundary effects, the MISE was calculated at the interval [0.1,0.9]. For estimating 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 the use of the following local MISE:

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

that measures the MISE between $\hat{m}(x)$ and $m(x)$ in the interval $[s - 0.05, s + 0.05]$.

For m_1 and m_2, we estimated both MISE and local MISE. For m_3, we did not estimate the local MISE because some jump points are located too close together, so the local MISE does not have much meaning as a criterion for comparison. We estimated both MISE and local MISE based on 1000 replications for various h_M and h values, and report the smallest $\widehat{\text{MISE}}$ of each estimation procedure along with the corresponding bandwidth in Table 3.1. We also report the smallest $\sum_s \widehat{\text{MISE}}_s$ of each estimation procedure along with the corresponding bandwidth in Table 3.2. In these tables, 'M0' and 'M1' stand for the local constant M-smoother and the local linear M-smoother respectively, and 'Ratio' is the ratio of the estimated (local) MISE of the direct approach to the

Table 3.2. The smallest $\sum_s \widehat{\text{MISE}}_s$ values along with their bandwidths based on 1000 replications

| | Direct Approach | | | | Indirect Approach | | | | |
| | $M0$ | | $M1$ | | M-DKE | | M-DLK | | Ratio |
	h_M	$\sum_s \widehat{\text{MISE}}_s$	h_M	$\sum_s \widehat{\text{MISE}}_s$	h	$\sum_s \widehat{\text{MISE}}_s$	h	$\sum_s \widehat{\text{MISE}}_s$	
m_1	0.01	0.0034	0.03	0.0031	0.025	0.0006	0.04	0.0007	4.76
m_2	0.02	0.0064	0.025	0.0062	0.025	0.0013	0.055	0.0010	5.84

estimated (local) MISE of the indirect approach. In each approach the smaller estimated (local) MISE was selected for computing the ratio.

Table 3.1 shows that the local linear based estimation methods have a smaller $\widehat{\text{MISE}}$ than the local constant based estimation methods in both the direct and indirect approach; this is what we expected. Because of the shapes of m_1 to m_3, the local linear fit should be preferable in all cases.

By the criterion of the overall good estimation of the regression function, the modified DLK procedure shows the best performance in both m_1 and m_2. The modified DLK procedure has overwhelming superiority over the local linear M-smoother in these cases. However, for m_3 the local linear M-smoother shows better performance than the modified DLK procedure. Depending on the shape of the true regression curves, we obtain quite different results. The function m_3 has a lot of jump points and some of them are located close together. When the adjacent jump points are too close, the jump detection procedure of (2.7) cannot detect all of them. For the indirect approach, the good estimation of jump locations is indispensable for getting a good estimation of regression curves. Thus, it seems that the inferior performance of the indirect approach in m_3 is caused by the limitation of the jump detection procedure.

In Table 3.2, we can check the jump-preserving performance around the given jump points of each method. Under the criterion of the jump-preserving property, it turns out that the local constant fit has a competitive advantage over the local linear fit in both the direct and indirect approach. This is a small surprise since the local constant fit is not preferable to the local linear fit in the estimation of the overall regression function. We also can see that the indirect approach shows the better performance than the direct approach in both m_1 and m_2.

4. Conclusion

The major goal of the indirect approach is not the estimation of the regression curve but the estimation of jump locations; therefore, it is actually presumed that the direct approach performs better than the indirect approach for obtaining the overall regression curve estimate. The point is whether the subsidiary result of the indirect approach is as good as the main result of the direct approach. According to the simulation results, it seems to be verified that the indirect approach is competitive method for estimating the discontinuous regression curve itself. The indirect approach outperforms the direct approach when there is only one jump point or the adjacent jump points are located some distance away.

There are many existing methods in both the direct and indirect approach, but a thorough comparison of the two approaches as the estimator of the discontinuous regression function itself has drawn limited attention. However, according to the simulation results, the indirect approach has a competitive advantage over the direct approach in some situations. An extensive numerical study (that includes all existing estimation procedures) is needed to form a more concrete conclusion.

References

Bowman, A .W. and Pope, A. (2006). Detecting discontinuities in nonparametric regression curves and surfaces, *Statistics & 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. and Goderniaux, A. C. (2004). Bandwidth selection for change point estimation in nonparametric regression, *Technometics*, **46**, 76–86

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.

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

Park, D. (2009a). Comparison of Jump-Preserving Smoothing and Smoothing based on Jump Detector, *Communications of the Korean Statistical Society*, **16**, 519–528.

Park, D. (2009b). Comparison of finite sample properties of jump detectors, Unpublished Manuscript.

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

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 processing by Chu, Glad, Godtliebsen, and Marron, *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.