Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2010)
72, *Part* 1, *pp.* 49–69

Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression

Bo Kai and Runze Li

Pennsylvania State University, University Park, USA

and Hui Zou

University of Minnesota, Minneapolis, USA

[Received September 2008. Revised March 2009]

Summary. Local polynomial regression is a useful non-parametric regression tool to explore fine data structures and has been widely used in practice. We propose a new non-parametric regression technique called *local composite quantile regression smoothing* to improve local polynomial regression further. Sampling properties of the estimation procedure proposed are studied. We derive the asymptotic bias, variance and normality of the estimate proposed. The asymptotic relative efficiency of the estimate with respect to local polynomial regression is investigated. It is shown that the estimate can be much more efficient than the local polynomial regression estimate for various non-normal errors, while being almost as efficient as the local polynomial regression estimate for normal errors. Simulation is conducted to examine the performance of the estimates proposed. The simulation results are consistent with our theoretical findings. A real data example is used to illustrate the method proposed.

Keywords: Asymptotic efficiency; Composite quantile regression estimator; Kernel function; Local polynomial regression; Non-parametric regression

1. Introduction

Consider the general non-parametric regression model

$$Y = m(T) + \sigma(T)\varepsilon, \qquad (1.1)$$

where Y is the response variable, T is a covariate, $m(T) = E(Y|T)$, which is assumed to be a smooth non-parametric function, and $\sigma(T)$ is a positive function representing the standard deviation. We assume that ε has mean 0 and variance 1. Local polynomial regression is a popular and successful method for non-parametric regression, and it has been well studied in the literature (Fan and Gijbels, 1996). By locally fitting a linear (or polynomial) regression model via adaptively weighted least squares, local polynomial regression can explore the fine features of the regression function and its derivatives. Although the least squares method is a popular and convenient choice in local polynomial fitting, we may consider the use of various local fitting methods. For example, in the presence of outliers, we may consider local least absolute deviation (LAD) polynomial regression (Fan *et al.*, 1994; Welsh, 1996). When the error follows a Laplacian distribution, the local LAD polynomial regression is more efficient than local least

Address for correspondence: Hui Zou, School of Statistics, University of Minnesota, 224 Church Street SE, Minneapolis, MN 55455, USA.
E-mail: hzou@stat.umn.edu

 1369–7412/10/72049

squares polynomial regression. Of course, local LAD polynomial regression can do much worse than local least squares polynomial regression in other different settings. The aim of this paper is to develop a new local estimation procedure that can significantly improve on classical local polynomial regression for a wide class of error distributions and has comparable efficiency in the worst case scenario.

Our proposal is built on the composite quantile regression (CQR) estimator that has recently been proposed by Zou and Yuan (2008) for estimating the regression coefficients in the classical linear regression model. Zou and Yuan (2008) showed that the relative efficiency of the CQR estimator compared with the least squares estimator is greater than 70% regardless of the error distribution. Furthermore, the CQR estimator could be much more efficient and sometimes arbitrarily more efficient than the least squares estimator. These nice theoretical properties of CQR in linear regression motivates us to construct local CQR smoothers as non-parametric estimates of the regression function and its derivatives.

We make several contributions in this paper.

(a) We propose the local linear CQR estimator for estimating the non-parametric regression function. We establish asymptotic theory for the local linear CQR estimator and show that, compared with the classical local linear least squares estimator, the new method can significantly improve the estimation efficiency of the local linear least squares estimator for commonly seen non-normal error distributions.
(b) We propose the local quadratic CQR estimator for estimating the derivative of the regression function. Asymptotic theory shows that the local quadratic CQR estimator can often drastically improve the estimation efficiency of its local least squares counterpart if the error distribution is non-normal and, at the same time, the loss in efficiency is at most 8.01% in the worst case scenario.
(c) General asymptotic theory for the local p-polynomial CQR estimator is established. Our theory does not require the error distribution to have a finite variance. Therefore, local CQR estimators can work well even when local polynomial regression fails because of the infinite variance in the noise.

It is a well-known fact that the local linear (polynomial) regression is the best linear smoother in terms of efficiency (Fan and Gijbels, 1996). There is no contradiction between this fact and our results, because the local CQR estimator proposed is a *non-linear* smoother.

The rest of this paper is organized as follows. In Section 2, we introduce local linear CQR for non-parametric regression and study its asymptotic properties. In Section 3, we propose local quadratic CQR for estimating the derivative of the non-parametric regression, which can further reduce the estimation bias by the local linear CQR. Monte Carlo study and a real data example are presented in Section 4. In Section 5 we present general theory for the local p-polynomial CQR and technical proofs.

2. Estimation of regression function

Suppose that (t_i, y_i), $i=1,\ldots,n$, is an independent and identically distributed random sample. Consider estimating the value of $m(T)$ at t_0. In local linear regression we first approximate $m(t)$ locally by a linear function $m(t)\approx m(t_0)+m'(t_0)(t-t_0)$ and then fit a linear model locally in a neighbourhood of t_0. Let $K(\cdot)$ be a smooth kernel function; the local linear regression estimator of $m(t_0)$ is $\hat{a}$, where

$$(\hat{a},\hat{b})=\arg\min_{a,b}\left[\sum_{i=1}^{n}\{y_i-a-b(t_i-t_0)\}^2K\left(\frac{t_i-t_0}{h}\right)\right], \tag{2.1}$$

where h is the smoothing parameter. Local linear regression enjoys many good theoretical properties, such as its design adaptation property and high minimax efficiency (Fan and Gijbels, 1992). However, local least squares regression breaks down when the error distribution does not have finite second moment, for the estimator is no longer consistent. Local LAD polynomial regression (Fan *et al.*, 1994; Welsh, 1996) replaces the least squares loss in equation (2.1) with the L_1-loss. By doing so, the local LAD estimator can deal with the infinite variance case, but for finite variance cases its relative efficiency compared with the local least squares estimator can be arbitrarily small.

We propose the local linear CQR estimator as an efficient alternative to the local linear regression estimator. Let $\rho_{\tau_k}(r) = \tau_k r - r\, I(r<0)$, $k=1,2,\ldots,q$, be q check loss functions at q quantile positions: $\tau_k = k/(q+1)$. In the linear regression model the CQR loss is defined as (Zou and Yuan, 2008)

$$\sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - a_k - bt_i).$$

CQR combines the strength across multiple quantile regressions with forcing a single parameter for the 'slope'. Since the non-parametric function is approximated by a linear model locally, we consider minimizing the locally weighted CQR loss

$$\sum_{k=1}^{q} \left[\sum_{i=1}^{n} \rho_{\tau_k}\{y_i - a_k - b(t_i - t_0)\} K\left(\frac{t_i - t_0}{h}\right) \right]. \tag{2.2}$$

Denote the minimizer of expression (2.2) by $(\hat{a}_1, \ldots, \hat{a}_q, \hat{b})$. Then we let

$$\begin{aligned} \hat{m}(t_0) &= \frac{1}{q} \sum_{k=1}^{q} \hat{a}_k, \\ \tilde{m}'(t_0) &= \hat{b}. \end{aligned} \tag{2.3}$$

We refer to $\hat{m}(t_0)$ as the local linear CQR estimator of $m(t_0)$. As an estimator of $m'(t_0)$, $\tilde{m}'(t_0)$ can be further improved by using the local quadratic CQR estimator which is discussed in the next section.

Remark 1. It is worth mentioning here that, although the check loss function is typically used to estimate the conditional quantile function of y given T (see Koenker (2005) and references therein), we simultaneously employ several check functions to estimate the regression (mean) function. So the local CQR smoother is conceptually different from non-parametric quantile regression by local fitting which has been studied in Yu and Jones (1998) and chapter 5 of Fan and Gijbels (1996).

Remark 2. In a short note Koenker (1984) studied the Hogg estimator as the minimizer of the weighted sum of check functions in the framework of parametric linear models. The focus there was to argue that the Hogg estimator is a different way to do L-estimation. The CQR loss can be regarded as a weighted sum of check functions with uniform weights and uniform quantiles $(\tau_k = k/(q+1), k=1,2,\ldots,q)$. When q is large, such a choice leads to nice oracle-like estimators in the oracle model selection theoretic framework (Zou and Yuan, 2008). Koenker (1984) did not discuss relative efficiency of the Hogg estimator relative to the least squares estimator. In this work we consider minimizing the locally weighted CQR loss and show that the local CQR smoothers have very interesting asymptotic efficiency properties. To our best knowledge, none of these has been studied in the literature.

2.1. Asymptotic properties

To see why local linear CQR is an efficient alternative to local linear regression, we establish the asymptotic properties of the local linear CQR estimator. Some notation is necessary for the discussion. Let $F(\cdot)$ and $f(\cdot)$ denote the density function and cumulative distribution function of the error distribution respectively. Denote by $f_T(\cdot)$ the marginal density function of the covariate T. We choose the kernel $K(\cdot)$ as a symmetric density function and let

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

Define

$$R_1(q) = \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\tau_{kk'}}{f(c_k)\, f(c_{k'})}, \tag{2.4}$$

where $c_k = F^{-1}(\tau_k)$ and $\tau_{kk'} = \tau_k \wedge \tau_{k'} - \tau_k \tau_{k'}$. In the following theorem, we present the asymptotic bias, variance and normality of $\hat{m}(t_0)$, whose proof is given in Section 5. Let $\mathbf{T}$ be the σ-field that is generated by $\{T_1, \ldots, T_n\}$.

Theorem 1. Suppose that t_0 is an interior point of the support of $f_T(\cdot)$. Under the regularity conditions (a)–(d) in Section 5, if $h \to 0$ and $nh \to \infty$, then the asymptotic conditional bias and variance of the local linear CQR estimator $\hat{m}(t_0)$ are given by

$$\text{bias}\{\hat{m}(t_0)|\mathbf{T}\} = \tfrac{1}{2} m''(t_0) \mu_2 h^2 + o_p(h^2), \tag{2.5}$$

$$\text{var}\{\hat{m}(t_0)|\mathbf{T}\} = \frac{1}{nh} \frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)} R_1(q) + o_p\left(\frac{1}{nh}\right). \tag{2.6}$$

Furthermore, conditioning on $\mathbf{T}$, we have

$$\sqrt{(nh)}\{\hat{m}(t_0) - m(t_0) - \tfrac{1}{2} m''(t_0) \mu_2 h^2\} \xrightarrow{\mathcal{L}} N\left\{0, \frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)} R_1(q)\right\}, \tag{2.7}$$

where $\rightarrow^{\mathcal{L}}$ stands for convergence in distribution.

Remark 3. In the proof that is given in Section 5 we assume that the error distribution is symmetric. Without such a condition the asymptotic bias will have a non-vanishing term. The asymptotic variance remains the same and the asymptotic normality still holds with a minor modification. In other words, the symmetric error distribution condition is only used to ensure that the quantity to which the local CQR estimator converges is the conditional mean function. This is similar to the situation when using the local LAD to estimate the conditional mean function. For that we need to assume that the mean and median of the error distribution coincide.

We see from theorem 1 that the leading term of the asymptotic bias for the local linear CQR estimator is the same as that for the local linear least squares estimator, whereas their asymptotic variances are different. The mean-squared error (MSE) of $\hat{m}(t_0)$ is

$$\text{MSE}\{\hat{m}(t_0)\} = \left\{\tfrac{1}{2} m''(t_0) \mu_2\right\}^2 h^4 + \frac{1}{nh} \frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)} R_1(q) + o_p\left(h^4 + \frac{1}{nh}\right).$$

By straightforward calculations we can see that the optimal variable bandwidth minimizing the asymptotic MSE of $\hat{m}(t_0)$ is

$$h^{\text{opt}}(t_0) = \left[\frac{\nu_0 \sigma^2(t_0)\, R_1(q)}{f_T(t_0) \{m''(t_0) \mu_2\}^2}\right]^{1/5} n^{-1/5}.$$

In practice, we may select a constant bandwidth by minimizing the mean integrated squared error $\text{MISE}(\hat{m}) = \int \text{MSE}\{\hat{m}(t_0)\}\, w(t)\,\mathrm{d}t$ for a weight function $w(t)$. Similarly, the optimal bandwidth minimizing the asymptotic $\text{MISE}(\hat{m})$ is

$$h^{\text{opt}} = \left\{ \frac{\nu_0\, R_1(q) \int \sigma^2(t)\, f_T^{-1}(t)\, w(t)\,\mathrm{d}t}{\mu_2^2 \int m''(t)^2 w(t)\,\mathrm{d}t} \right\}^{1/5} n^{-1/5}.$$

These calculations indicate that the local linear CQR estimator enjoys the optimal rate of convergence $n^{2/5}$.

2.2. Asymptotic relative efficiency

In this section, we study the asymptotic relative efficiency of the local linear CQR estimator with respect to the local linear least squares estimator by comparing their MSEs. The role of R_1 becomes clear in the relative efficiency study.

The local linear least squares estimator for $m(t_0)$ has MSE

$$\text{MSE}\{\hat{m}_{\text{LS}}(t_0)\} = \left\{ \frac{1}{2} m''(t_0)\mu_2 \right\}^2 h^4 + \frac{1}{nh} \frac{\nu_0}{f_T(t_0)} \sigma^2(t_0) + o_p\left(h^4 + \frac{1}{nh} \right),$$

and hence

$$h_{\text{LS}}^{\text{opt}}(t_0) = \left[\frac{\nu_0\, \sigma^2(t_0)}{f_T(t_0)\{m''(t_0)\mu_2\}^2} \right]^{1/5} n^{-1/5},$$

$$h_{\text{LS}}^{\text{opt}} = \left\{ \frac{\nu_0 \int \sigma^2(t)\, f_T^{-1}(t)\, w(t)\,\mathrm{d}t}{\mu_2^2 \int m''(t)^2\, w(t)\,\mathrm{d}t} \right\}^{1/5} n^{-1/5},$$

where $h_{\text{LS}}^{\text{opt}}(t_0)$ is the optimal variable bandwidth minimizing the asymptotic MSE and $h_{\text{LS}}^{\text{opt}}$ is the optimal bandwidth minimizing the asymptotic MISE. Therefore, we have

$$h^{\text{opt}}(t_0) = R_1(q)^{1/5}\, h_{\text{LS}}^{\text{opt}}(t_0), \qquad h^{\text{opt}} = R_1(q)^{1/5} h_{\text{LS}}^{\text{opt}}. \tag{2.8}$$

We use MSE_{opt} and MISE_{opt} to denote the MSE and MISE evaluated at their optimal bandwidth. Then by straightforward calculations we see that, as n approaches ∞,

$$\frac{\text{MSE}_{\text{opt}}\{\hat{m}_{\text{LS}}(t_0)\}}{\text{MSE}_{\text{opt}}\{\hat{m}(t_0)\}} \to R_1(q)^{-4/5},$$

$$\frac{\text{MISE}_{\text{opt}}\{\hat{m}_{\text{LS}}\}}{\text{MISE}_{\text{opt}}\{\hat{m}\}} \to R_1(q)^{-4/5}.$$

Thus, it is natural to define $\text{ARE}(\hat{m}, \hat{m}_{\text{LS}})$, the asymptotic relative efficiency of the local linear CQR estimator with respect to the local linear least squares estimator, as

$$\text{ARE}(\hat{m}, \hat{m}_{\text{LS}}) = R_1(q)^{-4/5}. \tag{2.9}$$

ARE depends only on the error distribution, although the dependence could be rather complex. However, for many commonly seen error distributions, we can directly compute the value of ARE. Table 1 displays $\text{ARE}(\hat{m}, \hat{m}_{\text{LS}})$ for some commonly seen error distributions.

Table 1. Comparisons of ARE($\hat{m}$, $\hat{m}_{LS}$)

Error distribution	*ARE($\hat{m}$, $\hat{m}_{LS}$) for the following values of q:*				
	$q=1$	$q=5$	$q=9$	$q=19$	$q=99$
$N(0,1)$	0.6968	0.9339	0.9659	0.9858	0.9980
Laplace	1.7411	1.2199	1.1548	1.0960	1.0296
t-distribution with 3 degrees of freedom	1.4718	1.5967	1.5241	1.4181	1.2323
t-distribution with 4 degrees of freedom	1.0988	1.2652	1.2377	1.1872	1.0929
$0.95\,N(0,1)+0.05\,N(0,3^2)$	0.8639	1.1300	1.1536	1.1540	1.0804
$0.90\,N(0,1)+0.10\,N(0,3^2)$	0.9986	1.2712	1.2768	1.2393	1.0506
$0.95\,N(0,1)+0.05\,N(0,10^2)$	2.6960	3.4577	3.4783	3.3591	1.3498
$0.90\,N(0,1)+0.10\,N(0,10^2)$	4.0505	4.9128	4.7049	3.5444	1.1379

Several interesting observations can be made from Table 1. Firstly, when the error distribution is $N(0,1)$ for which the local linear least squares estimator is expected to have the best performance, $\text{ARE}(\hat{m},\hat{m}_{LS})$ is very close to 1 regardless of the choice of q in the local linear CQR estimator. When $q=5$ the local linear CQR loses only at most 7% efficiency, whereas it performs as well as the local linear least squares estimator when $q=99$. Secondly, for all the other non-normal distributions that are listed in Table 1, the local linear CQR estimator can have higher efficiencies than the local linear least squares estimator when a small q is used. The mixture of two normal distributions is often used to model so-called contaminated data. For such distributions, $\text{ARE}(\hat{m},\hat{m}_{LS})$ can be as large as 4.9 and even more. Table 1 also indicates that, except for the Laplace error, the local CQRs with $q=5$ or $q=9$ are significantly better than that with $q=1$, which becomes the local LAD regression for these distributions. Finally, we observe that the ARE-values for a variety of distributions are very close to 1 when q is large ($q=99$). It turns out that this is true in general, as demonstrated in the following theorem.

Theorem 2. $\lim_{q\to\infty}\{R_1(q)\}=1$, and thus $\lim_{q\to\infty}\{\text{ARE}(\hat{m},\hat{m}_{LS})\}=1$.

Theorem 2 provides us insights into the asymptotic behaviour of the local linear CQR estimator and implies that the local linear CQR estimator is a safe competitor against the local linear least squares estimator, for it will not lose efficiency when using a large q. However, substantial gain in efficiency could be achieved by using a relatively small q such as $q=9$, as shown in Table 1.

3. Estimation of derivative

In many situations we are interested in estimating the derivative of $m(t)$. The local linear CQR also provides an estimator $\tilde{m}'(t_0)$ for the derivative of $m(t)$. The asymptotic bias and variance of the estimate $\tilde{m}'(t_0)$ in expression (2.3) are given in expressions (5.8) and (5.9) in Section 5. The local linear CQR estimator and the local linear regression estimator have the same leading bias term which depends on the intrinsic part $m'''(t_0)$ and the extra part $m''(t_0)\, f_T'(t_0)/f_T(t_0)$. In Chu and Marron (1991) and Fan (1992), it was already argued that the bias could be very large in many situations. So it may not be an ideal estimator because of the relatively large bias. Local quadratic regression is often preferred for estimating the derivative function, since it reduces the estimation bias without increasing the estimation variance (Fan and Gijbels, 1992). We show here that the same is true in local CQR smoothing.

We consider the local quadratic approximation of $m(t)$ in the neighbourhood of t_0: $m(t) \approx m(t_0) + m'(t_0)(t - t_0) + \frac{1}{2} m''(t_0)(t - t_0)^2$. Let $\mathbf{a} = (a_1, \ldots, a_q)$ and $\mathbf{b} = (b_1, b_2)$. We solve

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \arg\min_{\mathbf{a},\mathbf{b}} \left(\sum_{i=1}^{n} \left[\sum_{k=1}^{q} \rho_{\tau_k} \left\{ y_i - a_k - b_1(t_i - t_0) - \frac{1}{2} b_2 (t_i - t_0)^2 \right\} K\left(\frac{t_i - t_0}{h} \right) \right] \right). \tag{3.1}$$

Then the local quadratic CQR estimator for $m'(t_0)$ is given by

$$\hat{m}'(t_0) = \hat{b}_1. \tag{3.2}$$

3.1. Asymptotic properties

Denote

$$R_2(q) = \left(\sum_{k=1}^{q} \sum_{k'=1}^{q} \tau_{kk'} \right) \bigg/ \left\{ \sum_{k=1}^{q} f(c_k) \right\}^2. \tag{3.3}$$

The asymptotic bias, variance and normality are given in the following theorem.

Theorem 3. Suppose that t_0 is an interior point of the support of $f_T(\cdot)$. Under the regularity conditions (a)–(d) in Section 5, if $h \to 0$ and $nh^3 \to \infty$, then the asymptotic conditional bias and variance of $\hat{m}'(t_0)$, which are defined in equation (3.2), are given by

$$\mathrm{bias}\{\hat{m}'(t_0)|\mathbf{T}\} = \frac{1}{6} m'''(t_0) \frac{\mu_4}{\mu_2} h^2 + o_p(h^2), \tag{3.4}$$

$$\mathrm{var}\{\hat{m}'(t_0)|\mathbf{T}\} = \frac{1}{nh^3} \frac{\nu_2 \sigma^2(t_0)}{\mu_2^2 f_T(t_0)} R_2(q) + o_p\left(\frac{1}{nh^3} \right). \tag{3.5}$$

Furthermore, conditioning on $\mathbf{T}$, we have the following asymptotic normal distribution:

$$\surd(nh^3) \left\{ \hat{m}'(t_0) - m'(t_0) - \frac{1}{6} m'''(t_0) \frac{\mu_4}{\mu_2} h^2 \right\} \xrightarrow{\mathcal{L}} N\left\{ 0, \frac{\nu_2 \sigma^2(t_0)}{\mu_2^2 f_T(t_0)} R_2(q) \right\}. \tag{3.6}$$

Remark 4. In theorem 3 the symmetric error distribution assumption is used to obtain the asymptotic bias formula. Without that assumption, the asymptotic variance remains the same and the asymptotic normality still holds with a minor modification. It is also interesting to point out that when the variance function is homoscadastic the symmetric error distribution assumption is no longer needed for theorem 3.

Comparing equations (5.8) and (3.4), we see that the extra part $m''(t_0) f_T'(t_0)/f_T(t_0)$ is removed in the local quadratic CQR estimator. Comparing the local quadratic CQR and the local quadratic least squares estimators for $m'(t_0)$, we see that they have the same leading bias term, whereas their asymptotic variances are different.

From theorem 3, the MSE of the local quadratic CQR estimator $\hat{m}'(t_0)$ is given by

$$\mathrm{MSE}\{\hat{m}'(t_0)\} = \left\{ \frac{1}{6} m'''(t_0) \frac{\mu_4}{\mu_2} \right\}^2 h^4 + \frac{1}{nh^3} \frac{\nu_2 \sigma^2(t_0)}{\mu_2^2 f_T(t_0)} R_2(q) + o_p\left(h^4 + \frac{1}{nh^3} \right).$$

Thus, the optimal variable bandwidth minimizing $\mathrm{MSE}\{\hat{m}'(t_0)\}$ is

$$h^{\mathrm{opt}}(t_0) = R_2(q)^{1/7} \left[\frac{27 \nu_2 \sigma^2(t_0)}{f_T(t_0) \{ m'''(t_0) \mu_4 \}^2} \right]^{1/7} n^{-1/7}.$$

Furthermore, we consider the mean integrated squared error $\text{MISE}(\hat{m}') = \int \text{MSE}\{\hat{m}'(t)\}\, w(t)\, dt$ with a weight function $w(t)$. The optimal constant bandwidth minimizing the mean integrated squared error is given by

$$h^{\text{opt}} = R_2(q)^{1/7} \left\{ \frac{27\nu_2 \int \sigma^2(t)\, f_T^{-1}(t)\, w(t)\, dt}{\int m'''(t)^2\, w(t)\, dt \mu_4^2} \right\}^{1/7} n^{-1/7}.$$

These calculations indicate that the local quadratic CQR estimator enjoys the optimal rate of convergence $n^{2/7}$.

3.2. Asymptotic relative efficiency

In what follows we study the asymptotic relative efficiency of the local quadratic CQR estimator with respect to the local quadratic least squares estimator. Note that the MSE of local quadratic least squares estimator $\hat{m}'_{\text{LS}}(t_0)$ is given by

$$\text{MSE}\{\hat{m}_{\text{LS}}(t_0)\} = \left\{ \frac{1}{6} m'''(t_0) \frac{\mu_4}{\mu_2} \right\}^2 h^4 + \frac{1}{nh^3} \frac{\nu_2 \sigma^2(t_0)}{\mu_2^2 f_T(t_0)} + o_p\left(h^4 + \frac{1}{nh^3} \right),$$

and the mean integrated squared error is $\text{MISE}(\hat{m}'_{\text{LS}}) = \int \text{MSE}\{\hat{m}'_{\text{LS}}(t)\}\, w(t)\, dt$ with a weight function $w(t)$. Thus, by straightforward calculations, we note that

$$\begin{aligned} h^{\text{opt}}(t_0) &= h_{\text{LS}}^{\text{opt}}(t_0)\, R_2(q)^{1/7}, \\ h^{\text{opt}} &= h_{\text{LS}}^{\text{opt}}\, R_2(q)^{1/7}, \end{aligned} \tag{3.7}$$

where $h_{\text{LS}}^{\text{opt}}(t_0)$ and $h_{\text{LS}}^{\text{opt}}$ are the corresponding optimal bandwidths of the local quadratic least squares estimator. With the optimal bandwidths, we have

$$\frac{\text{MSE}_{\text{opt}}\{\hat{m}'_{\text{LS}}(t_0)\}}{\text{MSE}_{\text{opt}}\{\hat{m}'(t_0)\}} \rightarrow R_2(q)^{-4/7},$$

$$\frac{\text{MISE}_{\text{opt}}(\hat{m}'_{\text{LS}})}{\text{MISE}_{\text{opt}}(\hat{m}')} \rightarrow R_2(q)^{-4/7}.$$

Therefore, the asymptotic relative efficiency of the local quadratic CQR estimator $\hat{m}'$ with respect to the local quadratic least squares estimator $\hat{m}'_{\text{LS}}$ is defined to be

$$\text{ARE}(\hat{m}', \hat{m}'_{\text{LS}}) = R_2(q)^{-4/7}. \tag{3.8}$$

ARE depends only on the error distribution and it is scale invariant.

To gain insights into the asymptotic relative efficiency, we consider the limit when q is large. Zou and Yuan (2008) showed that

$$\lim_{q \to \infty} \{R_2(q)^{-1}\} > 6/\text{e}\pi = 0.7026.$$

Immediately, we know that, if using a large q, ARE is bounded below by $0.7026^{4/7} = 0.8173$. Having a universal lower bound is very useful because it prohibits severe loss in efficiency when replacing the local quadratic least squares estimator with the local quadratic CQR estimator. One of our contributions in this work is to provide an improved sharper lower bound as shown in the following theorem.

Theorem 4. Let $\mathcal{F}$ denote the class of error distributions with mean 0 and variance 1; then we have

Table 2. Comparisons of ARE($\hat{m}'$, $\hat{m}'_{\mathrm{LS}}$)

Error distribution	*ARE($\hat{m}'$, $\hat{m}'_{\mathrm{LS}}$) for the following values of q:*					
	$q=1$	$q=5$	$q=9$	$q=19$	$q=99$	$q=\infty$
$N(0,1)$	0.7726	0.9453	0.9625	0.9708	0.9738	0.9740
Laplace	1.4860	1.2812	1.2680	1.2625	1.2608	1.2607
t-distribution with 3 degrees of freedom	1.3179	1.4405	1.4435	1.4435	1.4430	1.4431
t-distribution with 4 degrees of freedom	1.0696	1.2038	1.2104	1.2123	1.2125	1.2125
$0.95\,N(0,1)+0.05\,N(0,3^2)$	0.9008	1.0867	1.1019	1.1073	1.1077	1.1077
$0.90\,N(0,1)+0.10\,N(0,3^2)$	0.9990	1.1869	1.1982	1.1999	1.1987	1.1987
$0.95\,N(0,1)+0.05\,N(0,10^2)$	2.0308	2.4229	2.4466	2.4482	2.4415	2.4415
$0.90\,N(0,1)+0.10\,N(0,10^2)$	2.7160	3.1453	3.1430	3.1135	3.1094	3.1093

$$\inf_{f\in\mathcal{F}}\lim_{q\to\infty}\{R_2(q)^{-1}\}=0.864. \tag{3.9}$$

The lower bound is reached if and only if the error follows the rescaled beta(2,2) distribution with mean 0 and variance 1. Thus

$$\lim_{q\to\infty}\{\mathrm{ARE}(\hat{m}',\hat{m}'_{\mathrm{LS}})\}\geqslant 0.9199. \tag{3.10}$$

It is interesting to note that theorem 4 provides us with the *exact* lower bound of $\mathrm{ARE}(\hat{m}',\hat{m}'_{\mathrm{LS}})$ as $q\to\infty$. Theorem 4 indicates that, if q is large, even in the worst scenario the potential efficiency loss for the local CQR estimator is only 8.01%.

Theorem 4 implies that the local quadratic CQR estimator is a safe alternative to the local quadratic least squares estimator. It concerns the worst case scenario. There are many optimistic scenarios as well in which the ARE can be much bigger than 1. We examine $\mathrm{ARE}(\hat{m}',\hat{m}'_{\mathrm{LS}})$ for the error distributions that are considered in Table 1. We also list the results in Table 2, where the column labelled $q=\infty$ shows the theoretical limit of $\mathrm{ARE}(\hat{m}',\hat{m}'_{\mathrm{LS}})$. Obviously, these limits are all larger than the lower bound 0.9199. The local quadratic CQR estimator loses only less than 4% efficiency when the error distribution is normal and $q=9$. It is interesting to see that for the other non-normal distributions $\mathrm{ARE}(\hat{m}',\hat{m}'_{\mathrm{LS}})$ is larger than 1 and its value is insensitive to the choice of q. For example, with $q=9$, the AREs are already very close to their theoretical limits.

4. Numerical comparisons and examples

In this section, we first use Monte Carlo simulation studies to assess the finite sample performance of the estimation procedures proposed and then demonstrate the application of the method proposed by using a real data example. Throughout this section we use the Epanechnikov kernel, i.e. $K(z)=\frac{3}{4}(1-z^2)_+$. We adopt the majorization-and-minimization algorithm that was proposed by Hunter and Lange (2000) for solving the local CQR smoothing estimator. All the numerical results are computed by using our MATLAB code, which is available on request.

4.1. Bandwidth selection in practical implementation

Bandwidth selection is an important issue in local smoothing. Here we briefly discuss the bandwidth selection issue in the local CQR smoothing estimator by using existing bandwidth selectors for the local polynomial regression. Here we consider two bandwidth selectors.

(a) *The 'pilot' selector*: the idea is to use a pilot bandwidth in local cubic CQR (defined in Section 5) to estimate $m''(t)$ and $m'''(t)$. The fitted residuals can be used to estimate $R_1(q)$ and $R_2(q)$. Thus, we can use the optimal bandwidth formula to estimate the optimal bandwidth and then refit the data.

(b) *A short-cut strategy*: in our numerical studies, we compare the local CQR and local least squares estimators. Note that in expressions (2.8) and (3.7) we obtain very neat relationships between the optimal bandwidths for the local CQR and local least squares estimators. The optimal bandwidth for the local least squares estimators can be selected by existing bandwidth selectors (see chapter 4 of Fan and Gijbels (1996)). In addition, we can infer the factors $R_1(q)$ and $R_2(q)$ from the residuals of the local least squares fit. Sometimes, we even know the exact values of the two factors (e.g. in simulations). Therefore, after fitting the local least squares estimator with the optimal bandwidth, we can estimate the optimal bandwidth for the local CQR estimator.

We used the short-cut strategy in our simulation examples. However, if the error variance is infinite or very large, then the local least squares estimator performs poorly. The pilot selector is a better choice than the short-cut strategy.

4.2. Simulation examples

In our simulation studies, we compare the performance of the newly proposed method with the local polynomial least squares estimate. The bandwidth is set to the optimal bandwidth in which $h_{\text{LS}}^{\text{opt}}$ is selected by a plug-in bandwidth selector (Ruppert *et al.*, 1995). The performance of estimators $\hat{m}(\cdot)$ and $\hat{m}'(\cdot)$ is assessed via the average squared errors ASE, defined by

$$\text{ASE}(\hat{g}) = \frac{1}{n_{\text{grid}}} \sum_{k=1}^{n_{\text{grid}}} \{\hat{g}(u_k) - g(u_k)\}^2,$$

with g equal to either $m(\cdot)$ or $m'(\cdot)$, where $\{u_k, k=1,\ldots,n_{\text{grid}}\}$ are the grid points at which the functions $\{\hat{g}(\cdot)\}$ are evaluated. In our simulation, we set $n_{\text{grid}}=200$ and grid points are evenly distributed over the interval at which the $m(\cdot)$ and $m'(\cdot)$ are estimated. We summarize our simulation results by using the ratio of average squared errors, $\text{RASE}(\hat{g}) = \text{ASE}(\hat{g}_{\text{LS}})/\text{ASE}(\hat{g})$ for an estimator $\hat{g}$, where $\hat{g}_{\text{LS}}$ is the local polynomial regression estimator under the least squares loss. We considered two simulation examples.

4.2.1. *Example 1*

We generated 400 data sets, each consisting of $n=200$ observations, from

$$Y = \sin(2T) + 2\exp(-16T^2) + 0.5\varepsilon, \tag{4.1}$$

where T follows $N(0,1)$. This model is adopted from Fan and Gijbels (1992). In our simulation, we considered five error distributions for ε: $N(0,1)$, Laplace, a t_3-distribution and a mixture of two normal distributions ($0.95\,N(0,1) + 0.05\,N(0,\sigma^2)$ with $\sigma=3, 10$). For the local polynomial CQR estimator, we consider $q=5, 9, 19$, and estimate $m(\cdot)$ and $m'(\cdot)$ over $[-1.5, 1.5]$. The mean and standard deviation of RASE over 400 simulations are summarized in Table 3. To see how the estimate proposed behaves at a typical point, Table 3 also depicts the biases and standard deviations of $\hat{m}(t)$ and $\hat{m}'(t)$ at $t=0.75$. In Table 3, CQR_5, CQR_9 and CQR_{19} correspond to the local CQR estimate with $q=5, 9, 19$ respectively.

4.2.2. *Example 2*

It is of interest to investigate the effect of heteroscedastic errors. For this, we generated 400

Table 3. Simulation results for example 1

	Results for $\hat{m}$			*Results for $\hat{m}'$*		
	RASE mean (standard deviation)	*t = 0.75*		*RASE mean (standard deviation)*	*t = 0.75*	
		Bias	*Standard deviation*		*Bias*	*Standard deviation*
Standard normal						
Least squares	—	−0.0239	0.1098	—	−0.0539	0.6871
CQR_5	$0.9314_{(0.1190)}$	−0.0224	0.1161	$0.9518_{(0.1087)}$	−0.0508	0.7257
CQR_9	$0.9588_{(0.0888)}$	−0.0236	0.1133	$0.9614_{(0.1019)}$	−0.0530	0.7165
CQR_{19}	$0.9802_{(0.0592)}$	−0.0228	0.1117	$0.9646_{(0.0998)}$	−0.0513	0.7178
Laplace						
Least squares	—	−0.0146	0.1215	—	−0.1108	0.6988
CQR_5	$1.1088_{(0.1985)}$	−0.0171	0.1155	$1.1014_{(0.1679)}$	−0.0774	0.6916
CQR_9	$1.0717_{(0.1351)}$	−0.0154	0.1195	$1.1025_{(0.1565)}$	−0.0834	0.6678
CQR_{19}	$1.0346_{(0.0856)}$	−0.0141	0.1214	$1.1005_{(0.1500)}$	−0.0934	0.6529
t-distribution with 3 degrees of freedom						
Least squares	—	−0.0214	0.1266	—	−0.0701	0.7254
CQR_5	$1.2752_{(0.5020)}$	−0.0182	0.1103	$1.2104_{(0.4584)}$	−0.0559	0.6635
CQR_9	$1.1712_{(0.3356)}$	−0.0158	0.1137	$1.2133_{(0.4526)}$	−0.0520	0.6537
CQR_{19}	$1.0710_{(0.2086)}$	−0.0186	0.1222	$1.2182_{(0.4403)}$	−0.0540	0.6431
0.95 N(0, 1) + 0.05 N(0, 9)						
Least squares	—	−0.0007	0.1256	—	−0.0382	0.8540
CQR_5	$1.0685_{(0.2275)}$	−0.0060	0.1202	$1.0479_{(0.1773)}$	−0.0182	0.8098
CQR_9	$1.0621_{(0.1740)}$	−0.0049	0.1219	$1.0531_{(0.1727)}$	−0.0154	0.8085
CQR_{19}	$1.0280_{(0.1125)}$	−0.0018	0.1251	$1.0532_{(0.1687)}$	−0.0198	0.8062
0.95 N(0, 1) + 0.05 N(0, 100)						
Least squares	—	0.0034	0.1283	—	−0.0456	0.8667
CQR_5	$2.1548_{(1.5318)}$	0.0002	0.0888	$1.7671_{(0.7607)}$	0.0022	0.5953
CQR_9	$1.5240_{(0.8360)}$	−0.0009	0.1181	$1.7527_{(0.7535)}$	0.0024	0.6030
CQR_{19}	$1.1600_{(0.8776)}$	0.0069	0.1365	$1.7560_{(0.7382)}$	0.0044	0.5927

simulation data sets, each consisting of $n = 200$ observations, from

$$Y = T\sin(2\pi T) + \sigma(T)\varepsilon, \tag{4.2}$$

where T follows $U(0,1)$, $\sigma(t) = \{2+\cos(2\pi t)\}/10$, and ε is the same as that in example 1. In this example, we estimate $m(t)$ and $m'(t)$ over [0,1]. The mean and standard deviation of RASE over 400 simulations are summarized in Table 4, in which we also show the biases and standard deviations of $\hat{m}(t)$ and $\hat{m}'(t)$ at $t = 0.4$. The notation of Table 4 is the same as that in Table 3.

Table 3 and Table 4 show very similar messages, although Table 4 indicates that the local CQR has more gains over the local least squares method. When the error follows the normal distribution, the RASEs of the local CQR estimators are slightly less than 1. For non-normal distributions, the RASEs of the local CQR estimators can be greater than 1, indicating the gain in efficiency. For estimating the regression function, CQR_5 and CQR_9 seem to have better overall performance than CQR_{19}. For estimating the derivative, all three CQR estimators perform

Table 4. Simulation results for example 2

	Results for $\hat{m}$			*Results for $\hat{m}'$*		
	RASE mean (standard deviation)	*t = 0.4*		*RASE mean (standard deviation)*	*t = 0.4*	
		Bias	*Standard deviation*		*Bias*	*Standard deviation*
Standard normal						
Least squares	—	−0.0177	0.0263	—	0.0329	0.2753
CQR_5	$0.9574_{(0.1699)}$	−0.0166	0.0271	$0.9376_{(0.3587)}$	0.0289	0.3019
CQR_9	$0.9783_{(0.1286)}$	−0.0165	0.0266	$0.9458_{(0.3092)}$	0.0283	0.3013
CQR_{19}	$0.9838_{(0.0815)}$	−0.0168	0.0266	$0.9491_{(0.2952)}$	0.0278	0.2962
Laplace						
Least squares	—	−0.0175	0.0249	—	0.0236	0.2718
CQR_5	$1.1938_{(0.3279)}$	−0.0145	0.0237	$1.2063_{(0.6794)}$	0.0106	0.2701
CQR_9	$1.1405_{(0.2523)}$	−0.0150	0.0243	$1.2046_{(0.6413)}$	0.0079	0.2719
CQR_{19}	$1.0857_{(0.1584)}$	−0.0157	0.0248	$1.2019_{(0.6035)}$	0.0098	0.2693
t-distribution with 3 degrees of freedom						
Least squares	—	−0.0167	0.0261	—	0.0025	0.3068
CQR_5	$1.5974_{(1.0324)}$	−0.0120	0.0229	$1.6099_{(1.7558)}$	0.0004	0.2503
CQR_9	$1.4247_{(0.8170)}$	−0.0132	0.0228	$1.5975_{(1.8047)}$	−0.0002	0.2560
CQR_{19}	$1.2111_{(0.4330)}$	−0.0140	0.0242	$1.5948_{(1.8291)}$	0.0006	0.2567
$0.95N(0, 1) + 0.05N(0, 9)$						
Least squares	—	−0.0175	0.0247	—	−0.0130	0.2916
CQR_5	$1.1788_{(0.6248)}$	−0.0157	0.0228	$1.2268_{(2.0608)}$	−0.0050	0.2778
CQR_9	$1.1507_{(0.4715)}$	−0.0157	0.0230	$1.2132_{(1.8791)}$	−0.0048	0.2754
CQR_{19}	$1.0835_{(0.2603)}$	−0.0159	0.0234	$1.2104_{(1.8546)}$	−0.0066	0.2742
$0.95N(0, 1) + 0.05N(0, 100)$						
Least squares	—	−0.0162	0.0260	—	0.0335	0.3728
CQR_5	$3.1661_{(2.4820)}$	−0.0077	0.0173	$3.0593_{(5.6699)}$	0.0245	0.2420
CQR_9	$2.4179_{(1.7012)}$	−0.0080	0.0171	$3.0287_{(5.3433)}$	0.0209	0.2533
CQR_{19}	$1.3469_{(0.5075)}$	−0.0085	0.0241	$3.0146_{(5.2728)}$	0.0234	0.2452

very similarly. These findings are consistent with the theoretical analysis of asymptotic relative efficiencies.

4.3. A real data example

As an illustration, we now apply the proposed local CQR methodology to the UK Family Expenditure Survey data subset with high net income, which consists of 363 observations. The scatter plot of data is depicted in Fig. 1(a). The data set was collected in the UK Family Expenditure Survey in 1973. Of interest is to study the relationship between food expenditure and the net income. Thus, we take the response variable Y to be the logarithm of food expenditure, and the predictor variable T is the net income.

We first estimated the regression function by using the local least squares estimator with the plug-in bandwidth selector (Ruppert *et al.*, 1995). We further employed the kernel density estimate to infer the error density $f(\cdot)$ based on the residuals from the local least squares estimator.

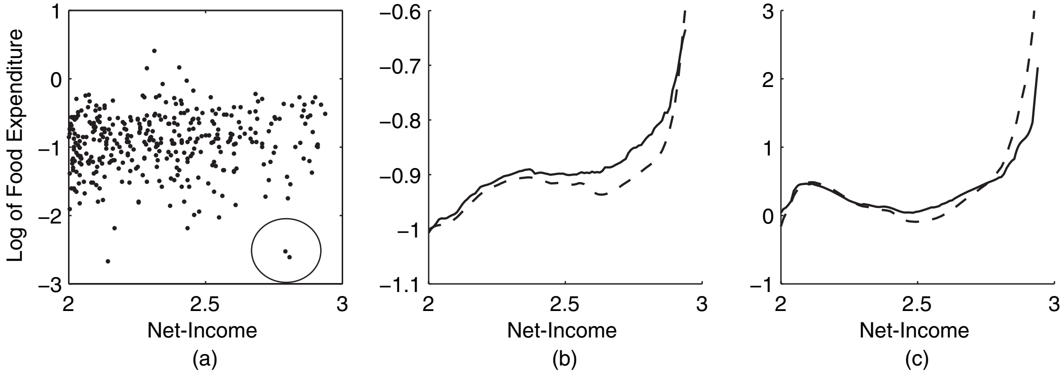

Fig. 1. (a) Scatter plot of data, (b) estimated regression function and (c) estimated derivative function: – – –, least squares; ———, CQR_9

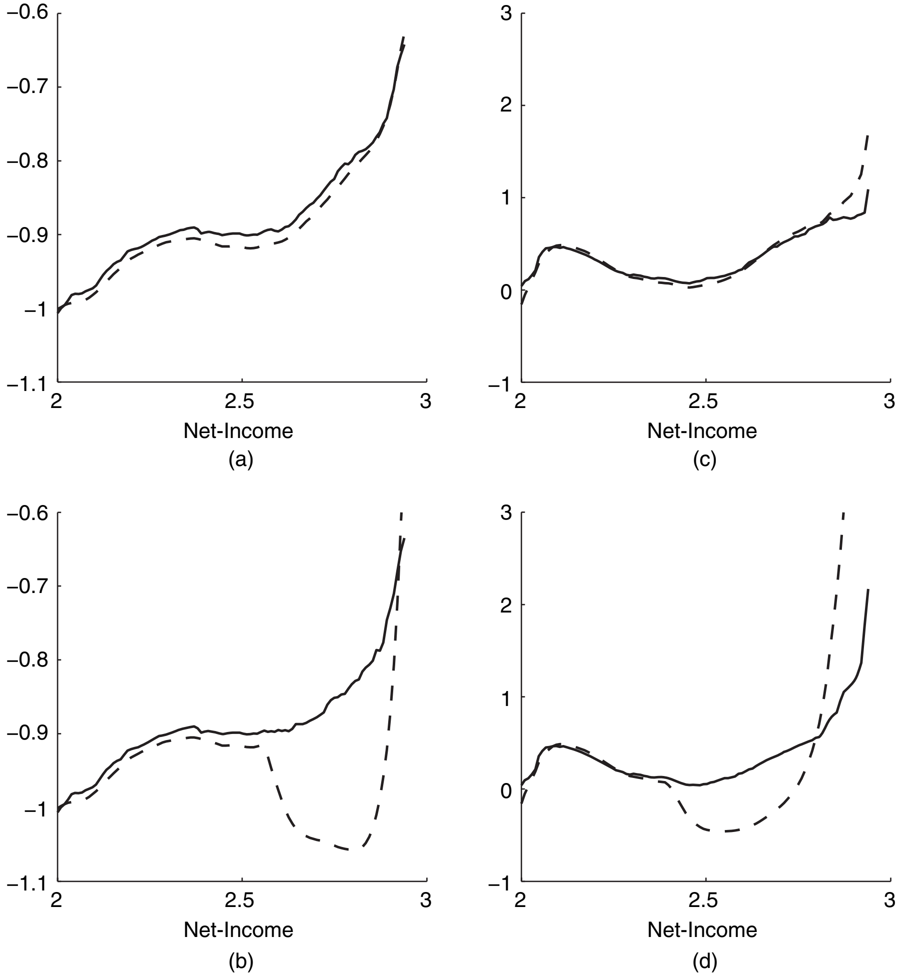

Fig. 2. (a), (b) Estimated regression function and (c), (d) its derivative (– – –, least squares; ———, CQR_9): (a), (c) estimate removing the two possible outliers; (b), (d) estimate moving the two possible outliers to more extreme cases

On the basis of the estimated density, we estimated both $R_1(q)$ and $R_2(q)$, which were used to compute the bandwidth selector for the CQR estimator. For this example, the estimated ratios are close to 1, so we basically use the same bandwidths for these two methods. The bandwidths selected are 0.24 for regression estimation and 0.4 for derivative estimation. The CQR estimates with $q=5, 9, 19$ with the bandwidths selected are evaluated. The CQR estimates with three different qs are very similar; we present only the CQR estimate with $q=9$ in Fig. 1.

It is interesting to see from Fig. 1 that the overall pattern of the local least squares and the local CQR estimate are the same. The difference between the local least squares estimate and the local CQR estimate of the regression function becomes large when the net income is around 2.8. From the scatter plot, there are two possible outlier observations: (2.7902, −2.5207) and (2.8063, −2.6105) (circled in the plot). To understand the influence of these two possible outliers, we re-evaluated the local CQR and the local least squares estimates after excluding these two possible outliers. The resulting estimates are depicted in Figs 2(a) and 2(c) from which we can see that the local CQR estimate remains almost the same, whereas the local least squares estimate changes considerably. We also note that, after removing these two possible outliers, the local least squares estimator becomes very close to the local CQR estimator. Furthermore, as a more extreme demonstration, we kept these two possible outliers in the data set and moved them to more extreme cases, i.e. we moved (2.7902, −2.5207) and (2.8063, −2.6105) to (2.7902, −6.5207) and (2.8063, −6.6105) respectively. After distorting the two observations, we recalculated the local CQR and the local least squares estimate. The resulting estimates are depicted in Figs 2(b) and 2(d) which clearly demonstrate that the local least squares estimate changes dramatically, whereas the local CQR estimate is nearly unaffected by the artificial data distortion.

5. Local p-polynomial composite quantile regression smoothing and proofs

In this section we establish asymptotic theory for the local p-polynomial CQR estimators. We then treat theorems 1 and 2 as two special cases of the general theory. As a generalization of the local linear and local quadratic CQR estimators, the local p-polynomial CQR estimator is constructed by minimizing

$$\sum_{k=1}^{q}\left[\sum_{i=1}^{n}\rho_{\tau_k}\left\{y_i-a_k-\sum_{j=1}^{p}b_j(t_i-t_0)^j\right\}K\left(\frac{t_i-t_0}{h}\right)\right], \tag{5.1}$$

and the local p-polynomial CQR estimators of $m(t_0)$ and $m^{(r)}(t_0)$ are given by

$$\begin{aligned}\hat{m}(t_0)&=\frac{1}{q}\sum_{k=1}^{q}\hat{a}_k,\\ \hat{m}^{(r)}(t_0)&=r!\hat{b}_r, \qquad r=1,\ldots,p.\end{aligned} \tag{5.2}$$

For the asymptotic analysis, we need the following regularity conditions.

(a) $m(t)$ has a continuous $(p+2)$th derivative in the neighbourhood of t_0.
(b) $f_T(\cdot)$, the marginal density function of T, is differentiable and positive in the neighbourhood of t_0.
(c) The conditional variance $\sigma^2(t)$ is continuous in the neighbourhood of t_0.
(d) Assume that the error has a symmetric distribution with a positive density $f(\cdot)$.

We choose the kernel function K such that K is a symmetric density function with finite support $[-M, M]$. The following notation is needed to present the asymptotic properties of the local p-polynomial CQR estimator. Let S_{11} be a $q\times q$ diagonal matrix with diagonal elements $f(c_k)$, $k=1,\ldots,q$, S_{12} be a $q\times p$ matrix with (k,j)-element $f(c_k)\mu_j$, $k=1,\ldots,q$ and $j=1,\ldots,p$, $S_{21}=S_{12}^{\mathrm{T}}$, and S_{22} be a $p\times p$ matrix with (j,j')-element $\Sigma_{k=1}^{q} f(c_k)\mu_{j+j'}$, for $j,j'=1,\ldots,p$. Similarly, Let Σ_{11} be a $q\times q$ matrix with (k,k')-element $\nu_0\tau_{kk'}, k, k'=1,\ldots,q, \Sigma_{12}$ be a $q\times p$ matrix with (k,j)-element $\nu_j\Sigma_{k'=1}^{q}\tau_{kk'}$, $k=1,\ldots,q$ and $j=1,\ldots,p, \Sigma_{21}=\Sigma_{12}^{\mathrm{T}}$, and Σ_{22} be a $p\times p$ matrix with (j,j')-element $(\Sigma_{k,k'=1}^{q}\tau_{kk'})\nu_{j+j'}$, for $j,j'=1,\ldots,p$. Define

$$S=\begin{pmatrix} S_{11} & S_{12}\\ S_{21} & S_{22}\end{pmatrix},$$

$$\Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.$$

Partition S^{-1} into four submatrices as follows:

$$S^{-1}=\begin{pmatrix} S_{11} & S_{12}\\ S_{21} & S_{22}\end{pmatrix}^{-1}=\begin{pmatrix} (S^{-1})_{11} & (S^{-1})_{12}\\ (S^{-1})_{21} & (S^{-1})_{22}\end{pmatrix},$$

where here and hereafter we use $(\cdot)_{11}$ to denote the top left-hand $q\times q$ submatrix and we use $(\cdot)_{22}$ to denote the bottom right-hand $p\times p$ submatrix.

Furthermore, let $u_k=\sqrt{(nh)}\{a_k-m(t_0)-\sigma(t_0)c_k\}, v_j=h^j\sqrt{(nh)}\{j!b_j-m^{(j)}(t_0)\}/j!$. Let $x_i=(t_i-t_0)/h, K_i=K(x_i)$ and

$$\Delta_{i,k}=\frac{u_k}{\sqrt{(nh)}}+\sum_{j=1}^{p}\frac{v_j x_i^j}{\sqrt{(nh)}}.$$

Write $d_{i,k}=c_k\{\sigma(t_i)-\sigma(t_0)\}+r_{i,p}$ with

$$r_{i,p}=m(t_i)-\sum_{j=0}^{p} m^{(j)}(t_0)(t_i-t_0)^j/j!.$$

Define $\eta_{i,k}^{*}$ to be $I\{\varepsilon_i\leqslant c_k-d_{i,k}/\sigma(t_i)-\tau_k\}$. Let $W_n^{*}=(w_{11}^{*},\ldots,w_{1q}^{*},w_{21}^{*},\ldots,w_{2p}^{*})^{\mathrm{T}}$ with

$$w_{1k}^{*}=\frac{1}{\sqrt{(nh)}}\sum_{i=1}^{n} K_i\eta_{i,k}^{*}$$

and

$$w_{2j}^{*}=\frac{1}{\sqrt{(nh)}}\sum_{k=1}^{q}\sum_{i=1}^{n} K_i x_i^j\eta_{i,k}^{*}.$$

The asymptotic properties of the local p-polynomial CQR estimator are based on the following theorem.

Theorem 4. Denote $\hat\theta_n=(\hat u_1,\ldots,\hat u_q,\hat v_1,\ldots,\hat v_p)$ as the minimizer of expression (5.1). Then, under the regularity conditions (a)–(c), we have

$$\hat\theta_n+\frac{\sigma(t_0)}{f_T(t_0)}S^{-1}E[W_n^{*}|\mathbf{T}]\xrightarrow{\mathcal{L}}\mathrm{MVN}\left\{\mathbf{0},\frac{\sigma^2(t_0)}{f_T(t_0)}S^{-1}\Sigma S^{-1}\right\}.$$

To prove theorem 3, we first establish lemmas 2 and 3.

Lemma 2. Minimizing expression (5.1) is equivalent to minimizing

$$\sum_{k=1}^{q}u_k\left\{\sum_{i=1}^{n}\frac{K_i\eta_{i,k}^{*}}{\sqrt{(nh)}}\right\}+\sum_{j=1}^{p}v_j\left\{\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_i x_i^j\eta_{i,k}^{*}}{\sqrt{(nh)}}\right\}+\sum_{k=1}^{q}B_{n,k}(\theta)$$

with respect to $\theta=(u_1,\ldots,u_q,v_1,\ldots,v_p)^{\mathrm{T}}$, where

$$B_{n,k}(\theta)=\sum_{i=1}^{n}\left(K_i\int_0^{\Delta_{i,k}}\left[I\left\{\varepsilon_i\leqslant c_k-\frac{d_{i,k}}{\sigma(t_i)}+\frac{z}{\sigma(t_i)}\right\}-I\left\{\varepsilon_i\leqslant c_k-\frac{d_{i,k}}{\sigma(t_i)}\right\}\right]\mathrm{d}z\right).$$

Proof. To apply the identity (Knight, 1998)

$$\rho_\tau(x-y)-\rho_\tau(x)=y\{I(x\leqslant 0)-\tau\}+\int_0^y\{I(x\leqslant z)-I(x\leqslant 0)\}\,\mathrm{d}z, \tag{5.3}$$

we write $y_i-a_k-\Sigma_{j=1}^{p}b_j(t_i-t_0)^j=\sigma(t_i)(\varepsilon_i-c_k)+d_{i,k}-\Delta_{i,k}$. Minimizing expression (5.1) is equivalent to minimizing

$$L_n(\theta)=\sum_{i=1}^{n}\left(K_i\sum_{k=1}^{q}[\rho_{\tau_k}\{\sigma(t_i)(\varepsilon_i-c_k)+d_{i,k}-\Delta_{i,k}\}-\rho_{\tau_k}\{\sigma(t_i)(\varepsilon_i-c_k)+d_{i,k}\}]\right).$$

Using identity (5.3) and with some straightforward calculations, it follows that

$$L_n(\theta)=\sum_{k=1}^{q}u_k\left\{\sum_{i=1}^{n}\frac{K_i\eta_{i,k}^*}{\surd(nh)}\right\}+\sum_{j=1}^{p}v_j\left\{\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_ix_i^j\eta_{i,k}^*}{\surd(nh)}\right\}+\sum_{k=1}^{q}B_{n,k}(\theta).$$

This completes the proof. □

Let $S_{n,11}$ be a $q\times q$ diagonal matrix with diagonal elements $f(c_k)\Sigma_{i=1}^nK_i/nh\sigma(t_i), k=1,\ldots,q$, $S_{n,12}$ be a $q\times p$ matrix with (k,j)-element $f(c_k)\Sigma_{i=1}^nK_ix_i^j/nh\sigma(t_i)$, $j=1,\ldots,p$, and $S_{n,22}$ be a $p\times p$ matrix with (j,j')-element $\Sigma_{k=1}^qf(c_k)\Sigma_{i=1}^nK_ix_i^{j+j'}/nh\sigma(t_i)$. Denote

$$S_n=\begin{pmatrix}S_{n,11} & S_{n,12}\\ S_{n,12}^{\mathrm{T}} & S_{n,22}\end{pmatrix}.$$

Lemma 3. Under conditions (a)–(c), $L_n(\theta)=\frac{1}{2}\theta^{\mathrm{T}}S_n\theta+(W_n^*)^{\mathrm{T}}\theta+o_p(1)$.

Proof. Write $L_n(\theta)$ as

$$L_n(\theta)=\sum_{k=1}^{q}u_k\left\{\sum_{i=1}^{n}\frac{K_i\eta_{i,k}^*}{\surd(nh)}\right\}+\sum_{j=1}^{p}v_j\left\{\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_ix_i^j\eta_{i,k}^*}{\surd(nh)}\right\}+\sum_{k=1}^{q}E_\varepsilon[B_{n,k}(\theta)|\mathbf{T}]+\sum_{k=1}^{q}R_{n,k}(\theta),$$

where $R_{n,k}(\theta)=B_{n,k}(\theta)-E_\varepsilon[B_{n,k}(\theta)|\mathbf{T}]$. Using $F(c_k+z)-F(c_k)=z\,f(c_k)+o(z)$, then

$$\begin{aligned}\sum_{k=1}^{q}E_\varepsilon[B_{n,k}(\theta)|\mathbf{T}]&=\sum_{k=1}^{q}\sum_{i=1}^{n}\left(K_i\int_0^{\Delta_{i,k}}\left[\frac{z}{\sigma(t_i)}f\left\{c_k-\frac{d_{i,k}}{\sigma(t_i)}\right\}+o(z)\right]\mathrm{d}z\right)\\&=\sum_{k=1}^{q}\sum_{i=1}^{n}\left[K_i\Delta_{i,k}^2\frac{f\{c_k-d_{i,k}/\sigma(t_i)\}}{2\,\sigma(t_i)}\right]+o_p(1)\\&=\sum_{k=1}^{q}\sum_{i=1}^{n}\left\{K_i\Delta_{i,k}^2\frac{f(c_k)}{2\,\sigma(t_i)}\right\}+o_p(1)=\frac{1}{2}\theta^{\mathrm{T}}S_n\theta+o_p(1).\end{aligned}$$

We now prove that $R_{n,k}(\theta)=o_p(1)$. It is sufficient to show that $\mathrm{var}_\varepsilon[B_{n,k}(\theta)|\mathbf{T}]=o_p(1)$. In fact,

$$\begin{aligned}\mathrm{var}_\varepsilon\{B_{n,k}(\theta)|\mathbf{T}\}&=\sum_{i=1}^{n}\mathrm{var}_\varepsilon\left\{\left(K_i\int_0^{\Delta_{i,k}}\left[I\left\{\varepsilon_i\leqslant c_k-\frac{d_{i,k}}{\sigma(t_i)}+\frac{z}{\sigma(t_i)}\right\}-I\left\{\varepsilon_i\leqslant c_k-\frac{d_{i,k}}{\sigma(t_i)}\right\}\right]\mathrm{d}z\right)|\mathbf{T}\right\}\\&\leqslant\sum_{i=1}^{n}E_\varepsilon\left[\left(K_i\int_0^{\Delta_{i,k}}I\left\{\varepsilon_i\leqslant c_k-\frac{d_{i,k}}{\sigma(t_i)}+\frac{z}{\sigma(t_i)}\right\}-I\left\{\varepsilon_i\leqslant c_k-\frac{d_{i,k}}{\sigma(t_i)}\right\}\right]\mathrm{d}z\right)^2|\mathbf{T}\right]\end{aligned}$$

$$\leqslant \sum_{i=1}^{n} K_i^2 \int_0^{|\Delta_{i,k}|} \int_0^{|\Delta_{i,k}|} \left[F\left\{ c_k - \frac{d_{i,k}}{\sigma(t_i)} + \frac{|\Delta_{i,k}|}{\sigma(t_i)} \right\} - F\left\{ c_k - \frac{d_{i,k}}{\sigma(t_i)} \right\} \right] \mathrm{d}z_1 \mathrm{d}z_2$$
$$= o\left(\sum_{i=1}^{n} K_i^2 \Delta_{i,k}^2 \right) = o_p(1).$$

5.1. Proof of theorem 5

Similarly to Parzen (1962), we have

$$\frac{1}{nh} \sum_{i=1}^{n} K_i x_i^j \xrightarrow{\mathrm{P}} f_T(t_0) \mu_j,$$

where $\rightarrow^{\mathrm{P}}$ stands for convergence in probability. Thus,

$$S_n \xrightarrow{\mathrm{P}} \frac{f_T(t_0)}{\sigma(t_0)} S = \frac{f_T(t_0)}{\sigma(t_0)} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

This, together with lemmas 2 and 3, leads to

$$L_n(\theta) = \frac{1}{2} \frac{f_T(t_0)}{\sigma(t_0)} \theta^{\mathrm{T}} S \theta + (W_n^*)^{\mathrm{T}} \theta + o_p(1).$$

Since the convex function $L_n(\theta) - (W_n^*)^{\mathrm{T}} \theta$ converges in probability to the convex function $\frac{1}{2}\{f_T(t_0)/\sigma(t_0)\} \theta^{\mathrm{T}} S \theta$, it follows from the convexity lemma (Pollard, 1991) that, for any compact set Θ, the quadratic approximation to $L_n(\theta)$ holds uniformly for θ in any compact set, which leads to

$$\hat{\theta}_n = -\frac{\sigma(t_0)}{f_T(t_0)} S^{-1} W_n^* + o_p(1).$$

Denote $\eta_{i,k} = I(\varepsilon_i \leqslant c_k) - \tau_k$ and $W_n = (w_{11}, \ldots, w_{1q}, w_{21}, \ldots, w_{2p})^{\mathrm{T}}$ with

$$w_{1k} = \frac{1}{\sqrt{(nh)}} \sum_{i=1}^{n} K_i \eta_{i,k}$$

and

$$w_{2j} = \frac{1}{\sqrt{(nh)}} \sum_{k=1}^{q} \sum_{i=1}^{n} K_i x_i^j \eta_{i,k}.$$

By the Cramer–Wald theorem, it is easy to see that the central limit theorem for $W_n|\mathbf{T}$ holds:

$$\frac{W_n|\mathbf{T} - E[W_n|\mathbf{T}]}{\sqrt{\{\mathrm{var}(W_n|\mathbf{T})\}}} \xrightarrow{\mathcal{L}} \mathrm{MVN}(\mathbf{0}, I_{(p+q)\times(p+q)}). \tag{5.4}$$

Note that $\mathrm{cov}(\eta_{i,k}, \eta_{i,k'}) = \tau_{kk'}$ and $\mathrm{cov}(\eta_{i,k}, \eta_{j,k'}) = 0$, if $i \neq j$. Similarly to Parzen (1962), we have $(1/nh)\Sigma_{i=1}^{n} K_i^2 x_i^j \rightarrow^{\mathrm{P}} f_T(t_0)\nu_j$, Therefore, $\mathrm{var}(W_n|\mathbf{T}) \rightarrow^{\mathrm{P}} f_T(t_0)\Sigma$. Combined with result (5.4), we have $W_n|\mathbf{T} \rightarrow^{\mathcal{L}} \mathrm{MVN}\{\mathbf{0}, f_T(t_0)\Sigma\}$. Moreover, we have

$$\mathrm{var}(w_{1k}^* - w_{1k}|\mathbf{T}) = \frac{1}{nh} \sum_{i=1}^{n} K_i^2 \mathrm{var}(\eta_{i,k}^* - \eta_{i,k}) \leqslant \frac{1}{nh} \sum_{i=1}^{n} K_i^2 \left[F\left\{ c_k + \frac{|d_{i,k}|}{\sigma(t_i)} \right\} - F(c_k) \right] = o_p(1)$$

and also

$$\mathrm{var}(w_{2j}^* - w_{2j}|\mathbf{T}) = \frac{1}{nh} \sum_{i=1}^{n} K_i^2 x_i^j \mathrm{var}\left(\sum_{k=1}^{q} \eta_{i,k}^* - \eta_{i,k} \right)$$
$$\leqslant \frac{q^2}{nh} \sum_{i=1}^{n} K_i^2 x_i^j \max_k \left[F\left\{ c_k + \frac{|d_{i,k}|}{\sigma(t_i)} \right\} - F(c_k) \right] = o_p(1);$$

thus $\mathrm{var}(W_n^* - W_n|\mathbf{T}) = o_p(1)$. So by Slutsky's theorem, conditioning on $\mathbf{T}$, we have $W_n^*|\mathbf{T} - E[W_n^*|\mathbf{T}] \to^{\mathcal{L}} \mathrm{MVN}\{\mathbf{0}, f_T(t_0)\Sigma\}$. Therefore,

$$\hat{\theta}_n + \frac{\sigma(t_0)}{f_T(t_0)} S^{-1}\, E[W_n^*|\mathbf{T}] \xrightarrow{\mathcal{L}} \mathrm{MVN}\left\{\mathbf{0}, \frac{\sigma^2(t_0)}{f_T(t_0)} S^{-1}\Sigma S^{-1}\right\}. \tag{5.5}$$

This completes the proof.

5.2. Proof of theorem 1

The asymptotic normality follows theorem 5 with $p=1$. Let us calculate the conditional bias and variance. Denote by $e_{q\times 1}$ the vector that contains q 1s. When $p=1$, S is a diagonal matrix with diagonal elements $f(c_1), \ldots, f(c_q)$ and $\mu_2 \Sigma_{k=1}^q f(c_k)$. So the asymptotic conditional bias of $\hat{m}(t_0) = (1/q)\Sigma_{k=1}^q \hat{a}_k$ is

$$\begin{aligned}\mathrm{bias}\{\hat{m}(t_0)|\mathbf{T}\} &= \frac{1}{q}\sigma(t_0)\sum_{k=1}^q c_k - \frac{1}{q\surd(nh)}\frac{\sigma(t_0)}{f_T(t_0)} e_{q\times 1}^{\mathrm{T}}(S^{-1})_{11}\, E[W_{1n}^*|\mathbf{T}] \\ &= \frac{1}{q}\sigma(t_0)\sum_{k=1}^q c_k - \frac{1}{qnh}\frac{\sigma(t_0)}{f_T(t_0)}\sum_{i=1}^n K_i \sum_{k=1}^q \frac{1}{f(c_k)}\left[F\left\{c_k - \frac{d_{i,k}}{\sigma(t_i)}\right\} - F(c_k)\right].\end{aligned}$$

Note that the error is symmetric; thus $\Sigma_{k=1}^q c_k = 0$ and, furthermore, it is easy to check that

$$\frac{1}{q}\sum_{k=1}^q \frac{1}{f(c_k)}\left[F\left\{c_k - \frac{d_{i,k}}{\sigma(t_i)}\right\} - F(c_k)\right] = -\frac{r_{i,p}}{\sigma(t_i)}\{1 + o_p(1)\}.$$

Therefore,

$$\mathrm{bias}\{\hat{m}(t_0)|\mathbf{T}\} = \frac{1}{nh}\frac{\sigma(t_0)}{f_T(t_0)}\sum_{i=1}^n K_i \frac{r_{i,p}}{\sigma(t_i)}\{1 + o_p(1)\}.$$

By using the fact that

$$\frac{1}{nh}\sum_{i=1}^n K_i \frac{r_{i,p}}{\sigma(t_i)} = \frac{f_T(t_0)\, m''(t_0)}{2\sigma(t_0)}\mu_2 h^2\{1 + o_p(1)\},$$

we obtain

$$\mathrm{bias}\{\hat{m}(t_0)|\mathbf{T}\} = \tfrac{1}{2} m''(t_0)\mu_2 h^2 + o_p(h^2). \tag{5.6}$$

Furthermore, the conditional variance of $\hat{m}(t_0)$ is

$$\begin{aligned}\mathrm{var}\{\hat{m}(t_0)|\mathbf{T}\} &= \frac{1}{nh}\frac{\sigma^2(t_0)}{f_T(t_0)}\frac{1}{q^2} e_{q\times 1}^{\mathrm{T}}(S^{-1}\Sigma S^{-1})_{11} e_{q\times 1} + o_p\left(\frac{1}{nh}\right) \\ &= \frac{1}{nh}\frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)} R_1(q) + o_p\left(\frac{1}{nh}\right).\end{aligned} \tag{5.7}$$

By using theorem 5, we can further derive the asymptotic bias and variance of $\tilde{m}'(t_0)$ that is given in expression (2.3):

$$\mathrm{bias}\{\tilde{m}'(t_0)|\mathbf{T}\} = \frac{1}{6}\left\{m'''(t_0) + 3m''(t_0)\frac{f_T'(t_0)}{f_T(t_0)}\right\}\frac{\mu_4}{\mu_2}h^2 + o_p(h^2), \tag{5.8}$$

$$\mathrm{var}\{\tilde{m}'(t_0)|\mathbf{T}\} = \frac{1}{nh^3}\frac{\nu_2 \sigma^2(t_0)}{\mu_2^2 f_T(t_0)} R_2(q) + o_p\left(\frac{1}{nh^3}\right). \tag{5.9}$$

5.3. Proof of theorem 2

Note that

$$\lim_{q\to\infty}\{R_1(q)\}=\int_0^1\int_0^1 \frac{s_1\wedge s_2-s_1s_2}{f\{F^{-1}(s_1)\}f\{F^{-1}(s_2)\}}\,\mathrm{d}s_1\,\mathrm{d}s_2$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F(z_1)\wedge F(z_2)-F(z_1)F(z_2)\}\,\mathrm{d}z_1\,\mathrm{d}z_2, \qquad (5.10)$$

by change of variables. Define two functions: $G(s)=\int_{-\infty}^{s}F(t)\,\mathrm{d}t$ and $H(s)=\int_{-\infty}^{s}G(t)\,\mathrm{d}t$. It is easy to verify that

$$G(s)=\int_{-\infty}^{s}(s-x)\,f(x)\,\mathrm{d}x=s\,F(s)-k_1(s), \qquad (5.11)$$

where $k_1(s)=\int_{-\infty}^{s}x\,f(x)\,\mathrm{d}x$. Similarly, we obtain

$$2\,H(s)=\int_{-\infty}^{s}(s-x)^2 f(x)\,\mathrm{d}x=s^2F(s)-2s\,k_1(s)+k_2(s), \qquad (5.12)$$

where $k_2(s)=\int_{-\infty}^{s}x^2 f(x)\,\mathrm{d}x$. Let I be the integral in equation (5.10). We have that

$$I=2\int_{-\infty}^{\infty}\left\{\int_{z_1}^{\infty}f(t)\,\mathrm{d}t\right\}G(z_1)\,\mathrm{d}z_1=2\int_{-\infty}^{\infty}f(t)\left\{\int_{-\infty}^{t}G(z_1)\,\mathrm{d}z_1\right\}\mathrm{d}t=\int_{-\infty}^{\infty}2\,f(t)\,H(t)\,\mathrm{d}t. \qquad (5.13)$$

By the definition of G and H, we know that $\mathrm{d}\{2\,H(t)\,F(t)-G^2(t)\}/\mathrm{d}t=2\,H(t)\,f(t)$, and combining equations (5.11) and (5.12) yields $2\,H(t)\,F(t)-G^2(t)=k_2(t)\,F(t)-k_1^2(t)$. Now it is easy to see that $I=1$, by the facts that $\int_{-\infty}^{\infty}x^2 f(x)\,\mathrm{d}x=E_F[\varepsilon^2]=1$ and $\int_{-\infty}^{\infty}x\,f(x)\,\mathrm{d}x=E_F[\varepsilon]=0$.

5.4. Proof of theorem 3

We apply theorem 5 to obtain the asymptotic normality. Denote by e_r the p-vector $(0,0,\ldots,1,0,\ldots,0)^{\mathrm{T}}$ with 1 on the rth position. When $p=2$, S_{12} has the form $S_{12}=(\mathbf{0}_{q\times1},\mu_2\{f(c_k)\}_{q\times1})$, and $S_{22}=\mathrm{diag}\{\mu_2\Sigma_{k=1}^{q}f(c_k),\mu_4\Sigma_{k=1}^{q}f(c_k)\}$. Since $S_{11}=\mathrm{diag}\{f(c_1),\ldots,f(c_q)\}$, $(S^{-1})_{22}=(S_{22}-S_{21}S_{11}^{-1}S_{12})^{-1}=\mathrm{diag}[\{\mu_2\Sigma_{k=1}^{q}f(c_k)\}^{-1},\{(\mu_4-\mu_2^2)\Sigma_{k=1}^{q}f(c_k)\}^{-1}]$. Note that $(S^{-1})_{21}=-(S^{-1})_{22}S_{21}S_{11}^{-1}$. Thus,

$$(S^{-1})_{21}=\left(\mathbf{0}_{q\times1},\left[\mu_2\Big/\left\{(\mu_4-\mu_2^2)\sum_{k=1}^{q}f(c_k)\right\}\right]\mathbf{1}_{q\times1}\right)^{\mathrm{T}}.$$

By theorem 4

$$\mathrm{bias}\{\hat{m}'(t_0)|\mathbf{T}\}=-\frac{\sigma(t_0)}{h\,f_T(t_0)}\frac{1}{\surd(nh)}e_1^{\mathrm{T}}\{(S^{-1})_{21}E[W_{1n}^*|\mathbf{T}]+(S^{-1})_{22}E[W_{2n}^*|\mathbf{T}]\}$$
$$=-\frac{\sigma(t_0)}{h\,f_T(t_0)}\frac{1}{\mu_2\sum_{k=1}^{q}f(c_k)}\frac{1}{\surd(nh)}E[w_{21}^*|\mathbf{T}].$$

Note that

$$E[w_{2j}^*|\mathbf{T}]=\frac{1}{\surd(nh)}\sum_{i=1}^{n}K_i x_i^j\sum_{k=1}^{q}\left[F\left\{c_k-\frac{d_{i,k}}{\sigma(t_i)}\right\}-F(c_k)\right].$$

Similarly, under condition (d), we have

$$\sum_{k=1}^{q}\left[F\left\{c_k-\frac{d_{i,k}}{\sigma(t_i)}\right\}-F(c_k)\right]=-\sum_{k=1}^{q}f(c_k)\frac{r_{i,p}}{\sigma(t_i)}\{1+o_p(1)\}.$$

Therefore,

$$\text{bias}\{\hat{m}'(t_0)|\mathbf{T}\}\frac{1}{nh^2}\frac{\sigma(t_0)}{f_T(t_0)}\sum_{i=1}^{n}K_i x_i \frac{r_{i,p}}{\sigma(t_i)}\{1+o_p(1)\}.$$

For $p=2$,

$$\frac{1}{nh}\sum_{i=1}^{n}K_i x_i \frac{r_{i,p}}{\sigma(t_i)}=\frac{f_T(t_0)\,m'''(t_0)}{6\sigma(t_0)}\frac{\mu_4}{\mu_2}h^3\{1+o_p(1)\},$$

we obtain

$$\text{bias}\{\hat{m}'(t_0)|\mathbf{T}\}=\frac{1}{6}m'''(t_0)\frac{\mu_4}{\mu_2}h^2+o_p(h^2). \tag{5.14}$$

Furthermore, the conditional variance of $\hat{m}(t_0)$ is

$$\begin{aligned}\text{var}\{\hat{m}'(t_0)|\mathbf{T}\}&=\frac{1}{nh^3}\frac{\sigma^2(t_0)}{f_T(t_0)}e_1^{\mathrm{T}}(S^{-1}\Sigma S^{-1})_{22}e_1+o_p\left(\frac{1}{nh^3}\right),\\ &=\frac{1}{nh^3}\frac{\nu_2\sigma^2(t_0)}{\mu_2^2 f_T(t_0)}R_2(q)+o_p\left(\frac{1}{nh^3}\right).\end{aligned} \tag{5.15}$$

which completes the proof.

5.5. *Proof of theorem 5*

From Zou and Yuan (2008), we know that

$$\lim_{q\to\infty}\left[\left\{\sum_{k=1}^{q}f(c_k)\right\}^2\right]\Big/\sum_{k=1}^{q}\sum_{k'=1}^{q}\tau_{kk'}=12\,E_F^2[f(\varepsilon)]=12\left\{\int f^2(x)\,\mathrm{d}x\right\}^2.$$

Thus $\lim_{q\to\infty}\{1/R_2(q)\}=12\{\int f^2(x)\,\mathrm{d}x\}^2$. We note that $12\{\int f^2(x)\,\mathrm{d}x\}^2$ is also the asymptotic Pitman efficiency of the Wilcoxon test relative to the t-test (Hodges and Lehmann, 1956). For the rest of the proof, readers are referred to Hodges and Lehmann (1956).

6. Discussion

In this paper our theoretical analysis deals with the classical setting in which t_0 is an interior point and the error distribution has finite variance. We should point out here that the same arguments hold for estimating boundary points and the methodology proposed is valid even when the error variance is infinite.

(a) *Automatic boundary correction*: for simplicity, consider $t\in[0,1]$ and $t_0=ch$ for some constant c. We show that the leading team of the asymptotic bias of the local linear or quadratic CQR estimator is the same as that of the local linear or quadratic least squares estimator, which indicates that the local CQR estimator enjoys the property of automatic boundary correction, which is a nice property of the local least squares estimator. Furthermore, the asymptotic relative efficiency remains exactly the same as that for interior points.

(b) *Infinite error variance*: we show that the local CQR estimator still enjoys the optimal rate of convergence and asymptotic normality even when the conditional variance is infinite. This property can be important for real applications, since we have no information on the error distribution in practice.

For a detailed theoretical proof of these claims, we refer interested readers to Kai *et al.* (2009), where we also provide additional simulation results to support the theory. We opt not to show these results here for brevity.

In this paper, we focus on the local CQR estimate for the non-parametric regression model. The methodology proposed and theory may be extended to settings in the presence of multivariate covariates by considering varying-coefficient models, additive models or semiparametric models. Such extensions are of great interest, and further research is needed for such extensions.

Finally, we would like to point out that the local CQR procedure is efficiently implemented by using the MM algorithm. Our experiences show that, for $q=9$ and sample size $n=7000$, the local CQR fit at a given location can be computed within 0.32 s on an Advanced Micro Devices 1.9-GHz machine. The MM implementation seems to be more efficient than standard linear programming. We discuss the computing algorithm in detail in a separate paper.

Acknowledgements

The authors are grateful to the Joint Editor, the Associate Editor and two referees for their helpful and constructive comments, which have greatly improved the quality of this paper.

Kai's research is supported by the National Institute on Drug Abuse, National Institutes of Health grants R21 DA024260 and P50 DA10075 as a research assistant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health. Li's research is supported by National Science Foundation grants DMS 0348869 and DMS 0722351. Zou's research is supported by National Science Foundation grant DMS 0706733.

References

Chu C.-K. and Marron J. S. (1991) Choosing a kernel regression estimator. *Statist. Sci.*, **6**, 404–436.
Fan, J. (1992) Design-adaptive nonparametric regression. *J. Am. Statist. Ass.*, **87**, 998–1004.
Fan, J. and Gijbels, I. (1992) Variable bandwidth and local linear regression smoothers. *Ann. Statist.*, **20**, 2008–2036.
Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
Fan, J., Hu, T. C. and Truong, Y. K. (1994) Robust non-parametric function estimation. *Scand. J. Statist.*, **21**, 433–446.
Hodges, J. and Lehmann, E. (1956) The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.*, **27**, 324–335.
Hunter, D. R. and Lange, K. (2000) Quantile regression via an MM algorithm. *J. Computnl Graph. Statist.*, **9**, 60–77.
Kai, B., Li, R. and Zou, H. (2009) Supplementary materials for "local cqr smoothing: an efficient and safe alternative to local polynomial regression". *Technical Report*. Pennsylvania State University, University Park. (Available from `http://www.stat.psu.edu/rli/research/Supplement-of-localCQR.pdf`.)
Knight, K. (1998) Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.*, **26**, 755–770.
Koenker, R. (1984) A note on l-estimates for linear models. *Statist. Probab. Lett.*, **2**, 323–325.
Koenker, R. (2005) *Quantile Regression*. Cambridge: Cambridge University Press.
Parzen, E. (1962) On estimation of a probability density function and model. *Ann. Math. Statist.*, **33**, 1065–1076.
Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometr. Theory*, **7**, 186–199.
Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Ass.*, **90**, 1257–1270.
Welsh, A. H. (1996) Robust estimation of smooth regression and spread functions and their derivatives. *Statist. Sin.*, **6**, 347–366.
Yu, K. and Jones, M. C. (1998) Local linear quantile regression. *J. Am. Statist. Ass.*, **93**, 228–237.
Zou, H. and Yuan, M. (2008) Composite quantile regression and the oracle model selection theory. *Ann. Statist.*, **36**, 1108–1126.