第51卷第5期 中国科学技术大学学报 Vol. 51, No. 5
2021年5月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA May 2021
Received: 2021-03-17; Revised: 2021-05-20 doi:10.52396/JUST-2021-0078

Inference of online updating approach to nonparametric smoothing of big data

CHEN Ziyu[1], GUO Xiao[2*]
1. Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China;
2. International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230601, China
* Corresponding author. E-mail: xiaoguo@ustc.edu.cn

Abstract: The online updating method (ONLINE) is an efficient analysis approach applied to big data. We prove the asymptotic properties and conduct statistical inference of the ONLINE models in kernel density and kernel regression. Several algorithms are proposed to solve the problems of the bandwidth selection in kernel density and regression respectively. We verify the asymptotic normality of the ONLINE density model in simulation and apply the ONLINE linear kernel regression to the Volatility Index (VIX) prediction. The empirical results show that the ONLINE linear kernel regression model achieves a comparable performance in continuously arriving option data streams prediction with significantly lower complexity than the classical local linear regression model.

Keywords: bandwidth; kernel estimator; online updating approach; statistical inference; VIX prediction
CLC number: O212.7 **Document code**: A
2020 Mathematics Subject Classification: 62G20

1 Introduction

Large organizations, such as financial companies and Internet giants, or other large enterprises, today receive data at a tremendous scale and speed due to the explosive growth of data sources that continuously generate streams of data. Whether option data comes from financial exchanges that generate millions of transactions in a short period, or log data in Internet companies application servers, or streaming data from Web sites and mobile phone clicks, all of these data streams require effective and efficient processing and analysis, which has become challenging problems in statistics and computer science. Several researchers have systematically discussed this issue related to the mining aspects of data streams[1-3]. Besides, nonparametric estimation methods have also been studied in data streams[4,5].

Considering the high diversity of streaming data, kernel density estimation and regression, as representatives of nonparametric methods, are very suitable to deal with estimation and prediction problems since they need not make excessive assumptions about the corresponding functions between the data[6-8]. On the other hand, rapid real-time inference of convective data in a few seconds is also expected for the need of practice while classical nonparametric methods have the weakness of computational heavily. Therefore, our focus is not only to perform statistical analysis with as little calculation and storage as possible but also to command more relaxed restrictions on the distribution of the data when it arrives in streams or blocks. Since requiring less storage and updating in real-time, the sequential updating approach is instrumental in the high-speed processing of massive data[9]. Much work has been done in this area[10-12]. Kong and Xia[12] go on to research this method and proved that the online updating method (ONLINE) models with index-special bandwidth are the optimal choice of minimizing the asymptotic mean square error among a very general class of online updating schemes of kernel density and nonparametric regression models.

Compared with the classical kernel estimation methods noted as OFFLINE, the ONLINE models are shown to use the obtained data block to estimate and update the bandwidth parameters, avoiding the complexity of re-estimation every time the data stream arrives. Although the ONLINE approach has sensational theoretical properties, further steps must be taken before attempting an application of the ONLINE method. Taking the density estimator as an example, when the information of the true density is unknown, how to

Citation: CHEN Ziyu, GUO Xiao. Inference of online updating approach to nonparametric smoothing of big data. J. Univ. Sci. Tech. China, 2021, 51(5): 390-403.

obtain the ONLINE estimators from the observed data is not studied[12], let alone the following inferences conducted.

In this paper, we establish the statistical inference for a wide variety of kernel-based nonparametric estimators coupled with index-special bandwidths based on the online updating approach and propose the available and practical algorithms for the applications of the models. More specifically, we prove the asymptotic normality with index-special bandwidths of the ONLINE estimators. Further, when the bandwidths are estimated reasonably from the data, the asymptotic normality of ONLINE estimators still holds. As follows, we also discuss their hypothesis tests and power functions. We design several algorithms of bandwidth selection for the ONLINE models, which are of great significance to the promotion and application of the model. The effectiveness of the ONLINE model is verified by numerical study and the ONLINE local linear model is used to predict the time series data of financial options.

This paper is organized as follows. In Section 2, we establish the asymptotic properties of the estimator in the ONLINE kernel density model. Then we establish a hypothesis test and discuss the power function. In Section 3, we discuss the ONLINE nonparametric regression models, including local constant and local linear model. In Section 4, the algorithms are constructed to solve the problems of the bandwidth selection in the ONLINE kernel density and regression model. Section 5 contains simulation studies to test the asymptotic normality of the ONLINE density estimator and we predict the Volatility Index by the ONLINE local linear regression models.

2 Density estimation and testing

The classical nonparametric density model and the sequential updating nonparametric density model are introduced in Section 2.1 and Section 2.2, respectively. And we establish the corresponding asymptotic properties in Section 2.3 and conduct the hypothesis test in Section 2.4.

2.1 Nonparametric density model

Throughout this paper, $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$ denote p-variate independent identically distributed random samples of density f. And we have $\boldsymbol{X}_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}$. The Parzen-Rosenblatt estimator of $f(\cdot)$ is defined as

$$\widehat{f}_n(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^{n}K_{h_n}(\boldsymbol{X}_i-\boldsymbol{x}) \tag{1}$$

where $K_{h_n}(\boldsymbol{u})=K(\boldsymbol{u}/h_n)/h_n^p$, and $K(\cdot)$ is a kernel density function. Intuitively, $\widehat{f}_n(\cdot)$ is the average of a set of weights. If the observation $\boldsymbol{X}_i$ is closer to $\boldsymbol{x}$ than other observations, then its weight is relatively larger. Conversely, if $\boldsymbol{X}_i$ is far away from $\boldsymbol{x}$, then the weight is small. The bandwidth h_n controls the degree of "closeness", so the choice of h_n is incredibly crucial. But for classical bandwidth selection methods, h_n is usually chosen based on the total of observations $(\boldsymbol{X}_i)_{i=1}^{n}$. Therefore, the bandwidth needs to be re-estimated when the observations arrive in batches.

2.2 The ONLINE nonparametric density model

To solve the problem, the ONLINE nonparametric density model introduced by Kong and Xia[12] is an improvement. Suppose $\widehat{f}_{n-1}$ is the current estimator after $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_{n-1}$ have been observed. Once $\boldsymbol{X}_n$ arrives, the estimator is then updated as

$$\widehat{f}_n(\boldsymbol{x})=(1-\beta_n)\widehat{f}_{n-1}(\boldsymbol{x})+\beta_n K_{h_n}(\boldsymbol{X}_n-\boldsymbol{x}),$$

where β_n is a weight coefficient, and it is worth noting that the choice of the bandwidth h_n is independent of all preceding observations. they have proved that $\widehat{f}_n(\boldsymbol{x})$ with $\beta=n^{-1}$ has the smallest asymptotic mean square error amongst a general class of weighting series. The ONLINE density model is

$$\tilde{f}_n(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x}) \tag{2}$$

The bandwidth $\tilde{h}_i=ci^{-\alpha}$, $i=1,2,\cdots,n$. By minimizing the mean square error of the ONLINE estimator, the parameters c and α have the following form:

$$\left.\begin{aligned}\alpha&=\frac{1}{4+p},\\ c&=\left(\frac{p(p+2)}{2(p+4)}\right)^{1/(4+p)}\left(\frac{R_2(K)f(\boldsymbol{x})}{[\operatorname{tr}\{\mathscr{H}_f(\boldsymbol{x})\}]^2}\right)^{1/(4+p)}\end{aligned}\right\} \tag{3}$$

here $\operatorname{tr}\{\mathscr{H}_f(\cdot)\}$ denotes the trace of Hessian matrix of $f(\cdot)$, and $\operatorname{tr}\{\mathscr{H}_f(\cdot)\}=\sum_{i=1}^{p}\partial^2 f/\partial x_i^2$, and $R_2(K)=\int K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$. As it can be seen in Equation (3), the estimator of c is not directly available. In Section 4, several algorithms are constructed to solve the difficulties of the estimation of c.

Compared with the classical model, the significant difference of Equation (2) is the form of index-special bandwidth. As we can see, in the ONLINE model, the bandwidth $\tilde{h}_i=ci^{-\alpha}$ is split into two parts: the constants c and α, and the variable $i^{-\alpha}$, where the constant c is related to $f(\cdot)$ and α depends on the dimension p of observations, and the variable term is only affected by the sample size n. The advantage of bandwidth splitting in this way is that if c can be accurately estimated from the initial data streams, it can be used for the subsequent data streams continuously, without the re-calculation of bandwidth, which is of great significance to reduce the amount of calculation of big data. In the next, we will study the asymptotic properties of $\tilde{f}_n(\boldsymbol{x})$

and establish the hypothesis test.

2.3 Asymptotic properties of the ONLINE density estimator

To discuss the asymptotic properties of the ONLINE density estimator, we make the following assumptions of kernel function and density function, respectively.

A1 The kernel function $K(\cdot)$ is symmetric, and we have $\int K(\boldsymbol{u})\mathrm{d}\boldsymbol{u}=1$, also we have $\int \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}K(\boldsymbol{u})\mathrm{d}\boldsymbol{u}=\boldsymbol{I}_p$ and $\int K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}=R_2(K)<\infty$.

A2 The density function $f(\cdot)$ is third order continuously differentiable at the interior point x of its support.

Under the above assumptions, we establish the asymptotic normality of $\tilde{f}_n(\boldsymbol{x})$ as follows.

Theorem 2.1 Assume Assumptions A1–A2 hold. Suppose that $\int |k(\boldsymbol{u})|^{2+\delta}\mathrm{d}\boldsymbol{u}<\infty$ for some $\delta>0$, then

$$\sqrt{\frac{(1+p\alpha)c^p}{f(\boldsymbol{x})R_2(K)n^{p\alpha-1}}}\left(\tilde{f}_n(\boldsymbol{x})-f(\boldsymbol{x})-\frac{c^2}{2(1-2\alpha)}[\mathrm{tr}\{\mathscr{H}_f(\boldsymbol{x})\}]n^{-2\alpha}\right)\xrightarrow{\mathrm{d}}N(0,1).$$

The proofs of all theorems and corollaries in this paper are presented in the Appendix.

The gradient vector and Hessian matrix of $f(\cdot)$ denote by $\nabla_f(\cdot)$ and $\mathscr{H}_f(\cdot)$, respectively. Given that Theorem 2.1 is not sufficient for the following hypothesis test, we consider proving the asymptotic normality of the estimated parameters.

Corollary 2.1 Assume Assumptions A1 – A2 hold. Suppose that $\int |k(\boldsymbol{u})|^{2+\delta}\mathrm{d}\boldsymbol{u}<\infty$ for some $\delta>0$, and $\mathrm{tr}\{\mathscr{H}_{\tilde{f}}(\boldsymbol{x})\}\xrightarrow{\mathrm{d}}\mathrm{tr}\{\mathscr{H}_f(\boldsymbol{x})\}$, then

$$\sqrt{\frac{(1+p\alpha)\tilde{c}^p}{\tilde{f}_n(\boldsymbol{x})R_2(K)n^{p\alpha-1}}}\left(\tilde{f}_n(\boldsymbol{x})-f(\boldsymbol{x})-\frac{\tilde{c}^2}{2(1-2\alpha)}[\mathrm{tr}\{\mathscr{H}_{\tilde{f}_n}(\boldsymbol{x})\}]n^{-2\alpha}\right)\xrightarrow{\mathrm{d}}N(0,1),$$

where

$$\tilde{c}=\{p(p+2)R_2(K)\tilde{f}_n(\boldsymbol{x})\}^{1/(4+p)}\cdot\{2(p+4)\mathrm{tr}^2\{\mathscr{H}_{\tilde{f}_n}(\boldsymbol{x})\}\}^{-1/(4+p)},$$

and $\mathrm{tr}\{\mathscr{H}_{\tilde{f}_n}(\boldsymbol{x})\}=\sum_{i=1}^{p}(\partial^2\tilde{f}_n/\partial x_i^2)|x_i$.

2.4 Hypothesis test of the ONLINE density estimation

Given the arrived observations following the distribution f, the estimator of $f(\boldsymbol{x}_0)$, which means the density of the fixed p dimensional point $\boldsymbol{x}_0$, can be easily computed. In this section, we aim to develop the hypothesis test of $f(\boldsymbol{x}_0)$ according to Corollary 2.1. Specifically, the hypothesis test for the true density of the given point $\boldsymbol{x}_0$ is as follows:

$$\mathrm{H}_0: f(\boldsymbol{x}_0)=\theta_0 \text{ vs } \mathrm{H}_1: f(\boldsymbol{x}_0)\neq\theta_0.$$

The estimator $\tilde{f}_n(\boldsymbol{x}_0)$, and its bias $\tilde{b}_n(\boldsymbol{x}_0)$ and variance $\tilde{\sigma}_n^2(\boldsymbol{x}_0)$ of $\boldsymbol{x}_0$ can be calculated according to the previous sections. Here

$$\tilde{b}_n(\boldsymbol{x}_0)=[2(1-2\alpha)]^{-1}[\tilde{c}^2\,\mathrm{tr}\{\mathscr{H}_{\tilde{f}_n}(\boldsymbol{x}_0)\}]n^{-2\alpha}$$

and

$$\tilde{\sigma}_n{}^2(\boldsymbol{x}_0)=[\tilde{f}_n(\boldsymbol{x}_0)R_2(K)][(1+p\alpha)\tilde{c}^p]^{-1}n^{p\alpha-1}.$$

According to Corollary 2.1, we have

$$\frac{\tilde{f}_n(\boldsymbol{x}_0)-f(\boldsymbol{x}_0)-\tilde{b}_n(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\xrightarrow{\mathrm{d}}N(0,1),$$

then the test statistic can be defined as

$$z_n=\frac{\tilde{f}_n(\boldsymbol{x}_0)-\theta_0-\tilde{b}_n(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\tag{4}$$

which follows the asymptotic standard normal distribution under H_0 is true. We reject H_0 if $|z_n|\geqslant u_{\alpha_0/2}$, where $u_{\alpha_0/2}$ is the upper $\alpha_0/2$ percentage point of the standard normal distribution and α_0 is the desired significance level of the test. Then we discuss the power function of the test. The power function is

$$\begin{aligned}\beta_n(\boldsymbol{x}_0)=&P\left(\left|\frac{\tilde{f}_n(\boldsymbol{x}_0)-\theta_0-\tilde{b}_n(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\right|>u_{\frac{\alpha_0}{2}}\right)=\\&1-P\left(-u_{\frac{\alpha_0}{2}}+\frac{\theta_0-f(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\leqslant\frac{\tilde{f}_n(\boldsymbol{x}_0)-(f(\boldsymbol{x}_0)+\tilde{b}_n(\boldsymbol{x}_0))}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\leqslant u_{\frac{\alpha_0}{2}}+\frac{\theta_0-f(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\right)\approx\\&2-\Phi\left(u_{\frac{\alpha_0}{2}}+\frac{f(\boldsymbol{x}_0)-\theta_0}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\right)-\Phi\left(u_{\frac{\alpha_0}{2}}+\frac{\theta_0-f(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)}\right).\end{aligned}$$

If $f(\boldsymbol{x}_0)$ converges to θ_0 fast enough such that $(f(\boldsymbol{x}_0)-\theta_0)/\tilde{\sigma}_n(\boldsymbol{x}_0)\to 0$, then the power $\beta_n(\boldsymbol{x}_0)$ converges to α_0 and the probability of Type Ⅱ error converges to $1-\alpha_0$, and the test is not able to discriminate the alternative from the null hypothesis. If $f(\boldsymbol{x}_0)$ converges to θ_0 at a slow rate such that $(f(\boldsymbol{x}_0)-\theta_0)/\tilde{\sigma}_n(\boldsymbol{x}_0)\to\infty$, then the power converges to 1 and the probability of Type Ⅱ error converges to 0, so the alternative is too easy. Intermediate rates seem to be the most interesting. Considering that $f(\boldsymbol{x}_0)=\theta_0+\delta_n$, δ_n converges to 0 as $n\to\infty$, the problem we consider is how fast δ_n converges to 0 can ensure the value of power function is between α_0 and 1. So δ_n converges to 0 at the same rate with $\tilde{\sigma}_n(\boldsymbol{x}_0)$, which is $O(n^{\frac{p\alpha-1}{2}})$. The curve of the power function is shown in Figure 1.

3 Nonparametric regression and testing

Regression analysis in nonparametric cases is suggested to be a widely used tool in applied data analysis. Therefore, in this section, we will discuss the ONLINE

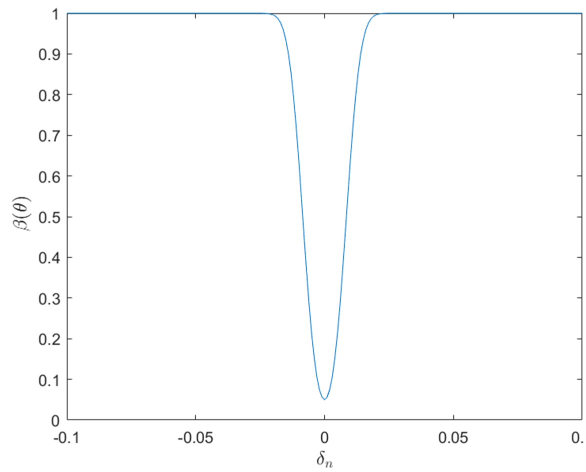


Figure 1. Power function.

model in nonparametric regression. Suppose we have observations $(\boldsymbol{X}_i, Y_i)_{i=1}^n$ which are generated according to

$$Y_i = m(\boldsymbol{X}_i) + \varepsilon_i,\ i = 1,\cdots,n \tag{5}$$

where $\boldsymbol{X}_i$ is a $p\times 1$ random vector with probability density function $f(\cdot)$, and $m(\cdot)$ is a smooth function. ε_i is an error item with $E(\varepsilon_i \mid \boldsymbol{X}_i) = 0$, $\mathrm{Var}(\varepsilon_i \mid \boldsymbol{X}_i) = \sigma_\varepsilon^2$. It is of interest to estimate $m(\cdot)$ based on $(\boldsymbol{X}_i, Y_i)_{i=1}^n$, after which we will conduct statistical inference.

3.1 ONLINE local constant estimator and its asymptotic properties

The ONLINE local constant estimator of $m(x)$ is addressed[12] by minimizing

$$\sum_{i=1}^n K_{\bar{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\{Y_i - m(\boldsymbol{x})\}^2.$$

Its solution is

$$\tilde{m}_{nw}(\boldsymbol{x}\mid\alpha,c) = \frac{\sum_{i=1}^n K_{\bar{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})Y_i}{\sum_{i=1}^n K_{\bar{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})} \tag{6}$$

again with index-specific bandwidths $h_i = ci^{-\alpha}$, $i = 1,\cdots,n$, for some constants $c > 0$, $\alpha > 0$. Similar with kernel estimator, by minimizing the mean square error of $\tilde{m}_{nw}(\boldsymbol{x}\mid\alpha,c)$, the constant c and α have the following form:

$$\alpha = \frac{1}{4+p},\ c = \left(\frac{p(p+2)}{2(p+4)}\right)^{1/(4+p)}\cdot \left(\frac{R_2(K)f(\boldsymbol{x})\sigma_\varepsilon^2}{[\mathrm{tr}\{\mathcal{H}_m(\boldsymbol{x})\}f(\boldsymbol{x}) + 2\nabla^{\mathrm{T}} m(\boldsymbol{x})\nabla f(\boldsymbol{x})]^2}\right)^{1/(4+p)},$$

where $\nabla m(\cdot)$ and $\mathcal{H}_m(\cdot)$ denote the gradient and Hession matrix of $m(\cdot)$. Under the same assumptions and denotations claimed in the previous section, the asymptotic distribution of $\tilde{m}_{nw}(\boldsymbol{x}\mid\alpha,c)$ can be stated by the following theorem.

Theorem 3.1 Under Assumptions A1−A2, we have

$$\sqrt{\frac{(1+p\alpha)c^p f(\boldsymbol{x})}{n^{p\alpha-1}R_2(K)\sigma_\varepsilon^2}}\{\tilde{m}_{nw}(\boldsymbol{x}) - m(\boldsymbol{x}) - B_{nw}(\boldsymbol{x})\}\xrightarrow{\mathrm{d}} N(0,1),$$

where

$$B_{nw}(\boldsymbol{x}) = c^2\{2(1-2\alpha)\}^{-1}\cdot [\mathrm{tr}\{\mathcal{H}_m(\boldsymbol{x})\} + 2\nabla^{\mathrm{T}} m(\boldsymbol{x})\nabla f(\boldsymbol{x})/f(\boldsymbol{x})]n^{-2\alpha}.$$

Similar to Corollary 2.1, the following corollary can be addressed.

Corollary 3.1 Under Assumptions A1−A2, we have

$$\sqrt{\frac{(1+p\alpha)\tilde{c}^p\tilde{f}_n(\boldsymbol{x})}{n^{p\alpha-1}R_2(K)\sigma_\varepsilon^2}}\{\tilde{m}_{nw}(\boldsymbol{x}) - m(\boldsymbol{x}) - \tilde{B}_{nw}(\boldsymbol{x})\}\xrightarrow{\mathrm{d}} N(0,1),$$

where

$$\tilde{B}_{nw}(\boldsymbol{x}) = \tilde{c}^2\{2(1-2\alpha)\}^{-1}\cdot [\mathrm{tr}\{\tilde{\mathcal{H}}_m(\boldsymbol{x})\} + 2\nabla^{\mathrm{T}}\tilde{m}(\boldsymbol{x})\nabla\tilde{f}_n(\boldsymbol{x})/\tilde{f}_n(\boldsymbol{x})]n^{-2\alpha}.$$

Here the notation $\tilde{\mathcal{H}}_m(x)$, $\nabla\tilde{m}(x)$ are the estimator of $\mathcal{H}_m(\boldsymbol{x})$ and $\nabla m(\boldsymbol{x})$ which satisfied $\tilde{\mathcal{H}}_m(x)\xrightarrow{\mathrm{d}}\mathcal{H}_m(x)$, and $\nabla\tilde{m}(x)\xrightarrow{\mathrm{d}}\nabla m(\boldsymbol{x})$. It's worth noting that $\mathcal{H}_{\tilde{m}}(x)$, the second derivative of the estimator of $m(\boldsymbol{x})$, is extremely complicated, so here we use $\tilde{\mathcal{H}}_m(x)$, the estimation of second derivative of $m(\boldsymbol{x})$ instead, which can be obtained by Taylor expansion.

3.2 ONLINE local linear estimator and its asymptotic properties

The ONLINE local linear estimator of $m(x)$ is addressed[12] by minimizing

$$\sum_{i=1}^n K_{\bar{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\{Y_i - b^{\mathrm{T}}\boldsymbol{X}_{\mathrm{in}}(\boldsymbol{x})\}^2 \tag{7}$$

with respect to $\boldsymbol{b} = [m(\boldsymbol{x}), \bar{h}_i\nabla^{\mathrm{T}} m(\boldsymbol{x})]^{\mathrm{T}}$, and $\boldsymbol{X}_{\mathrm{in}}(\boldsymbol{x}) = [1, (\boldsymbol{X}_i - \boldsymbol{x})^{\mathrm{T}}/\bar{h}_i]^{\mathrm{T}}$. Denote $\tilde{\boldsymbol{b}}$ as an estimator of $\boldsymbol{b}$. Denote $M(\boldsymbol{x}) = [m(\boldsymbol{x}), \nabla^{\mathrm{T}} m(\boldsymbol{x})]^{\mathrm{T}}$, and $\tilde{M}(\boldsymbol{x}) = [\tilde{m}(\boldsymbol{x}), \nabla^{\mathrm{T}}\tilde{m}(\boldsymbol{x})]^{\mathrm{T}}$. So we have

$$\tilde{\boldsymbol{b}} = [\tilde{S}_1(\boldsymbol{x})]^{-1}\tilde{S}_2(\boldsymbol{x}, Y) \tag{8}$$

where

$$\left.\begin{aligned}\tilde{S}_1(\boldsymbol{x}) &= \frac{1}{n}\sum_{i=1}^n K_{\bar{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\tilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})\tilde{\boldsymbol{X}}_{\mathrm{in}}{}^{\mathrm{T}}(\boldsymbol{x}),\\ \tilde{S}_2(\boldsymbol{x}, Y) &= \frac{1}{n}\sum_{i=1}^n K_{\bar{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\tilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})Y_i\end{aligned}\right\} \tag{9}$$

The asymptotic normality of $\tilde{M}(\boldsymbol{x})$ is established by the following theorem and corollary.

Theorem 3.2 Under Assumptions A1−A2, we have

$$n^{(1-p\alpha)/2}(\tilde{M}(\boldsymbol{x}) - M(\boldsymbol{x}) - B_{ll}(\boldsymbol{x}))\xrightarrow{\mathrm{d}} N(0,\Sigma),$$

where

$$B_{ll}(\boldsymbol{x}) = \frac{c^2 n^{-2\alpha}f(\boldsymbol{x})}{2(1-2\alpha)}\begin{pmatrix}\mathrm{tr}\{\mathcal{H}_m(\boldsymbol{x})\}\\ \boldsymbol{0}_{p\times 1}\end{pmatrix},$$

$$\Sigma = \frac{\sigma_\varepsilon^2}{c^p(1+p\alpha)f(\boldsymbol{x})}\begin{pmatrix} R_2(K) & \boldsymbol{0}_{1\times p} \\ \boldsymbol{0}_{p\times 1} & \tilde{R}_2(K)\end{pmatrix},$$

here $\tilde{R}_2(K) = \int \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$.

Corollary 3.2. Under Assumptions A1－A2, we have

$$[n^{1-p\alpha}\tilde{f}_n(\boldsymbol{x})]^{1/2}(\tilde{M}(\boldsymbol{x}) - M(\boldsymbol{x}) - \tilde{B}_{ll}(\boldsymbol{x})) \xrightarrow{\mathrm{d}} N(0,\Sigma),$$

where

$$\tilde{B}_{ll}(\boldsymbol{x}) = \frac{\tilde{c}^2 n^{-2\alpha}\tilde{f}_n(\boldsymbol{x})}{2(1-2\alpha)}\begin{pmatrix}\mathrm{tr}\{\tilde{\mathscr{H}}_m(\boldsymbol{x})\} \\ \boldsymbol{0}_{p\times 1}\end{pmatrix},$$

$$\Sigma = \frac{\sigma_\varepsilon^2}{c^p(1+p\alpha)}\begin{pmatrix} R_2(K) & \boldsymbol{0}_{1\times p} \\ \boldsymbol{0}_{p\times 1} & \tilde{R}_2(K)\end{pmatrix}.$$

Here the notation $\tilde{\mathscr{H}}_m(x)$, $\tilde{\nabla m}(x)$ are the estimator of $\mathscr{H}_m(x)$ and $\nabla m(\boldsymbol{x})$ which satisfied $\tilde{\mathscr{H}}_m(x) \xrightarrow{\mathrm{p}} \mathscr{H}_m(x)$, $\tilde{\nabla m}(x) \xrightarrow{\mathrm{p}} \nabla m(\boldsymbol{x})$.

3.3 Hypothesis test about the ONLINE constant estimator

When the observations $(\boldsymbol{X}_1, Y_1), \cdots, (\boldsymbol{X}_n, Y_n)$ arrive, the estimator of $m(\cdot)$ at a fixed p dimensional point $\boldsymbol{x}_0$ can be easily calculated. As the following, in this section, we aim to develop the hypothesis test of the $m(\boldsymbol{x}_0)$ according to Corollary 3.1.

$$\mathrm{H}_0: m(\boldsymbol{x}_0) = \theta_0 \text{ vs } \mathrm{H}_1: m(\boldsymbol{x}_0) \neq \theta_0.$$

The estimator $\tilde{m}_n(\boldsymbol{x}_0)$, and its bias $\tilde{b}_n(\boldsymbol{x}_0)$ and variance $\tilde{\sigma}_n^2(\boldsymbol{x}_0)$ can be derived according to the previous section. Here

$$\tilde{b}_{nw}(\boldsymbol{x}_0) = \tilde{c}^2\{2(1-2\alpha)\}^{-1}[\mathrm{tr}\{\tilde{\mathscr{H}}_m(\boldsymbol{x}_0)\} + 2\nabla^{\mathrm{T}}\tilde{m}(\boldsymbol{x}_0)\tilde{\nabla f}_n(\boldsymbol{x}_0)/\tilde{f}_n(\boldsymbol{x}_0)]n^{-2\alpha},$$

and

$$\tilde{\sigma}_{nw}^2(\boldsymbol{x}_0) = [R_2(K)\sigma_\varepsilon^2][(1+p\alpha)\tilde{c}^p\tilde{f}_n(\boldsymbol{x}_0)]^{-1}n^{p\alpha-1}.$$

According to Corollary 3.1, we have

$$\frac{\tilde{m}_{nw}(\boldsymbol{x}_0) - m(\boldsymbol{x}_0) - \tilde{b}_{nw}(\boldsymbol{x}_0)}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)} \xrightarrow{\mathrm{d}} N(0,1),$$

and then the test statistic can be defined as

$$z_n = \frac{\tilde{m}_{nw}(\boldsymbol{x}_0) - \theta_0 - \tilde{b}_n(\boldsymbol{x}_0)}{\tilde{\sigma}_n(\boldsymbol{x}_0)} \tag{10}$$

which follows the asymptotic standard normal distribution under H_0 is true. We reject H_0 if $|z_n| \geqslant u_{\alpha_0/2}$, where $u_{\alpha_0/2}$ is the upper $\alpha_0/2$ percentage point of the standard normal distribution and α_0 is the desired significance level of the test. The power function of the test is discussed as follows:

$$\beta_n(\boldsymbol{x}_0) = P\left(\left|\frac{\tilde{m}_{nw}(\boldsymbol{x}_0) - \theta_0 - \tilde{b}_{nw}(\boldsymbol{x}_0)}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)}\right| > u_{\frac{\alpha_0}{2}}\right) =$$
$$1 - P\left(-u_{\frac{\alpha_0}{2}} + \frac{\theta_0 - m(\boldsymbol{x}_0)}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)} \leqslant \frac{\tilde{m}_{nw}(\boldsymbol{x}_0) - (m(\boldsymbol{x}_0) + \tilde{b}_{nw}(\boldsymbol{x}_0))}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)} \leqslant u_{\frac{\alpha_0}{2}} + \frac{\theta_0 - m(\boldsymbol{x}_0)}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)}\right) \approx$$
$$2 - \Phi\left(u_{\frac{\alpha_0}{2}} + \frac{m(\boldsymbol{x}_0) - \theta_0}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)}\right) - \Phi\left(u_{\frac{\alpha_0}{2}} + \frac{\theta_0 - m(\boldsymbol{x}_0)}{\tilde{\sigma}_{nw}(\boldsymbol{x}_0)}\right).$$

Considering that $m(\boldsymbol{x}_0) = \theta_0 + \delta_n$, δ_n converges to 0 as $n \to \infty$, the problem we consider is how fast δ_n converges to 0 can ensure the value of power function is between α_0 and 1. So δ_n converges to 0 at the same rate with $\tilde{\sigma}_{nw}(\boldsymbol{x}_0)$, which is $O(n^{\frac{p\alpha-1}{2}})$.

4 Algorithms

In this section, we will discuss computational details on the bandwidth selection of the ONLINE models. It is generally accepted that the performance of the kernel estimator is mainly determined by the bandwidth. However, the bandwidth in the online model is not easy to obtain directly, and the difficulties lie in the determination of parameters c.

4.1 Solution of *c* in the ONLINE kernel density estimator

In the classical kernel density estimation, the main methods to solve bandwidth selection are rule of thumb[13], cross-validation[14-15] and plug-in[16,17]. There are drawbacks to blending the same bandwidth into all components[18,19]. As we can see, the parameter c is related to $\tilde{f}_n(x)$. Therefore, the calculation of c can be converted into the problem of searching a fixed point. If we plug in the $\tilde{f}_n(x)$ based on the previous $\tilde{c}_0$ to get a new $\tilde{c}_1$, $\tilde{c}$ can be addressed after iterating such a process for several times. More precisely, we state the steps in the following.

(Ⅰ) Based on the previous observations $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ using the cross-validation method, we obtain the initial bandwidth marked as h_n, then its ONLINE kernel density estimator $\tilde{f}_n(\cdot)$ can be calculated according to Equation (2).

(Ⅱ) The estimation of the initial parameter $\tilde{c}_0$ as follows:

$$\tilde{c}_0 = \left(\frac{p(p+2)}{2(p+4)}\right)^{1/(4+p)}\left(\frac{R_2(K)\tilde{f}_n(\boldsymbol{x})^2}{\mathrm{tr}\{\mathscr{H}_{\tilde{f}_n}(\boldsymbol{x})\}^2}\right)^{1/(4+p)}.$$

(Ⅲ) While the initial estimated value $\tilde{c}_0$ of c is obtained, we can calculate the first bandwidth $\tilde{h}_1 = \tilde{c}_0 i^{-\alpha}$. Then $\tilde{f}_n(\cdot)$ and $\mathrm{tr}\{\mathscr{H}_{\tilde{f}_n}(\cdot)\}$ according to $\tilde{h}_1$ can be calculated accordingly. Repeat this process until $\tilde{c}$ converges.

4.2 Solution of *c* in the ONLINE kernel regression estimator

In the ONLINE kernel regression models, the parameter c can also be estimated through the same idea in Section 4.1 theoretically while the second derivative of $\tilde{m}(\cdot)$

explicit form is not easy to access. Besides, the parameter c is determined by $f(\cdot)$, $m(\cdot)$ and σ_ε, which are complicated or even impossible to be derived in the practical problem. Moreover, the parameter c in regression models does not get closer to the true value with each iteration. So inspired by cross-validation that does not rely much on the calculation formula, we use the data-driven approach to figure out c. Let $\widehat{m}_{-i}(\boldsymbol{X}_i)$ denote the leave-one-out local linear estimator of $m(\boldsymbol{x})$ at $\boldsymbol{x} = \boldsymbol{X}_i$ by using all observations but $(\boldsymbol{X}_i, Y_i)$. And $\widehat{b}_{-i}(\boldsymbol{X}_i) = (\widehat{m}_{-i}(\boldsymbol{X}_i), \nabla^{\mathrm{T}}\widehat{m}_{-i}(\boldsymbol{X}_i))^{\mathrm{T}}$. We have

$$\widehat{b}_{-i}(\boldsymbol{X}_i) = \begin{pmatrix} \widehat{m}_{-i}(\boldsymbol{X}_i) \\ \nabla^{\mathrm{T}}\widehat{m}_{-i}(\boldsymbol{X}_i) \end{pmatrix} = \left[\sum_{j\neq i} K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\begin{pmatrix} 1 \\ \boldsymbol{X}_j - \boldsymbol{X}_i \end{pmatrix}(1, (\boldsymbol{X}_j - \boldsymbol{X}_i)^{\mathrm{T}})\right]^{-1} \cdot \sum_{j\neq i} K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\begin{pmatrix} 1 \\ \boldsymbol{X}_j - \boldsymbol{X}_i \end{pmatrix} Y_i.$$

Define a $(q+1)\times 1$ vector e_1 whose first element is one with all remaining elements being zero. The leave-one-out kernel estimator of $m(\boldsymbol{X}_i)$ is given by $\widehat{m}_{-i}(\boldsymbol{X}_i) = e_1^{\mathrm{T}}\widehat{m}_{-i}(\boldsymbol{X}_i)$, and we choose c to minimize the least-squares cross-validation function given by

$$\mathrm{CV}(c) = \sum_{i=1}^{n}[Y_i - \widehat{m}_{-i}(\boldsymbol{X}_i)]^2.$$

5 Simulation study

In this section, we analyze the performance of the ONLINE models in the density estimator. The ONLINE local linear model is applied to predict the Volatility Index.

5.1 Density estimator normality test on artificial dataset

In this part, we investigate the normality of the online density estimator by the following simulation based on a one-dimensional bimodal normal distribution. More specifically, the distribution has the following form:

$$f(x) = \frac{0.7}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \frac{0.3}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right),$$

where $\mu_1 = -3$, $\mu_2 = 3$, $\sigma_1^2 = \sigma_2^2 = 1$. Based on the distribution, we ran 500 simulations, each producing 500 observations. Four specialized points $x = \{-3.5, -1.5, 1.5, 3.5\}$ are considered respectively and Gaussian kernel is used. The bandwidth $\tilde{h}_i$ is determined by c, which could be estimated by the algorithm in Section 4.1. For 500 resulting estimator values for each point, we computed the sample percentiles which range from 1% to 99% and presented them with corresponding theoretical percentiles of standard normal distribution in quantile-quantile (Q-Q) plots as shown in Figure 2. Figure 2 shows Q-Q plots at $x=\{-3.5, -1.5, 1.5, 3.5\}$ drawn from Corollary 2.1. The plots deviate only slightly from 45-degree lines, indicating that the quantile-quantile of those points are very close to that of standard normality distribution.

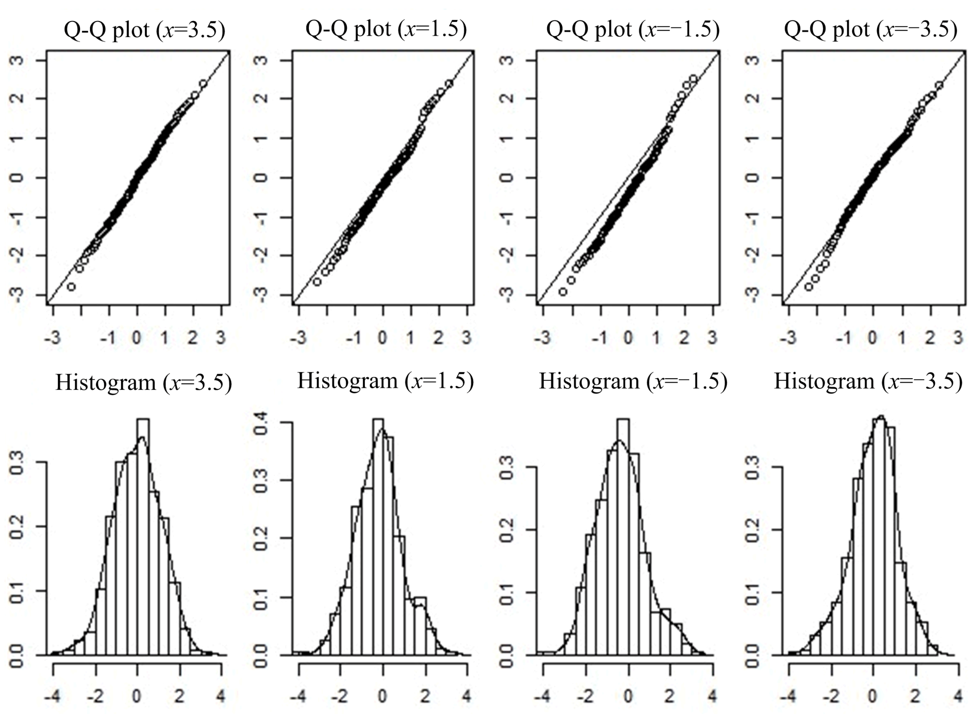


Figure 2. Quantile-Quantile plots and histograms at the points $x=\{-3.5, -1.5, 1.5, 3.5\}$ using the ONLINE model.

Table 1. Two types of errors of the ONLINE kernel density model.

x	$\alpha_0 = 0.01$		$\alpha_0 = 0.025$		$\alpha_0 = 0.05$	
	Type Ⅰ error	Type Ⅱ error	Type Ⅰ error	Type Ⅱ error	Type Ⅰ error	Type Ⅱ error
−3.5	0.012	0.988	0.024	0.972	0.038	0.966
−1.5	0.016	0.984	0.027	0.970	0.044	0.965
1.5	0.014	0.986	0.018	0.980	0.053	0.960
3.5	0.008	0.991	0.021	0.982	0.048	0.968

Besides, to show the performance of hypothesis testing, we calculate two types of errors at x according to Section 2.3. Repeated 500 times, the average probability of two types of errors can be obtained. It's worth noting that δ_n is the same order as $\widetilde{\sigma}_n(\cdot)$, so we have $\delta_n = 0.01$. Table 2 shows two types of errors at four points.

5.2 Volatility index prediction by ONLINE linear kernel regression model

In this section, we predict the Volatility Index (VIX) by using the ONLINE linear kernel regression model. The VIX is an index released by the Chicago Board of Options Exchange (CBOE) to reflect the expectation of future volatility of markets[20]. It provides market participants with an indicator that reflects the overall trend of the market, which is crucial in options trading[21]. The data used in this section consists of the CSI 300 stock index options on 2020-09-14, which is downloaded from Wind, and it is the snapshot of five-level market quotations of each option which are recorded per 500 milliseconds. As a result, a total number of 1085 observations are chosen. Considering that the 50 exchange-traded funds (50ETF) have been traded for a long time and the trading volume is extremely large, we also use the historical data of 50ETF options which are also included in the simulation. Note that the data is the difference value of two initial data whose time difference is the time interval of 10 ticks. Based on these values, the following model for the VIX difference sequence of CSI 300 stock index options is built,

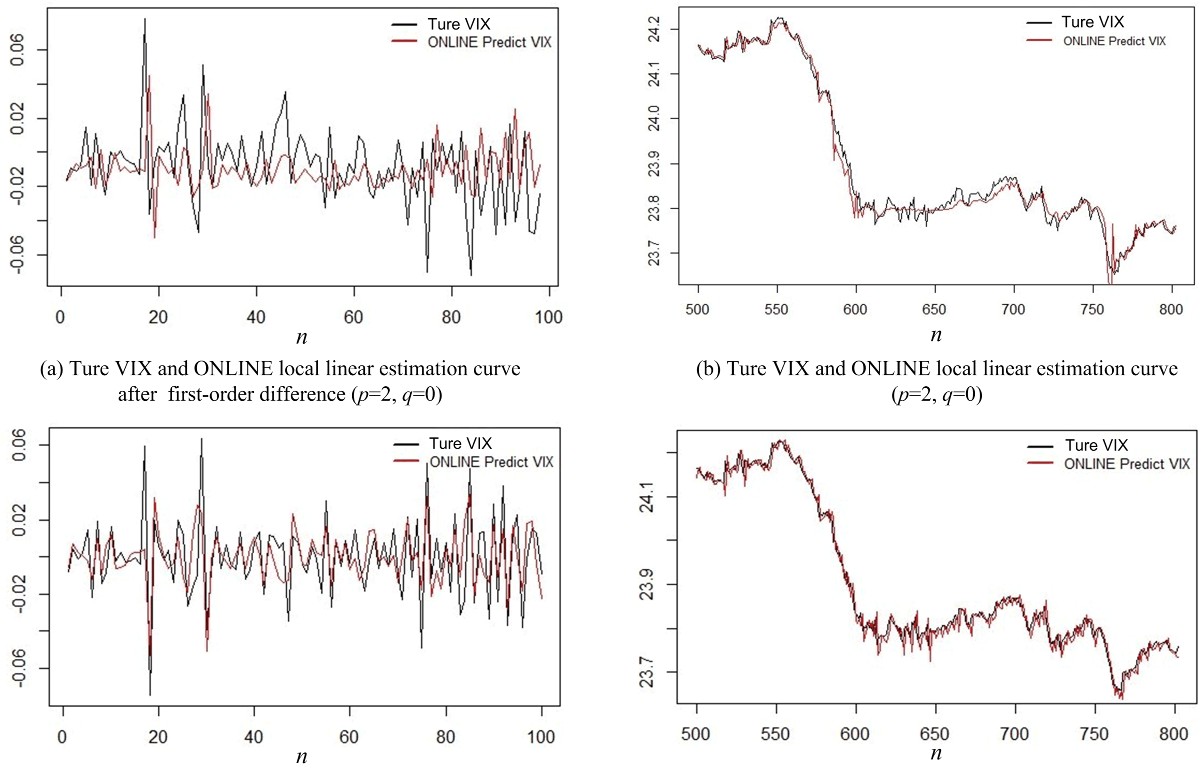
$$\Delta X_t = m(\Delta X_{t-1}, \Delta X_{t-2}, \cdots, \Delta X_{t-p}, \Delta X'_{t-1}, \Delta X'_{t-2}, \cdots, \Delta X'_{t-q}) + \varepsilon_t \quad (11)$$

Note that, X_{t-i} is the i-order lag term of CSI 300 stock index option, and X'_{t-j} is the j-order lag term of 50ETF option before the arrival moment. Here, the first 300 pieces of data are trained to select c value by the cross-validation method in Section 4.2, and the Gaussian kernel function is used. As a comparison, the bandwidth of the OFFLINE method is also selected by cross-validation.

Figure 3(a) and Figure 3(b) use a 2-dimensional regression model, $\Delta X_t = m(\Delta X_{t-1}, \Delta X_{t-2}) + \varepsilon_t$, and the ONLINE model is used to predict the left sample points with $c = 0.03$ which is selected from the first 300 sample points. Figure 3(a) is the real VIX curve and estimated curve of the ONLINE local linear model after the first difference. The first 100 points are intercepted for clearness of representation. The solid black line shows the curve of the ture VIX, while the solid red line is the ONLINE model's fitting curve. The volatility of the ture VIX is steep while the fitting curve of the ONLINE model is relatively smoother. As shown in Figure 3(b), in the intervals with obvious trends, such as [550,600], VIX has an obvious downward trend and the predicted value is relatively accurate. In the ranges where the VIX does not fluctuate sharply, such as [600, 700], the forecast curve is relatively smoother. Only using a 2-dimensional model, the overall estimated curve is too stable, so we consider adding another kind of 2-dimensional 50ETF option data. Figure 3(c) and Figure 3(d) use a 4-dimensional regression model, so we have $\Delta X_t = m(\Delta X_{t-1}, \Delta X_{t-2}, \Delta X'_{t-1}, \Delta X'_{t-2}) + \varepsilon_t$, and $c = 0.1$. Compared with the 2-dimensional model, the prediction curve is more accurate, and the fluctuation is more consistent with the ture VIX.

To further illustrate the result of the comparison, we compared the mean square error (MSE) and calculation consumption time of the ONLINE regression model with the OFFLINE counterpart. Here $p = 1, 2, 3, 5$, $n = 100, 300, 500$. As can be seen from Table 2, the mean square error of the ONLINE model is close to the OFFLINE. With the increase of sample size n and dimension p, the relative mean square error is also increasing, even approaching 1. However, Table 2 shows the ONLINE model achieves a comparable performance in continuously arriving option data streams prediction while it has significantly lower complexity than the traditional local linear predictive regression model.

In high-frequency trading, when the decision maker needs to consider whether to trade at some special price, it can be considered to conduct hypothesis tests for several points. Type Ⅰ error affects the transaction income, while Type Ⅱ error is related to risk. Therefore, the two types of errors are constructed together to provide contributions to transaction decisions.



(a) Ture VIX and ONLINE local linear estimation curve after first-order difference (p=2, q=0)

(b) Ture VIX and ONLINE local linear estimation curve (p=2, q=0)

(c) Ture VIX and ONLINE local linear estimation curve after first-order difference (p=2,q=2)

(d) Ture VIX and ONLINE local linear estimation curve (p=2, q=2)

Figure 3. Plots of ture VIX and ONLINE local linear estimation curve.

Table 2. MSE(10^{-4}) and computation time of ONLINE/OFFLINE regression models.

p	n	Time(ON)	MSE(ON)	Time(OFF)	MSE(OFF)	Relative time	Relative MSE
	100	0.92	0.60	90.37	0.50	98.23	0.83
1	300	1.12	0.82	338.91	0.71	302.60	0.87
	500	2.76	1.38	1421.46	1.24	515.02	0.90
	100	0.97	0.51	102.37	0.44	105.54	0.86
2	300	1.30	0.78	382.62	0.69	294.32	0.89
	500	3.02	1.46	1488.89	1.34	493.01	0.92
	100	1.56	0.51	160.49	0.45	102.88	0.89
3	300	1.72	0.81	524.67	0.72	305.04	0.93
	500	3.49	1.35	1674.85	1.28	479.90	0.95
	100	2.89	0.41	366.16	0.35	126.70	0.86
5	300	4.03	0.77	1385.03	0.69	343.68	0.90
	500	6.49	1.12	3809.76	1.05	587.02	0.94

6 Conclusions

In this paper, we prove the asymptotic properties of the ONLINE kernel density and ONLINE kernel regression, and establish hypothesis tests. Several algorithms are constructed to solve the problems of bandwidth selection in ONLINE kernel density and ONLINE nonparametric regression models. In addition, we verify the asymptotic normality of ONLINE density model in simulation, and apply this method to predict the Volatility Index (VIX) using the ONLINE linear kernel regression. It is of interest to conduct the inference of varying coefficient regression models in which the online method can also be applied. Besides, more datasets in different areas, such as finance, computer science, and engineering, can be considered to investigate the performance of the method in applications.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (12071452, 72091212, 11601500) and USTC Research Funds of the Double First-Class Initiative (YD2040002013).

Conflict of interest

The authors declare no conflict of interest.

Author information

CHEN Ziyu received the MS degree in statistics from the University of Science and Technology of China in 2021. Her research interests mainly focuses on nonparametric statistical inference.

GUO Xiao (corresponding author) is Associate Professor at the Department of Statistics and Finance, University of Science and Technology of China (USTC). He received the BS degree from Fudan University in 2009, and the PhD degree from University of Wisconsin-Madison in 2015. Afterwards, he joined the USTC. His research interests include high-dimensional statistical estimation and inference, robust estimation and large covariance matrix estimation.

References

[1] Aggarwal C C. Data Streams: Models and Algorithms. Heidelberg: Springer Science & Business Media, 2007.

[2] Aggarwal C C, Yu P S. A Survey of Synopsis Construction in Data Streams. Heidelberg: Springer Science & Business Media, 2007.

[3] Wrench C, Stahl F, Fatta G, et al. Data stream mining of event and complex event streams: A survey of existing and future technologies and applications in big data. In: Enterprise Big Data Engineering, Analytics, and Management. Hershey, PA: IGI Global, 2016.

[4] Cao Y, He H, Man H. SOMKE: Kernel density estimation over data streams by sequences of self-organizing maps. IEEE Transactions on Neural Networks Learning Systems, 2012, 23(8): 1254−1268.

[5] Min X, Ishibuchi H, Xin G, et al. Dm-KDE: Dynamical kernel density estimation by sequences of KDE estimators with fixed number of components over data streams. Frontiers of Computer Science, 2014, 8(4): 563−580.

[6] Fan J, Gijbels I. Local Polynomial Modelling and Its Applications. Boca Raton, FL: CRC Press, 1996.

[7] Härdle W. Applied Nonparametric Regression. Cambridge, UK: Cambridge University Press, 1990.

[8] Härdle W, Linton O. Nonparametric regression. SFB Discussion Papers, 1995, 3(2): 867−877.

[9] Elizabeth S D, Wu J, Wang C, et al. Online updating of statistical inference in the big data setting. Technometrics, 2016, 58(3): 393−403.

[10] Han B, Comaniciu D, Zhu Y, et al. Sequential kernel density approximation and its application to real-time visual tracking. IEEE Transactions on Pattern Analysis Machine Intelligence, 2008, 30(7): 1186−1197.

[11] Spurek P, Byrski K, Tabor J. Online updating of active function cross-entropy clustering. Pattern Analysis and Applications, 2019, 22(4): 1409−1425.

[12] Kong E, Xia Y. On the efficiency of online approach to nonparametric smoothing of big data. Statistica Sinica, 2019, 29(1): 185−201.

[13] Deheuvels P. Estimation nonparamétrique de la densité par histogrammes généralisés. Revue de Statistique Appliquée, 1977, 25(3): 5−42.

[14] Stone M. Cross-validation and multinomial prediction. Biometrika, 1974, 61(3): 509−515.

[15] Cao R, Cuevas A, Manteiga W G. A comparative study of several smoothing methods in density estimation. Computational Statistics Data Analysis, 1994, 17(2): 153−176.

[16] Woodroofe M. On choosing a delta-sequence. Annals of Mathematical Statistics, 1970, 41(5): 1665−1671.

[17] Scott D W, Tapia R A, Thompson J R. Kernel density estimation revisited. Nonlinear Analysis Theory Methods Applications, 1977, 1(4): 339−372.

[18] Cwik J, Koronacki J. A combined adaptive-mixtures/plug-in estimator of multivariate probability densities. Computational Statistics & Data Analysis, 1997, 26(2): 199−218.

[19] Vincent P, Bengio Y. Non-local manifold Parzen windows. In: Advances in Neural Information Processing Systems 15. Cambridge, MA: MIT Press, 2003: 849−856.

[20] Whaley R E. Understanding the VIX. The Journal of Portfolio Management, 2009, 35(3): 98−105.

[21] Jiang G J, Tian Y S. Extracting model-free volatility from option prices: An examination of the VIX index. The Journal of Derivatives, 2007, 14(3): 35−60.

关于大数据下非参数平滑化在线更新方法的推断

陈子玉[1],郭潇[2*]

1. 中国科学技术大学管理学院统计与金融系,安徽合肥 230026;2. 中国科学技术大学管理学院,国际金融研究院,安徽合肥 230601

*通讯作者. E-mail:xiaoguo@ ustc. edu. cn

摘要:在线更新方法是一种有效的大数据分析方法. 本文证明了核密度和核回归在线模型的渐近性质,并进行了相应的统计推断. 提出了几种算法分别解决了核密度和回归中带宽选择的困难. 在模拟中验证了在线核密度模型的渐近正态性,并将在线线性核回归模型应用于波动率指数(VIX)预测. 实证结果表明,与经典的局部线性回归模型相比,该模型在预测连续到达的期权数据流方面性能相当,但是计算复杂度显著降低.

关键词:带宽;核估计;在线更新方法;统计推断;波动率指数预测

Appendix

Lemma A.1 (Lyapounov CLT) Suppose that for each n, $\omega_{n1},\cdots,\omega_{nn}$ are independent, $E\omega_{ni}=0$ and $\sigma_{ni}^2=E\omega_{ni}^2<\infty$. Define $s_n^2=\sum_{i=1}^n \sigma_{ni}^2$. Suppose further that for some $\delta>0$ the following condition holds:

$$\lim_{n\to\infty}\sum_{i=1}^n \frac{E\mid \omega_{ni}\mid^{2+\delta}}{s_n^{2+\delta}}=0,$$

Then we have

$$\sum_{i=1}^n \omega_{ni}/s_n \xrightarrow{\mathrm{d}} N(0,1).$$

Proof of Theorem 2.1 Recall that

$$\tilde f_n(\boldsymbol{x})=n^{-1}\sum_{i=1}^n\frac{1}{\tilde h_i^p}K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{\tilde h_i}\right),$$

where $\tilde h_i=ci^{-\alpha}$ and $\alpha=(4+p)^{-1}$. To derive the asymptotic normality of $\tilde f_n(\boldsymbol{x})$, we rewrite $\tilde f_n(\boldsymbol{x})-f(\boldsymbol{x})$ as sum of two parts in the following form,

$$\tilde f_n(\boldsymbol{x})-f(\boldsymbol{x})=\{E[\tilde f_n(\boldsymbol{x})]-f(\boldsymbol{x})\}+\{\tilde f_n(\boldsymbol{x})-E[\tilde f_n(\boldsymbol{x})]\}\overset{\text{def}}{=\!=}B_n+V_n,$$

where B_n and V_n contribute to the bias and variance of the estimator. For the first part B_n, based on the notation above, we can expand it as follows:

$$B_n=E[\tilde f_n(\boldsymbol{x})]-f(\boldsymbol{x})=n^{-1}\sum_{i=1}^n\int_{-\infty}^{\infty}\frac{1}{\tilde h_i^p}K\left(\frac{\boldsymbol{z}-\boldsymbol{x}}{\tilde h_i}\right)f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}-f(\boldsymbol{x}),$$

Noted that $\boldsymbol{z}=\boldsymbol{u}h_i+\boldsymbol{x}$, then $\mathrm{d}z=h_i^p\mathrm{d}\boldsymbol{u}$, and we have

$$B_n=\frac{1}{n}\sum_{i=1}^n\int_{-\infty}^{\infty}K(\boldsymbol{u})f(\tilde h_i\boldsymbol{u}+\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}-f(\boldsymbol{x})=$$

$$\frac{1}{n}\sum_{i=1}^n\int_{-\infty}^{\infty}K(\boldsymbol{u})[f(\boldsymbol{x})+\tilde h_i\boldsymbol{u}^{\mathrm T}\nabla f(\boldsymbol{x})+\frac{\tilde h_i^2}{2}\boldsymbol{u}^{\mathrm T}\mathscr{H}_f(\boldsymbol{x})\boldsymbol{u}]\,\mathrm{d}\boldsymbol{u}+o(\frac{1}{n}\sum_{i=1}^n\tilde h_i^2)-f(\boldsymbol{x})=$$

$$\frac{1}{2n}\mathrm{tr}\{\mathscr{H}_f(\boldsymbol{x})\}\sum_{i=1}^n\tilde h_i^2+o(\frac{1}{n}\sum_{i=1}^n\tilde h_i^2)=\frac{c^2}{2}\mathrm{tr}\{\mathscr{H}_f(\boldsymbol{x})\}n^{-2\alpha}\int_0^1 t^{-2\alpha}\mathrm{d}t(1+o(1))+o(n^{-2\alpha})=$$

$$\frac{c^2}{2(1-2\alpha)}\mathrm{tr}\{\mathscr{H}_f(\boldsymbol{x})\}n^{-2\alpha}+o(n^{-2\alpha}).$$

For the second part V_n, we can also expand it as follows:

$$V_n=\tilde f_n(\boldsymbol{x})-E[\tilde f_n(\boldsymbol{x})]=\frac{1}{n}\sum_{i=1}^n\frac{1}{\tilde h_i^p}\left[K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{\tilde h_i}\right)-EK\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{\tilde h_i}\right)\right]\overset{\text{def}}{=\!=}\sum_{i=1}^n V_{n,i}.$$

It's clear that $E(V_{n,i})=0$, then

$$\mathrm{Var}(V_{n_i})=E(V_{n,i}^2)=\frac{1}{n^2\tilde h_i^{2p}}E\left[K\left(\frac{X_i-x}{\tilde h_i}\right)-EK\left(\frac{X_i-x}{\tilde h_i}\right)\right]^2=$$

$$\frac{1}{n^2\tilde h_i^{2p}}\left[EK^2\left(\frac{X_i-x}{\tilde h_i}\right)-E^2K\left(\frac{X_i-x}{\tilde h_i}\right)\right]=\frac{1}{n^2\tilde h_i^{2p}}\int_{-\infty}^{\infty}K^2\left(\frac{z-x}{\tilde h_i}\right)f(z)\mathrm{d}z-\left(\frac{1}{n\tilde h_i^{p}}\int_{-\infty}^{\infty}K\left(\frac{z-x}{\tilde h_i}\right)f(z)\mathrm{d}z\right)^2=$$

$$\frac{1}{n^2\tilde h_i^{p}}\int_{-\infty}^{\infty}K^2(u)f(\tilde h_i u+x)\mathrm{d}u-\left(\frac{1}{n}\int_{-\infty}^{\infty}K(u)f(\tilde h_i u+x)\mathrm{d}u\right)^2=$$

$$\frac{1}{n^2\tilde h_i^{p}}\int_{-\infty}^{\infty}K^2(u)\left[f(x)+\tilde h_i u^{\mathrm T}\nabla f(x)\right]\mathrm{d}u+o\left(\frac{1}{n^2\tilde h_i^{p}}\right)+O\left(\frac{1}{n^2}\right)=\frac{f(x)}{n^2\tilde h_i^{p}}\int_{-\infty}^{\infty}K^2(u)\mathrm{d}u+o\left(\frac{1}{n^2\tilde h_i^{p}}\right).$$

Denote $R_2(K)=\int_{-\infty}^{\infty}K^2(u)\mathrm{d}u$, it's easy to see that

$$\mathrm{Var}(V_n)=\sum_{i=1}^{n}\mathrm{Var}(V_{n,i})=f(x)R_2(K)n^{p\alpha-1}(1+p\alpha)^{-1}c^{-p}+o(n^{p\alpha-1}).$$

Now, by Jenson inequalities, we have

$$\sum_{i=1}^{n}E\,|\,V_{n,i}\,|^{2+\delta}=\sum_{i=1}^{n}\frac{1}{(n\tilde h_i^{p})^{2+\delta}}E\left|K\left(\frac{X_i-x}{\tilde h_i}\right)-EK\left(\frac{X_i-x}{\tilde h_i}\right)\right|^{2+\delta}\leqslant$$

$$\sum_{i=1}^{n}\frac{2^{2+\delta}}{(n\tilde h_i^{p})^{2+\delta}}E\left|K\left(\frac{X_i-x}{\tilde h_i}\right)\right|^{2+\delta}=\sum_{i=1}^{n}\frac{2^{1+\delta}}{n^{2+\delta}\tilde h_i^{p(1+\delta)}}\int|\,K(u)\,|^{2+\delta}f(\tilde h_i u+x)\mathrm{d}u=O(n^{-4\alpha(\delta+1)}).$$

Hence, we can derive the following equation:

$$\sum_{i=1}^{n}\frac{E\,|\,V_{n,i}\,|^{2+\delta}}{\mathrm{Var}(V_n)^{1+\delta/2}}=O(n^{-\frac{2\delta}{p+4}}).$$

Applying Lemma A.1 (the Liapounov CLT) yields

$$\frac{\sum_{i=1}^{n}V_{ni}}{\sqrt{\mathrm{Var}(V_n)}}=\frac{V_n}{\sqrt{\mathrm{Var}(V_n)}}=\frac{\tilde f_n(x)-f(x)-B_n}{\sqrt{\mathrm{Var}(V_n)}}\xrightarrow{\mathrm d}N(0,1).$$

Therefore the following result holds,

$$\sqrt{\frac{(1+p\alpha)c^p}{f(x)R_2(K)n^{p\alpha-1}}}\left(\tilde f_n(x)-f(x)-\frac{c^2}{2(1-2\alpha)}[\mathrm{tr}\{\mathscr{H}_f(x)\}]n^{-2\alpha}\right)\xrightarrow{\mathrm d}N(0,1).$$

Lemma A.2 Suppose A1−A2 hold and $n^{1-p\alpha}\to\infty$ as $n\to\infty$, $\tilde f_n(x)\xrightarrow{\mathrm P}f(x)$.

Proof By Chebyshev's inequality,

$$P(|\tilde f_n(x)-f(x)|>\varepsilon)\leqslant\frac{E(\tilde f_n(x)-f(x))^2}{\varepsilon^2}=\frac{E\tilde f_n^2(x)-f^2(x)}{\varepsilon^2}=\frac{c_1(x)}{\varepsilon^2n^{1-p\alpha}}+o(n^{p\alpha-1}),$$

where $c_1(x)=\{(1+p\alpha)c^p\}^{-1}f(x)R_2(K)$.

Proof of Corollary 2.1 Since $\tilde f_n(x)\xrightarrow{\mathrm p}f(x)$, $\mathrm{tr}\{\mathscr{H}_{\tilde f_n}(x)\}\xrightarrow{\mathrm p}\mathrm{tr}\{\mathscr{H}_f(x)\}$, and we have

$$\frac{\tilde c^2[\mathrm{tr}\{\mathscr{H}_{\tilde f_n}(x)\}]}{2(1-2\alpha)}n^{-2\alpha}\sqrt{\frac{(1+p\alpha)\tilde c^p}{\tilde f_n(x)R_2(K)n^{p\alpha-1}}}=\frac{(1+p\alpha)^{1/2}}{2(1-2\alpha)}\frac{\mathrm{tr}\{\mathscr{H}_{\tilde f_n}(x)\}}{(\tilde f_n(x)R_2(K))^{1/2}}\tilde c^{2+p/2}n^{2/(p+4)-2\alpha}=$$

$$\sqrt{\frac{p+4}{2(p+2)}}\frac{\mathrm{tr}\{\mathscr{H}_{\tilde f_n}(x)\}}{\sqrt{\tilde f_n(x)R_2(K)}}\sqrt{\frac{p(p+2)}{2(p+4)}\frac{\tilde f_n(x)R_2(K)}{\mathrm{tr}^2\{\mathscr{H}_{\tilde f_n}(x)\}}}n^0=\frac{\sqrt p}{2}.$$

Combining Slutsky theorem, Corollary 2.1 are derived.

Proof of Theorem 3.1 To begin with, we rewrite $\tilde m_{nw}(x)-m(x)$ as follows:

$$\tilde m_{nw}(x)-m(x)=\frac{[\tilde m_{nw}(x)-m(x)]\tilde f_n(x)}{\tilde f_n(x)}\overset{\mathrm{def}}{=\!=}\frac{\tilde M(x)}{\tilde f_n(x)}.$$

Since $Y_i=m(X_i)+\varepsilon_i$, we have

$$\tilde M(x)=[\tilde m_{nw}(x)-m(x)]\tilde f_n(x)=\frac{1}{n}\sum_{i=1}^{n}[Y_i-m(x)]K_{\tilde h_i}(X_i-x)=$$

$$\frac{1}{n}\sum_{i=1}^{n}[m(X_i)-m(x)]K_{\tilde h_i}(X_i-x)+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_iK_{\tilde h_i}(X_i-x)\overset{\mathrm{def}}{=\!=}b_n+v_n.$$

Here we consider the property of b_n as following:

$$E[b_n] = \frac{1}{n}\sum_{i=1}^{n}\int [m(\boldsymbol{X}_i) - m(\boldsymbol{x})]K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})f(\boldsymbol{X}_i)\mathrm{d}\boldsymbol{X}_i =$$

$$\frac{1}{n}\sum_{i=1}^{n}\int \frac{1}{\tilde{h}_i^p}K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{\tilde{h}_i}\right)[m(\boldsymbol{X}_i) - m(\boldsymbol{x})]f(\boldsymbol{X}_i)\mathrm{d}\boldsymbol{X}_i =$$

$$\frac{1}{n}\sum_{i=1}^{n}\int K(\boldsymbol{u})[m(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x}) - m(\boldsymbol{x})]f(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x})\mathrm{d}\boldsymbol{u} =$$

$$\frac{1}{n}\sum_{i=1}^{n}\int K(\boldsymbol{u})[\tilde{h}_i\boldsymbol{u}^{\mathrm{T}}\nabla m(\boldsymbol{x}) + \frac{\tilde{h}_i^2}{2}\boldsymbol{u}^{\mathrm{T}}\mathscr{H}_m(\boldsymbol{x})\boldsymbol{u}][f(\boldsymbol{x}) + \tilde{h}_i\boldsymbol{u}^{\mathrm{T}}\nabla f(\boldsymbol{x}) + \frac{\tilde{h}_i^2}{2}\boldsymbol{u}^{\mathrm{T}}\mathscr{H}_f(\boldsymbol{x})\boldsymbol{u}]\mathrm{d}\boldsymbol{u} + o(\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_i^2) =$$

$$\frac{1}{2n}\sum_{i=1}^{n}\tilde{h}_i^2[\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\}f(\boldsymbol{x}) + 2\nabla^{\mathrm{T}} m(\boldsymbol{x})\nabla f(\boldsymbol{x})] + o(\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_i^2) =$$

$$\frac{c^2}{2n}\sum_{i=1}^{n} i^{-2\alpha}[\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\}f(\boldsymbol{x}) + 2\nabla^{\mathrm{T}} m(\boldsymbol{x})\nabla f(\boldsymbol{x})] + o(n^{-2\alpha}) =$$

$$\frac{c^2}{2(1-2\alpha)}[\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\}f(\boldsymbol{x}) + 2\nabla^{\mathrm{T}} m(\boldsymbol{x})\nabla f(\boldsymbol{x})]n^{-2\alpha} + o(n^{-2\alpha}) = B_{nw}(\boldsymbol{x})f(\boldsymbol{x}) + o(n^{-2\alpha}),$$

where $B_{nw}(\boldsymbol{x}) = c^2\{2(1-2\alpha)\}^{-1}[\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\} + 2\nabla^{\mathrm{T}} m(\boldsymbol{x})\nabla f(\boldsymbol{x})/f(\boldsymbol{x})]n^{-2\alpha}$, and $\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\} = \sum_{i=1}^{p}\partial^2 m/\partial x_i^2$. Note that $\nabla m(\boldsymbol{x})$, $\nabla f(\boldsymbol{x})$ are gradient of $m(\boldsymbol{x})$ and $f(\boldsymbol{x})$. To continue,

$$\mathrm{Var}[b_n] = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}[m(\boldsymbol{X}_i) - m(\boldsymbol{x})]K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\right) = \sum_{i=1}^{n}\frac{1}{n^2\tilde{h}_i^{2p}}E\left[K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{\tilde{h}_i}\right)[m(\boldsymbol{X}_i) - m(\boldsymbol{x})]\right]^2 -$$

$$\sum_{i=1}^{n}\frac{1}{n^2}E^2\left[\frac{1}{\tilde{h}_i^p}K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{\tilde{h}_i}\right)[m(\boldsymbol{X}_i) - m(\boldsymbol{x})]\right] = \sum_{i=1}^{n}\frac{1}{n^2\tilde{h}_i^p}\int K^2(\boldsymbol{u})[m(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x}) - m(\boldsymbol{x})]^2 f(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x})\mathrm{d}\boldsymbol{u} -$$

$$\sum_{i=1}^{n}\frac{1}{n^2}\left[\int K(\boldsymbol{u})[m(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x}) - m(\boldsymbol{x})]f(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x})\mathrm{d}\boldsymbol{u}\right]^2 = O(\sum_{i=1}^{n}\frac{1}{n^2\tilde{h}_i^{p-2}}) = O(n^{-6\alpha}).$$

It's clear that

$$E[v_n] = E(E[v_n \mid \boldsymbol{x}]) = 0.$$

Denote $v_n = \sum_{i=1}^{n}\varepsilon_i K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})/n = \sum_{i=1}^{n} v_{ni}$, and it gives that

$$\mathrm{Var}[v_n] = \sum_{i=1}^{n}\frac{\sigma_\varepsilon^2}{n^2}E[K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})]^2 =$$

$$\sum_{i=1}^{n}\frac{\sigma_\varepsilon^2}{n^2\tilde{h}_i^p}\int K^2(\boldsymbol{u})f(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x})\mathrm{d}\boldsymbol{u} = \sum_{i=1}^{n}\frac{\sigma_\varepsilon^2}{n^2\tilde{h}_i^p}R_2(K)f(\boldsymbol{x}) + o(\sum_{i=1}^{n}\frac{1}{n^2\tilde{h}_i^p}) =$$

$$\frac{n^{p\alpha-1}R_2(K)f(\boldsymbol{x})\sigma_\varepsilon^2}{(1+p\alpha)c^p} + o(n^{p\alpha-1}).$$

Moerover,

$$\sum_{i=1}^{n}E\mid v_{ni}\mid^{2+\delta} = \frac{1}{n^{2+\delta}}\sum_{i=1}^{n}E\mid \varepsilon_i K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\mid^{2+\delta} = \frac{E(\varepsilon_i)^{2+\delta}}{n^{2+\delta}}\sum_{i=1}^{n}E\mid K_{\tilde{h}_i}(\boldsymbol{X}_i - \boldsymbol{x})\mid^{2+\delta} =$$

$$\frac{E(\varepsilon_i)^{2+\delta}}{n^{2+\delta}}\sum_{i=1}^{n}\int\frac{1}{\tilde{h}_i^{p(2+\delta)}}K^{2+\delta}\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{\tilde{h}_i}\right)f(\boldsymbol{X}_i)\mathrm{d}(\boldsymbol{X}_i) =$$

$$\frac{E(\varepsilon_i)^{2+\delta}}{n^{2+\delta}}\sum_{i=1}^{n}\frac{1}{\tilde{h}_i^{p+2\delta}}\int K^{2+\delta}(\boldsymbol{u})f(\tilde{h}_i\boldsymbol{u} + \boldsymbol{x})\mathrm{d}\boldsymbol{u} = O(n^{(p\alpha-1)(1+\delta)}),$$

which yields that

$$\sum_{i=1}^{n}\frac{E\mid v_{ni}\mid^{2+\delta}}{\mathrm{Var}(v_n)^{1+\delta/2}} = O(n^{(p\alpha-1)\delta/2}).$$

Consider that

$$T(\boldsymbol{x}) = \frac{b_n - B_{nw}(\boldsymbol{x})f(\boldsymbol{x})}{\sqrt{\mathrm{Var}(v_n)}} = \frac{b_n - E(b_n)}{\sqrt{\mathrm{Var}(v_n)}},$$

then the following equality holds,

$$P(|T(\boldsymbol{x})|>\varepsilon)\leqslant \frac{E(T(\boldsymbol{x}))^2}{\varepsilon^2}=\frac{\mathrm{Var}(b_n)}{\varepsilon^2\,\mathrm{Var}(v_n)}=o(n^{-2\alpha}).$$

According to the Lemma A.1, it's easy to find that

$$\frac{v_n}{\sqrt{\mathrm{Var}(v_n)}}=\frac{\widetilde{M}(\boldsymbol{x})-b_n}{\sqrt{\mathrm{Var}(v_n)}}=\frac{\widetilde{M}(\boldsymbol{x})-B_{nw}(\boldsymbol{x})f(\boldsymbol{x})+B_{nw}(\boldsymbol{x})f(\boldsymbol{x})-b_n}{\sqrt{\mathrm{Var}(v_n)}}\xrightarrow{\mathrm{d}}N(0,1).$$

Combined with the Slutsky theorem, Theorem 3.1 is easy to obtained.

Proof of Corollary 3.1 It's easy to see that

$$\sqrt{\frac{(1+p\alpha)\tilde{c}^p f(\boldsymbol{x})}{n^{p\alpha-1}R_2(K)\sigma_\varepsilon^2}}B_{nw}(\boldsymbol{x})=\sqrt{\frac{(1+p\alpha)f(\boldsymbol{x})}{2(1-2\alpha)R_2(K)\sigma_\varepsilon^2}}H(\boldsymbol{x})c^{p/2+2}n^{-2\alpha+(1-p\alpha)/2}=\frac{\sqrt{p}}{2},$$

where $H(\boldsymbol{x})=[\mathrm{tr}\{\mathscr{H}_{\tilde{m}}(\boldsymbol{x})\}+2\nabla^{\mathrm{T}}\tilde{m}(\boldsymbol{x})\nabla f(\boldsymbol{x})/f(\boldsymbol{x})]$. Since $\widetilde{\mathscr{H}}_m(x)\xrightarrow{\mathrm{p}}\mathscr{H}_m(x)$, and $\nabla\widetilde{m}^{\mathrm{T}}(x)\xrightarrow{\mathrm{p}}\nabla m(\boldsymbol{x})$, applying Slutsky theorem gives Corollary 3.1.

Proof of Theorem 3.2 Denote $\boldsymbol{b}=[m(\boldsymbol{x}),\tilde{h}_i\nabla^{\mathrm{T}}m(\boldsymbol{x})]^{\mathrm{T}}$, $\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})=[1,(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}/\tilde{h}_i]^{\mathrm{T}}$, and

$$\widetilde{S}_1(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}{}^{\mathrm{T}}(\boldsymbol{x}),\ \widetilde{S}_2(\boldsymbol{x},Y)=\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})Y_i.$$

So we can see that

$$\begin{gathered}\tilde{\boldsymbol{b}}-\boldsymbol{b}=\{\widetilde{S}_1(\boldsymbol{x})\}^{-1}\widetilde{S}_2(\boldsymbol{x},Y)-\{\widetilde{S}_1(\boldsymbol{x})\}^{-1}\widetilde{S}_1(\boldsymbol{x})\boldsymbol{b}=\\ \{\widetilde{S}_1(\boldsymbol{x})\}^{-1}\{\frac{1}{n}\sum_{i=1}^n[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})][Y_i-\widetilde{\boldsymbol{X}}_{\mathrm{in}}{}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{b}]\}=\\ \{\widetilde{S}_1(\boldsymbol{x})\}^{-1}\{\frac{1}{n}\sum_{i=1}^n[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})][m(\boldsymbol{X}_i)+\varepsilon_i-m(\boldsymbol{x})-(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}\nabla m(\boldsymbol{x})]\}=\\ \{\widetilde{S}_1(\boldsymbol{x})\}^{-1}\{\frac{1}{n}\sum_{i=1}^n[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})][\frac{1}{2}(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}\mathscr{H}_m(\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})+\varepsilon_i+R(\boldsymbol{x},\boldsymbol{X}_i)]\}\overset{\mathrm{def}}{=\!=}\\ \{\widetilde{S}_1(\boldsymbol{x})\}^{-1}\{B_n(\boldsymbol{x})+V_n(\boldsymbol{x})+R_n(\boldsymbol{x})\},\end{gathered}$$

where

$$\begin{gathered}B_n(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^n[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})][\frac{1}{2}(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}\mathscr{H}_m(\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})],\\ V_n(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^n[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})]\,\varepsilon_i,\\ R_n(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^n[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})]R(\boldsymbol{x},\boldsymbol{X}_i),\end{gathered}$$

and $R(\boldsymbol{x},\boldsymbol{X}_i)$ is the residual term after the second-order Taylor expansion of $m(\boldsymbol{X}_i)$. Moreover,

$$\begin{gathered}E[B_n(\boldsymbol{x})]=\frac{1}{n}\sum_{i=1}^n E[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})[1,(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}/\tilde{h}_i]^{\mathrm{T}}][\frac{1}{2}(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}\mathscr{H}_m(\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})]=\\ \frac{1}{2n}\sum_{i=1}^n\tilde{h}_i^2\int K(\boldsymbol{u})[1,\boldsymbol{u}^{\mathrm{T}}]^{\mathrm{T}}\boldsymbol{u}^{\mathrm{T}}\mathscr{H}_m\boldsymbol{u}\,f(\tilde{h}_i\boldsymbol{u}+\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}=\frac{c^2n^{-2\alpha}f(\boldsymbol{x})}{2(1-2\alpha)}\begin{pmatrix}\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\}\\ \boldsymbol{0}_{p\times1}\end{pmatrix}+o(\frac{1}{n}\sum_{i=1}^n\tilde{h}_i^2)=\\ \frac{c^2n^{-2\alpha}f(\boldsymbol{x})}{2(1-2\alpha)}\begin{pmatrix}\mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\}\\ \boldsymbol{0}_{p\times1}\end{pmatrix}+o(n^{-2\alpha}).\end{gathered}$$

Since $EV_n(\boldsymbol{x})=0$, it can be seen that

$$\begin{gathered}\mathrm{Var}(V_n(\boldsymbol{x}))=E(V_n(\boldsymbol{x}))^2=\frac{1}{n^2}\sum_{i=1}^n E[K_{\tilde{h}_i}(\boldsymbol{X}_i-\boldsymbol{x})\widetilde{\boldsymbol{X}}_{\mathrm{in}}(\boldsymbol{x})]^2\varepsilon_i^2=\\ \sum_{i=1}^n\frac{1}{n^2\tilde{h}_i^p}\int K^2(\boldsymbol{u})[1,\boldsymbol{u}^{\mathrm{T}}]^{\mathrm{T}}[1,\boldsymbol{u}^{\mathrm{T}}]f(\tilde{h}_i\boldsymbol{u}+\boldsymbol{x})\mathrm{d}\boldsymbol{u}=\begin{pmatrix}R_2(K)&\boldsymbol{0}_{1\times p}\\ \boldsymbol{0}_{p\times1}&\widetilde{R}_2(K)\end{pmatrix}\frac{\sigma_\varepsilon^2f(\boldsymbol{x})}{c^p(1+p\alpha)}n^{p\alpha-1}+o(n^{p\alpha-1}),\end{gathered}$$

where $\widetilde{R}_2(K)=\int\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$. In addition, if A and B are two fixed matrices of the same dimension and A^{-1} exists, then we have

$$(A+a_nB)^{-1}=A^{-1}-a_nA^{-1}BA^{-1}+O(a_n^2),$$

where $a_n\to0$. It is clear that

$$\tilde{S}_1(\boldsymbol{x}) = f(\boldsymbol{x})\boldsymbol{I}_{p+1} + O_p(n^{-\alpha}),$$

so we have

$$\{\tilde{S}_1(\boldsymbol{x})\}^{-1} = f^{-1}(\boldsymbol{x})\boldsymbol{I}_{p+1} + O_p(n^{-\alpha}).$$

Similar to the proof of the Theorem 2.1, by Chebyshev's inequality and the Lyapunov CLT, we have

$$n^{(1-p\alpha)/2}(\tilde{M}(\boldsymbol{x}) - M(\boldsymbol{x}) - B_{ll}(\boldsymbol{x})) \xrightarrow{\mathrm{d}} N(o, \Sigma),$$

where

$$B_{ll}(\boldsymbol{x}) = \frac{c^2 n^{-2\alpha} f(\boldsymbol{x})}{2(1-2\alpha)} \begin{pmatrix} \mathrm{tr}\{\mathscr{H}_m(\boldsymbol{x})\} \\ \boldsymbol{0}_{p\times 1} \end{pmatrix}, \quad \Sigma = \frac{\sigma_\varepsilon^2}{c^p(1+p\alpha)f(\boldsymbol{x})} \begin{pmatrix} R_2(K) & \boldsymbol{0}_{1\times p} \\ \boldsymbol{0}_{p\times 1} & \tilde{R}_2(K) \end{pmatrix},$$

Since one can easily verify that the conditions in Lemma A.1 are all satisfied. Therefore, we can apply the Liapounov CLT to conclude Theorem 3.2.

Proof of Corollary 3.2 The proof of Corollary 3.2 is analogous to that of Corollary 3.1. To avoid duplication, descriptions are not provided in this paper.

(Continued from p. 389)

Proof of Proposition 4.1 The score function of θ_k in the meTPR model is

$$s_k(\theta; y) = \frac{1}{2}\mathrm{Tr}((s_1\tilde{\Sigma}_n^{-1} yy^{\mathrm{T}}\tilde{\Sigma}_n^{-1} - \tilde{\Sigma}_n^{-1})\frac{\partial\tilde{\Sigma}_n}{\partial\theta_k}) \tag{A13}$$

Let l be the length of θ and $s_T(\theta; y) = (s_1(\theta; y), \cdots, s_l(\theta; y))^{\mathrm{T}}$. The score function becomes that under the GPR model when $s_1 = 1$. The impact factor $s_1 = (n+2v)/(2(v-1)+y^{\mathrm{T}}\tilde{\Sigma}_n^{-1}y)$ is very important for estimating θ. For example, when $y_j \to \infty$ for some j, the score function $s_T(\theta; y)$ is bounded, while that from the GPR model is not.

For a given parameter v, following Ref. [25] the influence function for the estimator $\hat{\theta}$ is

$$\mathrm{IF}(y; \hat{\theta}, F) = -(E(\frac{\partial^2 l(\theta; v)}{\partial\theta\partial\theta^{\mathrm{T}}}))^{-1} s_T(\theta; y) \tag{A14}$$

Note that the matrix $\partial^2 l(\theta; v)/(\partial\theta\partial\theta^{\mathrm{T}})$ is bounded for y, which indicates that the influence function of $\hat{\theta}$ is bounded under the meTPR model. Similarly, we can get that the score function is unbounded. So, for mGPR, the influence function of parameter estimation is also unbounded.

Proof of Proposition 4.2 Obviously $q^2 = (y-F_0(x))^{\mathrm{T}}(y-F_0(x))/\phi_0 = O(n)$. Under the condition of Lemma A.1 and the condition that $\| F_0 \|_k$ is bounded and $E_x(\log|I_n+\phi_0^{-1}\tilde{K}_n|) = o(n)$ is established. According to Lemma A.1, for a positive constant c and any $\varepsilon>0$, when n is large enough, we have

$$\frac{1}{n}E_x(D[p_{\phi_0}(y \mid F_0, x), p_{\phi_0,\hat{\theta}}(y \mid x)]) =$$

$$E_x\int \frac{1}{n}(-\log p_{\phi_0,\hat{\theta}}(y \mid x) + \log p_{\phi_0}(y \mid F_0, x))\mathrm{d}p_{\phi_0,\hat{\theta}}(y \mid x) \leqslant$$

$$E_x\int(\frac{1}{2n}\log|I_n + \phi_0^{-1}\tilde{K}_n| + \frac{q^2 + 2(v-1)}{2n(n+2v-2)}(\| F_0 \|_k^2 + c) + \frac{c}{n} + \varepsilon)\mathrm{d}p_{\phi_0,\hat{\theta}}(y \mid x) \tag{A15}$$

It gives

$$\frac{1}{n}E_x(D[p_{\phi_0}(y \mid F_0, x), p_{\phi_0,\hat{\theta}}(y \mid x)]) \to 0, \text{ as } n \to \infty \tag{A16}$$

Thus, the proposition holds.