Title
A new information criterion-based bandwidth selection method for non-parametric regressions

Authors
Xiang, S
Yao, W

Publication Date
2016

Peer reviewed

A New Information Criterion Based Bandwidth Selection Method for Nonparametric Regressions

Sijia Xiang [*], and Weixin Yao [†]

Abstract

Local linear estimator is a popularly used method to estimate the nonparametric regression functions, and many methods have been derived to estimate the smoothing parameter, or the bandwidth in this case. In this article, we propose an information criterion based bandwidth selection method, with the degrees of freedom originally derived for nonparametric inferences. Unlike the plug-in method, the new method does not require preliminary parameters to be chosen in advance, and is computationally efficient compared to the cross-validation method. Numerical study shows that the new method performs better or comparable to existing plug-in method or cross-validation method in terms of the estimation of the mean functions, and has lower variability than cross-validation selectors. Real data applications are also provided to illustrate the effectiveness of the new method.

Key words: Information criterion method; bandwidth selector; nonparametric regression.

[*]Corresponding Author, School of Mathematics and Statistics, Zhejiang University of Finance and Economics. E-mail: sjxiang@zufe.edu.cn. Xiang's research is supported by Zhejiang Provincial NSF of China grant LQ16A010002

[†]Department of Statistics, University of California, Riverside. E-mail: weixin.yao@ucr.edu. Yao's research is supported by NSF grant DMS-1461677.

1 Introduction

Given $\{(X_1, Y_1), ..., (X_n, Y_n)\}$ from the model

$$Y = m(X) + \epsilon, \tag{1.1}$$

where $m(x) = E(Y|X = x)$ is a smooth function, $E(\epsilon|X) = 0$ and $\mathrm{Var}(\epsilon|X = x) = \sigma^2(x)$. Local linear approximation, among other methods, has been proposed to estimate $m(x)$ and its features have been well studied. See, for example, Fan (1992, 1993) and Fan and Gijbels (1992). In a small neighborhood of x_0, $m(x) \approx m(x_0) + m'(x_0)(x - x_0) \equiv \beta_0 + \beta_1(x - x_0)$. The problem of estimating $m(x_0)$ is equivalent to the estimation of β_0, which is calculated by minimizing

$$\sum_{i=1}^{n}\{Y_i - \beta_0 - \beta_1(X_i - x_0)\}^2 K_h(X_i - x_0),$$

where $K_h(t) = K(t/h)/h$, $K(\cdot)$ is a kernel density function, and h is a smoothing parameter. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the solutions. Then $\hat{m}(x_0) = \hat{\beta}_0$.

The selection of the optimal smoothing parameter h is crucial to the estimation of $m(x)$. Classical bandwidth selection methods, such as cross-validation (CV), generalized cross-validation (GCV) or Akaike information criterion (AIC), and plug-in methods have been popularly used. CV (Härdle et al., 1988) and GCV (Craven and Wahba, 1979) try to minimize an unbiased estimator of mean average squared error (MASE), while AIC and improved AIC (AIC_c) (Hurvich, et al., 1998) are aiming at minimizing the expected Kullback-Leibler discrepancy. For example, CV selects the bandwidth that minimizes

$$CV(h) = n^{-1}\sum_{i=1}^{n}[Y_i - \hat{m}_{-i}(X_i)]^2 w(X_i),$$

where $\hat{m}_{-i}(X_i)$ is the "leave-one-out" estimate of $m(X_i)$, $w(X_i)$ is a weight function.

Classical bandwidth selectors are totally automatic, but tend to choose highly variable and under-smoothing parameters. As a result, plug-in methods have been developed, which minimizes a large sample approximation of MASE. For example, Ruppert et al. (1995) proposed the plug-in selector for the local linear estimator

$$h_{opt} = \left(\frac{R(K)}{\mu^2(K) \int m''(x)^2 f(x) dx} \right)^{-1/5} n^{-1/5},$$

where $R(K) = \int K(t)dt, \mu(K) = \int t^2 K(t)dt, f$ is the density of predictor variables, and n is the sample size. For plug-in method, we need to replace the unknown quantity $m''(x)$ by an estimator, such as based on a parametric fit. Compared to the classical bandwidth selector, plug-in methods yield more stable estimators, and does not tend to under-smooth in practice. However, plug-in methods have only been developed when the asymptotic optimal bandwidth has a simple form, and are criticized for not being able to minimize the average squared error (ASE) for the observed data set (Jones and Kappenman, 1991; Hall and Marron, 1991; Grund et al., 1994). In addition, the plug-in selector generally requires preliminary parameters to be chosen by the researchers, and the properties of the final estimator can be sensitive to those choices.

More recently, some new bandwidth selection methods have been proposed for more complicated model settings. For a nonparametric functional regression model with homoscedastic errors and unknown error density, Shang (2013) proposed a Bayesian bandwidth estimation procedure, which outperforms the likelihood CV for estimating the error density. Levine (2006) studied a possible bandwidth selection approach for difference-based variance estimators in the nonparametric regression, basing on the cross-validation idea adjusted for correlated data. For integrated time series data, Sun and Li (2011) suggested using the least squares CV (LS-CV) method to choose the smoothing parameter, and studied the asymptotic properties of both the local constant and local linear estimators.

In this article, we propose a new bandwidth selector for the local linear approximation of model (1.1), based on the idea of information criterion and the degrees of freedom proposed by Fan et al. (2001) for nonparametric regression. Unlike the plug-in method, the new method does not require preliminary parameters to be chosen in advance, and is computationally efficient compared to the cross-validation method. Numerical study shows that the new method performs better or comparable to existing plug-in method or cross-validation method in terms of the estimation of the mean functions, and has lower variability than cross-validation selectors. Some real data applications are also provided to illustrate the effectiveness of the new method.

The rest of the article is organized as follows. The derivations of the new method are given in Section 2 . In Section 3 and Section 4, we use simulation studies and real data examples to show the effectiveness of the new method, and compare it with existing bandwidth selectors. A few discussions are provided in Section 5.

2 Bayesian Information Criterion Based Method

In general, a Bayesian information criterion has the form:

$$n \log\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\} + df \times \log n, \tag{2.1}$$

where df is the degrees of freedom, amounting to the complexity of the model, and the first term is a measure of goodness-of-fit of the model. To implement the information criterion, a measure of the complexity of the model is needed. Unlike the parametric model, the model complexity is not well defined for nonparametric regression model. Here, we implement the degrees of freedom proposed by Fan et al. (2001), which is originally derived for nonparametric hypothesis testing. Based on Fan et al. (2001), the degrees of freedom of model (1.1) is:

$$df_N = r_K h^{-1}|\Omega|\{K(0) - \frac{1}{2}\int K^2(t)dt\}, \tag{2.2}$$

where Ω is the support of the covariate, $K(\cdot)$ is a kernel density function, and

$$r_K = \frac{K(0) - \frac{1}{2}\int K^2(t)dt}{\int\{K(t) - \frac{1}{2}K * K(t)\}^2 dt}.$$

Therefore, we propose to select the bandwidth which minimizes

$$BIC_N = n\log\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\} + df_N \times \log n, \tag{2.3}$$

where df_N is defined in (2.2). Note that the degrees of freedom and, therefore, BIC_N depend on the bandwidth h. We propose to apply the information criterion on a wide range of bandwidths, and select the bandwidth which minimizes the information criterion (2.3). In case the support of the covariate is not a closed interval, we propose to use the range of the sample to approximate $|\Omega|$.

Zhang (2003) also covered in detail the degrees of freedom of linear smoothers in non-parametric settings. For linear smoother $\mathbf{S}$, they proposed to use either $\mathrm{tr}(\mathbf{S})$, $\mathrm{tr}(\mathbf{S}^T\mathbf{S})$, or $\mathrm{tr}(2\mathbf{S} - \mathbf{S}^T\mathbf{S})$ as the degrees of freedom. Among these, they showed, under some conditions, $\mathrm{tr}(2\mathbf{S} - \mathbf{S}^T\mathbf{S}) = (2K - K * K)(0)|\Omega|/h\{1 + o(1)\}$, which is proportional to our df, with a multiplier $r_K/2$. However, the intuition behind the methods are quite different.

3 Simulation Study

In this section, we use Monte Carlo simulations to investigate the finite sample performance of the newly proposed bandwidth selection method, and compare it with some existing bandwidth selectors.

Table 1 contains the eight examples considered in the simulation study. In Figure 1, a random sample of size $n = 100$ is plotted for each example, accompanied by their mean functions. In Example 1 and Example 2, the covariates are from a closed interval and an open set, respectively. Example 3 and Example 4 were used by Fan (1992), where the mean function is approximately linear in Example 3, and the covariate in Example 4 is from a mixture of normal distributions. Example 5 - Example 8 were suggested by Hurvich et al. (1998), where Example 5 represents a case with less fine structure or trend, Example 6 a case with noticeably different degrees of curvature for different values of the predictor, Example 7 a case with a trend but no fine structure, and Example 8 non-differentiable at $x = 1/3$. For each model, sample sizes of $n = 100$, $n = 200$ and $n = 400$ are conducted over 500 repetitions.

Table 1: Models considered in the simulation study.

Mean functions	Density of covariate	Density of error
$2\sin(\pi x)$	$U(0,1)$	$N(0,2^2)$
$4-\sin(\pi x)$	$N(0,1)$	$N(0,0.2^2)$
$\sin(0.75x)$	$N(0,1)$	$N(0,0.6^2)$
$\sin(2.5x)$	$0.5N(-1,1)+0.5N(1.75,0.25)$	$N(0,0.6^2)$
$1-48x+218x^2-315x^3+145x^4$	$U(0,1)$	$N(0,4.5^2)$
$0.3\exp\{-64(x-0.25)^2\}+0.7\exp\{-256(x-0.75)^2\}$	$U(0,1)$	$N(0,1.25^2)$
$10\exp(-10x)$	$U(0,1)$	$N(0,8^2)$
$\exp(x-1/3)$ if $x<1/3$, $\exp\{-2(x-1/3)\}$ if $x\geq 1/3$	$U(0,1)$	$N(0,1.5^2)$

For each example, we assume that the data comes from model (1.1) and the local

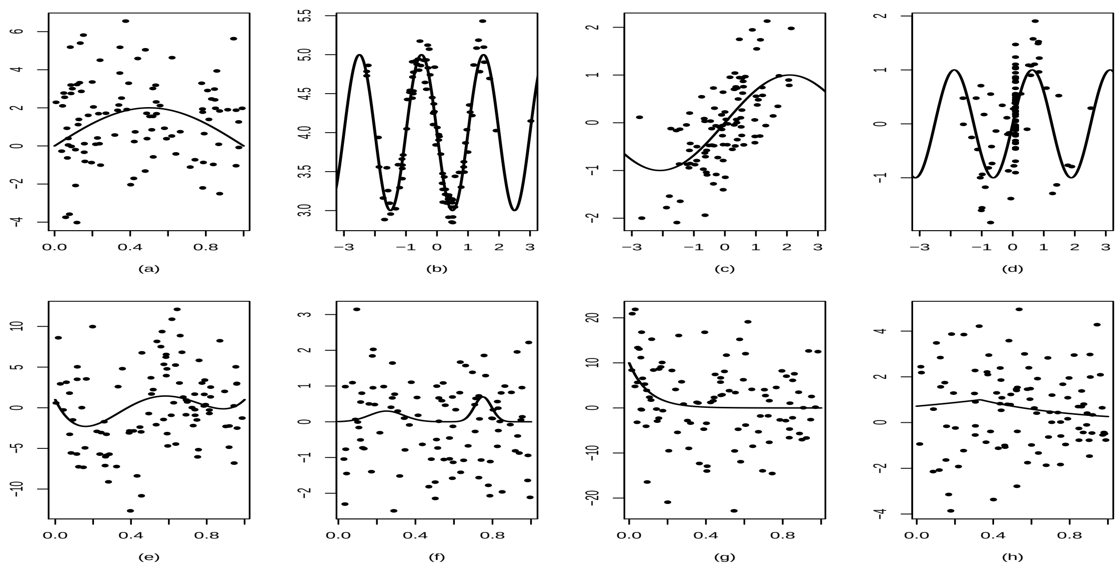

Figure 1: Simulation examples: random samples of size $n = 100$ and their corresponding density plots.

linear approximation with Gaussian kernel is used to estimate the mean functions. The performance of the new bandwidth selector is reported, and is compared with the plug-in method by Ruppert et al. (1995), leave-one-out cross-validation, and 10-fold cross-validation. For the plug-in method, the least squares quartic fit is used to approximate $m(x)$ and therefore $m''(x)$. If r is used to denote the range of predictors, then the grid of bandwidths is formed by taking 30 equally spaced points from $0.01r$ to $0.5r$.

To assess the performance of the bandwidth selectors, we report the average squared error (ASE) of the estimators:

$$ASE = \frac{1}{N} \sum_{i=1}^{N} \{m(u_i) - \hat{m}(u_i)\}^2, \tag{3.1}$$

where $\{u_1, ..., u_N\}$ is a set of equally spaced grid points, and N is the number of grid points. In the simulation, $N = 100$ is set for all examples.

Table 2 contains the mean and standard deviation of ASE of the four bandwidth selectors when $n = 100$, $n = 200$, and $n = 400$, based on 500 repetitions. Figure

2, Figure 3, and Figure 4 show the boxplots of ASE of Example 1 - Example 8 with $n = 100$, $n = 200$ and $n = 400$, respectively. We can see that the BIC_N is most often best, and usually not far away from best otherwise. The performance of 10-fold CV and leave-one-out CV are close, and are slightly better than the plug-in method in most cases.

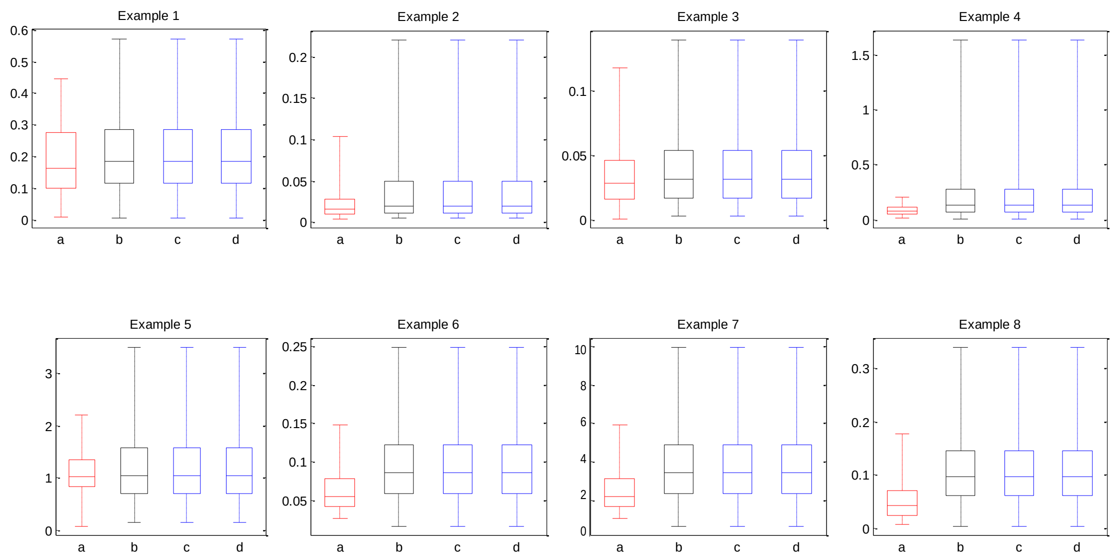

Figure 2: Boxplot of ASE of Example 1 - Example 8 with $n = 100$, and the bandwidth selected by a) BIC_N, b) plug-in, c) 10-fold CV, and d) leave-one-out CV.

In addition to the estimation of the mean functions, we also compare the computation efficiency among different bandwidth selectors. The simulation is done through Matlab on a personal laptop with an i7-3610QM CPU and 8GB of RAM. Table 3 reports the mean and standard deviation of calculation time (in seconds) of a repetition. As expected, the plug-in method is always the fastest method in all cases. The BIC_N method takes much less time than the 10-fold CV, and the leave-one-out cross-validation takes the longest time to compute.

To assess the variability of the selected bandwidths, we also report the variance of the selected bandwidths over 500 repetitions, in Table 4. Among the eight examples considered, Example 3 has significantly more variable bandwidths, due to the fact that

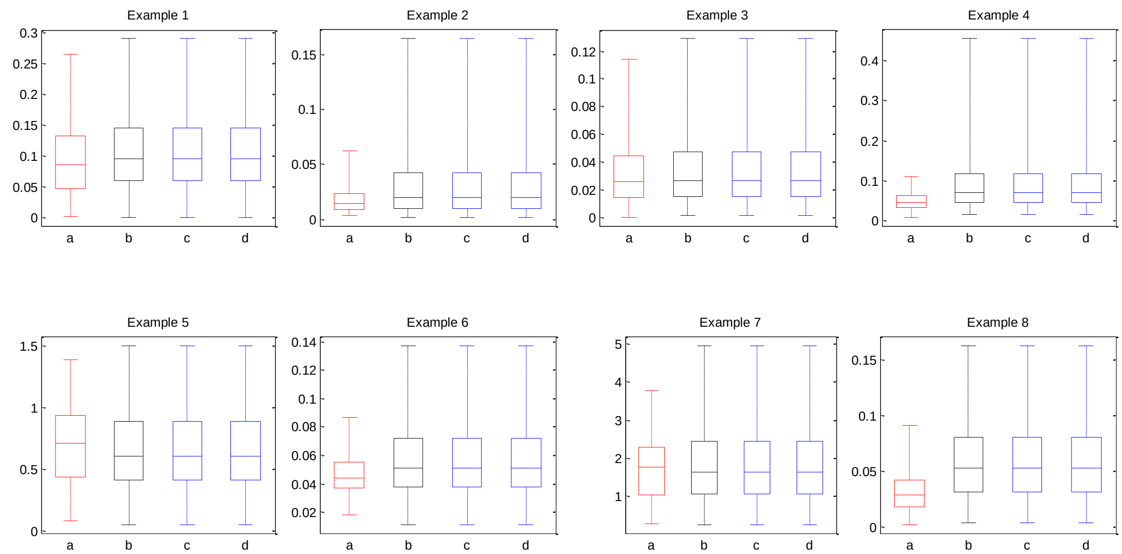

Figure 3: Boxplot of ASE of Example 1 - Example 8 with $n = 200$, and the bandwidth selected by a) BIC_N, b) plug-in, c) 10-fold CV, and d) leave-one-out CV.

its mean function is approximately linear. The plug-in selector has the least variable bandwidths in most cases. The BIC_N selector has relatively low variability compared to cross-validation when the sample size is small, and the performance of 10-fold or leave-one-out cross-validation are similar, in this respect.

Next, we consider the MSE and bias of the selected bandwidth. For each repetition, we find the optimal bandwidth which minimizes the ASE, defined in (3.1). The plug-in selector and cross-validation selector have a clear tendency towards undersmoothing, while the BIC_N tends to oversmooth in most cases. In terms of the magnitude of MSE and bias, the BIC_N gives most favorable result in Examples 2, 3, 4, 6, and 8, and comparable performance in other cases. The plug-in selector, in this cases, is the least satisfied bandwidth selector.

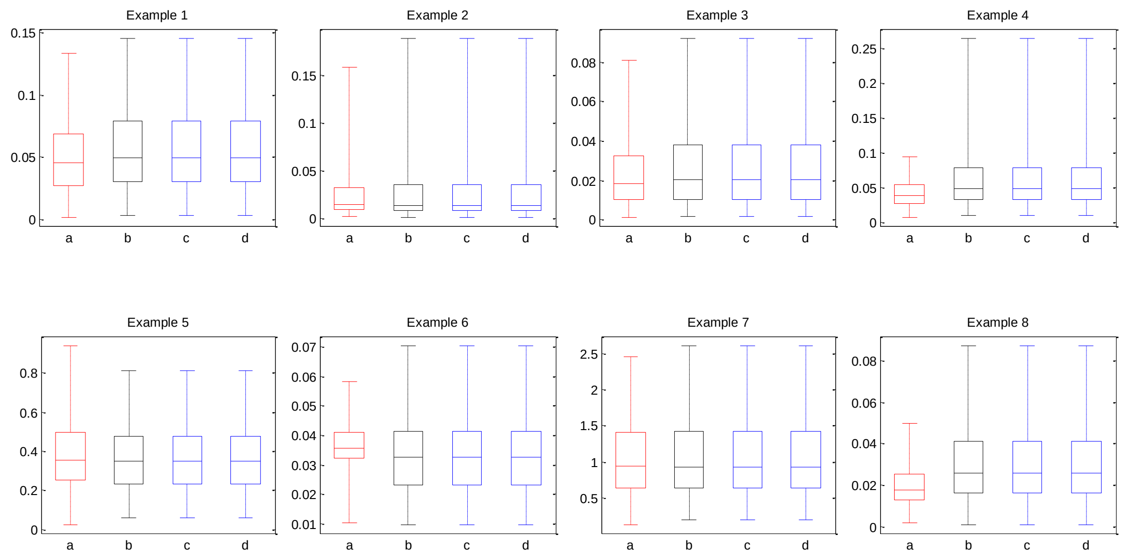

Figure 4: Boxplot of ASE of Example 1 - Example 8 with $n = 400$, and the bandwidth selected by a) BIC_N, b) plug-in, c) 10-fold CV, and d) leave-one-out CV.

4 Real data analysis

Example 1 (1995 British family expenditure data). We illustrate the application of the new bandwidth selector to the 1995 British family expenditure data, available from R package “np”. The data set consists of a random sample taken from the British Family Expenditure Survey for 1995. The households consist of married couples with an employed head-of-household between the ages of 25 and 55 years. There are 1655 household-level observations and 10 variables in the original data set. In this example, we use logarithm of total expenditure (logexp) as a covariate to predict for expenditure share on food (food).

Assuming the two variables follow model (1.1), we apply local linear approximation to the data, using each of the foregoing bandwidth selection method for optimal bandwidths. Figure 5 shows the scatter plot and fitted models based on the four bandwidth selectors.

To compare the newly proposed bandwidth selector to existing methods, since the true regression function is unknown, we use 10-fold cross-validation to check the prediction performance. The mean and standard deviation of the mean squared prediction error (MSPE) are reported in Table 6. The calculation time of each method is also reported.

It can be seen that the newly proposed bandwidth selector works comparable to 10-fold CV or leave-one-out CV in terms of prediction performance, but with much less computation time. Since 10-fold CV targets the minimization of the MSPE, it indicates that the bandwidth chosen by the new method has the optimal prediction performance. The plug-in selector, in this case, has similar prediction performance and with the least computation time.

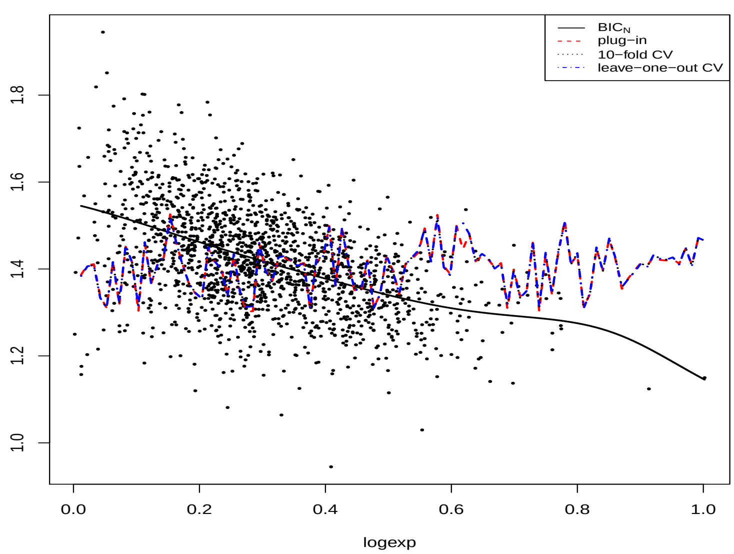

Figure 5: Scatter plot of the 1995 British family expenditure data, and its corresponding fitted models.

Example 2 (Canadian prestige data). Next, we apply the bandwidth selectors to the Canadian prestige data (Fox and Weisberg, 2011), using average education of occupational incumbents (in 1971) to predict for prestige score, which is from a social survey conducted in the mid-1960s. The data set has 102 observations, corresponding to occupations.

Table 6 shows the MSPE and calculation time of each of the bandwidth selection methods. Similar to the British family expenditure data, the new bandwidth selector and plug-in method obtain the optimal bandwidth in terms of prediction, but with much less calculation time, which is desirable in real data applications.

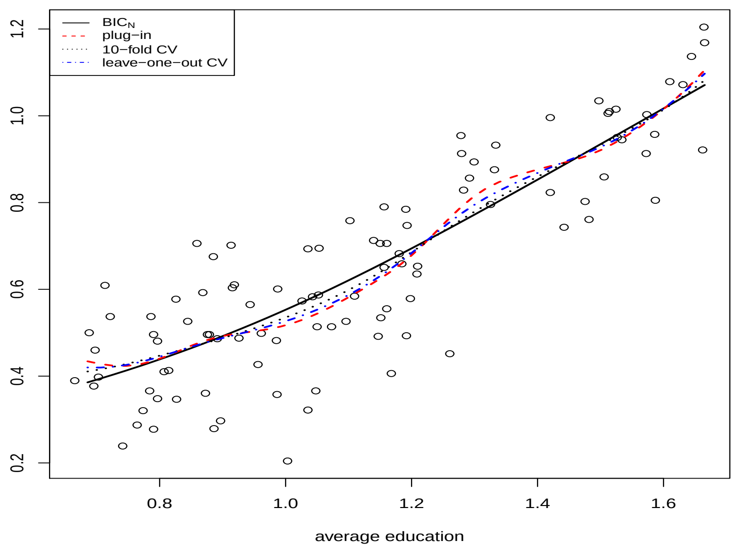

Figure 6: Scatter plot of the Canadian prestige data, and its corresponding fitted models.

5 Discussion

In this article, we proposed a bandwidth selector for local linear estimator, based on information criterion with the degrees of freedom originally derived for nonparametric inferences. The method can be easily implemented using any statistical software and is intuitively appealing. Simulation studies and real data examples show that the new selector outperforms the cross-validation method in terms of the estimation of the mean functions and calculation time, and is less variable in most cases. In addition, unlike the plug-in method, the new method does not require preliminary parameters to be chosen in advance, and therefore, is desirable in real data applications.

In this article, we only investigate the bandwidth selector of local linear estimator. It is also of great interest to extend our work to the local polynomial context based on the work of Zhang (2003). Applications to other nonparametric regression models, such as varying coefficient models and varying coefficient partial linear models, would also be valuable.

References

Craven, P. and Wahba G. (1979). Smoothing noisy data with spline functions. *Numerical Mathematics*, 31, 377-403.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *The Annuals of Statistics*, 21, 196-216.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annuals of Statistics*, 20, 2008-2036.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annuals of Statistics*, 29, 153-193.

Fox, J. and Weisberg, S. (2011). An R Companion to Applied Regression, Second Edition. Sage

Grund, B., Hall, P. and Marron, J. S. (1994). Loss and Risk in smoothing parameter selection. *Journal of Nonparametric Statistics*, 4, 107-132.

Hall, P. and Johnstone, I. (1992). Empirical functionals and efficient smoothing parameter selection (with discussion). *Journal of the Royal Statistical Society: Series B*, 54, 475-530.

Hall, P. and Marron, J. S. (1991). Lower bounds for bandwidth selection in density estimation. *Probability Theory and Related Fields*, 90, 149-173.

Härdel, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, 83, 86-99.

Hurvich, C.M., Simonoff, J.S. and Tsai, C. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society B*, 60(2), 271-293.

Jones, M. C. and Kappenman, R. F. (1991). On a class of kernel density estimate bandwidth selectors. *Scandinavian Journal of Statistics*, 19, 337-349.

Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: a possible approach. *Computational Statistics & Data Analysis*, 50, 3405-3431.

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

Shang, H. L. (2013). Bayestian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Computational Statistics and Data Analysis*, 67, 185-198.

Sun, Y. and Li, Q. (2011). Data-driven bandwidth selection for nonstationary semiparametric models. *Journal of Business & Economic Statistics*, 29(4), 541-551.

Table 2: Mean(Std) of ASE of the mean functions.

	Sample Size	BIC_N	plug-in	10-fold CV	leave-one-out CV
	$n = 100$	0.201(0.131)	0.235(0.189)	0.215(0.159)	0.212(0.157)
Example 1	$n = 200$	0.105(0.073)	0.117(0.082)	0.113(0.087)	0.113(0.088)
	$n = 400$	0.055(0.037)	0.063(0.045)	0.061(0.049)	0.062(0.051)
	$n = 100$	0.084(0.485)	0.135(0.598)	0.167(2.571)	0.172(2.572)
Example 2	$n = 200$	0.072(0.215)	0.096(0.620)	0.166(1.361)	0.144(1.212)
	$n = 400$	0.042(0.092)	0.077(0.262)	0.083(0.485)	0.083(0.485)
	$n = 100$	0.040(0.038)	0.069(0.377)	0.056(0.218)	0.070(0.555)
Example 3	$n = 200$	0.037(0.034)	0.041(0.041)	0.039(0.037)	0.039(0.043)
	$n = 400$	0.028(0.027)	0.032(0.044)	0.031(0.046)	0.030(0.029)
	$n = 100$	0.117(1.310)	0.748(5.685)	0.108(0.103)	0.102(0.078)
Example 4	$n = 200$	0.095(0.059)	0.397(3.902)	0.072(0.077)	0.082(0.130)
	$n = 400$	0.051(0.045)	0.185(1.755)	0.065(0.148)	0.057(0.069)
	$n = 100$	1.135(0.501)	1.345(1.027)	1.241(0.749)	1.255(0.784)
Example 5	$n = 200$	0.760(0.356)	0.718(0.409)	0.732(0.416)	0.724(0.408)
	$n = 400$	0.416(0.233)	0.393(0.207)	0.408(0.250)	0.408(0.257)
	$n = 100$	0.068(0.036)	0.106(0.069)	0.085(0.058)	0.087(0.065)
Example 6	$n = 200$	0.050(0.019)	0.062(0.034)	0.055(0.028)	0.055(0.027)
	$n = 400$	0.037(0.010)	0.036(0.017)	0.037(0.017)	0.038(0.019)
	$n = 100$	2.754(1.555)	4.197(4.176)	3.295(2.497)	3.410(3.144)
Example 7	$n = 200$	1.857(0.966)	2.054(1.468)	1.908(1.232)	1.999(1.617)
	$n = 400$	1.117(0.627)	1.142(0.698)	1.132(0.668)	1.117(0.774)
	$n = 100$	0.060(0.051)	0.126(0.104)	0.079(0.080)	0.082(0.084)
Example 8	$n = 200$	0.036(0.026)	0.065(0.047)	0.047(0.043)	0.047(0.043)
	$n = 400$	0.022(0.014)	0.033(0.024)	0.027(0.026)	0.028(0.029)

Table 3: Mean(Std) of calculation time (in seconds) of a repetition.

	Sample Size	BIC_N	plug-in	10-fold CV	leave-one-out CV
	$n = 100$	0.184(0.071)	0.029(0.017)	0.606(0.232)	4.746(1.764)
Example 1	$n = 200$	0.139(0.022)	0.019(0.003)	0.492(0.068)	6.128(0.805)
	$n = 400$	0.165(0.031)	0.020(0.004)	0.773(0.129)	12.482(2.013)
	$n = 100$	0.187(0.079)	0.032(0.033)	0.603(0.234)	4.758(1.771)
Example 2	$n = 200$	0.150(0.029)	0.023(0.022)	0.533(0.105)	6.666(1.260)
	$n = 400$	0.168(0.029)	0.023(0.008)	0.806(0.127)	12.934(2.058)
	$n = 100$	0.182(0.074)	0.027(0.011)	0.567(0.233)	4.467(1.802)
Example 3	$n = 200$	0.153(0.035)	0.022(0.005)	0.546(0.127)	6.782(1.537)
	$n = 400$	0.175(0.029)	0.022(0.004)	0.828(0.141)	13.116(2.021)
	$n = 100$	0.165(0.067)	0.025(0.010)	0.512(0.214)	4.016(1.634)
Example 4	$n = 200$	0.155(0.038)	0.023(0.005)	0.546(0.130)	6.792(1.508)
	$n = 400$	0.173(0.028)	0.023(0.004)	0.816(0.125)	13.083(2.078)
	$n = 100$	0.161(0.045)	0.024(0.006)	0.497(0.140)	3.921(1.101)
Example 5	$n = 200$	0.205(0.003)	0.026(0.015)	0.745(0.223)	5.519(1.225)
	$n = 400$	0.215(0.092)	0.025(0.011)	1.025(0.439)	16.662(7.227)
	$n = 100$	0.161(0.044)	0.024(0.006)	0.500(0.143)	3.955(1.104)
Example 6	$n = 200$	0.221(0.058)	0.029(0.010)	0.796(0.206)	9.962(2.550)
	$n = 400$	0.208(0.064)	0.025(0.007)	0.993(0.295)	16.007(4.568)
	$n = 100$	0.183(0.096)	0.027(0.014)	0.567(0.301)	4.462(2.363)
Example 7	$n = 200$	0.224(0.091)	0.029(0.012)	0.815(0.335)	10.219(4.139)
	$n = 400$	0.222(0.082)	0.026(0.010)	1.059(0.393)	17.284(6.442)
	$n = 100$	0.195(0.104)	0.029(0.014)	0.608(0.326)	4.803(2.556)
Example 8	$n = 200$	0.227(0.068)	0.030(0.012)	0.826(0.259)	10.331(3.159)
	$n = 400$	0.188(0.054)	0.022(0.007)	0.895(0.261)	14.629(3.814)

Table 4: Variance of selected bandwidths over 500 repetitions.

	Sample Size	BIC_N	plug-in	10-fold CV	leave-one-out CV
Example 1	$n = 100$	0.014	0.001	0.010	0.009
	$n = 200$	0.004	0.001	0.003	0.003
	$n = 400$	0.001	0.000	0.001	0.001
Example 2	$n = 100$	0.000	0.000	0.002	0.002
	$n = 200$	0.000	0.000	0.002	0.002
	$n = 400$	0.002	0.000	0.001	0.001
Example 3	$n = 100$	0.581	0.024	0.667	0.671
	$n = 200$	0.528	0.011	0.391	0.366
	$n = 400$	0.214	0.006	0.082	0.076
Example 4	$n = 100$	0.002	0.003	0.006	0.005
	$n = 200$	0.001	0.001	0.004	0.004
	$n = 400$	0.002	0.001	0.003	0.002
Example 5	$n = 100$	0.026	0.001	0.029	0.030
	$n = 200$	0.018	0.000	0.007	0.005
	$n = 400$	0.012	0.000	0.004	0.004
Example 6	$n = 100$	0.014	0.002	0.034	0.034
	$n = 200$	0.018	0.001	0.038	0.039
	$n = 400$	0.035	0.001	0.034	0.034
Example 7	$n = 100$	0.019	0.001	0.028	0.029
	$n = 200$	0.022	0.001	0.018	0.017
	$n = 400$	0.012	0.000	0.007	0.005
Example 8	$n = 100$	0.008	0.001	0.026	0.027
	$n = 200$	0.010	0.001	0.029	0.029
	$n = 400$	0.016	0.001	0.030	0.030

Table 5: MSE(Bias) of bandwidths.

	Sample Size	BIC_N	plug-in	10-fold CV	leave-one-out CV
	$n = 100$	0.018(0.082)	0.014(-0.102)	0.010(-0.006)	0.009(-0.010)
Example 1	$n = 200$	0.006(0.048)	0.004(-0.059)	0.004(-0.007)	0.004(-0.011)
	$n = 400$	0.001(0.029)	0.003(-0.040)	0.002(-0.006)	0.002(-0.009)
	$n = 100$	0.000(0.003)	0.003(-0.040)	0.003(-0.004)	0.003(-0.007)
Example 2	$n = 200$	0.001(0.024)	0.003(-0.043)	0.003(-0.013)	0.003(-0.016)
	$n = 400$	0.002(-0.014)	0.003(-0.045)	0.004(-0.040)	0.004(-0.046)
	$n = 100$	0.509(0.499)	0.757(-0.739)	0.667(0.035)	0.671(0.033)
Example 3	$n = 200$	0.321(0.341)	0.372(-0.483)	0.391(-0.005)	0.367(-0.040)
	$n = 400$	0.151(0.186)	0.113(-0.286)	0.084(-0.045)	0.079(-0.054)
	$n = 100$	0.005(0.050)	0.021(-0.125)	0.008(0.004)	0.006(-0.006)
Example 4	$n = 200$	0.002(0.035)	0.015(-0.108)	0.005(-0.003)	0.006(-0.014)
	$n = 400$	0.003(0.022)	0.011(-0.092)	0.005(-0.014)	0.004(-0.021)
	$n = 100$	0.037(0.126)	0.035(-0.160)	0.029(-0.009)	0.030(-0.005)
Example 5	$n = 200$	0.027(0.102)	0.013(-0.087)	0.018(0.003)	0.016(-0.004)
	$n = 400$	0.011(0.056)	0.003(-0.038)	0.004(-0.001)	0.004(-0.004)
	$n = 100$	0.019(0.073)	0.092(-0.281)	0.036(-0.049)	0.037(-0.051)
Example 6	$n = 200$	0.033(0.121)	0.076(-0.240)	0.039(-0.031)	0.040(-0.030)
	$n = 400$	0.053(0.157)	0.036(-0.145)	0.034(-0.021)	0.035(-0.021)
	$n = 100$	0.033(0.120)	0.041(-0.182)	0.028(-0.008)	0.029(-0.006)
Example 7	$n = 200$	0.030(0.109)	0.022(-0.129)	0.017(-0.005)	0.017(-0.011)
	$n = 400$	0.013(0.071)	0.007(-0.067)	0.007(0.000)	0.006(-0.008)
	$n = 100$	0.013(0.069)	0.097(-0.294)	0.026(-0.022)	0.027(-0.023)
Example 8	$n = 200$	0.021(0.100)	0.078(-0.261)	0.030(-0.020)	0.029(-0.008)
	$n = 400$	0.028(0.108)	0.059(-0.226)	0.030(-0.002)	0.030(-0.009)

Table 6: Mean(Std) of 10-fold CV.

	BIC_N	plug-in	10-fold CV	leave-one-out CV
1995 British family expenditure data				
MSPE	0.011(0.002)	0.011(0.002)	0.011(0.002)	0.011(0.002)
Time	0.183(0.008)	0.036(0.003)	3.843(0.041)	48.790(0.377)
Prestige data				
MSPE	0.017(0.009)	0.017(0.010)	0.017(0.010)	0.018(0.010)
Time	0.077(0.004)	0.014(0.001)	0.348(0.007)	2.627(0.024)