

Regression Estimation with Generalized Additive Noise

Jiahao Mou, Cong Wu

School of Science, Hubei University of Technology, Wuhan 430068, China

Abstract: In this paper, we study the regression estimation with generalized additive noise, which includes the classical regression estimation and regression estimation with additive noise. Based on the generalized additive noise model, we build a regression estimator and study its convergence rate.

Keywords: Regression estimation; Generalized additive noise model; Convergence rate

1. Introduction

Nonparametric regression models play an important role in medical statistics and economic statistics. Caron [1] examined the problem of estimating heterogeneous treatment effects using non-parametric regression-based methods, starting from an empirical study aimed at investigating the effect of participation in school meal programs on health indicators. Almanjahie [2] proposed a family of robust nonparametric estimators with unknown scale parameters for regression function based on the kernel method. The superiority of the proposed methods was shown through numerical and real data studies.

2. Construction of Kernel-Based Estimators for the Generalized Additive Noise Model

The goal of this paper is to 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

$$Y = m(X) + \varepsilon, W = X + B \cdot \delta \quad (1)$$

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 $\alpha = 1$, (1.1) reduces to the deconvolution model. For the study of errors-in-variables, Meister [3] studied regression estimation with one-dimensional data through kernel method. Guo [4] et al. extended Meister's theorems and studied the optimal convergence rate of the estimator with additive noise.

While $\alpha = 0$, (1.1) corresponds to the traditional regression model. Bouzebda [5] 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 [6] considered three common classes of kernel regression estimators and proved related properties. Ail [7] 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.

Research on kernel methods, Chu et al. [8] proposed five point-wise consistent and asymptotic normal estimators of the asymptotic variance function of the Nadaraya-Watson kernel estimator for nonparametric regression. Cleanthous [9] considered the density estimation problem of observations taking values in classical or non-classical spaces. Kernel and linear and nonlinear wavelet density estimators are studied. The convergence rates of these estimators are established and discussed.

In the study of wavelet estimation, Chesneau et al. [10] studied the estimation problem in a class of nonparametric regression models with multiplicative and additive noise. In this context, two kinds of wavelet estimation are proposed. And it is proved that they have faster convergence speed in Besov space. Kou et al. [11] studied a class of generalized nonparametric regression estimation models with multiplicative noise. And they proved the convergence rates of these regression estimators under point-wise error in Besov space.

The above scholars have studied the classical regression estimation and regression estimation with additive noise. In fact, the presence of noise within the data remains uncertain. Therefore, generalized additive noise model is introduced for the regression estimation. It can reduce to classical regression estimation as well as regression estimation with additive noise. More references are in [12-15].

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

$$h_w = (1 - \alpha)f + \alpha f * g \quad (2)$$

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,

when the function

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

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

has nonzeros 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 as in [4].

$$\zeta_{x,s} = \{(m, f_x) | \|f_x\|_\infty + \|m^2 f_x\|_\infty \leq C_1, m f_x \in L(R^d); f_x(x) \geq C_2; \} \quad (3)$$

f_x and $m f_x$ 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 $|\beta| = [s]$.

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

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

In the second section of the article, we estimate the regression function of generalized additive noise model through kernel method. We demonstrate the convergence rate $n^{-\frac{s}{2s+2\beta(\alpha)+d}}$ of the estimator with generalized additive noise.

3. The Convergence Rate of Regression Estimator with Generalized Additive Noise

This section aims to give a convergence rate of regression estimators $\hat{m}(x)$ for the function m in model (1).

To introduce our regression estimator, we firstly choose a kernel function $K(x)$ satisfying

$$(i) K \in L^1(R^d) \cap L^2(R^d) \text{ and } \text{supp} K^{ft} \subset [-1, 1]^d$$

$$(ii) \int_{R^d} K(x) x^\gamma dx = \delta_{0,|\gamma|} \text{ for } |\gamma| = 0, 1, \dots, n$$

$$(iii) \int_{R^d} K(x) |x|^s dx < \infty.$$

Defining $K_h(x) = \frac{1}{h} K(\frac{x}{h})$. Then, our regression estimator with generalized additive noise $\hat{m}$ is built as

$$\hat{m}(x) = \frac{\hat{p}(x)}{\hat{f}(x)} \quad (4)$$

where

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{(2\pi)^d} \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{it\omega_j}}{G_\alpha(t)} dt \quad (5)$$

$$\hat{p}(x) = \frac{1}{n} \sum_{j=1}^n Y_j \frac{1}{(2\pi)^d} \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{it\omega_j}}{G_\alpha(t)} dt \quad (6)$$

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 $B \lesssim A$.

A lemma is derived from [3], in order to obtain the convergence rate.

Lemma 1 Let $p := mf_X$ with $|m(x)| < +\infty$ and $f(x) > 0$. Then $\hat{m}(x)$ defined in (4)-(6) satisfies that $P[|\hat{m}(x) - m(x)|^2 > \varepsilon] \leq P[|\hat{p}(x) - p(x)|^2 \gtrsim \varepsilon] + P[|\hat{f}(x) - f(x)|^2 \gtrsim \varepsilon]$ for $\varepsilon > 0$ small enough.

Theorem 1 For the model (1.1), the estimator $\hat{m}$ in (4) satisfies that if $|G_\alpha(t)| \gtrsim (1+t^2)^{-\frac{\beta(\alpha)}{2}}$

with $\beta(\alpha) = \begin{cases} \beta, \alpha = 1 \\ 0, \alpha \in [0, 1) \end{cases}$ and $h \sim n^{-\frac{1}{2s+2\beta(\alpha)+d}}$.

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P[|\hat{m}(x) - m(x)|^2 \geq cn^{-\frac{2s}{2s+2\beta(\alpha)+d}}] = 0.$$

Proof, for $0 < \varepsilon_n \rightarrow 0$, Lemma 1 and Markov inequality imply that

$$\begin{aligned} \sup P[|\hat{m}(x) - m(x)|^2 \geq c\varepsilon_n] &\lesssim \\ (c\varepsilon_n)^{-1} \sup \{E|\hat{f}(x) - f(x)|^2 + E|\hat{p}(x) - p(x)|^2\} \end{aligned} \quad (7)$$

In order to estimate $E|\hat{p}(x) - p(x)|^2 + E|\hat{f}(x) - f(x)|^2$, one needs only to deal with the variance terms $Var\hat{p}(x), Var\hat{f}(x)$ and the bias terms $|E\hat{p}(x) - p(x)|^2, |E\hat{f}(x) - f(x)|^2$.

Since $W_1, W_2, \dots, W_n$ are i.i.d., the variance term

$$\begin{aligned} Var\hat{f}(x) &\leq (2\pi)^{-2d} n^{-1} E \left| \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{itW_1}}{G_\alpha(t)} dt \right|^2 \\ &= (2\pi)^{-2d} n^{-1} \int_{R^d} \left| \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{itw}}{G_\alpha(t)} dt \right|^2 h_w(w) dw. \end{aligned}$$

By $z = x - w, Var\hat{f}(x) \leq (2\pi)^{-2d} n^{-1} \cdot \int_{R^d} \left| \int_{R^d} \frac{e^{-itz} K^{ft}(ht)}{G_\alpha(t)} dt \right|^2 h_w(x-z) dz$. 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,

$$\text{Var} \hat{f}(x) \lesssim (2\pi)^{-2d} n^{-1} \int_{R^d} \left| \int_{R^d} \frac{e^{-itx} K^{ft}(ht) e^{itW_1}}{G_\alpha(t)} dt \right|^2 dz \quad (8)$$

Moreover, the above inequality becomes to

$$\text{Var} \hat{f}(x) \lesssim n^{-1} \int_{[-1/h, 1/h]^d} |K^{ft}(ht)/G_\alpha(t)|^2 dt \text{ thanks to Parseval identity and } \text{supp} K^{ft} \subset [-1, 1]^d.$$

Because $|G_\alpha(t)| \gtrsim (1 + t^2)^{-\frac{\beta(\alpha)}{2}}$,

$$\text{Var} \hat{f}(x) \lesssim n^{-1} h^{-(2\beta(\alpha)+d)} \quad (9)$$

Similar to $\text{Var} \hat{f}(x)$, one finds that

$$\begin{aligned} \text{Var} \hat{p}(x) &\leq (2\pi)^{-2d} n^{-1} E \left| Y_1 \int_{R^d} e^{-itx} K^{ft}(ht) e^{itW_1} / G_\alpha(t) dt \right|^2 \\ &\leq (2\pi)^{-2d} n^{-1} \int_{R^d} E(|Y_1|^2 | X_1 = u) \cdot E \left| \int_{R^d} e^{-it(x-u)} K^{ft}(ht) e^{it\alpha\delta_1} / G_\alpha(t) dt \right|^2 f_X(u) du \\ &\leq (2\pi)^{-2d} n^{-1} \|f_X(\cdot) E(|Y_1|^2 | X_1 = \cdot)\|_\infty \cdot \int_{R^d} E \left| \int_{R^d} e^{-it(x-u)} K^{ft}(ht) e^{it\alpha\delta_1} / G_\alpha(t) dt \right|^2 du. \end{aligned}$$

Obviously, we shall assume that the product of f_X and the conditional expectation of Y_1^2 given $X_1 = x$ is uniformly bounded for all $x \in R$. Hence,

$$\begin{aligned} &= (2\pi)^{-2d} n^{-1} \|f_X(\cdot) E(|Y_1|^2 | X_1 = \cdot)\|_\infty \cdot \int_{R^d} E |K^{ft}(ht) e^{it\alpha\delta_1} / G_\alpha(t)|^2 dt \\ &= (2\pi)^{-2d} n^{-1} \|f_X(\cdot) E(|Y_1|^2 | X_1 = \cdot)\|_\infty \cdot \int_{R^d} |K^{ft}(ht) / G_\alpha(t)|^2 dt. \end{aligned}$$

See [3]. By Parseval identity, this above inequality reduces to

$$\text{Var} \hat{p}(x) \lesssim n^{-1} \int_{R^d} |K^{ft}(ht) / G_\alpha(t)|^2 dt.$$

The same arguments as (9) lead to

$$\text{Var} \hat{p}(x) \lesssim n^{-1} h^{-(2\beta(\alpha)+d)} \quad (10)$$

To estimate the bias terms $|E \hat{p}(x) - p(x)|^2, |E \hat{f}(x) - f(x)|^2$, one knows by (2.2) that

$$\begin{aligned} E \hat{f}(x) &= (2\pi)^{-d} \int_{R^d} e^{-itx} K^{ft}(ht) E(e^{it\omega}) / G_\alpha(t) dt = (2\pi)^{-d} \int_{R^d} e^{-itx} K^{ft}(ht) h_\omega^{ft} / G_\alpha(t) dt \\ &= (2\pi)^{-d} \int_{R^d} e^{-itx} K^{ft}(ht) f^{ft}(t) dt \end{aligned}$$

Due to the inverse Fourier transform theorem, $E \hat{f}(x) = f_X * K_h(x)$. Similar to $E \hat{f}(x), E \hat{p}(x) = p * K_h(x)$. According to [4], $E|\hat{p}(x) - p(x)|^2 \lesssim h^{2s}$, $E|\hat{f}(x) - f(x)|^2 \lesssim h^{2s}$. Therefore, recombining the conclusions of (9) and (10).

$$\text{Var} \hat{f}(x) + |E \hat{f}(x) - f(x)|^2 + \text{Var} \hat{p}(x) + |E \hat{p}(x) - p(x)|^2 \lesssim n^{-1} h^{-(2\beta(\alpha)+d)} + h^{2s}$$

thanks to $|G_\alpha(t)| \gtrsim (1 + t^2)^{-\frac{\beta(\alpha)}{2}}$. Then it follows from the choices of h in Theorem 1 that

$$\text{Var} \hat{f}(x) + |E \hat{f}(x) - f(x)|^2 + \text{Var} \hat{p}(x) + |E \hat{p}(x) - p(x)|^2 \lesssim \varepsilon_n$$

where $\varepsilon_n = n^{-\frac{1}{2s+2\beta(\alpha)+d}}$ for the case.

Furthermore, $|E\hat{f}(x) - f(x)|^2 + |E\hat{p}(x) - p(x)|^2 \lesssim \varepsilon_n$. This with (2.4) leads to

$$\lim_{c \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \sup P[|\hat{m}(x) - m(x)|^2 \geq cn^{-\frac{2s}{2s+2\beta(\alpha)+d}}] \lesssim (c\varepsilon_n)^{-1} \varepsilon_n = c^{-1}$$

Finally, the desired conclusion is proved.

4. 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. Based on the generalized additive noise model, we build a regression estimator and study its convergence rate. It's the convergence rate of the estimator with generalized additive noise is $n^{-1}h^{-(2\beta(\alpha)+d)} + h^{2s}$. Clearly, when α represent different value, Theorem 1 can be reduced to Guo [4] and Politis [15] respectively.

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] Caron, A., Baio, G., Manolopoulou, I.: Estimating individual treatment effects using non-parametric regression models: A review. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY. 185, 1115–1149 (2022) <https://doi.org/10.1111/rssa.12824>
- [2] Almanjahie, I.M., Fetitah, O., Attouch, M.K., Louhab, H.: Asymptotic normality of the robust equivariant estimator for functional nonparametric models. MATHEMATICAL PROBLEMS IN ENGINEERING. 2022 (2022). <https://doi.org/10.1155/2022/8989037>
- [3] Alexander, M.: Deconvolution Problems in Nonparametric Statistics. Springer, Berlin (2009). <https://doi.org/10.1007/978-3-540-87557-4>
- [4] 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>
- [5] 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>
- [6] Pinelis, I.: Monotonicity preservation properties of kernel regression estimators. STATISTICS & PROBABILITY LETTERS. 177 (2021) <https://doi.org/10.1016/j.spl.2021.109157>
- [7] 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>
- [8] Chu, B.M., Jacho-Chavez, D.T., Linton, O.B.: Standard errors for nonparametric regression. ECONOMETRIC REVIEWS. 39, 674–690 (2020). <https://doi.org/10.1080/07474938.2020.1772563>
- [9] Cleanthous, G., Georgiadis, A.G., Kerkycharian, G., Petrushev, P., Picard, D.: Kernel and wavelet density estimators on manifolds and more general metric spaces. BERNOULLI. 26, 1832–1862 (2020) <https://doi.org/10.3150/19-BEJ1171>
- [10] Chesneau, C., El Kolei, S., Kou, J., Navarro, F.: Nonparametric estimation in a regression model with additive and multiplicative noise. JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS. 380 (2020).

<https://doi.org/10.1016/j.cam.2020.112971>

- [11] Chen, J., Kou, J.: Nonparametric pointwise estimation for a regression model with multiplicative noise. JOURNAL OF FUNCTION SPACES. 2021 (2021). <https://doi.org/10.1155/2021/1599286>
- [12] Delaigle, A., Hall, P.: Estimation of observation-error variance in errors-in-variables regression. STATISTICA SINICA. 21, 1023–1063 (2011). <https://doi.org/10.5705/ss.2009.039>
- [13] Kato, K., Sasaki, Y.: Uniform confidence bands for nonparametric errors-in-variables regression. JOURNAL OF ECONOMETRICS. 213, 516–555 (2019) <https://doi.org/10.1016/j.jeconom.2019.05.021>
- [14] Delaigle, A.: Nonparametric kernel methods with errors-in-variables: Constructing estimators, computing them, and avoiding common mistakes. AUSTRALIAN & NEW ZEALAND JOURNAL OF STATISTICS. 56, 105–124 (2014) <https://doi.org/10.1111/anzs.12066>
- [15] Politis, D.N., Vasiliev, V.A.: Non-parametric sequential estimation of a regression function based on dependent observations. SEQUENTIAL ANALYSIS-DESIGN METHODS AND APPLICATIONS. 32, 243–266 (2013) <https://doi.org/10.1080/07474946.2013.803398>