Journal of Econometrics 188 (2015) 502–513

Contents lists available at ScienceDirect

Journal of Econometrics

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

Optimal smoothing in nonparametric conditional quantile derivative function estimation☆

Wei Lin[a], Zongwu Cai[b,c], Zheng Li[d], Li Su[a,*]

[a] ISEM, Capital University of Economics and Business, Beijing, 100070, China

[b] Department of Economics, Kansas University, USA

[c] WISE and School of Economics, Xiamen University, 361005, China

[d] Department of Economics, Texas A&M University, USA

ARTICLE INFO

Article history:
Available online 14 March 2015

JEL classification:
C14
C18
C21

Keywords:
Gradient estimation
Local polynomial smoothing
Lease squares cross validation
Quantile regression

ABSTRACT

Marginal effect in nonparametric quantile regression is of special interest as it quantitatively measures how one unit change in explanatory variable heterogeneously affects dependent variable *ceteris paribus* at distinct quantiles. In this paper, we propose a data-driven bandwidth selection procedure based on the gradient of an unknown quantile regression function. Our method delivers the bandwidth with the oracle property in the sense that it is asymptotically equivalent to the optimal bandwidth if the true gradient were known. The results of Monte Carlo simulations are reported, and the finite sample performance of our proposed method confirms our theoretical analysis. An empirical application is also provided, showing that our proposed method delivers more reasonable and reliable quantile derivative estimates than traditional cross validation method.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Since the introduction of quantile regression by Koenker and Bassett's (1978) seminal paper and acknowledgment of its main advantages in robustness to outliers and in characterization of heterogeneous effects, quantile regression model has been widely used in a variety of research fields including economics, finance, medical and environmental sciences etc., and at the same time further developed in multiple important directions, such as censored data and time series data analysis etc. For example, Powell (1986) and Buchinsky and Hahn (1998) consider regression quantiles for censored data. Chernozhukov and Hansen (2004) study the effects of 401(k) participation on the wealth distribution using instrumental quantile regression model. Buchinsky (1994) uses quantile regression approach to describe conditional wage distribution and studied within-group wage inequality and heterogeneity of returns to schooling and experience. Cai (2002) considers quantile regressions with time series data. Engle and Manganelli (2004) propose estimating value at risk (VaR) with quantile regression. Koenker and Xiao (2002) propose quantile autoregressive (QAR) model, and study inference on unit root quantile regression models. Xiao (2009) proposes quantile cointegrating regression and studied the relationship between prices and market fundamentals using the US stock index data. For other quantile regression models and empirical applications in different areas, readers are referred to Yu et al. (2003), Koenker (2005), Wei et al. (2006) and Cade and Noon (2003).

In the analysis of conditional quantile regression, *marginal effect* is of special interest in economic analysis as it provides a good approximation to the amount of change of dependent variable at distinct quantiles in response to a unit change in the regressors. For example, Buchinsky (1994) finds that the returns to education are higher at the higher quantiles of conditional wage distribution, while the returns to experience are higher at the lower quantiles, especially during the early 1970s. Chernozhukov and Hansen (2004) find that there is significant heterogeneity in the effect of 401(k) participation on net financial assets.

Parametric quantile regression models employed in the literature have the common caveat that they may suffer from misspecification in the functional form of conditional quantiles. Leaving conditional quantile function unspecified in the model, nonparametric quantile regression is an immediate remedy for such misspecification. Among the many research directions of quantile

☆ Cai's research was supported, in part, by the National Nature Science Foundation of China grants #71131008 (Key Project).

* Corresponding author.
E-mail address: suli0917@163.com (L. Su).

http://dx.doi.org/10.1016/j.jeconom.2015.03.014
0304-4076/© 2015 Elsevier B.V. All rights reserved.

analysis, nonparametric techniques in quantile regression develops almost in parallel with the parametric approach. Stone (1977) introduces locally constant nonparametric quantile estimator and establishes its consistency and rates of convergence. Chaudhuri (1991) generalizes the results to locally polynomial quantile regression models using local Bahadur type representation. Chaudhuri et al. (1997) consider nonparametric estimation of average derivative of conditional quantile function. Their estimator is the weighted average of pointwise derivatives at each observations, and has root-n convergence rate. Yu and Jones (1998) consider local linear quantile regression using check function approach and double-kernel approach. Li and Racine (2008) introduce nonparametric estimation of conditional quantile functions based on estimates of cumulative distribution function when regressors include both continuous and discrete variables; and later on, Li et al. (2013) propose a data-driven least-square cross-validation method to optimally select smoothing parameters. Cai and Xu (2008) propose quantile regression methods for a class of smooth coefficient time series models. Lin and Li (2007) consider the estimation of quantile function with association dependence based on the L_1-norm kernel and establish asymptotic normality for their estimator. Hallin et al. (2009) propose spatial quantile regression model, which introduces nonparametric quantile method to spatial modeling framework. Cai and Xiao (2012) study both parametric and nonparametric quantile regression estimation for dynamic models with partially varying coefficients in which some coefficients are functions of covariates.

It is well known that the result of nonparametric estimation depends sensitively on the choice of smoothing parameter or bandwidth. For nonparametric conditional mean regression there are rich literature on bandwidth selection. See Rice (1984), Härdle and Vieu (1992), Hall et al. (1995), Xia and Li (2002), and Leung (2005) forcontinuous regressors case; see Li and Racine (2004), Racine and Li (2004), Racine and Li (2004), and Li, Simar and Zelenyuk (2014) for mixed continuous and categorical regressors case. However, there is not much work done on this issue for nonparametric estimate of conditional quantile function. Yu and Jones (1998) and Yu and Lu (2004) propose a selection method based on the *ad hoc* relationship of optimal bandwidths for conditional mean and quantile regression under the assumption that the second order derivatives of the quantile function are parallel. This assumption, although simplifying bandwidth selection for nonparametric quantile estimation, might not be valid for many applications because of nonlinear heteroskedasticity in conditional quantile functions. Alternatively, Cai and Xu (2008) construct a nonparametric version of the bias-corrected AIC to select the bandwidth. Their method address the structure of time series data and the overfitting or underfitting tendency. Each of the bandwidth selection methods mentioned above either depends on stringent restrictions on curvatures of quantile functions or is in need of pilot bandwidth under researcher's judgment. To our best knowledge, there is no published work on bandwidth selection for local polynomial quantile estimator or the quantile derivative estimator under check function approach with continuous covariates and mixed type (continuous and categorical) covariates. Recently, Li et al. (submitted for publication) propose a fully data driven cross validation bandwidth selection method for the local linear quantile estimator with mixed continuous and discrete data.

The contribution of this paper is to propose a fully automatic and data-driven bandwidth selection method for a local linear quantile derivative estimator. Borrowing the idea from Henderson et al. (2015), we propose to use local cubic derivative estimator to replace the unknown true gradient in the oracle LSCV setup for the gradient in Müller et al. (1987) with the local linear estimator. After this modification, the leading bias term in cross validation (CV) function is unchanged as the bias from local cubic estimator has a much smaller order than local linear estimator; and the leading variance in CV function becomes the variance of difference between the two estimators. After rescaling by a constant which only depends on kernel function, the variance of this difference is the same as the variance in the oracle LSCV setup with unknown oracle gradient. Therefore, the proposed bandwidth selection method possesses the oracle property that selected bandwidth is asymptotically equivalent to that selected with unknown oracle gradient.

The main advantage of this gradient based bandwidth selection method lies within the fact that it is fully data driven, and conveniently implemented without heavy computational burden or the need of pilot estimates/bandwidths under the researchers' subjective judgment which might strongly influences the results of nonparametric estimates. With a clear intuition, this selection method also readily applies to multivariate cases.

The remainder of this paper is organized as follows. Section 2 gives the formal details of our cross validation procedure and the asymptotic result of our proposed method. After that, we provide Monte Carlo simulation results in Section 3, comparing the finite sample performance of our bandwidth selection method to the oracle selection method for the estimation of nonparametric quantile derivative functions. An illustrative example is presented in Section 4, and Section 5 concludes this paper. All the technical details of derivations are relegated to the Appendix.

2. Methodology

Without loss of generality, we first consider the general univariate nonparametric quantile regression model

$$y_i = q_\tau(x_i) + e_i,$$

where the covariate x_i in the equation is a scalar variable, $q_\tau(x_i)$ is the τth quantile of y_i conditioned on x_i, i.e., $\Pr(y_i \le q_\tau(x_i)|x_i) = \tau$, and the functional form of quantile function $q_\tau(\cdot)$ is unknown and unspecified in the model. The multivariate case, in which x_i includes multiple variables, will be discussed at the end of Section 2. We are interested in the first order derivative of conditional quantile function, denoted as $\beta(x) \equiv dq_\tau(x)/dx$. Let $\hat{\beta}_{LL}(x)$ be the local linear estimator of $\beta(x)$ obtained from[1]

$$\min_{a,b} \sum_{i=1}^{n} \rho_\tau(y_i - a - b(x_i - x))K_{h,ix} \tag{1}$$

where $\rho_\tau(v) = v[\tau - \mathbf{1}(v \le 0)]$ with $0 < \tau < 1$, b estimates $dq_\tau(x)/dx$, $K_{h,ix} = h^{-1}K((x_i - x)/h)$ is kernel function, and h is the smoothing parameter.

Theoretically, the bandwidth h for derivative estimator should be chosen so that it minimizes the expected mean squared error $E\{[\hat{\beta}_{LL}(x) - \beta(x)]^2\}$. In practice, we can choose bandwidth h that minimizes its sample analogue,

$$CV(h) \equiv \frac{1}{n}\sum_{i=1}^{n}[\hat{\beta}_{LL}(x_i) - \beta(x_i)]^2 M(x_i), \tag{2}$$

where $M(\cdot)$ is a weight function with bounded support that trims out data that are close to the boundary of the support of x. The leading term of $CV(h)$ in (2), denoted as $CV_0(h)$, is

$$\begin{aligned} CV_0(h) &= \int E[\hat{\beta}_{LL}(x) - \beta(x)]^2 M(x)f(x)dx \\ &= \int \left[Bias_0^2(\hat{\beta}_{LL}(x)) + Var_0(\hat{\beta}_{LL}(x))\right] M(x)f(x)dx + (s.o.) \end{aligned} \tag{3}$$

[1] It is obvious that the derivative $\beta(x)$ and its estimate $\hat{\beta}_{LL}(x)$ depend on quantile level τ. The appearance of τ in the subscripts is suppressed in this paper for notational convenience.

where $f(x)$ is the density function of x, $Bias_0(\hat{\beta}_{LL}(x))$ and $Var_0(\hat{\beta}_{LL}(x))$ are the leading bias and leading variance terms of $\hat{\beta}_{LL}(x)$, and $(s.o.)$ are terms that have smaller order in probability than the first leading term, see Racine and Li (2004), Li and Racine (2006) and Hall et al. (2007). Although the true value of $\beta(x)$ is unknown, one can calculate the leading bias and leading variance terms of $\hat{\beta}_{LL}(x)$, which are[2]

$$Bias_0(\hat{\beta}_{LL}(x)) = h^2 \frac{(\mu_4 - \mu_2^2)\xi'(x)q_\tau''(x)}{2\mu_2\xi(x)} \equiv h^2 B_1(x) \tag{4}$$

$$Var_0(\hat{\beta}_{LL}(x)) = \frac{1}{nh^3}\frac{\tau(1-\tau)\nu_2}{\mu_2^2}\frac{f(x)}{\xi^2(x)} \equiv \frac{1}{nh^3}V_1\Omega(x) \tag{5}$$

where

$$B_1(x) = \frac{(\mu_4 - \mu_2^2)\xi'(x)q_\tau''(x)}{2\mu_2\xi(x)}, \qquad V_1 = \frac{\tau(1-\tau)\nu_2}{\mu_2^2},$$
$$\Omega(x) = \frac{f(x)}{\xi^2(x)}, \qquad \xi(x) = f_e(0|x)f(x), \tag{6}$$

$f_e(\cdot|x)$ is conditional pdf of e given x, $\mu_j = \int K(v)v^j dv$, and $\nu_j = \int K^2(v)v^j dv$. Therefore, it is easy to see that the optimal bandwidth $h_{0,opt}$ for derivative estimator that minimizes $CV_0(h)$ is

$$h_{0,opt} = \left[\frac{3V_1\int \Omega(x)M(x)f(x)dx}{4\int B_1^2(x)M(x)f(x)dx}\right]^{\frac{1}{7}} n^{-\frac{1}{7}}. \tag{7}$$

However, the result of $h_{0,opt}$ is infeasible and cannot be used directly, as there are unknown functions, such as $f(x)$, $\xi(x)$, $\xi'(x)$, and $q_\tau''(x)$, in the expression. One need to estimate these unknown functions with some initial bandwidths first, then plug those pilot estimates into the formula (7), which is known as the "plug-in" method proposed by Fan and Gijbels (1995). This method has two disadvantages. Theoretically, the "plug-in" formula in (7) can be too complicated to use especially when there are multiple regressors of different types (continuous and discrete) involved in the model. In practice, the "plug-in" method is not fully automatic as it depends on the initial bandwidths which are needed for estimating those unknown functions. If the initial selection of bandwidth is poor, the "plug-in" will also behave poorly.

In this paper, we borrow the idea from Henderson et al. (2015) to devise an alternative procedure for the selection of bandwidth, denoted by $\hat{h}_{0,opt}$, which is fully data-driven without selection of initial bandwidth, and it is equivalent to the infeasible $h_{0,opt}$ in the sense that it nearly minimizes the $CV_0(h)$ in (3), i.e., $\hat{h}_{0,opt}/h_{0,opt} = 1 + o_P(1)$.

The idea of constructing feasible version of $h_{0,opt}$ is clear: we use a consistent estimate $\hat{\beta}(x)$ of the derivative function $\beta(x)$ to replace the true value in (3). The objective function used to select h now becomes

$$CV_{LCB}(h) \equiv \frac{1}{n}\sum_{i=1}^{n}[\hat{\beta}_{LL}(x_i) - \hat{\beta}(x_i)]^2 M(x_i). \tag{8}$$

The natural candidate for $\hat{\beta}(x)$ can be local polynomial quantile derivative estimator with higher orders. In this paper, we choose local cubic estimator, i.e., $\hat{\beta}(x) = \hat{\beta}_{LCB}(x)$ (the subscript "LCB" stands for Local Cubic) obtained from

$$\min_{a,b}\sum_{i=1}^{n}\rho_\tau(y_i - a - b_1(x_i - x) - b_2(x_i - x)^2 - b_3(x_i - x)^3)K_{h,ix}$$

where b_1 estimates $dq_\tau(x)/dx$. Similarly, the leading term of $CV_{LCB}(h)$ is

$$CV_{0,LCB}(h) = \int MSE_0[\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x)]M(x)f(x)dx \tag{9}$$
$$= \int\Big[Bias_0^2(\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x)) + Var_0\Big(\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x)\Big)\Big]M(x)f(x)dx. \tag{10}$$

It is easy to see that $Bias_0(\hat{\beta}_{LL}(x))$ dominates the bias term $Bias_0(\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x))$ as the leading bias of local linear quantile derivative estimator is $O(h^2)$ while that of local cubic one is $O(h^4)$. In addition, to calculate the variance of $\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x)$, one needs to know the leading variances of both estimators and their covariance. Here we summarize the results

$$Bias_0[\hat{\beta}_{LCB}(x)] = O(h^4), \tag{11}$$

$$Var_0[\hat{\beta}_{LCB}(x)] = \frac{1}{nh^3}\frac{\tau(1-\tau)(\mu_6^2\nu_2 + \mu_4^2\nu_6 - 2\mu_4\mu_6\nu_4)}{(\mu_2\mu_6 - \mu_4^2)^2}\frac{f(x)}{\xi^2(x)} \equiv \frac{1}{nh^3}V_3\Omega(x), \tag{12}$$

$$Cov_0[\hat{\beta}_{LL}(x), \hat{\beta}_{LCB}(x)] = \frac{1}{nh^3}\frac{\tau(1-\tau)(\mu_6\nu_2 - \mu_4\nu_4)}{\mu_2(\mu_2\mu_6 - \mu_4^2)}\frac{f(x)}{\xi^2(x)} \equiv \frac{1}{nh^3}V_2\Omega(x), \tag{13}$$

where

$$V_2 = \frac{\tau(1-\tau)(\mu_6\nu_2 - \mu_4\nu_4)}{\mu_2(\mu_2\mu_6 - \mu_4^2)},$$
$$V_3 = \frac{\tau(1-\tau)(\mu_6^2\nu_2 + \mu_4^2\nu_6 - 2\mu_4\mu_6\nu_4)}{(\mu_2\mu_6 - \mu_4^2)^2}.$$

Substitute (4), (5), (11), (12), (13) into (10), we have that

$$CV_{0,LCB}(h) = \int\left(h^4B_1^2(x) + \frac{V_{1,3}}{nh^3}\Omega(x)\right)M(x)f(x)dx$$

where

$$V_{1,3} \equiv V_1 + V_3 - 2V_2 = \frac{\tau(1-\tau)\mu_4^2(\mu_2^2\nu_6 - 2\mu_2\mu_4\nu_4 + \mu_4^2\nu_2)}{\mu_2^2(\mu_2\mu_6 - \mu_4^2)^2}. \tag{14}$$

Correspondingly, the optimal choice of bandwidth h that minimizes $CV_{0,LCB}(h)$ is

$$h_{0,cubic} = \left[\frac{3V_{1,3}\int \Omega(x)M(x)f(x)dx}{4\int B_1^2(x)M(x)f(x)dx}\right]^{\frac{1}{7}} n^{-\frac{1}{7}}. \tag{15}$$

Note that the ratio of $h_{0,opt}$ to $h_{0,cubic}$ is a constant and free from unknown functions

$$\frac{h_{0,opt}}{h_{0,cubic}} = \left(\frac{V_1}{V_{1,3}}\right)^{\frac{1}{7}} = \left(\frac{\nu_2(\mu_2\mu_6 - \mu_4^2)^2}{\mu_4^2(\mu_2^2\nu_6 - 2\mu_2\mu_4\nu_4 + \mu_4^2\nu_2)}\right)^{\frac{1}{7}}. \tag{16}$$

Given such result, one can first obtain the bandwidth $\hat{h}_{cubic}$ by minimizing the feasible cross validation objective function (8), then rescale $\hat{h}_{cubic}$ by the ratio in (16) to obtain $\hat{h}_{0,opt}$ which nearly minimizes the infeasible cross validation objective function (2) in the sense that

$$\frac{\hat{h}_{0,opt}}{h_{0,opt}} = \frac{(V_1/V_{1,3})^{1/7}\hat{h}_{cubic}}{h_{0,opt}} = \frac{(V_1/V_{1,3})^{1/7}[h_{0,cubic} + o_P(h_{0,cubic})]}{h_{0,opt}} = 1 + o_P(1),$$

2 The derivations are in the Appendix.

where the second equality holds under some regularity conditions that are similar to those in Hall et al. (2007).

It is easy to compute the ratio $(V_1/V_{1,3})^{1/7}$ for commonly used kernel functions. For the Epanechnikov kernel, $K(v) = 0.75(1-v^2)\mathbf{1}(|v| \leq 1)$, $\mu_2 = 1/5$, $\mu_4 = 3/35$, $\mu_6 = 1/21$, $\nu_2 = 3/35$, $\nu_4 = 1/35$, $\nu_6 = 1/77$, and $(V_1/V_{1,3})^{1/7} = (44/135)^{1/7} \approx 0.8520$. For the Gaussian kernel, we have that $K(v) = (\sqrt{2\pi})^{-1}\exp(-v^2/2)$, $\mu_2 = 1$, $\mu_4 = 3$, $\mu_6 = 15$, $\nu_2 = 1/(4\sqrt{\pi})$, $\nu_4 = 3/(8\sqrt{\pi})$, $\nu_6 = 15/(16\sqrt{\pi})$, and $(V_1/V_{1,3})^{1/7} = (16/15)^{1/7} \approx 1.0093$, which, being very close to 1, implies that $\hat{h}_{cubic}$ can be almost regarded as $\hat{h}_{0,opt}$ without further adjustment. The two ratio values for the Gaussian and Epanechnikov kernels are the same as those in Henderson et al. (2015) which considers nonparametric estimation of derivatives of conditional mean function. This coincidence occurs not only for these two specific kernel functions but also for any other kernel functions, as we show in the Appendix that the expression for the ratio $(V_1/V_{1,3})^{1/7}$ given in (16) is the same as that in Henderson et al. (2015).

Although demonstrated in univariate case, after proper adaptation with heavier notations, this method can be also applied to multivariate cases. Because of space limitations, we only outline the bandwidth selection procedure for practical implementation, and report the main asymptotic results for the selected bandwidth.

Suppose that there are p regressors in the conditional quantile function $q_\tau(x)$ where $x = (x_{i1}, x_{i2}, \ldots, x_{ip})$. We want to choose bandwidth vector $\mathbf{h} = (h_1, \ldots, h_p)$ optimally in the sense that they minimize the mean squared error for the first order derivative quantile functions $q_\tau(x)$. Let $\beta_s(x) = \partial q_\tau(x)/\partial x_s$ for $s = 1, \ldots, p$ denote the first order partial derivative functions, and we consider each partial derivative separately and, without loss of generality, focus on the partial derivative with respect to the first regressor i.e. $s = 1$. The oracle (infeasible) cross validation function is

$$CV_1(\mathbf{h}) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\beta}_{1,LL}(x_i) - \beta_1(x_i)\right]^2 M(x_i). \tag{17}$$

Similar to the univariate case, replacing the true unknown partial derivative function with its consistent estimate from local cubic quantile regression, we propose to choose $\mathbf{h}$ to minimize the following feasible cross validation function,

$$CV_{1,f}(\mathbf{h}) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\beta}_{1,LL}(x_i) - \hat{\beta}_{1,LCB}(x_i)\right]^2 M(x_i) \tag{18}$$

where $\hat{\beta}_{1,LL}(x_i)$ and $\hat{\beta}_{1,LCB}(x_i)$ is the local linear and local cubic estimator of $\beta_1(x) = \partial q_\tau(x)/\partial x_1$. For $\hat{\beta}_{1,LL}(x_i)$, it can be obtained from

$$\min_{a,b}\sum_{i=1}^{n}\rho_\tau\left[y_i - a - b(x)'(x_i - x)\right]\mathbf{K}_{h,ix}$$

where b estimates $(\partial q_\tau(x)/\partial x_1, \ldots, \partial q_\tau(x)/\partial x_p)'$, the $p\times 1$ vector of first order derivative functions, $\mathbf{K}_{h,ix} = \prod_{s=1}^{p} h_s^{-1}K((x_{is}-x_s)/h_s)$ is the product kernel function. Before defining $\hat{\beta}_{1,LCB}(x_i)$, we introduce some additional notations that are in line with Masry (1996). Assume that the multivariate quantile function $q_\tau(x)$ has derivatives of total order 4 at the point x, and we can approximate $q_\tau(x)$ locally by a multivariate polynomial of total order 3,

$$q_\tau(x) = \sum_{0\leq|\mathbf{k}|\leq 3}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}q_\tau(x_0)(x-x_0)^{\mathbf{k}} \tag{19}$$

where $\mathbf{k} = \{k_1, \ldots, k_p\}$, $|\mathbf{k}| = \sum_{\ell=1}^{p} k_\ell$, $\mathbf{k}! = k_1! \times \cdots \times k_p!$, $x^{\mathbf{k}} = x_1^{k_1} \times \cdots \times x_p^{k_p}$,

$$\sum_{0\leq|\mathbf{k}|\leq 3} = \sum_{j=0}^{3}\sum_{k_1=0}^{j}\cdots\sum_{\substack{k_p=0\\ k_1+\cdots+k_p=j}}^{j}, \quad \text{and} \quad D^{\mathbf{k}}q_\tau(x) = \frac{\partial^{\mathbf{k}}q_\tau(x)}{\partial x_1^{k_1}\cdots\partial x_p^{k_p}}.$$

Given the notations above, the local cubic conditional quantile estimation solves the following minimization problem

$$\min_{b_0,\ldots,b_3}\sum_{i=1}^{n}\rho_\tau\left[y_i - \sum_{0\leq|\mathbf{k}|\leq 3} b_{\mathbf{k}}(x)'(x_i - x)^{\mathbf{k}}\right]\mathbf{K}_{h,ix},$$

where $\mathbf{k}! \cdot b_{\mathbf{k}}(x)$ estimates $D^{\mathbf{k}}q_\tau(x)$. Therefore, we use $\hat{\beta}_{1,LCB}(x_i)$ to denote the first component of the solution $b_{\mathbf{1}}$.

To simplify the notations, we assume that all the elements in the bandwidth vector are equal, i.e., $h_1 = h_2 = \cdots = h_p = h$. Let $h_{0,opt}$ and $h_{0,cubic}$ denote the values of the leading term in h that minimizes (17) and (18), and they have the following relationship with generalizes the univariate case in (16),

$$h_{0,opt}/h_{0,cubic} = (V_1/V_{1,3})^{1/(p+1)}. \tag{20}$$

Therefore, one first minimizes (18) to obtain the bandwidth $\hat{h}_{cubic}$, then rescale it by the ratio (20). Define $\hat{h}_{0,opt} = (V_1/V_{1,3})^{1/(p+1)}\hat{h}_{cubic}$ as the final selected bandwidth, which is asymptotically equivalent to the optimal bandwidth selected from the infeasible cross validation as if the true quantile partial derivative function were known, i.e., $\hat{h}_{0,opt}/h_{0,opt} = 1 + o_P(1)$. In practice, the ratio $(V_1/V_{1,3})^{1/(p+1)}$ is even closer to one than the univariate case as p increases, therefore no adjustment is needed if one uses the Gaussian kernel function.

3. Simulation

In this section, we numerically study the finite sample performance of local cubic based CV method. We consider the following two data generating processes (DGP),

$$\text{DGP1:} \quad y_i = 2 + \sin(x_i) + x_i u_i, \quad i = 1, \ldots, n,$$
$$\text{DGP2:} \quad y_i = 1 + \ln(1 + x_i) + [1 + \sin(x_i)]u_i, \quad i = 1, \ldots, n,$$

where $x_i \sim i.i.d.\text{Uniform}[0, \pi]$ and $u_i \sim i.i.d.N(0, 0.64)$. The number of replication is 500 and the sample sizes are 100 and 200. For each simulation, we calculate the mean squared errors and median squared errors as follows,

$$\text{Mean Squared Error} = \frac{1}{n}\sum_{i=1}^{n}[\hat{\beta}_{LL}(x_i) - \beta(x_i)]^2,$$

$$\text{Median Squared Error} = \underset{1\leq i\leq n}{\text{median}}\left\{[\hat{\beta}_{LL}(x_i) - \beta(x_i)]^2\right\}.$$

The reason why we consider median squared error is that there may be some outliers producing huge squared errors and skewing the mean, while median is robust to outliers.

For the purpose of comparison, we consider three different ways of selecting the bandwidth h:

(i) infeasible method: h is selected by minimizing $n^{-1}\sum_{i=1}^{n}[\hat{\beta}_{LL}(x_i) - \beta(x_i)]^2$. This method uses the true $\beta(x_i)$ in the objective function which is infeasible in practice. It serves as an optimal benchmark method.
(ii) Our proposed LL–Local Cubic based method which selects h by minimizing $n^{-1}\sum_{i=1}^{n}[\hat{\beta}_{LL}(x_i) - \beta_{LCB}(x_i)]^2$.
(iii) Least squares method which selects bandwidth h by minimizing

$$\sum_{i=1}^{n}\rho_\tau(y_i - \hat{\alpha}_{-i,LL}(x_i))$$

where $\hat{\alpha}_{-i,LL}$ is the leave-one-out estimator of $q_\tau(x_i)$.

We expect that the infeasible method to give the best estimation result, followed by our LL–Local Cubic based method because our method is asymptotically efficient in selecting h. The least

Table 1
Simulation results for DGP1.

CV method	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	Mean	Mean	Median	Median
$\tau = 0.25$				
Infeasible	0.167	0.103	0.059	0.037
Local Cubic	0.173	0.121	0.064	0.044
Least Squares	0.267	0.156	0.098	0.049
$\tau = 0.50$				
Infeasible	0.140	0.098	0.057	0.035
Local Cubic	0.145	0.111	0.060	0.038
Least Squares	0.190	0.162	0.071	0.048
$\tau = 0.75$				
Infeasible	0.171	0.100	0.068	0.038
Local Cubic	0.182	0.107	0.071	0.042
Least Squares	0.241	0.128	0.088	0.053

Table 2
Simulation results for DGP2.

CV method	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	Mean	Mean	Median	Median
$\tau = 0.25$				
Infeasible	0.046	0.024	0.030	0.015
Local Cubic	0.047	0.027	0.031	0.017
Least Squares	0.059	0.036	0.036	0.022
$\tau = 0.50$				
Infeasible	0.039	0.028	0.023	0.016
Local Cubic	0.041	0.033	0.025	0.018
Least Squares	0.054	0.039	0.029	0.021
$\tau = 0.75$				
Infeasible	0.108	0.067	0.066	0.036
Local Cubic	0.117	0.081	0.071	0.044
Least Squares	0.137	0.081	0.083	0.047

squares method selects an h that is asymptotically efficient for estimating the conditional quantile function, but it is not optimal for estimating the derivative function $\beta(x_i)$.

Tables 1 and 2 report the results for DGP1 and DGP2 respectively with the Gaussian kernel. For the Gaussian kernel, the ratio $(V_1/V_{1,3})^{1/7} = 1.009$ is very close to 1, which means it is not necessary for us to adjust the bandwidth selected by local cubic based CV method. Row 1–3, 4–6 and 7–9 are for $\tau = 0.25$, $\tau = 0.50$ and $\tau = 0.75$ respectively. For each quantile, we report the mean and median of squared errors of the three CV methods and the two sample sizes.

We can see that the median squared errors are much less than the mean squared errors, which confirms our concern of the outlier problem. However, our conclusion remains consistent no matter in consideration of mean squared errors or median squared errors. As a benchmark, infeasible CV method takes advantage of the true value of β and always gives the smallest MSEs, as expected. Without the information of true β, local cubic based CV method gives roughly 5% more of MSEs than infeasible CV method does. Meanwhile, LS-CV performs the worst. However, note that the comparisons here is for the derivatives of conditional quantile function, and the usual LS-CV method should outperform our proposed method since the optimal h for derivative estimation is not optimal for conditional quantile function estimation.

4. Empirical application

In order to complement the simulation results on finite sample performance, we give a simple illustrative example in univariate case, which compares the results with real data between traditional least squares cross validation (LSCV) and our proposed gradient based cross validation (GBCV) methods. For this purpose, we use a subset of hedonic housing price data from Anglin and Gençay (1996), in which the houses have less than 15,000 squared feet lots.

Table 3
Descriptive statistics.

	price[a]		*lotsize*[b]
Minimum	0.250		0.165
1st Quartile	0.490		0.360
Median	0.620		0.458
3rd Quartile	0.820		0.636
Maximum	1.900		1.320
Mean	0.680		0.511
Std. Dev.	0.2654		0.207
Correlation		0.532	
Skewness	1.208		1.059
Kurtosis	4.974		4.289
Sample size: 544			

[a] In 100,000 Canadian Dollars.
[b] In 10,000 squared feet.

Parmeter et al. (2007) studied the same data set in a nonparametric framework. Table 3 shows the descriptive statistics for the two variables in the data set, and it is observed that the two variables, *price* and *lotsize*, have a strong positive correlation.

In this application, we focus on the relationship between the housing price (*price*, measured in 100,000 Canadian Dollars) and the size of the lot the house resides on (*lotsize*, in 10,000 squared feet). Economic theory suggests that the housing price should increases as lot size increases. This monotonically increasing relationship implies that the gradient function should be positive everywhere. Therefore, in the local linear nonparametric estimation of the τth conditional quantile function of the price $y = price$, given $x = lotsize$,

$$y = q_\tau(x) + \varepsilon \quad \text{with } \tau \in (0, 1),$$

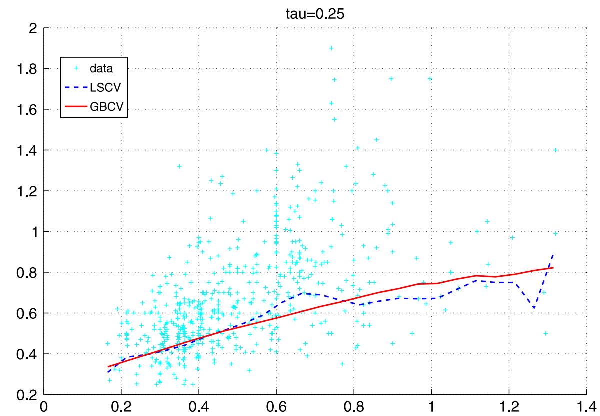
we expect that the derivative of the conditional quantile function $\beta(x) \equiv dq_\tau(x)/dx > 0$ for all x in the support of *lotsize* and $0 < \tau < 1$. Therefore, its local linear estimate $\hat{\beta}_{LL}(x)$, obtained from

$$(\hat{q}_\tau(x), \hat{\beta}_{LL}(x)) \equiv \underset{a,b}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - a - b(x_i - x))K_{h,ix} \quad (21)$$

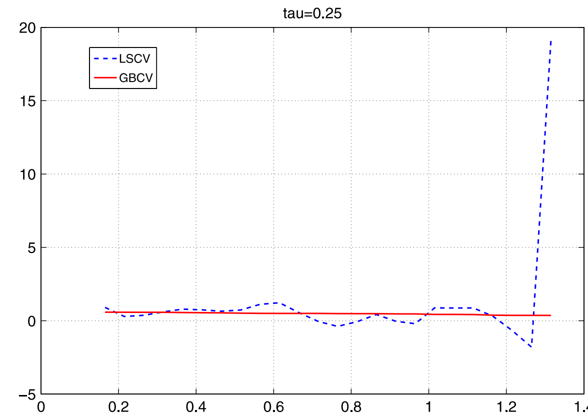
is also expected to be greater than zero, i.e., $\hat{\beta}_{LL}(x) > 0$, for all x and τ.

To compare the performance of LSCV and GBCV methods, we estimate the local linear conditional quantile model given in (21) at five different quantiles ($\tau = 0.1, 0.25, 0.5, 0.75$, and 0.9), select smoothing parameters h_{LS} and h_{GB} by LSCV and GBCV methods respectively,[3] and plot the estimated conditional quantile functions and their corresponding derivatives. Fig. 1 shows the results for the three quartiles, i.e., $\tau = 0.25, 0.5$, and 0.75; and Fig. 2 for two extreme quantiles, $\tau = 0.1$ and 0.9. First, on the left panel of the two figures, we observe that the estimated conditional quantile functions using GBCV method are much smoother than those using LSCV method. This is reasonable as the gradient based cross validation tends to deliver larger optimal smoothing parameter h_{GB} than least squares cross validation does, and larger smoothing parameter h_{GB} means smoother conditional quantile estimates. Second, on the right panel of the two figures, we observe that the estimated conditional quantile derivatives using GBCV method are also much smoother than those derivative estimates from LSCV method. This is especially the case when the quantiles are $\tau = 0.25$ and 0.5, as LSCV method delivers wiggly derivative estimates. Third, besides the smoothness, the derivative estimates from GBCV method

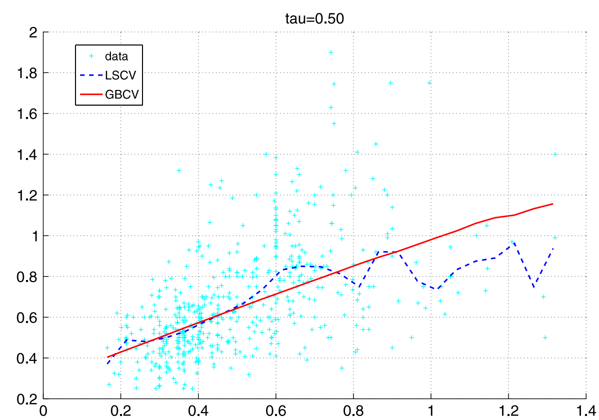
[3] Both cross validation methods select the smoothing parameter h for each quantile by univariate grid search on the range [0.02, 3]. In the selection range, the increment is 0.02 from 0.02 to 0.1, and 0.1 from 0.1 to 3. The values of selected smoothing parameters are reported in Figs. 1 and 2.



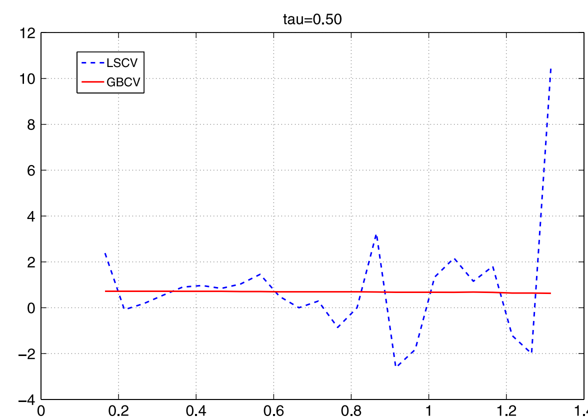
(a) Quantile functions ($\tau = 0.25, h_{LS} = 0.06, h_{GB} = 0.5$).



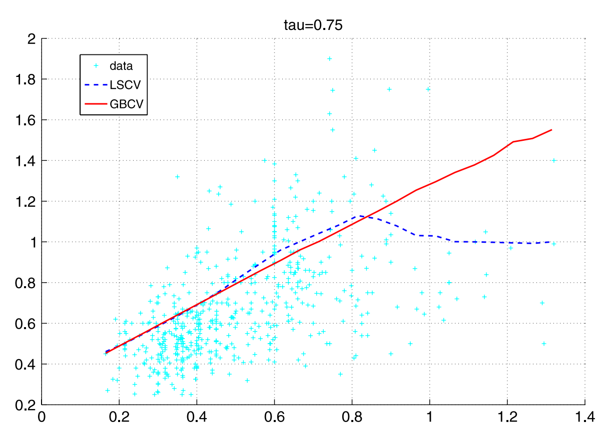
(b) Quantile derivatives ($\tau = 0.25$).



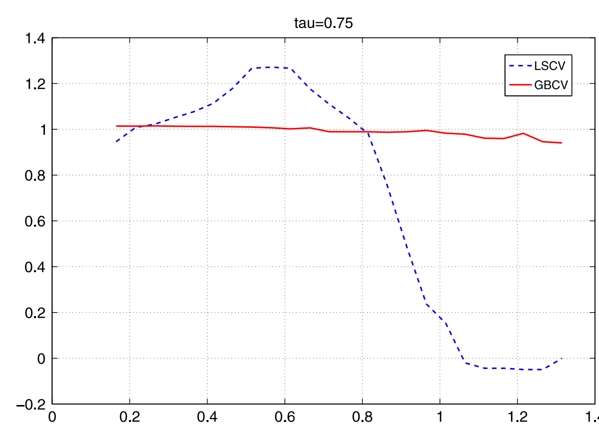
(c) Quantile functions ($\tau = 0.50, h_{LS} = 0.04, h_{GB} = 1.1$).



(d) Quantile derivatives ($\tau = 0.50$).



(e) Quantile functions ($\tau = 0.75, h_{LS} = 0.22, h_{GB} = 2.0$).

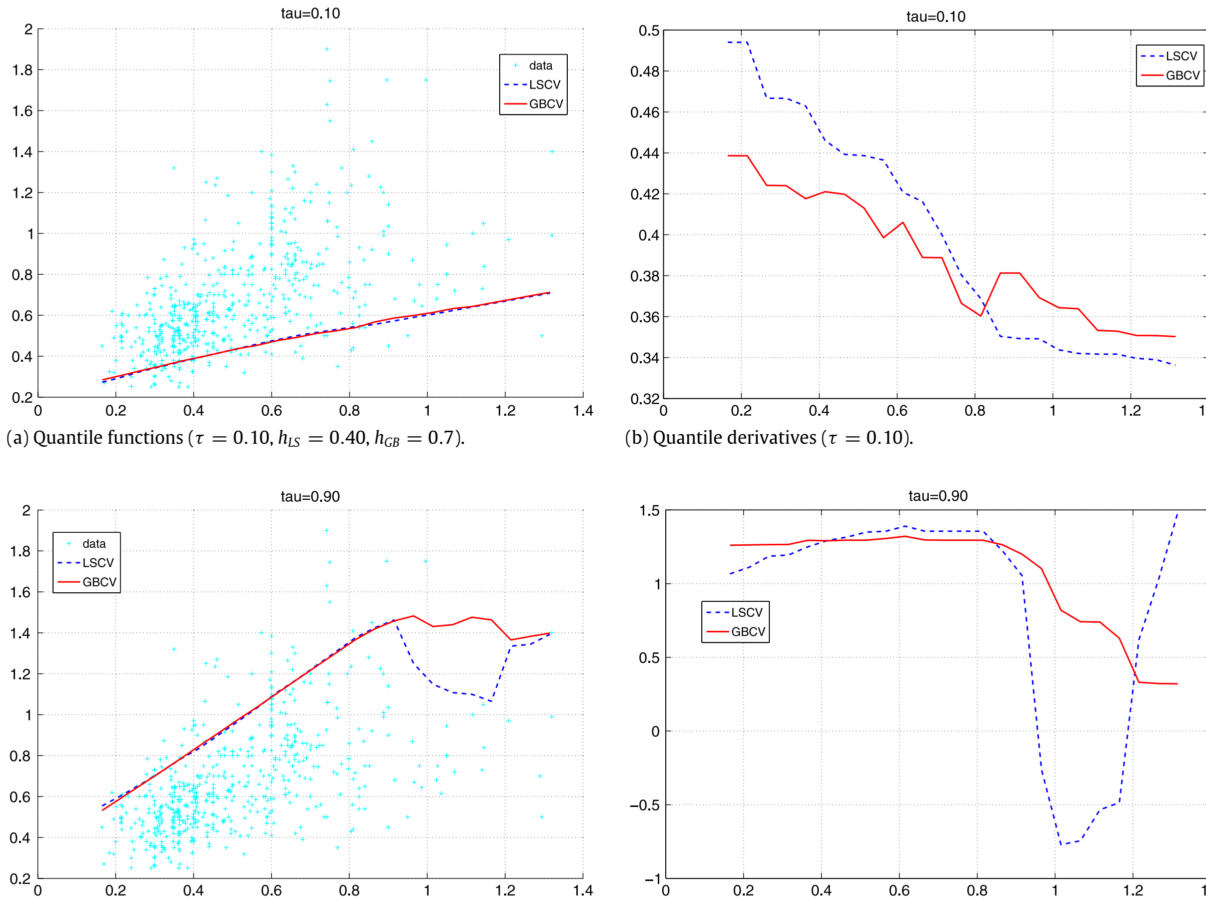


(f) Quantile derivatives ($\tau = 0.75$).

Fig. 1. Estimates of quantile functions and their derivatives on different quantiles with $\tau = 0.75$, 0.5, and 0.25. The cyan plus signs denote the actual data. h_{LS} and h_{GB} are selected smoothing parameters by the least squares cross validation (LSCV) and gradient based cross validation (GBCV) methods respectively. The broken blue curves denote the estimated conditional quantile functions using h_{LS} on the left panel and their corresponding derivatives on the right panel. The solid red curves denote those estimates using h_{GB}.

are all positive on the support of regression variable $x = lotsize$, which is consistent with the monotonically increasing relationship between the price and lot size in economic theory; while the derivative estimates from LSCV method at all quantiles except $\tau = 0.1$ have one or multiple negative value regions. At quantile level $\tau = 0.25$ and $\tau = 0.5$, the derivative estimates are negative in three intervals shown in Fig. 1(b) and 1(d) respectively; and there is one negative derivative region at each quantile level $\tau = 0.75$ and $\tau = 0.9$ shown in Fig. 1(f) and 2(d). The ranges of negative derivative regions at each quantile are summarized in Table 4. Only at low quantile $\tau = 0.1$, both LSCV and GBCV methods produce similar estimates of conditional quantile function and their derivative estimates are nonnegative.

This simple illustrative example implies that, in local linear conditional quantile regression, traditional least squares cross validation method tends to select a smoothing parameter h smaller than the optimal one for conditional quantile derivatives estimator, and hence produces unreliable estimation results because of undersmoothing with this too small and suboptimal h. Therefore, if the marginal effects of regression variables or the derivatives of condi-



(a) Quantile functions ($\tau = 0.10$, $h_{LS} = 0.40$, $h_{GB} = 0.7$).

(b) Quantile derivatives ($\tau = 0.10$).

(c) Quantile functions ($\tau = 0.90$, $h_{LS} = 0.14$, $h_{GB} = 0.3$).

(d) Quantile derivatives ($\tau = 0.90$).

Fig. 2. Estimates of quantile functions and their derivatives on different quantiles with $\tau = 0.1$ and 0.9. The cyan plus signs denote the actual data. h_{LS} and h_{GB} are selected smoothing parameters by the least squares cross validation (LSCV) and gradient based cross validation (GBCV) methods respectively. The broken blue curves denote the estimated conditional quantile functions using h_{LS} on the left panel and their corresponding derivatives on the right panel. The solid red curves denote those estimates using h_{GB}.

Table 4
Negative derivatives regions for two methods at different quantiles.

	LSCV method[a]	GBCV method
$\tau = 0.1$	None	None
$\tau = 0.25$	[0.7, 0.82], [0.9, 1], [1.18, 1.3]	None
$\tau = 0.5$	[0.6, 0.82], [0.9, 1], [1.18, 1.3]	None
$\tau = 0.75$	[1.05, 1.3]	None
$\tau = 0.9$	[0.95, 1.18]	None

[a] Negative value regions are identified approximately by visual inspection of Figs. 1 and 2.

tional quantiles are of interest, the gradient based cross validation should be preferred to the least squares cross validation for the selection of optimal smoothing parameter.

5. Conclusion

In this paper, we propose a fully data-driven optimal bandwidth selection procedure for local linear conditional quantile regression using check function. Different from traditional least squares cross validation method which selects optimal bandwidth for conditional quantiles, we propose gradient based cross validation method which selects optimal bandwidth for the derivatives of conditional quantiles. The estimation of marginal effect of regression variables at different quantiles, or derivative of quantile functions, is of special interest as it quantifies the amount of change of dependent variable at different quantiles in response to a unit change in regression variables. Theoretically, our procedure is shown to deliver bandwidths with the optimal rate for derivative estimators. Both simulation and empirical application results substantiate the superiority of our gradient based procedure in the estimation of quantile derivatives over standard least squares cross validation method which produces undersmoothed derivative estimates. Finally, our proposed method can be further extended in many different directions: such as selecting optimal smoothing parameter for estimating higher order derivatives, applying to the models with weakly dependent data and various semiparametric models etc.

Appendix

A.1. Leading bias and variance terms of local linear quantile derivative estimator

In this subsection, we derive the leading bias and variance of local linear quantile estimator presented in Eqs. (4) and (5).

The local linear estimator of $(q_\tau(x), q'_\tau(x))^T$ in quantile regression with check loss function can be obtained by solving the minimization problem in (1). To facilitate derivation, we define $b_h = bh$ so that

$$\begin{pmatrix} \hat{a}(x) \\ \hat{b}_h(x) \end{pmatrix} \equiv \operatorname*{argmin}_{a, b_h} S(a, b_h)$$

$$\equiv \operatorname*{argmin}_{a, b_h} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(y_i - a - \left(\frac{x_i - x}{h} \right) b_h \right) K_{h,ix}, \quad (22)$$

and the local linear quantile derivative estimator is

$\hat{\beta}_{LL}(x) = \hat{b}_h(x)/h.$

By Taylor expansion of

$$y_i = q_\tau(x_i) + e_i = q_\tau(x) + hq'_\tau(x)\left(\frac{x_i - x}{h}\right) + \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 + R_m(x, x_i) + e_i$$

and add-and-subtracting $n^{-1}\sum_{i=1}^n \rho_\tau(e_i)K_{h,ix}$ we can rewrite $S(a, b)$ as

$$\begin{aligned} S(a,b) &= \frac{1}{n}\sum_{i=1}^n \rho_\tau(e_i)K_{h,ix} \\ &\quad + \frac{1}{n}\sum_{i=1}^n \Bigg\{\rho_\tau\Bigg[q_\tau(x) + hq'_\tau(x)\left(\frac{x_i - x}{h}\right) \\ &\quad + \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 + R_m(x, X_i) + e_i \\ &\quad - a - \left(\frac{x_i - x}{h}\right)b_h\Bigg]K_{h,ix} - \rho_\tau(e_i)K_{h,ix}\Bigg\} \\ &\equiv S_1 + S_2(a,b), \end{aligned}$$

where only $S_2(a, b)$ depends on a and b. Also, note that following equality holds,

$$\rho_\tau(x - y) - \rho_\tau(x) = y[\mathbf{1}(x \le 0) - \tau] + \int_0^y [\mathbf{1}(x \le z) - \mathbf{1}(x \le 0)]dz,$$

and therefore, we obtain the equation given in Box I.

Substitute it into $S_2(a, b)$, we have the equation given in Box II.

In addition, we require the bias of the conditional quantile estimator converges to zero in probability, i.e., $a - q_\tau(x) + (b_h - hq'_\tau(x))\left(\frac{x_i - x}{h}\right) - \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 - R_m = o_p(1)$, so that

$$\begin{aligned} &E[S_{2,2}(a,b)|\mathcal{X}] \\ &= \frac{1}{2n}\sum_{i=1}^n K_{h,ix}f_e(0|x_i)\Bigg[a - q_\tau(x) + (b_h - hq'_\tau(x))\left(\frac{x_i - x}{h}\right) \\ &\quad - \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 - R_m\Bigg]^2 + (s.o.) \end{aligned}$$

and

$$\begin{aligned} &S_{2,1}(a,b) + S_{2,2}(a,b) = S_{2,1}(a,b) + E[S_{2,2}(a,b)|\mathcal{X}] + (s.o.) \\ &= \frac{1}{n}\sum_{i=1}^n \Bigg[a - q_\tau(x) + (b_h - hq'_\tau(x))\left(\frac{x_i - x}{h}\right) \\ &\quad - \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 - R_m\Bigg][\mathbf{1}(e_i \le 0) - \tau]K_{h,ix} \\ &\quad + \frac{1}{2n}\sum_{i=1}^n K_{h,ix}f_e(0|x_i)\Bigg[a - q_\tau(x) + (b_h - hq'_\tau(x))\left(\frac{x_i - x}{h}\right) \\ &\quad - \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 - R_m\Bigg]^2 + (s.o.). \end{aligned} \tag{23}$$

Minimizing $S(a, b)$ in (22) is equivalent to minimizing $S_{2,1}(a, b) + E[S_{2,2}(a, b)|\mathcal{X}]$ in (23). To simplify notation, define $\delta_0(x) \equiv a - q(x)$ and $\delta_1(x) \equiv b_h - hq'_\tau(x)$. The first order conditions for a and b are

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n [1_{[e_i\le 0]} - \tau]K_{h,ix} + \frac{1}{n}\sum_{i=1}^n K_{h,ix}f_e(0|x_i) \\ &\quad \times \left[\delta_0(x) + \delta_1(x)\left(\frac{x_i - x}{h}\right) - \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 - R_m\right] = 0 \\ &\frac{1}{n}\sum_{i=1}^n [1_{[e_i\le 0]} - \tau]K_{h,ix}\left(\frac{x_i - x}{h}\right) + \frac{1}{n}\sum_{i=1}^n K_{h,ix}f_e(0|x_i) \\ &\quad \times \left[\delta_0(x) + \delta_1(x)\left(\frac{x_i - x}{h}\right) - \frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 - R_m\right] \\ &\quad \times \left(\frac{x_i - x}{h}\right) = 0. \end{aligned}$$

In matrix form, we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n K_{h,ix}f_e(0|x_i)\begin{pmatrix} 1 & \dfrac{x_i - x}{h} \\ \dfrac{x_i - x}{h} & \left(\dfrac{x_i - x}{h}\right)^2 \end{pmatrix}\begin{pmatrix}\delta_0(x) \\ \delta_1(x)\end{pmatrix} \\ &= \frac{1}{n}\sum_{i=1}^n K_{h,ix}f_e(0|x_i)\begin{pmatrix} 1 \\ \dfrac{x_i - x}{h}\end{pmatrix}\frac{h^2 q''_\tau(x)}{2}\left(\frac{x_i - x}{h}\right)^2 \\ &\quad - \frac{1}{n}\sum_{i=1}^n [\mathbf{1}(e_i \le 0) - \tau]K_{h,ix}\begin{pmatrix} 1 \\ \dfrac{x_i - x}{h}\end{pmatrix} + (s.o.) \end{aligned}$$

where $(s.o.)$ are smaller order terms that do not contribute to the leading bias or variance. Under the assumption of independent and identical distribution, the summations in (24) converges to their expected values respectively, and the FOCs can be written as

$$\mathbf{A} \cdot \delta(x) = \mathbf{B} - \mathbf{C}$$

where

$$\begin{aligned} \mathbf{A} &= \mathbf{A}_0 + \mathbf{A}_h = \begin{pmatrix} \xi(x) & 0 \\ 0 & \mu_2\xi(x)\end{pmatrix} + \begin{pmatrix} 0 & h\mu_2\xi'(x) \\ h\mu_2\xi'(x) & 0\end{pmatrix} \\ &= \begin{pmatrix} \xi(x) & h\mu_2\xi'(x) \\ h\mu_2\xi'(x) & \mu_2\xi(x)\end{pmatrix}, \end{aligned}$$

$$\delta(x) = \begin{pmatrix}\delta_0(x) \\ \delta_1(x)\end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} h^2\mu_2\xi(x)q''_\tau(x)/2 \\ h^3\mu_4\xi'(x)q''_\tau(x)/2\end{pmatrix},$$

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^n [\mathbf{1}(e_i \le 0) - \tau]K_{h,ix}\begin{pmatrix} 1 \\ \dfrac{x_i - x}{h}\end{pmatrix}$$

where $\mu_j = \int K(v)v^j dv$ and $\xi(x) = f_e(0|x)f(x)$. Note that the expectations of those elements in $\mathbf{C}$ are zero. Therefore, when calculating the leading bias of the local linear quantile derivative estimator $Bias^0(\hat{\beta}_{LL}(x))$, we have

$$\begin{aligned} Bias^0(\hat{\beta}_{LL}(x)) &= E[\hat{\beta}_{LL}(x) - q'_\tau(x)] = h^{-1}E[\hat{b}_h(x) - hq'_\tau(x)] \\ &= h^{-1}E[\delta_1(x)] \\ &= h^{-1}\begin{pmatrix}0 & 1\end{pmatrix}\mathbf{A}^{-1}\mathbf{B} \\ &= h^2\frac{(\mu_4 - \mu_2^2)\xi'(x)q''(x)}{2\mu_2\xi(x)} \equiv h^2 B_1(x). \end{aligned}$$

Note that the $O(h)$ terms in $\mathbf{A}$ (i.e., $\mathbf{A}_h$) cannot be omitted in order to obtain the correct expression for $Bias^0(\hat{\beta}_{LL}(x))$.

The leading variance of the derivative estimator $Var^0(\hat{\beta}_{LL}(x))$ is

$$Var^0(\hat{\beta}_{LL}(x)) = Var^0\left(\begin{pmatrix}0 & 1\end{pmatrix}D_h\delta(x)\right) = h^{-2}Var^0\left(\begin{pmatrix}0 & 1\end{pmatrix}\mathbf{A}_0^{-1}\mathbf{C}\right)$$

$$\rho_\tau\left\{e_i-\left[a-q_\tau(x)+(b_h-hq'_\tau(x))\left(\frac{x_i-x}{h}\right)-\frac{h^2q''_\tau(x)}{2}\left(\frac{x_i-x}{h}\right)^2-R_m\right]\right\}-\rho_\tau(e_i)$$
$$=\left[a-q_\tau(x)+(b_h-hq'_\tau(x))\left(\frac{x_i-x}{h}\right)-\frac{h^2q''_\tau(x)}{2}\left(\frac{x_i-x}{h}\right)^2-R_m\right][\mathbf{1}(e_i\le 0)-\tau]$$
$$+\int_0^{a-q_\tau(x)+(b_h-hq'_\tau(x))\left(\frac{x_i-x}{h}\right)-\frac{h^2q''_\tau(x)}{2}\left(\frac{x_i-x}{h}\right)^2-R_m}[\mathbf{1}(e_i\le z)-\mathbf{1}(e_i\le 0)]dz$$

Box I.

$$S_2(a,b)=\frac{1}{n}\sum_{i=1}^n\left[a-q_\tau(x)+(b_h-hq'_\tau(x))\left(\frac{x_i-x}{h}\right)-\frac{h^2q''_\tau(x)}{2}\left(\frac{x_i-x}{h}\right)^2-R_m\right][\mathbf{1}(e_i\le 0)-\tau]K_{h,ix}$$
$$+\frac{1}{n}\sum_{i=1}^nK_{h,ix}\int_0^{a-q_\tau(x)+(b_h-hq'_\tau(x))\left(\frac{x_i-x}{h}\right)-\frac{h^2q''_\tau(x)}{2}\left(\frac{x_i-x}{h}\right)^2-R_m}[\mathbf{1}(e_i\le z)-\mathbf{1}(e_i\le 0)]dz$$
$$\equiv S_{2,1}(a,b)+S_{2,2}(a,b)$$

Box II.

$$=h^{-2}E\left[\frac{1}{\mu_2\xi(x)}\frac{1}{n}\sum_{i=1}^n[\mathbf{1}(e_i\le 0)-\tau]K_{h,ix}\left(\frac{x_i-x}{h}\right)\right]^2$$
$$=\frac{1}{nh^2}E\left[\frac{1}{\mu_2^2\xi^2(x)}[\mathbf{1}(e_i\le 0)-\tau]^2K_{h,ix}^2\left(\frac{X_i-x}{h}\right)^2\right]$$
$$=\frac{1}{nh^3}\frac{\tau(1-\tau)\nu_2}{\mu_2^2}\frac{f(x)}{\xi^2(x)}\equiv\frac{1}{nh^3}V_1\Omega(x)$$

where

$$D_h=\begin{pmatrix}1&0\\0&h^{-1}\end{pmatrix},\qquad \nu_j=\int K^2(v)v^jdv,$$
$$\Omega(x)=\frac{f(x)}{\xi^2(x)},\qquad V_1=\frac{\tau(1-\tau)\nu_2}{\mu_2^2}.$$

Note that the $O(h)$ terms in $\mathbf{A}$ do not enter the leading variance of the derivative estimator, and they can be omitted safely (i.e., regarded as zeros) to simplify the calculation.

A.2. Leading variance term of local cubic quantile derivative estimator

In this subsection, we show the proofs for (11) and (12), and (13).

The local cubic estimator of $(q_\tau(x), q'_\tau(x), q''_\tau(x), q'''_\tau(x))$ in quantile regression with check loss function solves the following minimization problem

$$\hat\gamma(x)\equiv\begin{pmatrix}\hat a(x)\\ \hat b_h(x)\end{pmatrix}=\operatorname*{argmin}_{a,b_h}S(a,b_h)$$
$$\equiv\operatorname*{argmin}_{a,b_h}\frac{1}{n}\sum_{i=1}^n\rho_\tau\left(y_i-a-\left(\frac{x_i-x}{h}\right)b_{1h}-\left(\frac{x_i-x}{h}\right)^2b_{2h}-\left(\frac{x_i-x}{h}\right)^3b_{3h}\right)K_{h,ix}\tag{24}$$

where the estimate of first, second, and third order derivatives are defined as

$$\hat\beta_{1,LCB}(x)=\hat b_{1h}(x)/h,\qquad \hat\beta_{2,LCB}(x)=\hat b_{2h}(x)/h^2,$$
$$\hat\beta_{3,LCB}(x)=\hat b_{3h}(x)/h^3.$$

Using Knight's (1998) identity and fourth order Taylor expansion of $q(x_i)$ around x, we can reformulate the objective function (24) in the minimization problem as $S_{2,1}(a,b)+E(S_{2,2}(a,b)|\mathcal{X})$

$$S_{2,1}(a,b)+E(S_{2,2}(a,b)|\mathcal{X})=\frac{1}{n}\sum_{i=1}^n\Bigg[\delta_0(x)+\delta_1(x)\left(\frac{x_i-x}{h}\right)+\delta_2(x)\left(\frac{x_i-x}{h}\right)^2+\delta_3(x)\left(\frac{x_i-x}{h}\right)^3-\frac{h^4q_\tau^{(4)}(x)}{24}\left(\frac{x_i-x}{h}\right)^4-R_m(x,x_i)\Bigg][\mathbf{1}(e_i\le 0)-\tau]K_{h,ix}$$
$$+\frac{1}{2n}\sum_{i=1}^nK_{h,ix}f_e(0|x_i)\Bigg[\delta_0(x)+\delta_1(x)\left(\frac{x_i-x}{h}\right)+\delta_2(x)\left(\frac{x_i-x}{h}\right)^2+\delta_3(x)\left(\frac{x_i-x}{h}\right)^3-\frac{h^4q_\tau^{(4)}(x)}{24}\left(\frac{x_i-x}{h}\right)^4-R_m(x,x_i)\Bigg]^2$$

where

$$\delta_c(x)=\left(\delta_{0c}(x)\delta_{1c}(x)\quad\delta_{2c}(x)\quad\delta_{3c}(x)\right)'$$
$$=\left(a-q_\tau(x)\quad b_{1h}-hq'_\tau(x)\quad b_{2h}-h^2q''_\tau(x)/2\quad b_{3h}-h^3q'''_\tau(x)/6\right)'$$

and $R_m(x,x_i)$ are higher order remainder terms in Taylor expansion. Minimizing $S(a,b)$ in (24) is asymptotically equivalent to minimizing $S_{2,1}(a,b)+E(S_{2,2}(a,b)|\mathcal{X})$. Therefore, similar to the local linear quantile estimator, the FOCs for (a,b) can be written is the matrix form

$$\frac{1}{n}\sum_{i=1}^nK_{h,ix}f_e(0|x_i)$$
$$\times\begin{pmatrix}1&\frac{x_i-x}{h}&\left(\frac{x_i-x}{h}\right)^2&\left(\frac{x_i-x}{h}\right)^3\\ \frac{x_i-x}{h}&\left(\frac{x_i-x}{h}\right)^2&\left(\frac{x_i-x}{h}\right)^3&\left(\frac{x_i-x}{h}\right)^4\\ \left(\frac{x_i-x}{h}\right)^2&\left(\frac{x_i-x}{h}\right)^3&\left(\frac{x_i-x}{h}\right)^4&\left(\frac{x_i-x}{h}\right)^5\\ \left(\frac{x_i-x}{h}\right)^3&\left(\frac{x_i-x}{h}\right)^4&\left(\frac{x_i-x}{h}\right)^5&\left(\frac{x_i-x}{h}\right)^6\end{pmatrix}$$

$$\times \begin{pmatrix} \delta_{0c}(x) \\ \delta_{1c}(x) \\ \delta_{2c}(x) \\ \delta_{3c}(x) \end{pmatrix}$$

$$= \frac{1}{n}\sum_{i=1}^{n} K_{h,ix} f_e(0|x_i) \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \\ \left(\frac{x_i - x}{h}\right)^2 \\ \left(\frac{x_i - x}{h}\right)^3 \end{pmatrix} \frac{h^4 q_\tau^{(4)}(x)}{24} \left(\frac{x_i - x}{h}\right)^4$$

$$- \frac{1}{n}\sum_{i=1}^{n} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \\ \left(\frac{x_i - x}{h}\right)^2 \\ \left(\frac{x_i - x}{h}\right)^3 \end{pmatrix}. \quad (25)$$

Under the assumption of independent and identical distribution, the summation terms in (25) converges in probability to their expectation values, and the FOC eqnarrays can be simplified as

$$\mathbf{A}_c \cdot \delta_c(x) = \mathbf{B}_c - \mathbf{C}_c$$

with

$$\mathbf{A}_c = \mathbf{A}_{c,0} + \mathbf{A}_{c,h} = \begin{pmatrix} \xi(x) & 0 & \mu_2\xi(x) & 0 \\ 0 & \mu_2\xi(x) & 0 & \mu_4\xi(x) \\ \mu_2\xi(x) & 0 & \mu_4\xi(x) & 0 \\ 0 & \mu_4\xi(x) & 0 & \mu_6\xi(x) \end{pmatrix}$$

$$+ \begin{pmatrix} 0 & h\mu_2\xi'(x) & 0 & h\mu_4\xi'(x) \\ h\mu_2\xi'(x) & 0 & h\mu_4\xi'(x) & 0 \\ 0 & h\mu_4\xi'(x) & 0 & h\mu_6\xi'(x) \\ h\mu_4\xi'(x) & 0 & h\mu_6\xi'(x) & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \xi(x) & h\mu_2\xi'(x) & \mu_2\xi(x) & h\mu_4\xi'(x) \\ h\mu_2\xi'(x) & \mu_2\xi(x) & h\mu_4\xi'(x) & \mu_4\xi(x) \\ \mu_2\xi(x) & h\mu_4\xi'(x) & \mu_4\xi(x) & h\mu_6\xi'(x) \\ h\mu_4\xi'(x) & \mu_4\xi(x) & h\mu_6\xi'(x) & \mu_6\xi(x) \end{pmatrix}$$

$$\delta_c(x) = \begin{pmatrix} \delta_0(x) \\ \delta_1(x) \\ \delta_2(x) \\ \delta_3(x) \end{pmatrix}, \qquad \mathbf{B}_c = \begin{pmatrix} h^4\mu_4\xi(x)q^{(4)}(x)/24 \\ h^5\mu_6\xi'(x)q^{(4)}(x)/24 \\ h^4\mu_6\xi(x)q^{(4)}(x)/24 \\ h^5\mu_8\xi'(x)q^{(4)}(x)/24 \end{pmatrix},$$

$$\mathbf{C}_c = \frac{1}{n}\sum_{i=1}^{n} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \\ \left(\frac{x_i - x}{h}\right)^2 \\ \left(\frac{x_i - x}{h}\right)^3 \end{pmatrix},$$

where the subscript c denotes local cubic estimation.

We are interested in the leading variance of local cubic quantile derivative estimator $\hat{\beta}_{LCB}(x)$ and leading covariance between $\hat{\beta}_{LCB}(x)$ and $\hat{\beta}_{LL}(x)$, i.e., $Var^0(\hat{beta}_{LCB}(x))$ and $Cov^0(\hat{\beta}_{LCB}(x), \hat{\beta}_{LL}(x))$. Define $D_{hc} = \text{diag}(1, h^{-1}, h^{-2}, h^{-3})$. The leading variance is,

$$Var^0(\hat{\beta}_{LCB}(x)) = Var^0\left(\begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix} D_{ch}\delta_c(x)\right)$$

$$= h^{-2}Var^0\left(\begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix} \mathbf{A}_{c,0}^{-1}\mathbf{C}_c\right)$$

$$= h^{-2}Var^0\left(\begin{pmatrix}0 & \frac{\mu_6}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} & 0 & \frac{-\mu_4}{(\mu_2\mu_6 - \mu_4^2)\xi(x)}\end{pmatrix} \mathbf{C}_c\right)$$

$$= h^{-2}Var^0\left[\frac{\mu_6}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} \frac{1}{n}\sum_{i=1}^{n} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \left(\frac{x_i - x}{h}\right)\right.$$

$$\left. - \frac{\mu_4}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} \frac{1}{n}\sum_{i=1}^{n} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \left(\frac{x_i - x}{h}\right)^3\right]$$

$$= \frac{1}{nh^2}E\left[\frac{\mu_6}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \left(\frac{X_i - x}{h}\right)\right]^2$$

$$+ \frac{1}{nh^2}E\left[\frac{\mu_4}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \left(\frac{X_i - x}{h}\right)^3\right]^2$$

$$- \frac{2}{nh^2}E\left[\frac{\mu_4\mu_6}{(\mu_2\mu_6 - \mu_4^2)^2\xi^2(x)} [\mathbf{1}(e_i \le 0) - \tau]^2 K_{h,ix}^2 \left(\frac{X_i - x}{h}\right)^4\right]$$

$$= \frac{1}{nh^3} \frac{\tau(1-\tau)(\mu_6^2\nu_2 + \mu_4^2\nu_6 - 2\mu_4\mu_6\nu_4)}{(\mu_2\mu_6 - \mu_4^2)^2} \frac{f(x)}{\xi^2(x)}$$

$$\equiv \frac{1}{nh^3} V_3 \Omega(x),$$

where the fourth equality holds because of the assumption of independent and identical distribution. The first expectation is

$$E\left[\frac{\mu_6}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} [\mathbf{1}(e_i \le 0) - \tau] K_{h,ix} \left(\frac{X_i - x}{h}\right)\right]^2$$

$$= \frac{\mu_6^2}{(\mu_2\mu_6 - \mu_4^2)^2\xi^2(x)} E\left[[\mathbf{1}(e_i \le 0) - \tau]^2 K_{h,ix}^2 \left(\frac{X_i - x}{h}\right)^2\right]$$

$$= \frac{\mu_6^2}{(\mu_2\mu_6 - \mu_4^2)^2\xi^2(x)} E$$

$$\times \left\{E\left[[\mathbf{1}(e_i \le 0) - \tau]^2 | X_i\right] K_{h,ix}^2 \left(\frac{X_i - x}{h}\right)^2\right\}$$

$$= \frac{\tau(1-\tau)\mu_6^2}{(\mu_2\mu_6 - \mu_4^2)^2\xi^2(x)} \int K_{h,ix}^2 \left(\frac{x_i - x}{h}\right)^2 f(x_i)dx_i$$

$$= \frac{\tau(1-\tau)\mu_6^2}{(\mu_2\mu_6 - \mu_4^2)^2\xi^2(x)} \frac{1}{h^2} \int K^2\left(\frac{x_i - x}{h}\right) \left(\frac{x_i - x}{h}\right)^2 f(x_i)dx_i$$

$$= \frac{1}{h} \frac{\tau(1-\tau)\mu_6^2}{(\mu_2\mu_6 - \mu_4^2)^2\xi^2(x)} \int K^2(v)v^2 f(vh + x)dv$$

$$= \frac{1}{h} \frac{\tau(1-\tau)\mu_6^2\nu_2}{(\mu_2\mu_6 - \mu_4^2)^2} \frac{f(x)}{\xi^2(x)},$$

and the other two expectations can be derived in the same way. Similarly, the leading covariance can be calculated as,

$$Cov^0(\hat{\beta}_{LL}(x), \hat{\beta}_{LCB}(x))$$

$$= h^{-2}Cov\left(\begin{pmatrix}0 & 1\end{pmatrix}\mathbf{A}^{-1}\mathbf{C}, \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}\mathbf{A}_{c,0}^{-1}\mathbf{C}_c\right)$$

$$= h^{-2}E\left\{\left[\frac{1}{\mu_2\xi(x)} \frac{1}{n}\sum_{i=1}^{n} [1_{[e_i \le 0]} - \tau] K_{h,ix} \left(\frac{x_i - x}{h}\right)\right]\right.$$

$$\times \left[\frac{\mu_6}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} \frac{1}{n}\sum_{i=1}^{n} [1_{[e_i \le 0]} - \tau] K_{h,ix} \left(\frac{x_i - x}{h}\right)\right.$$

$$\left.\left. - \frac{\mu_4}{(\mu_2\mu_6 - \mu_4^2)\xi(x)} \frac{1}{n}\sum_{i=1}^{n} [1_{[e_i \le 0]} - \tau] K_{h,ix} \left(\frac{x_i - x}{h}\right)^3\right]\right\}$$

$$= \frac{1}{nh^3} \frac{\tau(1-\tau)(\mu_6\nu_2 - \mu_4\nu_4)}{\mu_2(\mu_2\mu_6 - \mu_4^2)} \frac{f(x)}{\xi^2(x)}$$

$$\equiv \frac{1}{nh^3} V_2 \Omega(x)$$

where $\mathbf{A}_{c,0}^{-1}$ is given by the expression in Box III.

A.3. The ratio of bandwidth $h_{0,opt}$ to $h_{0,cubic}$

This subsection shows the proofs for (7) and (15).

$$\mathbf{A}_{c,0}^{-1} = \begin{pmatrix} \xi(x) & 0 & \mu_2\xi(x) & 0 \\ 0 & \mu_2\xi(x) & 0 & \mu_4\xi(x) \\ \mu_2\xi(x) & 0 & \mu_4\xi(x) & 0 \\ 0 & \mu_4\xi(x) & 0 & \mu_6\xi(x) \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{\mu_4}{(\mu_4-\mu_2^2)\xi(x)} & 0 & \frac{-\mu_2}{(\mu_4-\mu_2^2)\xi(x)} & 0 \\ 0 & \frac{\mu_6}{(\mu_2\mu_6-\mu_4^2)\xi(x)} & 0 & \frac{-\mu_4}{(\mu_2\mu_6-\mu_4^2)\xi(x)} \\ \frac{-\mu_2}{(\mu_4-\mu_2^2)\xi(x)} & 0 & \frac{1}{(\mu_4-\mu_2^2)\xi(x)} & 0 \\ 0 & \frac{-\mu_4}{(\mu_2\mu_6-\mu_4^2)\xi(x)} & 0 & \frac{\mu_2}{(\mu_2\mu_6-\mu_4^2)\xi(x)} \end{pmatrix}$$

Box III.

The bandwidth h chosen by the cross-validation minimizes the leading term of integrated mean square error,

$$\begin{aligned} CV_0(h) &= \int MSE_0\left[\hat{\beta}_{LL}(x) - q'_\tau(x)\right] M(x)f(x)dx \\ &= \int \left([Bias_0(\hat{\beta}_{LL}(x))]^2 + Var_0(\hat{\beta}_{LL}(x))\right) M(x)f(x)dx \\ &= \int \left(h^4 B_1^2(x) + \frac{V_1}{nh^3}\Omega(x)\right) M(x)f(x)dx \end{aligned} \tag{26}$$

where $M(x)$ is a weight function with bounded support that trims out data near the boundary of the support of x. Let $h_{0,opt}$ denote the values of h that minimizes $CV^0(h)$, and we have

$$h_{0,opt} = \left[\frac{3V_1 \int \Omega(x)M(x)f(x)dx}{4\int B_1^2(x)M(x)f(x)dx}\right]^{\frac{1}{7}} n^{-\frac{1}{7}}.$$

The value of $h_{0,opt}$ is unknown because of unknown functions in $\Omega(x)$ and $B_1^2(x)$.

Alternatively, we replace the unknown derivative function $q'_\tau(x)$ in $CV_0(h)$ with local cubic quantile derivative estimator $\hat{\beta}_{LCB}(x)$, a consistent estimate of $q'_\tau(x)$, we have

$$\begin{aligned} CV_{LCB,0}(h) &= \int MSE^0\left[\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x)\right] M(x)f(x)dx \\ &= \int \Big([Bias_0(\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x))]^2 \\ &\quad + Var_0(\hat{\beta}_{LL}(x) - \hat{\beta}_{LCB}(x))\Big) M(x)f(x)dx \\ &= \int \Big([Bias_0(\hat{\beta}_{LL}(x))]^2 + Var_0(\hat{\beta}_{LL}(x)) + Var_0(\hat{\beta}_{LCB}(x)) \\ &\quad - 2Cov_0(\hat{\beta}_{LL}(x), \hat{\beta}_{LCB}(x))\Big) M(x)f(x)dx \\ &= \int \left(h^4 B_1^2(x) + \frac{V_{1,3}}{nh^3}\Omega(x)\right) M(x)f(x)dx \end{aligned}$$

where

$$\begin{aligned} V_{1,3} &= V_1 + V_3 - 2V_2 \\ &= \frac{\tau(1-\tau)\nu_2}{\mu_2^2} + \frac{\tau(1-\tau)(\mu_6^2\nu_2 + \mu_4^2\nu_6 - 2\mu_4\mu_6\nu_4)}{(\mu_2\mu_6-\mu_4^2)^2} \\ &\quad -2\frac{\tau(1-\tau)(\mu_6\nu_2-\mu_4\nu_4)}{\mu_2(\mu_2\mu_6-\mu_4^2)} \\ &= \frac{\tau(1-\tau)\mu_4^2(\mu_2^2\nu_6 - 2\mu_2\mu_4\nu_4 + \mu_4^2\nu_2)}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2}. \end{aligned}$$

The optimal $h_{0,cubic}$ based on minimizing $CV_{LCB,0}(h)$ is

$$h_{0,cubic} = \left[\frac{3V_{1,3}\int \Omega(x)M(x)f(x)dx}{4\int B_1^2(x)M(x)f(x)dx}\right]^{\frac{1}{7}} n^{-\frac{1}{7}},$$

and the ratio $h_{0,opt}/h_{0,cubic}$ is a constant

$$\frac{h_{0,opt}}{h_{0,cubic}} = \left(\frac{V_1}{V_{1,3}}\right)^{\frac{1}{7}} = \left[\frac{\nu_2(\mu_2\mu_6-\mu_4^2)^2}{\mu_4^2(\mu_2^2\nu_6 - 2\mu_2\mu_4\nu_4 + \mu_4^2\nu_2)}\right]^{\frac{1}{7}}.$$

A.4. Simplification of the bandwidth ratio $h_{0,opt}/h_{0,cubic}$ in Henderson et al. (2015)

In this section, we show that the results of the bandwidth ratio $h_{0,opt}/h_{0,cubic}$ derived in Henderson et al. (2015) for conditional mean regression can be further simplified, and it coincides with the results in our paper for conditional quantile regression.

The two terms V_1 and $V_{1,3}$, which characterize the bandwidth ratio $h_{0,opt}/h_{0,cubic}$, are given in Eqs. (8) and (10) in Henderson et al. (2015) as follows,

$$V_1 = \frac{\nu_2}{\mu_2^2} \tag{27}$$

$$V_{1,3} = \frac{K_1}{K_2^2} + \frac{\nu_2}{\mu_2^2} - 2\frac{(\mu_4\mu_6 - \mu_2^2\mu_6)\nu_2 + (\mu_2^2\mu_4 - \mu_4^2)\nu_4}{\mu_2 K_2} \tag{28}$$

where $\mu_j = \int K(v)v^j dv$, $\nu_j = \int K^2(v)v^j dv$, $K(v)$ the kernel function; and the two constant terms K_1 and K_2 represent

$$\begin{aligned} K_1 &= (\mu_4\mu_6 - \mu_2^2\mu_6)^2\nu_2 + (\mu_2^2\mu_4 - \mu_4^2)^2\nu_6 \\ &\quad + 2(\mu_4\mu_6 - \mu_2^2\mu_6)(\mu_2^2\mu_4 - \mu_4^2)\nu_4 \\ &= (\mu_4-\mu_2^2)^2\mu_6^2\nu_2 + (\mu_4-\mu_2^2)^2\mu_4^2\nu_6 - 2(\mu_4-\mu_2^2)^2\mu_4\mu_6\nu_4 \\ &= (\mu_4-\mu_2^2)^2(\mu_6^2\nu_2 + \mu_4^2\nu_6 - 2\mu_4\mu_6\nu_4), \end{aligned}$$

and

$$\begin{aligned} K_2 &= \mu_2\mu_4\mu_6 - \mu_4^3 + \mu_2^2\mu_4^2 - \mu_2^3\mu_6 \\ &= \mu_2\mu_6(\mu_4-\mu_2^2) - \mu_4^2(\mu_4-\mu_2^2) \\ &= (\mu_4-\mu_2^2)(\mu_2\mu_6-\mu_4^2). \end{aligned}$$

Substituting the values of K_1 and K_2 into (28) and simplifying the expression of $V_{1,3}$, we have

$$\begin{aligned} V_{1,3} &= \frac{K_1}{K_2^2} + \frac{\nu_2}{\mu_2^2} - 2\frac{(\mu_4\mu_6 - \mu_2^2\mu_6)\nu_2 + (\mu_2^2\mu_4 - \mu_4^2)\nu_4}{\mu_2 K_2} \\ &= \frac{(\mu_6^2\nu_2 + \mu_4^2\nu_6 - 2\mu_4\mu_6\nu_4)\mu_2^2}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2} + \frac{\nu_2(\mu_2\mu_6-\mu_4^2)^2}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2} \\ &\quad -2\frac{(\mu_2\mu_6\nu_2 - \mu_2\mu_4\nu_4)(\mu_2\mu_6-\mu_4^2)}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2} \\ &= \frac{1}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2}(\mu_2^2\mu_6^2\nu_2 + \mu_2^2\mu_4^2\nu_6 - 2\mu_2^2\mu_4\mu_6\nu_4 \end{aligned}$$

$$
\begin{aligned}
&+\mu_2^2\mu_6^2\nu_2+\mu_4^4\nu_2-2\mu_2\mu_4^2\mu_6\nu_2\\
&-2\mu_2^2\mu_6^2\nu_2+2\mu_2\mu_4^2\mu_6\nu_2+2\mu_2^2\mu_4\mu_6\nu_4-2\mu_2\mu_4^3\nu_4)\\
=&\ \frac{\mu_2^2\mu_4^2\nu_6-2\mu_2\mu_4^3\nu_4+\mu_4^4\nu_2}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2}\\
=&\ \frac{\mu_4^2(\mu_2^2\nu_6-2\mu_2\mu_4\nu_4+\mu_4^2\nu_2)}{\mu_2^2(\mu_2\mu_6-\mu_4^2)^2}.
\end{aligned}
$$

After simplification, it is easy to see that the values V_1 and $V_{1,3}$ in (27) and (28) for conditional mean regression, after rescaling by the same constant $\tau(1-\tau)$, equal to V_1 and $V_{1,3}$ in (6) and (14) derived for conditional quantile regression. Therefore, the bandwidth ratio $h_{0,opt}/h_{0,cubic}$ in Henderson et al. (2015) is

$$
\frac{h_{0,opt}}{h_{0,cubic}}=\left(\frac{V_1}{V_{1,3}}\right)^{\frac{1}{7}}=\left[\frac{\nu_2(\mu_2\mu_6-\mu_4^2)^2}{\mu_4^2(\mu_2^2\nu_6-2\mu_2\mu_4\nu_4+\mu_4^2\nu_2)}\right]^{\frac{1}{7}},
$$

which coincides with the bandwidth ratio $h_{0,opt}/h_{0,cubic}$ for conditional quantile regression in our paper.

References

Anglin, P.M., Gençay, R., 1996. Semiparametric estimation of a hedonic price function. J. Appl. Econometrics 11, 633–648.
Buchinsky, M., 1994. Changes in the U.S. wage structure 1963–1987: application of quantile regression. Econometrica 62, 405–458.
Buchinsky, M., Hahn, J., 1998. An alternative estimator for the censored quantile regression model. Econometrica 66, 653–671.
Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. 1, 412–420.
Cai, Z., 2002. Regression quantile for time series. Econometric Theory 18, 169–192.
Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. J. Econometrics 167, 413–425.
Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. J. Amer. Statist. Assoc. 103 (484), 1595–1608.
Chaudhuri, P., 1991. Nonparametric estimates of regression quantiles and their local bahadur representation. Ann. Statist. 2, 760–777.
Chaudhuri, P., Doksum, K., Samarov, A., 1997. On average derivative quantile regression. Ann. Statist. 25 (2), 715–744.
Chernozhukov, V., Hansen, C., 2004. The effects of 401(k) participation on the wealth distribution: an instrumental quantile regression analysis. Rev. Econom. Statist. 86 (3), 735–751.
Engle, R.F., Manganelli, S., 2004. CAViaR: conditional autoregressive Value at Risk by regression quantile. J. Bus. Econom. Statist. 22, 367–381.
Fan, J., Gijbels, I., 1995. Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. J. Roy. Statist. Soc. Ser. B 57 (2), 371–394.
Hall, P., Lahiri, S.N., Polzehl, J., 1995. On bandwidth choice in nonparametric regression with both short-and long-range dependent errors. Ann. Statist. 23 (6), 1921–1936.
Hall, P., Li, Q., Racine, J.S., 2007. Nonparametric estimation of regression functions in the presence of irrelevant regressors. Rev. Econom. Statist. 89 (4), 784–789.
Hallin, M., Lu, Z., Yu, K., 2009. Local linear spatial quantile regression. Bernoulli 15, 659–686.
Härdle, W., Vieu, P., 1992. Kernel regression smoothing of time series. J. Time Series Anal. 13 (3), 209–232.
Henderson, D.J., Li, Q., Parmeter, C.F., Yao, S., 2015. Gradient-based smoothing parameter selection for nonparametric regression estimation. J. Econometrics 184 (2), 233–241.
Knight, K., 1998. Limiting distributions for L_1 regression estimators under general conditions. Ann. Statist. 26, 755–770.
Koenker, R., 2005. Quantile Regression. Cambridge University Press.
Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33–50.
Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. Econometrica 70, 1583–1612.
Leung, D.H.Y., 2005. Cross-validation in nonparametric regression with outliers. Ann. Statist. 2291–2310.
Li, Q., Racine, J., 2004. Cross-validated local linear nonparametric regression. Statist. Sinica 14 (2), 485–512.
Li, D., Li, Q., Li, Z., 2002. Nonparametric estimation of conditional quantile regression with mixed discrete and continuous data, unpublished manuscript (submitted for publication).
Li, Q., Lin, J., Racine, J., 2013. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. J. Bus. Econom. Statist. 31, 57–65.
Lin, Z., Li, D., 2007. Asymptotic normality of L1norm kernel estimator of conditional median under association dependence. J. Multivariate Anal. 98, 1214–1230.
Li, Q., Racine, J.S., 2006. Nonparametric Econometrics: Theory and Practice. Princeton University Press.
Li, Q., Racine, J.S., 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. J. Bus. Econom. Statist. 26 (4), 423–434.
Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. J. Time Ser. Anal. 17, 571–599.
Müller, H.-G., Stadmuller, U., Schmitt, T., 1987. Bandwidth choice and confidence intervals for derivatives of noisy data. Biometrika 74 (4), 743–749.
Parmeter, C.F., Henderson, D.J., Kumbhakar, S.C., 2007. Nonparametric estimation of a hedonic price function. J. Appl. Econometrics 22, 695–699.
Powell, J.L., 1986. Censored regression quantiles. J. Econometrics 32, 143–155.
Racine, J.S., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data. J. Econometrics 119 (1), 99–130.
Rice, J., 1984. Boundary modification for kernel regression. Commun. Stat., A 13 (7), 893–900.
Stone, C.J., 1977. Consistent nonparametric regression. Ann. Statist. 4, 595–645.
Wei, Y., Pere, A., Koenker, R., He, X., 2006. Quantile regression methods for reference growth charts. Stat. Med. 25, 1369–1382.
Xiao, Z., 2009. Quantile cointegrating regression. J. Econometrics 150, 248–260.
Xia, Y., Li, W.K., 2002. Asymptotic behavior of bandwidth selected by the cross-validation method for local polynomial fitting. J. Multivariate Anal. 83 (2), 265–287.
Yu, K., Jones, M.C., 1998. Local linear quantile regression. J. Amer. Statist. Assoc. 93 (441), 228–237.
Yu, K., Lu, Z., 2004. Local linear additive quantile regression. Scand. J. Stat. 31, 333–346.
Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: applications and current research areas. The Statistician 52 (3), 331–350.