Journal of Statistical
Theory & Methods
1994, Vol. 5, pp11 ~ 20

Variable Bandwidth Selection for Kernel Regression [1]

Daehak Kim [2]

ABSTRACT

In recent years, nonparametric kernel estimation of regresion function are abundant and widely applicable to many areas of statistics. Most of modern researches concerned with the fixed global bandwidth selection which can be used in the estimation of regression function with all the same value for all x. In this paper, we propose a method for selecting locally varing bandwidth based on bootstrap method in kernel estimation of fixed design regression. Performance of proposed bandwidth selection method for finite sample case is conducted via Monte Carlo simulation study.

1. INTRODUCTION

Let $Y_1, Y_2, \cdots, Y_n$ be observations on the unknown regression function $\theta(\cdot)$ with a model

[1]This paper was supported in part by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1992

[2]Department of Statistics, College of Natural Sciences, Hyosung Women's University

$$Y_i = \theta(x_i) + \epsilon_i, \quad i = 1, \cdots, n \tag{1}$$

where ϵ_i are iid from unknown F with mean 0 and finite variance σ^2 and x_i are fixed design points. Without loss of generality, we assume the regression function θ is defined on the closed interval $[0, 1]$ and x_i are equally spaced.

We consider Nadaraya (1964) and Watson (1964) type kernel estimate of regression function of the form

$$\widehat{\theta}_n(x : h) = \sum_{i=1}^{n} \frac{K((x - x_i)/h)}{\sum_{j=1}^{n} K((x - x_j)/h)} Y_i \tag{2}$$

where $K(\cdot)$ is kernel and global bandwidth h represents over all amount of smoothing which depends on n and tends to 0 as $n \to \infty$ but $nh \to \infty$.

It is well known that bandwidth selection is a crucial problem. The main reason is that a small bandwidth yields large variance while large bandwidth yields large bias of the kernel estimate. Various proposals for appropriate selection of global bandwidth have been made. Wong (1983) had shown the strong consistency of cross validated bandwidth for (2). Bootstrap method for the selection of global bandwidth for kernel regression was also developed by Kim(1993).

However the kernel estimate of the form (2) exhibit an increased bias near peaks of θ. Since the bias near peaks may contribute especially a large portion to mean squared error of the estimate (MSE), MSE might be reduced by decreasing the bandwidth near peaks and increasing the bandwidth in flat parts of the curve. When estimating the regression function at a particular point, it would be helpful to use variable bandwidth (local bandwidth) for the nature of the data. For example, near a peak a relatively small value of smoothing parameter is appropriate, wheras on an approximately linear section, a large value of smoothing should be used.

So, we consider the choice of locally varying bandwidths. The corresponding kernel estimate is

$$\widehat{\theta}_n(x : h_x) = \sum_{i=1}^{n} \frac{K((x - x_i)/h_x)}{\sum_{j=1}^{n} K((x - x_j)/h_x)} Y_i \tag{3}$$

where h_x is local bandwidth satisfing $nh_x \to \infty$ but $h_x \to 0$ as $n \to \infty$.

In this paper, we propose a data dependent local bandwidth selection method based on bootstrap method for the kernel estimate of the form (3).

For bootstrap method, we estimate the quantity,

$$E[\widehat{\theta}_n(x : h_x) - \theta(x)]^2 \tag{4}$$

for $\forall x$ using bootstrap method via Monte Carlo simulation and find data dependent bandwidth h_x which minimise the bootstrap estimate of (4). To study the performance of proposed local bandwidth selections, Monte Carlo simulation is conducted.

2. LOCAL BANDWIDTH SELECTION

2.1 Preliminaries

In this section, we propose a local bandwidth selection methods based on bootstrap method. We assume the following resonalbly mild conditions.

Assumptions

1. $h_x \to 0$ and $nh_x \to \infty$ as $n \to \infty$
2. The kernel K is symmetric with finite support $[-A, A]$
3. $\theta(x)$ is continuous and $\theta''(x)$ exists

Under the assumption 1.,2., and 3. we have the limiting bias and variance

$$E_F\widehat{\theta}_n(x : h_x) = \theta(x) + \frac{h_x^2}{2}K_2(x)\theta''(x) + o(h_x^2) \tag{5}$$

$$\mathrm{Var}_F\widehat{\theta}_n(x : h_x) = \frac{1}{nh_x}\sigma^2 \int K^2(x)dx + o(\frac{1}{nh_x}) \tag{6}$$

where $K_2 = \int x^2K(x)dx$. These asymptotic expression indicate that appropriate choice of local bandwidth h_x should be influenced by $\theta''(x)$. When $|\theta''(x)|$ is large, since the bias is greatly affected by $\theta''(x)$, small values of h_x are required to keep the bias low, whereas when $|\theta''(x)|$ is small, large values of h_x are appropriate to deflate variance. Variable bandwidth selection method aims to balance these effects in a way that is appropriate for each particular location.

By simple calculation, we can get locally asymptotic optimal bandwidth by

$$\begin{aligned} h_x^{omm} &= n^{-1/5}\sigma^{2/5}[\int K^2(x)dx]^{1/5}[\theta''(x)]^{-2/5}K_2^{-2/5} \\ &= cn^{-1/5} \end{aligned}$$

in MSE sense. Of course, we don't know σ and θ'', but rate of h_x.

Since the variance of $\widehat{\theta}_n(x : h_x)$ converges to 0 at the rate $o(1/nh_x)$, we consider the normalized process

$$Z_n(x : cn^{-1/5}) = n^{2/5}(\widehat{\theta}_n(x : cn^{-1/5}) - \theta(x)). \tag{7}$$

For the asymptotic distribution of $Z_n(x : cn^{-1/5})$, see Kim(1993). It is important to know that if h_x is chosen to balance the bias and standard deviation of $\widehat{\theta}_n(x, h_x)$, then the variance and squared bias will have the same rate of convergence. Therefore it is necessary to ensure that this behavior is mirrored in the distribution of bootstrap estimator.

2.2 Bootstrapping kernel regression

The bootstrap method in kernel regression function estimation is to estimate MSE from the given sample and then find the bandwidth minimising the bootstrap estimate of MSE. But it requires the initial regression function estimate $\widehat{\theta}_n(x : h_0)$ of $\theta(x)$ for the residual estimation. So we take cross-validatory bandwidth h_0 as an objective initial estimate of global bandwidth h. Hardle and Bowman(1988) considered bootstrapping kernel regression for Priestly and Chao(1972) estimator

$$\widehat{\theta}_n(x : h_x) = n^{-1}h^{-1}\sum_{i=1}^{n} K((x - x_i)/h)y_i$$

They used kernel estimate of $\theta''(x)$ in the estimation of bias and considered bias corrected bootstrapping. Moreover they discarded estimated large residuals in resampling. We, in this paper, consider bootstrapping kernel regression for the original process. It can be another method.

Let $\widehat{\theta}_n(x)$ be some initial regression function estimate of $\theta(x)$ with a given data set $Y_1, Y_2, \cdots, Y_n$. From this initial estimate of regression function, we can get an estimated residuals $\widetilde{\epsilon}_i$ by $\widetilde{\epsilon}_i = Y_i - \widehat{\theta}_n(x_i)$, $i = 1, 2, \cdots, n$ and let $\widehat{\epsilon}_i$ be centered residual of $\widetilde{\epsilon}_i$. That is, $\widehat{\epsilon}_i = \widetilde{\epsilon}_i - \frac{1}{n}\sum_{i=1}^{n}\widetilde{\epsilon}_i$, $i = 1, 2, \cdots, n$

Now let $\widehat{F}_n$ be empirical distribution function of centered residuals $\widehat{\epsilon}_i$ and let $\widehat{\epsilon}_1^*, \widehat{\epsilon}_2^*, \cdots, \widehat{\epsilon}_n^*$ be i.i.d. sample from the $\widehat{F}_n$ i.e. by resampling with replacement from the empirical distribution $\widehat{F}_n$ of $\widehat{\epsilon}_i$. From these conditionally independent random variable $\widehat{\epsilon}_i^*$, we can get a new data set Y_i^* by

$$Y_i^* = \widehat{\theta}_n(x_i) + \widehat{\epsilon}_i^* \quad i = 1, 2, \cdots, n$$

With those Y_i^* , we construct the bootstrapped regression function estimate

$$\widehat{\theta}_n^*(x : h_x) = \sum_{i=1}^{n} \frac{K((x - x_i)/h_x)}{\sum_{i=1}^{n} K((x - x_i)/h_x)} Y_i^*.$$

The following theorem is helpful in understanding bootstrapping.

Theorem For fixed $c > 0$ and with the assumptions 1.,2. and 3.,

$$Z_n^*(x : cn^{-1/5}) = n^{2/5}(\widehat{\theta}_n^*(x : cn^{-1/5}) - \widehat{\theta}_n(x))$$

has the same asymptotic distribution of $Z_n(x : cn^{-1/5})$.

Proof. See Kim(1993)

Remark 1. The limiting bias and variance of bootstrapped regression function estimate $\widehat{\theta}_n^*(x : h_x)$ are the same with those of asymtotics (5),(6). It means that the behaviour of bootstrap distribution is similar to that of true asymptotic distribution. So, we can expect that large values of h_x is obtained in flat part whereas small values of h_x are obtained in peak part of the function.

Corollary Let c^θ and c^b be the optimal choices minimising the limits

$$\lim_{n\to\infty} EZ_n(x : cn^{-1/5}) \quad \text{and} \quad \lim_{n\to\infty} E^*Z_n^*(x : cn^{-1/5})$$

respectively where E^* denotes conditional expectation under $\widehat{F}_n$, then by the theorem

$$\lim_{n\to\infty} n^{2/5}E[\widehat{\theta}_n(x : c^\theta n^{-1/5}) - \theta(x)]^2 = \lim_{n\to\infty} n^{2/5}E[\widehat{\theta}_n(x : c^b n^{-1/5}) - \theta(x)]^2$$

Remark 2. By the Corollary, we can get a data dependent bandwidth not a asymptotic one which has asymptotic optimal properties.

2.3 Local bandwidth choice

Based on the consistency of bootstrap method, we can get the bootstrap estimates of $MSE = E(\widehat{\theta}_n(x : h_x) - \theta(x))^2$. Let $\widehat{\epsilon}_1^*, \widehat{\epsilon}_2^*, \cdots, \widehat{\epsilon}_n^*$ be i.i.d. samples from the $\widehat{F}_n$ where $\widehat{F}_n$ is empirical distribution function of estimated and centered residuals from the initial regression estimates. With these residuals, we obtain the resampled Y_i^* and construct the bootstrapped regression function estimate

$$\widehat{\theta}_{n:j}^*(x : h_x) = \sum_{i=1}^{n} \frac{K((x - x_i)/h_x)}{\sum_{i=1}^{n} K((x - x_i)/h_x)} Y_i^* \quad \text{for} \quad j = 1, 2, \cdots, B.$$

Then, bootstrap estmates of MSE would be

$$\begin{aligned} BMSE &= \int (\widehat{\theta}_n^*(x : h_x) - \widehat{\theta}_n(x : h_0))^2 d\widehat{F}_n \\ &= \frac{1}{B}\sum_{j=1}^{B} [\widehat{\theta}_{n:j}^*(x : h_x) - \widehat{\theta}_n(x : h_0)]^2 \end{aligned}$$

where h_0 is the initial global bandwidth chosen by cross-validation method.

So we can get the local bandwidth h_x for $\forall x$ by minimising $BMSE$.

3. SIMULATION STUDY

In this chapter, performance of proposed local bandwidth selection methods for fixed sample size is conducted thru Monte Carlo simulation. For simulations, the regression function estimate $\widehat{\theta}_n(x : h_x)$ of the form (3) is considered.

Our test regression function considered were the following two different types of functions.

$$\theta_1(x) = \sin(4\pi x) \text{ (large } \theta''(x), \text{ periodic)}$$
$$\theta_2(x) = h(0.25, 0.05) + h(0.5, 0.1) \text{ (bimodal function)}$$

where $h(\mu, \sigma) = \exp(-(x-\mu)^2/2\sigma^2)/\sigma\sqrt{2\pi}$. For reasons of computational efficiency, we used the Epanechnikov kernel (1969)

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{for } |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

These curves were computed on [0,1] with sample size n=100. Errors were generated from the standard normal distribution and transformed to have standard deviation 0.1 and 0.8 respectively. Due to the computational limit, we only considered standard normal distribution as an error distribution. All computation was carried out by Workstation(Sun-10) and random numbers were generated thru subroutine RNNOR and RNUND in IMSL.

$B = 100$ bootstrap sample were used. Any increase in the number of bootstrap samples could only improve the estimates. We had 100 replications for each test function.

Figure 1. Estimated curve for $\theta_1(x)$

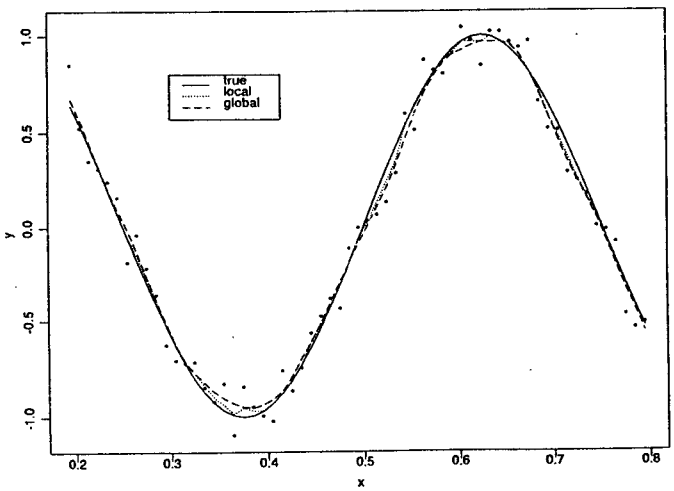


Figure 2. Bandwidth plot for $\theta_1(x)$

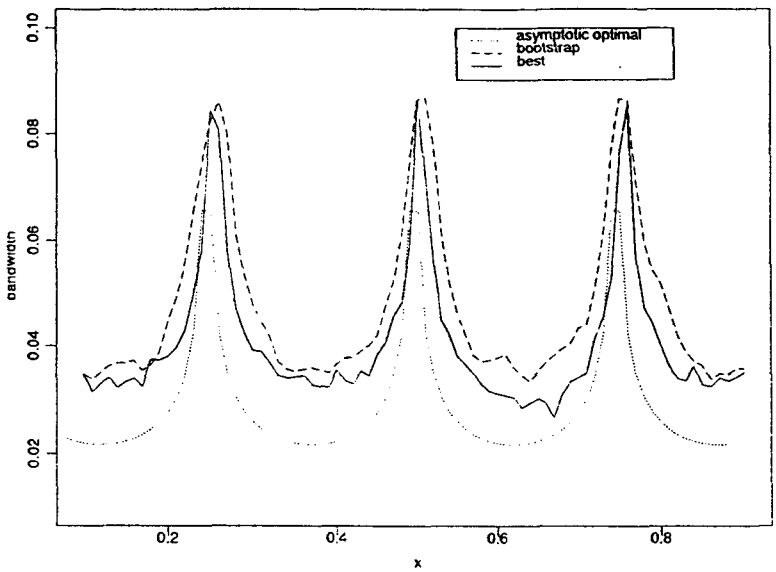
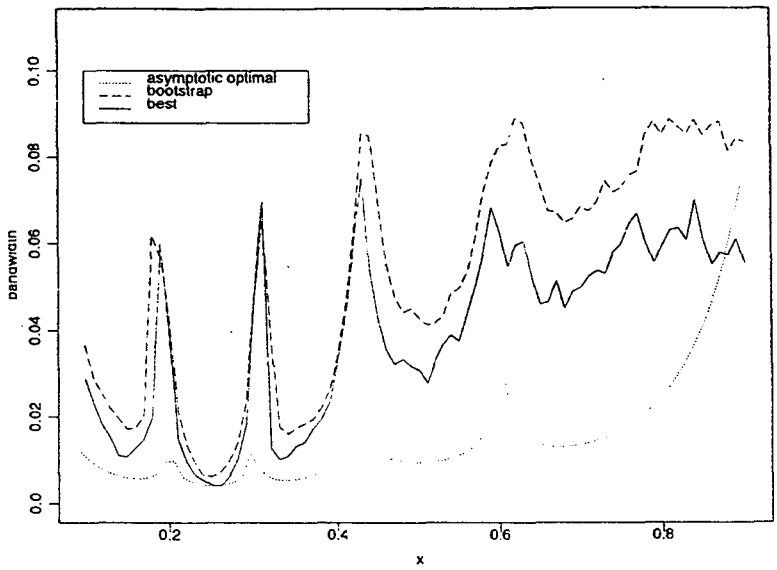


Figure 3. Bandwidth plot for $\theta_2(x)$



In figure 1, we represent the data, true function and estimated function based on local and global bandwidth respectively. As we can see, estimated curve with local bandwidth are closer to the true curve than the curve with the global bandwidth for all x. Figure 2 represents estimated bandwidth for the function $\theta_1(x)$. Asymptotic optimal bandwidths

Figure 4. Local bias for $\theta_2(x)$

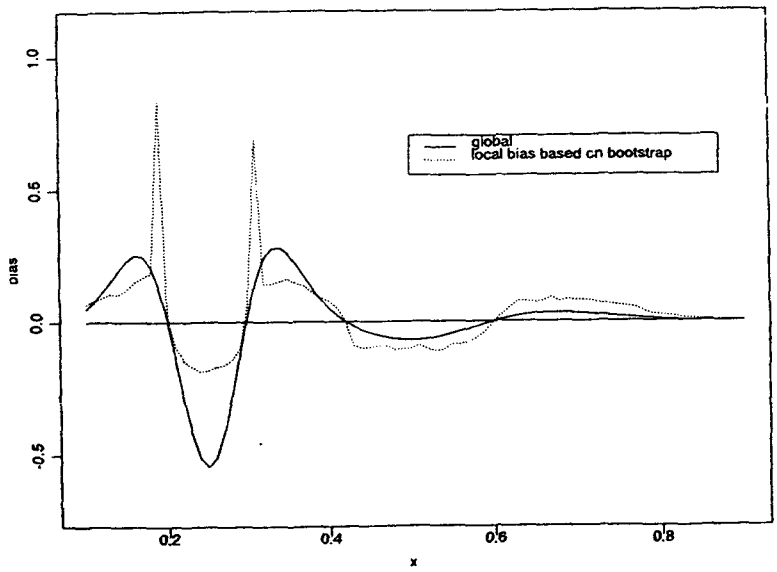
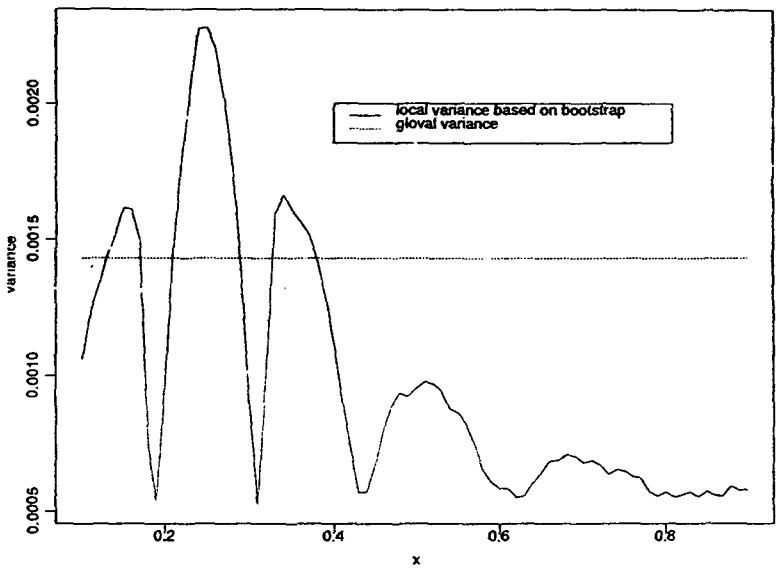


Figure 5. Local variance for $\theta_2(x)$



are evaluated with known θ and σ. Where, best means the estimated local bandwidth which we presume the knowledge of true function. So it is the best bandwidth which can be achived from the sample. There seems to be a general tendency for the local bandwidth based on bootstrap method. We

note that in the peak part of the curve estimated bandwidths are small and in flat part large values of bandwidth is obtained. Bootstrap based bandwidths are closer to the best bandwidths than asymptotic ones. We have presented a asymptotic optimal, estimated local bandwidth in figure 3. In this plot, the same tendency for the local bandwidth are appeared.

Figure 6. Estimated curve for $\theta_2(x)$

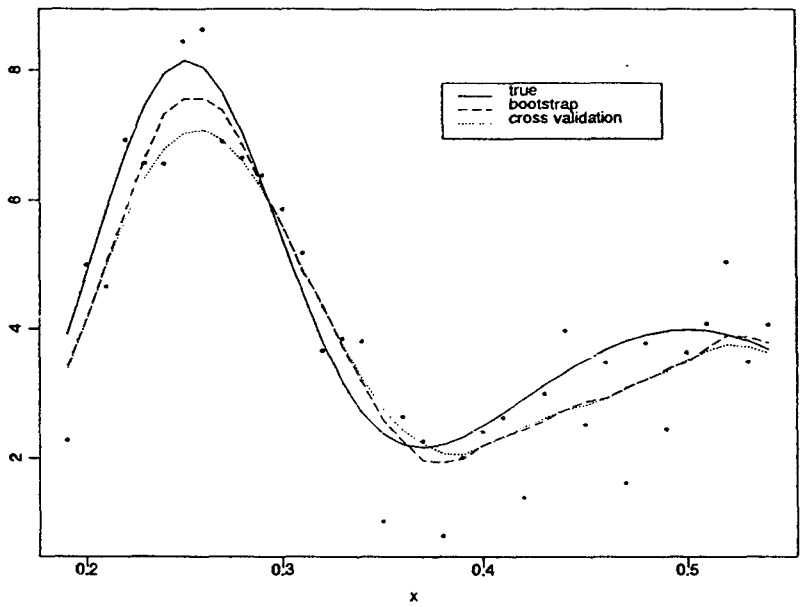


Figure 4 and 5 represents the estimated local bias and local variance for $\theta_2(x)$. From this plot, we can check that biases in peak part of the curve are greatly reduced whereas variances are deflated. Figure 6 represents true function and estimated curve for $\theta_2(x)$.

4. DISCUSSION AND CONCLUSION

From the plots of estimated local bandwidth in Figure 1 thru Figure 5, we can see that the local bandwidths varies considerably. Also we can see that bootstrap based local bandwidths which minimise bootstrap estiamte of MSE are a little large than locally asymptotic optimal bandwidths.

The bootstrap method for the choice local of bandwidth generally provides an improvement in performance where cross-validatory bandwidth is used as

an objective initial bandwidth. For all the two functions considered and for this sample size, the bootstrap based local bandwidth choice yields closer estimate to the true curve than cross-validated global bandwidth.

We proposed a local bandwidth selection method for selecting the data driven bandwidth for the fixed design kernel regression function estimation. It would seem that the major benefit of local bandwidth selection method is to provide accurate estimate for the peak part of unknown regression function. Where computational consideration permit, bootstrap compares favorably with cross-validation for these small sample size.

REFERENCES

(1) Beran, R. (1985). Bootstrap Methods in Statistics., *Jber. Math. Verein.*, 86, 14-30.

(2) Epanechnikov, V.A. (1969). Nonparametric estimation of multidimensional probability density., *Theory of Probability and its Applications*, 14, 153-158.

(3) Hardle, W. and Bowman, W.(1988). Bootstrapping in nonparametric regression : Local adaptive smoothing and confidence bands. *Journal of American Statistical Association*, 83, 102-110.

(4) Kim, D. (1993). Nonparametric Kernel Regression Function Estimation with Bootstrap Method., *Journal of Korean Statistical Society*, 22, 361-368.

(5) Nadaraya, E.A. (1964,a). On estimating regression., *Theory of Probability and its Applications*, 9, 141-142.

(6) Priestly, M.B. and Chao, M. J. (1972). Nonparametric function fitting., *Journal of Royal Statistical Society*, Ser. B, 34, 385-392

(7) Watson, G.S. (1964). Smooth regression Analysis., *Shankya, Ser A.*, 26, 359-372.

(8) Wong, W.H. (1983). On the consistency of cross-validation in kernel nonparametric regression. *Annals of Statistics*, 11, 1136-1141.