The Korean Communications
in Statistics Vol. 4, No. 1, 1997
pp. 177-183

Selection of Data-adaptive Polynomial Order in Local Polynomial Nonparametric Regression

Jaekeun Jo[1)]

Abstract

A data-adaptive order selection procedure is proposed for local polynomial nonparametric regression. For each given polynomial order, bias and variance are estimated and the adaptive polynomial order that has the smallest estimated mean squared error is selected locally at each location point. To estimate mean squared error, empirical bias estimate of Ruppert (1995) and local polynomial variance estimate of Ruppert, Wand, Holst and Hössjer (1995) are used. Since the proposed method does not require fitting polynomial model of order higher than the model order, it is simpler than the order selection method proposed by Fan and Gijbels (1995b).

1. Introduction

Let X and Y be random variables which can be modelled by

$$Y = m(X) + \varepsilon, \qquad E\varepsilon = 0 \qquad \text{and} \qquad \mathrm{Var}(\varepsilon) = v(X),$$

where $m(x)$ and $v(x)$ are smooth functions specifying the conditional mean and variance functions of Y given $X = x$. It is of interest to estimate regression function $m(x) = E(Y|X = x)$ based on a random sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ from (X, Y). Monographs such as Härdle (1990), Wand and Jones (1995) and Fan and Gijbels (1996) provide a good deal of various nonparametric curve fitting procedures. Among them, local polynomial regression method is considered in this paper.

For u in a neighborhood of a location point x, assume that $m(u)$ can be modelled locally by a polynomial of order p,

$$m(u) \approx m(x) + m'(x)(u-x) + \cdots + m^{(p)}(x)(u-x)^p/p!. \tag{1.1}$$

Then the local polynomial regression method estimates regression function and its derivatives by minimizing

1) Associate Professor, Department of Computer Science and Statistics, Kyungsung University, Pusan, 608-736, Korea.

$$\sum_{i=1}^{n}\left\{ Y_i - \sum_{j=0}^{p}\beta_j(X_i-x)\right\}^2 K\left(\frac{X_i-x}{h}\right) \tag{1.2}$$

with respect to $\beta_0, \cdots, \beta_p$, where h is a bandwidth and K is a symmetric kernel function. From (1.1), it is clear that $j!\,\widehat{\beta_j}$ estimates $m^{(j)}(x)$, for $j=0, \cdots, p$. If we let $\boldsymbol{W}_h(x)$ be the $n\times n$ diagonal matrix whose ith diagonal element is $K((X_i-x)/h)$, and $\boldsymbol{X}_p(x)$ be the $n\times(p+1)$ design matrix whose $(k,l)^{th}$ element is $(X_k-x)^{l-1}$, the weighted least squares problem (1.2) can be rewritten in matrix form as minimizing

$$(\boldsymbol{y}- \boldsymbol{X}_p(x)\boldsymbol{\beta})^T \boldsymbol{W}_h(x)(\boldsymbol{y}- \boldsymbol{X}_p(x)\boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$, where $\boldsymbol{y}=(Y_1, \cdots, Y_n)^T$ and $\boldsymbol{\beta}=(\beta_0, \cdots, \beta_p)^T$. Then the ordinary least squares theory leads to the solution $\widehat{\boldsymbol{\beta}}=(\boldsymbol{X}_p(x)^T \boldsymbol{W}_h(x)\boldsymbol{X}_p(x))^{-1}\boldsymbol{X}_p(x)^T \boldsymbol{W}_h(x)\boldsymbol{y}$ and the local polynomial estimator of $m(x)$ is $\widehat{m}(x;h,p)=\widehat{\beta_0}$. And the bias and variance of $\widehat{m}(x;h,p)$ are given by

$$\begin{aligned}
&\text{Bias}\ (\widehat{m}(x;h,p))\\
&\quad= \boldsymbol{e}_1^T(\boldsymbol{X}_p(x)^T \boldsymbol{W}_h(x)\boldsymbol{X}_p(x))^{-1}\boldsymbol{X}_p(x)\boldsymbol{W}_h(x)(\boldsymbol{m}- \boldsymbol{X}_p(x)\boldsymbol{\beta}),\\
&\text{Var}\ (\widehat{m}(x;h,p))\\
&\quad= \boldsymbol{e}_1^T(\boldsymbol{X}_p(x)^T \boldsymbol{W}_h(x)\boldsymbol{X}_p(x))^{-1}(\boldsymbol{X}_p(x)^T\boldsymbol{\Sigma}\,\boldsymbol{X}_p(x))\\
&\qquad\cdot(\boldsymbol{X}_p(x)^T \boldsymbol{W}_h(x)\boldsymbol{X}_p(x))^{-1}\boldsymbol{e}_1,
\end{aligned}$$

where $\boldsymbol{m}=\{m(X_1), \cdots, m(X_n)\}^T$, $\boldsymbol{e}_1=(1,0,\cdots,0)^T$ and $\boldsymbol{\Sigma}=\text{diag}\{K^2((X_i-x)/h)v(X_i)\}$ (see e. g. Wand and Jones (1995), Fan and Gijbels (1996)). Thus the mean squared error of estimator $\widehat{m}(x;h,p)$ is

$$\text{MSE}(\widehat{m}(x;h,p))= \text{Bias}^2(\widehat{m}(x;h,p))+ \text{Var}(\widehat{m}(x;h,p)).$$

Several estimates of bias and variance terms have been proposed by authors. Researchers have investigated bandwidth selection problem through the estimation of mean squared error (Ruppert, Sheather, and Wand (1995)), but selection of polynomial order has not been studied widely. Recently Fan and Gijbels (1995a) discussed the issue of local variable bandwidth and Fan and Gijbels (1995b) considered selection of adaptive polynomial order for local polynomial regression. For fixed polynomial order p, bandwidth h can be selected locally to capture varying curvatures of the unknown regression curve. On the contrary for fixed bandwidth, selection of polynomial order locally (different order polynomial at different location point x) is another way to capture curvature of regression curve. In this paper, an adaptive polynomial order selection procedure is considered. The proposed procedure is different from procedure suggested by Fan and Gijbels (1995a, 1995b) in the estimations of bias and variance terms. Empirical bias estimation procedure of Ruppert (1995) is used to estimate bias. Ruppert used this procedure to select local bandwidth and called this procedure EBBS (empirical-bias bandwidth selector). A distinct feature of EBBS is that it uses empirical estimates rather than asymptotic expression of bias. In addition, it is different from procedures of Fan and Gijbels in that it does not require fitting of local polynomial of order higher than the model order. Besides bias estimate, estimate proposed by Ruppert, Wand, Holst and Hössjer (1995) is used to estimate variance term. On the contrary, variance estimate of Fan and Gijbels is obtained by fitting extra higher order polynomial fitting which can result collinearity problem. Combining these estimates, we can estimate mean squared error. For each polynomial order, mean squared errors are estimated and the order that yields the smallest estimated mean squared error is selected as the adaptive polynomial order at the location point and at the bandwidth where we are to estimate regression function.

2. Polynomial Order Selection

Ruppert (1995) proposed an estimator of bias term to select local bandwidth for fixed order local polynomial regression. A similar bandwidth selection algorithm has been proposed for Priestly-Chao kernel estimator by Cha and Lee (1994). It is well known that standard asymptotics lead to the asymptotic bias of $\widehat{m}(x;h,p)$ which can be written as

$$\text{Bias}(\widehat{m}(x;h,p)) \approx C_{p+1}h^{p+1}+C_{p+2}h^{p+2}+O(h^{p+3}),$$

where C_{p+1}, C_{p+2} are constant factors which depend on m, K and the design density (see e. g. Wand and Jones (1995), Fan and Gijbels (1996)). From this approximation, the bias is estimated empirically using estimates at several bandwidths to fit a model for $\mathrm{E}(\widehat{m}(x;h,p))$ as a function of h. At a fixed point x and at a fixed bandwidth h_0 where we are to estimate $\mathrm{E}(\widehat{m}(x;h_0,p))$, choose bandwidths $h_0^{(1)},\dots,h_0^{(J)}$ $(J\geq 1)$ in the neighborhood of h_0 and calculate $\widehat{m}(x,h_0^{(j)},p)$, $(j=1,\dots,J)$ and then for some $t\geq 1$, fit the curve

$$\widehat{m}(x,h,p)\approx C_0+C_{p+1}h^{p+1}+\dots+C_{p+t}h^{p+t} \tag{2.1}$$

to the $\{(h_0^{(j)},\widehat{m}(x,h_0^{(j)},p),\ j=1,\dots,J\}$ by the ordinary least squares. Then the bias of estimator $\widehat{m}(x;h_0,p)$ is estimated by

$$\widehat{C}_{p+1}h_0^{p+1}+\dots+\widehat{C}_{p+t}h_0^{p+t}. \tag{2.2}$$

Assuming local homoscedasticity $v(X_i)\approx v(x)$ for X_i in a neighborhood of x, variance of $\widehat{m}(x;h,p)$ can be approximated by (see e. g. Wand and Jones (1995), Fan and Gijbels (1996))

$$\begin{aligned} \text{Var}\ (\widehat{m}(x;h,p)) &\approx e_1^{\ T}(\ \boldsymbol{X}_p(x)^T\ \boldsymbol{W}_h(x)\ \boldsymbol{X}_p(x)\ )^{-1}(\ \boldsymbol{X}_p(x)^T\ \boldsymbol{W}_h(x)^2\ \boldsymbol{X}_p(x)) \\ &\quad \cdot (\ \boldsymbol{X}_p(x)^T\ \boldsymbol{W}_h(x)\ \boldsymbol{X}_p(x)\ )^{-1} v(x)\ \boldsymbol{e}_1 . \end{aligned} \tag{2.3}$$

We can estimate $v(x)$ using the normalized weighted residual sum of squares of Fan and Gijbels (1995a) or we can utilize another variance function estimator that smooth squared residuals by local polynomial fitting of different order and bandwidth (Ruppert, Wand, Holst and Hössjer (1995)).

Let p_1 and h_1 be the polynomial order and bandwidth at the first stage (smoothing $\boldsymbol{y}$) and let $\boldsymbol{r}=(r_1,\cdots,r_n)^T=\boldsymbol{y}-\widehat{\boldsymbol{y}}$, where $\widehat{\boldsymbol{y}}=(\widehat{m}(X_1,h_1,p_1),\cdots,\widehat{m}(X_n,h_1,p_1))^T$, be the residual vector from the fitting. And at the second stage, squared residuals are smoothed similarly with polynomial of order p_2 and bandwidth h_2. If S_1 is an $n\times n$ smoother matrix at the first step ($\widehat{\boldsymbol{y}}=S_1\ \boldsymbol{y}$,), the variance estimate proposed by Ruppert et al. is

$$\widehat{v}(x)=\frac{e_1^T\{X_{p_2}(x)^T W_{h_2}(x) X_{p_2}(x)\}^{-1} X_{p_2}(x)^T W_{h_2}(x)\ \boldsymbol{r}^2}{1+e_1^t\{X_{p_2}(x)^T W_{h_2}(x) X_{p_2}(x)\}^{-1} X_{p_2}(x)^T W_{h_2}(x)\Delta}, \tag{2.4}$$

where $\boldsymbol{r}^2=(r_1^2,\cdots,r_n^2)^T$, and $\Delta=\text{diag}(S_1S_1^T-2S_1)$.

Combining (2.2)-(2.4), MSE($\widehat{m}(x;h_0,p)$) can be estimated by

$$\begin{aligned} \widehat{\text{MSE}}\ (\widehat{m}(x;h_0,p)) &= (\ \widehat{C}_{p+1}h_0^{p+1}+\dots+\ \widehat{C}_{p+t}h_0^{p+t})^2 \\ &\quad +\ e_1^{\ T}(\ \boldsymbol{X}_p(x)^T\ \boldsymbol{W}_{h_0}(x)\ \boldsymbol{X}_p(x)\ )^{-1}(\ \boldsymbol{X}_p(x)^T\ \boldsymbol{W}_{h_0}(x)^2\ \boldsymbol{X}_p(x)) \\ &\quad \cdot (\ \boldsymbol{X}_p(x)^T\ \boldsymbol{W}_{h_0}(x)\ \boldsymbol{X}_p(x)\ )^{-1}\widehat{v}(x)\ \boldsymbol{e}_1 . \end{aligned} \tag{2.5}$$

By repeating above calculations for each $p\in\{0,1,\cdots,p_{\max}\}$ at fixed x and h_0, we can choose order of polynomial which has the smallest $\widehat{\text{MSE}}(\widehat{m}(x;h_0,p))$.

3. An Example

Since the justifications for the adaptive order selection and extensive examples can be found in Fan and Gijbels (1995b), we present only one simulated example to illustrate the performance of our order selection procedure. The simulated example is

$$Y_i=\sin^3(2\pi X_i^3)+\varepsilon_i,$$

where $X\sim U(0,1)$, $\varepsilon\sim N(0,(0.1)^2)$. We take a sample of size n=256 and we use Epanechnikov kernel at every smoothing steps and we estimated regression curve at 100 equally spaced location points.

At the first step, bias is estimated empirically at each location points. Regression curves in (2.1) are computed at 10 bandwidths ($J=10$) in the neighborhood of h_0 with $t=3$, and bias is estimated by (2.2). And at the second step, variance function is estimated by (2.4). For computational simplicity, local linear estimation with equal bandwidths are used in two steps of variance function estimation ($p_1=p_2=1$, $h_1=h_2=0.1$). And finally, adopting algorithm of Fan and Gijbels (1995b), mean squared error at each location point is obtained by the weighted average of mean squared errors at points in the neighborhood of that location point.

Figure 1 (a)-(f) shows estimated curves by adaptive order selection. For comparison, curves by local linear fitting are also given. We used several bandwidths h_0=0.07, 0.1, 0.12, 0.15, 0.18 and 0.2. As we can see in Figure 1, adaptive order selection outperform local linear fitting especially near the peak and valley for various bandwidths. And contrary to the changes in local linear fitting for the varying bandwidths, changes of adaptive order fitting is robust to the bandwidth. This result coincides with results of simulations in Fan and Gijbels (1995b).

Although adaptive order selection procedure is robust to the bandwidths, bandwidth selection strategy more refined than the rule-of-thumb method of Fan and Gijbels (1995b) deserves further study.

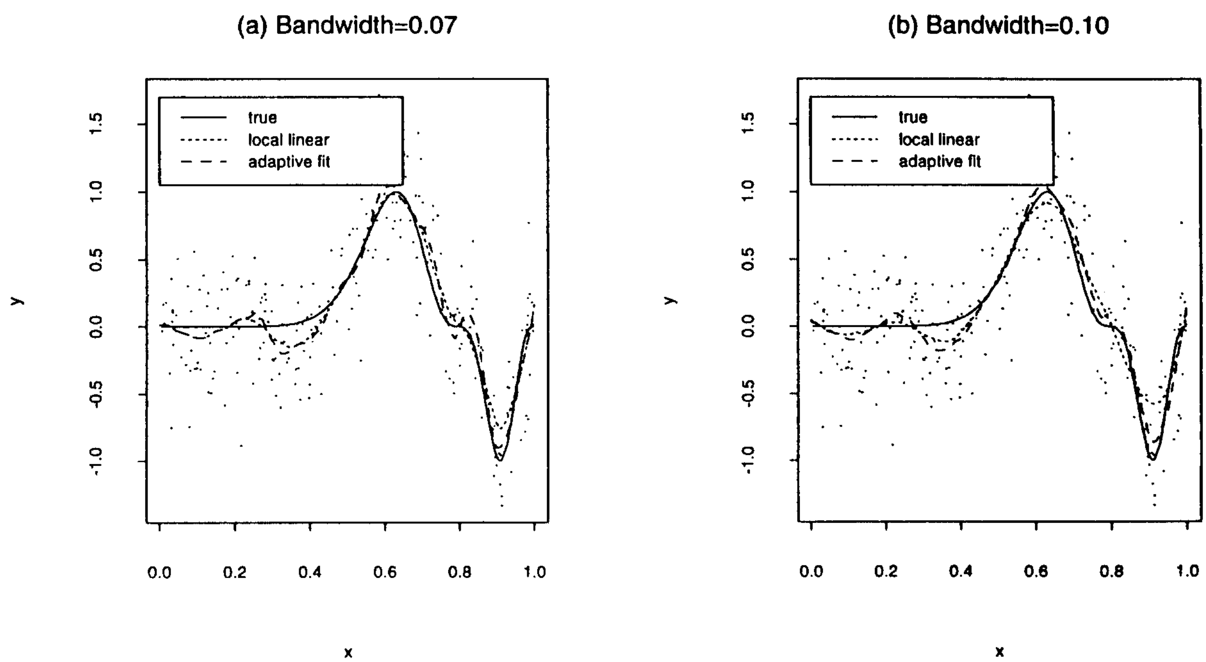

Figure 1: The comparison of local polynomial regression by adaptive order selection and local linear regression

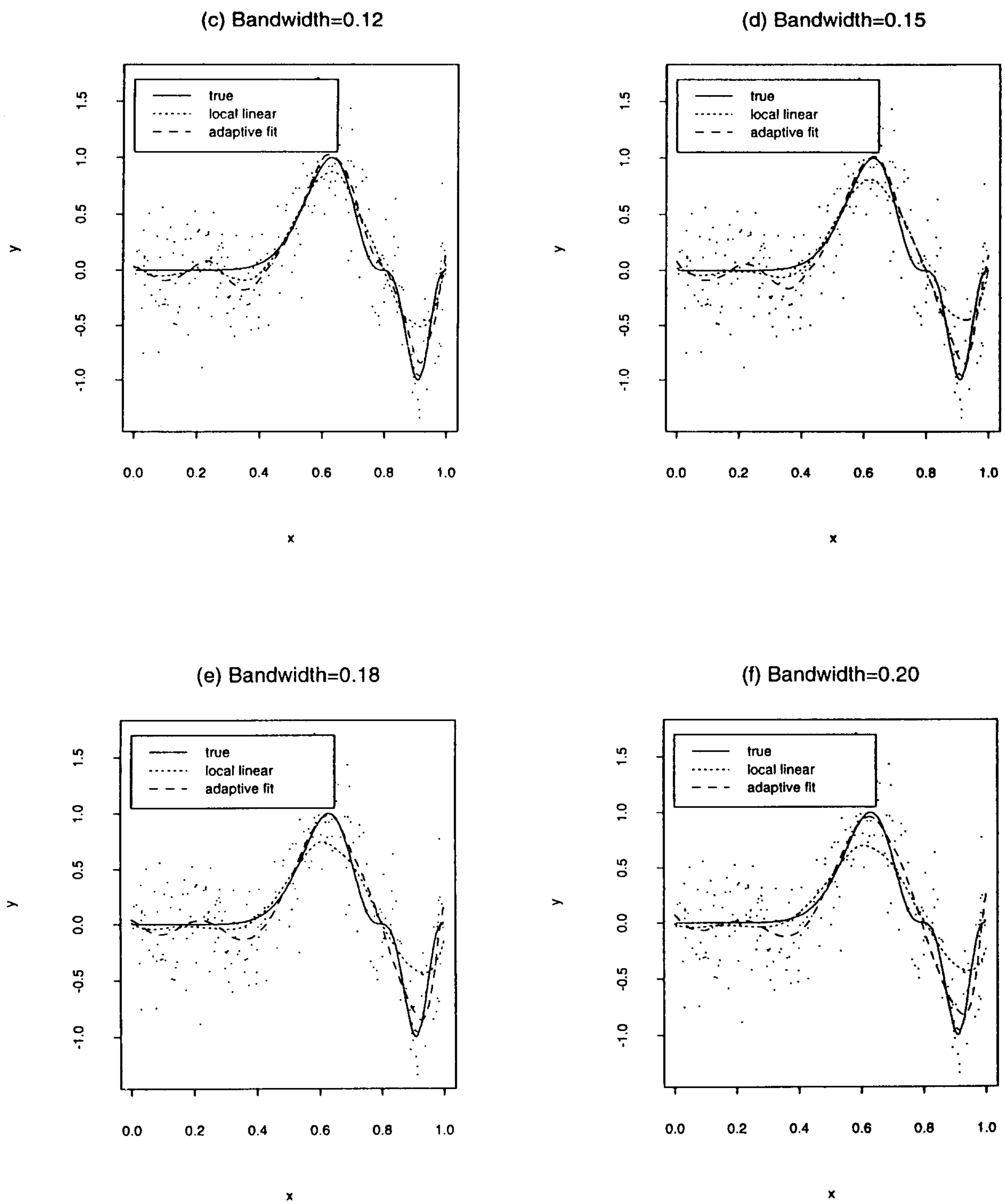

Figure 1(continued)

REFERENCES

[1] Cha, K-J. and Lee, S-W. (1994). An adaptive bandwidth selection algorithm in nonparametric regression, *Korean Journal of Applied Statistics*, Vol. 7, 149-158.

[2] Fan, J. and Gijbels, I. (1995a). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation, *Journal of the Royal Statistical Society, Series B*, Vol. 57, 371-394.

[3] Fan, J. and Gijbels, I. (1995b). Adaptive order polynomial fitting: bandwidth robustification and bias reduction, *Journal of Computational and Graphical Statistics*, Vol. 4, 213-227.

[4] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

[5] Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, New York.

[6] Ruppert, D. (1995). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation, manuscript.

[7] Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, Vol. 90, 1257-1270.

[8] Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1995). Local polynomial variance function estimation, manuscript.

[9] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman & Hall, London.