Nonparametric Quantile Regression Estimation with Mixed Discrete and Continuous Data

Degui Li[*] University of York

Qi Li[†] Texas A&M University

Zheng Li[‡] North Carolina State University

February 3, 2020

Abstract

In this paper, we investigate the problem of nonparametrically estimating a conditional quantile function with mixed discrete and continuous covariates. A local linear smoothing technique combining both continuous and discrete kernel functions is introduced to estimate the conditional quantile function. We propose using a fully data-driven cross-validation approach to choose the bandwidths, and further derive the asymptotic optimality theory. In addition, we also establish the asymptotic distribution and uniform consistency (with convergence rates) for the local linear conditional quantile estimators with the data-dependent optimal bandwidths. Simulations show that the proposed approach compares well with some existing methods. Finally, an empirical application with the data taken from the IMDb website is presented to analyze the relationship between box office revenues and online rating scores.

Keywords: Bandwidth selection, Discrete regressors, Local linear smoothing, Nonparametric estimation, Quantile regression

JEL Classification: C13, C14, C35

[*]Department of Mathematics, University of York, York, YO10 5DD, UK. Email: degui.li@york.ac.uk.

[†]The corresponding author. Department of Economics, Texas A&M University, College Station, TX, 77843, USA. Email: qi-li@tamu.edu.

[‡]Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC, 27695, USA. Email: zli42@ncsu.edu.

1 Introduction

In recent years, there has been an increasing interest on nonparametric estimation of the regression models, as the nonparametric approach allows the data "speak for themselves" and thus has the ability to detect regression structures which may be difficult to uncover by traditional parametric modelling approaches. Various nonparametric methods have attracted much attention of statisticians and econometricians (c.f., Green and Silverman, 1994; Wand and Jones, 1995; Fan and Gijbels, 1996; Pagan and Ullah, 1999; Horowitz, 2009). One of the most commonly-used nonparametric estimation methods is the local linear smoothing method as it has advantages over the traditional Nadaraya-Watson kernel approach, such as higher asymptotic efficiency, design adaption and automatic boundary correction. We refer to the book by Fan and Gijbels (1996) for a detailed account on this subject.

Most of the aforementioned literature focuses on the nonparametric estimation with continuous regressors. However, in practice, it is not uncommon that some of the regressors might be discrete (e.g., gender, race and religious belief). Although in principle one can split the whole sample into many cells (determined by values of the discrete regressors) to handle the discrete variables, in practice such a sample-splitting method quickly becomes infeasible when the number of discrete cells is large. Indeed as pointed out by Li and Racine (2004), the naive splitting method may perform poorly when the number of subgroups is relatively large and the number of observations in some subgroups is small. To address this problem, they consider a nonparametric kernel-based method which smoothes both continuous and discrete covariates. Such a method works well in practice.

It is well known that the conditional mean function may not be a good representative of the impact of the explanatory variables on the response variable. Hence, it is often of interest to model conditional quantiles when studying the regression relationship between a response variable and some explanatory variables. Since the seminal paper by Koenker and Bassett (1978), the quantile regression method has been widely used in many disciplines such as economics, finance, political science and other social science fields. The quantile regression serves as a robust alternative to the mean regression. Recent developments on parametric and nonparametric quantile estimation and inference include Yu and Jones (1998), Cai (2002), Yu and Lu (2004), Koenker and Xiao (2006), Cai and Xu (2008), Hallin, Lu and Yu (2009), Escanciano and Velasco (2010), Kong, Linton and Xia (2010), Belloni and Chernozhukov (2011), Galvao (2011), Guerre and Sabbah (2012), Cai and Xiao (2012), Galvao, Lamarche and Lima (2013), Li, Lin and Racine (2013), Spokoiny, Wang and Härdle (2013), Escanciano and Goh (2014), Qu and Yoon (2015), Racine and Li (2017), Belloni *et al* (2019) and Zhu *et al* (2019). Koenker (2005) and Koenker *et al* (2017) give a comprehensive overview on various methodologies in quantile regression and their applications. In particular, the paper by

Li, Lin and Racine (2013) is among the first to estimate the conditional cumulative distribution function (CDF) nonparametrically by smoothing both the discrete and continuous covariates, and then obtain quantile regression function estimation by inverting the estimated conditional CDF at the desired quantiles. The bandwidths are chosen optimally in estimating the nonparametric CDF, and therefore, they may not be optimal for estimating for estimating the conditional quantile regression function.

In this paper, we propose a different nonparametric method to estimate the conditional quantile function via minimizing a local linear weighted "check function" (objective function) defined in Section 2. To tackle a general setting with mixed continuous and discrete regressors, we construct the local linear smoothing check function with both the continuous and discrete kernel functions involved. As the numerical performance of the local linear estimation is sensitive to the bandwidths or smoothing parameters, we further study the choice of smoothing parameters in the local linear quantile estimation, proposing a completely data-driven rescaled cross-validation (RCV) approach to directly choose the optimal smoothing parameters and deriving their asymptotic optimality property. As pointed out by Yao and Tong (1998), the RCV method is computationally faster than the conventional "leave-one-out" cross-validation method in selecting the optimal smoothing parameters. This advantage would be more significant in nonparametric quantile regression estimation which lacks a closed-form solution in the estimation procedure.

The check function based quantile regression estimation method proposed in this paper has at least three advantages over the inverse-CDF approach proposed by Li, Lin and Racine (2013). First, it is computationally much more efficient than the inverse-CDF approach. Our check-function-based RCV method requires $O(n^2)$ computations, whereas the inverse-CDF-based CV method involves $O(n^3)$ computations, where n is the sample size. This is a significant improvement in the computational aspect. The second advantage of using our method is that, besides the estimated conditional quantile function, we also obtain the derivative function (of the conditional quantile function with respect to the continuous components) estimate and its asymptotic theory, while it seems difficult to obtain derivative function estimation and to derive the related asymptotic theory if one uses the inverse-CDF method. The third advantage is that, in practice the optimal smoothing parameters often vary over τ, and our method addresses this issue well by providing optimal smoothing parameters for each specific quantile $\tau \in (0,1)$. In contrast, the inverse-CDF method gives the same smoothing parameters for all $\tau \in (0,1)$, and does not have the flexibility of choosing τ-dependent optimal smoothing parameters to conditional quantile function estimation. Under some mild conditions, we prove the point-wise asymptotic normal distribution and uniform convergence rates for the developed local linear quantile estimators using the data-dependent bandwidths determined by the RCV approach. We use simulations to illustrate the finite-sample behavior of the proposed method and compare our approach with some existing methods. Finally we apply the proposed nonparametric quantile regression estimation method to study the relationship between box office revenues and online rating scores using the dataset collected from the IMDb website.

2 Local linear quantile regression estimation

Suppose that (Y_i, X_i, Z_i), $i = 1, \cdots, n$, are the observations independently drawn from an identical distribution, where $Y_i \in \mathcal{R}$ is univariate, $X_i \in \mathcal{D}_x$ is a p-dimensional continuous random vector and $Z_i \in \mathcal{D}_z$ is a q-dimensional discrete random vector, $\mathcal{D}_x = \left\{(x_1, \cdots, x_p)^\intercal : \underline{c}_j \leqslant x_j \leqslant \overline{c}_j,\ j = 1, \cdots, p\right\}$ is a bounded subset of $\mathcal{R}^p$ with $\underline{c}_j < \overline{c}_j$, $j = 1, \cdots, p$, being bounded constants, and $\mathcal{D}_z$ is a finite support of the discrete vector Z_i. For expositional simplicity, we assume that each component of Z_i only takes non-negative integer values. We estimate the conditional quantile function and its derivatives (with respect to the continuous components) by minimizing a local linear weighted objective function, and then introduce a completely data-driven method to choose appropriate smoothing parameters involved.

For $y \in \mathcal{R}$, $x = (x_1, \cdots, x_p)^\intercal \in \mathcal{D}_x$ and $z = (z_1, \cdots, z_q)^\intercal \in \mathcal{D}_z$, we denote $\mathsf{F}(y|x, z)$ as the conditional CDF of the response variable Y_i (evaluated at y) given the covariates $X_i = x$ and $Z_i = z$. For $0 < \tau < 1$, we let $\mathsf{Q}_\tau(x, z)$ be the conditional τ-quantile regression function of Y_i given $X_i = x$ and $Z_i = z$, i.e., $\mathsf{Q}_\tau(x, z) = \inf\{y \in \mathcal{R} : \mathsf{F}(y|x, z) \geqslant \tau\}$, or equivalently

$$\mathsf{Q}_\tau(x, z) = \arg\min_{a \in \mathcal{R}} \mathsf{E}\left[\rho_\tau(Y_i - a)\middle| X_i = x, Z_i = z\right], \tag{2.1}$$

where $\rho_\tau(\cdot)$ is the τ-quantile check function (or loss function) defined as $\rho_\tau(y) = y\,(\tau - \mathsf{I}\{y < 0\})$ with $\mathsf{I}\{\mathcal{A}\}$ being the indicator function of the set $\mathcal{A}$.

We apply the local linear smoothing approach to estimate the τ-quantile regression function $\mathsf{Q}_\tau(x_0, z_0)$ based on the definition given in (2.1), where $x_0 = (x_{0,1}, \cdots, x_{0,p})^\intercal \in \mathcal{D}_x$ and $z_0 = (z_{0,1}, \cdots, z_{0,q})^\intercal \in \mathcal{D}_z$. Due to mixture of discrete and continuous data in the regressors, two types of kernel-weights are required to construct the locally weighted loss function. For the continuous regressors, we use a conventional kernel weight $\mathbf{K}_h(X_i - x_0)$ defined by

$$\mathbf{K}_h(X_i - x_0) = \frac{1}{h_1 \cdots h_p}\mathbf{K}\left(\frac{X_i - x_0}{h}\right) = \prod_{j=1}^{p} \frac{1}{h_j} K\left(\frac{X_{i,j} - x_{0,j}}{h_j}\right),$$

where $h = (h_1, \cdots, h_p)^\intercal$, h_j is the bandwidth for the j-th continuous covariate $X_{i,j}$, and $K(\cdot)$ is a univariate kernel function. For the discrete covariates, we use the following discrete kernel (with the convention that 0^0 equals to 1):

$$\Lambda_\lambda(Z_i, z_0) = \prod_{j=1}^{q} \lambda_j^{I\{Z_{i,j} \neq z_{0,j}\}},$$

where $\lambda = (\lambda_1, \cdots, \lambda_q)^\intercal$, $\lambda_j \in [0, 1]$ is the bandwidth for the j-th discrete covariate $Z_{i,j}$. The local linear estimates of $Q_\tau(x_0, z_0)$ and its derivatives (with respect to the continuous components) $Q'_{\tau,j}(x_0, z_0)$, $j = 1, 2, \cdots, p$, are obtained by minimizing the weighted loss function

$$L_n(\alpha, \beta; x_0, z_0) = \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \alpha - (X_i - x_0)^\intercal \beta \right) \mathbf{K}_h(X_i - x_0) \Lambda_\lambda(Z_i, z_0) \tag{2.2}$$

with respect to α and $\beta = (\beta_1, \cdots, \beta_p)^\intercal$. We denote the minimizers by

$$\hat{\alpha} \equiv \widehat{Q}_\tau(x_0, z_0), \qquad \hat{\beta}_j \equiv \widehat{Q}'_{\tau,j}(x_0, z_0), \ \ j = 1, \cdots, p.$$

The above check function based local linear conditional quantile estimator with mixed discrete and continuous covariates was previously considered in Li and Racine (2008). However, they did not provide asymptotic analysis on the selection of optimal smoothing parameters by some data-driven methods, which is the main task of the present paper.

When the smoothing parameter in the discrete kernel is chosen as a vector of zeros, the above approach reduces to the traditional local linear quantile estimation method which splits the full sample into several groups (sub-samples) according to different values that the discrete covariates can assume. Then one may directly apply the local linear quantile estimation methodology and theory developed in the literature for the case of purely continuous regressors (c.f., Yu and Jones, 1998; Cai and Xu, 2008). However, as pointed out by Li and Racine (2004), such a naive sample-splitting method may increase the estimation variance. In particular, it is well known that if the sample size in the subgroup is too small, one cannot expect to get reliable estimation results with the sample-splitting local linear quantile estimation method.

It is of crucial importance to appropriately select smoothing parameters in the nonparametric local linear smoothing procedure. In this paper we propose to use a completely data-driven method to choose the optimal bandwidth vectors h and λ. The cross-validatory bandwidth selection criterion has been extensively studied in the context of local kernel-based mean regression estimation with continuous regressors (c.f., Rice, 1984; Hall, Lahiri and Polzehl, 1995; Xia and Li, 2002; Leung, 2005). In recent years, there has also been an increasing interest in extending this bandwidth selection approach to the case with mixed continuous and discrete regressors (c.f., Li and Racine, 2004). However, most of the existing literature focuses on the bandwidth selection in the kernel-based estimation in the context of conditional mean regression. Extension of the cross-validation bandwidth selection method to the conditional quantile regression is non-trivial and the derivation of the asymptotic optimality property is challenging as there is no closed-form expression for the local linear quantile estimator. Li, Lin and Racine (2013) studied the bandwidth selection in nonparametric quantile regression estimation. Their optimal bandwidths for both the continuous and discrete regressors are chosen when estimating the nonparametric CDF. As a result, the chosen bandwidths are optimal for the CDF estimation, but not for the quantile regression estimation. In this paper, we introduce a data-driven method to directly select bandwidth vectors which are optimal for quantile regression estimation.

We split the full sample into two sets: the training set $\mathcal{M}_1 = \{(Y_i, X_i, Z_i), i = 1, \cdots, m\}$, and the validation set $\mathcal{M}_2 = \{(Y_i, X_i, Z_i), i = m+1, \cdots, n\}$, where m has the same order as n (say $m = \lfloor n/2 \rfloor$). For $j = m+1, \cdots, n$, let $\widehat{Q}_{\mathcal{M}_1}(X_j, Z_j; h, \lambda) \equiv \widehat{Q}_{\tau,\mathcal{M}_1}(X_j, Z_j; h, \lambda)$ be the local linear estimated value of $Q_\tau(X_j, Z_j)$ with bandwidth vectors h and λ, which are obtained by minimizing (2.2) with (x_0, z_0) and $\sum_{i=1}^n$ replaced by (X_j, Z_j) and $\sum_{i=1}^m$, respectively. Define the following objective function:

$$\mathsf{CV}(h, \lambda) = \frac{1}{n-m} \sum_{j=m+1}^{n} \rho_\tau \left(Y_j - \widehat{Q}_{\mathcal{M}_1}(X_j, Z_j; h, \lambda) \right) M(X_j), \tag{2.3}$$

where $M(\cdot)$ is a weight function that trims out observations whose continuous components are close to the boundary. Let

$$\left(\widehat{h}_m, \widehat{\lambda}_m \right) = \arg \min_{(h,\lambda) \in \mathcal{H}_m} \mathsf{CV}(h, \lambda), \tag{2.4}$$

where $\mathcal{H}_m$ is a set of grid points $[h(k), \lambda(k)]$, $k = 1, \cdots, L_m$, satisfying Assumption 5(i)(ii) in Appendix A and that

$$\max_{1 \leqslant k \leqslant L_m} \min_{j \neq k} \|h(k) - h(j)\| + \max_{1 \leqslant k \leqslant L_m} \min_{j \neq k} \|\lambda(k) - \lambda(j)\| \leqslant \gamma_m, \tag{2.5}$$

$\|\cdot\|$ denotes the Euclidean norm. Assumption 5(iii) in Appendix A gives some restrictions on L_m and γ_m, ensuring that the grid points are sufficiently dense in the set $\mathcal{H}_m$. With $\widehat{h}_m$ and $\widehat{\lambda}_m$, we do the re-scaling to obtain the RCV bandwidth vectors for the full sample as

$$\widehat{h} = \left(\widehat{h}_1, \cdots, \widehat{h}_p \right)^\intercal = \widehat{h}_m \left(\frac{m}{n} \right)^{1/(4+p)} \quad \text{and} \quad \widehat{\lambda} = \left(\widehat{\lambda}_1, \cdots, \widehat{\lambda}_q \right)^\intercal = \widehat{\lambda}_m \left(\frac{m}{n} \right)^{2/(4+p)}. \tag{2.6}$$

The motivation for constructing the RCV optimal bandwidth vectors (for full sample) in (2.6) is relevant to asymptotic order of the theoretically optimal bandwidths, which will be made clear later in the section. As the training set size in the RCV bandwidth selection is usually smaller than the size $n-1$ in the conventional "leave-one-out" cross-validation, we expect that the computational time for the RCV method would be faster in particular when the full sample size is large. The RCV method was studied by Yao and Tong (1998) for bandwidth selection in kernel regression estimation with univariate continuous regressor, and a similar idea was recently used by Fan, Guo and Hao (2012) and Chen, Fan and Li (2018) for error variance estimation in high-dimensional mean regression models.

To derive the asymptotic optimality for the RCV selected bandwidths $\widehat{h}$ and $\widehat{\lambda}$, we need some additional notation. Let $e_i = Y_i - Q_\tau(X_i, Z_i)$ and assume that $\mathsf{P}(e_i \leqslant 0|X_i = x, Z_i = z) = \tau$ for $x \in \mathcal{D}_x$ and $z \in \mathcal{D}_z$. Define $f_{xz}(\cdot,\cdot)$ as the joint probability density function of X_i and Z_i and $f_e(\cdot|x,z)$ as the conditional density function of e_i given $X_i = x$ and $Z_i = z$. Let $Q''_{\tau,j}(x,z)$, $j = 1,2,\cdots,p$, be the second-order derivative function of $Q_\tau(\cdot,\cdot)$ with respect to the j-th continuous component evaluated at the point (x,z). Define

$$\begin{aligned} b_m(X_i, Z_i; h, \lambda) &= \frac{\mu_2}{2}\sum_{j=1}^{p} h_j^2 Q''_{\tau,j}(X_i, Z_i) + \sum_{z\in\mathcal{D}_z}\sum_{j=1}^{q} I_j(z, Z_i)\lambda_j \frac{f(X_i, z)}{f(X_i, Z_i)}\left[Q_\tau(X_i, z) - Q_\tau(X_i, Z_i)\right], \\ \sigma_m^2(X_i, Z_i; h) &= \frac{1}{mH}\cdot\frac{\tau(1-\tau)\nu_0^p}{\left[f_e(0|X_i, Z_i)\right]^2 f_{xz}(X_i, Z_i)}, \end{aligned}$$

where $\mu_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$ for $j = 0,1,2,\cdots$, $f(x,z) = f_e(0|x,z)f_{xz}(x,z)$, $I_j(\bar{z},\widetilde{z}) = \mathsf{I}\{\bar{z}_j \neq \widetilde{z}_j\}\prod_{k=1,\neq j}^{q}\mathsf{I}\{\bar{z}_k = \widetilde{z}_k\}$ for $\bar{z} = (\bar{z}_1,\cdots,\bar{z}_q)^\intercal$ and $\widetilde{z} = (\widetilde{z}_1,\cdots,\widetilde{z}_q)^\intercal$, and $H = \prod_{j=1}^{p} h_j$. The following theorem gives the uniform asymptotic expansion of $\mathsf{CV}(h,\lambda)$, which is crucial to derive the asymptotic optimality of $\widehat{h}$ and $\widehat{\lambda}$ defined in (2.6).

THEOREM 2.1. *Suppose that Assumptions 1–5 in Appendix A are satisfied and there exists a constant* $0 < \varpi < 1$ *such that* $m/n \to \varpi$*. Then, we have that, uniformly over* $(h,\lambda) \in \mathcal{H}_m$

$$\mathsf{CV}(h,\lambda) = \mathsf{CV}_* + \frac{1}{2}\mathsf{E}\left\{\left[b_m^2(X_i, Z_i; h, \lambda) + \sigma_m^2(X_i, Z_i; h)\right] M(X_i) f_e(0|X_i, Z_i)\right\} + \text{s.o.}, \tag{2.7}$$

where $\mathsf{CV}_* = (n-m)^{-1}\sum_{i=m+1}^{n}\rho_\tau(e_i)M(X_i)$ *is unrelated to* h *and* λ*, and "s.o." represents terms with smaller asymptotic probability orders than the second term on the right hand side of (2.7).*

The proof of Theorem 2.1 is provided in Appendix B.1. Define the following mean squared errors of the local linear quantile regression estimation for the validation set $\mathcal{M}_2$:

$$\mathsf{MSE}_{\mathcal{M}_2}(h,\lambda) = \frac{1}{n-m}\sum_{i=m+1}^{n}\left[Q_\tau(X_i, Z_i) - \widehat{Q}_{\mathcal{M}_1}(X_i, Z_i; h, \lambda)\right]^2 M(X_i) f_e(0|X_i, Z_i), \tag{2.8}$$

and its asymptotic leading term:

$$\mathsf{MSE}^{L}_{\mathcal{M}_2}(h,\lambda) = \mathsf{E}\left\{\left[b^2_m(X_i,Z_i;h,\lambda) + \sigma^2_m(X_i,Z_i;h)\right] M(X_i)f_e(0|X_i,Z_i)\right\}. \tag{2.9}$$

Let $a = (a_1, \cdots, a_p)^\intercal$ and $d = (d_1, \cdots, d_q)^\intercal$ with $a_j = h_j \cdot m^{1/(4+p)}$ and $d_j = \lambda_j \cdot m^{2/(4+p)}$. Then (2.9) can be written as $\mathsf{MSE}^{L}_{\mathcal{M}_2}(h,\lambda) = m^{-4/(4+p)} g(a, d)$, where

$$g(a,d) = \mathsf{E}\left\{\left[\overline{b}^2(X_i,Z_i;a,d) + \overline{\sigma}^2(X_i,Z_i;a)\right] M(X_i)f_e(0|X_i,Z_i)\right\} \tag{2.10}$$

with $\overline{b}(X_i, Z_i; a, d) = \frac{\mu_2}{2}\sum_{j=1}^{p} a_j^2 Q''_{\tau,j}(X_i, Z_i) + \sum_{z\in\mathcal{D}_z}\sum_{j=1}^{q} I_j(z, Z_i) d_j \frac{f(X_i,z)}{f(X_i,Z_i)}\left[Q_\tau(X_i, z) - Q_\tau(X_i, Z_i)\right]$, and $\overline{\sigma}^2(X_i, Z_i; a) = \frac{1}{a_1\cdots a_p}\cdot\frac{\tau(1-\tau)\nu_0^p}{[f_e(0|X_i,Z_i)]^2 f_{xz}(X_i,Z_i)}$. Note that the function $g(a, d)$ does not depend on m. We assume that there exist unique positive constants a_j^0, $j = 1, \cdots, p$, and non-negative constants d_k^0, $k = 1, \cdots, q$, that minimize $g(a, d)$ defined in (2.10). Li and Zhou (2005) discussed some sufficient conditions on existence and uniqueness of these constants in the context of optimal bandwidth selection in nonparametric kernel-based mean regression. With some minor modifications, their conditions are applicable to our setting. The theoretical optimal bandwidths are defined as $h_j^0 = a_j^0 \cdot n^{-1/(4+p)}$, $j = 1, \cdots, p$, and $\lambda_k^0 = d_k^0 \cdot n^{-2/(4+p)}$, $k = 1, \cdots, q$. The following theorem shows that the RCV selected bandwidths $\widehat{h}$ and $\widehat{\lambda}$ are asymptotically optimal.

THEOREM 2.2. *Suppose that the conditions of Theorem 2.1 are satisfied, and let $\widehat{h}$ and $\widehat{\lambda}$ be defined in (2.6). Then,*

$$n^{1/(4+p)}\widehat{h}_j \xrightarrow{p} a_j^0 \ \text{ for } \ j = 1, 2, \cdots, p, \quad \textit{and} \quad n^{2/(4+p)}\widehat{\lambda}_k \xrightarrow{p} d_k^0 \ \text{ for } \ k = 1, 2, \cdots, q. \tag{2.11}$$

The proof of Theorem 2.2 is given in Appendix B.2. Theorem 2.2 above shows that the data-driven RCV selected bandwidth vectors $\widehat{h}$ and $\widehat{\lambda}$ are asymptotically equivalent to the theoretically optimal ones $h^0 = \left(h_1^0, \cdots, h_p^0\right)^\intercal$ and $\lambda^0 = \left(\lambda_1^0, \cdots, \lambda_q^0\right)^\intercal$, extending some existing asymptotic optimality results on the CV nonparametric estimation with mixed continuous and discrete regressors from the mean regression setting (c.f., Theorem 3.1 in Li and Racine, 2004) to the quantile regression setting. The convergence results (2.11) plays a critical role in deriving the point-wise asymptotic normal distribution and uniform consistency of the local linear quantile estimator using the data-dependent bandwidth vectors $\widehat{h}$ and $\widehat{\lambda}$.

3 Asymptotic theory with the RCV selected bandwidths

In this section, we provide the point-wise asymptotic normal distribution and uniform consistency for the local linear quantile function estimator defined in Section 2 with the data-driven RCV selected bandwidths. As in Section 2, we define the following asymptotic bias term:

$$b(x_0, z_0; h, \lambda) = \frac{\mu_2}{2} \sum_{k=1}^{p} h_k^2 Q''_{\tau,k}(x_0, z_0) + \sum_{z \in \mathcal{D}_z} \sum_{k=1}^{q} I_k(z, z_0)\lambda_k \frac{f(x_0, z)}{f(x_0, z_0)} \left[Q_\tau(x_0, z) - Q_\tau(x_0, z_0)\right].$$

We start with a point-wise asymptotic normal distribution theory for the local linear quantile estimator $\widehat{Q}_\tau(x_0, z_0; \widehat{h}, \widehat{\lambda})$, where $\widehat{h}$ and $\widehat{\lambda}$ are defined in (2.6).

THEOREM 3.1. *Suppose that the conditions of Theorem 2.2 are satisfied. Then, we have*

$$\left(n\widehat{H}\right)^{1/2} \left[\widehat{Q}_\tau(x_0, z_0; \widehat{h}, \widehat{\lambda}) - Q_\tau(x_0, z_0) - b(x_0, z_0; \widehat{h}, \widehat{\lambda})\right] \xrightarrow{d} N\left[0, V(x_0, z_0)\right], \tag{3.1}$$

where $\widehat{H} = \prod_{j=1}^{p} \widehat{h}_j$, *and* $V(x_0, z_0) = \frac{\tau(1-\tau)\nu_0^p}{[f_e(0|x_0,z_0)]^2 f_{xz}(x_0,z_0)}$.

The proof of Theorem 3.1 is provided in Appendix B.3. The normalization rate in (3.1) is random as $\widehat{H}$ is a product of data-driven RCV selected bandwidths. Theorem 3.1 above can be seen as an extension of the corresponding results from the continuous regressors case (c.f., Yu and Jones, 1998; Cai and Xu, 2008; Hallin, Lu and Yu, 2009) to the mixed continuous and discrete regressors case. It is easy to find that the discrete kernel in the local linear quantile estimation does not contribute to the asymptotic variance, but it influences the form of the asymptotic bias, see, for example, the second term of $b(x_0, z_0; h, \lambda)$. This finding is similar to that obtained by Li, Lin and Racine (2013). In addition, as in Li and Li (2010), although the data-dependent bandwidths $\widehat{h}$ and $\widehat{\lambda}$ are used, the limit distribution in (3.1) remains the same as that using the deterministic optimal bandwidths.

In order to make use of the above limit distribution theory to conduct point-wise statistical inference on the quantile regression curves, we need to estimate the asymptotic bias and variance, both of which contain some unknown quantities. In general, there are several commonly-used approaches to construct their estimates. The first one is to use the plug-in method, which directly replaces the unknown quantities in the asymptotic bias and variance by appropriate estimated values. For example, the second-order derivatives of the quantile regression function $Q''_{\tau,j}(x_0, z_0)$ can be estimated through the local quadratic quantile regression (e.g., Cheng and Peng, 2002) and the density function $f_{xz}(x_0, z_0)$ can be consistently estimated by the kernel density estimation method, see Appendix B.5. The second way is to use the bootstrap method to obtain the estimates of the estimation bias and variance (e.g., Zhou, 2010). The third approach is to use undersmoothed bandwidths so that estimation bias terms are asymptotically negligible. When the sample is of small or medium size, the bootstrap method is usually preferred to the plug-in method. With

Theorem 3.1, we next briefly discuss using the plug-in method, i.e., to replace unknown quantities in the asymptotic bias and variance by consistent estimators in constructing point-wise confidence intervals. Letting $\widehat{b}(x_0, z_0; \widehat{h}, \widehat{\lambda})$ and $\widehat{V}(x_0, z_0)$ be estimators of $b(x_0, z_0; \widehat{h}, \widehat{\lambda})$ and $V(x_0, z_0)$ which are defined in Appendix B.5, then for $\alpha \in (0, 1)$, the $100(1-\alpha)\%$ confidence interval of $Q_\tau(x_0, z_0)$ is given by

$$\left[\widehat{Q}_\tau(x_0, z_0; \widehat{h}, \widehat{\lambda}) - \widehat{b}(x_0, z_0; \widehat{h}, \widehat{\lambda}) - c_{1-\alpha/2}\sqrt{\frac{\widehat{V}(x_0, z_0)}{n\widehat{H}}},\ \widehat{Q}_\tau(x_0, z_0; \widehat{h}, \widehat{\lambda}) - \widehat{b}(x_0, z_0; \widehat{h}, \widehat{\lambda}) + c_{1-\alpha/2}\sqrt{\frac{\widehat{V}(x_0, z_0)}{n\widehat{H}}}\right], \tag{3.2}$$

where $c_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of a standard normal random variable. However, the convergence rate of the bias estimation $\widehat{b}(x_0, z_0; \widehat{h}, \widehat{\lambda})$ is often slow (in particular when the sample size is relatively small), and the estimated variance used in construction of the confidence interval should account for the variability of both the conditional quantile function estimation and bias estimation. Calonico, Cattaneo and Titiunik (2014) and Calonico, Cattaneo, and Farrell (2018) derived new distribution theories to tackle this issue and introduced a robust confidence interval construction in the context of conditional mean regression model with continuous regressors. A further extension of this theory and methodology to the conditional quantile regression setting can be found in Qu and Yoon (2018), Chiang, Hsu and Sasaki (2019) and Chiang and Sasaki (2019). It would be interesting to apply this technique to modify the confidence interval defined in (3.2) and give the relevant theoretical justification. This is left as a future research topic. In Section 4.2 we discuss the construction of bootstrap confidence intervals (uniformly over quantile levels) and examine its finite-sample numerical performance.

Next, we present the uniform consistency with convergence rate for the conditional quantile estimator $\widehat{Q}_\tau(x, z; \widehat{h}, \widehat{\lambda})$ over $x \in \mathcal{D}_x(\epsilon)$, where $\mathcal{D}_x(\epsilon)$ is defined in Assumption 4.

THEOREM 3.2. *Suppose that the conditions of Theorem 2.2 are satisfied. Then, we have*

$$\sup_{x \in \mathcal{D}_x(\epsilon)} \left|\widehat{Q}_\tau(x, z; \widehat{h}, \widehat{\lambda}) - Q_\tau(x, z)\right| = O_P\left(n^{-2/(4+p)} \log^{1/2} n\right). \tag{3.3}$$

The proof of Theorem 3.2 is provided in Appendix B.4. The uniform convergence rate in (3.3) is the same as the conventional uniform convergence rates for the kernel-based estimators when the theoretically (non-random) optimal bandwidths are used. It is also of interest to further study the uniform distribution theory over τ and $x \in \mathcal{D}_x(\epsilon)$. For the case of purely continuous regressors, this problem has been explored in the existing literature. For example, the uniform convergence and distribution theory (over x but with τ fixed) for the kernel-based quantile estimation is studied by Härdle and Song (2010), and the weak convergence (uniformly over τ but with x fixed, and with deterministic bandwidth) for the local linear quantile estimation is considered by Qu and Yoon (2015). Their results can be used to construct simultaneous confidence bands for the quantile regression functions, facilitating the relevant uniform inference. In Appendix D of the supplemental document we extend Qu and Yoon (2015)'s uniform convergence result to the case of mixed continuous and discrete regressors. An important open question is whether the weak convergence result presented in Proposition E.1 of the supplementary document is still valid if we replace the deterministic h_τ and λ_τ by the RCV selected bandwidths $\widehat{h}_\tau$ and $\widehat{\lambda}_\tau$. We conjecture that the answer to this open question is 'affirmative', however, we are unable to prove this conjecture. We report a small-scale simulation study in Section 4.2 to examine the performance of a uniform bootstrap confidence interval procedure. The simulation results support our conjecture. We leave the challenging work of verifying this conjecture to a future research topic.

4 Simulation

In Section 4.1, we use simulations to examine the finite-sample performance of our proposed estimator and several existing methods; in Section 4.2, we discuss how to construct uniform bootstrap confidence intervals and evaluate its finite-sample performance via simulation.

4.1 Quantile estimation mean squared errors and bandwidth selection

We consider the following four conditional quantile function estimation methods: (i) the proposed check-function-based conditional quantile function estimator with the RCV selected bandwidths, denoted as "Check (RCV)"; (ii) the check-function-based conditional quantile function estimator with the bandwidths chosen by the conventional leave-one-out CV method, denoted as "Check (LOOCV)"; (iii) the traditional check-function-based quantile estimation that only smoothes the continuous covariate (thus splitting the sample into cells according to different values of the discrete covariate), denoted as "Check (non-smooth)"; and (iv) the nonparametric inverse-CDF estimation with the bandwidths for nonparametric CDF estimation chosen by the least-squares CV method suggested by Li, Lin and Racine (2013), denoted as "Inverse-CDF".

We design the data generating process (DGP) to capture general patterns of the empirical data to be used in Section 5 below. There are three major patterns: (i) the conditional quantile curves are nonlinear in the continuous covariate; (ii) the distribution of the response variable is stretched as the value of the continuous covariate increases; (iii) the distribution of the response variable conditional on the continuous covariate is not symmetric. The following DGP serves our purpose:

$$Y_i = X_i^2 + Z_i + \sqrt{X_i} \cdot u_i, \quad i = 1, \cdots, n, \tag{4.1}$$

where $X_i \sim \mathsf{Uniform}[0,4]$ and $Z_i \in \{0,1\}$ is a binary variable with $P(Z_i = 0) = 0.7$ and $P(Z_i = 1) = 0.3$, and the error u_i follows a shifted $F(10,10)$ distribution with zero mean, resulting in asymmetric distributional pattern. The quantile levels we consider in the simulation are $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$. We examine three sample sizes: $n = 100$, $n = 200$ and $n = 400$. For each simulation set up, the number of replications is 1000.

Table 1 reports the simulation results of the average mean squared error (MSE) for the four conditional quantile estimation methods. The first column of the table specifies the methods used in the conditional quantile estimation, and the next five columns specify the quantile levels we estimate at. The upper four rows of the MSEs are for the sample size of 100, the middle four rows are for the sample size of 200, and the lower four rows are for the sample size of 400. From the table, we find that the two check-function-based estimation methods: "Check (RCV)" and "Check (LOOCV)" smoothing over both the continuous and discrete covariates, generally perform better than the "Inverse-CDF" method and the "Check (non-smooth)" method which only smoothes over the continuous covariate. This advantage becomes more significant at the extreme quantile levels (say, $\tau = 0.1$ or 0.9), whereas the performance is similar among the four estimation methods when $\tau = 0.25$, 0.50, or 0.75. The advantage of the "Check (RCV)" and "Check (LOOCV)" methods over the "Inverse-CDF" method is mainly due to the fact that the optimal smoothing parameters for the "Inverse-CDF" method are not quantile-specific, lacking the flexibility of selecting different smoothing parameters at different quantile levels. In contrast, the optimal smoothing parameters for the proposed check-function-based estimation method are adaptive to the quantile levels. We also observe that "Check (LOOCV)" performs slightly better than "Check (RCV)" for almost all cases. This suggests that the asymptotic theory of the leave-one-out CV smoothing parameter selection is, although challenging to establish, a worthwhile future research topic.

The advantage of the "Check (RCV)" and "Check (LOOCV)" methods over the "Check (non-smooth)" method observed from Table 1 shows that smoothing the discrete variable can borrow data from the neighbouring cells to significantly reduce estimation variance, while only introduce mild bias. As a result, the MSE of the estimation can be reduced in finite samples. We further investigate the differences between the "Check (RCV)" and "Check (non-smooth)" methods by decomposing the MSE into squared bias and variance. Tables 2 and 3 report the average squared estimation bias and variance, respectively. Table 2 shows that the estimation bias of "Check (RCV)" is not larger than that of "Check (non-smooth)". The main reason is that the "Check (non-smooth)" method leads to smaller estimation bias and larger variance by setting the smoothing parameter for discrete covariate as 0. Consequently, to minimize the MSE and balance the squared bias and variance, the CV method tends to increase the optimal smoothing parameter for the continuous covariate. To confirm this, we further report the average value of the selected smoothing parameter for continuous covariate in Table 4. From the table, we find that the average bandwidth value of

Table 1: Average MSE values among the four quantile estimation methods

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			$n = 100$		
Check (RCV)	0.263	0.218	0.284	0.688	2.121
Check (LOOCV)	0.235	0.203	0.286	0.679	1.807
Check (non-smooth)	0.419	0.268	0.396	1.028	3.129
Inverse-CDF	0.423	0.210	0.285	0.740	2.994
			$n = 200$		
Check (RCV)	0.154	0.128	0.172	0.417	1.101
Check (LOOCV)	0.129	0.121	0.170	0.380	0.983
Check (non-smooth)	0.219	0.145	0.217	0.563	1.935
Inverse-CDF	0.249	0.125	0.166	0.406	1.647
			$n = 400$		
Check (RCV)	0.077	0.070	0.102	0.238	0.647
Check (LOOCV)	0.071	0.067	0.097	0.225	0.594
Check (non-smooth)	0.086	0.078	0.114	0.276	0.882
Inverse-CDF	0.135	0.069	0.095	0.238	0.927

"Check (non-smooth)" is indeed larger than that of "Check (RCV)" in almost all cases. On the other hand, Table 3 reveals that the "Check (RCV)" method has smaller estimation variance than the "Check (non-smooth)" method, coinciding with our expectation.

Table 2: Squared bias comparison between "Check (RCV)" and "Check (non-smooth)"

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			$n = 100$		
Check (RCV)	0.056	0.032	0.045	0.153	0.386
Check (non-smooth)	0.155	0.039	0.057	0.238	0.459
			$n = 200$		
Check (RCV)	0.038	0.022	0.041	0.119	0.310
Check (non-smooth)	0.092	0.032	0.054	0.159	0.327
			$n = 400$		
Check (RCV)	0.017	0.012	0.027	0.067	0.173
Check (non-smooth)	0.023	0.015	0.032	0.074	0.199

Finally we compare the RCV-selected smoothing parameters with the infeasible optimal smoothing parameters which are estimated by minimizing the infeasible MSE in (2.8), where X_i and Z_i are the simulated random sample and $Q_\tau(\cdot,\cdot)$ is the true quantile regression function defined in (4.1). We only consider the case when $\tau = 0.5$, and repeat the process for 1000 times and estimate the empirical distribution of the minimizers. Figure 1 reports the distributions of the infeasible optimal and RCV-selected smoothing parameters for the continuous covariate, respectively. The left, center

Table 3: Variance comparison between "Check (RCV)" and "Check (non-smooth)"

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			$n = 100$		
Check (RCV)	0.207	0.186	0.239	0.535	1.735
Check (non-smooth)	0.264	0.229	0.338	0.790	2.670
			$n = 200$		
Check (RCV)	0.116	0.106	0.131	0.298	0.791
Check (non-smooth)	0.127	0.113	0.163	0.404	1.608
			$n = 400$		
Check (RCV)	0.060	0.058	0.075	0.172	0.474
Check (non-smooth)	0.063	0.063	0.082	0.202	0.683

Table 4: Average of the selected bandwidths

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			$n = 100$		
Check (RCV)	0.156	0.183	0.278	0.206	0.210
Check (non-smooth)	0.189	0.206	0.311	0.225	0.240
			$n = 200$		
Check (RCV)	0.109	0.132	0.189	0.181	0.203
Check (non-smooth)	0.140	0.146	0.189	0.198	0.210
			$n = 400$		
Check (RCV)	0.088	0.116	0.173	0.139	0.158
Check (non-smooth)	0.095	0.120	0.173	0.154	0.184

and right panels correspond to $n = 100$, 200 and 400, respectively. As expected, the RCV selected smoothing parameter has larger variation than the infeasible optimally smoothing parameter. Nevertheless, both distributions are unimode and are approximately peaked at the same position. The rate of convergence of the RCV-selected smoothing parameter to the theoretically optimal smoothing parameter appears to be slow, which is a well-known fact even for the conditional mean function estimation setting (c.f., Härdle, Hall and Marron, 1988). Figure 2 reports the distributions of the infeasible optimal and RCV-selected smoothing parameters for the discrete covariate, from which we observe similar patterns as those in Figure 1.

4.2 Uniform bootstrap confidence interval

In this section, we discuss using a resampling bootstrap method to construct uniform confidence intervals at a given point (x_0, z_0), over quantile $\tau \in [0, 1]$. There are several advantages of using the bootstrap method for inference over the asymptotic approximation method: (i) it avoids calculation of the complicated estimation variances; (ii) by using undersmoothed bandwidth parameters,

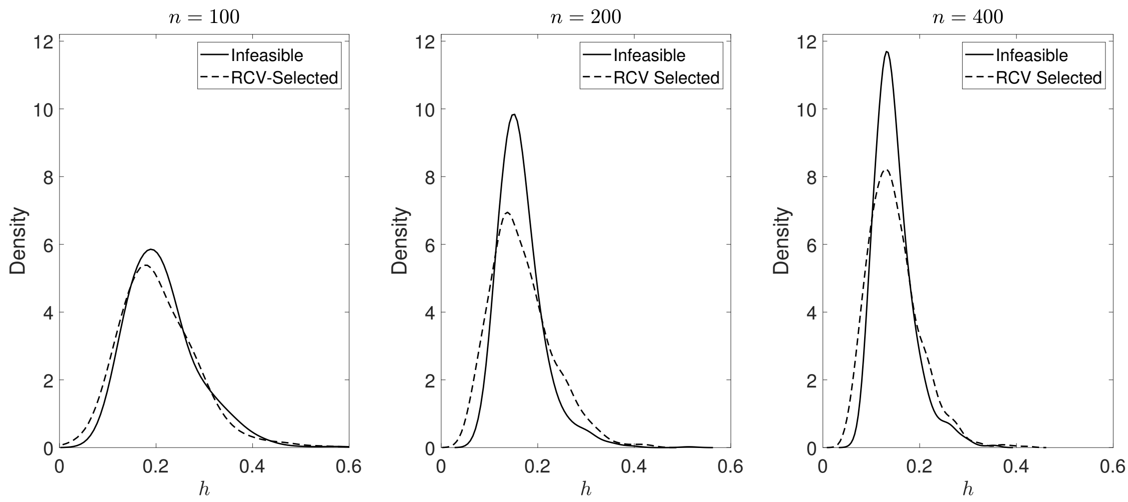

Figure 1: Infeasible optimal and RCV selected smoothing parameters for the continuous covariate

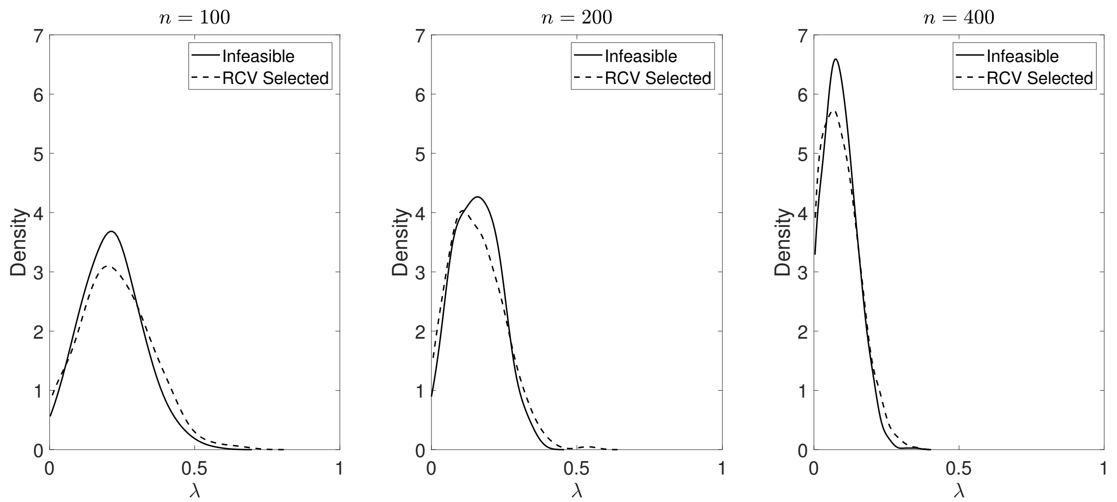

Figure 2: Infeasible optimal and RCV selected smoothing parameters for the discrete covariate

we also avoid calculation of the leading bias terms in the conditional quantile estimation; (iii) the bootstrap method often provides more accurate approximation result in particular when the sample size is small (Hall, 1992). The procedure of computing the uniform bootstrap confidence interval is as follows:

1. Use pair-wise bootstrap to resample from the original sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, and denote the bootstrap sample as $\{Y_i^*, X_i^*, Z_i^*\}_{i=1}^n$.

2. Partition the support of quantile τ into a grid $\{\tau_1, \tau_2, \cdots, \tau_m\}$, where m is the number of grid points. Using the bootstrap sample $\{Y_i^*, X_i^*, Z_i^*\}_{i=1}^n$, we estimate the conditional quantile function at (x_0, z_0) for each τ_j with bandwidth h^* and λ^*. Denote the estimate as $\widehat{Q}^*_{\tau_j}(x_0, z_0; h^*, \lambda^*)$,

$j = 1, 2, \cdots, m$, and calculate the maximum distance between $\widehat{Q}^*_{\tau_j}(x_0, z_0; h^*, \lambda^*)$ and $\widehat{Q}_{\tau_j}(x_0, z_0; h, \lambda)$ over τ_j, as follows:

$$D^*(x_0, z_0) = \max_{\tau \in \{\tau_1, \tau_2, \cdots, \tau_m\}} \left| \widehat{Q}^*_\tau(x_0, z_0; h^*, \lambda^*) - \widehat{Q}_\tau(x_0, z_0; h, \lambda) \right|,$$

where $\widehat{Q}_{\tau_j}(x_0, z_0; h, \lambda)$ is the conditional τ_j-quantile estimator using the original sample $\{Y_i, X_i, Z_i\}_{i=1}^n$ and the bandwidths h and λ.

3. Repeat Steps 1 and 2 for B times and obtain B estimates of $D^*(x_0, z_0)$. Calculate the α-th percentile of these B estimates, which is denoted as $D^*_\alpha(x_0, z_0)$. The $100(1-\alpha)\%$ uniform confidence interval for $Q_\tau(x_0, z_0)$ is obtained by

$$\left[\widehat{Q}_{\tau_j}(x_0, z_0; h, \lambda) - D^*_\alpha(x_0, z_0), \widehat{Q}_{\tau_j}(x_0, z_0; h, \lambda) + D^*_\alpha(x_0, z_0) \right].$$

In the simulation, we choose the grid points as $\{0.1, 0.2, \cdots, 0.9\}$ in Step 2. We further compare the performances of two bandwidth selection methods: (i) h, h^*, λ and λ^* are the RCV selected smoothing parameters, i.e., $h = h^* = \widehat{h}$, and $\lambda = \lambda^* = \widehat{\lambda}$; (ii) h, h^*, λ and λ^* are undersmoothed, i.e., $h = h^* = \widehat{h} n^{-\alpha_1}$ with $\alpha_1 = \frac{1}{20}$, and $\lambda = \lambda^* = \widehat{\lambda} n^{-\alpha_2}$ with $\alpha_2 = \frac{1}{10}$. The performance is evaluated using the coverage ratio. We construct 90% and 95% uniform confidence intervals at two evaluation points, $(x_0, z_0) = (2, 1)$ and $(x_0, z_0) = (3, 0)$. The sample sizes are $100, 200$ and 400, the number of replication is 1000, and the number of B in the bootstrap is 500.

Table 5 reports the coverage ratios. The upper and lower blocks are for the two evaluation points: $(x_0, z_0) = (2, 1)$ and $(x_0, z_0) = (3, 0)$, respectively. Within each block, the three rows correspond to the three sample sizes $n = 100, 200$, and 400, respectively. The left panel and the right panel of Table 5 are for the RCV and undersmoothed bandwidths, respectively. Within each panel, the first and second columns correspond to the coverage ratios of the 90% and 95% uniform confidence intervals, respectively. As the sample size increases, the coverage ratios, for the case with undersmoothed bandwidths, approach the corresponding nominal coverage probabilities. In particular, we find that for $n = 400$, the bootstrap confidence intervals using the undersmoothed bandwidths outperform those using the RCV bandwidths.

5 An empirical application

The conditional quantile regression plays an important role in empirical studies, especially in the circumstance where the conditional mean regression may not well reflect the nonlinear relationship among the variables of interest. For example, in the studies of individual income, the mean estimate

Table 5: Coverage ratio of uniform confidence intervals

Bandwidth	RCV		Undersmoothed	
Nominal Coverage	90%	95%	90%	95%
	Evaluated at $(x_0 = 2, z_0 = 1)$			
$n = 100$	0.850	0.895	0.856	0.892
$n = 200$	0.869	0.917	0.869	0.901
$n = 400$	0.883	0.929	0.910	0.941
	Evaluated at $(x_0 = 3, z_0 = 0)$			
$n = 100$	0.873	0.927	0.874	0.927
$n = 200$	0.871	0.927	0.895	0.940
$n = 400$	0.847	0.914	0.905	0.949

may be distorted by outliers, while the median is robust. In this section, we apply our method to study the relationship between word-of-mouth and box office revenue. As the internet penetrates our life, more and more online rating websites become prosperous. They cover movies, hotels, restaurants, etc., and help spread word-of-mouth at an unprecedented speed. The relationship between word-of-mouth and product sales is of great interest to researchers. Dellarocas, Zhang and Awad (2007) studied the forecasting power of online reviews on future box office revenue; Duan, Gu and Whinston (2008) divided the effects of online reviews into persuasive effect and awareness effect, and showed that the awareness effect is the major driver of box office revenue; Gilchrist and Sands (2016) documented the social spillover effect in box office revenue. We provide further empirical evidence to this subject, applying our proposed method to estimate the quantiles of box office revenue conditional on online rating score and showing clear upward trends in each examined quantile as online rating score increases.

5.1 Data description

Our dataset is obtained from the IMDb website (www.imdb.com). For each movie, we observe characteristics including gross revenue, average IMDb rating, release date, genre and Motion Picture Association of America (MPAA) film ratings. We collect movies that are listed on IMDb with release date between 2011 and 2014. Genre takes four categories: "Action", "Animation," "Comedy" and "Drama", and MPAA rating is either "PG" (Parental Guidance Suggested) or "PG-13" (Parents Strongly Cautioned). The number of movies in the sample is 227.

Table 6 reports the summary statistics of the continuous variables that we use in the following empirical analysis. The gross revenue variable has a similar pattern to the individual income in that there are possible outliers with extremely high values. The conditional mean regression may not reflect the behavior of gross revenue accurately, so we prefer using the conditional quantile regression to examine the relationship between online rating scores and movie box office revenues. The average IMDb rating has a distribution that centers between 6 and 7. The probability density declines as the average rating deviates from the center. Table 7 shows the summary statistics of the discrete variables. For MPAA status, “PG-13” is much more frequent than “PG”. For genre, almost half of the movies are “Action”. The rest three categories have similar shares.

Table 6: Summary statistics of the continuous variables

	Mean	Std	Min	Max
Gross revenue (in million dollars)	91.51	89.41	7.02	623.36
Average IMDb rating	6.32	0.95	3.1	8.5

Table 7: Summary statistics of the discrete variables

	MPAA Status			
Possible Value	PG	PG-13		
Frequency	0.28	0.72		
	Genre			
Possible Value	Action	Animation	Comedy	Drama
Frequency	0.43	0.17	0.22	0.18

In the quantile regression model, the dependent variable y is the gross revenue, the continuous explanatory variable x is the average IMDb rating, and the discrete explanatory variables z_1 and z_2 are genre and MPAA status, respectively. We estimate the quantiles of y conditional on x, z_1 and z_2 with our proposed method at the quantile levels $10\%, 25\%, 50\%, 75\%$ and 90%. We also compare the performance of our method with that of the inverse-CDF method.

5.2 Empirical results

The left panel of Figure 3 shows the estimated conditional quantile curves of gross revenue conditional on the average IMDb rating, with genre fixed at $z_1 =$ “action”, and MPAA status fixed at $z_2 =$ “PG-13”. Using the proposed check-function-based method which smoothes both the continuous and discrete covariates, we plot five curves from top to bottom corresponding to the 90%, 75%, 50%, 25% and 10% quantiles, respectively. The figure shows two patterns: (i) as the average IMDb rating increases, all five quantiles increase; and (ii) the distribution of the gross revenue is stretched, i.e., the 10% quantile increases slowly whereas the 90% quantile increases fast. The right panel of Figure 3 shows the estimated conditional quantile curves by the inverse-CDF method. The estimated curves in the right panel are quite similar to those in the left panel, which is not surprising since both estimators are consistent with the same convergence rate. It

Table 8: The estimated smoothing parameters in the empirical study

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			Smoothing parameter for x		
Check Function	0.313	0.290	0.416	0.377	0.642
Inverse-CDF			0.399 (uniform over quantiles)		
			Smoothing parameter for z_1		
Check Function	0.534	0.398	0.505	0.322	0.313
Inverse-CDF			0.650 (uniform over quantiles)		
			Smoothing parameter for z_2		
Check Function	0.502	0.191	0.352	0.158	0.165
Inverse-CDF			0.593 (uniform over quantiles)		

is worth mentioning again that our check-function-based method has computational advantage against the inverse-CDF method. Figures 4 reports the point-wise 95% confidence interval for the conditional median function ($\tau = 50\%$). We also report the 95% confidence intervals for the 10%, 25%, 75% and 90% conditional quantile functions which are available in Appendix E of the supplementary document. These confidence intervals are constructed via the standard (resampling with replacement) bootstrap method with undersmoothed bandwidth parameters. We then use the bootstrap method introduced in Section 4.2 to construct the uniform confidence intervals for the conditional quantile functions at two evaluation points: (1) $x = 6, z_1 =$"action"$, z_2 =$"PG-13" and (2) $x = 7, z_1 =$"action"$, z_2 =$"PG-13". Figure 5 shows the two uniform confidence intervals. The uniform confidence intervals are wider than the corresponding point-wise confidence intervals to accommodate extreme quantiles.

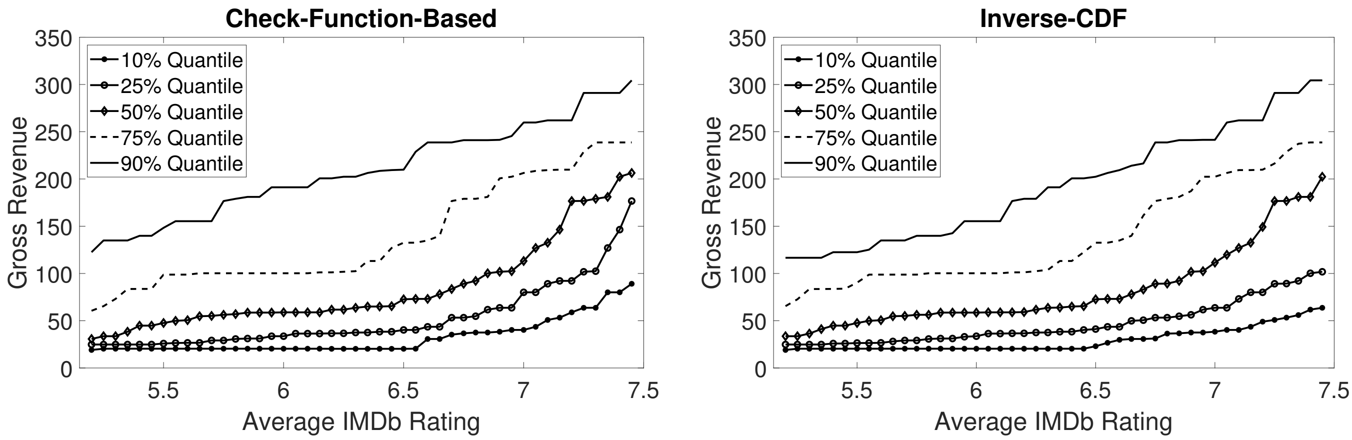

Figure 3: Estimated conditional quantile curves using the check-function-based and inverse-CDF methods

Next, we report values of all the selected optimal smoothing parameters in Table 8. The top block of the table corresponds to the selected smoothing parameters for the continuous covariate,

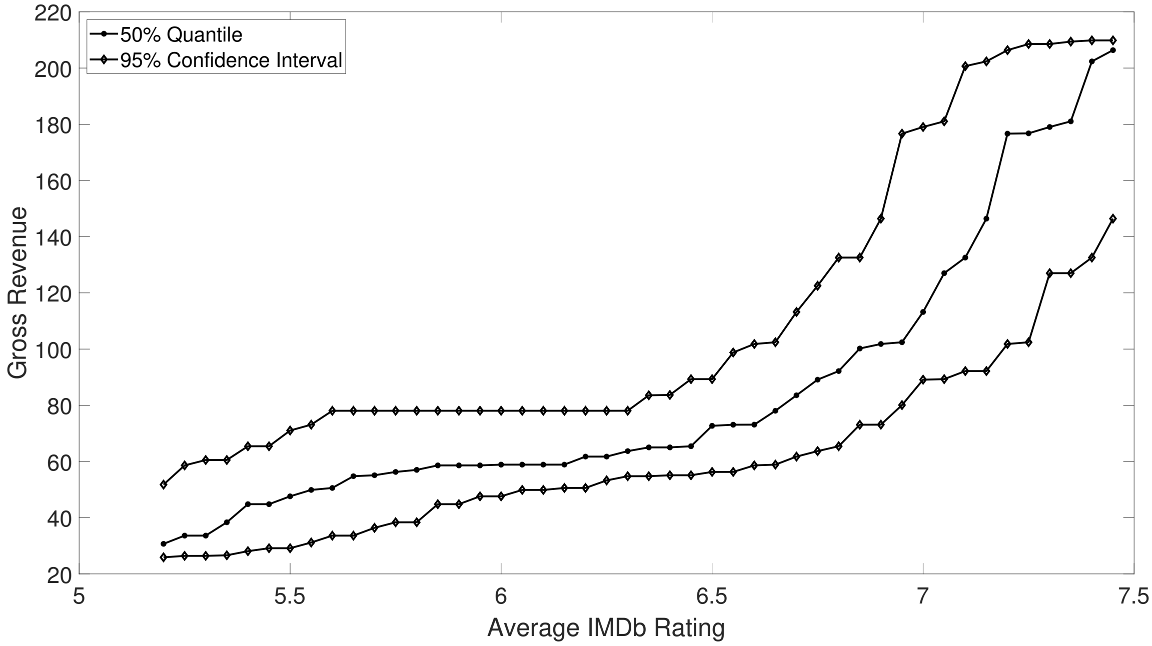

Figure 4: The 95% point-wise confidence interval of the conditional median function

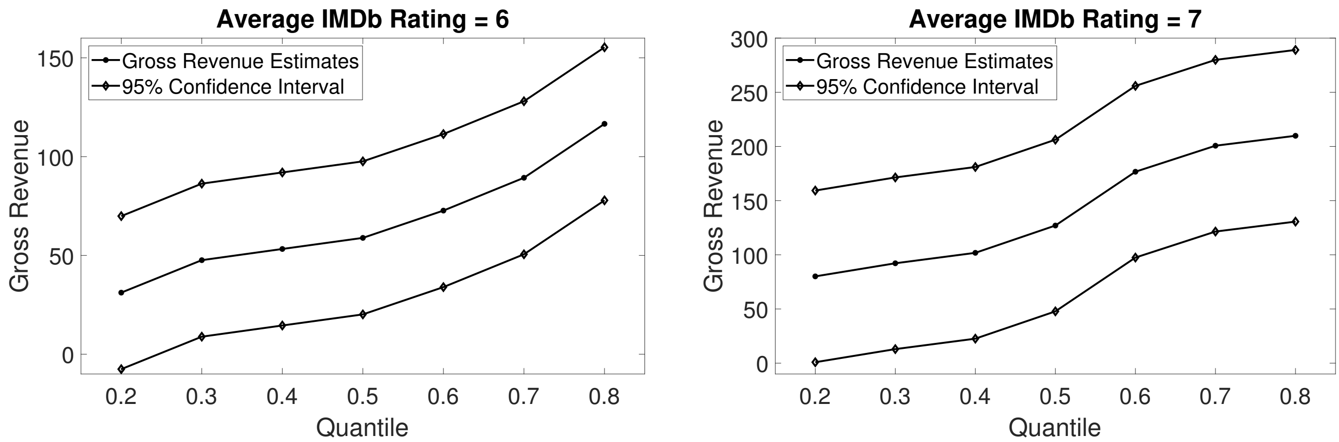

Figure 5: The 95% uniform confidence intervals over quantiles with z_1 ="action" and z_2 ="PG-13"

and the lower two blocks of the table correspond to the selected smoothing parameters for the two discrete covariates, respectively. The inverse-CDF method has the same smoothing parameters across all quantiles, whereas the check-function-based method selects quantile-specific smoothing parameters.

To further compare the performance of our method with that of inverse-CDF method, we assess their out-of-sample prediction performance. We randomly split the sample into estimation sample (90% of the sample size = 204) and validation sample (10% of the sample size = 23). We use the estimation sample to forecast the conditional quantiles in the validation sample. The out-of-sample prediction accuracy is measured by the average check-function value (over the validation sample), which is defined as

$$\frac{1}{23}\sum_{i=1}^{23}\rho_\tau\left(y_i^o-\widehat{Q}_\tau(x_i^o,z_{i1}^o,z_{i2}^o)\right)\qquad \text{with } \tau=0.10,0.25,0.50,0.75,0.90,$$

where y_i^o, x_i^o, z_{i1}^o and z_{i2}^o are the i-th data point in the validation sample, and $\widehat{Q}_\tau(\cdot)$ is the estimated conditional τ-quantile function either by the check-function-based method or the inverse-CDF method using the estimation sample. We repeat the above process for 1000 times. Table 9 reports the means (over the 1000 replications) of the average check function value, from which we find that the check-function-based method outperforms the inverse-CDF method in the out-of-sample quantile forecasting for all the five levels especially at the right-tail quantile level, i.e., at the 90 percentile the loss function value of the former is less than one third of the latter.

Table 9: Comparison of out-of-sample forecasting performance

	$\tau=0.10$	$\tau=0.25$	$\tau=0.50$	$\tau=0.75$	$\tau=0.90$
Check Function	8.09	16.12	25.05	23.54	16.11
Inverse-CDF	9.97	18.05	27.78	36.60	49.80

6 Conclusion

In this paper, we study the nonparametric conditional quantile function estimation when the covariates include both continuous and discrete components. We combine the quantile check function and the local linear smoothing technique with the mixed continuous and discrete kernel function to directly estimate the conditional quantile function, making it substantially different from the method proposed in Li, Lin and Racine (2013) which first estimates the CDF nonparametrically and then inverts the estimated CDF to obtain the quantile estimation. One advantage of our method is that it selects quantile (τ) dependent optimal smoothing parameters, while the inverse-CDF method does not have this flexibility. The smoothing parameters involved in the proposed nonparametric quantile regression estimation are determined by a completely data-driven RCV criterion. We derive the asymptotic optimality property of the selected smoothing parameters, and establish the point-wise asymptotic normal distribution theory and uniform consistency for the local linear quantile regression estimator using the data-dependent smoothing parameters determined by the RCV criterion. The simulation results show that our method has better small-sample performance than the naive local linear quantile estimation without smoothing the discrete regressors and the inverse-CDF method. In addition, the selected smoothing parameters are very close to the theoretically optimal ones in finite samples. Furthermore, the proposed nonparametric quantile estimation method is used to study the relationship between box office revenues and online rating scores. Our empirical results suggest that as the average online rating score increases, the examined quantiles of box office revenue increases, and the higher quantile regression functions increases faster than the lower ones.

We derive the asymptotic theory of the local linear quantile regression estimator for fixed $0 < \tau < 1$, see Theorems 3.1 and 3.2. A possible extension is to combine the local linear smoothing method introduced in Section 2 with the linear interpolation technique (Qu and Yoon, 2015) to obtain quantile regression estimates for all $\tau \in \mathcal{T} \equiv [\underline{\tau}, \overline{\tau}]$, where $0 < \underline{\tau} < \overline{\tau} < 1$. Specifically, consider a set of grid points (with equal distance) arranged in the increasing order: $\{\tau_0, \tau_1, \cdots, \tau_{r_n}\}$ with $\tau_0 = \underline{\tau}$ and $\tau_{r_n} = \overline{\tau}$, where r_n is a positive integer which may be divergent to infinity as n increases. For each given τ_k, $0 \leqslant k \leqslant r_n$, we minimize the loss function in (2.2) with respect to α, and obtain the local linear estimate of the τ_k-quantile regression function $\mathsf{Q}_{\tau_k}(x_0, z_0)$, denoted by $\widehat{\mathsf{Q}}_{\tau_k}(x_0, z_0; h_{\tau_k}, \lambda_{\tau_k})$, where we make the dependence of h_τ and λ_τ on τ explicitly. Then, we apply the technique of linear interpolation between $\widehat{\mathsf{Q}}_{\tau_k}(x_0, z_0; h_{\tau_k}, \lambda_{\tau_k})$, i.e.,

$$\widehat{\mathsf{Q}}_{\tau}^{\diamond}(x_0, z_0) = \frac{\tau_{k+1} - \tau}{\tau_{k+1} - \tau_k} \cdot \widehat{\mathsf{Q}}_{\tau_k}(x_0, z_0; h_{\tau_k}, \lambda_{\tau_k}) + \frac{\tau - \tau_k}{\tau_{k+1} - \tau_k} \cdot \widehat{\mathsf{Q}}_{\tau_{k+1}}(x_0, z_0; h_{\tau_{k+1}}, \lambda_{\tau_{k+1}}) \tag{6.1}$$

for $\tau_k \leqslant \tau \leqslant \tau_{k+1}$. To ensure monotonicity of the estimated quantile regression curves (with respect to τ), we may further apply the re-arrangement technique to $\widehat{\mathsf{Q}}_{\tau}^{\diamond}(x_0, z_0)$ defined in (6.1), see, for example, Chernozhukov, Fernández-Val and Galichon (2010). It would be an interesting future topic to derive the uniform convergence result for $\widehat{\mathsf{Q}}_{\tau}^{\diamond}(x_0, z_0)$ using the RCV selected smoothing parameters.

While the proposed RCV method to select the smoothing parameters has computational advantage over the conventional "leave-one-out" CV method which is commonly used in practice, the simulation results reported in Section 4 show that the latter slightly outperforms the RCV method in finite samples (see Table 1). Therefore, a further extension of the theoretical justification of this paper to cover the case of "leave-one-out" CV method (possibly with a different mathematical technique) would be another important future research topic.

Acknowledgements

The authors would like to thank a Co-Editor, an Associate Editor and two anonymous reviewers for their insightful comments that substantially improve an early version of the paper.

Appendix A Assumptions

Throughout the appendix, we write $a_n \sim b_n$ and $a_n \overset{P}{\sim} b_n$ to denote that $a_n = b_n(1 + o(1))$ and $a_n = b_n(1 + o_P(1))$, respectively; let $a_n \asymp b_n$ denote that $a_n = O(b_n)$ and $b_n = O(a_n)$ hold jointly. Below we give some regularity conditions which are used to prove the main theoretical results of the paper.

ASSUMPTION 1. *The kernel function* $K(\cdot)$ *is a Lipschitz continuous and symmetric probability density function with a compact support* $[-1, 1]$.

ASSUMPTION 2. *The sequence of* $\{(Y_i, X_i, Z_i)\}$ *is composed of independent and identically distributed (i.i.d.) random vectors.*

ASSUMPTION 3. *The conditional density function of* $e_i \equiv Y_i - Q_\tau(X_i, Z_i)$ *for given* $X_i = x$ *and* $Z_i = z$, $f_e(\cdot|x, z)$, *exists and has continuous first-order derivative at point zero. Furthermore,* $f_e(0|x, z)$ *is continuous with respect to* x *and bounded away from infinity and zero uniformly for* $(x, z) \in \mathcal{D}_x \times \mathcal{D}_z$. *The conditional CDF of* e_i *for given* $X_i = x$ *and* $Z_i = z$, $F_e(\cdot|x, z)$, *is continuous with respect to* x. *The joint probability density function of* X_i *and* Z_i, $f_{xz}(\cdot, \cdot)$, *is bounded away from infinity and zero over* $\mathcal{D}_x \times \mathcal{D}_z$.

ASSUMPTION 4. *The quantile regression function* $Q_\tau(\cdot, z)$ *is twice continuously differentiable on* $\mathcal{D}_x$. *The weight function* $M(\cdot)$ *is continuous on* $\mathcal{D}_x(\epsilon)$, *and* $M(x) = 0$ *for* $x \notin \mathcal{D}_x(\epsilon)$, *where* $\mathcal{D}_x(\epsilon) = \{(x_1, \cdots, x_p) : \underline{c}_j + \epsilon \leqslant x_j \leqslant \bar{c}_j - \epsilon,\ j = 1, \cdots, p\}$ *with* ϵ *being a very small positive constant.*

ASSUMPTION 5. *(i) For* $(h, \lambda) \in \mathcal{H}_m$ *with* $h = (h_1, \cdots, h_p)^\intercal$ *and* $\lambda = (\lambda_1, \cdots, \lambda_q)^\intercal$, $h_j \to 0$ *and* $\lambda_j \to 0$ *as* $n \to \infty$.

(ii) Letting $H = \prod_{j=1}^p h_j$, *for* $(h, \lambda) \in \mathcal{H}_m$, $\sum_{j=1}^p h_j^2 + \sum_{j=1}^q \lambda_j = o\left(\left(\frac{1}{n^{1/2}H}\right) \wedge \left(\frac{\log n}{nH}\right)^{1/3}\right)$, *and* $\frac{\log n}{nH} = o\left(\sum_{j=1}^p h_j^2 + \sum_{j=1}^q \lambda_j\right)$.

(iii) The number of grid points in $\mathcal{H}_m$, L_m, *diverges at a polynomial rate of* m, *and in addition,* $\gamma_m = o(m^{-2/(4+p)})$.

Assumption 1 imposes some mild conditions on the continuous kernel function in the nonparametric kernel-based smoothing, and several commonly-used kernel functions such as the uniform kernel and the Epanechnikov kernel satisfy these conditions (c.f., Fan and Gijbels, 1996). In Assumption 2, we impose the *i.i.d.* condition on the observations, a commonly-used setting in the literature on nonparametric estimation with mixed discrete and continuous data (c.f., Li and Racine, 2004). The asymptotic results of this paper can be generalized to the general stationary and weakly dependent processes at the cost of more lengthy proofs. There is no moment condition on e_i to estimate the conditional quantile regression, indicating that the error distribution is allowed to have heavy tails. Assumptions 3 and 4 give some smoothness conditions on the (conditional) density functions, the weight function and the quantile regression function, which are often imposed in studying asymptotic behavior of nonparametric kernel-based estimators. Assumption 5 imposes some standard restrictions on the smoothing parameters. Assumption 5(ii) is crucial to guarantee that the second term on the right hand side of (2.7) is the asymptotic leading term to determine the optimal bandwidth vectors and Assumption 5(iii) indicates that the grid points are sufficiently dense in the set $\mathcal{H}_m$. The theoretically optimal bandwidths h^0 and λ^0 satisfy Assumption 5(ii), and by Assumption 5(iii), there exists a grid point in $\mathcal{H}_m$ which is sufficiently close to (h^0, λ^0).

Appendix B Proofs of the main results

In this appendix, we give the detailed proofs of the main theoretical results in Sections 2 and 3. In order to simplify the presentation, we first introduce some notation. For $x = (x_1, \cdots, x_p)^\intercal \in \mathcal{D}_x$ and $z = (z_1, \cdots, z_q)^\intercal \in \mathcal{D}_z$, we let

$$\Delta_{ni}(\alpha, \beta; x, z) = \frac{1}{\sqrt{nH}} \left[u_n(\alpha; x, z) + \sum_{j=1}^{p} v_{n,j}(\beta_j; x, z) \left(\frac{X_{i,j} - x_j}{h_j} \right) \right]$$

with $u_n(\alpha; x, z) = \sqrt{nH}\,[\alpha - Q_\tau(x, z)]$ and $v_{n,j}(\beta_j; x, z) = \sqrt{nH} h_j \left[\beta_j - Q'_{\tau,j}(x, z) \right]$, and $b_i(x, z) = Q_\tau(X_i, Z_i) - Q_\tau(x, z) - \sum_{j=1}^{p} Q'_{\tau,j}(x, z)(X_{i,j} - x_j)$. With the above notation, it is easy to see that

$$Y_i - \alpha - (X_i - x)^\intercal \beta = e_i - \Delta_{ni}(\alpha, \beta; x, z) + b_i(x, z).$$

B.1 Proof of Theorem 2.1

Let $\zeta_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) = \widehat{Q}_{\mathcal{M}_1}(X_j, Z_j; h, \lambda) - Q_\tau(X_j, Z_j)$ with $\widehat{Q}_{\mathcal{M}_1}(X_j, Z_j; h, \lambda)$ defined as in (2.3). Then

$$\begin{aligned} \mathsf{CV}(h, \lambda) &= \frac{1}{n-m} \sum_{j=m+1}^{n} \rho_\tau \left(e_j + Q_\tau(X_j, Z_j) - \widehat{Q}_{\mathcal{M}_1}(X_j, Z_j; h, \lambda) \right) M(X_j) \\ &\equiv \mathsf{CV}_* + \mathsf{CV}_2(h, \lambda), \end{aligned} \tag{B.1}$$

where CV_* (defined in Theorem 2.1) does not rely on the smoothing parameters h and λ (so would not play any role in choosing the optimal bandwidth vectors), and

$$\mathsf{CV}_2(h, \lambda) = \frac{1}{n-m} \sum_{j=m+1}^{n} \left[\rho_\tau(e_j - \zeta_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)) - \rho_\tau(e_j) \right] M(X_j).$$

We only need to study $\mathsf{CV}_2(h, \lambda)$. Using the following identity result (e.g., Knight, 1998):

$$\rho_\tau(u - v) - \rho_\tau(u) = v \left(\mathrm{I}\{u \leqslant 0\} - \tau \right) + \int_0^v \left(\mathrm{I}\{u \leqslant w\} - \mathrm{I}\{u \leqslant 0\} \right) dw, \tag{B.2}$$

we have

$$\rho_\tau\big(e_j - \zeta_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)\big) - \rho_\tau(e_j) = \zeta_{\mathcal{M}_1,h,\lambda}(X_i, Z_i)\left(\mathsf{I}\{e_j \leqslant 0\} - \tau\right) + \int_0^{\zeta_{\mathcal{M}_1,h,\lambda}(X_j,Z_j)} \left(\mathsf{I}\{e_j \leqslant w\} - \mathsf{I}\{e_j \leqslant 0\}\right) dw.$$

Define

$$\begin{aligned}
\mathsf{CV}_{21}(h,\lambda) &= \frac{1}{n-m}\sum_{j=m+1}^{n} \zeta_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)\left(\mathsf{I}\{e_j \leqslant 0\} - \tau\right) M(X_j), \\
\mathsf{CV}_{22}(h,\lambda) &= \frac{1}{n-m}\sum_{j=m+1}^{n} M(X_j)\int_0^{\zeta_{\mathcal{M}_1,h,\lambda}(X_j,Z_j)} \left(\mathsf{I}\{e_j \leqslant w\} - \mathsf{I}\{e_j \leqslant 0\}\right) dw.
\end{aligned}$$

Then, we readily have that $\mathsf{CV}_2(h,\lambda) = \mathsf{CV}_{21}(h,\lambda) + \mathsf{CV}_{22}(h,\lambda)$.

We first derive the asymptotic order for $\mathsf{CV}_{22}(h,\lambda)$ and show that it is the asymptotic leading term of $\mathsf{CV}_2(h,\lambda)$. As $m \asymp n$, by Propositions C.1 and C.2 in Appendix C of the supplemental document and Assumption 5(ii), we have

$$\begin{aligned}
\mathsf{CV}_{22}(h,\lambda) &= \mathsf{CV}^*_{22}(h,\lambda) + O_P\left(\chi_1^{5/2}(h) + \chi_1(h)\chi_2^{3/2}(h,\lambda) + \chi_3(h,\lambda)\right) \\
&= \mathsf{CV}^*_{22}(h,\lambda) + o_P\left(\chi_2^2(h,\lambda) + \frac{1}{mH}\right) \qquad \text{(B.3)}
\end{aligned}$$

uniformly over $(h,\lambda) \in \mathcal{H}_m$, where $\chi_1(h) = (\log n)^{1/2}(nH)^{-1/2}$, $\chi_2(h,\lambda) = \sum_{j=1}^{p} h_j^2 + \sum_{j=1}^{q}\lambda_j$, $\chi_3(h,\lambda) = (\log n/n)^{1/2}\left[\chi_1(h) + \chi_2(h,\lambda)\right]$, and

$$\mathsf{CV}^*_{22}(h,\lambda) = \frac{1}{2(n-m)}\sum_{j=m+1}^{n}\left[\zeta^*_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)\right]^2 f_e(0|X_j, Z_j) M(X_j) \qquad \text{(B.4)}$$

in which $\zeta^*_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) = f^{-1}(X_j, Z_j) W_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)$, $f(x,z) = f_{xz}(x,z) f_e(0|x,z)$, and

$$W_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) = m^{-1}\sum_{i=1}^{m}\eta_i(X_j, Z_j)\mathbf{K}_h(X_i - X_j)\mathbf{\Lambda}_\lambda(Z_i, Z_j)$$

with $\eta_i(x,z) = \tau - \mathsf{I}\{e_i \leqslant -b_i(x,z)\}$.

Letting $\tilde\eta_i = \tau - \mathsf{I}(e_i \leqslant 0)$, we rewrite $W_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)$ as

$$\begin{aligned}
W_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) &= \frac{1}{m}\sum_{i=1}^{m}\left[\eta_i(X_j, Z_j) - \tilde\eta_i\right]\mathbf{K}_h(X_i - X_j)\mathbf{\Lambda}_\lambda(Z_i, Z_j) + \frac{1}{m}\sum_{i=1}^{m}\tilde\eta_i\mathbf{K}_h(X_i - X_j)\mathbf{\Lambda}_\lambda(Z_i, Z_j) \\
&\equiv \overline{W}_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) + \widetilde{W}_{\mathcal{M}_1,h,\lambda}(X_j, Z_j). \qquad \text{(B.5)}
\end{aligned}$$

Define $B_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) = f^{-1}(X_j, Z_j)\overline{W}_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)$ and $V_{\mathcal{M}_1,h,\lambda}(X_j, Z_j) = f^{-1}(X_j, Z_j)\widetilde{W}_{\mathcal{M}_1,h,\lambda}(X_j, Z_j)$, by

(B.4) and (B.5), we readily have that

$$\begin{aligned}
\mathsf{CV}_{22}^{*}(h,\lambda) &= \frac{1}{2(n-m)}\sum_{j=m+1}^{n} B_{\mathcal{M}_1,h,\lambda}^{2}(X_j,Z_j)f_e(0|X_j,Z_j)M(X_j)+\\
&\quad \frac{1}{2(n-m)}\sum_{j=m+1}^{n} V_{\mathcal{M}_1,h,\lambda}^{2}(X_j,Z_j)f_e(0|X_j,Z_j)M(X_j)+\\
&\quad \frac{1}{n-m}\sum_{j=m+1}^{n} B_{\mathcal{M}_1,h,\lambda}(X_j,Z_j)V_{\mathcal{M}_1,h,\lambda}(X_j,Z_j)f_e(0|X_j,Z_j)M(X_j)\\
&\equiv \mathsf{CV}_{22,B}^{*}(h,\lambda)+\mathsf{CV}_{22,V}^{*}(h,\lambda)+\mathsf{CV}_{22,BV}^{*}(h,\lambda). \qquad \text{(B.6)}
\end{aligned}$$

We consider the three terms on the right hand side of (B.6) separately. We start with $\mathsf{CV}_{22,B}^{*}(h,\lambda)$. Note that

$$\begin{aligned}
&\frac{1}{2(n-m)}\sum_{j=m+1}^{n} B_{\mathcal{M}_1,h,\lambda}^{2}(X_j,Z_j)f_e(0|X_j,Z_j)M(X_j)\\
=&\ \frac{1}{2m^2(n-m)}\sum_{j=m+1}^{n}\frac{f_e(0|X_j,Z_j)M(X_j)}{f^2(X_j,Z_j)}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\left[\eta_{i_1}(X_j,Z_j)-\tilde{\eta}_{i_1}\right]\left[\eta_{i_2}(X_j,Z_j)-\tilde{\eta}_{i_2}\right]\\
&\ \mathbf{K}_h(X_{i_1}-X_j)\mathbf{\Lambda}_\lambda(Z_{i_1},Z_j)\mathbf{K}_h(X_{i_2}-X_j)\mathbf{\Lambda}_\lambda(Z_{i_2},Z_j)\\
=&\ \frac{1}{2m^2(n-m)}\sum_{j=m+1}^{n}\frac{f_e(0|X_j,Z_j)M(X_j)}{f^2(X_j,Z_j)}\sum_{i=1}^{m}\left[\eta_i(X_j,Z_j)-\tilde{\eta}_i\right]^2\mathbf{K}_h^2(X_i-X_j)\mathbf{\Lambda}_\lambda^2(Z_i,Z_j)+\\
&\ \frac{1}{2m^2(n-m)}\sum_{j=m+1}^{n}\frac{f_e(0|X_j,Z_j)M(X_j)}{f^2(X_j,Z_j)}\sum_{i_1=1}^{m}\sum_{i_2=1,\neq i_1}^{m}\left[\eta_{i_1}(X_j,Z_j)-\tilde{\eta}_{i_1}\right]\left[\eta_{i_2}(X_j,Z_j)-\tilde{\eta}_{i_2}\right]\\
&\ \mathbf{K}_h(X_{i_1}-X_j)\mathbf{\Lambda}_\lambda(Z_{i_1},Z_j)\mathbf{K}_h(X_{i_2}-X_j)\mathbf{\Lambda}_\lambda(Z_{i_2},Z_j)\\
\equiv&\ \mathsf{CV}_{22,B,1}^{*}(h,\lambda)+\mathsf{CV}_{22,B,2}^{*}(h,\lambda). \qquad \text{(B.7)}
\end{aligned}$$

By (B.7) and Proposition C.3, we have

$$\mathsf{CV}_{22,B}^{*}(h,\lambda)=\frac{1}{2(n-m)}\sum_{j=m+1}^{n} b^2(X_j,Z_j;h,\lambda)f_e(0|X_j,Z_j)M(X_j)+o_P\left(\chi_2^2(h,\lambda)+\frac{1}{mH}\right) \qquad \text{(B.8)}$$

uniformly over $(h,\lambda)\in\mathcal{H}_m$. Next, we consider $\mathsf{CV}_{22,V}^{*}(h,\lambda)$ on the right hand side of (B.6). Note that

$$\begin{aligned}
&\frac{1}{2(n-m)}\sum_{j=m+1}^{n} V_{\mathcal{M}_1,h,\lambda}^{2}(X_j,Z_j)f_e(0|X_j,Z_j)M(X_j)\\
=&\ \frac{1}{2m^2(n-m)}\sum_{j=m+1}^{n}\frac{f_e(0|X_j,Z_j)M(X_j)}{f^2(X_j,Z_j)}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\tilde{\eta}_{i_1}\tilde{\eta}_{i_2}\mathbf{K}_h(X_{i_1}-X_j)\mathbf{\Lambda}_\lambda(Z_{i_1},Z_j)\\
&\ \mathbf{K}_h(X_{i_2}-X_j)\mathbf{\Lambda}_\lambda(Z_{i_2},Z_j)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2m^2(n-m)} \sum_{j=m+1}^{n} \frac{f_e(0|X_j,Z_j)M(X_j)}{f^2(X_j,Z_j)} \sum_{i=1}^{m} \tilde{\eta}_i^2 \mathbf{K}_h^2(X_i - X_j)\boldsymbol{\Lambda}_\lambda^2(Z_i,Z_j) + \\
&\quad \frac{1}{2m^2(n-m)} \sum_{j=m+1}^{n} \frac{f_e(0|X_j,Z_j)M(X_j)}{f^2(X_j,Z_j)} \sum_{i_1=1}^{m} \sum_{i_2=1,\neq i_1}^{m} \tilde{\eta}_{i_1}\tilde{\eta}_{i_2}\mathbf{K}_h(X_{i_1} - X_j)\boldsymbol{\Lambda}_\lambda(Z_{i_1},Z_j) \\
&\quad \mathbf{K}_h(X_{i_2} - X_j)\boldsymbol{\Lambda}_\lambda(Z_{i_2},Z_j) \\
&\equiv \mathsf{CV}^*_{22,V,1}(h,\lambda) + \mathsf{CV}^*_{22,V,2}(h,\lambda). \qquad \text{(B.9)}
\end{aligned}
$$

By (B.9) and Proposition C.4, we may show that

$$
\mathsf{CV}^*_{22,V}(h,\lambda) = \frac{1}{2(n-m)} \sum_{j=m+1}^{n} \sigma_m^2(X_j,Z_j;h,\lambda) f_e(0|X_j,Z_j)M(X_j) + o_P\left(\frac{1}{mH}\right) \qquad \text{(B.10)}
$$

uniformly over $(h,\lambda) \in \mathcal{H}_m$. By Proposition C.5, we have

$$
\mathsf{CV}^*_{22,BV}(h,\lambda) = o_P\left(\chi_2^2(h,\lambda) + \frac{1}{mH}\right) \qquad \text{(B.11)}
$$

uniformly over $(h,\lambda) \in \mathcal{H}_m$. Then, with (B.3), (B.6), (B.8), (B.10) and (B.11), we can prove that

$$
\begin{aligned}
\mathsf{CV}_{22}(h,\lambda) &= \frac{1}{2(n-m)} \sum_{j=m+1}^{n} \left[b^2(X_j,Z_j;h,\lambda) + \sigma_m^2(X_j,Z_j;h,\lambda)\right] f_e(0|X_j,Z_j)M(X_j) + \text{s.o.} \\
&= \frac{1}{2}\mathsf{E}\left\{\left[b_m^2(X_i,Z_i;h,\lambda) + \sigma_m^2(X_i,Z_i;h)\right] M(X_i) f_e(0|X_i,Z_i)\right\} + \text{s.o.} \qquad \text{(B.12)}
\end{aligned}
$$

uniformly over $(h,\lambda) \in \mathcal{H}_m$.

Finally, following the proof of Proposition C.2, we may show that

$$
\mathsf{CV}_{21}(h,\lambda) = O_P\left(\chi_3(h,\lambda)\right) = o_P\left(\chi_2^2(h,\lambda) + \frac{1}{mH}\right) \qquad \text{(B.13)}
$$

uniformly over $(h,\lambda) \in \mathcal{H}_m$, where $\chi_3(h,\lambda)$ is defined as in (B.3). By (B.12) and (B.13), we complete the proof of (2.7) in Theorem 2.1. ■

B.2 Proof of Theorem 2.2

Let $h^0_{m,j} = a^0_j \cdot m^{-1/(4+p)}$ for $j = 1, \cdots, p$, and $\lambda^0_{m,k} = d^0_k \cdot m^{-2/(4+p)}$ for $k = 1, \cdots, q$. Then, we readily have that $h^0_j = h^0_{m,j} \cdot \left(\frac{m}{n}\right)^{1/(4+p)}$ for $j = 1, \cdots, p$, and $\lambda^0_k = \lambda^0_{m,k} \cdot \left(\frac{m}{n}\right)^{2/(4+p)}$ for $k = 1, \cdots, q$. By (2.6), we only need to prove that $\widehat{h}_{m,j} - h^0_{m,j} = o_P(h^0_{m,j})$ for $j = 1, 2, \cdots, p$, and $\widehat{\lambda}_{m,k} - \lambda^0_{m,k} = o_P(\lambda^0_{m,k})$ for $k = 1, 2, \cdots, q$, where $\widehat{h}_{m,j}$ is the j-th element of $\widehat{h}_m$ and $\widehat{\lambda}_{m,k}$ is the k-th element of $\widehat{\lambda}_m$.

From the asymptotic representation of the CV function in Theorem 2.1, the term CV_* does not rely on the bandwidth vectors h and λ. The second term on the right hand side of (2.7) is the asymptotic leading term for the optimal bandwidth selection. Therefore, the optimal bandwidth vectors $\widehat{h}_m$ and $\widehat{\lambda}_m$ satisfy that $\widehat{h}_m = h_m^\diamond + o_P(h_m^\diamond)$ and $\widehat{\lambda}_m = \lambda_m^\diamond + o_P(\lambda_m^\diamond)$, where $h_m^\diamond$ and $\lambda_m^\diamond$ are the bandwidth vectors that minimize the second term on the right hand side of (2.7) uniformly over $(h, \lambda) \in \mathcal{H}_m$. Note that the latter equals to $\frac{1}{2} \cdot \mathsf{MSE}^{L}_{\mathcal{M}_2}(h, \lambda)$, see the definition in (2.9). Under the assumption that there exist uniquely determined $a_j^0 > 0$, $j = 1, \cdots, p$, and $d_k^0 \geqslant 0$, $k = 1, \cdots, q$, that minimize $g(a, d)$ defined in (2.10), using Assumption 5(iii), we must have that $h_m^\diamond = h_m^0 + o(h_m^0)$ and $\lambda_m^\diamond = \lambda_m^0 + o(\lambda_m^0)$, where $h_m^0 = \left(h_{m,1}^0, \cdots, h_{m,p}^0\right)^\intercal$ and $\lambda_m^0 = \left(\lambda_{m,1}^0, \cdots, \lambda_{m,q}^0\right)^\intercal$. This completes the proof of Theorem 2.2. ■

B.3 Proof of Theorem 3.1

Let $\zeta(x_0, z_0; h, \lambda) = (nH)^{1/2}\left[\widehat{Q}_\tau(x_0, z_0; h, \lambda) - Q_\tau(x_0, z_0)\right]$, where we make the dependence of the local linear quantile estimators on the bandwidths h and λ explicit. We first derive the asymptotic normality for $\zeta(x_0, z_0; h, \lambda)$ when the bandwidth vectors h and λ are chosen as the deterministic optimal bandwidths defined in Theorem 2.2, i.e., $h = h^0$, $\lambda = \lambda^0$ and $H^0 = \prod_{k=1}^p h_k^0$. By Proposition C.1, we have the following Bahadur representation:

$$\zeta(x_0, z_0; h^0, \lambda^0) = \frac{W_n(x_0, z_0; h^0, \lambda^0)}{f(x_0, z_0)}(1 + o_P(1)), \tag{B.14}$$

where $f(x_0, z_0) = f_{xz}(x_0, z_0) f_e(0|x_0, z_0)$ and

$$W_n(x_0, z_0; h^0, \lambda^0) = \frac{H^0}{\sqrt{nH^0}} \sum_{i=1}^n \eta_i(x_0, z_0) \mathbf{K}_{h^0}(X_i - x_0) \mathbf{\Lambda}_{\lambda^0}(Z_i, z_0).$$

Thus, to establish the asymptotic distribution theory of $\zeta(x_0, z_0; h^0, \lambda^0)$, we only need to derive the limiting distribution of $W_n(x_0, z_0; h^0, \lambda^0)$.

Let $\widetilde{W}_n(x_0, z_0; h^0, \lambda^0)$ be defined as $W_n(x_0, z_0; h^0, \lambda^0)$ but with $\eta_i(x_0, z_0)$ replaced by $\widetilde{\eta}_i = \tau - \mathsf{I}\{e_i \leqslant 0\}$. Then, we have

$$\begin{aligned}
& W_n(x_0, z_0; h^0, \lambda^0) - \mathsf{E}\left[W_n(x_0, z_0; h^0, \lambda^0)\right] \\
= \; & \widetilde{W}_n(x_0, z_0; h^0, \lambda^0) - \mathsf{E}\left[\widetilde{W}_n(x_0, z_0; h^0, \lambda^0)\right] + W_n(x_0, z_0; h^0, \lambda^0) - \widetilde{W}_n(x_0, z_0; h^0, \lambda^0) - \\
& \mathsf{E}\left[W_n(x_0, z_0; h^0, \lambda^0) - \widetilde{W}_n(x_0, z_0; h^0, \lambda^0)\right].
\end{aligned} \tag{B.15}$$

Note that

$$\begin{aligned}
& \mathsf{Var}\left[W_n(x_0, z_0; h^0, \lambda^0) - \widetilde{W}_n(x_0, z_0; h^0, \lambda^0)\right] \\
\leqslant \; & \mathsf{E}\left[\left|W_n(x_0, z_0; h^0, \lambda^0) - \widetilde{W}_n(x_0, z_0; h^0, \lambda^0)\right|^2\right] \\
= \; & \mathsf{E}\left\{\mathsf{E}\left[\left|W_n(x_0, z_0; h^0, \lambda^0) - \widetilde{W}_n(x_0, z_0; h^0, \lambda^0)\right|^2 \Big| (\mathcal{X}_n, \mathcal{Z}_n)\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= O\left(\frac{H^0}{n}\sum_{i=1}^{n}\mathsf{E}\left\{\mathbf{K}^2_{h^0}(X_i - x_0)\boldsymbol{\Lambda}^2_{\lambda^0}(Z_i, z_0)\mathsf{E}\left[(\eta_i(x_0, z_0) - \widetilde{\eta}_i)^2\Big|(\mathcal{X}_n, \mathcal{Z}_n)\right]\right\}\right) \\
&= o\left(\frac{1}{nH^0}\sum_{i=1}^{n}\mathsf{E}\left[\mathbf{K}^2\left(\frac{X_i - x_0}{h^0}\right)\boldsymbol{\Lambda}^2_{\lambda^0}(Z_i, z_0)\right]\right) = o(1), \qquad \text{(B.16)}
\end{aligned}
$$

where $\mathcal{X}_n = \sigma(X_1, \cdots, X_n)$ and $\mathcal{Z}_n = \sigma(Z_1, \cdots, Z_n)$. This implies that it is sufficient to show

$$
\widetilde{W}_n(x_0, z_0; h^0, \lambda^0) - \mathsf{E}\left[\widetilde{W}_n(x_0, z_0; h^0, \lambda^0)\right] \xrightarrow{d} \mathsf{N}\left[0, V_\star(x_0, z_0)\right], \qquad \text{(B.17)}
$$

where $V_\star(x_0, z_0) = \tau(1-\tau)\nu_0^p f_{xz}(x_0, z_0)$. By the classical Central Limit Theorem for the *i.i.d.* random variables, we can complete the proof of (B.17). In view of (B.15)–(B.17), we have that

$$
W_n(x_0, z_0; h^0, \lambda^0) - \mathsf{E}\left[W_n(x_0, z_0; h^0, \lambda^0)\right] \xrightarrow{d} \mathsf{N}\left[0, V_\star(x_0, z_0)\right]. \qquad \text{(B.18)}
$$

Meanwhile, by some elementary calculations, one can prove that

$$
\frac{1}{\sqrt{nH^0}}\frac{\mathsf{E}\left[W_n(x_0, z_0; h^0, \lambda^0)\right]}{f(x_0, z_0)} \sim b(x_0, z_0; h^0, \lambda^0). \qquad \text{(B.19)}
$$

By (B.18) and (B.19), we readily have that

$$
\zeta(x_0, z_0; h^0, \lambda^0) - \sqrt{nH^0}b(x_0, z_0; h^0, \lambda^0) \xrightarrow{d} \mathsf{N}\left[0, V(x_0, z_0)\right], \qquad \text{(B.20)}
$$

where $V(x_0, z_0)$ is defined as in (3.1).

Let $a = (a_1, \cdots, a_p)^\intercal$ and $d = (d_1, \cdots, d_q)^\intercal$ with $a_j = h_j \cdot n^{1/(4+p)}$ and $d_k = \lambda_k \cdot n^{2/(4+p)}$ as defined in Section 2. Similarly, let $\widehat{a} = (\widehat{a}_1, \cdots, \widehat{a}_p)^\intercal$ and $\widehat{d} = \left(\widehat{d}_1, \cdots, \widehat{d}_q\right)^\intercal$ with $\widehat{a}_j = \widehat{h}_j \cdot n^{1/(4+p)}$ and $\widehat{d}_k = \widehat{\lambda}_k \cdot n^{2/(4+p)}$, and let $a^0 = \left(a^0_1, \cdots, a^0_p\right)^\intercal$ and $d^0 = \left(d^0_1, \cdots, d^0_q\right)^\intercal$ with a^0_j and d^0_k defined as in Theorem 2.2. Define

$$
\overline{\zeta}(x_0, z_0; a, d) = \zeta(x_0, z_0; h, \lambda), \quad \overline{b}(x_0, z_0; a, d) = b(x_0, z_0; h, \lambda).
$$

when $h = a \cdot n^{-1/(4+p)}$ and $\lambda = d \cdot n^{-2/(4+p)}$. Let $\mathcal{S}_{c_1}(a^0)$ and $\mathcal{S}_{c_2}(d^0)$ be two neighborhoods of a^0 and d^0 with radius $c_1 > 0$ and $c_2 > 0$, respectively. Furthermore, let $\bar{\mathcal{A}}$ be a set of grid points $a(k)$ and $\bar{\mathcal{D}}$ a set of grid points $d(k)$ such that $[h(k), \lambda(k)] \in \mathcal{H}_m$ with $h(k) = a(k) \cdot m^{-1/(4+p)}$ and $\lambda(k) = d(k) \cdot m^{-2/(4+p)}$. By Proposition C.1 in Appendix C, one can show that

$$
\overline{\zeta}(x_0, z_0; a, d) = \frac{\overline{W}_n(x_0, z_0; a, d)}{f(x_0, z_0)} + o_P(1) \qquad \text{(B.21)}
$$

uniformly over $a \in \bar{\mathcal{A}}$ and $d \in \bar{\mathcal{D}}$, where $\overline{W}_n(x_0, z_0; a, d) = W_n(x_0, z_0; h, \lambda)$. Define

$$
\overline{W}^c_n(x_0, z_0; a, d) = \overline{W}_n(x_0, z_0; a, d) - \mathsf{E}\left[\overline{W}_n(x_0, z_0; a, d)\right],
$$

and $\bar{\mathcal{S}}_{c_1}(a^0) = \mathcal{S}_{c_1}(a^0) \cap \bar{\mathcal{A}}$ and $\bar{\mathcal{S}}_{c_2}(d^0) = \mathcal{S}_{c_2}(d^0) \cap \bar{\mathcal{D}}$. With (B.21), Theorem 2.2, the definition of the RCV selected smoothing parameters, and Theorem 3.1 in Li and Li (2010), we only have to show that

$$\mathsf{E}\left[\left|\overline{W}_n^c(x_0, z_0; a, d) - \overline{W}_n^c(x_0, z_0; a', d')\right|^2\right] \leqslant c_3 \cdot \|a - a'\|^2, \tag{B.22}$$

where $a, a' \in \bar{\mathcal{S}}_{c_1}(a^0)$, $d, d' \in \bar{\mathcal{S}}_{c_2}(d^0)$ and c_3 is a positive constant, and

$$\frac{\mathsf{E}\left[\overline{W}_n(x_0, z_0; a, d)\right]}{f(x_0, z_0)} - \sqrt{nH} \cdot \overline{b}(x_0, z_0; a, d) = o(1) \tag{B.23}$$

uniformly over $a \in \bar{\mathcal{S}}_{c_1}(a^0)$ and $d \in \bar{\mathcal{S}}_{c_2}(d^0)$. The proof of (B.23) is straightforward as its left hand wide is non-random and one can use the standard Taylor expansion of the quantile regression function to prove this result. We therefore only need to prove (B.22).

Without loss of generality, we consider $p = 1$ and define

$$\varpi_i(x_0, z_0; a, d) = \eta_i(x_0, z_0) K\left(\frac{X_i - x_0}{h}\right) \boldsymbol{\Lambda}_\lambda(Z_i, z_0)$$

with $h = a \cdot n^{-1/(4+p)}$ and $\lambda = d \cdot n^{-2/(4+p)}$. Note that

$$\begin{aligned}
& \mathsf{E}\left\{\left[\varpi_i(x_0, z_0; a, d) - \varpi_i(x_0, z_0; a', d')\right]^2\right\} \\
= \;& \int \mathsf{E}\left[\eta_i^2(x_0, z_0) | X_i = x, Z_i = z_0\right] \left[K\left(\frac{x - x_0}{h}\right) - K\left(\frac{x - x_0}{h'}\right)\right]^2 dx + O\left(\sum_{k=1}^q \lambda_k + \sum_{k=1}^q \lambda_k'\right) \\
= \;& h \int \mathsf{E}\left[\eta_i^2(x_0, z_0) | X_i = x_0 + hw, Z_i = z_0\right] \left[K(w) - K(hw/h')\right]^2 dw + O\left(\sum_{k=1}^q \lambda_k + \sum_{k=1}^q \lambda_k'\right) \\
\leqslant \;& c_4 h \left[1 - (h/h')\right]^2 = c_4 [a/(a')^2](a - a')^2,
\end{aligned} \tag{B.24}$$

where $h' = a' \cdot n^{-1/(4+p)}$ and $\lambda' = d' \cdot n^{-2/(4+p)}$ and c_4 is a positive constant. Using (B.24) and following the argument in the proof of Example 4.1 in Li and Li (2010), one can readily establish (B.22). This completes the proof of Theorem 3.1. ■

B.4 Proof of Theorem 3.2

Without loss of generality, we let $p = 1$ and $q = 1$. Let (h^0, λ^0) and $\left(\widehat{h}, \widehat{\lambda}\right)$ be the deterministic optimal and RCV selected bandwidth vectors, respectively. Define

$$\mathcal{S}_{\bar{\varepsilon}}(h^0, \lambda^0) = \left\{(h, \lambda) \in \mathcal{H}_n : \left|\frac{h - h^0}{h^0}\right| < \bar{\varepsilon}, \left|\frac{\lambda - \lambda^0}{\lambda^0}\right| < \bar{\varepsilon}\right\},$$

where $\mathcal{H}_n$ is a set of grid points $[h(k), \lambda(k)]$ such that $\left[(n/m)^{1/(4+p)}h(k), (n/m)^{2(4+p)}\lambda(k)\right] \in \mathcal{H}_m$. Let a^0 and d^0 satisfy $h^0 = a^0 \cdot n^{-1/5}$ and $\lambda^0 = d^0 \cdot n^{-2/5}$ and define

$$\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0, d^0) = \left\{(a, d) : \; a \in \bar{\mathcal{A}}, \; d \in \bar{\mathcal{D}}, \; \left|a - a^0\right| < \bar{\varepsilon}, \; \left|d - d^0\right| < \bar{\varepsilon}\right\},$$

where $\bar{\mathcal{A}}$ and $\bar{\mathcal{D}}$ are defined as in the proof of Theorem 3.1. By Theorem 2.2, for any $\bar{\varepsilon} > 0$, we readily have that

$$\mathsf{P}\left(\left(\widehat{h}, \widehat{\lambda}\right) \in \mathcal{S}_{\bar{\varepsilon}}(h^0, \lambda^0)\right) \to 1 \;\text{ and }\; \mathsf{P}\left(\left(\widehat{a}, \widehat{d}\right) \in \bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0, d^0)\right) \to 1, \tag{B.25}$$

where $\widehat{a}$ and $\widehat{d}$ are defined such that $\widehat{h} = \widehat{a} \cdot n^{-1/5}$ and $\widehat{\lambda} = \widehat{d} \cdot n^{-2/5}$.

For notational simplicity, we let $\widehat{\mathsf{Q}}_\tau(x, z; a, d) = \widehat{\mathsf{Q}}_\tau(x, z; h, \lambda)$, where $h = an^{-1/5}$ and $\lambda = dn^{-2/5}$. With (B.25), it is sufficient to prove that

$$\mathsf{P}\left(\sup_{(a,d)\in\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0,d^0)} \sup_{x\in\mathcal{D}_x(\epsilon)} \left|\widehat{\mathsf{Q}}_\tau(x, z; a, d) - \mathsf{Q}_\tau(x, z)\right| > c_5 \cdot \iota_n\right) \to 0, \tag{B.26}$$

where c_5 is a sufficiently large positive constant and $\iota_n = n^{-2/5}\log^{1/2} n$. Combining Proposition C.1 and the proof of Example 2.1 in Li and Li (2010), we readily have the following Bahadur representation:

$$\widehat{\mathsf{Q}}_\tau(x, z; a, d) - \mathsf{Q}_\tau(x, z) = f^{-1}(x, z)w_n(x, z; a, d)(1 + o_P(1)), \tag{B.27}$$

uniformly over $x \in \mathcal{D}_x(\epsilon)$ and $(a, d) \in \bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0, d^0)$, where

$$w_n(x, z; a, d) = \frac{1}{(nh)^{1/2}}\overline{W}_n(x, z; a, d) = \frac{1}{nh}\sum_{i=1}^{n} \eta_i(x, z)K\left(\frac{X_i - x}{h}\right)\Lambda_\lambda(Z_i, z)$$

with $h = an^{-1/5}$ and $\lambda = dn^{-2/5}$. By Assumption 3, $f(x, z)$ is strictly larger than a positive constant uniformly over x and z. On the other hand, we may prove that

$$\sup_{(a,d)\in\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0,d^0)} \sup_{x\in\mathcal{D}_x(\epsilon)} \left|\mathsf{E}\left[w_n(x, z; a, d)\right]\right| = O(n^{-2/5}) = o(\iota_n). \tag{B.28}$$

Hence, by (B.27) and (B.28), in order to prove (B.26), we only need to show that

$$\sup_{(a,d)\in\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0,d^0)} \sup_{x\in\mathcal{D}_x(\epsilon)} \left|w_n(x, z; a, d) - \mathsf{E}\left[w_n(x, z; a, d)\right]\right| = O_P(\iota_n). \tag{B.29}$$

Consider covering the set $\mathcal{D}_x(\epsilon)$ by some disjoint intervals $\mathcal{D}_1, \cdots, \mathcal{D}_{L_1}$. Denote the center point of $\mathcal{D}_l$ by x_l and let the radius of $\mathcal{D}_l$ be of order $\iota_n n^{-2/5}$. Then, the order of the number L_1 is $\iota_n^{-1}n^{2/5}$. Let L_2 be the number of grid points in $\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0, d^0)$ which diverges at a polynomial rate of n by Assumption 5(iii). Note that

$$\sup_{(a,d)\in\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0,d^0)} \sup_{x\in\mathcal{D}_x(\epsilon)} \left|w_n(x, z; a, d) - \mathsf{E}\left[w_n(x, z; a, d)\right]\right|$$

$$\leqslant \max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\left|w_n(x_{l_1},z;a_{l_2},d_{l_2})-\mathsf{E}\left[w_n(x_{l_1},z;a_{l_2},d_{l_2})\right]\right|+$$
$$\max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\sup_{x\in\mathcal{D}_{l_1}}\left|w_n(x,z;a_{l_2},d_{l_2})-w_n(x_{l_1},z;a_{l_2},d_{l_2})\right|+$$
$$\max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\sup_{x\in\mathcal{D}_{l_1}}\left|\mathsf{E}\left[w_n(x,z;a_{l_2},d_{l_2})\right]-\mathsf{E}\left[w_n(x_{l_1},z;a_{l_2},d_{l_2})\right]\right|, \tag{B.30}$$

where (a_l,d_l), $1\leqslant l\leqslant L_2$, are the grid points in the set $\bar{\mathcal{S}}_{\bar{\varepsilon}}(a^0,d^0)$. As the radius of $\mathcal{D}_l$ has the order of $\iota_n n^{-2/5}$, by the smoothness condition on $K(\cdot)$ in Assumption 1, we may show that

$$\max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\sup_{x\in\mathcal{D}_{l_1}}\left|w_n(x,z;a_{l_2},d_{l_2})-w_n(x_{l_1},z;a_{l_2},d_{l_2})\right|=O_P(\iota_n) \tag{B.31}$$

and

$$\max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\sup_{x\in\mathcal{D}_{l_1}}\left|\mathsf{E}\left[w_n(x,z;a_{l_2},d_{l_2})\right]-\mathsf{E}\left[w_n(x_{l_1},z;a_{l_2},d_{l_2})\right]\right|=O(\iota_n). \tag{B.32}$$

On the other hand, by the Bonferroni inequality and Bernstein inequality for independent sequence (e.g., van der Vaart and Wellner, 1996), we can prove that

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\left|w_n(x_{l_1},z;a_{l_2},d_{l_2})-\mathsf{E}\left[w_n(x_{l_1},z;a_{l_2},d_{l_2})\right]\right|>c_5\iota_n\right)\\
\leqslant\ &\sum_{l_1=1}^{L_1}\sum_{l_2=1}^{L_2}\mathsf{P}\left(\left|w_n(x_{l_1},z;a_{l_2},d_{l_2})-\mathsf{E}\left[w_n(x_{l_1},z;a_{l_2},d_{l_2})\right]\right|>c_5\iota_n\right)\\
\leqslant\ &O\left(L_1\cdot L_2\cdot\exp\{-\bar{c}_5\log n\}\right)=o(1),
\end{aligned} \tag{B.33}$$

where $\bar{c}_5$ would be a sufficiently large positive constant if c_5 is large enough. Hence, we have

$$\max_{1\leqslant l_1\leqslant L_1}\max_{1\leqslant l_2\leqslant L_2}\left|w_n(x_{l_1},z;a_{l_2},d_{l_2})-\mathsf{E}\left[w_n(x_{l_1},z;a_{l_2},d_{l_2})\right]\right|=O_P(\iota_n). \tag{B.34}$$

By (B.30)–(B.32) and (B.34), we can prove (B.29), completing the proof of (3.3) in Theorem 3.2. ■

B.5 Estimators of bias and variance

In this appendix, we discuss the construction of the estimators of local linear quantile estimation bias and variance. Following the discussion in Section 3, the estimator of $b(x_0,z_0;\widehat{h},\widehat{\lambda})$ can be obtained by replacing unknown functions by their consistent estimates as follows

$$\widehat{b}(x_0,z_0;\widehat{h},\widehat{\lambda})=\frac{\mu_2}{2}\sum_{j=1}^{p}\widehat{h}_j^2\widehat{Q}''_{\tau,j}(x_0,z_0)+\sum_{z_1\in\mathcal{D}_z}\sum_{k=1}^{q}I_k(z_1,z_0)\widehat{\lambda}_k\frac{\widehat{f}(x_0,z_1)}{\widehat{f}(x_0,z_0)}\left[\widehat{Q}_\tau(x_0,z_1)-\widehat{Q}_\tau(x_0,z_0)\right],$$

where $\widehat{f}(x_0, z_0) = \widehat{f}_e(0|x_0, z_0)\widehat{f}_{xz}(x_0, z_0)$ with

$$\widehat{f}_e(0|x_0, z_0) = \frac{\sum_{i=1}^n K_{h_0}(\widehat{e}_i)\mathbf{K}_h(X_i - x_0)\mathbf{\Lambda}_\lambda(Z_i, z_0)}{\sum_{i=1}^n \mathbf{K}_h(X_i - x_0)\mathbf{\Lambda}_\lambda(Z_i, z_0)}, \tag{B.35}$$

$$\widehat{f}_{xz}(x_0, z_0) = \frac{1}{nH}\sum_{i=1}^n \mathbf{K}_h(X_i - x_0)\mathbf{\Lambda}_\lambda(Z_i, z_0), \tag{B.36}$$

$K_{h_0}(\widehat{e}_i) = h_0^{-1}K(\widehat{e}_i/h_0)$, $\widehat{e}_i = Y_i - \widehat{Q}_\tau(X_i, Z_i; \widehat{h}, \widehat{\lambda})$, the bandwidth h_0 in (B.35) can be determined by the rule of thumb: $h_0 = s_{\widehat{e}}n^{-1/(5+p)}$ with $s_{\widehat{e}}$ being the sample standard deviation of $\widehat{e}_i$, and $\widehat{Q}''_{\tau,j}(x_0, z_0)$ is the j-th diagonal element of the $p \times p$ matrix $\widehat{\Gamma}$. The matrix $\widehat{\Gamma}$, along with $\widehat{\alpha}$ and $\widehat{\beta}$, is defined by minimizing the following objective function (e.g., Cheng and Peng, 2002):

$$\sum_{i=1}^n \rho_\tau\left(Y_i - \alpha - (X_i - x_0)^\top\beta - (\frac{1}{2})(X_i - x_0)^\top\Gamma(X_i - x_0)\right)\mathbf{K}_h(X_i - x_0)\mathbf{\Lambda}_\lambda(Z_i, z_0).$$

The consistent estimator of $V(x_0, z_0)$ is given by

$$\widehat{V}(x_0, z_0) = \frac{\tau(1-\tau)\nu_0^p}{\left[\widehat{f}_e(0|x_0, z_0)\right]^2\widehat{f}_{xz}(x_0, z_0)},$$

where $\widehat{f}_e(0|x_0, z_0)$ and $\widehat{f}_{xz}(x_0, z_0)$ are defined in (B.35) and (B.36), respectively.

References

Belloni, A. and Chernozhukov, V. (2011). l_1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*, 39, 82–130.

Belloni, A., Chernozhukov, V., Chetverikov, D. and Fernández-Val, I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, 213, 4–29.

Cai, Z. (2002). Regression quantile for time series. *Econometric Theory*, 18, 169–192.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167, 413–425.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103, 1596–1608.

Calonico, E., Cattaneo, M.D. and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82, 2295-2326.

Calonico, E., Cattaneo, M.D. and and Farrell, M.H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of American Statistical Association*, 113, 767-779.

Chen, Z., Fan, J. and Li, R. (2018). Error variance estimation in ultrahigh dimensional additive models. *Journal of the American Statistical Association*, 113, 315–327.

Cheng, M. and Peng, L. (2002). Regression modeling for nonparametric estimation and distribution and quantile functions. *Statistica Sinica*, 12, 1043–1060.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78, 1093–1125.

Chiang, H.D., Hsu, Y.C. and Sasaki, Y. (2019). Robust uniform inference for quantile treatment effects in regression discontinuity designs. *Journal of Econometrics*, 211, 589–618.

Chiang, H.D. and Sasaki, Y. (2019). Causal inference by quantile regression kink designs. *Journal of Econometrics*, 210, 405–433.

Dellarocas, C., Zhang, X. and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.

Duan, W., Gu, B. and Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016.

Escanciano, J. C. and Goh, C. (2014). Specification analysis of linear quantile models. *Journal of Econometrics*, 178, 495–507.

Escanciano, J. C. and Velasco, C. (2010). Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, 159, 209–221.

Fan J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J., Guo, S. and Hao, N. (2012). Variance estimatino using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society Series B*, 74, 37–65.

Galvao, A. (2011) Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164, 142–157.

Galvao, A., Lamarche, C. and Lima, L. (2013). Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association*, 108, 1075–1089.

Gilchrist, D. S. and Sands, E. G. (2016). Something to talk about: Social spillovers in movie consumption. *Journal of Political Economy*, 124(5), 1339–1382.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Monographs on Statistics and Applied Probability 58, Chapman and Hall.

Guerre, E. and Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, 28, 87–129.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Härdle, W., Hall, P. and Marron, J.S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of American Statistical Association*, 83, 86-101.

Härdle, W. and Song, S. (2010). Confidence bands in quantile regression. *Econometric Theory*, 26, 1180–1200.

Hall, P., Lahiri, S. N. and Polzehl, J. (1995). Bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Annals of Statistics*, 23, 1921–1936.

Hall, P., Li, Q. and Racine, J. (2007). Nonparametric estimation of regression functions in the presence of irrelevant regressors. *The Review of Economics and Statistics*, 89, 784–789.

Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli*, 15, 659–686.

Horowitz, J. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Springer Series in Statistics, Springer.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. and Bassett, G. W. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

Koenker, R., Chernozhukov, V., He, X. and Peng L. (2017). *Handbook of Quantile Regression*. Chapman and Hall/CRC.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101, 980–990.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics*, 26, 755–770.

Kong, E., Linton, O. and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, 26, 1529–1564.

Leung, D. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics*, 33, 2291–2310.

Li, D. and Li, Q. (2010). Nonparametric/semiparametric estimation and testing of econometric models with data dependent smoothing parameters. *Journal of Econometrics*, 157, 179–190.

Li, Q., Lin, J. and Racine, J. (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business and Economic Statistics* 31, 57–65.

Li, Q. and Racine, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14, 485–512.

Li, Q. and Racine, J. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, 26, 423–434.

Li, Q. and Zhou, R. (2005). The uniqueness of cross-validation selected smoothing parameters in kernel estimation of nonparametric models. *Econometric Theory* 21, 1017–1025

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Qu, Z. and Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185, 1–19.

Qu, Z. and Yoon, J. (2018). Uniform inference on quantile effects under sharp regression discontinuity designs. Forthcoming in *Journal of Business and Economic Statistics*.

Racine, J. and Li, K. (2017). Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. *Journal of Econometrics*, 201, 72–94.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics*, 12, 1215–1230.

Spokoiny, V., Wang, W. and Härdle, W. (2013). Local quantile regression. *Journal of Statistical Planning and Inference*, 143, 1109–1129.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall.

Xia, Y. and Li, W. K. (2002). Asymptotic behavior of bandwidth selected by cross-validation method under dependence. *Journal of Multivariate Analysis*, 83, 265–287.

Yao, Q. and Tong, H. (1998). Cross-validatory bandwidth selections for regression estimation based on dependent data. *Journal of Statistical Planning and Inference*, 68, 387–415.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93, 228–238.

Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, 31, 333–346.

Zhou, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Annals of Statistics*, 38, 2187–2217.

Zhu, X., Wang, W., Wang, H. and Härdle, W. (2019). Network quantile autoregression. *Journal of Econometrics*, 212, 345–358.