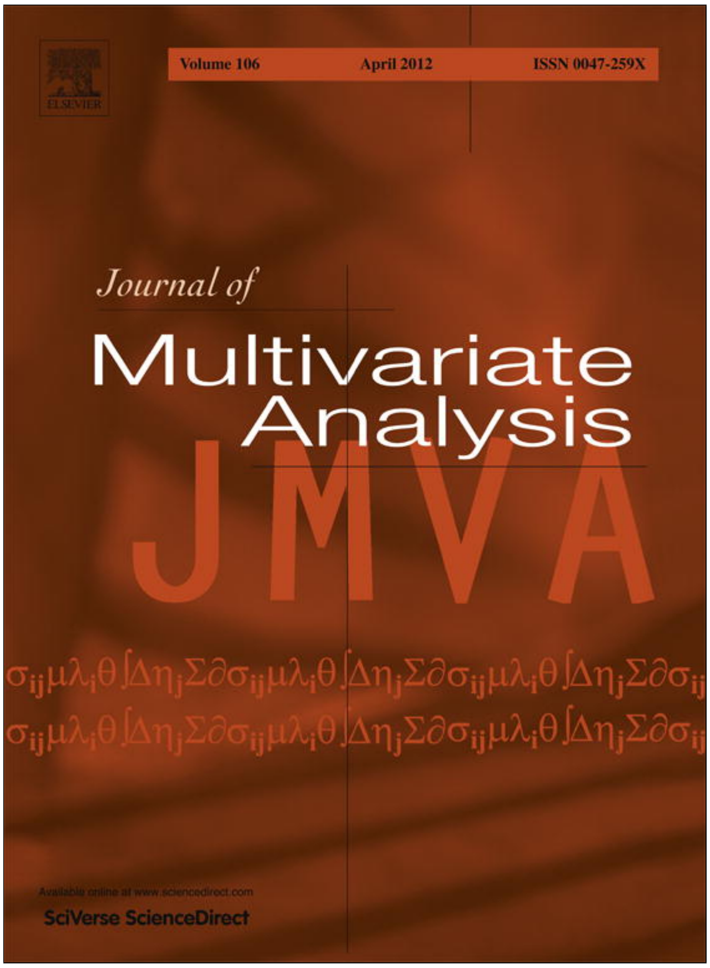
Volume 106
April 2012
ISSN 0047-259X
Journal of
Multivariate Analysis
JMVA
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Journal of Multivariate Analysis 106 (2012) 36–48

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Local Walsh-average regression

Long Feng, Changliang Zou, Zhaojun Wang *

LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, China

ARTICLE INFO

Article history:
Received 14 September 2010
Available online 16 December 2011

AMS 2000 subject classifications:
62G05
62G08
62G20

Keywords:
Asymptotic efficiency
Local composite quantile estimator
Local polynomial regression
Robust nonparametric regression
Walsh-average regression

ABSTRACT

Local polynomial regression is widely used for nonparametric regression. However, the efficiency of least squares (LS) based methods is adversely affected by outlying observations and heavy tailed distributions. On the other hand, the least absolute deviation (LAD) estimator is more robust, but may be inefficient for many distributions of interest. Kai et al. (2010) [13] propose a nonparametric regression technique called local composite quantile regression (LCQR) smoothing to improve local polynomial regression further. However, the performance of LCQR depends on the choice of the number of quantiles to combine, a meta parameter which plays a vital role in balancing the performance of LS and LAD based methods. To overcome this issue, we propose a novel method termed the local Walsh-average regression (LWAR) estimator by minimizing a locally Walsh-average based loss function. Under the same assumptions in Kai et al. (2010) [13], we theoretically show that the proposed estimator is highly efficient across a wide spectrum of distributions. Its asymptotic relative efficiency with respect to the LS based method is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test. Both of the theoretical and numerical results demonstrate that the performance of the new approach and LCQR is at least comparable in estimating the nonparametric regression function or its derivatives and in some cases the new approach performs better than the LCQR with commonly recommended number of quantiles, especially for estimating the regression function.

1. Introduction

Nonparametric regression is a smoothing method for recovering the regression function and its characteristics from noisy data. Consider the general nonparametric regression model

$$Y = m(T) + \sigma(T)\varepsilon, \tag{1}$$

where Y is the response variable, T is a covariate, $m(T) = E(Y \mid T)$, which is assumed to be a smooth non-parametric function, and $\sigma(T)$ is a positive function representing the standard deviation. Suppose that $\{y_i, t_i\}_{i=1}^n$, is an independent and identically distributed random sample, and the errors ε_i's are independent and identically distributed (i.i.d.) from a distribution G (its density denoted by g). The aim of nonparametric regression is to estimate the value of $m(t_0)$ or its derivatives $m^{(r)}(t_0)$ for $r \geq 1$.

Due to both simplicity of computation and nice asymptotic properties, the local least squares (LLS) polynomial estimator is one of the popular smoothing methods and has been well studied in the literature [2,4]. To be more specific, the LLS

* Corresponding author.
E-mail address: zjwangnk@yahoo.com (Z. Wang).

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.12.003

estimators are constructed according to the following criterion: find $(\hat{a}, \hat{b}_1, \ldots, \hat{b}_p)$ so as to minimize

$$\sum_{i=1}^{n} \left\{ y_i - a - \sum_{k=1}^{p} b_k (t_i - t_0)^k \right\}^2 K\left(\frac{t_i - t_0}{h}\right), \tag{2}$$

where $K(\cdot)$ is a smooth kernel function and h is the smoothing parameter.

Although the LLS method is a successful and standard choice in nonparametric fitting, it may suffer when the error follows a heavy-tailed distribution or in the presence of outliers. Specially, when the error distribution does not have finite second moment which occurs for instance when the random error has a Cauchy distribution, the LLS is no longer consistent. Thus, considerable efforts have been devoted to construct some robust nonparametric polynomial smoothers. One way is to use a general function $\zeta(\cdot)$ instead of quadratic loss in (2) to form local M-type polynomial regression estimators, say

$$\sum_{i=1}^{n} \zeta \left(y_i - a - \sum_{k=1}^{p} b_k (t_i - t_0)^k \right) K\left(\frac{t_i - t_0}{h}\right).$$

It is interesting to point out that the concept of robust local polynomial regression was previously introduced by Cleveland [1], leading to the so-called LOWESS or Locally Weighted Smoothing Scatterplots. When p is chosen to be 0 and an outlier-resistant function $\zeta(\cdot)$ [12] is specified, minimizing the objective function above would lead to a robust local constant estimator. See [1,9], for example. Fan et al. [6], Fan and Jiang [8], and Welsh [20] pointed out that, the local M-type polynomial regression estimators with $p \geq 1$ cannot only have advantage of reducing bias but also cope well with edge effects and be effective methods for derivative estimation.

One of the most popular local M-type regression estimators may be the local least absolute deviation (LLAD) polynomial estimator,

$$\sum_{i=1}^{n} \left| y_i - a - \sum_{k=1}^{p} b_k (t_i - t_0)^k \right| K\left(\frac{t_i - t_0}{h}\right),$$

which can be more robust when g deviates from the normal (see [5,6]). Nevertheless, the efficiency of LLAD compared to LLS is proportional to the density at the median. For the Gaussian error case, the distribution of the greatest interest, this quantity is only about 0.697 $(0.637^{4/5})$. And worse still, the efficiency can be arbitrarily small if $g(0)$ is close to zero.

To overcome some of the above issues, Kai et al. [13] proposed the local composite quantile (LCQR) regression estimator by averaging multiple local quantile regressions [21]. The LCQR estimator is based on solving

$$\arg \min_{a_1,\ldots,a_q,b_1,\ldots,b_p} \sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k} \left\{ y_i - a_k - \sum_{j=1}^{p} b_j (t_i - t_0)^j \right\} K\left(\frac{t_i - t_0}{h}\right), \tag{3}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile regression function, and $\tau_1, \ldots, \tau_q$ are the quantiles to combine. Note that there are $q + p$ parameters to optimize since there are q intercepts included in this formulation. Indeed, the LCQR can be seen as a tradeoff case of LLS ($q = \infty$) and LLAD ($q = 1$) by choosing an appropriate $q \in (1, \infty)$. Kai et al. [13] showed that LCQR can be more robust in various circumstances, with little loss in efficiency compared to LLS in estimating the regression function $m(t_0)$ or its derivative $m'(t_0)$. However, the performance of LCQR depends on the choice of q, the number of quantiles to combine, a meta parameter which plays a vital role in balancing the performance of LLS and LLAD.

Theoretically speaking, the nice properties of LCQR lie on the assumption that the error distribution is symmetric (see the discussions in [13]). Note that for the symmetric location model in the context of nonparametric statistics, the median of Walsh-average of the data, which is also well known as Hodges–Lehmann estimate [11], can balance robustness and efficiency in estimating the location parameter. This motivates us to develop a novel Walsh-average estimation procedure that is highly efficient, robust and computationally simple. The new approach can significantly improve on classical local polynomial regression for a wide class of error distributions and has comparable efficiency in the worst case scenario. Theoretical analysis reveals that the asymptotic relative efficiency (ARE), measured by the asymptotic mean squared error (or the asymptotic mean integrated squared error), of the local Walsh-average estimator in comparison with the LLS estimator has an expression that is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test. By both theoretical and numerical analysis, we find that the performance of the new approach and LCQR is at least comparable in estimating $m(t_0)$ or $m'(t_0)$ and in many situations the new approach performs better than the LCQR with commonly used parameter q (5 or 9 recommended by Kai et al. [13]). In contrast with LCQR, the new approach does not need to choose the meta parameter q and the minimization problem is just $p + 1$ rather than $p + q$ as in LCQR.

The remainder of this paper is organized as follows: our proposed methodology is described in detail in Section 2, including the estimation procedure, algorithm and main theoretical properties. Its numerical performance is investigated in Section 3. Several remarks draw the paper to its conclusion in Section 4. The technical details are provided in the Appendix A.

2. Local *p*-polynomial Walsh-average regression

2.1. The estimation procedure

Firstly, consider the classical univariate location model,

$$X = \theta + \varepsilon,$$

where θ is the location parameter and we assume that the distribution of ε is symmetric around 0. Given a sample of observations $X_1, \ldots, X_n$, the Hodges–Lehmann estimate of θ is accordingly defined by the median of Walsh-average of $\{X_i\}_{i=1}^n$, that is

$$\widehat{\theta} = \text{Median}_{i \le j} \left\{ \frac{X_i + X_j}{2} \right\}.$$

This is also equivalent to the minimization problem that

$$\widehat{\theta} = \arg\min_{\theta} \sum_{i \le j} \left| \frac{X_i + X_j}{2} - \theta \right|. \tag{4}$$

Recall the model (1). In local p-polynomial regression we first approximate $m(t)$ locally by a p-polynomial function

$$m(t) \approx m(t_0) + m'(t_0)(t - t_0) + \cdots + \frac{1}{p!} m^{(p)}(t_0)(t - t_0)^p$$

and then fit a polynomial model locally in a neighborhood of t_0. Naturally, we may define the local weighted Walsh-average loss function by combining the local polynomial regression and the Walsh-average minimization problem (4)

$$L_n(a, b_1, \ldots, b_p) = \frac{1}{2n(n+1)} \sum_{i \le j} \left| e_i + e_j \right| K_h(t_i - t_0) K_h(t_j - t_0), \tag{5}$$

where $e_i = y_i - a - \sum_{k=1}^{p} b_k (t_i - t_0)^k$ and $K_h(\cdot) = K(\cdot/h)/h$. Denote the minimizer of (5) by $(\hat{a}, \hat{b}_1, \ldots, \hat{b}_p)$ and the local p-polynomial Walsh-average estimator of $m(t_0)$ and $m^{(r)}(t_0)$ are accordingly given by

$$\hat{m}(t_0) = \hat{a}$$
$$\hat{m}^{(r)}(t_0) = r! \hat{b}_r, \quad r = 1, \ldots, p.$$

Hereafter, we refer to these estimates as the local Walsh-average regression (LWAR) estimators.

Remark 1. Wang et al. [18] recently proposed an estimation procedure for the varying coefficient models

$$Y = a_0(U) + \mathbf{X}^T \mathbf{a}(U) + \varepsilon$$

based on local ranks. In their procedure, the location smooth function $a_0(U)$ is estimated separately by LLAD after the $\mathbf{a}(U)$ is estimated by using the local rank approach. This is because the intercept is not identifiable in the global rank estimation of linear model [10]. The symmetric assumption is not needed there because of the use of LLAD. In contrast, the LCQR and LWAR estimators of $m(t_0)$ can achieve consistency due to the assumption that the error distribution is symmetric. See [13] and Section 2.3 for further discussions.

2.2. Computational algorithm

Wang et al. [18] provide a pseudo-observation algorithm to implement the local rank procedure for varying coefficient model. Similar to their idea, the local Walsh-average smoothing estimator of $(a, b_1, \ldots, b_p)^T$ can be solved by fitting a weighted L_1 regression on $n(n+1)/2$ pseudo-observations $\{\boldsymbol{x}_i + \boldsymbol{x}_j, y_i + y_j\}$ with weights $w_{ij} = K((t_i - t_0)/h)K((t_j - t_0)/h)$, where $\boldsymbol{x}_i = (1, t_i - t_0, \ldots, (t_i - t_0)^p)^T$, $1 \le i \le j \le n$. Such an algorithm is quite efficient and reliable in practice because subroutines or functions for carrying out weighted L_1 regression are available in most of statistical software packages. In our numerical studies, we use the function *rq* in the R package *quantreg*. The R code for implementing the proposed scheme are available from the authors upon request. When the sample size is small, the minimization algorithm for LWAR is usually faster than the LCQR with large q because it has less parameters. In contrast, when the sample size is large, LCQR with a small or moderate q would be faster to implement than LWAR because there is a double-sum in the objective function for LWAR.

2.3. Asymptotic properties

In this subsection, we establish the asymptotic properties of the local Walsh-average regression. Denote by $f(\cdot)$ the marginal function of the covariate T. We choose the kernel function K such that K is a symmetric density function with a finite support and let

$$\mu_j = \int u^j K(u)du \quad \text{and} \quad \nu_j = \int u^j K^2(u)du, \quad j = 0, 1, 2, \ldots.$$

Without loss of generality, we assume that the variance of error distribution to be 1. Otherwise, we can conveniently transform the error term in model (1) to $[\sigma\sigma(T)] \times (\sigma^{-1}\varepsilon)$ when the variance is σ^2. For the asymptotic analysis, we need the following regularity conditions.

(C1) $m(t)$ has a continuous $(p+2)$-th derivative in the neighborhood of t_0.
(C2) $f(\cdot)$ is differentiable and positive in the neighborhood of t_0.
(C3) The conditional variance $\sigma^2(t)$ is continuous in the neighborhood of t_0.
(C4) The error has a symmetric distribution with a positive density $g(\cdot)$ which is Lipschitz continuous.

The following notations are needed to present the asymptotic properties of the local p-polynomial Walsh-average estimator. Let $\mathbf{S}$ and $\mathbf{S}^*$ be $(p+1)\times(p+1)$ matrices with $(l+1, k+1)$-elements $\mu_{l+k}+\mu_l\mu_k$ and $\nu_{l+k}+\mu_l\nu_k+\mu_k\nu_l+\mu_l\mu_k\nu_0$, $0 \le l, k \le p$, respectively. Similarly, let $\boldsymbol{c}_p$ and $\tilde{\boldsymbol{c}}_p$ be $(p+1)$-variate vectors with $\mu_{p+i+1}+\mu_{p+1}\mu_i$ and $\mu_{p+i+2}+\mu_{p+2}\mu_i$, $0 \le i \le p$, respectively. Let $\mathbf{T}$ be the σ-field that is generated by $\{t_1, \ldots, t_n\}$. Moreover, we use notation $m^{(0)}(t_0)$ for $m(t_0)$ in certain circumstance for simplicity.

Theorem 1. *Suppose that t_0 is an interior point of the support of $f(\cdot)$ and Conditions (C1)–(C4) hold. If $h \to 0$ and $nh^2 \to \infty$, then, for $0 \le r \le p$,*

(i) *The asymptotic conditional variance of $\hat{m}^{(r)}(t_0)$ is given by*

$$\mathrm{var}(\hat{m}^{(r)}(t_0) \mid \mathbf{T}) = \boldsymbol{e}_{r+1}^T \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\boldsymbol{e}_{r+1} \frac{r!^2\sigma^2(t_0)}{12\tau^2 f(t_0) nh^{1+2r}} + o_p\left(\frac{1}{nh^{1+2r}}\right),$$

where $\tau = \int g^2(x)dx$;

(ii) *If $p-r$ is odd, then the asymptotic conditional bias of $\hat{m}^{(r)}(t_0)$ is given by*

$$\mathrm{bias}(\hat{m}^{(r)}(t_0) \mid \mathbf{T}) = \boldsymbol{e}_{r+1}^T \mathbf{S}^{-1}\boldsymbol{c}_p \frac{r!}{(p+1)!} m^{(p+1)}(t_0) h^{p+1-r} + o_p(h^{p+1-r}).$$

If $p-r$ is even, then

$$\mathrm{bias}(\hat{m}^{(r)}(t_0) \mid \mathbf{T}) = \boldsymbol{e}_{r+1}^T \mathbf{S}^{-1}\tilde{\boldsymbol{c}}_p \frac{r!}{(p+1)!}\left[m^{(p+2)}(t_0) + (p+2)m^{(p+1)}(t_0)\frac{f'(t_0)}{f(t_0)}\right] h^{p+2-r} + o_p(h^{p+2-r}),$$

where $\boldsymbol{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ with the kth element being one;

(iii) *Furthermore, conditioning on $\mathbf{T}$, we have*

$$\sqrt{nh^{1+2r}}\{\hat{m}^{(r)}(t_0) - m^{(r)}(t_0) - \mathrm{bias}(\hat{m}^{(r)}(t_0))\} \to N\left(0, \boldsymbol{e}_{r+1}^T \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\boldsymbol{e}_{r+1} \frac{r!^2\sigma^2(t_0)}{12\tau^2 f(t_0)}\right),$$

in distribution as $n \to \infty$.

2.4. Asymptotic relative efficiency

In this subsection, we study the asymptotic relative efficiency of the local Walsh-average estimators with that of the local least squares estimators for estimating $m(t_0)$ and $m'(t_0)$. Parallel to [13], we also consider to use the local linear $(p = 1)$ and quadratic $(p = 2)$ Walsh-average regression for estimating the regression function $m(\cdot)$ and the derivative function $m'(\cdot)$, respectively. Similar to LCQR and LLS [13,3], we will see these two LWAR estimators enjoy the ideal feature that the asymptotic conditional biases of the estimators do not depend on $f(\cdot)$. We summarize the result into the following corollaries. We only provide some key points regarding the proofs in the Appendix A since they follow from Theorem 1 and some straightforward algebra.

Hereafter we denote the local estimators based on least squares, least Walsh-average and least composite quantile regression with q quantiles by $\hat{m}_{\mathrm{LS}}$, $\hat{m}_{\mathrm{WA}}$ and $\hat{m}_{\mathrm{CQR}_q}$, respectively. The dependence on the degrees of polynomial in each estimator is not made explicit in notation for simplicity, which should not cause any confusion. We only need keep in mind that $m(t_0)$ and $m'(t_0)$ are respectively estimated through local linear and quadratic regression for all the methods.

Corollary 1. *Suppose the assumptions given in Theorem 1 all hold. Then, the asymptotic conditional bias and variance of the local linear Walsh-average estimator $\hat{m}_{\mathrm{WA}}(t_0)$ are given by*

$$\mathrm{bias}\{\hat{m}_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{2}m''(t_0)\mu_2 h^2 + o_p(h^2),$$

$$\mathrm{var}\{\hat{m}_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{nh}\frac{\upsilon_0\sigma^2(t_0)}{12\tau^2 f(t_0)} + o_p\left(\frac{1}{nh}\right).$$

Furthermore, conditioning on **T**, *we have*

$$\sqrt{nh}\left\{\hat{m}_{\mathrm{WA}}(t_0) - m(t_0) - \frac{1}{2}m''(t_0)\mu_2 h^2\right\} \to N\left(0, \frac{\upsilon_0\sigma^2(t_0)}{12\tau^2 f(t_0)}\right),$$

in distribution as $n \to \infty$.

Corollary 2. *Suppose the assumptions given in Theorem 1 all hold. Then, the asymptotic conditional bias and variance of the local quadratic Walsh-average estimator $\hat{m}'_{\mathrm{WA}}(t_0)$ are given by*

$$\mathrm{bias}\{\hat{m}'_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{6}m'''(t_0)\frac{\mu_4}{\mu_2}h^2 + o_p(h^2),$$

$$\mathrm{var}\{\hat{m}'_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{nh^3}\frac{\upsilon_2\sigma^2(t_0)}{12\tau^2\mu_2^2 f(t_0)} + o_p\left(\frac{1}{nh^3}\right).$$

Furthermore, conditioning on **T**, *we have*

$$\sqrt{nh^3}\left\{\hat{m}'_{\mathrm{WA}}(t_0) - m'(t_0) - \frac{1}{6}m'''(t_0)\frac{\mu_4}{\mu_2}h^2\right\} \to N\left(0, \frac{\upsilon_2\sigma^2(t_0)}{12\tau^2\mu_2^2 f(t_0)}\right),$$

in distribution as $n \to \infty$.

Remark 2. It is worth noting that the symmetric distribution assumption is no longer needed for the asymptotic normality of $\hat{m}'_{\mathrm{WA}}(t_0)$ given in Corollary 2. From the proof of Corollary 2, we can see that without that assumption, the asymptotic bias remains the same and the asymptotic variance is given by

$$\frac{1}{nh^3}\frac{\xi\upsilon_2\sigma^2(t_0)}{\delta^2\mu_2^2 f(t_0)},$$

where $\delta = \int g(-u)g(u)du$ and $\xi = \int (G(-u) - 1/2)^2 g(u)du$. Under the symmetric assumption, we have $\delta = \tau$ and $\xi = 1/12$ and consequently the asymptotic variance in Corollary 2 follows. Thus, the symmetric assumption is only used for asymptotic relative efficiency comparison between LCQR and LWAR but the consistency and asymptotic normality of local quadratic Walsh-average for estimating m' still hold without this assumption. This feature is mainly due to $\int uK(u)du = 0$ and also shared by LCQR as mentioned in [13].

By two corollaries above, we now are ready to compare the estimation efficiency of LWAR estimators with that of LLS estimators for estimating $m(t_0)$ and $m'(t_0)$. The local linear least squares estimator for $m(t_0)$ has mean square error (MSE)

$$\mathrm{MSE}(\hat{m}_{\mathrm{LS}}(t_0)) = \left\{\frac{1}{2}m''(t_0)\mu_2\right\}^2 h^4 + \frac{1}{nh}\frac{\upsilon_0\sigma^2(t_0)}{f(t_0)} + o_p\left(h^4 + \frac{1}{nh}\right),$$

and hence

$$h_{\mathrm{LS}}^{\mathrm{opt}}(t_0) = \left\{\frac{\upsilon_0\sigma^2(t_0)}{f(t_0)[m''(t_0)\mu_2]^2}\right\}^{1/5} n^{-1/5}, \quad (6)$$

in which $h_{\mathrm{LS}}^{\mathrm{opt}}(t_0)$ is the optimal bandwidth minimizing the asymptotic MSE for estimating $m(t_0)$. The local Walsh-average estimator for $m(t_0)$ has MSE

$$\mathrm{MSE}(\hat{m}_{\mathrm{WA}}(t_0)) = \left\{\frac{1}{2}m''(t_0)\mu_2\right\}^2 h^4 + \frac{1}{nh}\frac{\upsilon_0\sigma^2(t_0)}{12\tau^2 f(t_0)} + o_p\left(h^4 + \frac{1}{nh}\right),$$

and correspondingly

$$h_{\mathrm{WA}}^{\mathrm{opt}}(t_0) = \left\{\frac{\upsilon_0\sigma^2(t_0)}{12\tau^2 f(t_0)[m''(t_0)\mu_2]^2}\right\}^{1/5} n^{-1/5}. \quad (7)$$

Table 1
The asymptotic efficiency comparisons of LLS, LLAD, LCQR and LWAR for estimating m and m'. $t(d)$: student's t-distribution with d degrees of freedom. $T(\rho;\sigma)$: Tukey contaminated normal with CDF $F(x) = (1-\rho)\Phi(x) + \rho\phi(x/\sigma)$ where $\Phi(x)$ is the CDF of a standard normal distribution and $\rho \in [0, 1]$ is the contamination proportion.

Error	ARE(LLS, LLAD)		ARE(LLS, LCQR_5)		ARE(LLS, LCQR_9)		ARE(LLS, LWAR)	
	m	m'	m	m'	m	m'	m	m'
$N(0, 1)$	0.697	0.773	0.934	0.945	0.966	0.963	0.964	0.974
Laplace	1.741	1.486	1.220	1.281	1.155	1.268	1.383	1.261
$t(3)$	1.472	1.318	1.597	1.441	1.524	1.444	1.670	1.443
$t(4)$	1.099	1.070	1.265	1.204	1.238	1.210	1.310	1.213
$T(0.05; 3)$	0.864	0.901	1.130	1.087	1.157	1.102	1.154	1.108
$T(0.10; 3)$	0.999	0.999	1.271	1.187	1.279	1.198	1.289	1.199
$T(0.05; 10)$	2.696	2.031	3.458	2.423	3.478	2.447	3.489	2.442
$T(0.10; 10)$	4.051	2.716	4.913	3.145	4.705	3.143	4.845	3.109

This allows us to calculate the local asymptotic relative efficiency

$$\frac{\text{MSE}_{\text{opt}}(\hat{m}_{\text{LS}}(t_0))}{\text{MSE}_{\text{opt}}(\hat{m}_{\text{WA}}(t_0))} \to (12\tau^2)^{4/5},$$

as $n \to \infty$, where we use MSE_{opt} to denote the MSE evaluate at its optimal bandwidth. Alternatively, we may consider the ARE obtained by comparing the MISE, which is defined as $\text{MISE}(h) = \int E[\hat{m}(t_0) - m(t_0)]^2 w(t)dt$ with a weighted function $w(\cdot)$. This provides a global measurement. With the theoretical optimal bandwidths, we can obtain the ARE based on MISE is still

$$\text{ARE}(\hat{m}_{\text{LS}}, \hat{m}_{\text{WA}}) = (12\tau^2)^{4/5}.$$

Similarly, by Corollary 2, we can derive the ARE of the local quadratic Walsh-average estimator with respect to local quadratic least squares estimator is given by

$$\text{ARE}(\hat{m}'_{\text{LS}}, \hat{m}'_{\text{WA}}) = (12\tau^2)^{4/7}.$$

Note that the above ARE is essentially related to the ARE of the signed-rank Wilcoxon test in comparison with the t-test and it has a lower bound 0.8896 ($0.864^{4/5}$; cf., Section 1.7.2 in [10]). It is also worth noting that $\hat{m}'_{\text{WA}}$ should have similar performance to the LCQR with infinity quantiles, say, $\hat{m}'_{\text{CQR}_\infty}$, because they have the same asymptotic MSE and relative efficiency with respect to LLS according to Theorem 4 in [13] and Corollary 2.

The ARE of local linear Walsh-average versus local linear least squares, together with that of LCQR and LLAD, are documented in Table 1 for a number of distributions, including some commonly used distributions such as the normal distribution, t distribution, and Tukey's contaminated normal distribution. Note that the values for LCQR are just copied from the values given in Table 1 of [13]. From this table, LWAR is asymptotically much more efficient than LLAD for the t distribution and Tukey's contaminated normal model. Compared to LLS, LWAR is almost as efficient for the normal case, but can much more efficient for heavy tailed distributions, such as t and Tukey's contaminated normals. Regarding the comparison between LWAR and LCQR, they have very similar AREs in estimating derivatives. The LWAR outperforms LCQR with $q = 5$ or 9 in the cases of t and Laplace distributions when the nonparametric function m is of interest, whereas they are comparable under Tukey contaminated normals.

2.5. Bandwidth selection

As the same with any nonparametric regression procedure, an important choice to be made is the amount of local averaging performed to obtain the regression estimate. For the local polynomial regression estimator, bandwidth selection rules were considered in [4,16], among others. Although we have derived the asymptotically optimal bandwidth for the LWAR estimator as in (7), it is difficult to use the plug-in method to estimate it due to many unknown quantities. Kai et al. [13] suggest to use the so-called "short-cut" strategy to obtain a neat but reasonable bandwidth. In this paper, we adopt their method and recommend the following bandwidth selection procedure. From (6) and (7), for estimating the regression function $m(\cdot)$, we see that

$$h_{\text{WA}}^{\text{opt}} = (12\tau^2)^{-1/5} h_{\text{LS}}^{\text{opt}}. \tag{8}$$

Thus, we can firstly use existing bandwidth selectors [2] to estimate $h_{\text{LS}}^{\text{opt}}$. The coefficient τ can be robustly estimated by using the approach given in [10] and easily be estimated by the function *wilcoxontau* in the R package [17]. Finally, we plug in this estimator into (8) to obtain the bandwidth for the LWAR estimator. For estimating the derivative $m'(\cdot)$ through local quadratic Walsh-average regression, a similar strategy could be employed. It is worth noting that the short-cut strategy for bandwidth selection may not work well if the error variance is very large or infinite. In such situations, one may consider the so-called "pilot" selector [4] which is a better choice than the short-cut strategy. Readers may refer to [13] for some related discussion.

Table 2
Simulation results for comparison of RASEs for estimating m.

Method Model	LLAD (I)	$LCQR_5$	$LCQR_9$	LWAR
Normal	$0.695_{(0.290)}$	$0.931_{(0.119)}$	$0.959_{(0.089)}$	$0.974_{(0.097)}$
Laplace	$1.345_{(0.648)}$	$1.109_{(0.199)}$	$1.072_{(0.135)}$	$1.202_{(0.233)}$
$t(3)$	$1.172_{(0.551)}$	$1.275_{(0.502)}$	$1.171_{(0.336)}$	$1.431_{(0.953)}$
$T(0.05, 3)$	$0.969_{(0.511)}$	$1.069_{(0.228)}$	$1.062_{(0.174)}$	$1.105_{(0.230)}$
$T(0.05, 10)$	$2.001_{(1.089)}$	$2.155_{(1.532)}$	$1.524_{(0.836)}$	$2.360_{(1.370)}$
Model	(II)			
Normal	$0.683_{(0.289)}$	$0.957_{(0.170)}$	$0.978_{(0.129)}$	$0.962_{(0.182)}$
Laplace	$1.314_{(0.766)}$	$1.194_{(0.328)}$	$1.141_{(0.252)}$	$1.281_{(0.452)}$
$t(3)$	$1.417_{(0.968)}$	$1.597_{(1.032)}$	$1.425_{(0.817)}$	$1.633_{(1.561)}$
$T(0.05, 3)$	$0.865_{(0.480)}$	$1.179_{(0.625)}$	$1.151_{(0.472)}$	$1.170_{(0.418)}$
$T(0.05, 10)$	$2.323_{(1.408)}$	$3.166_{(2.482)}$	$2.418_{(1.701)}$	$3.634_{(3.188)}$

Note: standard deviations are given in parentheses.

Table 3
Simulation results for comparison of RASEs for estimating m'.

Method Model	LLAD (I)	$LCQR_5$	$LCQR_9$	LWAR
Normal	$0.797_{(0.244)}$	$0.952_{(0.109)}$	$0.961_{(0.102)}$	$0.985_{(0.052)}$
Laplace	$1.323_{(0.528)}$	$1.101_{(0.168)}$	$1.103_{(0.157)}$	$1.142_{(0.130)}$
$t(3)$	$1.163_{(0.616)}$	$1.210_{(0.458)}$	$1.213_{(0.453)}$	$1.212_{(0.423)}$
$T(0.05, 3)$	$0.934_{(0.328)}$	$1.048_{(0.177)}$	$1.053_{(0.173)}$	$1.080_{(0.097)}$
$T(0.05, 10)$	$1.507_{(0.886)}$	$1.767_{(0.761)}$	$1.753_{(0.754)}$	$1.849_{(1.072)}$
χ_2^2	$0.989_{(0.821)}$	$1.474_{(1.121)}$	$1.469_{(0.999)}$	$1.466_{(0.986)}$
Model	(II)			
Normal	$0.627_{(0.784)}$	$0.938_{(0.359)}$	$0.946_{(0.309)}$	$0.971_{(0.380)}$
Laplace	$1.301_{(1.237)}$	$1.206_{(0.679)}$	$1.205_{(0.641)}$	$1.184_{(0.804)}$
$t(3)$	$1.321_{(1.801)}$	$1.610_{(1.756)}$	$1.598_{(1.805)}$	$1.336_{(1.006)}$
$T(0.05, 3)$	$0.829_{(0.892)}$	$1.227_{(2.061)}$	$1.213_{(1.879)}$	$1.048_{(0.644)}$
$T(0.05, 10)$	$2.056_{(2.184)}$	$3.059_{(5.670)}$	$3.029_{(5.343)}$	$2.481_{(3.781)}$
χ_2^2	$1.033_{(0.858)}$	$1.646_{(1.535)}$	$1.628_{(1.477)}$	$1.607_{(1.025)}$

Note: standard deviations are given in parentheses.

3. Numerical studies and examples

3.1. Simulation examples

We study the finite-sample performance of LWAR in this section via a simulation study. Throughout this section we use the Epanechnikov kernel and 1000 replications for each considered example. For a clear comparison, we adopt all the settings used in [13]. To be more specific, the bandwidth for the LLS method is chosen by a plug-in bandwidth selector [16] while the short-cut strategy illustrated in Section 2.5 is considered for LWAR. One example is the regression

$$Y = \sin(2T) + 2\exp(-16T^2) + 0.5\varepsilon,$$

where T follows $N(0, 1)$. We estimate $m(\cdot)$ and $m'(\cdot)$ over $[-1.5, 1.5]$; The other example is the one with heteroscedastic errors, with the form

$$Y = T\sin(2\pi T) + \sigma(T)\varepsilon,$$

where T follows $U(0, 1)$, $\sigma(t) = [2 + \cos(2\pi t)]/10$. In this example, we estimate $m(\cdot)$ and $m'(\cdot)$ over $[0, 1]$.

Five error distributions for ε are considered: $N(0, 1)$, Laplace, $t(3)$, Tukey contaminated normal $T(0.05; 3)$ and $T(0.05; 10)$. The performance of estimators is assessed via the average square errors (ASE) defined on 200 equally spaced grid points within given estimation intervals. We summarize simulation results by using the ratio of average square errors, $\text{RASE}(\hat{m}) = \text{ASE}(\hat{m}_{\text{LS}})/\text{ASE}(\hat{m})$ for an estimator $\hat{m}$, where $\hat{m}_{\text{LS}}$ is the LLS estimator. Kai et al. [13] have shown that the LCQR is an efficient and safe alternative to LLS in finite-sample cases and thus should be an ideal benchmark in our comparison. The means and standard deviations (in parentheses) of RASEs for estimating the function m and its derivative are summarized in Tables 2 and 3, respectively. In the simulation studies in [13], $LCQR_5$ and $LCQR_9$ seem to have better overall performance. So, we only present the results of these two estimators for the comparison use and all the values regarding LCQR given in this table are just copied from [13]. The results for the LLAD estimator are also presented in these two tables.

A few observations can be made from Tables 2 and 3. Firstly, the proposed LWAR method is highly efficient for all the distributions under consideration. Its efficiency compared to LLS is high and its RASEs are all greater than 1 except for

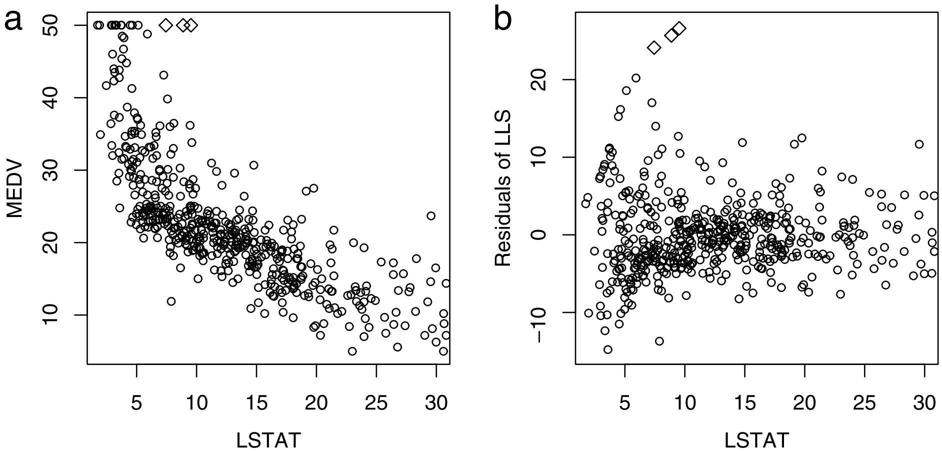

Fig. 1. (a) Scatter plot of data; (b) Residuals of local least squares regression.

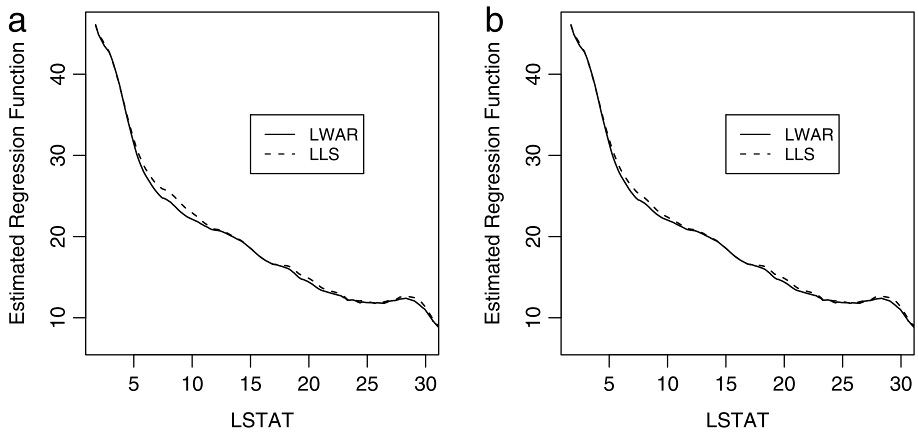

Fig. 2. Estimated regression function: (a) Using the whole data; (b) Using the data removing the three outliers.

the normal distribution. Even for normal, its RASEs are merely slightly smaller than 1. Secondly, the proposed bandwidth selection method for choosing h performs satisfactorily and conforms to the asymptotic results shown in Table 1. In other words, the theoretical ARE and simulated RASE of LWAR are always quite matched. Thirdly, the LWAR performs better than LCQRs with $q = 5$ and $q = 9$ in most situations when estimating m, while the LCQR outperforms LWAR for estimating the derivative in the heteroscedastic example. Moreover, both the LWAR and LCQR generally outperform the LLAD to certain degree in most cases. In addition, we add an asymmetric example in Table 3, the chi-square distribution with two degrees of freedom, to examine the finite sample performance of $\hat{m}'$ since the symmetric assumption is no longer required as we have mentioned before. The errors are centralized so that they have zero expectation. We can see that both the LCQR and LWAR perform significantly better than LLS and LLAD as we can expect and the two methods provide similar estimation ability in such case. We also examine other error variance magnitudes for both models and the conclusion is similar.

3.2. *A real-data example*

We applied our method to the Boston housing data, which has been widely studied in various contexts in the literature, for instance, [7,19], etc. The Boston housing data contains 506 observations, and is publicly available in the R package *mlbench*. The response variable Y is the median value of owner-occupied homes (MEDV) in each of the 506 census tracts in the Boston Standard Metropolitan Statistical Areas, and there are thirteen predictor variables. In this study, we are interested in the relationship between MEDV and one of the most important predictors, the percentage of lower status of the population (LSTAT). The scatter plot of data (MEDV versus LSTAT) is depicted in Fig. 1(a). We first estimate the regression function by using the local least squares estimator with the plug-in bandwidth selector. The residuals of LLS regression are shown in Fig. 1(b). Then we use the bandwidth selection method provided in Section 2.5 to select the bandwidth for LWAR. The unknown parameter τ in (8) can be estimated by using those residuals from LLS. The bandwidths selected are 2.09 and 2.22 for LLS and LWAR, respectively, when the regression function is of interest. For derivative estimation, the bandwidths are given by 3.55 and 3.71, respectively. The resulting estimated functions and derivative functions are displayed in Figs. 2(a) and 3(a), respectively.

It is readily seen from Figs. 2(a) and 3(a) that the two curves in each panel have similar patterns. The difference between the LLS estimates and the LWAR estimates is considerably large when the LSTAT is around (7,10). This is not surprising to us

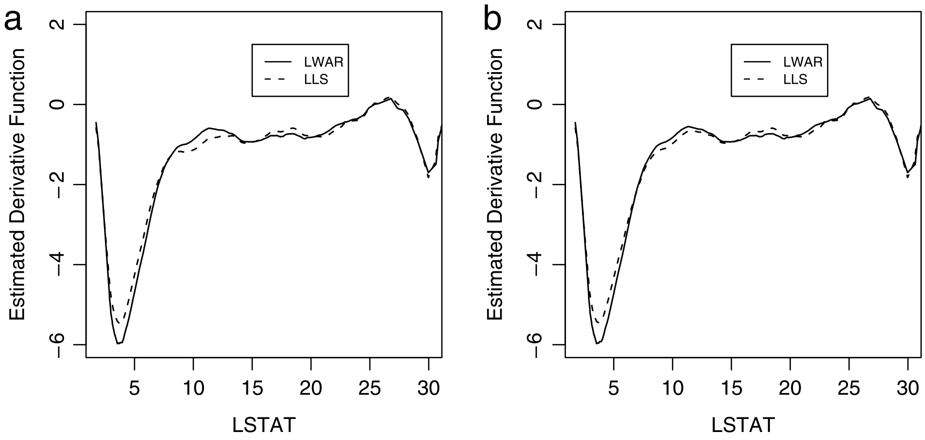

Fig. 3. Estimated derivative function: (a) Using the whole data; (b) Using the data removing the three outliers.

because from the scatter plot we can observe that there are three possible outlier observations, (7.44, 50), (8.88, 50) and (9.53, 50), which have been symboled by "diamond". These three outlying observations resulted in rather large residuals when the LLS method is employed, as shown in Fig. 1(b) (the corresponding three residuals are also symboled by "diamond").

To appreciate the robustness of LWAR and the influence of these three possible outliers, we obtain the LWAR and the LLS estimates after excluding these three possible outliers. The resulting estimates are displayed in Figs. 2(b) and 3(b). We can clearly see that the LWAR estimate remains almost the same, whereas the LLS estimate changes by certain degrees when LSTAT is around (7,10). Besides, after excluding these three outlying observations, the LLS estimate becomes much more close to the LWAR estimate.

4. Discussion

Both the LWAR and LCQR require a fundamental assumption that the error distribution is symmetric. Without such a condition their asymptotic biases will become a non-vanishing term. This is obviously stronger than the assumption in the situation when using the LLAD or LLS. This is the underlying key to the successes of LWAR and LCQR. Of course, such a condition is not too limited and could be valid in practice. A direct application of LWAR is to estimate the location part $a_0(U)$ in the varying-coefficient model mentioned in Remark 1 instead of LLAD after the $\mathbf{a}(U)$ is estimated by using [18]'s local rank approach. In such situation, the asymptotic relative efficiencies with respect to the LS based method for both the location and coefficient parts would be the same. In addition, as discussed by Remark 2, the symmetric error distribution assumption is no longer needed when estimating the derivatives $m'(\cdot)$. How to extend LWAR to other models, such as partly linear models and varying coefficient models, and how about the assumptions of consistency, are questions left for future research.

Acknowledgments

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This research was supported by the NNSF of China Grants No. 11001138, 11071128, 11131002, 11101306 and RFDP of China Grant No. 20110031110002. Zou also thanks the support of Nankai Young Grant No. 65010731.

Appendix A. Proofs of theorems

The proof follows along the same lines of the proof of Theorem 3.2 of [18], although part of details differs much. Here we only present a sketch of proof of theorems. A detailed technical report is available from an on-line supplemental file. Throughout this section, we will use the following notations for ease of exposition. Denote $\gamma_n = (nh)^{-1/2}$, $\boldsymbol{\beta} = (a, b_1, \ldots, b_p)^T$,

$$\boldsymbol{z}_i = \left[1, \frac{t_i - t_0}{h}, \ldots, \left(\frac{t_i - t_0}{h}\right)^p\right]^T,$$

$$\boldsymbol{\beta}^* = \gamma_n^{-1}\left\{a - m(t_0), h\left[b_1 - m'(t_0)\right], \ldots, h^p\left[p!b_p - m^{(p)}\right]/p!\right\}^T,$$

$$\Delta_i = m(t_i) - m(t_0) - \sum_{k=1}^{p} \frac{1}{k!} m^{(k)}(t_0).$$

With these notations, we rewrite $L_n(\boldsymbol{\beta})$ as follow

$$L_n^*(\boldsymbol{\beta}^*) = \frac{1}{2n(n+1)} \sum_{i \le j} \left| \sigma(t_i)\varepsilon_i + \sigma(t_j)\varepsilon_j + \Delta_i + \Delta_j - \gamma_n(\mathbf{z}_i + \mathbf{z}_j)^T \boldsymbol{\beta}^* \right| K_h(t_i - t_0) K_h(t_j - t_0).$$

Moreover, we use $\boldsymbol{\omega}_n(\boldsymbol{\beta}^*)$ to denote the gradient function of $L_n^*(\boldsymbol{\beta}^*)$. More specifically,

$$\begin{aligned} \boldsymbol{\omega}_n(\boldsymbol{\beta}^*) &= \nabla L_n^*(\boldsymbol{\beta}^*) \\ &= -\frac{\gamma_n}{2n(n+1)} \sum_{i \le j} \mathrm{sgn}[\sigma(t_i)\varepsilon_i + \sigma(t_j)\varepsilon_j + \Delta_i + \Delta_j - \gamma_n(\mathbf{z}_i + \mathbf{z}_j)^T \boldsymbol{\beta}^*] \times (\mathbf{z}_i + \mathbf{z}_j) K_h(t_i - t_0) K_h(t_j - t_0), \end{aligned}$$

where $\mathrm{sgn}(\cdot)$ is the sign function.

In order to prove the theorems, we firstly state a few necessary lemmas.

Lemma 1. *Suppose that Conditions (C1)–(C4) hold, $h \to 0$, and $nh^2 \to \infty$. Then,*

$$\gamma_n^{-1}[\boldsymbol{\omega}_n(\boldsymbol{\beta}^*) - \boldsymbol{\omega}_n(\mathbf{0})] = 4\tau f^2(t_0)\sigma^{-1}(t_0)\gamma_n \mathbf{S}\boldsymbol{\beta}^* + o_p(1).$$

Proof. Let $U_n = \gamma_n^{-1}[\boldsymbol{\omega}_n(\boldsymbol{\beta}^*) - \boldsymbol{\omega}_n(\mathbf{0})] = [2n(n+1)]^{-1} \sum_{i \le j} W_n(\mathbf{d}_i, \mathbf{d}_j)$, where

$$\begin{aligned} W_n(\mathbf{d}_i, \mathbf{d}_j) &= \big\{ \mathrm{sgn}[\sigma(t_i)\varepsilon_i + \sigma(t_j)\varepsilon_j + \Delta_i + \Delta_j] - \mathrm{sgn}[\sigma(t_i)\varepsilon_i + \sigma(t_j)\varepsilon_j + \Delta_i + \Delta_j \\ &\quad - \gamma_n(\mathbf{z}_i + \mathbf{z}_j)^T \boldsymbol{\beta}^*] \big\} \times (\mathbf{z}_i + \mathbf{z}_j) K_h(t_i - t_0) K_h(t_j - t_0), \end{aligned}$$

and $\mathbf{d}_i = (t_i, \epsilon_i)^T$. Clearly, $U_n = [2n(n+1)]^{-1} \sum_{i \le j} W_n(\mathbf{d}_i, \mathbf{d}_j)$ is of the form of U-statistic since $W_n(\cdot, \cdot)$ is symmetric in its arguments. Note that

$$\begin{aligned} E[\|W_n(\mathbf{d}_i, \mathbf{d}_j)\|^2] &\le 4h^{-4} E\left\{ [\mathbf{z}_i + \mathbf{z}_j]^T [\mathbf{z}_i + \mathbf{z}_j] K^2\left(\frac{t_i - t_0}{h}\right) K^2\left(\frac{t_j - t_0}{h}\right) \right\} \\ &= O(h^{-2}) = o(n), \end{aligned}$$

by using the assumption $nh^2 \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm. Thus, $U_n = E[W_n(\mathbf{d}_i, \mathbf{d}_j)] + o_p(1)$ by Lemma A.1 of [18]. Note that

$$\begin{aligned} E[W_n(\mathbf{d}_i, \mathbf{d}_j)] &= -2h^{-2} E\left\{ \int \left[G\left(\frac{\Delta_i + \Delta_j + \epsilon}{\sigma(t_i)}\right) - G\left(\frac{-\gamma_n(\mathbf{z}_i + \mathbf{z}_j)^T \boldsymbol{\beta}^* + \Delta_i + \Delta_j + \epsilon}{\sigma(t_i)}\right) \right] \right. \\ &\quad \left. \times g\left(\frac{\epsilon}{\sigma(t_j)}\right) d\left(\frac{\epsilon}{\sigma(t_j)}\right) [\mathbf{z}_i + \mathbf{z}_j] K\left(\frac{t_i - t_0}{h}\right) K\left(\frac{t_j - t_0}{h}\right) \right\} \\ &= 2h^{-2}\gamma_n E\left\{ \int \sigma^{-1}(t_i) g\left(\frac{\sigma(t_j)\epsilon + \Delta_i + \Delta_j}{\sigma(t_i)}\right) g(\epsilon) d\epsilon [\mathbf{z}_i + \mathbf{z}_j][\mathbf{z}_i + \mathbf{z}_j]^T \right. \\ &\quad \left. \times K\left(\frac{t_i - t_0}{h}\right) K\left(\frac{t_j - t_0}{h}\right) \right\} \boldsymbol{\beta}^*(1 + o(1)) \\ &= 4\tau f^2(t_0)\sigma^{-1}(t_0)\gamma_n \mathbf{S}\boldsymbol{\beta}^* + o(1), \end{aligned}$$

where the last equality is followed by a simple calculation similar to Parzen [14]. □

Moreover, we consider the following quadratic function:

$$B_n(\boldsymbol{\beta}^*) = \gamma_n^{-1} \boldsymbol{\beta}^{*T} \boldsymbol{\omega}_n(\mathbf{0}) + 2\tau f^2(t_0)\sigma^{-1}(t_0)\gamma_n \boldsymbol{\beta}^{*T} \mathbf{S}\boldsymbol{\beta}^* + \gamma_n^{-1} L_n^*(\mathbf{0}).$$

Lemma 2. *Suppose that Conditions (C1)–(C4) hold. Then for any $\varepsilon > 0$ and $c > 0$*

$$P\left\{ \sup_{\|\boldsymbol{\beta}^*\| \le c} \left| \gamma_n^{-1} L^*(\boldsymbol{\beta}^*) - B_n(\boldsymbol{\beta}^*) \right| \ge \varepsilon \right\} \to 0.$$

Proof. By using Lemma 1, we have

$$\nabla[\gamma_n^{-1} L_n^*(\boldsymbol{\beta}^*) - B_n(\boldsymbol{\beta}^*)] = \gamma_n^{-1}[\boldsymbol{\omega}_n(\boldsymbol{\beta}^*) - \boldsymbol{\omega}_n(\mathbf{0})] - 4\tau f^2(t_0)\sigma^{-1}(t_0)\gamma_n \mathbf{S}\boldsymbol{\beta}^* = o_p(1).$$

Using similar arguments of diagonal subsequencing and convexity in the proof of Theorem A.3.7 of [10], we can complete the proof of this lemma. Details are omitted. □

Proof of Theorem 1. Denote $\hat{\boldsymbol\beta}$ be the minimizer of $L_n^*(\boldsymbol\beta^*)$ and $\tilde{\boldsymbol\beta}$ be the minimizer of $B_n(\boldsymbol\beta^*)$. Since the convex function $\gamma_n^{-1}\left[L_n^*(\boldsymbol\beta^*)-L_n^*(\mathbf{0})\right]$ converges in probability to the convex function $2\tau f^2(t_0)\sigma^{-1}(t_0)\gamma_n\boldsymbol\beta^{*T}\mathbf{S}\boldsymbol\beta^*+\gamma_n^{-1}\boldsymbol\beta^{*T}\boldsymbol\omega_n(\mathbf{0})$ (by Lemma 2), it follows from the convexity lemma [15] that,

$$\hat{\boldsymbol\beta}=\tilde{\boldsymbol\beta}+o_p(1)=-\gamma_n^{-2}[4\tau f^2(t_0)\sigma^{-1}(t_0)]^{-1}\mathbf{S}^{-1}\boldsymbol\omega_n(\mathbf{0})+o_p(1). \tag{A.1}$$

We next establish the asymptotic normality of $\hat{\boldsymbol\beta}$. Let us re-express $-\gamma_n^{-2}\boldsymbol\omega_n(0)$ as

$$\begin{aligned}-\gamma_n^{-2}\boldsymbol\omega_n(\mathbf{0}) &= \gamma_n^{-2}\frac{\gamma_n}{n(n+1)}\sum_{i\le j}[I(\sigma(t_i)\varepsilon_i+\sigma(t_j)\varepsilon_j+\Delta_i+\Delta_j>0)-1/2]\zeta_{ij}(t_0)\\ &= \frac{1}{\gamma_n n(n+1)}\sum_{i\le j}[I(\sigma(t_i)\varepsilon_i+\sigma(t_j)\varepsilon_j>0)-1/2]\zeta_{ij}(t_0)\\ &\quad+\frac{1}{\gamma_n n(n+1)}\sum_{i\le j}[I(\sigma(t_i)\varepsilon_i+\sigma(t_j)\varepsilon_j+\Delta_i+\Delta_j>0)-I(\sigma(t_i)\varepsilon_i+\sigma(t_j)\varepsilon_j>0)]\zeta_{ij}(t_0)\\ &:=\boldsymbol\omega_{n1}(\mathbf{0})+\boldsymbol\omega_{n2}(\mathbf{0}),\end{aligned}$$

where we denote $\zeta_{ij}(t_0)=(\boldsymbol z_i+\boldsymbol z_j)K_h(t_i-t_0)K_h(t_j-t_0)$ for simplicity.

We first prove that

$$\boldsymbol\omega_{n1}(\mathbf{0})\to N_{p+1}\left(\mathbf{0},\frac{4}{3}f^3(t_0)\upsilon_0\mathbf{S}^*\right),$$

in distribution as $n\to\infty$. Note that we can write $\boldsymbol\omega_{n1}(\mathbf{0})=\sqrt{n}[n(n+1)]^{-1}\sum_{i\le j}H_n(\boldsymbol d_i,\boldsymbol d_j)$, where

$$H_n(\boldsymbol d_i,\boldsymbol d_j)=h^{-3/2}[I(\sigma(t_i)\varepsilon_i+\sigma(t_j)\varepsilon_j>0)-1/2]\zeta_{ij}(t_0).$$

Similarly to the arguments in the proof of Lemma 1, it can be shown that $E\|H_n(\boldsymbol d_i,\boldsymbol d_j)\|^2=o(n)$. Furthermore, according to the symmetric of the error distribution, we have $E[H_n(\boldsymbol d_i,\boldsymbol d_j)]=0$. This implies that $\boldsymbol\omega_{n1}=n^{-1}\sum_{i=1}^n r_n(\boldsymbol d_i)+o_p(1)$, where $r_n(\boldsymbol d_i)=E[H_n(\boldsymbol d_i,\boldsymbol d_j)\mid \boldsymbol d_i]$. Then

$$\begin{aligned}r_n(\boldsymbol d_i) &= E[H_n(\boldsymbol d_i,\boldsymbol d_j)\mid\boldsymbol d_i]\\ &= h^{-3/2}\left[G\left(\frac{\sigma(t_i)\varepsilon_i}{\sigma(t_j)}\right)-1/2\right]K\left(\frac{t_i-t_0}{h}\right)\times E\left[(\boldsymbol z_i+\boldsymbol z_j)K\left(\frac{t_j-t_0}{h}\right)\mid\boldsymbol d_i\right]\\ &= h^{-1/2}\left[G\left(\frac{\sigma(t_i)\varepsilon_i}{\sigma(t_j)}\right)-1/2\right]K\left(\frac{t_i-t_0}{h}\right)\\ &\quad\times\left[\left(\int K(t)f(t_0+th)dt\right)\boldsymbol z_i+\int E(\boldsymbol z_j\mid t_j=t_0+th)K(t)f(t_0+th)dt\right].\end{aligned}$$

Furthermore,

$$\begin{aligned}E[r_n(\boldsymbol d_i)r_n(\boldsymbol d_i)^T] &= \int\left[G\left(\frac{\epsilon}{\sigma(t_j)}\right)-1/2\right]^2dG\left(\frac{\epsilon}{\sigma(t_i)}\right)h^{-1}E\Bigg\{K^2\left(\frac{t_i-t_0}{h}\right)\\ &\quad\times\left[\left(\int K(t)f(t_0+th)dt\right)\boldsymbol z_i+\int E(\boldsymbol z_j\mid t_j=t_0+th)K(t)f(t_0+th)dt\right]\\ &\quad\times\left[\left(\int K(t)f(t_0+th)dt\right)\boldsymbol z_i^T+\int E(\boldsymbol z_j^T\mid t_j=t_0+th)K(t)f(t_0+th)dt\right]\Bigg\}\\ &\to\frac{4}{3}f^3(t_0)\upsilon_0\mathbf{S}^*.\end{aligned}$$

By applying the Lindeberg–Feller central limit theorem, we can obtain the asymptotic normality of $\boldsymbol\omega_{n1}(\mathbf{0})$.

Next, we show that

$$\boldsymbol\omega_{n2}(\mathbf{0})=\frac{4h^{p+1}}{\gamma_n}[\tau f^2(t_0)\sigma^{-1}(t_0)\boldsymbol c_p m^{(p+1)}(t_0)+o(1)]+o_p(1),$$

when $p-r$ is odd. Similarly to the above procedure, we rewrite

$$\boldsymbol\omega_{n2}(\mathbf{0})=[n(n+1)]^{-1}\sum_{i\le j}H_n^*(\boldsymbol d_i,\boldsymbol d_j),$$

where

$$H_n^*(\boldsymbol{d}_i, \boldsymbol{d}_j) = nh^{-1}\gamma_n[I(\sigma(t_i)\varepsilon_i + \sigma(t_j)\varepsilon_j + \Delta_i + \Delta_j > 0) - I(\sigma(t_i)\varepsilon_i + \sigma(t_j)\varepsilon_j > 0)]\zeta_{ij}(t_0).$$

Note that

$$\begin{aligned}\Delta_i + \Delta_j = &\frac{1}{(p+1)!}((t_i - t_0)^{p+1} + (t_j - t_0)^{p+1})m^{(p+1)}(t_0)\\ &+ \frac{1}{(p+2)!}((t_i - t_0)^{p+2} + (t_j - t_0)^{p+2})m^{(p+2)}(t_0) + o((t_i - t_0)^{p+2}) + o((t_j - t_0)^{p+2}).\end{aligned}$$

Similarly, it can be shown that $\boldsymbol{\omega}_{n2}(\mathbf{0}) = E[H_n^*(\boldsymbol{d}_i, \boldsymbol{d}_j)] + o_p(1)$. Furthermore,

$$\begin{aligned}E[H_n^*(\boldsymbol{d}_i, \boldsymbol{d}_j)] &= nh^{-1}\gamma_n E\left\{\int \left[G\left(\frac{\epsilon + \Delta_i + \Delta_j}{\sigma(t_i)}\right) - G\left(\frac{\epsilon}{\sigma(t_i)}\right)\right] g\left(\frac{\epsilon}{\sigma(t_j)}\right) d\left(\frac{\epsilon}{\sigma(t_j)}\right) \times \zeta_{ij}(t_0)\right\}\\ &= nh^{-1}\gamma_n[\tau + O(h)]E\left\{\sigma^{-1}(t_i)(\Delta_i + \Delta_j) \times \zeta_{ij}(t_0)\right\}(1 + o(1))\\ &= \frac{4h^{p+1}}{\gamma_n}[\tau f^2(t_0)\sigma^{-1}(t_0)\mathbf{c}_p m^{(p+1)}(t_0) + o(1)] + o_p(1).\end{aligned}$$

Finally, by a similar procedure, we also have

$$\boldsymbol{\omega}_{n2}(\mathbf{0}) = \frac{4h^{p+2}}{\gamma_n}\left\{\tau f^2(t_0)\sigma^{-1}(t_0)\tilde{\mathbf{c}}_p\left[m^{(p+2)}(t_0) + (p+2)m^{(p+1)}(t_0)\frac{f'(t_0)}{f(t_0)}\right] + o(1)\right\} + o_p(1)$$

when $p - r$ is even. By using the approximation given in (A.1), we obtain the theorem. □

Proof of Corollary 1. In local linear Walsh-average estimation, $m(t) \approx m(t_0) + m'(t_0)(t - t_0)$ is used. Note that when $p = 1$,

$$\mathbf{S} = \begin{pmatrix} 2 & 0 \\ 0 & \mu_2 \end{pmatrix}, \qquad \mathbf{S}^* = \begin{pmatrix} 4\upsilon_0 & 0 \\ 0 & \upsilon_2 \end{pmatrix}$$

so, by Theorem 1,

$$\mathrm{var}\{\hat{m}_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{nh}\frac{\upsilon_0\sigma^2(t_0)}{12\tau^2 f(t_0)} + o_p\left(\frac{1}{nh}\right).$$

Furthermore, note that $\mathbf{c}_1 = (\mu_2, \mu_4 + \mu_2)$, then

$$\mathrm{bias}\{\hat{m}_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{2}m''(t_0)\mu_2 h^2 + o_p(h^2).$$

Thus, the corollary is immediately obtained. □

Proof of Corollary 2. For general purpose, we highlight here the results of $\hat{m}'_{\mathrm{WA}}(t_0)$ without the symmetric distribution assumption. Recall the notations δ and ξ defined in Remark 2. Without the symmetric error distribution, taking the same procedure as in the proof of Theorem 1, we have

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + o_p(1) = -\gamma_n^{-2}[4\delta f^2(t_0)\sigma^{-1}(t_0)]^{-1}\mathbf{S}^{-1}\boldsymbol{\omega}_n(\mathbf{0}) + o_p(1) \tag{A.2}$$

and

$$\begin{aligned}&\boldsymbol{\omega}_{n1}(\mathbf{0}) \to N_3\left(\mathbf{c}, 16\xi f^3(t_0)\upsilon_0\mathbf{S}^*\right),\\ &\boldsymbol{\omega}_{n2}(\mathbf{0}) = \frac{4h^3}{\gamma_n}[\delta f^2(t_0)\mathbf{c}_p m'''(t_0) + o(1)] + o_p(1),\end{aligned}$$

where $\mathbf{c} = E[\boldsymbol{\omega}_{n1}(\mathbf{0})] = (a, 0, b)^T$, and a, b are some constants. Note that when $p = 2$,

$$\mathbf{S} = \begin{pmatrix} 2 & 0 & 2\mu_2 \\ 0 & \mu_2 & 0 \\ 2\mu_2 & 0 & \mu_4 + \mu_2^2 \end{pmatrix}, \qquad \mathbf{S}^* = \begin{pmatrix} 4\upsilon_0 & 0 & 2\upsilon_2 + 2\mu_2\upsilon_0 \\ 0 & \upsilon_2 & 0 \\ 2\upsilon_2 + 2\mu_2\upsilon_0 & 0 & \upsilon_4 + 2\mu_2\upsilon_2 + \mu_2^2\upsilon_0, \end{pmatrix}$$

and $\mathbf{c_2} = (0, \mu_4, 0)^T$. By (A.2) and simple calculations, we have

$$\begin{aligned}&\mathrm{bias}\{\hat{m}'_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{6}m'''(t_0)\frac{\mu_4}{\mu_2}h^2 + o_p(h^2),\\ &\mathrm{var}\{\hat{m}'_{\mathrm{WA}}(t_0) \mid \mathbf{T}\} = \frac{1}{nh^3}\frac{\xi\upsilon_2\sigma^2(t_0)}{\delta^2\mu_2^2 f(t_0)} + o_p\left(\frac{1}{nh^3}\right).\end{aligned}$$

On the other hand, when the error distribution is symmetric, $\delta = \tau$ and $\xi = 1/12$, and then the corollary follows from Theorem 1. $\square$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at doi:10.1016/j.jmva.2011.12.003.

References

[1] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Amer. Statist. Assoc. 74 (1979) 829–836.
[2] J. Fan, Local linear regression smoothers and their minimax efficiencies, Ann. Statist. 21 (1993) 196–216.
[3] J. Fan, I. Gijbels, Design-adaptive nonparametric regression, J. Amer. Statist. Assoc. 87 (1992) 998–1004.
[4] J. Fan, I. Gijbels, Local Polynomial Modeling and its Applications, Chapman and Hall, London, 1996.
[5] J. Fan, P. Hall, On cureve estimation by minimizing mean absolute deviation and its implications, Ann. Statist. 22 (1994) 867–885.
[6] J. Fan, T. Hu, Y. Truong, Robust non-parametric function estimation, Scandinavian J. Statist. 21 (1994) 433–446.
[7] J. Fan, T. Huang, Profile likelihood inferences on semiparametric varying-coefficient partially linear models, Bernoulli 11 (2005) 1031–1057.
[8] J. Fan, J. Jiang, Variable bandwidth and one-step local M-estimator, Sci. China Ser. A 43 (2000) 65–81.
[9] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, New York, 1990.
[10] T.P. Hettmansperger, J.W. McKean, Robust Nonparametric Statistical Methods, Arnold, London, 1998.
[11] J.L. Hodges Jr., E.L. Lehmann, Estimates of location based on rank tests, Ann. Math. Statist. 34 (1963) 598–611.
[12] P.J. Huber, Robust Estimation, Wiley, New York, 1981.
[13] B. Kai, R. Li, H. Zou, Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression, J. R. Stat. Soc. Ser. B 72 (2010) 49–69.
[14] E. Parzen, On estimation of a probability density function and model, Ann. Math. Statist. 33 (1962) 1065–1076.
[15] D. Pollard, Asymptotics for least absolute deviation regression estimators, Econom. Theory 7 (1991) 186–199.
[16] D. Ruppert, S.J. Sheather, M.P. Wand, An effective bandwidth selector for local least squares regression, J. Amer. Statist. Assoc. 90 (1995) 1257–1270.
[17] J. Terpstra, J. McKean, Rank-based analysis of linear models using R, J. Stat. Softw. 14 (2005) 1–26.
[18] L. Wang, B. Kai, R. Li, Local rank inference for varying coefficient models, J. Amer. Statist. Assoc. 488 (2009) 1631–1645.
[19] H. Wang, Y. Xia, Shrinkage estimation of the varying coefficient model, J. Amer. Statist. Assoc. 104 (2009) 747–757.
[20] A.H. Welsh, Robust estimation of smooth regression and spread functions and their derivatives, Statist. Sinica 6 (1996) 347–366.
[21] H. Zou, M. Yuan, Composite quantile regression and the oracle model selection theory, Ann. Statist. 36 (2007) 1108–1126.