Local CQR Smoothing: An Efficient and Safe Alternative to Local Polynomial Regression

Bo Kai[*], **Runze Li [Professor of Statistics]**[†], and **Hui Zou [Associate Professor of Statistics]** [‡]

[*]Department of Statistics, The Pennsylvania State University. E-mail: bokai@psu.edu

[†]Department of Statistics, The Pennsylvania State University. E-mail: rli@stat.psu.edu

[‡]School of Statistics, University of Minnesota. E-mail: hzou@stat.umn.edu

Abstract

Local polynomial regression is a useful nonparametric regression tool to explore fine data structures and has been widely used in practice. In this paper, we propose a new nonparametric regression technique called *local composite-quantile-regression (CQR) smoothing* in order to further improve local polynomial regression. Sampling properties of the proposed estimation procedure are studied. We derive the asymptotic bias, variance and normality of the proposed estimate. Asymptotic relative efficiency of the proposed estimate with respect to the local polynomial regression is investigated. It is shown that the proposed 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 proposed estimates. The simulation results are consistent with our theoretical findings. A real data example is used to illustrate the proposed method.

Key words and phrases

Asymptotic efficiency; CQR estimator; Kernel function; Local polynomial regression; Nonparametric regression

1 Introduction

Consider the general nonparametric regression model

$$Y = m(T) + \sigma(T)\epsilon, \tag{1.1}$$

where Y is the response variable, T is a covariate, $m(T) = E(Y|T)$, which is assumed to be a smooth nonparametric function, and $\sigma(T)$ is a positive function representing the standard deviation. We assume ϵ has mean 0 and variance 1. Local polynomial regression is a popular and successful method for nonparametric regression, and it has been well studied in the literature (Fan & Gijbels 1996). By locally fitting a linear (or polynomial) regression model via adaptively weighted least squares, local polynomial regression is able to 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 using different

Correspondence to: Hui Zou.

local fitting methods. For example, in the presence of outliers, one may consider local least absolute deviation (LAD) polynomial regression (Fan, Hu & Truong 1994, Welsh 1996). When the error follows a Laplacian distribution, the local LAD polynomial regression is more efficient than the local least squares polynomial regression. Of course, the local LAD polynomial regression can do much worse than the 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 upon the classical local polynomial regression for a wide class of error distributions, and has comparable efficiency in the worst case scenario.

Our proposal is built upon the composite-quantile-regression (CQR) estimator recently proposed by Zou & Yuan (2008) for estimating the regression coefficients in the classical linear regression model. Zou & Yuan (2008) show that the relative efficiency of the CQR estimator compared to the least squares estimator is greater than 70% regardless 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 the local CQR smoothers as nonparametric estimates of the regression function and its derivatives.

We make several contributions in this paper.

- We propose the local linear CQR estimator for estimating the nonparametric regression function. We establish the asymptotic theory of 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 used non-normal error distributions.

- We propose the local quadratic CQR estimator for estimating the derivative of the regression function. The 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.

- The general asymptotic theory of 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 due to 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 & Gijbels 1996). There is no contradiction between this fact and our results, because the proposed local CQR estimator is a *nonlinear* smoother.

The rest of this paper is organized as follows. In section 2, we introduce the local linear CQR for the nonparametric regression and study its asymptotic properties. In section 3, we propose the local quadratic CQR for estimating the derivative of the nonparametric regression, which is able to 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 the general theory of the local p-polynomial CQR and technical proofs.

2 Estimation of regression function

Suppose that (t_i, y_i), $i = 1, \cdots, n$, 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 neighborhood of t_0. Let $K(\cdot)$ be a smooth kernel function, the local linear regression estimators of $m(t_0)$ is â, where

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

(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 & 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. The local least absolute deviation (LAD) polynomial regression (Fan et al. 1994, Welsh 1996) replaces the least squares loss in (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 to 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 $\rho_{\tau_k}(r) = \tau_k r - rI(r < 0)$, $k = 1, 2, \ldots, q$, be q check loss functions at q quantiles positions: $\tau_k = \dfrac{k}{q+1}$. In the linear regression model the CQR loss is defined as (Zou & Yuan 2008)

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

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

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

(2.2)

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

$$\widehat{m}(t_0) = \frac{1}{q}\sum_{k=1}^{q}\widehat{a}_k, \text{ and } \tilde{m}'(t_0) = \widehat{b}.$$

(2.3)

We refer $\hat{m}(t_0)$ to 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 nonparametric quantile regression by local fitting which has been studied in Yu & Jones (1998) and chapter 5 of Fan & 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 is 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 $(\tau_k = \frac{k}{q+1}, k = 1, 2, \ldots, q)$. When q is large, such a choice leads to nice oracle-like estimators in the oracle model selection theoretic framework (Zou & 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

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

Define

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

(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 **T** be the σ-field generated by $\{T_1, \cdots, T_n\}$.

Theorem 2.1—*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*

$$Bias(\widehat{m}(t_0)|\mathbf{T}) = \frac{1}{2} m''(t_0) \mu_2 h^2 + o_p(h^2),$$

(2.5)

$$Var(\widehat{m}(t_0)|\mathbf{T}) = \frac{1}{nh} \frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)} R_1(q) + o_p(\frac{1}{nh}).$$

(2.6)

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

$$\sqrt{nh}\{\widehat{m}(t_0) - m(t_0) - \frac{1}{2} m''(t_0) \mu_2 h^2\} \xrightarrow{\mathscr{L}} N\left(0, \frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)} R_1(q)\right).$$

(2.7)

where $\xrightarrow{\mathscr{L}}$ stands for convergence in distribution.

Remark 3—In the proof given in section 5 we assume the error distribution is symmetric. Without such a condition the asymptotic bias will have a non-varnishing 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. This is similar to the situation when using the local LAD to estimate the conditional mean function where we need to assume the mean and median of the error distribution coincide.

We see from Theorem 2.1 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, while their asymptotic variances are different. The mean squared error of $\widehat{m}(t_0)$ is

$$\mathrm{MSE}\{\widehat{m}(t_0)\}=\{\frac{1}{2}m''(t_0)\mu_2\}^2 h^4+\frac{1}{nh}\frac{v_0\sigma^2(t_0)}{f_T(t_0)}R_1(q)+o_p(h^4+\frac{1}{nh}).$$

By straightforward calculations we can see that the optimal variable bandwidth minimizing the asymptotic mean squared error of $\widehat{m}(t_0)$ is

$$h^{\mathrm{opt}}(t_0)=\left[\frac{v_0\sigma^2(t_0)R_1(q)}{f_T(t_0)\{m''(t_0)\mu_2\}^2}\right]^{1/5}n^{-1/5}.$$

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

$$h^{\mathrm{opt}}=\left[\frac{v_0R_1(q)\int\sigma^2(t)f_T^{-1}(t)w(t)dt}{\mu_2^2\int\{m''(t)\}^2w(t)dt}\right]^{1/5}n^{-1/5}.$$

The above 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 mean squared errors. The role of R_1 becomes clear in the relative efficiency study.

The local linear least squares estimator for $m(t_0)$ has the mean squared error

$$\mathrm{MSE}\{\widehat{m}_{\mathrm{LS}}(t_0)\}=\{\frac{1}{2}m''(t_0)\mu_2\}^2 h^4+\frac{1}{nh}\frac{v_0}{f_T(t_0)}\sigma^2(t_0)+o_p(h^4+\frac{1}{nh}),$$

and hence

$$h_{\text{LS}}^{\text{opt}}(t_0) = \left[\frac{\nu_0 \sigma^2(t_0)}{f_T(t_0)\{m''(t_0)\mu_2\}^2} \right]^{1/5} n^{-1/5}, \quad h_{\text{LS}}^{\text{opt}} = \left[\frac{\nu_0 \int \sigma^2(t) f_T^{-1}(t) w(t) dt}{\mu_2^2 \int \{m''(t)\}^2 w(t) dt} \right]^{1/5} n^{-1/5},$$

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

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

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 ∞,

$$\frac{\text{MSE}_{\text{opt}}\{\widehat{m}_{\text{LS}}(t_0)\}}{\text{MSE}_{\text{opt}}\{\widehat{m}(t_0)\}} \rightarrow R_1(q)^{-4/5}, \quad \frac{\text{MISE}_{\text{opt}}\{\widehat{m}_{\text{LS}}\}}{\text{MISE}_{\text{opt}}\{\widehat{m}\}} \rightarrow R_1(q)^{-4/5}.$$

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

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

The ARE only depends 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 the $\text{ARE}(\widehat{m}, \widehat{m}_{\text{LS}})$ for some commonly seen error distributions.

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, the $\text{ARE}(\widehat{m}, \widehat{m}_{\text{LS}})$ is very close to 1 regardless the choice of q in the local linear CQR estimator. When $q = 5$ the the local linear CQR only loses at most 7% efficiency, while it performs as well as the local linear least squares estimator when $q = 99$. Secondly, for all the other non-normal distributions 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 normals is often used to model the so-called contaminated data. For such distributions, the $\text{ARE}(\widehat{m}, \widehat{m}_{\text{LS}})$ can be as large as 4.9 and even more. Table 1 also indicates that, except for the Laplace error, the local CQR with $q = 5$ or $q = 9$ are significantly better than the one with $q = 1$, which becomes the local LAD 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 phenomenon is true in general, as demonstrated in the following theorem.

Theorem 2.2—$\lim_{q \to \infty} R_1(q) = 1$, *and thus* $\lim_{q \to \infty} ARE(\widehat{m}, \widehat{m}_{LS}) = 1$.

Theorem 2.2 provides us insights into the asymptotic behavior 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. On the other hand, substaintial 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 $\tilde{m}'(t_0)$ to the derivative of $m(t)$. The asymptotic bias and variance of the estimate $\tilde{m}'(t_0)$ in (2.3) are given in (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 & Marron (1991) and Fan (1992), the authors 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. The local quadratic regression is often preferred for estimating the derivative function, since it reduces the estimation bias without increasing the estimation variance (Fan & Gijbels 1992). We show here that the same phenomenon is true in local CQR smoothing.

We consider the local quadratic approximation of $m(t)$ in the neighborhood 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 $\mathbf{a} = (a_1, \cdots, a_q)$ and $\mathbf{b} = (b_1, b_2)$. We solve

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

(3.1)

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

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

(3.2)

3.1 Asymptotic properties

Denote

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

(3.3)

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

Theorem 3.1—*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 $h \rightarrow 0$ and $nh^3 \rightarrow \infty$, then the asymptotic conditional bias and variance of $\tilde{m}'(t_0)$, defined in (3.2) is given by*

$$Bias(\widehat{m'}(t_0)|\mathbf{T})=\frac{1}{6}m'''(t_0)\frac{\mu_4}{\mu_2}h^2+o_p(h^2),$$

(3.4)

$$Var(\widehat{m'}(t_0)|\mathbf{T})=\frac{1}{nh^3}\frac{v_2\sigma^2(t_0)}{\mu_2^2 f_T(t_0)}R_2(q)+o_p(\frac{1}{nh^3}).$$

(3.5)

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

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

(3.6)

Remark 4—In Theorem 3.1 the symmetric-error-distribution assumption is used to get 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 3.1.

Comparing (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, while their asymptotic variances are different.

From Theorem 3.1, the mean squared error of local quadratic CQR estimator $\widehat{m}'(t_0)$ is given by

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

Thus, the optimal variable bandwidth minimizing $\mathrm{MSE}\{\widehat{m}'(t_0)\}$ 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 $\mathrm{MISE}\widehat{m}'=\int \mathrm{MSE}\{\widehat{m}'(t)\}w(t)\,dt$ with a weight function $w(t)$. The optimal constant bandwidth minimizing the mean integrated squared error is given by

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

The above 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 mean squared error of local quadratic least squares estimator $\widehat{m}'_{\mathrm{LS}}(t_0)$ is given by

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

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

$$h^{\mathrm{opt}}(t_0)=h^{\mathrm{opt}}_{\mathrm{LS}}(t_0)R_2(q)^{1/7}, \; h^{\mathrm{opt}}=h^{\mathrm{opt}}_{\mathrm{LS}}R_2(q)^{1/7}, \tag{3.7}$$

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

$$\frac{\mathrm{MSE}_{\mathrm{opt}}\{\widehat{m}'_{\mathrm{LS}}(t_0)\}}{\mathrm{MSE}_{\mathrm{opt}}\{\widehat{m}'(t_0)\}} \to R_2(q)^{-4/7}, \; \frac{\mathrm{MISE}_{\mathrm{opt}}(\widehat{m}'_{\mathrm{LS}})}{\mathrm{MISE}_{\mathrm{opt}}(\widehat{m}')} \to R_2(q)^{-4/7}.$$

Therefore, the asymptotic relative efficiency (ARE) of the local quadratic CQR estimator ($\widehat{m}'$) with respect to the local quadratic least squares estimator ($\widehat{m}'_{\mathrm{LS}}$) is defined to be

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

The ARE only depends 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 & Yuan (2008) showed that

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

Immediately, we know that if using a large q, the 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 3.2

Let $\mathscr{F}$ denote the class of error distributions with mean 0 and variance 1, then we have

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

The lower bound is reached if and only if the error follows the rescaled Beta(2,2) distribution with mean zero and variance one. Thus

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

It is interesting to note that Theorem 3.2 provides us the *exact* lower bound of $\mathrm{ARE}(\widehat{m}',\widehat{m}'_{\mathrm{LS}})$ as $q\to\infty$. Theorem 3.2 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 3.2 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 the $\mathrm{ARE}(\widehat{m}', \widehat{m}'_{\mathrm{LS}})$ for the error distributions considered in Table 1. We also list the results in Table 2, where the column labeled $q = \infty$ shows the theoretical limit of the $\mathrm{ARE}(\widehat{m}', \widehat{m}'_{\mathrm{LS}})$. Obviously, these limits are all larger than the lower bound 0.9199. The local quadratic CQR estimator only loses 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 the $\mathrm{ARE}(\widehat{m}', \widehat{m}'_{\mathrm{LS}})$ 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 proposed estimation procedures and then demonstrate the application of the proposed method 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 MM algorithm proposed by (Hunter & Lange 2000) for solving the local CQR smoothing estimator. All the numerical results are computed using our MATLAB code, which is available upon 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 selector for the local polynomial regression. Here we consider two bandwidth selectors.

1. *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.

2. *A short-cut strategy.* In our numerical studies, we compare the local CQR and local least squares estimators. Note that in (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 & Gijbels (1996)). In addition, we are able to 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 one in which the $h_{\mathrm{LS}}^{\mathrm{opt}}$ is selected by a plug-in bandwidth selector (Ruppert, Sheather & Wand 1995). The performance of estimator $\widehat{m}(\cdot)$ and $\widehat{m}'(\cdot)$ is assessed via the average squared errors (ASE), defined by $\mathrm{ASE}(\widehat{g}) = \frac{1}{n_{\mathrm{grid}}} \sum_{k=1}^{n_{\mathrm{grid}}} \{\widehat{g}(u_k) - g(u_k)\}^2$, with g equals 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 using the ratio of average squared

errors (RASE), $\text{RASE}(\hat{g}) = \dfrac{\text{ASE}_{(\hat{g}_{\text{LS}})}}{\text{ASE}_{(\hat{g})}}$ 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.

Example 4.1—We generated 400 data set, each consisting of $n = 200$ observations, from

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

where T follows $N(0, 1)$. This model is adopted from Fan & Gijbels (1992). In our simulation, we considered five error distributions for ϵ: $N(0, 1)$, Laplace, t_3 distribution, a mixture of two normals $(0.95N(0,1) + 0.05N(0, \sigma^2))$ with $\sigma = 3,10)$. For the local polynomial CQR estimator, we consider $q = 5, 9$ and 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 proposed estimate 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 $q = 5, 9$ and 19, respectively.

Example 4.2—It is of interest to investigate the effect of heteroscedastic errors. To this end, we generated 400 simulation data sets, each consisting of $n = 200$ observations, from

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

where T follows $U(0, 1)$, $\sigma(t) = \{2 + \cos(2\pi t)\}/10$, and ϵ is the same as that in Example 4.1. 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 one. For non-normal distributions, the RASEs of the local CQR estimators can be greater than one, 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 very similarly. These findings are consistent with the theoretical analysis of AREs.

4.3 A real data example

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

We first estimated the regression function 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. Based on 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 selected bandwidths are 0.24 for regression estimation and 0.4 for derivative estimation. The CQR estimates with $q = 5,9$ and 19 with the selected bandwidths are evaluated. The CQR estimates with three different q's are very similar, we only present the CQR estimate with $q = 9$ in Figure 1.

It is interesting to see from Figure 1 that the overall patten 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 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 impact 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 the top panel of Figure 2, from which we can see that the local CQR estimate remains almost the same, while the local least squares estimate changes a lot. 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 $(2.7902, (2.7902, -26.5207)$ and $(2.8063, -6.6105)$, respectively. After distorting the two observations, we re-calculated the local CQR and the local least squares estimate. The resulting estimates are depicted in the bottom panel of Figure 2, which clearly demonstrates that the local least squares estimate changes dramatically, while the local CQR estimate is nearly un-affected by the artificial data distortion.

5 Local *p*-polynomial CQR smoothing and proofs

In this section we establish asymptotic theory of the local p-polynomial CQR estimators. We then treat Theorems 2.1 and 3.1 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

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

(5.1)

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

$$\widehat{m}(t_0) = \frac{1}{q} \sum_{k=1}^{q} \widehat{a}_k, \text{ and } \widehat{m}^{(r)}(t_0) = r! \widehat{b}_r, \ r = 1, \cdots, p.$$

(5.2)

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

  A. $m(t)$ has a continuous $(p + 2)^{th}$ derivative in the neighborhood of t_0.

  B. $f_T(\cdot)$, the marginal density function of T, is differentiable and positive in the neighborhood of t_0.

  C. The conditional variance $\sigma^2(t)$ is continuous in the neighborhood 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 $f(c_k)$,

$k = 1, \cdots, q$, S_{12} be a $q \times p$ matrix with (k, j)-element being $f(c_k)\mu_j$, $k = 1, \cdots, q$ and $j = 1, \cdots, p$, $S_{21} = S_{12}^T$, and S_{22} be a $p \times p$ matrix with (j, j')-element being $\sum_{k=1}^{q} f(c_k)\mu_{j+j'}$, for $j, j' = 1, \cdots, p$. Similarly, Let Σ_{11} be a $q \times q$ matrix with (k, k')-element $v_0 \tau_{kk'}$, $k, k' = 1, \cdots, q$, Σ_{12} be a $q \times p$ matrix with (k, j)-element being $v_j \sum_{k'=1}^{q} \tau_{kk'}$, $k = 1, \cdots, q$ and $j = 1, \cdots, p$, $\Sigma_{21} = \Sigma_{12}^T$, and Σ_{22} be a $p \times p$ matrix with (j, j')-element being $(\sum_{k,k'=1}^{q} \tau_{kk'})v_{j+j'}$, for $j, j' = 1, \cdots, p$. Define

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

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 and hereafter, we use $(\cdot)_{11}$ to denote the left-top $q \times q$ submatrix and use $(\cdot)_{22}$ to denote the right-bottom $p \times p$ 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 $\Delta_{i,k} = \dfrac{u_k}{\sqrt{nh}} + \sum_{j=1}^{p} \dfrac{v_j x_i^j}{\sqrt{nh}}$. Write $d_{i,k} = c_k[\sigma(t_i) - \sigma(t_0)] + r_{i,p}$ with $r_{i,p} = m(t_i) - \sum_{j=0}^{p} m^{(j)}(t_0)(t_i - t_0)^j/j!$. $\eta_{i,k}^*$ to be $I(\epsilon_i \leq c_k - \dfrac{d_{i,k}}{\sigma(t_i)}) - \tau_k$. let $W_n^* = (w_{11}^*, \cdots, w_{1q}^*, w_{21}^*, \cdots, w_{2p}^*)^T$ with $w_{1k}^* = \dfrac{1}{\sqrt{nh}}\sum_{i=1}^{n} K_i \eta_{i,k}^*$ and $w_{2j}^* = \dfrac{1}{\sqrt{nh}}\sum_{k=1}^{q}\sum_{i=1}^{n} K_i x_i^j \eta_{i,k}^*$.

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

Theorem 5.1

Denote $\hat{\theta}_n = (\hat{u}_1, \ldots, \hat{u}_1, \hat{v}_1, \ldots, \hat{v}_p)$ be the minimizer of (5.1). Then under the regularity conditions (A)—(C), we have

$$\widehat{\theta}_n + \frac{\sigma(t_0)}{f_T(t_0)} S^{-1} E(W_n^* | \mathbf{T}) \xrightarrow{\mathscr{L}} MVN(0, \frac{\sigma^2(t_0)}{f_T(t_0)} S^{-1} \Sigma S^{-1}).$$

To prove theorem 5.1, we first establish Lemmas 5.2—5.3.

Lemma 5.2

Minimizing (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 $\theta = (u_1, \cdots, u_q, v_1, \cdots, v_p)^T$, where

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

Proof. To apply the identity (Knight 1998)

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

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

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

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

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

This completes the proof.

Let $S_{n,11}$ be a $q \times q$ diagonal matrix with diagonal elements $f(c_k) \sum_{i=1}^{n} \frac{K_i}{nh\sigma(t_i)}, k=1, \cdots, q$;

$S_{n,12}$ be a $q \times p$ matrix with (k, j)-element $f(c_k) \sum_{i=1}^{n} \frac{K_i x_i^j}{nh\sigma(t_i)}, j=1, \cdots, p$; $S_{n,22}$ be a $p \times p$ matrix

with (j, j') element $\sum_{k=1}^{q} f(c_k) \sum_{i=1}^{n} \frac{K_i x_i^{j+j'}}{nh\sigma(t_i)}$. Denote

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

Lemma 5.3

Under Conditions (A) − (C), $L_n(\theta) = \frac{1}{2} \theta^T S_n \theta + (W_n^*)^T \theta + o_p(1)$.

Proof. Write $L_n(\theta)$ as

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

where $R_{n,k}(\theta) = B_{n,k}(\theta) - E_\epsilon[B_{n,k}(\theta)|\mathbf{T}]$. Using $F(c_k + z) - F(c_k) = zf(c_k) + o(z)$, then $\sum_{k=1}^{q} E_\epsilon[B_{n,k}(\theta)|\mathbf{T}]$ equals

$$\sum_{k=1}^{q} \sum_{i=1}^{n} \left[K_i \int_0^{\Delta_{i,k}} \left\{ \frac{z}{\sigma(t_i)} f(c_k - \frac{d_{i,k}}{\sigma(t_i)}) + o(z) \right\} dz \right]$$

$$= \sum_{k=1}^{q} \sum_{i=1}^{n} \left[K_i \Delta_{i,k}^2 \frac{f(c_k - \frac{d_{i,k}}{\sigma(t_i)})}{2\sigma(t_i)} \right] + o_p(1)$$

$$= \sum_{k=1}^{q} \sum_{i=1}^{n} \left[K_i \Delta_{i,k}^2 \frac{f(c_k)}{2\sigma(t_i)} \right] + o_p(1) = \frac{1}{2} \theta^T S_n \theta + o_p(1)$$

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

$$Var_\epsilon[B_{n,k}(\theta)|\mathbf{T}]$$

$$= \sum_{i=1}^{n} Var_\epsilon \left[\left\{ K_i \int_0^{\Delta_{i,k}} \left[I(\epsilon_i \le c_k - \frac{d_{i,k}}{\sigma(t_i)} + \frac{z}{\sigma(t_i)}) - I(\epsilon_i \le c_k - \frac{d_{i,k}}{\sigma(t_i)}) \right] dz \right\} | \mathbf{T} \right]$$

$$\le \sum_{i=1}^{n} E_\epsilon \left[\left\{ K_i \int_0^{\Delta_{i,k}} \left[I(\epsilon_i \le c_k - \frac{d_{i,k}}{\sigma(t_i)} + \frac{z}{\sigma(t_i)}) - I(\epsilon_i \le c_k - \frac{d_{i,k}}{\sigma(t_i)}) \right] dz \right\}^2 | \mathbf{T} \right]$$

$$\le \sum_{i=1}^{n} K_i^2 \int_0^{|\Delta_{i,k}|} \int_0^{|\Delta_{i,k}|} \left[F(c_k - \frac{d_{i,k}}{\sigma(t_i)} + \frac{|\Delta_{i,k}|}{\sigma(t_i)}) - F(c_k - \frac{d_{i,k}}{\sigma(t_i)}) \right] dz_1 dz_2$$

$$= o(\sum_{i=1}^{n} K_i^2 \Delta_{i,k}^2) = o_p(1).$$

Proof of Theorem 5.1—Similar to Parzen (1962), we have

$\frac{1}{nh} \sum_{i=1}^{n} K_i x_i^j \xrightarrow{P} f_T(t_0) \mu_j$, where $\xrightarrow{P}$ stands for convergence in probability. Thus,

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

This, together with Lemmas 5.2, 5.3, leads to

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

Since the convex function $L_n(\theta) - (W_n^*)^T \theta$ converges in probability to the convex function $\frac{1}{2} \frac{f_T(t_0)}{\sigma(t_0)} \theta^T S \theta$, 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

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

Denote $\eta_{i,k} = I(\epsilon_i \le c_k) - \tau_k$ and $W_n = (w_{11}, \cdots, w_{1q}, w_{21}, \cdots, w_{2p})^T$ 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 CLT for $W_n | \mathbf{T}$ holds

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

Note that $Cov(\eta_{i,k}, \eta_{i,k'}) = \tau_{kk'}, Cov(\eta_{i,k}, \eta_{j,k'}) = 0$, if $i \ne j$. Similar to Parzen (1962), we have

$\frac{1}{nh} \sum_{i=1}^{n} K_i^2 x_i^j \xrightarrow{P} f_T(t_0) \nu_j$, Therefore, $Var[W_n | \mathbf{T}] \xrightarrow{P} f_T(t_0) \Sigma$. Combined with (5.4), we have $W_n | \mathbf{T} \xrightarrow{\mathscr{L}} MVN(0, f_T(t_0)\Sigma)$. Moreover, we have

$Var(w_{1k}^* - w_{1k} | \mathbf{T})$

$$= \frac{1}{nh} \sum_{i=1}^{n} K_i^2 Var(\eta_{i,k}^* - \eta_{i,k}) \le \frac{1}{nh} \sum_{i=1}^{n} K_i^2 \{ F(c_k + \frac{|d_{i,k}|}{\sigma(t_i)}) - F(c_k) \} = o_p(1) \text{ and also } Var(w_{2j}^*$$

$$-w_{2j} | \mathbf{T}) = \frac{1}{nh} \sum_{i=1}^{n} K_i^2 x_i^j Var(\sum_{k=1}^{q} \eta_{i,k}^* - \eta_{i,k}) \le \frac{q^2}{nh} \sum_{i=1}^{n} K_i^2 x_i^j \max_k \{ F(c_k + \frac{|d_{i,k}|}{\sigma(t_i)}) - F(c) \} = o_p(1), \text{ thus } Var(W_n^*$$

$$- W_n | \mathbf{T}) = o_p(1)$$

. So by Slutsky's theorem, conditioning on $\mathbf{T}$, we have

$W_n^* | \mathbf{T} - E(W_n^* | \mathbf{T}) \xrightarrow{\mathscr{L}} MVN(0, f_T(t_0)\Sigma)$. Therefore,

$$\widehat{\theta}_n + \frac{\sigma(t_0)}{f_T(t_0)} S^{-1} E(W_n^* | \mathbf{T}) \xrightarrow{\mathscr{L}} MVN(0, \frac{\sigma^2(t_0)}{f_T(t_0)} S^{-1} \Sigma S^{-1}). \tag{5.5}$$

This completes the proof.

Proof of Theorem 2.1—The asymptotic normality follows Theorem 5.1 with $p = 1$. Let us calculate the conditional bias and variance, respectively. Denote by $e_{q \times 1}$ the vector that contains q 1's. When $p = 1$, S is a diagonal matrix with diagonal elements

$f(c_1), \cdots, f(c_q), \mu_2 \sum_{k=1}^{q} f(c_k)$. So the asymptotic conditional bias of $\widehat{m}(t_0) = \frac{1}{q} \sum_{k=1}^{q} \widehat{a}_k$ is

$$Bias(\widehat{m}(t_0) | \mathbf{T}) = \frac{1}{q} \sigma(t_0) \sum_{k=1}^{q} c_k - \frac{1}{q \cdot \sqrt{nh}} \frac{\sigma(t_0)}{f_T(t_0)} e_{q \times 1}^T (S^{-1})_{11} E(W_{1n}^* | \mathbf{T})$$

$$= \frac{1}{q} \sigma(t_0) \sum_{k=1}^{q} c_k - \frac{1}{q \cdot nh} \frac{\sigma(t_0)}{f_T(t_0)} \sum_{i=1}^{n} K_i \sum_{k=1}^{q} \frac{1}{f(c_k)} \{ F(c_k - \frac{d_{i,k}}{\sigma(t_i)}) - F(c_k) \}.$$

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

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

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

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

$$Bias(\widehat{m}(t_0)|\mathbf{T}) = \frac{1}{2} m''(t_0) \mu_2 h^2 + o_p(h^2).$$

(5.6)

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

$$Var(\widehat{m}(t_0)|\mathbf{T}) = \frac{1}{nh} \frac{\sigma^2(t_0)}{f_T(t_0)} \frac{1}{q^2} e_{q\times1}^T (S^{-1}\Sigma S^{-1})_{11} e_{q\times1} + o_p(\frac{1}{nh})$$
$$= \frac{1}{nh} \frac{v_0\sigma^2(t_0)}{f_T(t_0)} R_1(q) + o_p(\frac{1}{nh}).$$

(5.7)

By using Theorem 5.1, we can further derive the asymptotic bias and variance of $\tilde{m}'(t_0)$ given in (2.3):

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

(5.8)

$$Var(\tilde{m}'(t_0)|\mathbf{T}) = \frac{1}{nh^3} \frac{v_2\sigma^2(t_0)}{\mu_2^2 f_T(t_0)} R_2(q) + o_p(\frac{1}{nh^3}).$$

(5.9)

Proof of Theorem 2.2—Note that

$$\lim_{q\to\infty} R_1(q) = \int_0^1 \int_0^1 \frac{s_1 \wedge s_2 - s_1 s_2}{f(F^{-1}(s_1))f(F^{-1}(s_2))} ds_1 ds_2$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(z_1) \wedge F(z_2) - F(z_1)F(z_2)) \, dz_1 dz_2.$$

(5.10)

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

$$G(s) = \int_{-\infty}^{s} (s - x)f(x)dx = sF(s) - k_1(s),$$

(5.11)

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

$$2H(s) = \int_{-\infty}^{s}(s-x)^2 f(x)dx = s^2 F(s) - 2sk_1(s) + k_2(s),$$

(5.12)

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

$$2\int_{-\infty}^{\infty}\left(\int_{z_1}^{\infty} f(t)dt\right)G(z_1)dz_1 = 2\int_{-\infty}^{\infty} f(t)\left(\int_{-\infty}^{t} G(z_1)dz_1\right)dt = \int_{-\infty}^{\infty} 2f(t)H(t)dt.$$

(5.13)

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

Proof of Theorem 3.1—We apply Theorem 5.1 to get the asymptotic normality. Denote by e_r the p-vector $(0, 0, \cdots, 1, 0, \cdots, 0)^T$ with 1 on the rth position. When $p = 2$, S_{12} has the following

$$S_{12} = (0_{q\times 1},$$

$$\mu_2(f(c_k))_{q\times 1}), \quad \text{and } S_{22} = \text{diag}(\mu_2\sum_{k=1}^{q} f(c_k),$$

$$\mu_4\sum_{k=1}^{q} f(c_k)). \text{ Since } S_{11} = \text{diag}(f(c_1),$$

$$\cdots, f(c_q)), \quad (S^{-1})_{22}$$

$$= (S_{22} - S_{21}S_{11}^{-1}S_{12})^{-1}$$

$$= \text{diag}((\mu_2\sum_{k=1}^{q} f(c_k))^{-1},$$

forms $\qquad \{(\mu_4 - \mu_2^2)\sum_{k=1}^{q} f(c_k)\}^{-1})$.

Note that $(S^{-1})_{21} = -(S^{-1})_{22}S_{21}S_{11}^{-1}$. Thus, $(S^{-1})_{21}$ equals $(0_{q\times 1}, [\mu_2/\{(\mu_4 - \mu_2^2)\sum_{k=1}^{q} f(c_k)\}]1_{q\times 1})^T$. By Theorem 5.1

$$\begin{aligned}
Bias(\widehat{m}'(t_0)|\mathbf{T}) &= -\frac{\sigma(t_0)}{hf_T(t_0)}\frac{1}{\sqrt{nh}}e_1^T\{(S^{-1})_{21}E(W_{1n}^*|\mathbf{T}) + (S^{-1})_{22}E(W_{2n}^*|\mathbf{T})\} \\
&= -\frac{\sigma(t_0)}{hf_T(t_0)}\frac{1}{\mu_2\sum_{k=1}^{q} f(c_k)}\frac{1}{\sqrt{nh}}E(w_{21}^*|\mathbf{T}).
\end{aligned}$$

Note that $E(w_{2j}^*|\mathbf{T}) = \dfrac{1}{\sqrt{nh}}\sum_{i=1}^{n} K_i x_i^j \sum_{k=1}^{q}\{F(c_k - \dfrac{d_{i,k}}{\sigma(t_i)}) - F(c_k)\}$. Similarly, under condition (D), we have $\sum_{k=1}^{q}\{F(c_k - \dfrac{d_{i,k}}{\sigma(t_i)}) - F(c_k)\} = -\sum_{k=1}^{q} f(c_k)\cdot\dfrac{r_{i,p}}{\sigma(t_i)}\{1+o_p(1)\}$. Therefore, $Bias$ $(\widehat{m}'(t_0|\mathbf{T})$ is equal to $\dfrac{1}{nh^2}\dfrac{\sigma(t_0)}{f_T(t_0)}\sum_{i=1}^{n} K_i x_i\dfrac{r_{i,p}}{\sigma(t_i)}\{1+o_p(1)\}$. 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

$$Bias(\widehat{m}'(t_0)|\mathbf{T}) = \frac{1}{6}m'''(t_0)\frac{\mu_4}{\mu_2}h^2 + o_p(h^2).$$

(5.14)

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

$$
\begin{aligned}
Var(\widehat{m}'(t_0)|\mathbf{T}) &= \frac{1}{nh^3}\frac{\sigma^2(t_0)}{f_T(t_0)}e_1^T(S^{-1}\Sigma S^{-1})_{22}e_1 + o_p(\frac{1}{nh^3}),\\
&= \frac{1}{nh^3}\frac{\nu_2\sigma^2(t_0)}{\mu_2^2 f_T(t_0)}R_2(q) + o_p(\frac{1}{nh^3}).
\end{aligned}
$$

(5.15)

which completes the proof.

Proof of Theorem 3.2—From Zou & Yuan (2008), we know that

$$\lim_{q\to\infty}\left(\sum_{k=1}^{q}f(c_k)\right)^2 \bigg/ \left(\sum_{k=1}^{q}\sum_{k'=1}^{q}\tau_{kk'}\right) = 12E_F^2[f(\epsilon)] = 12\left(\int f^2(x)dx\right)^2.$$

Thus $\lim_{q\to\infty}\frac{1}{R_2(q)} = 12(\int f^2(x)dx)^2$. We notice that $12(\int f^2(x)dx)^2$ is also the asymptotic Pitman efficiency of the Wilcoxon test relative to the t-test Hodges & Lehmann (1956). For the rest of the proof, readers are referred to Hodges & 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 proposed methodology is valid even when the error variance is infinite.

- **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/quadratic CQR estimator is the same as that of the local linear/quadratic LS estimator, which indicates that the local CQR estimator enjoys the property of automatical boundary correction, a nice property of local LS estimator. Furthermore, the asymptotic relative efficiency remains exactly the same as that for interior points.

- **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 detailed theoretical proof of the above claims, we refer interested readers to a supplementary file (Kai, Li & Zou 2009) of this paper, where we also provide additional simulation results to support the theory. We opt not to show these results here due to space limit.

In this paper, we focus on the local CQR estimate for the nonparametric regression model. The proposed methodology and theory may be extended to the settings in 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 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 seconds on an AMD 1.9GHz machine. The MM implementation seems to be more efficient than the standard linear programming. We discuss the computing algorithm in details in a separate article.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Chu C-K, Marron JS. Choosing a kernel regression estimator. Statist. Sci 1991;6(4):404–436.

Fan J. Design-adaptive nonparametric regression. J. Amer. Statist. Assoc 1992;87(420):998–1004.

Fan J, Gijbels I. Variable bandwidth and local linear regression smoothers. Ann. Statist 1992;20(4):2008–2036.

Fan, J.; Gijbels, I. Local polynomial modelling and its applications. London: Chapman & Hall; 1996.

Fan J, Hu TC, Truong YK. Robust non-parametric function estimation. Scand. J. Statist 1994;21(4):433–446.

Hodges J, Lehmann E. The efficiency of some nonparametric competitors of the t-test. Ann. Math. Stat 1956;27(2):324–335.

Hunter DR, Lange K. Quantile regression via an MM algorithm. Journal of Computational and Graphical Statistics 2000;9(1):60–77.

Kai B, Li R, Zou H. Supplementary materials for "local cqr smoothing: An efficient and safe alternative to local polynomial regression". 2009 Technical report, http://www.stat.psu.edu/rli/research/Supplement-of-localCQR.pdf.

Knight K. Limiting distributions for L_1 regression estimators under general conditions. Ann. Statist 1998;26(2):755–770.

Koenker R. A note on l-estimates for linear models. Stat. and Prob. Letters 1984;2(6):323–325.

Koenker, R. Quantile regression. Cambridge: Cambridge University Press; 2005.

Parzen E. On estimation of a probability density function and mode. Ann. Math. Statist 1962;33:1065–1076.

Pollard D. Asymptotics for least absolute deviation regression estimators. Econometric Theory 1991;7(2):186–199.

Ruppert D, Sheather SJ, Wand MP. An effective bandwidth selector for local least squares regression. J. Amer. Statist. Assoc 1995;90(432):1257–1270.

Welsh AH. Robust estimation of smooth regression and spread functions and their derivatives. Statist. Sinica 1996;6(2):347–366.

Yu K, Jones MC. Local linear quantile regression. J. Amer. Statist. Assoc 1998;93(441):228–237.

Zou H, Yuan M. Composite quantile regression and the oracle model selection theory. The Annals of Statistics 2008;36(3):1108–1126.
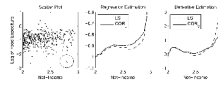
Figure 1.
The left panel is the scatter plot of data, the middle panel is the estimated regression function, and the right panel is the estimated derivative function.

Figure 2.
Plot of estimated regression function and its derivative. The top panel is for the estimate removing the two possible outliers, and bottom panel is for the estimate moving the two possible outliers to more extreme cases. The left panel is for the estimated regression function, and the right panel is the estimated derivative function.
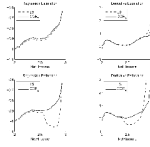
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Comparisons of ARE $\hat{m}$, $\hat{m}_{LS}$

Error Distribution	ARE ($\hat{m}$, $\hat{m}_{LS}$)					
	$q = 1$	$q = 5$	$q = 9$	$q = 19$	$q = 99$	
$N(0, 1)$	0.6968	0.9339	0.9659	0.9858	0.9980	
Laplace	1.7411	1.2199	1.1548	1.0960	1.0296	
t-distribution with $df = 3$	1.4718	1.5967	1.5241	1.4181	1.2323	
t-distribution with $df = 4$	1.0988	1.2652	1.2377	1.1872	1.0929	
.95N(0, 1) + .05N(0, 3²)	0.8639	1.1300	1.1536	1.1540	1.0804	
.90N(0, 1) + .10N(0, 3²)	0.9986	1.2712	1.2768	1.2393	1.0506	
.95N(0, 1) + .05N(0, 10²)	2.6960	3.4577	3.4783	3.3591	1.3498	
.90N(0, 1) + .10N(0, 10²)	4.0505	4.9128	4.7049	3.5444	1.1379	

Table 2

Comparisons of $\mathrm{ARE}(\widehat{m}', \widehat{m}'_{LS})$

Error Distribution	$\mathrm{ARE}(\widehat{m}', \widehat{m}'_{LS})$						
	$q = 1$	$q = 5$	$q = 9$	$q = 19$	$q = 99$	$q = \infty$
$N(0, 1)$	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 $df = 3$	1.3179	1.4405	1.4435	1.4435	1.4430	1.4431
t-distribution with $df = 4$	1.0696	1.2038	1.2104	1.2123	1.2125	1.2125
$.95N(0, 1) + .05N(0, 3^2)$	0.9008	1.0867	1.1019	1.1073	1.1077	1.1077
$.90N(0, 1) + .10N(0, 3^2)$	0.9990	1.1869	1.1982	1.1999	1.1987	1.1987
$.95N(0, 1) + .05N(0, 10^2)$	2.0308	2.4229	2.4466	2.4482	2.4415	2.4415
$.90N(0, 1) + .10N(0, 10^2)$	2.7160	3.1453	3.1430	3.1135	3.1094	3.1093

Table 3

Simulation Results for Example 4.1

	$\hat{m}$			$\hat{m}'$		
	RASE	t = 0.75		RASE	t = 0.75	
	Mean(SD)	Bias	Std	Mean(SD)	Bias	Std
Standard Normal						
LS	—	−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						
LS	—	−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 df = 3						
LS	—	−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
.95N(0, 1) + .05N(0,9)						
LS	—	−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
.95N(0, 1) + .05N(0,100)						
LS	—	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

	$\hat{m}$				$\hat{m}'$		
	RASE **Mean(SD)**	**$t = 0.75$**		**RASE** **Mean(SD)**	**$t = 0.75$**		
		Bias	**Std**		**Bias**	**Std**	
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

Table 4

Simulation Results for Example 4.2.

	$\hat{m}$			$\hat{m}'$		
	RASE	$t = 0.4$		RASE	$t = 0.4$	
	Mean(SD)	Bias	Std	Mean(SD)	Bias	Std
Standard Normal						
LS	—	−0.0177	0.0263	—	0.0329	0.2753
CQR5	0.9574(0.1699)	−0.0166	0.0271	0.9376(0.3587)	0.0289	0.3019
CQR9	0.9783(0.1286)	−0.0165	0.0266	0.9458(0.3092)	0.0283	0.3013
CQR19	0.9838(0.0815)	−0.0168	0.0266	0.9491(0.2952)	0.0278	0.2962
Laplace						
LS	—	−0.0175	0.0249	—	0.0236	0.2718
CQR5	1.1938(0.3279)	−0.0145	0.0237	1.2063(0.6794)	0.0106	0.2701
CQR9	1.1405(0.2523)	−0.0150	0.0243	1.2046(0.6413)	0.0079	0.2719
CQR19	1.0857(0.1584)	−0.0157	0.0248	1.2019(0.6035)	0.0098	0.2693
t-distribution with df = 3						
LS	—	−0.0167	0.0261	—	0.0025	0.3068
CQR5	1.5974(1.0324)	−0.0120	0.0229	1.6099(1.7558)	0.0004	0.2503
CQR9	1.4247(0.8170)	−0.0132	0.0228	1.5975(1.8047)	−0.0002	0.2560
CQR19	1.2111(0.4330)	−0.0140	0.0242	1.5948(1.8291)	0.0006	0.2567
.95N(0, 1) + .05N(0,9)						
LS	—	−0.0175	0.0247	—	−0.0130	0.2916
CQR5	1.1788(0.6248)	−0.0157	0.0228	1.2268(2.0608)	−0.0050	0.2778
CQR9	1.1507(0.4715)	−0.0157	0.0230	1.2132(1.8791)	−0.0048	0.2754
CQR19	1.0835(0.2603)	−0.0159	0.0234	1.2104(1.8546)	−0.0066	0.2742
.95N(0, 1) + .05N(0,100)						
LS	—	−0.0162	0.0260	—	0.0335	0.3728
CQR5	3.1661(2.4820)	−0.0077	0.0173	3.0593(5.6699)	0.0245	0.2420

	RASE Mean(SD)	$t = 0.4$ Bias	Std
	$\hat{m}$		
CQR$_9$	2.4179$_{(1.7012)}$	−0.0080	0.0171
CQR$_{19}$	1.3469$_{(0.5075)}$	−0.0085	0.0241

	RASE Mean(SD)	$t = 0.4$ Bias	Std
	$\hat{m}'$		
	3.0287$_{(5.3443)}$	0.0209	0.2533
	3.0146$_{(5.2728)}$	0.0234	0.2452