Bandwidth Selection for Local Smoothing Jump Detector

Dongryeon Park[1,a]

[a]Department of Information and Statistics

Abstract

Local smoothing jump detection procedure is a popular method for detecting jump locations and the performance of the jump detector heavily depends on the choice of the bandwidth. However, little work has been done on this issue. In this paper, we propose the bootstrap bandwidth selection method which can be used for any kernel-based or local polynomial-based jump detector. The proposed bandwidth selection method is fully data-adaptive and its performance is evaluated through a simulation study and a real data example.

Keywords: Bandwidth, bootstrap, discontinuous regression, local smoothing jump detection.

1. Introduction

Nonparametric regression analysis for the estimation of the discontinuous regression function has received much attention over the years. Quite a number of researches on this topic have been done. These researches may be classifies into two categories. The first approach, which is usually called jump-preserving smoothing, estimates the regression function directly without detecting the jump points explicitly. Reference on this approach includes Chu *et al.* (1998), Rue *et al.* (2002), Polzehl and Spokoiny (2000), Qiu (2003) and Gijbels *et al.* (2007), among others.

The second approach estimates the location of jump points first using one of various jump detection procedures and then applies the ordinary smoothing technique to each smooth part of regression curve separately. In this approach, many of jump detectors are based on kernel-based estimation methods or local polynomial estimation methods. The literature on this approach includes Müller (1992), Wu and Chu (1993), Qiu *et al.* (1991), Gijbels *et al.* (1999) and Park (2008), among others.

The major issue of the second approach is to obtain good estimates of jump locations. Once we have correct estimates of jump locations, we have a good chance of getting accurate estimate of regression function. Thus excellent performance of jump detector is crucial and it is quite clear that if the jump detector is based on kernel-based or local polynomial-based method, then the performance of the jump detector heavily depends on the choice of the bandwidth. Therefore, the bandwidth selection method for jump detector is really important issue for the second approach, but little work has been done on this issue. Gijbels and Goderniaux (2004) proposed the bootstrap procedure for selecting bandwidth for the jump detection procedure of Gijbels *et al.* (1999), but their method is rather complicated.

In this paper, we propose the bandwidth selection method which can be used for any kernel-based or local polynomial-based jump detector. Our bandwidth selection method is based on the bootstrap, but it is much simpler than Gijbels and Goderniaux (2004).

2. Bandwidth Selection Method

2.1. New measure of performance for local smoothing jump detector

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. 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]. Throughout the paper, we assume $p = 1$. The extension of the proposed bandwidth selection method to the case of unknown p seems straightforward, but we do not deal with it in this paper.

The common procedure of the local smoothing jump detector is to fit nonparametrically smooth curves to the left and to the right of given point, and the basic information of the presence of a jump discontinuity is contained in the difference between the left fit and the right fit. For example, the jump detector proposed by Park (2009) is defined by standardizing the difference as follows.

$$T(x) = \frac{\hat{m}_+(x) - \hat{m}_-(x)}{\sqrt{\text{Var}\,(\hat{m}_+(x) - \hat{m}_-(x))}} \tag{2.2}$$

where $\hat{m}_+(x)$ and $\hat{m}_-(x)$ is the local linear right-sided estimator and the local linear left-sided estimator by using the data lying to right and left of x, respectively. Here x should be located inside of $[h, 1-h]$, and it means that the local smoothing jump detection procedures have the "boundary problem" in the sense that they can not detect jumps in the border regions of the design interval which are $[0, h]$ and $[1-h, 1]$. Now the estimator $\hat{s}_1$ of jump location s_1 is defined as

$$\hat{s}_1 = \arg \max_{x \in [h, 1-h]} |T(x)|$$

It is common to measure the closeness of the estimator $\hat{s}_1$ to its target parameter s_1 by $\text{MSE}(\hat{s}_1) = E(\hat{s}_1 - s_1)^2$. The possible approach for the choice of the value of bandwidth h is based on the analog of the MSE. The jump detector proposed by Wu and Chu (1993) is based on the Gasser-Müller estimator. They derived the asymptotic mean squared error(AMSE) of $\hat{s}_1$ and tried to obtain the optimal bandwidth by minimizing AMSE, but they failed. Besides, they assumed that $s_1 \in [\delta, 1 - \delta]$ where $\delta > 0$ is arbitrary small constant. We do not have the exact expression of the AMSE($\hat{s}_1$) for any other existing jump detectors including $T(x)$ of (2.2) yet, so it seems that the asymptotic approach for the choice of the value of h is not available for now.

The alternative to the asymptotic approach may be Monte Carlo simulation. We investigate the finite sample properties of MSE($\hat{s}_1$). Figure 1 displays the Monte Carlo MSE($\hat{s}_1$) for various h based on 2000 repeated samples for the following model.

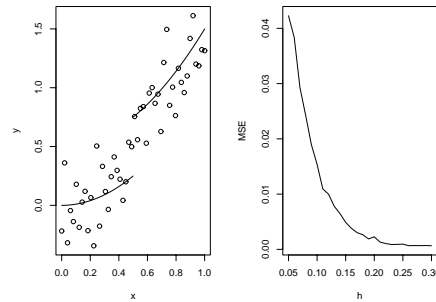$$Y_i = x_i^2 + 0.5 \cdot I(x_i \geq 0.5) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.3}$$

Figure 1: *Left panel: typical data set and true regression curve of model (2.3). Right panel: Monte Carlo MSE($\hat{s}_1$) for various h.*

where $n = 50$, $x_i = (i - 1)/(n - 1)$, and ϵ_i's are independent and normally distributed with mean 0 and standard deviation $\sigma = 0.2$. As we can see, the Monte Carlo MSE($\hat{s}_1$) is a decreasing function of h. We have tried several different models and we have found that the pattern of Figure 1 is universal. Thus according to the Monte Carlo approach, we should choose the value of h as large as possible, but this is a kind of contradictory result with Wu and Chu (1993). Even though they failed to derive the optimal bandwidth, they argued that the optimal value of h is of the minimum order of the value of h satisfying some regularity condition.

This contradiction may be caused by the different assumption about the border region. In Wu and Chu (1993), the border regions are $[0, \delta]$ and $[1 - \delta, 1]$, so the length of the border region is fixed. However, in our paper, the length of border region is increasing with h. That is to say, we assume that the local smoothing jump detector can detect the jump points only if they are located inside of $[h, 1 - h]$. Thus it is always true that $h \leq \hat{s}_1 \leq 1 - h$, so large value of h reduces the range of $\hat{s}_1$ and it has a direct effect on large reduction of the Monte Carlo Var($\hat{s}_1$) and this also entails large reduction of the Monte Carlo MSE($\hat{s}_1$) as long as $s_1 \in [h, 1 - h]$. Therefore, we should penalize the large value of h for computing MSE($\hat{s}_1$). In this point of view, we propose the new measure of performance for the local smoothing jump detector. We define the penalized mean squared error(PMSE) as follows.

$$\text{PMSE} = \log\left[E(\hat{s}_1 - s_1)^2\right] + \frac{\lambda}{\log(n)(1 - 2h)}. \tag{2.4}$$

In the second term of PMSE, the constant λ controls the amount of penalization, $(1 - 2h)$ is the range of $\hat{s}_1$, and $\log(n)$ is used as a kind of weight. The choice of h usually depend on n. When n is small we usually prefer to select large value of h in order to get more information. Thus, by using $\log(n)$ as a weight, we can give more penalty to the case of smaller n for selecting larger value of h.

We do not obtain any theoretical results about PMSE yet, but through the Monte Carlo simulation we investigate the finite sample properties of PMSE. With the model (2.3), the Monte Carlo PMSE($\hat{s}_1$) based on 2000 repeated samples for $n = 50$ with 4 different values of λ are displayed in Figure 2. Note that both $\lambda = 10$ and $\lambda = 15$ provide almost identical pattern of the Monte Carlo PMSE($\hat{s}_1$). During the simulation, we have found that the Monte Carlo PMSE($\hat{s}_1$) shows almost identical pattern with quite wide rage of λ, and $\lambda = 10$ always provides satisfactory results. Thus we decided to choose $\lambda = 10$ throughout the paper.

As we can see in Figure 2, the Monte Carlo PMSE($\hat{s}_1$) has a global minimum point. Therefore, it is possible to select the value of h by minimizing PMSE($\hat{s}_1$). Even though we have verified that PMSE($\hat{s}_1$) has a global minimum through simulation study, we are not quite sure yet whether the value of h which minimized PMSE($\hat{s}_1$) is really 'good' bandwidth. This is about the appropriateness

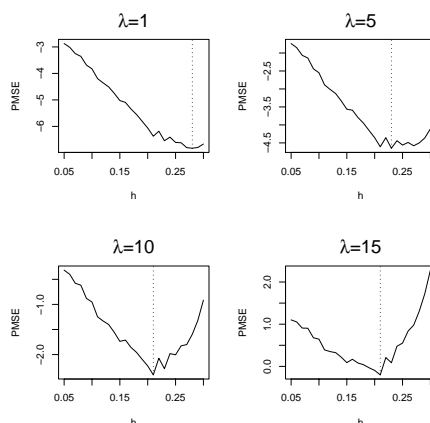Figure 2: *Monte Carlo PMSE($\hat{s}_1$) for various h and λ for model (2.3).*

of PMSE as a measure of performance. Because of above-mentioned boundary problem of local smoothing jump detector, there is no clear way to measure the performance of $\hat{s}_1$. Maybe this is why we do not have well established bandwidth selection method for jump detector yet. In Section 3, we verify the appropriateness of PMSE as a measure of performance of $\hat{s}_1$ numerically. The theoretical investigation needs further study.

2.2. Bootstrap bandwidth selection method

Since PMSE of (2.4) includes the unknown parameter s_1, we should figure out how to estimate PMSE in order to construct the bandwidth selection method based on PMSE. Now the idea is to use a bootstrap procedure to estimate PMSE and to select the bandwidth for which this estimated PMSE is minimal.

Bootstrap algorithm

Step 1 From the observed sample $\chi = \{(x_1, Y_1), \ldots, (x_n, Y_n)\}$ we randomly draw a pair (x_i, Y_i) with replacement to form $\chi^* = \{(x_1^*, Y_1^*), \ldots, (x_n^*, Y_n^*)\}$.

Step 2 With a range of value of h, we compute $\hat{s}_1^*$ for the resample χ^*.

Step 3 With B bootstrap replications of step 1 and 2, we have B values of $\hat{s}_1^*$, denoted by $\hat{s}_1^{(b)}$, $b = 1, \ldots, B$ for each value of h.

Step 4 We compute the bootstrap estimate of PMSE by

$$\text{PMSE}(\hat{s}_1) = \log \left[\frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{s}_1^{(b)} - \overline{\hat{s}_1^{(*)}} \right)^2 + \left(\overline{\hat{s}_1^{(*)}} - \hat{s}_1 \right)^2 \right] + \frac{\lambda}{\log(n)(1-2h)},$$

where $\overline{\hat{s}_1^{(*)}} = 1/B \sum_{b=1}^{B} \hat{s}_1^{(b)}$ and $\hat{s}_1$ is the estimated jump location computed from the original observed sample.

Step 5 We select the value of h for which the bootstrap estimate of PMSE($\hat{s}_1$) is minimal.

With the model (2.3), we compute the bootstrap estimate PMSE($\hat{s}_1$) based on 2000 independent samples for $n = 50$ with $\lambda = 10$ case. In each sample, we set $B = 200$. The results are displayed in Figure 3. Note that PMSE($\hat{s}_1$) is well estimated by the bootstrap procedure.
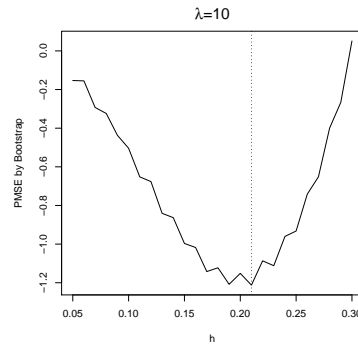
Figure 3: *Monte Carlo bootstrap estimate of PMSE($\hat{s}_1$) for model (2.3).*

Table 1: Simulation results

Interval	$\sigma = 0.15$				$\sigma = 0.2$			
	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
[0.00, 0.15)	0	1	0	0	0	2	2	1
[0.15, 0.25)	0	1	0	0	6	5	3	0
[0.25, 0.35)	0	2	0	967	16	14	10	863
[0.35, 0.45)	3	7	9	17	26	24	26	64
[0.45, 0.55)	986	976	981	13	895	898	909	55
[0.55, 0.65)	3	10	6	2	39	37	37	8
[0.65, 0.75)	5	3	4	0	16	15	10	6
[0.75, 0.85)	3	0	0	1	2	5	3	3
[0.85, 1.00]	0	0	0	0	0	0	0	0
mean of $\hat{s}_1$	0.510	0.509	0.509	0.311	0.507	0.507	0.508	0.329
sd of $\hat{s}_1$	0.020	0.023	0.016	0.032	0.048	0.052	0.045	0.067

3. Numerical Study

In this section, we evaluate the bootstrap bandwidth selection method through a simulation study and provide an example with real data.

3.1. Simulation study

We considered the following discontinuous regression models:

$$m_1(x) = 2 + 0.5I(x \geq 0.5),$$
$$m_2(x) = x + 0.5I(x \geq 0.5),$$
$$m_3(x) = x^2 + 0.5I(x \geq 0.5),$$
$$m_4(x) = \sin(2\pi x) + 1I(x \geq 0.3).$$

The design points were given by $x_i = (i - 1)/(n - 1)$ for $i = 1, \ldots, n$. The sample size considered was $n = 50$. The error terms were generated from $N(0, \sigma^2)$ where $\sigma = 0.15, 0.2$. The typical data sets of each regression model for $\sigma = 0.2$ along with the true regression lines are presented in Figure 4. As a jump detection procedure, we used the jump detector of (2.2). In all studies we considered 1000 simulations and the number of bootstrap replicates was $B = 200$.

The simulation results are summarized in Table 1. The number of times that the estimated values of $\hat{s}_1$ fall in the specific intervals out of 1000 simulation samples are presented. In the two bottom lines of the table we also list the means and standard deviations of $\hat{s}_1$ across the 1000 simulation samples.
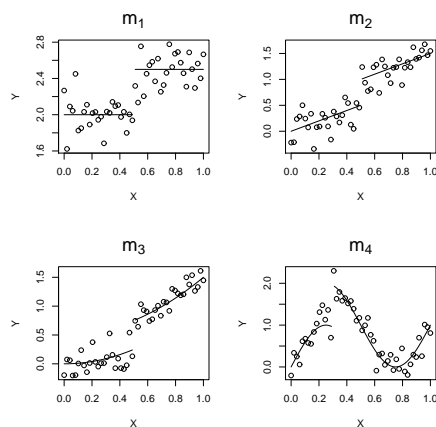
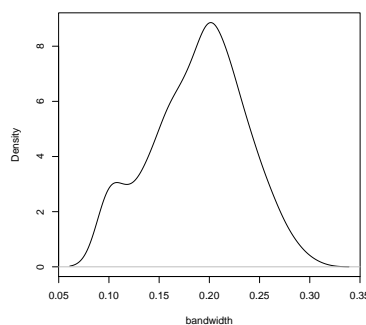Figure 4: *The true regression lines with a typical simulated data set.*

Figure 5: *The kernel density estimate of the bandwidth selected by the bootstrap algorithm based on 1000 simulations for m_3 and $\sigma = 0.2$.*

Even though we consider rather small sample size, $n = 50$ case only, the performance of the proposed fully data-adaptive bandwidth selection method is quite excellent.

For the case of m_3 and $\sigma = 0.2$, we display a graph of a kernel density estimate of 1000 bandwidths selected by the proposed bootstrap algorithm in Figure 5. This graph shows that the density function of bootstrap selected bandwidth is concentrated around 0.2.

3.2. Example with real data

It is well known that the December sea level pressure during 1921–1992 in Bombay, India has a discontinuity point around 1960 and 1961. The small dots in the left panel of Figure 6 represent the sea level pressure. These data are given in Qiu (2005).

The right panel of Figure 6 is the bootstrap estimate of PMSE, so the bootstrap procedure tells us that the optimal bandwidth is $h = 13$. The jump detector of (2.2) with $h = 13$ indicates that the abrupt change occurs around 1961, and the final estimate of regression curve is represented by the solid curves in the left panel of Figure 6. It seems that the proposed bandwidth selection method works quite well in this example.
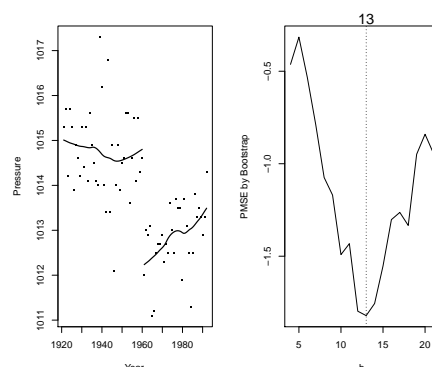
Figure 6: *The December sea level pressure during 1921 to 1992 in Bombay, India. Left: The small dots represent sea level pressure and the solid curves represent nonparametric estimate. Right: The bootstrap estimate of PMSE for various h.*

4. Conclusion

When we want to estimate jump discontinuity locations of the unknown regression function, local smoothing jump detection procedure is very popular. It is quite clear that the performance of any kernel-based or polynomial-based jump detector heavily depends on the choice of the bandwidth. Thus the bandwidth selection method for local smoothing jump detector is the important issue, but little attention has been paid so far.

In this paper, we proposed the bootstrap bandwidth selection method for local smoothing jump detector. Through the simulation and real data example we investigated the performance of the fully data-adaptive bandwidth selection method, and it seems that the proposed method performs very well.

We defined the PMSE as the measure of performance for local smoothing jump detector and investigated the empirical properties only. Theoretical investigation about the appropriateness of PMSE should be done, and it needs further study.

Acknowledgment

I gratefully thank the two referees for their many valuable comments which substantially improved the presentation.

References

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., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves, *Annals of the Institute of Statistical Mathematics*, **51**, 231–251.

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.

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.

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

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

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

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.

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.