Journal of Statistical Planning and Inference 211 (2021) 423–438

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Robust estimation of nonparametric function via addition sequence

WenWu Wang [a,*], Wei Shen [b], Tiejun Tong [b]

[a] School of Statistics, Qufu Normal University, Jining, Shandong, China
[b] Department of Mathematics, Hong Kong Baptist University, Hong Kong

ARTICLE INFO

Article history:
Received 24 April 2018
Received in revised form 22 March 2019
Accepted 22 July 2020
Available online 21 August 2020

Keywords:
Robust nonparametric function estimation
Differenced method
Median smoothing
Platykurtic and heavy-tailed errors

ABSTRACT

In this paper, we propose a robust method for the estimation of regression function. By symmetric addition, we change platykurtic errors into leptokurtic errors; and then estimate the nonparametric function by the local polynomial least absolute deviation regression. Different from the local polynomial least squares estimator, the new estimator is robust for outliers and heavy-tailed errors even if the error variance does not exist; different from the usual local polynomial least absolute deviation estimator and the composite quantile regression estimator, it does not depend on the finite density values at chosen quantile points, but relies on the expectation of the error density. To improve the finite sample performance, two bias-reduced versions are further proposed under different smoothness conditions. For the equidistant designs, the asymptotic properties are established. In simulations, the new estimator has the less mean square errors than its main competitors in the presence of platykurtic and heavy-tailed errors.

1. Introduction

In nonparametric estimation of regression function, the local polynomial least squares (LS) regression is a successful and popular method, and its asymptotic theory has been well studied in the literature (Fan and Gijbels, 1996). If the errors are normally distributed, the LS estimator is the most efficient estimator and has the likelihood interpretation (Fan et al., 1998). Facing the outliers and heavy-tailed errors, the LS estimator is not robust and its efficiency cannot be guaranteed, i.e., for the Cauchy distribution. Therefore, it is important to develop robust and efficient estimation methods for many applications such as finance and economics (Zhao and Xiao, 2014).

Traditionally, there were two kinds of robust procedures: locally weighted LS regression and local least absolute deviation (LAD) regression. Locally weighted LS procedure aims to reduce the influence of outliers by assigning the down-weight to outliers, and in spirit is locally linear, including locally weighted polynomial LS fitting (Cleveland, 1979), kernel M-smoother method (Härdle and Gasser, 1984), and spline smoother (Silverman, 1985). Local LAD regression is a different procedure. Tukey (1977) proposed various modifications of local median smoothing, Fan and Hall (1994) proposed a framework and gave its asymptotic efficiency. Wang and Scott (1994) proposed the local polynomial LAD regression to further reduce the bias. Welsh (1996) considered the estimation of the regression function and spread functions and their derivatives. Brown et al. (2008) further proposed the wavelet median regression. These methods are robust for outliers and heavy-tailed errors, meanwhile need to satisfy an implied condition to keep the estimation efficiency: the error density

* Corresponding author.
E-mail address: wengewsh@sina.com (W. Wang).

https://doi.org/10.1016/j.jspi.2020.07.007
0378-3758/

is leptokurtic. If the density function value at median becomes smaller, the estimation variance becomes larger and thus the estimation efficiency is lower.

To improve the efficiency, the quantile regression (Koenker and Bassett, 1978; Koenker, 2005) provides a useful technique. Yu and Jones (1998) and Härdle et al. (2013) proposed the local quantile regression. For certain distributions, a quantile estimator at non-median may deliver a more efficient estimator than the LAD estimator. Since quantile regression provides a way of exploiting the distribution information, combining information over multiple quantiles may improve the estimation efficiency. Kai et al. (2010) proposed the local composite quantile regression (CQR) estimator by simple averaging multiple quantile regression estimators, which is asymptotically equivalent to the local LS estimator as the number of quantiles goes to infinity. Fan et al. (2018) generalized CQR to the single-index model. Zhao and Xiao (2014) proposed the locally weighted quantile average (WQA) estimator by optimally weighting multiple quantile regression estimators, which achieves the Cramer–Rao lower bound of variance as the number of quantiles goes to infinity. However, the asymptotic properties of CQR and WQA estimators depend on the information at all quantile points. In practical applications, both estimators are a generalization of the LAD estimator, from one quantile to finite quantiles. The estimation efficiency depends on the error density values at finite quantiles.

In this paper, we propose a new and robust method to deal with platykurtic errors and outliers. By the additive transformation of the original data, we change platykurtic errors into peak errors; then we use the local LAD regression to estimate the regression function. The new function estimation is robust for outliers and heavy-tailed errors, meanwhile improve the estimation efficiency for platykurtic errors.

2. Estimation methodology

Consider the nonparametric regression model

$$Y_i = m(x_i) + \epsilon_i, \quad 0 \le i \le n, \tag{1}$$

where $x_i = i/n$ are the equidistant design points, $m(\cdot)$ is an unknown smooth regression function, and ϵ_i are independent and identically distributed (i.i.d.) random errors with median 0 and a symmetric density function $f(\cdot)$.

2.1. An illustration

In the past ten years, the China house prices change dramatically, which go through three-time fast increasing. Facing the dramatic changes, Wang et al. (2019) proposed a robust and efficient method to estimate the relative growth rate, which is equivalent to estimate the first-order derivative. In this paper, we propose a new method to estimate the price curves robustly (see Section 5).

To show the new idea, we consider four error distributions: Uniform$[-1, 1]$ and $50\%N(-d, 1) + 50\%N(d, 1)$ with $d = 1, 2, 3$. The four distributions have small density values at median 0. For independent ϵ_1, ϵ_2 in each of the four distributions, we construct the new error $\eta = (\epsilon_1 + \epsilon_2)/2$. The new errors η with median 0 have larger peak values in Fig. 1. This toy example shows that the error with platy kurtosis can change into the peak kurtosis error by addition.

2.2. Estimation methodology

Define the symmetric (about i) first-order addition sequence as

$$Y_{ij}^{(1)} = (Y_{i-j} + Y_{i+j})/2, \quad 1 \le j \le k, \tag{2}$$

where k is a positive integer with $k+1 \le i \le n-k$. Assume that $m(\cdot)$ is continuously differentiable. Then the first-order Taylor expansions of $m(x_{i\pm j})$ around x_i are

$$m(x_{i\pm j}) = m(x_i) \pm m^{(1)}(x_i)\frac{j}{n} + o\left(\frac{j}{n}\right).$$

We decompose $Y_{ij}^{(1)}$ into two parts as follows

$$\begin{aligned} Y_{ij}^{(1)} &= \frac{m(x_{i-j}) + m(x_{i+j})}{2} + \frac{\epsilon_{i-j} + \epsilon_{i+j}}{2} \\ &= m(x_i) + \eta_{ij} + o\left(\frac{j}{n}\right), \end{aligned}$$

where $\eta_{ij} = (\epsilon_{i-j} + \epsilon_{i+j})/2$ with median$(\eta_{ij}) = 0$. Thus the median of $Y_{ij}^{(1)}$ is

$$\text{Median}[Y_{ij}^{(1)}] = m(x_i) + o\left(\frac{j}{n}\right) \approx m(x_i),$$

where $j = o(n)$. We estimate the constant by the local LAD regression.

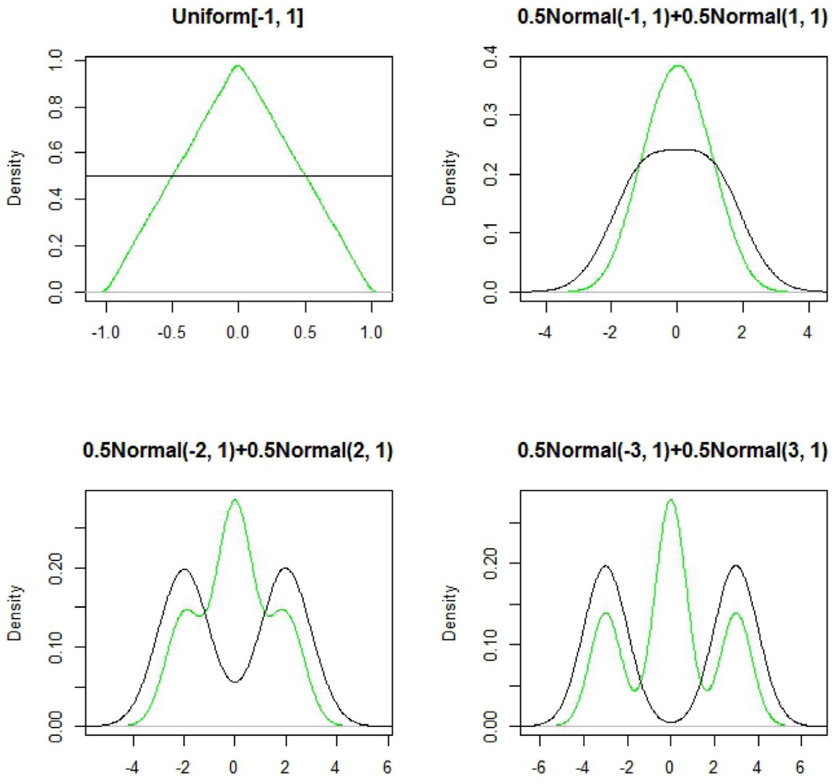

Fig. 1. Densities of original errors (black line) and new errors (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to deduce the accurate estimation bias, we further assume that $m(\cdot)$ is two times continuously differentiable. Following the paradigm of Wang and Scott (1994) and Wang et al. (2019), we discard the higher-order terms of $m(\cdot)$ and locally assume the approximate model is

$$Y_{ij}^{(1)} = \beta_{i20} + \beta_{i22}x_j^2 + \eta_{ij}, \quad 1 \le j \le k, \tag{3}$$

where $\beta_{i2} = (\beta_{i20}, \beta_{i22})^T = (m(x_i), m^{(2)}(x_i)/2)^T$ are the unknown coefficients of the true underlying quadric function. Now we apply the local constant LAD regression to estimate the constant as

$$\hat{b}_{i0} = \arg\min_{b_{i0}} \sum_{j=1}^{k} |Y_{ij}^{(1)} - b_{i0}|,$$

and define the regression function estimator as

$$\hat{m}(x_i) = \hat{b}_{i0}. \tag{4}$$

The robust estimator $\hat{m}(x_i)$ will not converge to $m(x_i)$, but rather to β_{i1}, where

$$\beta_{i1} \triangleq \arg\min_{c_{i1}} \sum_{j=1}^{k} \left| c_{i1} - (\beta_{i20} + \beta_{i22}x_j^2) \right|.$$

Note that as $k \to \infty$, β_{i1} is close to $m(x_i)$. Next we first establish the asymptotic normality of $\hat{m}(x_i)$.

Theorem 1. *Assume that ϵ_i are i.i.d. random errors with median 0 and a continuous symmetric density $f(\cdot)$. For the equidistant design and true model (3), as $k \to \infty$ and $k/n \to 0$, the robust error estimator $\hat{m}(x_i)$ in (4) is normally distributed*

$$2k^{1/2}g(0)(\hat{m}(x_i) - \beta_{i1}) \xrightarrow{d} N(0, 1),$$

where $g(0) = 2\int_{-\infty}^{\infty} f^2(x)\mathrm{d}x$, and $\xrightarrow{d}$ denotes convergence in distribution.

Remark 1. We assume that the error distribution is symmetric. Such a condition is only used to ensure that the quantity to which the new estimator converges is the conditional mean function. This is similar to the case when using the local LAD with median 0 to estimate the conditional mean function. In other words, we need assume that the mean and median of the error distribution coincide. The same assumption is used in Kai et al. (2010).

Now we give the asymptotic properties of $\hat{m}(x_i)$.

Theorem 2. *Under the assumptions of Theorem 1, the bias and variance of the robust estimator in (4) are, respectively,*

$$\text{Bias}[\hat{m}(x_i)] \triangleq \text{E}[\hat{m}(x_i)] - m(x_i) \approx \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}, \quad \text{Var}[\hat{m}(x_i)] \approx \frac{1}{4g(0)^2 k}.$$

The optimal k that minimizes the asymptotic mean square error (AMSE) is

$$k_{opt} \approx \left(\frac{9}{4g(0)^2 m^{(2)}(x_i)^2}\right)^{1/5} n^{4/5},$$

and, consequently, the minimum AMSE is

$$\text{AMSE}[\hat{m}(x_i)] \approx 0.27\left(\frac{m^{(2)}(x_i)^2}{g(0)^8}\right)^{1/5} n^{-4/5}.$$

Remark 2. As for the choice of the tuning parameter, there are two standard approaches: plug-in approach and cross-validation approach. Since our estimation method leaves Y_i out to estimate $m(x_i)$, we can adopt the leave-one-out cross-validation to estimate k. The criterion is

$$\hat{h} = \arg\min \sum_{i=1}^{n}(Y_i - \hat{m}(x_i))^2.$$

The more details refer to Härdle and Marron (1985) and Li and Racine (2004).

Since the higher-order terms have no effect on the asymptotic results including the bias and the variance, thus the robust estimator (4) has asymptotic property in the nonparametric model (1) .

Corollary 1. *For the nonparametric regression model (1), as $k \to \infty$ and $k/n \to 0$, the robust estimator in (4) is normal distributed*

$$2k^{1/2}g(0)(\hat{m}(x_i) - m(x_i) - \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}) \xrightarrow{d} N(0,1),$$

as $k \to \infty$ and $k^5/n^4 \to 0$, the robust estimator in (4) is normal distributed

$$2k^{1/2}g(0)\left(\hat{m}(x_i) - m(x_i)\right) \xrightarrow{d} N(0,1).$$

Remark 3. As for the choice of addition sequence, we can define the difference sequence $Y_{ij}^{(2)} = (Y_{i-j} - 2Y_i + Y_{i+j})/2$. Decompose $Y_{ij}^{(2)}$ into two parts and simplify it to

$$\begin{aligned} Y_{ij}^{(2)} &= \frac{m(x_{i-j}) - 2Y_i + m(x_{i+j})}{2} + \frac{\epsilon_{i-j} - 2\epsilon_i + \epsilon_{i+j}}{2} \\ &= -\epsilon_i + \eta_{ij} + o\left(\frac{j}{n}\right). \end{aligned}$$

Since i is fixed as j changes, we estimate ϵ_i as

$$\hat{\epsilon}_i = \arg\min_{\epsilon_i} \sum_{j=1}^{k} |Y_{ij}^{(2)} - (-\epsilon_i)|.$$

Define the regression function estimator as

$$\hat{m}(x_i) = Y_i - \hat{\epsilon}_i, \tag{5}$$

due to $m(x_i) = Y_i - \epsilon_i$ (Wang et al., 2017; Wang and Yu, 2017). The same asymptotic results for $\hat{m}(x_i)$ in (5) can be established as in Theorem 2 according to $\hat{m}(x_i) - m(x_i) = -(\hat{\epsilon}_i - \epsilon_i)$.

2.3. Estimation methodology with random addition

In order to improve the estimation efficiency, we define the first-order random addition sequence as

$$Y_{ijl}^{(1)} = \frac{Y_{i+j} + Y_{i+l}}{2}, \quad -k \leq j < l \leq k, \tag{6}$$

somewhat like random difference for derivative estimation in Wang et al. (2019). For the first-order Taylor expansion, we simplify it to

$$Y_{ijl}^{(1)} = m(x_i) + m^{(1)}(x_i)x_{ijl} + \eta_{ijl} + o\left(\frac{j}{n} + \frac{l}{n}\right), \quad -k \leq j < l \leq k, \tag{7}$$

where $\eta_{ijl} = (\epsilon_{i+j} + \epsilon_{i+l})/2$ and $x_{ijl} = \frac{x_{i+j}+x_{i+l}-2x_i}{2} = \frac{x_j+x_l}{2}$. Since i is fixed as j and l change and $\mathrm{Median}(\eta_{ijl}) = 0$, then

$$\mathrm{Median}(Y_{ijl}^{(1)}) \approx m(x_i) + m^{(1)}(x_i)x_{ijl}.$$

Now we adapt the local LAD to estimate parameters

$$(\hat{\alpha}_{i0}, \hat{\alpha}_{i1})^T = \arg\min_{\alpha_{i0},\alpha_{i1}} \sum_{-k\leq j<l\leq k} |Y_{ijl}^{(1)} - \alpha_{i0} - \alpha_{i1}x_{ijl}|, \tag{8}$$

and define the estimator of $m(x_i)$ as

$$\hat{m}(x_i) = \hat{\alpha}_{i0}. \tag{9}$$

The asymptotic results are as follows.

Theorem 3. *Under the assumptions of Theorem 1, the bias and variance of the robust estimator in (9) are, respectively,*

$$\mathrm{Bias}[\hat{m}(x_i)] \triangleq \mathrm{E}[\hat{m}(x_i)] - m(x_i) \approx \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}, \quad \mathrm{Var}[\hat{m}(x_i)] \approx \frac{1}{6g(0)^2k}.$$

Remark 4. From Theorems 2 and 3, the bias of the estimation by random addition is the same as the estimation by symmetric addition; while the variance of the estimation by random addition reduces to 2/3 of the estimation by symmetric addition. Thus random addition is helpful to improve the efficiency. This idea in spirit is similar to the one-sample Hodges–Lehmann estimator of the center of a distribution (Hodges and Lehmann, 1956; Hershberger, 2011), which was defined as the median of the set of $n(n+1)/2$ Walsh averages, where each Walsh average is the arithmetic average of two observations.

3. Comparison to existing local linear estimators

For the overall comparison, we briefly review some popular nonparametric estimation methods. For simplicity, we adopt the uniform kernel for all local methods and then $h = k/n$.

Fan and Gijbels (1996) proposed the local linear LS regression to estimate the nonparametric function,

$$(\hat{\beta}_{i0}^{ls}, \hat{\beta}_{i1}^{ls})^T = \arg\min \sum_{j=-k}^{k} \left(Y_{i+j} - \beta_{i0} - \beta_{i1}x_j\right)^2.$$

Define the local linear LS estimator as

$$\hat{m}_{ls}(x_i) = \hat{\beta}_{i0}^{ls}. \tag{10}$$

Based on the technique of Wang and Lin (2015), we have the following corollary under equidistant design. When the errors are Gaussian, the local LS estimator is the most efficient corresponding to the local likelihood criterion (Fan et al., 1998).

Corollary 2. *Assume that ϵ_i are i.i.d. random errors with mean 0 and variance σ^2 for the nonparametric regression model (1). Then the bias and variance of the LS estimator in (10) are, respectively,*

$$\mathrm{Bias}[\hat{m}_{ls}(x_i)] = \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}, \quad \mathrm{Var}[\hat{m}_{ls}(x_i)] \approx \frac{\sigma^2}{2k}.$$

Wang and Scott (1994) proposed the local linear LAD method such that

$$(\hat{\beta}_{i0}^{lad}, \hat{\beta}_{i1}^{lad})^T = \arg\min \sum_{j=-k}^{k} \left|Y_{i+j} - \beta_{i0} - \beta_{i1}x_j\right|,$$

where the similar procedure was proposed by Fan and Hall (1994) and Welsh (1996). Define the LAD estimator as

$$\hat{m}_{lad}(x_i) = \hat{\beta}_{i0}^{lad}. \tag{11}$$

The asymptotic properties are given as follows.

Corollary 3. *Assume that ϵ_i are i.i.d. random errors with median 0 and a symmetric density function $f(\cdot)$ for the nonparametric regression model (1). Then the bias and variance of the LAD estimator in (11) are, respectively,*

$$\text{Bias}[\hat{m}_{lad}(x_i)] = \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}, \quad \text{Var}[\hat{m}_{lad}(x_i)] \approx \frac{1}{8kf(0)^2}.$$

Quantile regression technique (Koenker and Bassett, 1978; Koenker, 2005) exploits the distribution information to improve the estimation efficiency. Based on the CQR (Zou and Yuan, 2008), Kai et al. (2010) proposed the local linear CQR such that

$$(\{\hat{\beta}_{i0h}\}_{h=1}^{q}, \hat{\beta}_{i1})^T = \arg\min \sum_{h=1}^{q}\left(\sum_{j=-k}^{k}\rho_{\tau_h}(y_i - \beta_{i0h} - \beta_{i1}x_{i+j})\right),$$

where $\tau_h = h/(q+1)$, and $\rho_\tau(x) = \tau x - xI(x<0)$ is the check loss function. Define the CQR estimator as

$$\hat{m}_{cqr}(x_i) = \frac{1}{q}\sum_{h=1}^{q}\hat{\beta}_{i0h}. \tag{12}$$

Based on Theorem 1 of Kai et al. (2010), we have the following corollary.

Corollary 4. *Assume that ϵ_i are i.i.d. random errors with mean 0, a symmetric density function $f(\cdot)$ and the cumulative distribution function $F(\cdot)$ for the nonparametric regression model (1). Then the bias and variance of the CQR estimator in (12) are, respectively,*

$$\text{Bias}[\hat{m}_{cqr}(x_i)] = \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}, \quad \text{Var}[\hat{m}_{cqr}(x_i)] \approx \frac{R_1(q)}{2k},$$

where $R_1(q) = \frac{1}{q^2}\sum_{l=1}^{q}\sum_{l'=1}^{q}\frac{\tau_{ll'}}{f(c_l)f(c_{l'})}$, $c_l = F^{-1}(\tau_l)$, and $\tau_{ll'} = \min\{\tau_l, \tau_{l'}\} - \tau_l\tau_{l'}$. As $q \to \infty$, we have that

$$R_1(q) \to \sigma^2,$$

assuming the error variance σ^2 exists.

Corollary 4 implies the local linear CQR estimator has the same asymptotic behavior with the local linear LS estimator as the quantile number q becomes larger. For the CQR estimator with fixed q, its behavior depends on finite density function values from the expression of $R_1(q)$.

Zhao and Xiao (2014) proposed the WQA estimator to further improve the estimation efficiency. Firstly, they use the local linear quantile regression (Yu and Jones, 1998) to estimate the τ_h quantile of $m(x_i)$

$$(\hat{\beta}_{i0h}^{wqa}, \hat{\beta}_{i1h}^{wqa}) \triangleq \arg\min \sum_{j=-k}^{k}\rho_{\tau_h}(y_i - \beta_{i0h} - \beta_{i1h}x_{i+j}),$$

and then define the WQA estimator as

$$\hat{m}_{wqa}(x_i) = \sum_{h=1}^{q} w_h\hat{\beta}_{i0h}^{wqa}, \tag{13}$$

where $w_h = \frac{H^{-1}e_q}{e_q^T H^{-1}e_q}$ is the optimal weights, $H = \left\{\frac{\min\{\tau_l,\tau_{l'}\}-\tau_l\tau_{l'}}{f(c_l)f(c_{l'})}\right\}$ depends on the density function $f(\cdot)$, and $e_q = (1,\ldots,1)_{q\times 1}$. The asymptotic properties are as follows.

Corollary 5. *Assume that ϵ_i are i.i.d. random errors with mean 0, a symmetric density function $f(\cdot)$ for the nonparametric regression model (1). Then the bias and variance of the WQA estimator in (13) are, respectively,*

$$\text{Bias}[\hat{m}_{wqa}(x_i)] = \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}, \quad \text{Var}[\hat{m}_{wqa}(x_i)] \approx \frac{R_2(q)}{2k},$$

where $R_2(q) = w_q^T H w_q = (e_q^T H^{-1} e_q)^{-1}$. As $q \to \infty$, we have that

$$R_2(q) \to I(f_\epsilon)^{-1},$$

where $I(f_\epsilon)$ is the Fisher information of f_ϵ.

For the LS, LAD, CQR, WQA and new estimators, the asymptotic biases are all the same; while the asymptotic variances are $\frac{\sigma^2}{2k}$, $\frac{1}{8kf(0)^2}$, $\frac{R_1(q)}{2k}$, $\frac{R_2(q)}{2k}$, and $\frac{1}{4\ kg(0)^2}$, respectively. In fact, the comparative components are σ^2, $\frac{1}{4f(0)^2}$, $R_1(q)$, $R_2(q)$, and $\frac{1}{2g(0)^2}$.

As $q \to \infty$, we have $R_1(q) \to \sigma^2$ and $R_2(q) \to I(f_\epsilon)^{-1}$. The CQR estimator is asymptotically equal to the LS estimator; the WQA estimator is the most efficient, just like the local maximum likelihood estimator in Fan et al. (1998). However, the WQA needs to know the error density function in advance or to be estimated accurately; otherwise, it is invalid.

When q is fixed, the CQR and WQA estimators depend on finite values of unknown density function f_ϵ. Thus the estimation efficiency is uncertain. For the LS, LAD and new estimators, their asymptotic efficiencies do not depend on q quantiles but keep constant. In the presence of platykurtic and heavy-tailed errors, the new estimator will have high efficiency. In addition, the new estimator relies on $g(0) = 2E[f(x)]$, which includes all information of the density $f(\cdot)$. The term $E[f(x)]$ appears in Theorem 3.1 of Zou and Yuan (2008) for the regression coefficients estimation except the constant term. For the nonparametric regression function estimation in Kai et al. (2010) as well as Zhao and Xiao (2014), there is no similar result.

4. Bias-reduced estimators

In order to improve the finite-sample performance for high-oscillate nonparametric functions, we further consider the higher-order Taylor expansion to reduce the estimation bias. Assume that $m(\cdot)$ is four times continuously differentiable and the true model is

$$Y_{ij}^{(2)} = \beta_{i40} + \beta_{i42}d_j^2 + \beta_{i44}d_j^4 + \eta_{ij}, \tag{14}$$

where $\beta_{i4} = (\beta_{i40}, \beta_{i42}, \beta_{i44})^T = (-\epsilon_i, m^{(2)}(x_i)/2, m^{(4)}(x_i)/24)^T$ are the unknown coefficients. Now we use the local LAD quadratic regression to estimate the error, that is

$$\hat{b}_{i2} \triangleq \arg\min_{b_{i2}} \sum_{j=1}^{k} |Y_{ij}^{(2)} - D_{2j}^T b_{i2}|,$$

where $D_{2j} = (1, d_j^2)^T$, and $b_{i2} = (b_{i20}, b_{i22})^T$.

Define the estimators of ϵ_i and $m(x_i)$, respectively, as

$$\hat{\epsilon}_i = -\hat{b}_{i20}, \quad \hat{m}(x_i) = Y_i + \hat{b}_{i20}. \tag{15}$$

The robust error estimator $\hat{\epsilon}_i$ will converge to $-\beta_{i30}$, where

$$\beta_{i3} \triangleq (\beta_{i30}, \beta_{i32})^T \triangleq \arg\min_{c_{i20}, c_{i22}} \sum_{j=1}^{k} |(c_{i20} + c_{i22}d_j^2) - (\beta_{i40} + \beta_{i42}d_j^2 + \beta_{i42}d_j^4)|.$$

Next we establish the asymptotic normality of $\hat{\epsilon}_i$ in (15), and derive the asymptotic properties of $\hat{m}(x_i)$.

Corollary 6. *Assume that ϵ_i are i.i.d. random errors with median 0 and a continuous, symmetric density $f(\cdot)$. For the equidistant design and the true model (14), as $k \to \infty$ and $k/n \to 0$, the bias and variance of the robust estimator in (15) are, respectively,*

$$\text{Bias}[\hat{m}(x_i)] \approx \frac{m^{(4)}(x_i)}{280}\frac{k^4}{n^4}, \quad \text{Var}[\hat{m}(x_i)] \approx \frac{9}{4}\frac{1}{4g(0)^2 k}.$$

The optimal k that minimizes the asymptotic mean square error (AMSE) is

$$k_{opt} \approx 2.60\left(\frac{1}{g(0)^2 m^{(4)}(x_i)^2}\right)^{1/9} n^{8/9},$$

and consequently, the minimum AMSE is

$$\text{AMSE}[\hat{m}(x_i)] \approx 0.21\left(\frac{m^{(4)}(x_i)^2}{g(0)^{16}}\right)^{1/9} n^{-8/9}.$$

In the nonparametric regression model (1), the robust estimator $\hat{m}(x_i)$ in (15) is normal distributed.

Further assume that $m(\cdot)$ is six times continuously differentiable, and the true model is

$$Y_{ij}^{(2)} = \beta_{i60} + \beta_{i62}d_j^2 + \beta_{i64}d_j^4 + \beta_{i66}d_j^6 + \eta_{ij},$$

where $\beta_{i6} = (\beta_{i60}, \beta_{i62}, \beta_{i64})^T = (-\epsilon_i, m^{(2)}(x_i)/2, m^{(4)}(x_i)/24, m^{(6)}(x_i)/720)^T$. Now we use the local LAD quartic regression

$$\hat{b}_{i4} \triangleq \arg\min_{b_{i4}} \sum_{j=1}^{k} |Y_{ij}^{(2)} - D_{4j}^T b_{i4}|,$$

Table 1
The bias and variance for three new estimators.

	1	3	5
Bias	$\frac{m^{(2)}}{6}\frac{k^2}{n^2}$	$\frac{m^{(4)}}{280}\frac{k^4}{n^4}$	$\frac{m^{(6)}}{33264}\frac{k^6}{n^6}$
Variance	$\frac{1}{4g(0)^2k}$	$\frac{9}{4}\frac{1}{4g(0)^2k}$	$\frac{255}{64}\frac{1}{4g(0)^2k}$

where $D_{4j} = (1, d_j^2, d_j^4)^T$, and $b_{i4} = (b_{i40}, b_{i42}, b_{i44})^T$. Define the estimator of $m(x_i)$ as

$$\hat{m}(x_i) = Y_i + \hat{b}_{i40}. \tag{16}$$

The asymptotic normality is valid by a similar discussion above. Next we only give the asymptotic bias and variance.

Corollary 7. *The bias and variance of the robust estimator in (16) are, respectively,*

$$\text{Bias}[\hat{m}(x_i)] \approx \frac{m^{(6)}(x_i)}{33264}\frac{k^6}{n^6}, \quad \text{Var}[\hat{m}(x_i)] \approx \frac{225}{64}\frac{1}{4g(0)^2k}.$$

For convenience, we summary the bias and variance for three estimators in Table 1.

5. Simulation studies and data analysis

5.1. Simulation studies

In this subsection, we conduct some simulations to compare the performance of the following estimators: LS, LAD (Wang and Scott, 1994), Huber (based on Huber loss function), Tukey (based on Tukey's bisquare loss function) and our NEW-S (based on symmetric difference) and NEW-R (based on random difference). We specify the regression function as the periodic function

$$m(x) = \sin(2\pi x). \tag{17}$$

The error distributions are as follows: the normal distribution $E_1 = N(0, 1^2)$, the mixed normal distributions $E_2 = 0.9N(0, 1^2) + 0.1N(0, 10^2)$ and $E_3 = 0.5N(-3, 1^2) + 0.5N(3, 1^2)$, the mixed Laplace distribution $E_4 = 0.5Laplace(-2, 1) + 0.5Laplace(2, 1)$, and the asymmetric distribution $E_5 = \chi^2(4) - 4$ with mean zero. As for the sample size, we consider $n = 50, 200$. To compare the performance of these estimators, we adopt the adjusted mean absolute error (AMAE) as

$$AMAE(k) = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} |\hat{m}(x_i) - m(x_i)|,$$

with fixed $k = n/10$.

Table 2 reports the results of simulations based on 1000 repetitions, in which the numbers outside and inside the brackets represent the mean and standard deviation of the AMAEs, respectively. In summary, our proposed NEW-R is robust and efficient for all errors; the NEW-R has a better performance uniformly than the NEW-S, which coincides with our theoretical results. For the error E_1, the LS has a little better performance than the others, and the others have the similar performance except LAD. For the heavy-tailed error E_2, the LS estimator performs worse than the others, and the Tukey is better. For the platykurtic errors E_3, the NEW-R is the best, and the LAD breaks down. For the infinite-variance error E_4, the Tukey is the best, and the LS breaks down. For asymmetric error E_5 with mean zero, the LS is best due to its unbiasedness, and the others perform a litter worse due to their systematic deviations.

To show the importance of bias-correction, we specify that the regression function is $m(x) = 10\sin(2\pi x)$, which corresponds to the high-oscillation function. The error is $0.8N(0, 0.1^2) + 0.2N(0, 1^2)$ in this setting. Fig. 2 shows that the new bias-reduced estimator has the better performance, where $NEW1$, $NEW2$ and $NEW3$ are the new estimators with Taylor expansion of first-order, third-order and fifth-order, respectively. To choose p and k simultaneously, we provide the following criterion for real data

$$\arg\min_{p,k} \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{m}(x_i; p, k)|.$$

5.2. House price of China in latest ten years

In reality, there are many data sets recorded by equidistance and equitime. For example, image data is recorded by cell-grid, and temperature is recorded by hour, day or month. In this section, we apply our new method to the data set of house price in the capital of China, Beijing. The monthly data is borrowed from Wang et al. (2019), which last from January 2008 to July 2018 and have size 127.

Table 2
AMAE for the existing estimators with $n = 50, 200$.

n	Error	LS	NEW − S	NEW − R
50	E_1	0.237 (0.223)	0.302 (0.300)	0.248 (0.239)
	E_2	0.744 (0.326)	0.399 (0.115)	0.348 (0.127)
	E_3	0.770 (0.216)	0.897 (0.236)	0.805 (0.274)
	E_4	3.380 (5.250)	1.035 (0.235)	0.890 (0.271)
	E_5	0.664 (0.204)	0.863 (0.170)	0.713 (0.197)
200	E_1	0.131 (0.038)	0.162 (0.031)	0.135 (0.037)
	E_2	0.404 (0.141)	0.190 (0.036)	0.159 (0.041)
	E_3	0.408 (0.147)	0.382 (0.129)	0.326 (0.144)
	E_4	4.282 (15.22)	0.518 (0.113)	0.428 (0.113)
	E_5	0.367 (0.112)	0.523 (0.129)	0.466 (0.148)
n	Error	LAD	Huber	Tukey
50	E_1	0.305 (0.290)	0.247 (0.239)	0.253 (0.242)
	E_2	0.351 (0.096)	0.336 (0.119)	0.309 (0.100)
	E_3	1.566 (0.294)	0.867 (0.270)	0.931 (0.297)
	E_4	1.076 (0.227)	0.882 (0.260)	0.849 (0.234)
	E_5	0.929 (0.214)	0.713 (0.200)	0.786 (0.229)
200	E_1	0.163 (0.043)	0.135 (0.037)	0.136 (0.037)
	E_2	0.175 (0.044)	0.156 (0.041)	0.142 (0.042)
	E_3	1.230 (0.262)	0.416 (0.156)	0.447 (0.167)
	E_4	0.683 (0.159)	0.426 (0.129)	0.408 (0.119)
	E_5	0.691 (0.195)	0.461 (0.145)	0.522 (0.162)

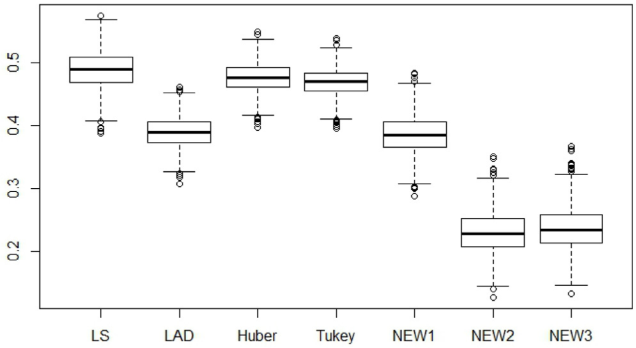

Fig. 2. The comparison among the existing estimators.

With the rapid development of China's economy, people's demand for housing quality is becoming higher and higher. As a result, the house prices also grew rapidly. Meanwhile, the Chinese government has promulgated a number of housing policies to keep the stability of house prices in latest ten years. The house price data has large oscillation amplitude locally, and our new method is suitable for the house price data. We apply our method to estimate the regression function in Fig. 3. In the last ten years, the house price goes through tricycle fast increasing, and the increasing rate of the latter is faster than that of the former.

6. Conclusion and discussion

In the presence of the platykurtic and heavy-tailed errors, we propose an addition-sequence method for the robust estimation of regression function. The new method consists of three main steps: construct a sequence of symmetric or random addition, estimate the errors using the LAD regression, and obtain the robust function estimation. Under different smoothness conditions, we proposed three estimators for improving the finite-sample performance.

In this paper, we focus on the nonparametric function estimation with equidistant designs. The method can be extended to random designs with minor changes, which is similar to fixed design except the kernel estimation method. Assume the X_i is a random variable with density function $f(\cdot)$. We focus on the estimation of $m(x_0)$ at the design point x_0. Firstly, we define the first-order random addition sequence as

$$Y_{jl}^{(1)} = \frac{Y_j + Y_l}{2}, \quad \{j, l : |X_j - x_0| \leq h, |X_l - x_0| \leq h\},$$

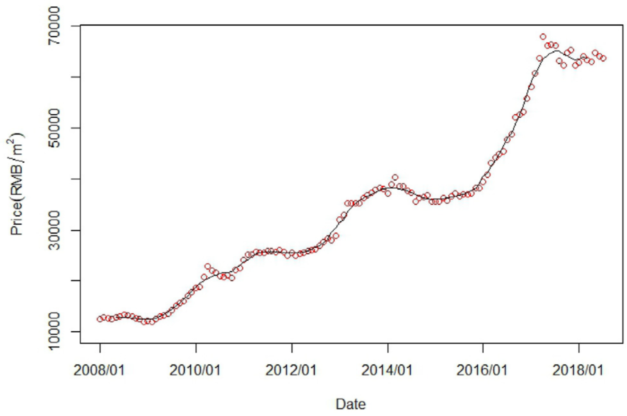

Fig. 3. The robust estimator for the house price of Beijing.

where $h = o(1)$ is a bandwidth. Secondly, assume that $m(\cdot)$ is one-time continuously differentiable. Then the first-order Taylor expansions of $m(X_j)$ and $m(X_l)$ around x_0 are

$$m(X_j) = m(x_0) + m^{(1)}(x_0)(X_j - x_0) + o(X_j - x_0),$$
$$m(X_l) = m(x_0) + m^{(1)}(x_0)(X_l - x_0) + o(X_l - x_0).$$

Thus, in the neighborhood of x_0 with radius h, we have

$$Y_{jl}^{(1)} = m(x_0) + m'(x_0)x_{jl} + \eta_{jl} + o\left(x_{jl}\right),$$

where $\eta_{jl} = (\epsilon_j + \epsilon_l)/2$ with $\text{Median}(\eta_{jl}) = 0$, and $x_{jl} = \frac{X_j+X_l-2x_0}{2}$ with $|x_{jl}| \le h$, and

$$\text{Median}(Y_{jl}^{(1)}) \approx m(x_0) + m'(x_0)x_{jl}.$$

Thirdly, we adopt the locally weighted least absolute deviation regression to estimate parameters

$$(\hat{\alpha}_{i0}, \hat{\alpha}_{i1})^T = \arg\min_{\alpha_{i0},\alpha_{i1}} \sum_{1\le j<l\le n} |Y_{jl}^{(1)} - \alpha_{i0} - \alpha_{i1}x_{jl}| K_h(X_j - x_0)K_h(X_l - x_0),$$

where $K_h(\cdot) = K(\cdot/h)$ is a kernel function with the bandwidth h. Define the estimator of $m(x_0)$ as

$$\hat{m}(x_0) = \hat{\alpha}_{i0}.$$

It is interesting to consider the robustness properties in nonparametric models: the breakdown point (Hampel, 1968; Huber and Ronchetti, 2009) and the influence function (Hampel, 1974; Ichimura and Newey, 2017).

Acknowledgment

We would like to thank two anonymous reviewers and associate editor for their constructive comments on improving the quality of the paper. Wang's work is support by the fund XKJJC201901 of Qufu Normal University, China.

Appendix A. Lemmas

Proposition 1. *Assume that the errors ϵ_i are i.i.d. with median 0 and a continuous density function $f(\cdot)$, then $\eta_{ij} = (\epsilon_{i-j} + \epsilon_{i+j})/2$ ($j = 1, \ldots, k$) are i.i.d. with median 0 and a continuous density $g(\cdot)$, where*

$$g(x) = 2\int_{-\infty}^{\infty} f(2x - \epsilon)f(\epsilon)d\epsilon = \text{E}[f(2x - \epsilon)].$$

Further assume that $f(\cdot)$ is symmetric, then

$$g(0) = 2\int_{-\infty}^{\infty} f^2(\epsilon)d\epsilon = 2\text{E}[f(\epsilon)].$$

Proof. The distribution of $\eta_{ij} = (\epsilon_{i-j} + \epsilon_{i+j})/2$ is

$$F_{\eta_{ij}}(x) = P((\epsilon_{i-j} + \epsilon_{i+j})/2 \le x)$$

$$
\begin{aligned}
&= \iint_{\epsilon_{i+j} \le 2x - \epsilon_{i-j}} f(\epsilon_{i+j}) f(\epsilon_{i-j}) \mathrm{d}\epsilon_{i+j} \mathrm{d}\epsilon_{i-j} \\
&= \int_{-\infty}^{\infty} \{ \int_{-\infty}^{2x-\epsilon_{i-j}} f(\epsilon_{i+j}) \mathrm{d}\epsilon_{i+j} \} f(\epsilon_{i-j}) \mathrm{d}\epsilon_{i-j} \\
&= \int_{-\infty}^{\infty} F_\epsilon(2x - \epsilon_{i-j}) f(\epsilon_{i-j}) \mathrm{d}\epsilon_{i-j}.
\end{aligned}
$$

Then the density of η_{ij} is

$$
g(x) \triangleq \frac{\mathrm{d}F_{\tilde{\delta}_{ij}}(x)}{\mathrm{d}x} = 2 \int_{-\infty}^{\infty} f(2x - \epsilon_{i-j}) f(\epsilon_{i-j}) \mathrm{d}\epsilon_{i-j}.
$$

By the symmetry of the density function, we have

$$
\begin{aligned}
F_{\eta_{ij}}(0) &= \int_{-\infty}^{\infty} F_\epsilon(-\epsilon_{i-j}) f(\epsilon_{i-j}) \mathrm{d}\epsilon_{i-j} \\
&= (F_\epsilon - \frac{1}{2} F_\epsilon^2(\epsilon_{i-j})) \mid_{-\infty}^{\infty} \\
&= \frac{1}{2}, \\
g(0) &= 2 \int_{-\infty}^{\infty} f^2(\epsilon_{i-j}) \mathrm{d}\epsilon_{i-j}. \quad \square
\end{aligned}
$$

Lemma 1. *Suppose that the errors ϵ_i are iid with median 0 and a continuous positive density $f(\cdot)$ in a neighborhood of zero. Define the function*

$$
M(s, t) = \mathrm{E}[\left|\epsilon_i + s - t\right| - \left|\epsilon_i + s\right|].
$$

Then M has a unique minimum at point $(0, 0)$, and furthermore in a neighborhood of $(0, 0)$,

$$
M(s, t) = f(0)t^2 - 2f(0)st + o(s^2 + t^2).
$$

Proof. See Wang and Scott (1994). $\square$

Lemma 2 (*Convexity Lemma*)**.** *Let $\{\lambda_i(\theta) : \theta \in \Theta\}$ be a sequence of random convex functions defined on a convex open subset Θ of R^d. Suppose that λ is a real-valued function on Θ for which $\lambda_i(\theta) \xrightarrow{p} \lambda(\theta)$, for each θ in Θ. Then for every compact subset K of Θ,*

$$
\sup_{\theta \in K} \left|\lambda_i(\theta) - \lambda(\theta)\right| \xrightarrow{p} 0.
$$

The function λ is necessarily convex on Θ.

Proof. See Pollard (1991). $\square$

Appendix B. Proof of Theorem 1

Following the techniques by Pollard (1991) and Wang and Scott (1994), we have an analogous proof. The details are as follows. Write $z_j = (k)^{-1/2}$, then $\sum_{j=1}^k z_j^2 = 1$ and $\max |z_j| \to 0$. Define $u_{ij} = \beta_{i20} - \beta_{i1} + \beta_{i22} d_j^2$ and

$$
G_k(\theta_i) = \sum_{j=1}^{k} (|\eta_{ij} + u_{ij} - z_j\theta_i| - |\eta_{ij} + u_{ij}|)
$$

Due to Lemmas 1 and 2, we have

$$
\hat{\theta}_i = (k)^{1/2}(\hat{b}_{i0} - b_{i0}) = \arg\min G_k(\theta_i),
$$

where $\hat{b}_{i0} = \arg\min \sum_{j=1}^k |Y_{ij}^{(2)} - b_{i0}|$.

Define

$$\begin{aligned}\Gamma(\theta_i) &= \mathrm{E}[G_k(\theta_i)] \\ &= \sum_{j=1}^{k} M(z_j\theta_i, u_{ij}) \\ &= g(0)\theta_i^2 - 2g(0)k^{-1/2}\theta_i \sum_{j=1}^{k} u_{ij}.\end{aligned}$$

Let $D_j = \chi[\eta_{ij} + u_{ij} < 0] - \chi[\eta_{ij} + u_{ij} > 0]$, where χ denotes the indicator function. Then define

$$R_{ij}(\theta_i) = |\eta_{ij} + u_{ij} - z_j\theta_i| - |\eta_{ij} + u_{ij}| - (D_j - \mathrm{E}[D_j])z_j\theta_i,$$

and $W = \sum_{j=1}^{k}(D_j - \mathrm{E}[D_j])z_j$. Due to $\mathrm{E}[W] = 0$, we finally write

$$G_k(\theta_i) = \Gamma(\theta_i) + W\theta_i + \sum_{j=1}^{k}(R_j(\theta_i) - \mathrm{E}[R_j(\theta_i)])z_j.$$

For the fixed θ_i, the sum of centered terms $\zeta_{ij} = R_{ij}(\theta_i) - \mathrm{E}[R_{ij}(\theta_i)]$ contributes only an $o_p(1)$ to G_k. Define

$$\tilde{R}_{ij}(\theta_i) = (|\eta_{ij} + u_{ij} - z_j\theta_i| - |\eta_{ij} + u_{ij}| - D_jz_j\theta_i) - \mathrm{E}[|\eta_{ij} + u_{ij} - z_j\theta_i| - |\eta_{ij} + u_{ij}| - D_jz_j\theta_i],$$

and then $\zeta_{ij} = \tilde{R}_{ij}(\theta_i) - \mathrm{E}[\tilde{R}_{ij}(\theta_i)]$. Due to the Cauchy–Schwarz inequality and the fact that $\max|z_j| \to 0$, we have

$$\begin{aligned}\mathrm{E}[|\sum_{j=1}^{k}\tilde{R}_{ij}^2(\theta_i)|] &= \sum_{j=1}^{k}\mathrm{Var}[\tilde{R}_{ij}(\theta_i)] \\ &\leq 4\sum_{j=1}^{k}\theta_i^2\mathrm{E}\{\chi[|\eta_{ij} + u_{ij}| \leq |z_j\theta_i|]\} \\ &\leq 4\sum_{j=1}^{k}\theta_i^2 U(|z_j\theta_i|) \\ &\leq 4\sum_{j=1}^{k}\theta_i^2 U(|\theta_i|\max|z_j|) \\ &\to 0.\end{aligned}$$

where $U(t) \to 0$ as $t \to 0$.

Rewrite

$$G_k(\theta_i) = g(0)\theta_i^2 - 2g(0)k^{-1/2}\theta_i\sum_{j=1}^{k}u_{ij} + o(1) + W\theta_i + o_p(1).$$

Define

$$\begin{aligned}\lambda_k(\theta_i) &= G_k(\theta_i) + 2g(0)k^{-1/2}\theta_i - W\theta_i, \\ \lambda(\theta_i) &= g(0)\theta_i^2,\end{aligned}$$

we establish the uniform convergence on a compact set due to Convexity Lemma. Finally define the approximating function as

$$\phi(\theta_i) = g(0)\theta_i^2 - 2g(0)k^{-1/2}\theta_i\sum_{j=1}^{k}u_{ij}, \sum_{j=1}^{k}u_{ij} + W\theta_i,$$

and find the minimizer $\tilde{\theta}_i = -W/2g(0) + k^{-1/2}\sum_{j=1}^{k}u_{ij}$ by taking the derivative and setting it equal to 0

$$\frac{\partial\phi(\theta_i)}{\partial\theta_i} = 0.$$

The central limit theorem ensures that W has asymptotically the normal distribution.

Rewrite W in terms of the minimizer $\tilde{\theta}_i$ as

$$W = -2g(0)\theta_i + 2g(0)k^{-1/2}\sum_{j=1}^{k}u_{ij},$$

and thus

$$G_k(\theta_i) = g(0)(\theta_i - \tilde{\theta}_i)^2 - g(0)\tilde{\theta}_i^2 + r(\theta_i),$$

where for each compact set K, $\sup_K |r(\theta_i)| \xrightarrow{p} 0$. We verify that $\tilde{\theta}_i$ lies close enough to $\tilde{\theta}_i$ to be asymptotically normal, thus we show that

$$\tilde{\theta}_i - \hat{\theta}_i \to 0.$$

Now we prove that for every δ, $P(|\tilde{\theta}_i - \hat{\theta}_i|) \to 0$. On the close ball $B(\tilde{\theta}_i, \delta)$ with center $\tilde{\theta}_i$ and radius δ, Because $\tilde{\theta}_i$ converges in distribution, it is stochastically bounded. The compact set K can be chosen to contain $B(\tilde{\theta}_i, \delta)$ with probability arbitrarily close to 1, which implies

$$\Delta(\theta_i) = \sup_{B(\tilde{\theta}_i,\delta)} |r(\theta_i)| \xrightarrow{p} 0.$$

Appendix C. Proof of Theorem 2

Define $\phi_{ij} = \beta_{i10} - (\beta_{i00} + \beta_{i02}d_j^2)$. Using the similar discussion in Wang and Scott (1994), we have

$$\mathrm{E}[\hat{b}_{i0} \mid \epsilon_i] - k^{-1} \sum_{1\le j\le k} \phi_{ij} \xrightarrow{p} \beta_{i10},$$

which simplifies to

$$\mathrm{E}[\hat{b}_{i0} \mid \epsilon_i] - \beta_{i00} \xrightarrow{p} k^{-1} \sum_{1\le j\le k} \beta_{i02}d_j^2.$$

Thus the bias of $\hat{m}(x_i)$ is

$$\mathrm{Bias}[\hat{m}(x_i)] = \mathrm{E}[\hat{m}(x_i)] - m(x_i) = E[\hat{b}_{i0} \mid \epsilon_i] - \beta_{i00} \approx \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2}.$$

From Theorem 1, we have the variance of $\hat{m}(x_i)$ is

$$\mathrm{Var}[\hat{m}(x_i)] = \mathrm{Var}[\hat{b}_{i0} \mid \epsilon_i] \approx \frac{1}{4g(0)^2k}.$$

Appendix D. Proof of Theorem 3

Following the techniques by Wang et al. (2019), we have an analogous proof. Rewrite the objective function as a U-process

$$S_n(\alpha) = \sum_{-k\le j<l\le k} f_n(Y_{i+j}, Y_{i+l}|\alpha),$$

where

$$f_n(Y_{i+j}, Y_{i+l}|\alpha) = \left|\frac{Y_{i+j} + Y_{i+l}}{2} - \alpha_{i0} - \alpha_{i1}\frac{x_j + x_l}{2}\right|\frac{1}{h^2}1(|x_j| \le h)1(|x_l| \le h),$$

with $\alpha = (\alpha_{i0}, \alpha_{i1})^T$ and $h = k/n$. Define $U_n(\alpha) = \frac{2}{n(n-1)}S_n(\alpha)$, $H = diag\{1, h\}$, and $X_{jl} = (1, \frac{x_j+x_l}{2})^T$. Note that

$$\arg\min S_n(\alpha) = \arg\min U_n(\alpha) = \arg\min[U_n(\alpha) - U_n(\beta)],$$

where $\beta = (m(x_i), m'(x_i))^T$.

We first show that $H(\hat{\alpha} - \beta) = o_p(1)$. We use Lemma 4 of Porter and Yu (2015) to show the consistency. Essentially, we need to show that

(i) $\sup_{\alpha\in\mathcal{B}} |U_n(\alpha) - U_n(\beta) - \mathrm{E}[U_n(\alpha) - U_n(\beta)]| \xrightarrow{p} 0$, where $\mathcal{B}$ is a compact parameter space for α;
(ii) $\inf_{\|H(\hat{\alpha}-\beta)\|>\delta} \mathrm{E}[U_n(\alpha) - U_n(\beta)] > \epsilon$ for n large enough, where δ and ϵ are fixed positive small numbers.

We use Theorem $A.2$ of Ghosal et al. (2000) to show (i), where

$$\mathcal{F}_n = \{f_n(Y_{i+j}, Y_{i+l}|\alpha) - f_n(Y_{i+j}, Y_{i+l}|\beta)|\alpha \in \mathcal{B}\}.$$

Note that $\mathcal{F}_n$ forms an Euclidean-class of functions by applying Lemma 2.13 of Pakes and Pollard (1989), where $\alpha = 1$, $f(\cdot, t_0) = 0$, $\phi(\cdot) = \|X_{jl}\|\frac{1}{h^2}1(|x_j| \le h)1(|x_l| \le h)$ and the envelope function is $F_n(\cdot) = M\phi(\cdot)$ for some finite constant M. It follows that

$$N(\epsilon\|F_n\|_{Q,2}, \mathcal{F}, L_2(Q)) \lesssim \epsilon^{-C}$$

for any probability measure Q and some positive constant C, where $\lesssim$ means the left side is bounded by a constant times the right side. Hence,

$$\frac{1}{n}\mathrm{E}[\int_0^\infty \log N(\epsilon\|F_n\|_{Q,2}, \mathcal{F}, L_2(U_2^n))] \lesssim \frac{1}{n}(\mathrm{E}[F_n^2])^{1/2}\int_0^\infty \log\frac{1}{\epsilon}\mathrm{d}\epsilon = O(\frac{1}{n}),$$

where U_2^n is the random discrete measure putting mass $\frac{1}{n(n-1)}$ on each of the points (Y_{i+j}, Y_{i+l}). Next, by Lemma $A.2$ of Ghosal et al. (2000), the projections

$$\bar{f}_n(Y_{i+j}|\alpha) = \int f_n(Y_{i+j}, Y_{i+l}|\alpha)\mathrm{d}F_{Y_{i+l}}$$

satisfy

$$\sup_Q N(\epsilon\|\bar{F}_n\|_{Q,2}, \bar{\mathcal{F}}_n, L_2(Q)) \lesssim \epsilon^{-2C},$$

where $\bar{\mathcal{F}}_n = \{\bar{f}_n(Y_{i+j}|\alpha) - \bar{f}_n(Y_{i+j}|\beta)|\alpha \in \mathcal{B}\}$, and $\bar{F}_n$ is an envelope of $\bar{\mathcal{F}}_n$. Thus,

$$n^{-1/2}\mathrm{E}[\int_0^\infty \log N(\epsilon, \bar{\mathcal{F}}, L_2(P_n))]\mathrm{d}\epsilon \lesssim n^{-1/2}(\mathrm{E}[\bar{F}_n^2])^{1/2}\int_0^\infty \log\frac{1}{\epsilon}\mathrm{d}\epsilon = O(\frac{1}{n^{1/2}}).$$

By Theorem $A.2$ and Markov's inequality, condition (i) is satisfied.

As for condition (ii), by Proposition 1 of Wang and Scott (1994),

$$\begin{aligned}
&\mathrm{E}[U_n(\alpha) - U_n(\beta)]\\
\approx&\frac{2}{n(n-1)}\sum_{j<l}\frac{g(0)}{2}[X_{jl}^T H^{-1}H(\alpha-\beta)]^2\frac{1}{h^2}1(|x_j| \le h)1(|x_l| \le h)\\
-&\frac{2}{n(n-1)}\sum_{j<l}g(0)[m(x_{i+j}) - m(x_{i+l}) - X_{jl}^T\beta][X_{jl}^T H^{-1}H(\alpha-\beta)]\\
&\frac{1}{h^2}1(|x_j| \le h)1(|x_l| \le h)\\
\gtrsim&\delta^2 - h^5\delta.
\end{aligned}$$

Next, we derive the asymptotic distribution of $\sqrt{nh}H(\hat{\alpha}-\beta)$. Firstly, we approximate the first-order conditions by Theorem $A.1$ of Ghosal et al. (2000). Secondly, we derive the asymptotic distribution of $\sqrt{nh}H(\hat{\alpha}-\beta)$ by the empirical process.

Firstly, the first-order conditions can be written as

$$\frac{2}{n(n-1)}\sum_{j<l} sign(Y_{i+j} + Y_{i+l} - Z_{jl}^T H\hat{\alpha})Z_{jl}\frac{\sqrt{h}}{h^2}1(|x_j| \le h)1(|x_l| \le h) = o_p(1),$$

where $Z_{jl} = H^{-1}X_{jl}$. By Example 2.9 of Pakes and Pollard (1989), $\mathcal{F}_n'$ forms an Euclidean-class of functions with envelope $F_n' = \|Z_{jl}\|\frac{\sqrt{h}}{h^2}1(|x_j| \le h)1(|x_l| \le h)$, where

$$\mathcal{F}_n' = \{f_n'(Y_{i+j}, Y_{i+l}|\alpha) : \alpha \in \mathcal{B}\},$$

so

$$N(\epsilon\|F_n'\|_{Q,2}, \mathcal{F}_n', L_2(Q)) \lesssim \epsilon^{-V}$$

for any probability measure Q and some positive constant V. By Theorems $A.1$ and $A.2$ of Ghosal et al. (2000), it follows that

$$\begin{aligned}
&n\mathrm{E}\{\sup_{f_n'\in\mathcal{F}_n'}|U_2^n f_n' - 2P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\alpha)]) - \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\alpha)]|\}\\
\lesssim&\mathrm{E}[\int_0^\infty \log N(\epsilon, \mathcal{F}_n', L_2(U_2^n))\mathrm{d}\epsilon] \lesssim \int_0^1 \log(\epsilon^{-V})\mathrm{d}\epsilon\sqrt{\mathrm{E}[(F_n')^2]} \lesssim h^{-1/2},
\end{aligned}$$

where P_n is the empirical measure of the original data, $\bar{\mathrm{E}}_2[\cdot]$ takes expectation on Y_{i+l} and also averages over d_l, and $\bar{\mathrm{E}}[\cdot]$ takes expectation on (Y_{i+j}, Y_{i+l}) and also averages over (X_j, X_l). Thus,

$$n^{1/2}\sup_{f_n'\in\mathcal{F}_n'}|U_2^n f_n' - 2P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\alpha)]) + \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\alpha)]| = o_p(1),$$

which implies

$$n^{1/2}\{-2P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})]) + \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})]\} = o_p(1),$$

where

$$\begin{aligned}&\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\alpha)]\\=&\frac{\sqrt{h}}{nh^2}\sum_l [2F_\epsilon(Y_{i+j} - m(d_{i+l}) - Z_{jl}^T H\alpha) - 1]Z_{jl}1(|x_j| \le h)1(|x_l| \le h)\end{aligned}$$

with $F_\epsilon(\cdot)$ being the cumulative distribution function of ϵ. In other words,

$$2\mathbb{G}_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})]) + \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})] = o_p(1),$$

where $\mathbb{G}_n(f) = \sqrt{n}(P_n - P)f$ is the standardized empirical process. By Lemma 2.13 of Pakes and Pollard (1989), $\mathcal{F}_{1n}' = \{\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\alpha) : \alpha \in \mathcal{B}]\}$ is Euclidean with envelope $F_{1n} = \frac{\sqrt{h}}{h^2} Z_{jl}\sum_l \|Z_{jl}\| 1(|x_j| \le h)1(|x_l| \le h)$, so by Lemma 2.17 of Pakes and Pollard (1989), and $H(\hat{\alpha} - \beta) = o_p(1)$, we have

$$\mathbb{G}_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})]) = \mathbb{G}_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) + o_p(1).$$

Thus,

$$\begin{aligned}&2\mathbb{G}_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) + \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})]\\=&2\sqrt{n}P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) - 2\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]\\&+ \sqrt{n}(\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})] - \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) + \sqrt{n}\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]\\=&2\sqrt{n}P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})]) + 2P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\beta)] - \bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})])\\&+ \sqrt{n}(\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})] - \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) + \sqrt{n}\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]\\=&2\sqrt{n}P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})]) + \sqrt{n}(\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})] - \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)])\\&+ \sqrt{n}\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]\\=&o_p(1),\end{aligned}$$

where

$$\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})] = \frac{\sqrt{h}}{nh^2}\sum_l [2F_\epsilon(\epsilon_{i+j}) - 1]Z_{jl}1(|x_j| \le h)1(|x_l| \le h)$$

satisfies $\mathrm{E}[\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})]] = 0$, and the second to last equality is from

$$\sqrt{n}\{P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) - P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})])\} \approx \sqrt{n}\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)].$$

Since

$$\begin{aligned}\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\alpha)] = &\frac{\sqrt{h}}{nh^2}\sum_{l,j} Z_{jl}1(|x_j| \le h)1(|x_l| \le h)\\&\times [2\int F_\epsilon(\epsilon + m(x_{i+j}) + m(x_{i+l}) - Z_{jl}^T H\alpha) - 1]f(\epsilon)\mathrm{d}\epsilon,\end{aligned}$$

we have

$$\sqrt{n}(\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\hat{\alpha})] - \bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)]) \approx -\sqrt{nh}g(0)(\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}Z_{jl}^T)H(\hat{\alpha} - \beta),$$

$$\sqrt{n}\bar{\mathrm{E}}[f_n'(Y_{i+j}, Y_{i+l}|\beta)] \approx \sqrt{nh}g(0)\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}(x_j^2 + x_l^2)\frac{m^{(2)}(x_i)}{2!}.$$

In summary,

$$\begin{aligned}&\sqrt{nh}\{H(\hat{\alpha} - \beta) - (\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}Z_{jl}^T)^{-1}(\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}(x_j^2 + x_l^2))\frac{m^{(2)}(x_i)}{2!}\}\\\approx&2g(0)^{-1}(\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}Z_{jl}^T)^{-1}\sqrt{n}P_n(\bar{\mathrm{E}}_2[f_n'(Y_{i+j}, Y_{i+l})]).\end{aligned}$$

Thus, the asymptotic bias is

$$e^T H^{-1}(\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}Z_{jl}^T)^{-1}(\frac{1}{n^2h^2}\sum_{l,j} Z_{jl}(x_j^2 + x_l^2))\frac{m^{(2)}(x_i)}{2!} = \frac{m^{(2)}(x_i)}{6}\frac{k^2}{n^2},$$

and the asymptotic variance is

$$\frac{4}{kg(0)^2} e^T H^{-1} G^{-1} V G^{-1} H^{-1} e = \frac{1}{6g(0)^2 k},$$

where $e = (1, 0)^T$, $G = \frac{1}{k^2} \sum_{l,j} Z_{jl} Z_{jl}^T$, and $V = \frac{1}{3k} \sum_j (\frac{1}{k} \sum_l Z_{jl})(\frac{1}{k} \sum_l Z_{jl})^T$ with $\mathrm{Var}[2F_\epsilon(\epsilon_{i+j}) - 1] = 1/3$.

References

Brown, L., Cai, T., Zhou, H., 2008. Robust nonparametric estimation via wavelet median regression. Ann. Statist. 36 (5), 2055–2084.
Cleveland, W., 1979. Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc. 74 (368), 829–836.
Fan, J., Farmen, M., Gijbels, I., 1998. Local maximum likelihood estimation and inference. J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (3), 591–608.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Cahpman & Hall, London.
Fan, J., Hall, P., 1994. On curve estimation by minimizing mean absolute deviation and its implications. Ann. Statist. 22 (2), 867–885.
Fan, Y., Hardle, W., Wang, W., Zhu, L., 2018. Composite quantile regression for the single-index model. J. Bus. Econom. Statist. 22 (2), 867–885.
Ghosal, S., Sen, A., vander Vaart, A., 2000. Testing monotonicity of regression. Ann. Statist. 28 (4), 1054–1082.
Hampel, F., 1968. Contributions to the Theory of Robust Estimation (Ph.d. thesis). University of California, Berkeley.
Hampel, F., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 62, 1179–1186.
Härdle, W., Gasser, T., 1984. Robust non-parametric function fitting. J. R. Stat. Soc. Ser. B Stat. Methodol. 46 (1), 42–51.
Härdle, W., Marron, S., 1985. Optimal bandwidth selection in nonparametric regression function estimation. Ann. Statist. 13 (4), 1465–1481.
Härdle, W., Spokoiny, V., Wang, W., 2013. Local quantile regression. J. Statist. Plann. Inference 143 (7), 1109–1129.
Hershberger, S., 2011. Hodges-lehmann estimators. In: Lovric, M. (Ed.), International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg.
Hodges, J., Lehmann, E., 1956. The efficiency of some nonparametric competitors of the t-test. Ann. Math. Stat. 27, 324–335.
Huber, P., Ronchetti, E., 2009. Robust Statistics, second ed. Wiley, New York.
Ichimura, H., Newey, W., 2017. The influence function of semiparametric estimators.
Kai, B., Li, R., Zou, H., 2010. Local composite quantile regression smoothing: an efficient and sate alternative to local polynomial regression. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (1), 49–69.
Koenker, R., 2005. Quantile Regression. Cambridge University Press, New York.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Li, Q., Racine, J., 2004. Cross-validated local linear nonparametric regression. Statist. Sinica 14, 485–512.
Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.
Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. Econometric Theory 7, 186–199.
Porter, J., Yu, P., 2015. Regression discontinuity designs with unknown discontinuity points. J. Econometrics 189, 132–147.
Silverman, B.W., 1985. Some aspects of the spline smoothing appraoch to nonparametric regression curve fitting (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 47, 1–52.
Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MA.
Wang, W., Lin, L., 2015. Derivative estimation based on difference sequence via locally weighted least squares regression. J. Mach. Learn. Res. 16, 2617–2641.
Wang, W., Lin, L., Yu, L., 2017. Optimal variance estimation based on lagged second-order difference in nonparametric regression. Comput. Statist. 32, 1047–1063.
Wang, F., Scott, D., 1994. The L_1 method for robust nonparametric regression. J. Amer. Statist. Assoc. 89 (425), 65–76.
Wang, W., Yu, P., 2017. Asymptotically optimal differenced estimators of error variance in nonparametric regression. Comput. Statist. Data Anal. 125–143.
Wang, W., Yu, P., Lin, L., Tong, T., 2019. Robust estimation of derivatives using locally weighted least absolute deviation regression. J. Mach. Learn. Res. 20 (60), 1–49.
Welsh, A., 1996. Robust estimation of smooth regression and spread functions and their derivatives. Statist. Sinica 347–366.
Yu, K., Jones, M., 1998. Local linear quantile regression. J. Amer. Statist. Assoc. 93, 228–237.
Zhao, Z., Xiao, Z., 2014. Efficient regression via optimally combining quantile information. Econometric Theory 30 (6), 1272–1314.
Zou, H., Yuan, M., 2008. Composite quantile regression and the oracle model selection theory. Ann. Statist. 36 (3), 1108–1126.