A Data Driven Selecting Rule for the Bandwidth of Regression Estimation with Generalized Additive Noise

Jiahao Mou, Cong Wu
School of Science, Hubei university of technology, Wuhan, 430068, China

Abstract: For the problem of bandwidth selection in nonparametric regression, most of the existing methods carried out theoretical analysis and numerical calculation with a fixed bandwidth. However, the selection of an appropriate bandwidth and the guarantee of good performance depend heavily on the parameters of the smoothness of regression function, which are difficult to calculate in practice. To overcome this problem, a novel and efficient data-driven selecting rule is proposed to adaptively determine the appropriate bandwidth. It turns out that the bandwidth only loses the $\ln n$ factor in terms of convergence rate by selecting rule.

Keywords: Data-driven; Regression estimation; The selecting rule of bandwidth

1. Introduction

The goal of this paper is to propose a data-driven selecting rule to adaptively determine the appropriate bandwidth. Firstly, we estimate a regression function m from independent and identically distributed (i.i.d) data $(W_j, Y_j)(j = 1,2,\dots,n)$ generated by the model

$$\mathrm{Y} = \mathrm{m}(\mathrm{X}) + \varepsilon, \mathrm{W} = \mathrm{X} + \mathrm{B} \cdot \delta \tag{1}$$

where X stands for a real-valued random variable with unknown probability density

f on R^d, δ denotes an independent random noise with the probability density g and $\varepsilon \in \{0,1\}$ Bernoulli random variable with $P(B = 1) = \alpha$, $\alpha \in [0,1]$.

When $\alpha = 1$, Eq. (1) reduces to the deconvolution model. For the study of errors-in-variables, Meister [1] studied regression estimation with one-dimensional data through kernel method. Guo [2] et al. extended Meister's theorems and studied the optimal convergence rate of the estimator with additive noise.

While $\alpha = 0$, Eq. (1) corresponds to the traditional regression model. Bouzebda [3] studied the bandwidth consistency of the kernel-type estimator in the case of weaker

kernel conditions and extended existing uniform bounds on kernel regression estimator. Pinelis [4] considered three common classes of kernel regression estimators and proved related properties. Ail [5] proposed a new improvement of the Nadaraya-Watson kernel nonparametric regression estimator. The bandwidth of this new improvement was obtained depending on the three different statistical indicators. Including comparisons with four others kernel estimators. The proposed estimator in the case of harmonic mean is more accurate than all classical methods.

Clearly, the density h_w of W in Eq. (1) satisfies

$$h_w = (1-\alpha)f + \alpha f * g \tag{2}$$

because $P\{W < t\} = P\{B = 0\}P\{X + B \cdot \delta|_{B=0} < t\} + P\{B = 1\}P\{X + B \cdot \delta|_{B=1} < t\} = (1 -$

$\alpha)P\{X < t\} + \alpha P\{X + \delta < t\}$. Furthermore,

$$f^{ft}(t) = [(1-\alpha) + \alpha g^{ft}(t)]^{-1} h_w^{ft}(t) := G_\alpha^{-1}(t) h_w^{ft}(t).$$

when the function

$$G_\alpha(t) = 1 - \alpha + \alpha g^{ft}(t)$$

has nonzero on R^d, where f^{ft} is the Fourier transform of $f \in L^1(R^d)$ defined by

$$f^{ft}(t) := \int_{R^d} f(x) e^{itx} dx.$$

To estimate the regression function, we need the assumption of function space [4].

$\zeta_{x,s} = \{(m, f_X) | \|f_X\|_\infty + \|m^2 f_X\|_\infty \le C_1, mf_X \in L(R^d); f_X(x) \ge C_2; \}$ (1.3)

f_X and mf_X satisfy local Holder condition of order s at $x \in R^d$ for some positive constants $C_1, C_2 > 0$. A function f is said to satisfy local Holder condition of orders, if there exists a constant $C > 0$ such that for each β with $|\beta| = \lfloor s \rfloor$,

$$\left|\partial^\beta f(y) - \partial^\beta f(z)\right| \le C|y - z|^{s-|\beta|} \text{for} y, z \in Q(x, r)$$

where $Q(x, r) := \{y = (y_1, y_2, \ldots, y_d) | y_i \in (x_i - r, x_i + r), i = 1,2,\ldots,d\}$with $r > 0$.

For the research of data-driven, Wu et al. [6] used a linear wavelet estimator to obtain a point-wise optimal estimation. Then they used a data-driven method to obtaining adaptive and near-optimal estimation and showed the logarithmic factor necessary to get the adaptivity. Kong et al. [7] proposed the local polynomial regression based on bimodal kernels for the derivative estimation under correlated errors. Based on the asymptotic mean integrated squared error, they also provided a data-driven bandwidth selection criterion.

Unlike traditional research, Chen [8] constructed a linear wavelet estimator of the anisotropic regression function, and proposed a regression estimator based on the scale parameter data-driven selection rule. It turned out that results attain the optimal convergence rate of nonparametric pointwise estimation. Kim et al. [9] studied selecting the number of change points in segmented line regression and proposed a new method based on two Schwarz type criteria. The proposed method is computationally much more efficient than previous ones. For l(0) penalized (nonlinear) regression problems, Li [10] proposed a novel and efficient data-driven line search rule to adaptively determine the appropriate step size based on the idea of support detection and root finding. Liu [11]-[12] provided a new data-driven bandwidth selection method for kernel quantile estimators. The effectiveness of this rule is confirmed by numerical experiments. Bagkavos [13] studied asymptotic consistency and distribution of the practically useful (data-driven) version of the bandwidth rule. The potential of the method as a data-analytic tool is illustrated by application. El-Dakkak [14] studied probabilistic models with adaptive algorithms that accurately fit wind speed distributions and a non-parametric combinatorial method is implemented. It is worthwhile mentioning that the implemented procedure is adaptive (i.e. data driven) and robust. Wylupek [15] developed a new solution of the general nonparametric k-sample problem for independent continuous random variables. They solved the testing problem for members of this approximating net by a data-driven test. Simulations show that the new omnibus test has power comparable to existing k-sample tests in case of changes of location or scale.

2. Data-Driven

In this section, we use data-driven rule to select a optimal bandwidth h and only lose the $\ln n$ factor in terms of convergence rate.

We introduce some notations used frequently later on.

For two variables A and B, $A \lesssim B$ denotes $A \leq CB$ for some positive constant C. $A \gtrsim B$ means $A \lesssim B$. We use $A \sim B$ to stand for both $A \lesssim B$ and $B \lesssim A$.

We introduce some lemmas used later on.

Lemma 1 (Bernstein inequality) Let $i.i.d$ data $X_1,\ldots,X_n$ satisfy $EX_i = 0$ and $|X_i| \leq \|X\|_\infty$. Then

$$P|\frac{1}{n}\sum_{i=1}^{n} X_i| \geq \gamma \leq 2\exp(-\frac{n\gamma^2}{2[E{X_i}^2 + \|X\|_\infty \gamma/3]}).$$

The regression kernel estimator is defined by

$$\widehat{m_h}(x) = \frac{\hat{p}_h(x)}{\hat{f}_h(x)} = \frac{\sum_{j=1}^{n} Y_j \int_{R^d} e^{-itx} K^{ft}(ht) e^{it\omega_j}/G_\alpha(t)\,dt}{\sum_{j=1}^{n} \int_{R^d} e^{-itx} K^{ft}(ht) e^{it\omega_j}/G_\alpha(t)\,dt}.$$

Lemma 2 Let $p := mf_X$ with $|m(x)| < +\infty$ and $f(x) > 0$. Then $\widehat{m}(x)$ defined in above satisfies that

$$P[|\widehat{m}(x) - m(x)|^2 > \varepsilon] \leq P[|\hat{p}(x) - p(x)|^2 \gtrsim \varepsilon] + P[|\hat{f}(x) - f(x)|^2 \gtrsim \varepsilon]$$

for $\varepsilon > 0$ small enough.

Then we have proved the following result.

Lemma 3 For $h_1 = 2^{-k_0} \sim n^{-\frac{1}{2s+2\beta(\alpha)+d}}, k_0 \in N$ and $1 \leq p < \infty$,

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} \sup P[\left|\widehat{m_{h_1}}(x) - m(x)\right|^2 \geq cn^{-\frac{2s}{2s+2\beta(\alpha)+d}}] = 0.$$

Lemma 4 Since $W_1, W_2, \ldots, W_n$ are i.i.d., the variance term

$$Var\hat{f}(x) \leq (2\pi)^{-2d} n^{-1} E\left|\int_{R^d} e^{-itx} K^{ft}(ht) e^{itW_1}/G_\alpha(t)\,dt\right|^2$$

$$= (2\pi)^{-2d} n^{-1} \int_{R^d} \left|\int_{R^d} e^{-itx} K^{ft}(ht) e^{itw}/G_\alpha(t)\,dt\right|^2 h_w(w)dw.$$

By $z = x - w, Var\hat{f}(x) \leq (2\pi)^{-2d} n^{-1} \int_{R^d} \left|\int_{R^d} e^{-itz} K^{ft}(ht)/G_\alpha(t)\,dt\right|^2 h_w(x-z)dz$. By $(m, f_X) \in \zeta_{x,s}$, $\|f_X\|_\infty \leq C_1$ and $\|f_X * g\|_\infty \leq \|f_X\|_\infty \|g\|_1 \leq C_1 \|g\|_1$.

Hence,

$$Var\hat{f}(x) \lesssim (2\pi)^{-2d} n^{-1} \int_{R^d} \left|\int_{R^d} e^{-itx} K^{ft}(ht) e^{itW_1}/G_\alpha(t)\,dt\right|^2 dz$$

Moreover, the above inequality becomes to

$$Var\hat{f}(x) \lesssim \mathrm{n}^{-1} \int_{[-1/\mathrm{h},1/\mathrm{h}]^{\mathrm{d}}} \left|\mathrm{K}^{\mathrm{ft}}(\mathrm{ht})/\mathrm{G}_\alpha(\mathrm{t})\right|^2 \mathrm{dt}$$

thanks to Parseval identity and supp $K^{ft} \subset [-1,1]^d$. Because $|G_\alpha(t)| \gtrsim (1+\mathrm{t}^2)^{-\frac{\beta(\alpha)}{2}}, Var\hat{f}(x) \lesssim n^{-1} h^{-(2\beta(\alpha)+d)}$.

In order to introduce data-driven, denote h, $h_* := \max\{h, h_*\}$ and

$$\hat{\mu}_h := \lambda \sqrt{\frac{h^{-(2\beta(\alpha)+d)} \ln n}{n}} \tag{3}$$

with the constant λ specified late on.

Let $\Omega := \{2^{-k}, \ldots, 2^{-1}, 1\}$, $k = max\{\mathrm{k} \cdot 2^{\mathrm{k}} \leq \mathrm{n}\}$, $x_+ := \max\{0, x\}$, and the h_0 be defined by the following rule.

$$\widehat{\xi_h}(x_0) := max[|\widehat{m}_{h,h_*}(x_0) - \widehat{m}_{h_*}(x_0)| - \widehat{\mu_{h_*}} - \widehat{\mu_h}]_+$$
$$\hat{\xi}_{h_0}(x_0) + 2\hat{\mu}_{h_0} := min[\hat{\xi}_h(x_0) + 2\hat{\mu}_h]$$

Theorem 1 For h_0 given as above,

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[|\widehat{m_{h_0}}(x) - m(x)|^2 \geq c(lnn)n^{-\frac{2s}{2s+2\beta(\alpha)+d}} \tag{4}$$

Proof Let h_1 be given in Lemma 3. Then

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[|\widehat{m_{h_1}}(x) - m(x)|^2 \geq cn^{-\frac{2s}{2s+2\beta(\alpha)+d}} \tag{5}$$

By (i) and (ii), $|\widehat{m}_{h_1,h_0} - \widehat{m}_{h_0}| + |\widehat{m}_{h_1,h_0} - \widehat{m}_{h_1}| \leq (\widehat{\xi_{h_1}} + \widehat{\mu_{h_0}} + \widehat{\mu_{h_1}}) + (\widehat{\xi_{h_0}} + \widehat{\mu_{h_0}} + \widehat{\mu_{h_1}}) = (\widehat{\xi_{h_0}} + 2\widehat{\mu_{h_0}}) + (\widehat{\xi_{h_1}} + 2\widehat{\mu_{h_1}}) \leq 2(\widehat{\xi_{h_1}} + 2\widehat{\mu_{h_1}}) \lesssim \widehat{\xi_{h_1}} + \widehat{\mu_{h_1}}$.

According to Eq. (3) and the choice of h_1, $\hat{\mu}^2{}_{h_1} \lesssim (lnn)n^{-\frac{2s}{2s+2\beta(\alpha)+d}}$. On the other hand, $|\widehat{m}_{h_1} - m|^2$ is estimated by Lemma 3. Hence, it suffices for the desired conclusion Eq. (4) to show

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP\left[\widehat{\xi_{h_1}}^2 \geq c(lnn)n^{-\frac{2s}{2s+2\beta(\alpha)+d}}\right] = 0 \tag{6}$$

Note that $\widehat{m}_{h_1,h_*} = \widehat{m}_{h_*}$ for $h_* \geq h_1$ and $\widehat{m}_{h_1,h_*} = \widehat{m}_{h_1}$ for $h_* < h_1$. Then $\widehat{\xi_{h_1}} = max(|\widehat{m}_{h_1,h_*} - \widehat{m}_{h_*}| - \widehat{\mu_{h_*}} - \widehat{\mu_{h_1}})_+ = max(|\widehat{m}_{h_1,h_*} - \widehat{m}_{h_*}| - \widehat{\mu_{h_*}} - \widehat{\mu_{h_1}})_+$. Moreover,

$$\widehat{\xi_{h_1}} \leq max(|\widehat{m}_{h_1} - m| + |m - \widehat{m}_{h_*}| - \widehat{\mu_{h_*}} - \widehat{\mu_{h_1}})_+$$

On the other hand, $(f + g + h)_+ \leq f_+ + g_+ + h_+$. Hence,

$$\widehat{\xi_{h_1}} \leq (|\widehat{m}_{h_1} - m| - \hat{\mu}_{h_1})_+ + max(|m - \widehat{m}_{h_*}| - \hat{\mu}_{h_*})_+.$$

Furthermore, we only need to prove that for $h \in \Omega$ and $2^{-k} \leq h \leq h_1$,

$$\operatorname{Lim}_{c\to\infty} \overline{\lim_{n\to\infty}} supP\left[(|\widehat{m}_h - m| - \hat{\mu}_h)_+ \geq \frac{c}{2}(lnn)^{\frac{1}{2}} n^{-\frac{s}{2s+2\beta(\alpha)+d}}\right] = 0 \tag{7}$$

According to Lemma 2 and Eq. (3)

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[(|\hat{p}_h(x) - p(x)|) \geq cC_1\hat{\mu}_h] + P[(|\hat{f}_h(x) - f(x)|) \geq cC_1\hat{\mu}_h] = 0.$$

For the $|\hat{f}_h(x) - f(x)|$, one knows that

$$|\hat{f}_h(x) - f(x)| = |\hat{f}_h(x) - E\hat{f}_h + E\hat{f}_h - f(x)| \leq |\hat{f}_h(x) - E\hat{f}_h| + |E\hat{f}_h - f(x)|.$$

Hence,

$$P[|\widehat{f_h}(x) - f(x)| \geq cC_1\widehat{\mu_h}] \leq P[|\widehat{f_h}(x) - E\widehat{f_h}(x)| \geq cC_1\widehat{\mu_h}] + P[|E\widehat{f_h}(x) - f(x)| \geq cC_1\widehat{\mu_h}]$$

By the proof of Theorem 1, $|E\widehat{f_h}(x) - f(x)| \leq C_2h^s \leq C_2h_1^s = C_2n^{-\frac{s}{2s+2\beta(\alpha)+d}}$

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[(|\widehat{Ef}_h(x) - f(x)|) \geq cC_1\hat{\mu}_h] = 0.$$

According to the definition of $\hat{f}_h(x)$,

$$\hat{f}_h(x) = \frac{1}{n}\sum_{j=1}^{n} \int_{R^d} \frac{e^{-itx}K^{ft}(ht)e^{it\omega_j}}{G_\alpha(t)dt}$$

and $\hat{f}_h(x) - E\hat{f}_h(x) = \frac{1}{n}\sum_{j=1}^{n} M_j$ with

$$M_j := \int_{R^d} \frac{e^{-itx}K^{ft}(ht)[e^{it\omega_j} - E(e^{it\omega_j})]}{G_\alpha(t)dt}$$

Clearly, $\{M_j\}$ are i.i.d and $EM_j = 0$. According to lemma 4, it is easy to see that

$$EM_j^2 \leq \int \left| \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{it\omega_j}}{G_\alpha(t) dt} \right|^2 h_\omega(\omega) d\omega \leq C_3 h^{-(2\beta(\alpha)+d)}$$

and $\|M_j\|_\infty \leq \left\| \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{it\omega_j}}{G_\alpha(t) dt} \right\|_\infty \leq C_4 h^{-(\beta(\alpha)+d)}$. Using Lemma 1 and the definition of $\hat{\mu}_h$, one knows that

$$P[|\hat{f}_h(x) - E\hat{f}_h(x)| \geq C_1 c\hat{\mu}_h] \leq 2\exp\{-\frac{n(C_1 c\hat{\mu}_h)^2}{2\left[C_3 h^{-(2\beta(\alpha)+d)} + \frac{C_4 h^{-(\beta(\alpha)+d)} C_1 c\hat{\mu}_h}{3}\right]}\} \tag{8}$$

Since $\hat{\mu}_h \lesssim 1$ thanks to (2.1). Furthermore,

$$2\left[C_3 h^{-(2\beta(\alpha)+d)} + \frac{C_4 h^{-(\beta(\alpha)+d)} C_1 c\hat{\mu}_h}{3}\right] \leq C_5 c h^{-(2\beta(\alpha)+d)}$$

and (2.6) reduce to

$$P[|\hat{f}_h(x) - E\hat{f}_h(x)| \geq C_1 c\hat{\mu}_h] \leq 2\exp\{-\frac{{C_1}^2 \lambda^2}{C_5} c(\ln n)$$

thanks to (2.1), choose λ such that $\lambda^2 > \frac{C_5}{{C_1}^2}$. Then

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[(|\hat{f}_h(x) - E\hat{f}_h|) \geq cC_1\hat{\mu}_h] = 0.$$

Therefore, $\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[(|\hat{f}_h(x) - f(x)|) \geq cC_1\hat{\mu}_h] = 0$. Similar to $P[|\hat{f}_h(x) - f(x)|]$, $\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP[(|\hat{p}_h(x) - p(x)|) \geq cC_1\hat{\mu}_h] = 0$. Finally, it follows Eq. (8) and the choice of h_1 that

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} supP\left[\widehat{\xi_{h_1}}^2 \geq c(lnn) n^{-\frac{2s}{2s+2\beta(\alpha)+d}}\right] = 0$$

which Eq. (4). The proof is done.

In our Theorem 1, the convergence rate is consistent with Theorem 1 in only lose the $\ln n$ factor by data-driven. The data-driven methods of regression estimation and density estimation [6] are based on the selection rules of estimators. Due to the particularity of the proof in regression estimation, the splitting of m should be after applying the selection rules.

3. Conclusion

In this paper, we study the regression estimation with generalized additive noise, which includes the classical regression estimation and regression estimation with additive noise. However, the selection of an appropriate bandwidth and the guarantee of good performance depend heavily on the parameters of the smoothness of regression function, which are difficult to calculate in practice. Therefore, a novel and efficient data-driven selecting rule is proposed to adaptively determine the appropriate bandwidth. It turns out that the bandwidth only loses the $\ln n$ factor in terms of convergence rate by selecting rule.

Acknowledge

This paper is supported by National Natural Science Foundation of China (NO.12201194). The authors would like to thank the referees for their important comments and suggestions.

References

[1] Alexander, M.: Deconvolution Problems in Nonparametric Statistics. Springer, Berlin (2009). https://doi.org/10.1007/978-3-540-87557-4

[2] Guo, H., Liu, Y.: Convergence rates of multivariate regression estimators with errors-in-variables. NUMERICAL FUNCTIONAL ANALYSIS AND OPTIMIZATION. 38, 1564–1588 (2017). https://doi.org/10.1080/01630563.2017.1349145

[3] Bouzebda, S., El-hadjali, T.: Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. JOURNAL OF NONPARAMETRIC STATISTICS. 32, 864–914 (2020). https://doi.org/10.1080/10485252.2020.1834107

[4] Pinelis, I.: Monotonicity preservation properties of kernel regression estimators. STATISTICS & PROBABILITY LETTERS. 177 (2021) https://doi.org/10.1016/j.spl.2021.109157

[5] Ali, T.H.: Modification of the adaptive nadaraya-watson kernel method for nonparametric regression (simulation study). COMMUNICATIONS IN STATISTICS-SIMULATION AND COMPUTATION. 51, 391–403 (2022) https://doi.org/10.1080/03610918.2019.1652319

[6] Liu, Y., Wu, C.: Point-wise wavelet estimation in the convolution structure density model. JOURNAL OF FOURIER ANALYSIS AND APPLICATIONS. 26 (2020) https://doi.org/10.1007/s00041-020-09794-y

[7] Kong, D., Zhao, S., Wang, W.: Nonparametric derivative estimation with bimodalkernels under correlated errors. COMPUTATIONAL STATISTICS. 39, 1847– 1865 (2024) https://doi.org/10.1007/s00180-023-01419-4

[8] Chen, J., Kou, J.: Pointwise estimation of anisotropic regression functions using wavelets with data-driven selection rule. MATHEMATICS. 12

[9] Kim, H.-J., Chen, H.-S., Midthune, D., Wheeler, B., Buckman, D.W., Green, D., Byrne, J., Luo, J., Feuer, E.J.: Data-driven choice of a model selection method in joinpoint regression. JOURNAL OF APPLIED STATISTICS. 50, 1992–2013 (2023) https://doi.org/10.1080/02664763.2022.2063265

[10] Li, P., Jiao, Y., Lu, X., Kang, L.: A data-driven line search rule for support recovery in high-dimensional data analysis. COMPUTATIONAL STATISTICS & DATA ANALYSIS. 174 (2022) https://doi.org/10.1016/j.csda.2022.107524

[11] Liu, X., Song, Y., Cheng, H.-F., Zhang, K.: A bootstrap-based bandwidth selection rule for kernel quantile estimators. COMPUTATIONAL STATISTICS. 40, 4037–4058 (2025) https://doi.org/10.1007/s00180-024-01582-2

[12] Liu, X., Song, Y., Zhang, K.: An exact bootstrap-based bandwidth selection rule for kernel quantile estimators. COMMUNICATIONS IN STATISTICS-SIMULATION AND COMPUTATION. 53, 3699–3720 (2024) https://doi.org/10.1080/03610918.2022.2110595

[13] Bagkavos, D., Patil, P.N., Ramanathan, T.V.: Local linear estimation and plug-in bandwidth selection for the ratio of two hazard rate functions. JOURNAL OF NONPARAMETRIC STATISTICS. (2025) https://doi.org/10.1080/10485252.2025.2466650

[14] El-Dakkak, O., Feng, S., Wahbah, M., EL-Fouly, T.H.M., Zahawi, B.: Combinatorial method for bandwidth selection in wind speed kernel density estimation. IET RENEWABLE POWER GENERATION. 13, 1670–1680 (2019) https://doi.org/10.1049/iet-rpg.2018.5643

[15] Wylupek, G.: Data-driven k-sample tests. TECHNOMETRICS. 52, 107–123 (2010) https://doi.org/10.1198/TECH.2009.08101