Article:

Chen, Xirong, Li, Degui orcid.org/0000-0001-6802-308X, Li, Qi et al. (1 more author) (2019) Nonparametric Estimation of Conditional Quantile Functions in the Presence of Irrelevant Covariates. Journal of Econometrics. pp. 433-450. ISSN 0304-4076

https://doi.org/10.1016/j.jeconom.2019.04.037

Nonparametric Estimation of Conditional Quantile Functions in the Presence of Irrelevant Covariates

Xirong Chen* Degui Li† Qi Li‡ Zheng Li§

February, 2019

Abstract

Allowing for the existence of irrelevant covariates, we study the problem of estimating a conditional quantile function nonparametrically with mixed discrete and continuous data. We estimate the conditional quantile regression function using the check-function-based kernel method and suggest a data-driven cross-validation (CV) approach to simultaneously determine the optimal smoothing parameters and remove the irrelevant covariates. When the number of covariates is large, we first use a screening method to remove the irrelevant covariates and then apply the CV criterion to those that survive the screening procedure. Simulations and an empirical application demonstrate the usefulness of the proposed methods.

Keywords: Cross-validation, Discrete regressors, Irrelevant covariates, Nonparametric quantile regression, Screening

JEL Classification: C13, C14, C35

*School of International Trade and Economics, University of International Business and Economics, China. Email: cxr1989@gmail.com

†Department of Mathematics, University of York, UK. Email: degui.li@york.ac.uk

‡ The corresponding author. Department of Economics, Texas A&M University, USA. Email: qi-li@tamu.edu

§Department of Agricultural and Resource Economics, NC State University, USA. Email: zli42@ncsu.edu

1 Introduction

Nonparametric estimation of conditional mean and/or quantile functions has received increasing attention among econometricians and statisticians in recent decades (c.f., Fan and Gijbels, 1996; Ghysels and Ng, 1998; Pagan and Ullah, 1999; Cai, 2002; Ai and Chen, 2003; Fan and Yao, 2003; Belloni, Chernozhukov and Fernández-Val, 2011; Fan and Park, 2012; Fan and Liu, 2016). Compared with a conditional mean regression function, a conditional quantile regression function, when evaluated at different quantiles, can provide a more comprehensive picture of the impact of covariates on the response variable and thus reveal an entire distributional relationship between the covariates and the response variable. Among various nonparametric estimation techniques, the kernel-based smoothing method is probably the most commonly-used one for applied researchers. It is well known that the numerical performance of nonparametric kernel estimation relies on the choice of smoothing (or bandwidth) parameters. Hence, there has been a large literature on data-driven methods to select optimal smoothing parameters in estimating conditional mean functions and conditional density functions, see Hall and Marron (1987), Härdle, Hall and Marron (1988), Marron, Jones and Sheather (1996) and Hall, Racine and Li (2004), among others.

In contrast, there is relatively sparse literature on developing data-driven methods to select optimal smoothing parameters in estimating conditional quantile functions, which seems more challenging than that in estimating conditional mean functions. This is partly due to that the check-function-based conditional quantile estimation involves minimization of a non-smooth objective function and the resulting estimation lacks a closed form. One may avoid the problem of non-smooth objective function by first estimating a conditional cumulative distribution function (CDF) and then inverting the conditional CDF to obtain the conditional quantile function (e.g., Cai, 2002; Li, Lin and Racine, 2013). The optimal smoothing parameters can be selected in the first step of nonparametric CDF estimation according to certain data-driven criterion. However, the smoothing parameters chosen in this way are usually not optimal for the conditional quantile estimation, see the simulation studies in Li, Li and Li (2018). In addition, such an inverted-CDF-based method is difficult to be extended to a more general setting in particular when one is interested in estimating derivatives of the conditional quantile function. The latter issue is addressed in a recent paper by Li, Li and Li (2018) which uses the local linear smoothing method in minimizing the non-smooth objective function and then obtains the conditional quantile function estimation. They introduce a data-driven cross-validation (CV) method to directly select the smoothing parameters for the nonparametric quantile regression estimation and derive the asymptotical optimality property for the CV selected smoothing parameters.

However, Li, Li and Li (2018)'s paper restricts attention to the case that all the discrete and continuous covariates are relevant in the sense that they all significantly affect the conditional quantile function. Such an assumption becomes inappropriate when one faces a large number of candidate covariates in econometric modelling, which is not uncommon in practical applications. It is likely that some of the covariates are redundant in the sense that they do not have any impact on the response variable, and should be removed to improve the estimation efficiency. Hall, Racine and Li (2004) and Hall, Li and Racine (2007) explore this issue in the context of conditional density function estimation and conditional mean function estimation, where they show that, through the least squares CV method, the "irrelevant covariates" can be removed by over-smoothing. For other relevant developments on the CV model selection in parametric, nonparametric and semiparametric models, we refer to Zhang (1991), Shao (1993), Gao and Tong (2004), Gao *et al* (2017) and the references therein.

Most of the aforementioned literature focuses on the CV method to remove the irrelevant covariates and select the optimal smoothing parameters associated with the significant ones within the conditional mean regression framework. It remains an open problem to develop a completely data-driven method to simultaneously select optimal smoothing parameters and remove redundant covariates in nonparametric conditional quantile regression. The current paper fills this gap. We use the local constant check-function-based method to estimate the conditional quantile regression function, where the discrete and continuous kernel functions are combined to deal with the mixed discrete and continuous regressors. A completely data-driven CV approach is applied to jointly determine the optimal smoothing parameters and remove the irrelevant covariates (via over-smoothing). Under some mild conditions, the CV selected smoothing parameters are proved to be asymptotically optimal with convergence rates comparable to those obtained by Racine and Li (2004). In addition, the irrelevant covariates (which can be either continuous or discrete) are over-smoothed and thus removed with probability approaching one, indicating the consistency of covariate selection. The asymptotic normal distribution of the local kernel quantile estimation using the CV selected smoothing parameters is also established, complementing the results derived in Li and Li (2010). Furthermore, we generalize the model setting and methodology to the case when the dimension of covariates is large (growing with the sample size n) and introduce a two-step procedure: (i) use a kernel-based quantile screening technique to remove the irrelevant continuous and discrete covariates, and (ii) apply the CV criterion to those that survive in the first step of screening and further select the significant covariates and determine the optimal smoothing parameters. Note that the existing literature on variable or feature selection in high-dimensional quantile regression only considers the case of purely continuous covariates (e.g., He, Wang and Hong, 2013; Ma, Li and Tsai, 2017; Xia, Li and Fu, 2018). The present paper considers a more general setting which contains both the discrete and continuous covariates. Our simulation studies show that the proposed procedure has a reasonably good small-sample performance. In the empirical application, we apply the developed method to analyze the data taken from the National Longitudinal Survey of Youth 1997, and find that while men's dating experience is positively correlated with their median wage, women's dating experience is smoothed out after using the CV method, indicating that women's dating experience is irrelevant to their median wage.

The rest of the paper is organized as follows. The local constant check-function-based estimation method and the CV method are introduced in Section 2. The technical assumptions and the main asymptotic results are given in Section 3. Methodology and theory for the case of high-dimensional covariates are presented in Section 4. Section 5 reports the simulation results and Section 6 presents an empirical application. Section 7 concludes the paper. The proofs of the main results are given in Appendix A, and the proofs of the technical lemmas are provided in Appendix B contained in a supplemental document.

2 Conditional Quantile Estimation

In this section, we describe the nonparametric kernel-based smoothing method to estimate the conditional quantile regression function with mixed discrete and continuous covariates, and then introduce the CV method to select the optimal bandwidth parameters. Since the seminal paper by Koenker and Bassett (1978), the parametric and nonparametric quantile regression modelling has experienced rapid developments (c.f., Jones and Hall, 1990; Yu and Jones, 1998; Cai, 2002; Chernozhukov and Hong, 2002; Koenker, 2005; Angrist, Chernozhukov and Fernández-Val, 2006; Koenker *et al*, 2017; Racine and Li, 2017; Li, Li and Li, 2018). In this paper, we suppose that $\left(Y_i, \bar{\mathbf{X}}_i^c, \tilde{\mathbf{X}}_i^c, \bar{\mathbf{X}}_i^d, \tilde{\mathbf{X}}_i^d\right)$, $i = 1, \cdots, n$, are independent and identically distributed as $\left(Y, \bar{\mathbf{X}}^c, \tilde{\mathbf{X}}^c, \bar{\mathbf{X}}^d, \tilde{\mathbf{X}}^d\right)$, where Y is univariate, $\bar{\mathbf{X}}^c = \left(\bar{X}_1^c, \bar{X}_2^c, \cdots, \bar{X}_{d_1}^c\right)^\intercal$ is a d_1-dimensional *relevant* continuous covariate vector, $\tilde{\mathbf{X}}^c = \left(\tilde{X}_1^c, \tilde{X}_2^c, \cdots, \tilde{X}_{d_2}^c\right)^\intercal$ is a d_2-dimensional *irrelevant* continuous covariate vector, $\bar{\mathbf{X}}^d = \left(\bar{X}_1^d, \bar{X}_2^d, \cdots, \bar{X}_{d_3}^d\right)^\intercal$ is a d_3-dimensional *relevant* discrete covariate vector and $\tilde{\mathbf{X}}^d = \left(\tilde{X}_1^d, \tilde{X}_2^d, \cdots, \tilde{X}_{d_4}^d\right)^\intercal$ is a d_4-dimensional *irrelevant* discrete covariate vector. Without loss of generality, we assume that $\bar{\mathbf{X}}_j^d \in \bar{\mathcal{D}}_j \stackrel{\text{def}}{=} \{0, 1, \cdots, \bar{c}_j - 1\}$ for $j = 1, \cdots, d_3$ and $\tilde{\mathbf{X}}_j^d \in \tilde{\mathcal{D}}_j \stackrel{\text{def}}{=} \{0, 1, \cdots, \tilde{c}_j - 1\}$ for $j = 1, \cdots, d_4$, where $\bar{c}_j$ and $\tilde{c}_j$ are bounded positive integers. Let $\mathcal{S} = \bar{\mathcal{S}} \times \tilde{\mathcal{S}}$ with $\bar{\mathcal{S}}$ and $\tilde{\mathcal{S}}$ being the compact supports of $\bar{\mathbf{X}}^c$ and $\tilde{\mathbf{X}}^c$, respectively. In this section, we consider the simple case when all the dimensions, d_i, $1 \leqslant i \leqslant 4$, are fixed. Extension of methodology to the more general setting with diverging dimensions will be studied in Section 4.

The irrelevant covariates are assumed to be independent of the response variable and the relevant covariates, i.e.,

$$\left(Y, \bar{\mathbf{X}}^c, \bar{\mathbf{X}}^d\right) \perp \left(\tilde{\mathbf{X}}^c, \tilde{\mathbf{X}}^d\right), \tag{2.1}$$

where the notation $A \perp B$ means that A and B are independent with each other. Note that the condition (2.1) implies that, for any fixed

$$\bar{\mathbf{x}}_0 = \begin{pmatrix} \bar{\mathbf{x}}_0^c \\ \bar{\mathbf{x}}_0^d \end{pmatrix} \text{ and } \tilde{\mathbf{x}}_0 = \begin{pmatrix} \tilde{\mathbf{x}}_0^c \\ \tilde{\mathbf{x}}_0^d \end{pmatrix},$$

we have $F(y|\bar{\mathbf{x}}_0, \tilde{\mathbf{x}}_0) = F(y|\bar{\mathbf{x}}_0)$, where $F(y|\bar{\mathbf{x}}_0)$ is the conditional CDF of the response variable Y (evaluated at y) given the covariates $\bar{\mathbf{X}}^c = \bar{\mathbf{x}}_0^c$ and $\bar{\mathbf{X}}^d = \bar{x}_0^d$, where $\bar{\mathbf{x}}_0^c \in \bar{\mathcal{S}}$, $\tilde{\mathbf{x}}_0^c \in \tilde{\mathcal{S}}$, $\bar{\mathbf{x}}_0^d \in \prod_{j=1}^{d_3} \bar{\mathcal{D}}_j$ and $\tilde{\mathbf{x}}_0^d \in \prod_{j=1}^{d_4} \tilde{\mathcal{D}}_j$ are vectors of dimensions d_1, d_2, d_3 and d_4, respectively. For given $0 < \tau < 1$, we use $Q_\tau(\bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d)$ to denote the conditional τ-quantile function of the response variable Y given $\bar{\mathbf{X}}^c = \bar{\mathbf{x}}_0^c$ and $\bar{\mathbf{X}}^d = \bar{\mathbf{x}}_0^d$, i.e.,

$$Q_\tau(\bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d) = \inf\left\{y \in \mathcal{R}: \; F(y|\bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d) \geqslant \tau\right\}, \tag{2.2}$$

or equivalently,

$$Q_\tau(\bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d) = \arg\min_{a \in \mathcal{R}} \mathsf{E}\left[\rho_\tau(Y - a)\middle|\bar{\mathbf{X}}^c = \bar{\mathbf{x}}_0^c, \bar{\mathbf{X}}^d = \bar{\mathbf{x}}_0^d\right], \tag{2.3}$$

where $\rho_\tau(\cdot)$ is the check function $\rho_\tau(y) = y\left[\tau - I(y < 0)\right]$, and $I(\mathcal{A})$ is the indicator function of the event $\mathcal{A}$.

In practice, the prior information on the irrelevant covariates $\tilde{\mathbf{X}}^c$ and $\tilde{\mathbf{X}}^d$ is usually unknown. Hence, we have to use the full sample containing both the relevant and irrelevant covariates in the initial local kernel-based estimation of the quantile regression function. For notational simplicity, let

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}_i^c \\ \mathbf{X}_i^d \end{pmatrix}, \; \mathbf{X}_i^c = \begin{pmatrix} \bar{\mathbf{X}}_i^c \\ \tilde{\mathbf{X}}_i^c \end{pmatrix}, \; \mathbf{X}_i^d = \begin{pmatrix} \bar{\mathbf{X}}_i^d \\ \tilde{\mathbf{X}}_i^d \end{pmatrix},$$

and accordingly

$$\mathbf{x}_0 = \begin{pmatrix} \mathbf{x}_0^c \\ \mathbf{x}_0^d \end{pmatrix}, \ \mathbf{x}_0^c = \begin{pmatrix} \bar{\mathbf{x}}_0^c \\ \tilde{\mathbf{x}}_0^c \end{pmatrix}, \ \mathbf{x}_0^d = \begin{pmatrix} \bar{\mathbf{x}}_0^d \\ \tilde{\mathbf{x}}_0^d \end{pmatrix}.$$

Let $h = (\bar{h}, \tilde{h})$ and $\lambda = (\bar{\lambda}, \tilde{\lambda})$, where $\bar{h}$, $\tilde{h}$, $\bar{\lambda}$ and $\tilde{\lambda}$ are row vectors of smoothing parameters with dimensions d_1, d_2, d_3 and d_4, respectively. These smoothing parameters correspond to the covariate vectors $\bar{\mathbf{X}}^c$, $\tilde{\mathbf{X}}^c$, $\bar{\mathbf{X}}^d$ and $\tilde{\mathbf{X}}^d$, respectively. Since there are both the continuous and discrete covariates, we need to use different types of kernel functions to smooth them. For the continuous covariates, we use the product kernel defined by

$$K_h(\mathbf{X}_i^c - \mathbf{x}_0^c) = \prod_{s=1}^{d_1} \bar{h}_s^{-1} k\left(\frac{\bar{X}_{is}^c - \bar{x}_{0s}^c}{\bar{h}_s}\right) \prod_{s=1}^{d_2} \tilde{h}_s^{-1} k\left(\frac{\tilde{X}_{is}^c - \tilde{x}_{0s}^c}{\tilde{h}_s}\right), \tag{2.4}$$

where $k(\cdot)$ is a univariate kernel function, $\bar{X}_{is}^c$ and $\bar{x}_{0s}^c$ are the s-th element of $\bar{\mathbf{X}}_i^c$ and $\bar{\mathbf{x}}_0^c$, respectively, $\tilde{X}_{is}^c$ and $\tilde{x}_{0s}^c$ are defined similarly, $\bar{h} = (\bar{h}_1, \cdots, \bar{h}_{d_1})$ and $\tilde{h} = (\tilde{h}_1, \cdots, \tilde{h}_{d_2})$. For the discrete covariates, we use a different kernel function which is defined by

$$\Lambda_\lambda(\mathbf{X}_i^d, \mathbf{x}_0^d) = \prod_{s=1}^{d_3} \bar{\lambda}_s^{I\left(\bar{X}_{is}^d \neq \bar{x}_{0s}^d\right)} \prod_{s=1}^{d_4} \tilde{\lambda}_s^{I\left(\tilde{X}_{is}^d \neq \tilde{x}_{0s}^d\right)}, \tag{2.5}$$

where $\bar{X}_{is}^d$ and $\bar{x}_{0s}^d$ are the s-th element of $\bar{\mathbf{X}}_i^d$ and $\bar{\mathbf{x}}_0^d$, respectively, $\tilde{X}_{is}^d$ and $\tilde{x}_{0s}^d$ are defined similarly, $\lambda = (\bar{\lambda}, \tilde{\lambda}) = (\bar{\lambda}_1, \cdots, \bar{\lambda}_{d_3}, \tilde{\lambda}_1, \cdots, \tilde{\lambda}_{d_4}) \in [0,1]^{d_3+d_4}$ are the bandwidth parameters for the discrete covariates, and the convention of $0^0 = 1$ is used.

The local kernel estimate of the quantile regression function $Q_\tau(\bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d)$ is obtained as the minimizer to the following kernel-weighted objective function:

$$\mathcal{L}_n(\alpha; \mathbf{x}_0^c, \mathbf{x}_0^d) = \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \alpha) K_h(\mathbf{X}_i^c - \mathbf{x}_0^c) \Lambda_\lambda(\mathbf{X}_i^d, \mathbf{x}_0^d). \tag{2.6}$$

We denote the minimizer to the objective function $\mathcal{L}_n(\alpha; \mathbf{x}_0^c, \mathbf{x}_0^d)$ (with respect to α) by $\widehat{Q}_\tau(\mathbf{x}_0^c, \mathbf{x}_0^d; h, \lambda) = \widehat{Q}_\tau(\mathbf{x}_0; h, \lambda)$. When λ, the smoothing parameter vector in the discrete kernel, is chosen as a vector of zeros, the quantile estimator above reduces to the conventional local constant quantile estimator (c.f., Jones and Hall, 1990), splitting the full sample into many groups or sub-samples according to different values of the discrete covariates. On the other hand, if the j-th element of λ is chosen as one, the corresponding discrete covariate would not have any influence on the response variable. Such a discrete covariate is deemed to be irrelevant and should be deleted (c.f., Hall, Li and Racine, 2007). Therefore, we restrict the range for each component of the discrete smoothing parameter vector to be $[0,1]$, i.e., $0 \leqslant \bar{\lambda}_j \leqslant 1$ for all $j = 1, \cdots, d_3$, and $0 \leqslant \tilde{\lambda}_j \leqslant 1$ for all $j = 1, \cdots, d_4$.

We next introduce the CV method to determine the optimal values for the smoothing parameter vectors h and λ involved in the local constant kernel smoothing. Let $\widehat{Q}_{(-j)}(\mathbf{X}_j; h, \lambda)$ be the leave-one-out local constant estimate of $Q_\tau(\mathbf{X}_j) = Q_\tau(\bar{\mathbf{X}}_j^c, \bar{\mathbf{X}}_j^d)$ with bandwidths h and λ, which can be obtained as a minimizer to (2.6) with $\mathbf{x}_0$ and $\sum_{i=1}^n$ being replaced by $\mathbf{X}_j$ and $\sum_{i=1, i\neq j}^n$, respectively. We suggest choosing the bandwidth parameters (h, λ) by minimizing the following CV objective function:

$$\mathsf{CV}(h, \lambda) = \frac{1}{n} \sum_{j=1}^{n} \rho_\tau \left(Y_j - \widehat{Q}_{(-j)}(\mathbf{X}_j; h, \lambda) \right) W(\mathbf{X}_j), \tag{2.7}$$

where $W(\cdot)$ is a weight function that trims out boundary observations to avoid the well-known boundary effect in kernel-based estimation. Throughout the paper, we use h^* and λ^* to denote the CV selected bandwidths that minimize $\mathsf{CV}(h, \lambda)$ defined in (2.7).

3 Asymptotic Theory

In this section, we state the main asymptotic results for the methods proposed in Section 2. Let $f_e(\nu|\bar{\mathbf{x}})$ and $F_e(\nu|\bar{\mathbf{x}})$ denote the conditional density function and CDF of $e_i \stackrel{\text{def}}{=} Y_i - Q_\tau(\bar{\mathbf{X}}_i)$ evaluated at $e_i = \nu$ given $\bar{\mathbf{X}}_i = \bar{\mathbf{x}}$, respectively. We start with some regularity conditions which are needed to derive the asymptotic theory.

ASSUMPTION 1. *(i) The random vectors* $(Y_i, \mathbf{X}_i^\intercal)^\intercal$, $i = 1, \cdots, n$, *are independent and identically distributed (i.i.d.).*

(ii) The conditional density function of $\bar{\mathbf{X}}_i^c = \bar{\mathbf{x}}^c$ *given* $\bar{\mathbf{X}}_i^d = \bar{\mathbf{x}}^d$, $f(\bar{\mathbf{x}}^c|\bar{\mathbf{x}}^d)$, *is continuous and bounded away from infinity and zero for* $\bar{\mathbf{x}}^c \in \bar{\mathcal{S}}$ *and* $\bar{\mathbf{x}}^d \in \bar{\mathcal{D}}$, *where* $\bar{\mathcal{S}}$ *is the compact support of the relevant continuous covariates* $\bar{\mathbf{X}}_i^c$ *and* $\bar{\mathcal{D}} = \prod_{j=1}^{d_3} \bar{\mathcal{D}}_j$ *is the support of the relevant discrete covariates* $\bar{\mathbf{X}}_i^d$.

ASSUMPTION 2. *(i) For each* $\bar{\mathbf{x}} \in \bar{\mathcal{S}} \times \bar{\mathcal{D}}$, *the conditional density function* $f_e(\cdot|\bar{\mathbf{x}})$ *is strictly positive and has continuous first-order derivative at point zero.*

(ii) Both $f_e(\nu|\bar{\mathbf{x}})$ *and* $F_e(\nu|\bar{\mathbf{x}})$ *are positive and continuous with respect to* $\bar{\mathbf{x}}^c$, *where* ν *is in a small neighborhood of* 0. *In addition,* $F_e(0|\bar{\mathbf{x}}) = \tau$ *for all* $\bar{\mathbf{x}} \in \bar{\mathcal{S}} \times \bar{\mathcal{D}}$.

ASSUMPTION 3. *(i) The conditional quantile function* $Q_\tau(\cdot, \bar{\mathbf{x}}^d)$ *is twice continuously differentiable on* $\bar{\mathcal{S}}$ *for all* $\bar{\mathbf{x}}^d \in \bar{\mathcal{D}}$.

(ii) The weight function $W(\mathbf{x}) = W(\mathbf{x}^c, \mathbf{x}^d)$ *is bounded with* $W(\mathbf{x}^c, \mathbf{x}^d) = 0$ *if* $\mathbf{x}^c$ *is in a given small neighborhood of the boundary points of* $\mathcal{S}$*, where* $\mathcal{S}$ *is the compact support of the continuous covariates.*

ASSUMPTION 4. *(i) The univariate kernel function* $k(\cdot)$ *is a Lipschitz continuous and symmetric probability density function with a compact support, and* $k(0) \geqslant c_k > 0$.

(ii) Let $\bar{h}_s \to 0$ *for* $s = 1, \cdots, d_1$*, and there exists a bounded constant* $c > 0$ *such that* $n^{-c} < \tilde{h}_s < n^c$ *for* $s = 1, \cdots, d_2$.

(iii) Defining $H = \prod_{s=1}^{d_1} \bar{h}_s \prod_{s=1}^{d_2} (\tilde{h}_s \wedge 1)$, $n^{\epsilon - 1} \leqslant H \leqslant n^{-\epsilon}$ *with* $0 < \epsilon < 1/(d_1 + d_2 + 4)$*, where* $\wedge$ *denotes minimum. In addition, there exists a sequence of positive numbers* $\{m_n\}$ *such that* $m_n \geqslant \sqrt{\log n}$, $m_n^2 = o(nH)$*, and*

$$(nH) \left(\sum_{s=1}^{d_1} \bar{h}_s^4 + \sum_{s=1}^{d_3} \bar{\lambda}_s^2 \right) = O(m_n^2). \tag{3.1}$$

(iv) Let $\bar{\lambda}_s \to 0$, $s = 1, \cdots, d_3$*, and* $\tilde{\lambda}_s \in [0, 1]$, $s = 1, \cdots, d_4$.

REMARK 3.1. Assumption 1(i) imposes the *i.i.d.* condition on the random observations, which has been commonly used in the literature on nonparametric kernel estimation (c.f., Härdle, Hall and Marron, 1988; Marron, Jones and Sheather, 1996; Racine and Li, 2004). Note that there is no moment condition on e_i to estimate the conditional quantile function, indicating that the heavy-tail distribution for e_i is allowed. Assumption 1(ii) imposes mild restriction on the conditional density function of the relevant continuous covariates given the relevant discrete covariates. Assumptions 2 and 3 give some smoothness conditions on the (conditional) density function, CDF function and the conditional quantile functions, respectively, which are standard assumptions for kernel smoothing estimation of the conditional quantile function. In particular, Assumption 3(ii) ensures that the random observations with observed values of continuous covariates very close to the boundary points would be automatically trimmed out in the CV method, circumventing the well-known boundary effect in the kernel estimation. Assumption 4(i) imposes some mild conditions on the kernel function $k(\cdot)$, which can be satisfied by several commonly-used kernel functions such as the uniform kernel and the Epanechnikov kernel. Assumption 4(ii)–(iv) further imposes some restrictions on the smoothing parameters. For the relevant continuous and discrete covariates, all the associated smoothing parameters converge to 0 as $n \to \infty$. However, for the irrelevant continuous covariates, the associated smoothing parameters take values in a larger range which may be either convergent to 0 or divergent to ∞ as $n \to \infty$. For the irrelevant discrete covariates we only need that $\tilde{\lambda}_s \in [0,1]$ for all $s = 1, \cdots, d_4$ without any further restriction. The condition (3.1) in Assumption 4(iii) is mainly used to control the bias term and derive the uniform convergence results, where m_n is usually chosen as $\sqrt{\log n}$.

Before presenting the main results, we need to introduce some further notation. Let $\bar{x}_s^c$ and $\bar{x}_s^d$ be the s-th elements of the vectors $\bar{\mathbf{x}}^c$ and $\bar{\mathbf{x}}^d$, respectively. For $s = 1, \cdots, d_1$, let $Q_\tau^{(s)}(\bar{\mathbf{x}})$ and $f_{\bar{\mathbf{X}}}^{(s)}(\bar{\mathbf{x}})$ be the first-order derivative functions of $Q_\tau(\cdot)$ and $f_{\bar{\mathbf{X}}}(\cdot)$ with respect to $\bar{x}_s^c$, respectively, and let $Q_\tau^{(ss)}(\bar{\mathbf{x}})$ be the second-order derivative function of $Q_\tau(\cdot)$ with respect to the $\bar{x}_s$. Define

$$
\begin{aligned}
b(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda}) \quad = \quad & \frac{\mu_2}{2} \sum_{s=1}^{d_1} \bar{h}_s^2 \left[Q_\tau^{(ss)}(\bar{\mathbf{X}}_i) + 2 Q_\tau^{(s)}(\bar{\mathbf{X}}_i) \xi^{(s)}(\bar{\mathbf{X}}_i)/\xi(\bar{\mathbf{X}}_i) \right] + \\
& \sum_{\bar{\mathbf{x}}^d \in \bar{\mathcal{D}}} \frac{\xi(\bar{\mathbf{X}}_i^c, \bar{\mathbf{x}}^d)}{\xi(\bar{\mathbf{X}}_i)} \sum_{s=1}^{d_3} \bar{\lambda}_s I_s(\bar{\mathbf{x}}^d, \bar{\mathbf{X}}_i^d) \left[Q_\tau(\bar{\mathbf{X}}_i^c, \bar{\mathbf{x}}^d) - Q_\tau(\bar{\mathbf{X}}_i) \right], \qquad (3.2)
\end{aligned}
$$

where $\mu_2 = \int u^2 k(u) du$, $\xi(\bar{\mathbf{x}}) = f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}) f_e(0|\bar{\mathbf{x}})$, $\xi^{(s)}(\bar{\mathbf{x}})$ is the first-order derivative of $\xi(\bar{\mathbf{x}})$ with respect to $\bar{x}_s^c$, $I_s(\bar{\mathbf{x}}^d, \bar{\mathbf{X}}_i^d) = I\left(\bar{X}_{is}^d \neq \bar{x}_s^d\right) \prod_{k=1,\neq s}^{d_3} I\left(\bar{X}_{ik}^d = \bar{x}_k^d\right)$ and $\bar{\mathcal{D}}$ is defined in Assumption 1(ii). Let

$$
\sigma^2(\bar{\mathbf{X}}_i; \bar{h}) = \frac{1}{n\bar{H}} \frac{\tau(1-\tau)\nu_0}{f_{\bar{\mathbf{X}}}(\bar{\mathbf{X}}_i) f_e^2(0|\bar{\mathbf{X}}_i)}, \qquad (3.3)
$$

where $\bar{H} = \bar{h}_1 \cdots \bar{h}_{d_1}$ and $\nu_0 = \left[\int k^2(u) du\right]^{d_1}$. We will show in Appendix A that $b(\cdot; \bar{h}, \bar{\lambda})$ is the leading estimation bias term, whereas $\sigma^2(\cdot; \bar{h})$ is the leading estimation variance term which does not rely on $\bar{\lambda}$. Define the estimation mean squared error (MSE) as

$$
\mathrm{MSE}(h, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \left[Q_\tau(\bar{\mathbf{X}}_i) - \widehat{Q}_{(-i)}(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda}) \right]^2 W(\mathbf{X}_i) f_e(0|\bar{\mathbf{X}}_i). \qquad (3.4)
$$

Through the proofs in Appendix A, we show that the leading term of $\mathrm{MSE}(h, \lambda)$ is

$$
\begin{aligned}
\mathrm{MSE}_L(\bar{h}, \bar{\lambda}) \quad & = \quad \mathrm{E}\left\{ \left[b^2(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{X}}_i; \bar{h}) \right] \bar{W}(\bar{\mathbf{X}}_i) f_e(0|\bar{\mathbf{X}}_i) \right\} \\
& = \quad \int_{\bar{\mathcal{S}} \times \bar{\mathcal{D}}} \left[b^2(\bar{\mathbf{x}}; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{x}}; \bar{h}) \right] \bar{W}(\bar{\mathbf{x}}) f_e(0|\bar{\mathbf{x}}) f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}) d\bar{\mathbf{x}}, \qquad (3.5)
\end{aligned}
$$

where $\bar{W}(\bar{\mathbf{x}}) = \int_{\tilde{\mathcal{S}} \times \tilde{\mathcal{D}}} W(\mathbf{x}) f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}$, $f_{\tilde{\mathbf{X}}}(\cdot)$ is the density function of $\tilde{\mathbf{X}}_i$, $\tilde{\mathcal{S}}$ is the compact support of the irrelevant continuous covariates $\tilde{\mathbf{X}}_i^c$ and $\tilde{\mathcal{D}} = \prod_{j=1}^{d_3} \tilde{\mathcal{D}}_j$ is the support of the irrelevant discrete covariates $\tilde{\mathbf{X}}_i^d$. Choosing $\bar{h} = \bar{a} \cdot n^{-1/(d_1+4)}$ and $\bar{\lambda} = \bar{b} \cdot n^{-2/(d_1+4)}$ with $\bar{a} = (\bar{a}_1, \cdots, \bar{a}_{d_1})$ and $\bar{b} = \left(\bar{b}_1, \cdots, \bar{b}_{d_3}\right)$, and letting

$$
\mathrm{MSE}_L^\star(\bar{a}, \bar{b}) = \int_{\bar{\mathcal{S}} \times \bar{\mathcal{D}}} \left[b^2(\bar{\mathbf{x}}; \bar{a}, \bar{b}) + \frac{1}{\bar{a}_1 \cdots \bar{a}_{d_1}} \frac{\tau(1-\tau)\nu_0}{f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}) f_e^2(0|\bar{\mathbf{x}})} \right] \bar{W}(\bar{\mathbf{x}}) f_e(0|\bar{\mathbf{x}}) f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}) d\bar{\mathbf{x}},
$$

we readily have that

$$\mathrm{MSE}_L(\bar{h},\bar{\lambda}) = n^{-4/(d_1+4)} \cdot \mathrm{MSE}_L^\star(\bar{a},\bar{b}). \tag{3.6}$$

Let $h_s^0 = a_s^0 \cdot n^{-1/(d_1+4)}$, $s = 1,\cdots,d_1$, and $\lambda_s^0 = b_s^0 \cdot n^{-2/(d_1+4)}$, $s = 1,\cdots,d_3$, where $a^0 = \left(a_1^0,\cdots,a_{d_1}^0\right)$ and $b^0 = \left(b_1^0,\cdots,b_{d_3}^0\right)$ are the minimizers to $\mathrm{MSE}_L^\star(\bar{a},\bar{b})$. In the context of local constant mean regression estimation, Li and Zhou (2005) discuss some sufficient conditions for existence and uniqueness of a^0 and b^0. Their conditions are applicable to the nonparametric quantile regression setting with some minor modification. Letting $\{\bar{h}_s^*\}_{s=1}^{d_1}$, $\{\tilde{h}_s^*\}_{s=1}^{d_2}$, $\{\bar{\lambda}_s^*\}_{s=1}^{d_3}$ and $\{\tilde{\lambda}_s^*\}_{s=1}^{d_4}$ denote the CV selected smoothing parameters defined in Section 2, we next present their asymptotic optimality property.

THEOREM 3.1. *Suppose Assumptions 1–4 are satisfied, and $m_n = o(n^\iota)$ for any $\iota > 0$.*

(i) When $d_1 = 1$, we have

$$\frac{\bar{h}_1^* - h_1^0}{h_1^0} = O_P\left(m_n^2 n^{-1/10}\right), \tag{3.7}$$

$$\bar{\lambda}_s^* - \lambda_s^0 = O_P\left(m_n^2 n^{-1/2}\right), \quad s = 1,\cdots,d_3, \tag{3.8}$$

$$P\left(\tilde{h}_s^* > C\right) \to 1 \text{ for all } C > 0,\ s = 1,\cdots,d_2, \tag{3.9}$$

$$\tilde{\lambda}_s^* = 1 + o_P(1),\ s = 1,\cdots,d_4. \tag{3.10}$$

(ii) When $d_1 \geqslant 2$, we have (3.9), (3.10),

$$\frac{\bar{h}_s^* - h_s^0}{h_s^0} = O_P\left(m_n^{5/2} n^{-1/(d_1+4)}\right), \quad s = 1,\cdots,d_1, \tag{3.11}$$

$$\bar{\lambda}_s^* - \lambda_s^0 = O_P\left(m_n^{5/2} n^{-3/(d_1+4)}\right), \quad s = 1,\cdots,d_3. \tag{3.12}$$

REMARK 3.2. We prove Theorem 3.1 under the assumption that the irrelevant covariates satisfy (2.1). A weaker condition would be to assume that

$$\text{conditional on } (\bar{\mathbf{X}}^c, \bar{\mathbf{X}}^d)\text{, the covariables } (\tilde{\mathbf{X}}^c, \tilde{\mathbf{X}}^d) \text{ and } Y \text{ are independent.} \tag{3.13}$$

Though (3.13) is more appealing than (2.1), condition (3.13) creates technical hurdles, so we are only able to prove our main results under (2.1). However, simulations reported in Section 5 (see Table 3) show that our methodology works in finite samples under (3.13).

The convergence results in (3.7), (3.8), (3.11) and (3.12) show the asymptotic optimality of $\bar{h}^* = \left(\bar{h}_1^*,\cdots,\bar{h}_{d_1}^*\right)$ and $\bar{\lambda}^* = \left(\bar{\lambda}_1^*,\cdots,\bar{\lambda}_{d_3}^*\right)$ associated with the relevant continuous and discrete covariates, respectively. The convergence rates in Theorem 3.1 are comparable to those in the literature derived for optimal bandwidth selection in kernel density or mean regression estimation. For the case of $d_1 = 1$, letting $m_n = \sqrt{\log n}$, the convergence rate in (3.7) is close to the rates $O_P(n^{-1/10})$ obtained by Hall and Marron (1987) (for kernel density estimation) and Racine and Li (2004) (for kernel mean regression estimation). It is nearly optimal up to a logarithmic factor. The convergence rate in (3.8) is comparable to that in Theorem 2.2(i) of Racine and Li (2004). However, the convergence rates shown in (3.11) and (3.12) when $d_2 \geqslant 2$ are a bit slower than those in the literature. For example, Racine and Li (2004) obtain the rates $O_P(n^{-2/(d_1+4)})$ and $O_P(n^{-4/(d_1+4)})$ for the estimated smoothing parameters associated with the relevant continuous and discrete covariates, respectively. The slower convergence rates in Theorem 3.1(ii) are mainly due to the fact that the kernel quantile regression estimation does not have a closed form and the approximation rate in the uniform Bahadur presentation (see Lemma A.1 in Appendix A) affects the convergence rates of the CV selected optimal smoothing parameters. Similarly to some existing results in the context of conditional mean regression estimation with irrelevant covariates (e.g., Theorem 2.1 in Hall, Li and Racine, 2007), (3.9) and (3.10) indicate that the irrelevant continuous and discrete covariates can be smoothed out with probability approaching one, achieving the consistency of variable selection.

We next give the asymptotic distribution theory for the kernel quantile estimation with the data-dependent CV selected smoothing parameter vectors h^* and λ^* as in Li and Li (2010).

THEOREM 3.2. *Suppose Assumptions 1–4 are satisfied. Then, we have*

$$\sqrt{n\bar{H}^*}\left[\widehat{Q}_\tau(\mathbf{x}_0; h^*, \lambda^*) - Q_\tau(\bar{\mathbf{x}}_0) - b(\bar{\mathbf{x}}_0; \bar{h}^*, \bar{\lambda}^*)\right] \xrightarrow{d} N\left[0, \sigma_*^2(\bar{\mathbf{x}}_0)\right], \tag{3.14}$$

where $\bar{H}^* = \bar{h}_1^* \dots \bar{h}_{d_1}^*$, $b(\bar{\mathbf{x}}_0; \bar{h}^*, \bar{\lambda}^*)$ *is defined as in (3.2) but with* $\bar{\mathbf{X}}_i$, $\bar{h}$ *and* $\bar{\lambda}$ *replaced by* $\bar{\mathbf{x}}_0$, $\bar{h}^*$ *and* $\bar{\lambda}^*$, *respectively, and* $\sigma_*^2(\bar{\mathbf{x}}_0) = \frac{\tau(1-\tau)\nu_0}{f_e^2(0|\bar{\mathbf{x}}_0) f_{\bar{X}}(\bar{\mathbf{x}}_0)}$.

REMARK 3.3. Theorem 3.2 above extends the distribution result in Example 4.3 of Li and Li (2010) to a more general setting with mixed discrete and continuous regressors, and extends Theorem 2.2 in Hall, Li and Racine (2007) from mean regression to quantile regression. Although we use the full sample containing the irrelevant covariates in the kernel mean regression estimation procedure, the optimal smoothing parameters via the CV method could automatically smooth out the irrelevant covariates, making the developed quantile estimation weakly converge to $Q_\tau(\bar{\mathbf{x}}_0)$ (only dependent on the relevant covariates) with the conventional normal limit distribution.

4 Extension to High-Dimensional Setting

In practical applications, it may be the case that the number of candidate covariates in quantile regression is large or even exceeds the sample size. The so-called sparsity assumption is usually imposed on the model structure in order to develop feasible estimation and inferential methodologies. The sparsity assumption means that the number of significant covariates in high-dimensional quantile regression is relatively small (either fixed or divergent to infinity at a slow rate). Variable or feature selection in high-dimensional linear quantile regression has been extensively studied in the literature and various shrinkage and screening techniques have been introduced to identify these significant covariates(e.g., Belloni and Chernozhukov, 2011; Wang, Wu and Li, 2012; Fan, Fan and Barut, 2014; Zheng, Peng and He, 2015; Ma, Li and Tsai, 2017). For extensions to high-dimensional nonparametric quantile regression, we refer to Belloni, Chernozhukov and Fernández-Val (2011), He, Wang and Hong (2013) and Xia, Li and Fu (2018). In this section, we consider a general nonparametric quantile regression setting which contains high-dimensional mixed continuous and discrete covariates. To the best of our knowledge, this topic has not been tackled in the literature.

Recall that $\mathbf{X}_i^c$ and $\mathbf{X}_i^d$ denote the vectors of continuous and discrete covariates, respectively. Let X_{is}^c and X_{is}^d be the s-th element of $\mathbf{X}_i^c$ and $\mathbf{X}_i^d$, respectively. Note that the local quantile regression estimation method and the CV smoothing parameter selection criterion proposed in Section 2 are only applicable to the low-dimensional case, i.e., d_i, $1 \leqslant i \leqslant 4$, are fixed. When the dimension of covariates is large, we need to first screen out many irrelevant covariates and reduce the number of continuous and discrete covariates to a size which is feasible to implement the methods developed in Section 2. A natural idea is to rank the importance of each covariate by evaluating its marginal effect on the response. If Y_i and X_{is}^c are independent, we readily have that

$$Q_{\tau,s}^c(X_{is}^c) = Q_\tau \quad \text{a.s.}\ \forall\, 0 < \tau < 1,$$

where $Q_{\tau,s}^c(X_{is}^c)$ is the τ-th marginal quantile regression of Y_i given the s-th continuous covariate X_{is}^c, Q_τ is the τ-th unconditional quantile of Y_i and a.s. denotes "almost surely". Let $\widehat{Q}_{\tau,s}^c(x)$ be the local kernel estimate of $Q_{\tau,s}^c(x)$ obtained by minimizing the kernel-weighted objective function:

$$\mathcal{L}_{n,s}^c(\alpha) = \frac{1}{nb_1} \sum_{i=1}^{n} \rho_\tau(Y_i - \alpha) k\left(\frac{X_{is}^c - x}{b_1}\right) \tag{4.1}$$

with respect to α, where $k(\cdot)$ is a kernel function satisfying Assumption 4(i) and b_1 is a bandwidth. We then calculate the following weighted L_1-quantity for the s-th continuous covariate:

$$\widehat{D}^c_{\tau,s} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{Q}^c_{\tau,s}(X^c_{is}) - \widehat{Q}_\tau\right| w_s(X^c_{is}), \tag{4.2}$$

where $\widehat{Q}_\tau$ is the τ-th sample quantile function using only the response observations and $w_s(\cdot)$ is a univariate positive weight function trimming out boundary observations of the s-th continuous covariate. The construction in (4.2) is similar to that in He, Wang and Hong (2013) who use the sieve quantile estimation method and an L_2-distance measure. With $\widehat{D}^c_{\tau,s}$, we define the following index set which is the estimate of the index set containing all the indices corresponding to the significant continuous covariates in the τ-th quantile regression:

$$\widehat{\mathcal{M}}^c_\tau = \left\{1 \leqslant s \leqslant d_1 + d_2 : \ \widehat{D}^c_{\tau,s} \geqslant \gamma^c_n\right\}, \tag{4.3}$$

where γ^c_n is a pre-determined thresholding parameter.

The same screening procedure can also be applied to the discrete covariates. Let $Q^d_{\tau,s}(x)$ be the τ-th marginal quantile regression of Y_i given the s-th discrete covariate $X^d_{is} = x$. We estimate $Q^d_{\tau,s}(x)$ by $\widehat{Q}^d_{\tau,s}(x)$, which is obtained by minimizing

$$\mathcal{L}^d_{n,s}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(Y_i - \alpha)\cdot b_2^{I(X^d_{is}\neq x)} \tag{4.4}$$

with respect to α, where b_2 is a smoothing parameter. With the kernel quantile estimates $\widehat{Q}^d_{\tau,s}(X^d_{is})$, we may construct

$$\widehat{D}^d_{\tau,s} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{Q}^d_{\tau,s}(X^d_{is}) - \widehat{Q}_\tau\right|, \tag{4.5}$$

and consequently obtain the following estimated index set for significant discrete covariates:

$$\widehat{\mathcal{M}}^d_\tau = \left\{1 \leqslant s \leqslant d_3 + d_4 : \ \widehat{D}^d_{\tau,s} \geqslant \gamma^d_n\right\}, \tag{4.6}$$

where γ^d_n is a pre-specified thresholding parameter. In order to save computational burden in the above kernel screening procedure, we select the smoothing parameters via the rule of thumb, i.e., $b_1 = \alpha_1 \cdot n^{-1/5}$ and $b_2 = \alpha_2 \cdot n^{-2/5}$, where α_1 and α_2 are two positive constants. Such a choice of smoothing parameters is theoretically sensible following Theorem 3.1(i).

Let
$$D^c_{\tau,s} = \int_{\mathcal{S}_s} \left|Q^c_{\tau,s}(x) - Q_\tau\right| f^c_s(x) w_s(x) dx \tag{4.7}$$
and
$$D^d_{\tau,s} = \sum_{x_i \in \mathcal{D}_s} \left|Q^c_{\tau,s}(x_i) - Q_\tau\right| p^d_s(x_i), \tag{4.8}$$
where $f^c_s(\cdot)$ is the marginal density function of the s-th continuous covariate, $p^d_s(\cdot)$ is the probability mass function of the s-th discrete covariate, $\mathcal{S}_s$ and $\mathcal{D}_s$ denote the supports for the s-th continuous and discrete covariates, respectively. Throughout this section, we use $\mathcal{M}^c_\tau$ and $\mathcal{M}^d_\tau$ to denote the index sets for significant continuous and discrete covariates, respectively. In order to derive the well-known sure screening property, we need the following technical assumptions.

ASSUMPTION 5. *(i) For each* $s = 1, \cdots, d_1 + d_2$, *the marginal quantile regression function* $Q^c_{\tau,s}(x)$ *has continuous second-order derivative. In addition, their first and second-order derivative functions are bounded uniformly over* s.

(ii) For any $x \in \mathcal{S}_s$, $s = 1, \cdots, d_1 + d_2$, *the conditional density function of* $e^c_{is} \stackrel{\text{def}}{=} Y_i - Q^c_{\tau,s}(X^c_{is})$ *given* $X^c_{is} = x$, $f^c_{e,s}(\cdot|x)$, *is strictly positive and has continuous first derivative at point zero. For* ν *in a small neighborhood of* 0 *and* $s = 1, \cdots, d_1 + d_2$, $f^c_{e,s}(\nu|x)$ *is positive and continuous with respect to* x. *In addition, the marginal density function of* X^c_{is}, $f^c_s(\cdot)$, *is strictly positive and has continuous first derivative.*

(iii) For any $x \in \mathcal{D}_s$, $s = 1, \cdots, d_3 + d_4$, *the conditional density function of* $e^d_{is} \stackrel{\text{def}}{=} Y_i - Q^d_{\tau,s}(X^d_{is})$ *given* $X^d_{is} = x$, $f^d_{e,s}(\cdot|x)$, *is strictly positive and has continuous first derivative at point zero. In addition, when* ν *is in a small neighborhood of* 0, $Q^d_{\tau,s}(x)$ *and* $f^d_{e,s}(\nu|x)$ *are uniformly bounded over* $x \in \mathcal{D}_s$ *and* $s = 1, \cdots, d_3 + d_4$.

(iv) The weight functions $w_s(\cdot)$ *is bounded uniformly over* $s = 1, \cdots, d_1 + d_2$, *and, in addition,* $w_s(x) = 0$ *when* x *is in a given small neighborhood of the boundary points of* $\mathcal{S}_s$.

ASSUMPTION 6. *(i) There exists a positive constant* ν *such that* $d_i = O(n^\nu)$ *for* $i = 1, \cdots, 4$.

(ii) Letting $\varpi_n > 0$ *satisfy* $n^{-2/5}\sqrt{\log n} = o(\varpi_n)$,
$$\min_{s \in \mathcal{M}^c_\tau} D^c_{\tau,s} \geqslant \varpi_n, \quad \min_{s \in \mathcal{M}^d_\tau} D^d_{\tau,s} \geqslant \varpi_n.$$

REMARK 4.1. The smoothness conditions in Assumption 5 are similar to those in Assumptions 2 and 3, and are necessary to derive the uniform consistency of the marginal quantile regression estimation. Assumption 6(i) shows that the dimensions diverge to infinity at a polynomial rate of n, and may exceed the sample size when $\nu > 1$. In fact, by slightly modifying the proofs, the methodology and theory developed in this section are still applicable when d_i diverges at a slow exponential rate of n. Assumption 6(ii) is crucial to distinguish between the relevant and irrelevant covariates and allows $D^c_{\tau,s}$ and $D^d_{\tau,s}$ to be close to zero at an appropriate rate.

The following theorem gives the sure screening property, i.e., $\mathcal{M}^c_\tau \subset \widehat{\mathcal{M}}^c_\tau$ and $\mathcal{M}^d_\tau \subset \widehat{\mathcal{M}}^d_\tau$ hold with probability approaching one.

THEOREM 4.1. *Suppose that Assumption 1(i), 4(i), 5 and 6 are satisfied. Choosing* $\gamma^c_n = \gamma^d_n = \varpi_n/2$*, and letting* $b_1 = \alpha_1 \cdot n^{-1/5}$ *and* $b_2 = \alpha_2 \cdot n^{-2/5}$ *with* α_1 *and* α_2 *being two positive constants, we have*

$$\mathsf{P}\left(\mathcal{M}^c_\tau \subset \widehat{\mathcal{M}}^c_\tau,\ \mathcal{M}^d_\tau \subset \widehat{\mathcal{M}}^d_\tau\right) \to 1. \tag{4.9}$$

The above theorem complements some existing sure screening properties in high-dimensional quantile estimation (c.f., He, Wang and Hong, 2013; Ma, Li and Tsai, 2017; Xia, Li and Fu, 2018). An alternative kernel screening procedure is to conduct the leave-one-out kernel estimation for each marginal quantile regression and then use the data-driven CV method to determine the optimal smoothing parameter. From (3.9) and (3.10) in Theorem 3.1, if the optimal bandwidth for the continuous covariate exceeds a pre-determined sufficiently large positive constant or the optimal smoothing parameter for the discrete covariate is very close to 1, we expect that the corresponding covariate is irrelevant and should be removed. However, due to computational burden of implementing the CV method, such a screening method would be very time-consuming in particular when the dimension of the candidate covariates is very large.

5 Monte-Carlo Studies

In this section, we use Monte-Carlo simulations to investigate the finite-sample performance of the methods proposed in Sections 2 and 4, and compare our methods with some existing methods. We first examine the numerical performance of some kernel-based quantile estimation methods when the dimension of covariates is fixed, followed by the performance of kernel-based dimension reduction and estimation with high-dimensional covariates, and finally compare our modelling method with a semiparametric partially linear modelling method.

5.1 Low-Dimensional Nonparametric Quantile Estimation

Consider the following data generating process

$$\textbf{DGP1}: \qquad Y_i = 3\cos(\bar{X}_i^c) + \frac{1}{2}\bar{X}_i^d + \frac{1}{2}(\bar{X}_i^c)^2 \cdot u_i, \;\; i = 1, \cdots, n,$$

where $\bar{X}_i^c \sim \text{Uniform}(-2, 2)$, $\bar{X}_i^d \sim B(2, 0.5)$ (sum of 2 Bernoulli trials with success probability 0.5 for each trial), i.e., $\bar{X}_i^d \in \{0, 1, 2\}$ with $P(\bar{X}_i^d = 0) = 0.5^2 = 1/4$, $P(\bar{X}_i^d = 1) = 2(0.5)^2 = 1/2$, $P(\bar{X}_i^d = 2) = (0.5^2) = 1/4$. We consider two distributions for the error term u_i: the standard normal distribution $N(0, 1)$, and the student's t-distribution with 5 degrees of freedom denoted by $t(5)$. The conditional quantiles to be estimated in our simulation are at $\tau = 0.10, 0.25, 0.50, 0.75$ and 0.90. The sample sizes are $n = 100$, 200, 400, and the number of replications for each setup is 1000.

In this simulation study, we compare our method proposed in Section 2 with the traditional check-function-based kernel quantile estimation which only smoothes the continuous covariate $\bar{X}_i^c$ (thus splitting the full sample into cells according to the three different values of the discrete covariate $\bar{X}_i^d$), and the nonparametric inverted-CDF estimation with the bandwidths chosen by the method suggested in Li, Lin and Racine (2013). Tables 1 and 2 (corresponding to the standard normal distribution and t-distribution for u_i, respectively) report the simulation results of the average MSE over 1000 replications under DGP1. For each panel in the two tables, the first row reports the results of the proposed estimator that smoothes both the continuous and discrete covariates (denote it as "Check (smooth)"); the second row gives the results of the check-function-based conditional quantile estimator that does not smooth the discrete covariate (denote it as "Check (non-smooth)"); and the third row presents the results of the inverted-CDF approach introduced by Li, Lin and Racine (2013) (denote it as "Inverted-CDF"). From Tables 1 and 2, we find that our method that smoothes over both the continuous and discrete covariates performs significantly better than the naive method which only smoothes the continuous covariate but not the discrete covariate. This is similar to the finding in the context of conditional mean function estimation (c.f., Hall, Li and Racine, 2007). The main reason is that smoothing a discrete covariate can borrow the data information from neighborhoods to reduce estimation variance while introducing only mild estimation bias. Consequently, the finite-sample MSE can be reduced. Meanwhile, we also find from Tables 1 and 2 that the proposed estimation method with the CV selected smoothing parameters outperforms the inverted-CDF method, especially at the extreme quantiles, analogous to the findings in Li, Li and Li (2018).

Table 1: Average MSE in DGP1 with normal distribution errors

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			$n = 100$		
Check (smooth)	0.486	0.306	0.223	0.237	0.281
Check (non-smooth)	0.658	0.425	0.299	0.296	0.297
Inverted-CDF	0.529	0.339	0.240	0.244	0.290
			$n = 200$		
Check (smooth)	0.342	0.205	0.144	0.148	0.201
Check (non-smooth)	0.423	0.260	0.182	0.175	0.208
Inverted-CDF	0.368	0.215	0.157	0.164	0.237
			$n = 400$		
Check (smooth)	0.233	0.129	0.086	0.094	0.118
Check (non-smooth)	0.260	0.144	0.101	0.104	0.122
Inverted-CDF	0.237	0.130	0.098	0.107	0.160

5.2 Nonparametric Dimension Reduction and Estimation

We next examine the numerical performance of the proposed nonparametric dimension reduction methods in both low- and high-dimensional settings. In the low-dimensional setting, we show the ability of the proposed CV method to smooth out the irrelevant covariates; and in the high-dimensional setting, we demonstrate that the proposed kernel

Table 2: Average MSE in DGP1 with t-Distribution errors

Method	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
			$n = 100$		
Check (smooth)	0.845	0.397	0.279	0.323	0.515
Check (non-smooth)	1.151	0.595	0.396	0.429	0.580
Inverted-CDF	0.946	0.466	0.289	0.375	0.691
			$n = 200$		
Check (smooth)	0.569	0.238	0.162	0.184	0.314
Check (non-smooth)	0.808	0.325	0.205	0.231	0.362
Inverted-CDF	0.732	0.262	0.168	0.229	0.539
			$n = 400$		
Check (smooth)	0.392	0.159	0.096	0.118	0.190
Check (non-smooth)	0.509	0.182	0.114	0.134	0.199
Inverted-CDF	0.463	0.161	0.104	0.134	0.352

screening method can correctly identify the relevant covariates with high probability. We consider the following data generating process

$$\textbf{DGP2}: \qquad Y_i = 2\ln(1 + (\bar{X}_i^c)^2 + 2\bar{X}_i^d) + \exp(\bar{X}_i^d - (\bar{X}_i^c)^2)u_i, \;\; i = 1, \cdots, n,$$

where $\bar{X}_i^c \sim \text{Uniform}(-2, 2)$, $\bar{X}_i^d \sim B(1, 0.5)$ (1 Bernoulli trial with success probability 0.5), i.e., $\bar{X}_i^d \in \{0, 1\}$ with $P(\bar{X}_i^d = 0) = P(\bar{X}_i^d = 1) = 0.5$, and $u_i \sim N(0, 1)$. In addition to the relevant variables, we add two irrelevant variables into our dataset: $\tilde{X}_i^c$ and $\tilde{X}_i^d$, following the same distributions with $\bar{X}_i^c$ and $\bar{X}_i^d$, respectively. We consider two situations: (i) the four covariates, $\bar{X}_i^c, \bar{X}_i^d, \tilde{X}_i^c, \tilde{X}_i^d$, are independent with each other; (ii) $\bar{X}_i^d$ and $\tilde{X}_i^d$ are independent with other covariates but $\bar{X}_i^c$ and $\tilde{X}_i^c$ are correlated with correlation coefficient 0.5. We consider three sample sizes, $n = 100$, 200, 400, and the number of replications is 500. The conditional quantile regression functions are estimated at $\tau = 0.25$, 0.50 and 0.75.

We use the CV method to select the optimal smoothing parameters $\bar{h}^*, \tilde{h}^*, \bar{\lambda}^*, \tilde{\lambda}^*$ for $\bar{X}_i^c, \tilde{X}_i^c, \bar{X}_i^d, \tilde{X}_i^d$, respectively. Table 3 reports both means and standard deviations (in parenthesis) of the CV-selected bandwidths over 500 replications for $\tau = 0.5$. The upper block of Table 3 corresponds to situation (i) when $\bar{X}_i^c$ and $\tilde{X}_i^c$ are independent, while the lower block of Table 3 corresponds to situation (ii) when $\bar{X}_i^c$ and $\tilde{X}_i^c$ are correlated. For both cases (i) and (ii), we observe that $\bar{h}^*$ and $\bar{\lambda}^*$ decrease to 0 as sample size increases, while $\tilde{h}^*$ diverges and $\tilde{\lambda}^*$ approaches 1. Thus, our simulations suggest that the CV method can detect and remove irrelevant covariates under the weak condition (3.13) even though we can only prove the theory under assumption (2.1). When $\tau = 0.25$ or 0.75, the results on the CV-selected bandwidths are similar to those in Table 3, and thus we omit them to save space.

Table 3: Mean and standard deviation of the CV-selected bandwidths

	$\bar{h}^*$	$\tilde{h}^*$	$\bar{\lambda}^*$	$\tilde{\lambda}^*$
		$\bar{X}_i^c$ and $\tilde{X}_i^c$ are independent		
$n = 100$	0.22 (0.10)	73.10 (71.35)	0.23 (0.22)	0.68 (0.31)
$n = 200$	0.20 (0.07)	115.84 (108.95)	0.13 (0.10)	0.72 (0.30)
$n = 400$	0.16 (0.06)	144.31 (130.69)	0.06 (0.05)	0.75 (0.27)
		$\bar{X}_i^c$ and $\tilde{X}_i^c$ are correlated		
$n = 100$	0.21 (0.09)	63.21 (62.06)	0.20 (0.17)	0.67 (0.29)
$n = 200$	0.19 (0.07)	102.44 (95.60)	0.12 (0.10)	0.70 (0.30)
$n = 400$	0.15 (0.06)	137.98 (126.95)	0.06 (0.06)	0.77 (0.25)

In order to conduct high-dimensional variable selection in simulation, we next consider DGP2 in situation (i), but replace one irrelevant continuous variable $\tilde{X}_i^c$ in the above simulation by a set of 50 i.i.d. irrelevant continuous variables $\left\{\tilde{X}_{i,1}^c, \tilde{X}_{i,2}^c, \cdots, \tilde{X}_{i,50}^c\right\}$ and replace one irrelevant categorical variable $\tilde{X}_i^d$ by a set of 50 i.i.d. irrelevant categorical variables $\left\{\tilde{X}_{i,1}^d, \tilde{X}_{i,2}^d, \cdots, \tilde{X}_{i,50}^d\right\}$. We employ the kernel screening method proposed in Section 4 to rank the importance of the continuous/categorical variables, according to the values of $\widehat{D}_{\tau,s}^c$ and $\widehat{D}_{\tau,s}^d$ defined in (4.2) and (4.5), respectively. To evaluate the finite-sample performance of the proposed screening method, we compute the following two frequencies (out of 500 replications): (i) the true significant continuous/categorical covariate ranks as "first" among all continuous/categorical variables; (ii) the true significant continuous/categorical covariate ranks as "top 2" among all continuous/categorical variables.

Table 4 reports the relevant results. The upper and lower panels correspond to the continuous and categorical variable selection, respectively. From the table, we find that as the sample size increases, the frequency of detecting significant covariates approaches 1, supporting the sure screening property derived in Section 4. Note that the marginal effects of $\bar{X}_i^c$ and $\bar{X}_i^d$ on the conditional quantile depend on the quantile τ. Among the three quantiles (0.25, 0.50 and 0.75) being estimated, $\bar{X}_i^c$ has the largest absolute marginal effect when $\tau = 0.25$, whereas $\bar{X}_i^d$ has the largest absolute marginal effect when $\tau = 0.75$.

Table 4: Frequencies of selecting significant covariates

	$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.75$	
	Continuous covariate selection					
	First	Top 2	First	Top 2	First	Top 2
$n = 100$	0.994	1	0.994	1	0.672	0.754
$n = 200$	1	1	1	1	0.914	0.952
$n = 400$	1	1	1	1	0.996	0.998
	Categorical covariate selection					
	First	Top 2	First	Top 2	First	Top 2
$n = 100$	0.850	0.898	0.988	0.996	0.998	1
$n = 200$	0.986	0.998	1	1	1	1
$n = 400$	1	1	1	1	1	1

5.3 Comparison with Semiparametric Quantile Regression

We next compare our kernel-based nonparametric quantile regression method with the semiparametric partially linear quantile regression (c.f., Cai and Xiao, 2012). When the categorical variables enter into the DGP in the additive and linear form, we may write the quantile regression function as

$$Q_\tau(\mathbf{X}) = Q_\tau^c(\mathbf{X}^c) + \beta_\tau \mathbf{X}^d,$$

where β_τ is a vector of unknown coefficients for $\mathbf{X}^d$. The above partially linear model structure enables us to apply some commonly-used variable selection methods, such as LASSO, to deal with irrelevant categorical variables. We expect that the partially linear quantile regression is more efficient than the nonparametric quantile regression when the partially linear model assumption holds, since it utilizes the semiparametric functional structure. However, if the partially linear assumption fails, the semiparametric partially linear quantile regression estimation becomes inconsistent.

We consider the following two DGPs:

$$\begin{aligned} \mathbf{DGP3}: &\qquad Y_i = \sin\left[(\bar{X}_i^c)^2\right]\bar{X}_i^d + u_i,\ \ i = 1, \cdots, n, \\ \mathbf{DGP4}: &\qquad Y_i = \sin\left[(\bar{X}_i^c)^2\right] + \bar{X}_i^d + u_i,\ \ i = 1, \cdots, n, \end{aligned}$$

where $\bar{X}_i^c \sim \text{Uniform}(-2, 2)$, $\bar{X}_i^d \in \{-1, 1\}$ with $P(\bar{X}_i^d = -1) = P(\bar{X}_i^d = 1) = 0.5$, and $u_i \sim N(0, 1)$. DGP3 has a non-separable regression form while DGP4 has a partially linear structure. There are 50 irrelevant categorical variables $\tilde{X}_{i,1}^d, \tilde{X}_{i,2}^d, \cdots, \tilde{X}_{i,50}^d$, independently following the same distribution as $\bar{X}_i^d$. There is no irrelevant continuous covariate involved. As in the previous two subsections, the sample sizes are $n = 100$, 200 and 400. The conditional quantile is estimated at $\tau = 0.5$, and the number of replications is 500.

We compare the performance between (i) partial linear quantile regression combined with LASSO variable selection and (ii) our nonparametric quantile regression combined with kernel-based screening method. The first step is to reduce the dimension of categorical variables from 51 to 2, using LASSO and the screening method, respectively. The second step is to estimate the conditional quantile regression based on the low dimensional dataset, using the semiparametric partially linear quantile regression and our nonparametric quantile regression, respectively.

Table 5 reports the MSEs for both the nonparametric and semiparametric quantile regression estimation. The proposed nonparametric method performs similarly between DGP3 and DGP4 with the MSEs decreasing as the sample size increases. For DGP3, the semiparametric partially linear quantile estimation method has much larger estimation MSE than the nonparametric method (especially when n is large), which is not surprising as the partially linear model is misspecified in DGP3. In contrast, for DGP4, the semiparametric partially linear method outperforms the nonparametric method, indicating that the correct semiparametric functional structure helps improve estimation efficiency in finite samples.

Table 5: MSE comparison between nonparametric and semiparametric partially linear methods

	DGP3		DGP4	
	Nonparametric	Semiparametric	Nonparametric	Semiparametric
$n = 100$	0.259	0.390	0.219	0.194
$n = 200$	0.147	0.289	0.137	0.101
$n = 400$	0.079	0.244	0.079	0.053

All the simulated data in this section are generated following some location-scale conditional quantile functions whose representation may be restrictive. Alternatively, as suggested by a referee, one can use the so-called Skorohod's representation to define a (possibly) non-separable nonlinear conditional quantile regression function, making use of the equivalent representation: $F(Y|\mathbf{X}) = U|\mathbf{X}$ with U being distributed as uniform $[0, 1]$. In fact, through some small-scale simulations with data generated via the Skorohod's representation, we obtain numerical results similar to those in Section 5.1. To save space, we do not report the detailed results in the paper. They are available from the authors upon request.

6 An Empirical Application

In this section, we apply the proposed methods to study the effect of dating experience on wages for men and women. The gender gap in wages has drawn significant attention from economists, and has been extensively studied in the literature. Blau and Kahn (1996) analyze microdata from ten industrialized nations and claim that the wage structure plays an important role in gender gap. O'Neill and Polachek (1993) examine the factors underlying the narrowing gender gap during the 1980s and find that around one third to one half of the narrowing can be explained by the converging work-related characteristics. Bagues and Esteve-Volart (2010) study the gender composition of recruiting committee and find that a job candidate's probability to be hired is negatively affected if the majority of the committee members are within the same gender membership as the candidate. The gender gap is due to the fundamental difference between men and women. This empirical application aims to shed some lights on such a difference.

We use the data from the National Longitudinal Survey of Youth 1997 (NLSY97) in our empirical study. NLSY97 is a nationally representative data set of approximately 9000 American youths aged between 12 and 17 years when first interviewed in 1997. The survey interviewed these youths annually from 1997 to 2011 and biennially after 2011. The survey includes standard demographic information and ASVAB[1] math and verbal score percentile information. It also asks whether respondents had been on a date during 2007 to 2008 (when the respondents were between 23 and 26 years old and not married), their total income in 2013, and their total working hours in 2013. With the survey information, we generate the continuous response variable, hourly wage (in 2013), and the discrete explanatory variable, ever date (during 2007 to 2008). The continuous explanatory variable, ability, is measured by the ASVAB percentile. Because income is highly correlated with age, we restrict our sample to respondents who were 29, 30, and 31 years old by the (survey) year of 2013. The sample size is 765. In order to isolate the dating effect, we apply our method separately to men and women. Table 6 presents the summary statistics of the variables: hourly wage, ever date, ability and gender.

Table 6: Summary statistics for the real data

Variables	Hourly Wage	Ever Date	Ability	Gender
Mean	18.73	0.52	48.13	0.55
Standard Deviation	16.29	0.50	30.15	0.50

We estimate a conditional median function ($\tau = 0.5$) of hourly wage given the covari-

[1]The Armed Services Vocational Aptitude Battery (ASVAB) measures the respondent's knowledge and skills in the topical areas including math and reading.

ates ability and ever date for men and women separately. First we use the proposed CV method to select the smoothing parameters with the result given in Table 7, where $\hat{c}^*$ is related to $\hat{h}^*$ via $\hat{h}^* = \hat{c}^* s_a n^{-1/5}$ with s_a being the sample standard error of the continuous covariate ability. In Table 7, we can see that for women, the ever date covariate is smoothed out, indicating that women's median income in 2013 does not depend on whether a woman had the dating experience during 2007 to 2008 or not. However, there is a different story for men: the ever date covariate is not smoothed out, so it is deemed to be a relevant covariate for determining men's conditional median income. These results show that the dating experience plays a significant role in determining wages for men in later years but not for women.

Table 7: The CV selected smoothing parameters

Men		Women	
Covariate	Bandwidth	Covariate	Bandwidth
Ability	$\hat{c}^* = 1.92$	Ability	$\hat{c}^* = 1.18$
Ever Date	$\hat{\lambda}^* = 0.55$	Ever Date	$\hat{\lambda}^* = 1$

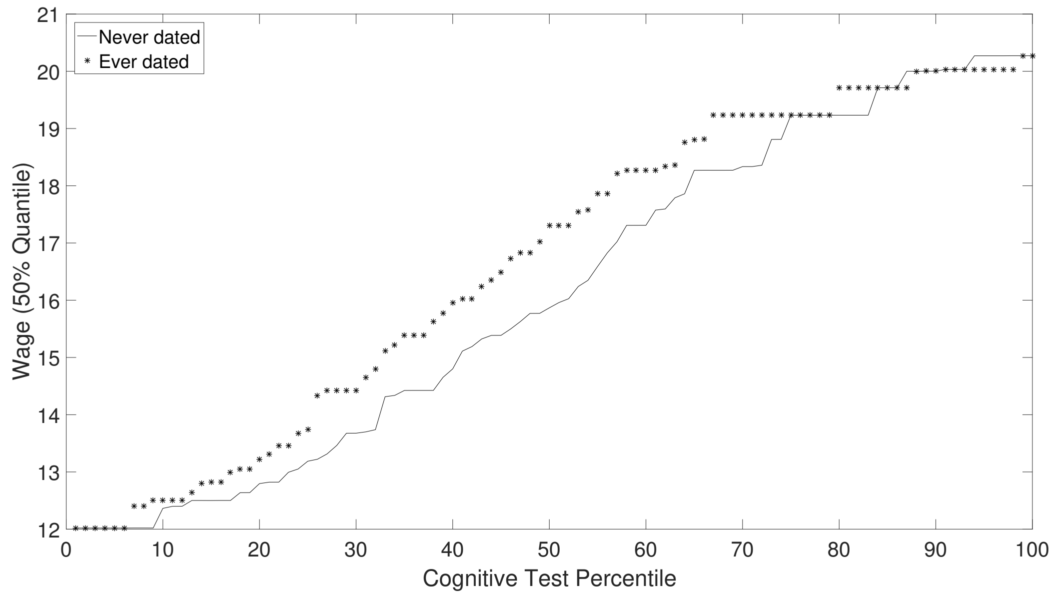

Figure 1: Median income for men

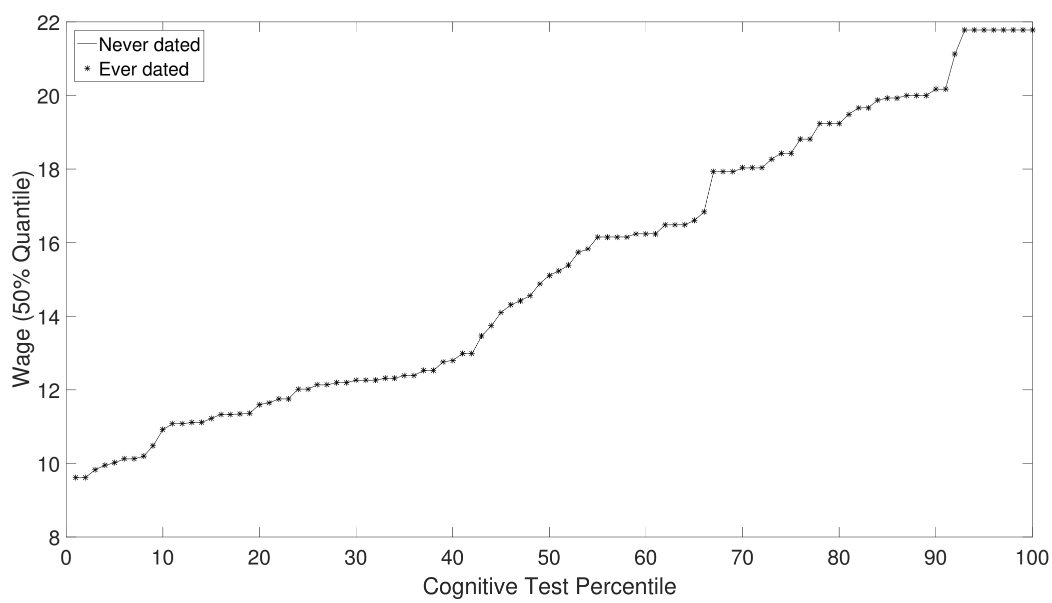

Figure 2: Median income for women

Figures 1 and 2 show the estimated median wage for both men and women with the smoothing parameters selected by the CV method. For men, we see that individuals with dating experience have higher wages than those without dating experience, and the wage gap between ever-dating and never-dating is large when the cognitive test percentile is around 50%. As the cognitive test percentile approaches 100%, the gap tends to vanish. The fact that the ever-dating covariate plays a positive role in determining men' median wages suggests that the male individuals with dating experience are likely to be more sociable or more out-going than those without such experience, and these characteristics are positively correlated with wage. Thus, the ever date covariate may serve as a proxy for an individual's sociability and can be used to help predict men's future wage. For women, because the CV selected smoothing parameter for the ever date covariate takes an upper bound of 1, it implies that the ever date covariate is unrelated to women's median wage. Therefore, the two curves for ever-dating and never-dating coincide with each other (they become one identical curve) for women. Finally as expected, for both men and women, we see an upward-sloping curve for median wage versus the cognitive ability.

7 Conclusions

In this paper, we study the problem of nonparametrically estimating a conditional quantile regression function, where the covariates include both continuous and discrete components. Unlike the recent paper by Li, Li and Li (2018), our paper allows the presence of irrelevant discrete and continuous covariates. We combine the quantile check function and the local smoothing technique with the mixed continuous and discrete kernel functions to directly estimate the conditional quantile function. In order to select the optimal smoothing parameters, we use the data-driven CV method, which can also automatically detect and remove the irrelevant covariates by over-smoothing them. The CV selected smoothing parameters are proved to be asymptotically optimal (with convergence rates) and the irrelevant covariates can be smoothed out asymptotically (with probability approaching one). Furthermore, we establish the asymptotic normal distribution theory for the proposed conditional quantile estimator with data-dependent smoothing parameters, generalizing the existing results that only deal with the case of relevant covariates in quantile regression. In the high-dimensional setting when the number of covariates is comparable to (or exceeds) the sample size, we suggest using a kernel-based quantile screening method to remove the irrelevant continuous and discrete covariates and then apply the CV method to those that survive the kernel screening procedure. Simulation studies provide a numerical examination of the finite-sample behavior of the proposed method as well as its comparison with some existing methods. An empirical application using the NLSY97 data to study the relationship between dating experience and median income suggests that women's dating experience is independent of their median wage as the dating experience covariate is automatically removed by the data-driven CV method.

Acknowledgements

The authors would like to thank a co-editor, an associate editor and two reviewers for their valuable comments, which substantially improved our paper.

Appendix A: Proofs of the Asymptotic Results

In this appendix, we give the detailed proofs of the main asymptotic results in Sections 3 and 4. The proofs of the technical lemmas are available in a supplemental document. Throughout the proof, we use $a_n \approx b_n$ to denote that $a_n = b_n(1 + o(1))$. For notational simplicity, we let $\mathcal{K}_{h,\lambda}(\mathbf{X}_i, \mathbf{x}) = K_h(\mathbf{X}_i^c - \mathbf{x}^c)\Lambda_\lambda(\mathbf{X}_i^d, \mathbf{x}^d)$, $\bar{H} = \prod_{s=1}^{d_1} \bar{h}_s$ and $\tilde{H} = \prod_{s=1}^{d_2}(\tilde{h}_s \vee 1)$, where $\vee$ denotes maximum. We start with two technical lemmas, which are key to the proof of Theorem 3.1. The first lemma gives the Bahadur representation for the kernel quantile regression estimation uniformly over $\mathbf{x}$ and (h, λ), which is of independent interest and complements the results derived by Su and White (2012) and Kong and Xia (2017) both of which only consider the case of continuous regressors.

LEMMA A.1. *Let $Q_\tau(\bar{\mathbf{x}})$ be the conditional τ-quantile regression function evaluated at $\bar{\mathbf{x}}$ and $\widehat{Q}_\tau(\mathbf{x}; h, \lambda)$ be the corresponding local kernel estimate using the smoothing parameters h and λ. Suppose that Assumptions 1, 2, 3(i) and 4 are satisfied. Then we have*

$$(nH)^{1/2}\left[\widehat{Q}_\tau(\mathbf{x}; h, \lambda) - Q_\tau(\bar{\mathbf{x}})\right] = \left[V(\mathbf{x}; \tilde{h}, \tilde{\lambda})\right]^{-1}\left[(nH)^{1/2}\tilde{H} \cdot U_n(\mathbf{x}; h, \lambda)\right] + O_P(m_n^{3/2}(nH)^{-1/4}) \tag{A.1}$$

uniformly over $\mathbf{x} \in \mathcal{S}^\star \times \mathcal{D}$ and (h, λ) satisfying Assumption 4(ii)–(iv), where $\mathcal{S}^\star \subset \mathcal{S}$ such that

$W(\mathbf{x}^c, \mathbf{x}^d) \neq 0$ *for* $\mathbf{x}^c \in \mathcal{S}^\star$, m_n *and* H *are defined as in Assumption 4(iii),*

$$V(\mathbf{x}; \tilde{h}, \tilde{\lambda}) = f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}}) f_e(0|\bar{\mathbf{x}}) \cdot \mathsf{E}\left[\tilde{H} \cdot \tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_i, \tilde{\mathbf{x}})\right]$$

with $\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_i, \tilde{\mathbf{x}}) = \prod_{s=1}^{d_2} \tilde{h}_s^{-1} k\left(\frac{\tilde{X}_{is}^c - \tilde{x}_{0s}^c}{\tilde{h}_s}\right) \prod_{s=1}^{d_4} \tilde{\lambda}_s^{I\left(\tilde{X}_{is}^d \neq \tilde{x}_{0s}^d\right)}$, *and*

$$U_n(\mathbf{x}; h, \lambda) = \frac{1}{n} \sum_{i=1}^n \eta_i(\bar{\mathbf{x}}) \mathcal{K}_{h,\lambda}(\mathbf{X}_i, \mathbf{x}),$$

with $\eta_i(\bar{\mathbf{x}}) = \tau - I\left(Y_i - Q_\tau(\bar{\mathbf{x}}) < 0\right)$.

LEMMA A.2. *Suppose that Assumptions 1–4 are satisfied. Then, we have*

$$\begin{aligned} CV(h, \lambda) &= CV_1 + \frac{1}{2n} \sum_{i=1}^n \left[b^2(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda}) + \sigma_\diamond^2(\mathbf{X}_i; h, \lambda)\right] W(\mathbf{X}_i) f_e(0|\bar{\mathbf{X}}_i) \\ &\quad + O_P\left(m_n^{5/2}/(nH)^{5/4} + m_n^2/(nH^{1/2})\right) \end{aligned} \tag{A.2}$$

uniformly over (h, λ) *satisfying Assumption 4(ii)–(iv), where*

$$CV_1 \stackrel{\mathrm{def}}{=} \frac{1}{n} \sum_{i=1}^n \rho_\tau(e_i) W(\mathbf{X}_i)$$

is unrelated to the smoothing parameters h *and* λ, $b^2(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda})$ *is defined in (3.2),*

$$\sigma_\diamond^2(\mathbf{X}_i; h, \lambda) = \frac{1}{n\bar{H}} \cdot \frac{\tau(1-\tau)\nu_0}{f_{\bar{\mathbf{X}}}(\bar{\mathbf{X}}_i) f_e^2(0|\bar{\mathbf{X}}_i)} \cdot R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda})$$

with $R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda}) = \mathsf{E}\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}^2(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right] / \mathsf{E}^2\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right]$.

Using the above two lemmas, we next prove the main theoretical results in Section 3.

PROOF OF THEOREM 3.1. Note that $\sigma_\diamond^2(\mathbf{X}_i; h, \lambda) = \sigma^2(\bar{\mathbf{X}}_i, \bar{h}) R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda})$. The smoothing parameters for the irrelevant covariates, $\tilde{h}$ and $\tilde{\lambda}$, only appear in the term $R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda})$. Since $\sigma^2(\bar{\mathbf{X}}_i, \bar{h})$ is always non-negative, to minimize $\sigma^2(\mathbf{X}_i; h, \lambda)$, we first choose $\tilde{h}$ and $\tilde{\lambda}$ to minimize $R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda})$. Recall that $R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda}) = \mathsf{E}\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}^2(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right] / \mathsf{E}^2\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right]$, and note that $\mathsf{E}\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}^2(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right]$ in the numerator is the conditional expectation of $\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}^2(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)$, while $\mathsf{E}^2\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right]$ in the denominator is the squared conditional expectation of $\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)$. It is straightforward to show that

$$\mathsf{E}\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}^2(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right] \geqslant \mathsf{E}^2\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j, \tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right] \quad \text{a.s.},$$

indicating that

$$R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda}) \geqslant 1 \ \text{a.s.}$$

uniformly over i. It is easy to see that $R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda})$ reaches the minimum value 1 if and only if $\tilde{h}_s \to \infty$ for all $s = 1, \cdots, d_2$, and $\tilde{\lambda}_s \to 1$ for all $s = 1, \cdots, d_4$ as $n \to \infty$, which are feasible due to Assumption 4(ii)(iv). Therefore, we prove (3.9) and (3.10) in Theorem 3.1.

With (3.9) and (3.10), we replace $R(\tilde{\mathbf{X}}_i; \tilde{h}, \tilde{\lambda})$ and H by 1 and $\bar{H}$, respectively, in the subsequent proof. By Lemma A.2, we have

$$\mathrm{CV}(h, \lambda) = \mathrm{CV}_1 + \frac{1}{2}\mathrm{CV}^\star(\bar{h}, \bar{\lambda}) + O_P(\chi_n) \tag{A.3}$$

uniformly over $(\bar{h}, \bar{\lambda})$ satisfying Assumption 4(ii)–(iv), where $\chi_n = m_n^{5/2}/(n\bar{H})^{5/4} + m_n^2/(n\bar{H}^{1/2})$,

$$\mathrm{CV}^\star(\bar{h}, \bar{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\left[b^2(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{X}}_i; \bar{h})\right] W(\mathbf{X}_i) f_e(0|\bar{\mathbf{X}}_i).$$

Letting $\kappa_n = \sum_{s=1}^{d_1} \bar{h}_s^2 + \sum_{s=1}^{d_3} \bar{\lambda}_s$, we note that

$$\begin{aligned}
\mathrm{CV}^\star(\bar{h}, \bar{\lambda}) &= \mathrm{E}\left\{\left[b^2(\bar{\mathbf{X}}_i; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{X}}_i; \bar{h})\right] W(\mathbf{X}_i) f_e(0|\bar{\mathbf{X}}_i)\right\} + O_P\left(n^{-1/2}\left(\kappa_n^2 + (n\bar{H})^{-1}\right)\right) \\
&= \int_{\mathcal{S}\times\mathcal{D}} \left[b^2(\bar{\mathbf{x}}; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{x}}; \bar{h})\right] W(\mathbf{x}) f_e(0|\bar{\mathbf{x}}) f_X(\mathbf{x}) d\mathbf{x} + o_P(\chi_n) \\
&= \int_{\bar{\mathcal{S}}\times\bar{\mathcal{D}}} \left[b^2(\bar{\mathbf{x}}; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{x}}; \bar{h})\right] f_e(0|\bar{\mathbf{x}}) f_{\bar{X}}(\bar{\mathbf{x}}) \left[\int_{\tilde{\mathcal{S}}\times\tilde{\mathcal{D}}} W(\mathbf{x}) f_{\tilde{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}\right] d\bar{\mathbf{x}} + o_P(\chi_n) \\
&= \int_{\bar{\mathcal{S}}\times\bar{\mathcal{D}}} \left[b^2(\bar{\mathbf{x}}; \bar{h}, \bar{\lambda}) + \sigma^2(\bar{\mathbf{x}}; \bar{h})\right] \bar{W}(\bar{\mathbf{x}}) f_e(0|\bar{\mathbf{x}}) f_{\bar{X}}(\bar{\mathbf{x}}) d\bar{\mathbf{x}} + o_P(\chi_n) \\
&\stackrel{\text{def}}{=} \mathrm{MSE}_L(\bar{h}, \bar{\lambda}) + o_P(\chi_n).
\end{aligned} \tag{A.4}$$

By (A.3) and (A.4), we readily have that

$$\mathrm{CV}(h, \lambda) = \mathrm{CV}_1 + \frac{1}{2}\mathrm{MSE}_L(\bar{h}, \bar{\lambda}) + O_P\left(m_n^{5/2}/(n\bar{H})^{5/4} + m_n^2/(n\bar{H}^{1/2})\right) \tag{A.5}$$

uniformly over $(\bar{h}, \bar{\lambda})$ satisfying Assumption 4(ii)–(iv). This shows that the CV selected smoothing parameters $\bar{h}^*$ and $\bar{\lambda}^*$ asymptotically minimize $\mathrm{MSE}_L(\bar{h}, \bar{\lambda})$ as CV_1 is not related to the smoothing parameters. When $d = 1$, as $m_n^{5/2}/(n\bar{H})^{5/4} = o\left(m_n^2/(n\bar{H}^{1/2})\right)$ with $\bar{h}_1 \asymp h_1^0$, using (A.5), we have

$$\mathrm{CV}(h, \lambda) = \mathrm{CV}_1 + \frac{1}{2}\mathrm{MSE}_L(\bar{h}, \bar{\lambda}) + O_P\left(m_n^2 \bar{H}^{1/2} \cdot \mathrm{MSE}_L(\bar{h}, \bar{\lambda})\right). \tag{A.6}$$

From (3.6) and (A.6), we may show that when $d_1 = 1$,

$$\bar{h}_1^* - h_1^0 = O_P\left(h_1^0 \cdot m_n^2 n^{-1/10}\right); \tag{A.7}$$

and

$$\bar{\lambda}_s^* - \lambda_s^0 = O_P\left(m_n^2 n^{-1/2}\right), \quad s = 1, \cdots, d_3. \tag{A.8}$$

When $d \geqslant 2$, as $m_n \to \infty$, we have $m_n^2/(n\bar{H}^{1/2}) = o\left(m_n^{5/2}/(n\bar{H})^{5/4}\right)$ with $\bar{h}_s \approx h_s^0$, which indicates that

$$CV(h, \lambda) = CV_1 + \frac{1}{2}MSE_L(\bar{h}, \bar{\lambda}) + O_P\left(m_n^{5/2}(n\bar{H})^{-1/4} \cdot MSE_L(\bar{h}, \bar{\lambda})\right). \tag{A.9}$$

From (3.6) and (A.9), we can prove that when $d_1 \geqslant 2$,

$$\bar{h}_s^* - h_s^0 = O_P\left(h_s^0 \cdot m_n^{5/2} n^{-1/(d_1+4)}\right), \quad s = 1, \cdots, d_1; \tag{A.10}$$

and

$$\bar{\lambda}_s^* - \lambda_s^0 = O_P(m_n^{5/2} n^{-3/(d_1+4)}), \quad s = 1, \cdots, d_3. \tag{A.11}$$

By (A.7), (A.8), (A.10) and (A.11), we can complete the proof of Theorem 3.1. □

PROOF OF THEOREM 3.2. From the stochastic equicontinuity argument in Li and Li (2010) and Theorem 3.1, it suffices to prove Theorem 3.2 with the CV selected smoothing parameters $\bar{h}^*$ and $\bar{\lambda}^*$ being replaced by the corresponding non-random optimal smoothing parameters $\bar{h}^0$ and $\bar{\lambda}^0$ that minimize $MSE_L(\bar{h}, \bar{\lambda})$ defined in (3.5). By (3.9) and (3.10) in Theorem 3.1, the irrelevant continuous and discrete covariates are smoothed out. Consequently, $\mathcal{L}_n(\alpha; \mathbf{x}_0^c, \mathbf{x}_0^d)$ defined in (2.6) can be simplified to

$$\mathcal{L}_n(\alpha; \bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d) = \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - \alpha) K_{\bar{h}^0}(\bar{\mathbf{X}}_i^c - \bar{\mathbf{x}}_0^c) \Lambda_{\bar{\lambda}^0}(\bar{\mathbf{X}}_i^d, \bar{\mathbf{x}}_0^d),$$

where

$$K_{\bar{h}^0}(\bar{\mathbf{X}}_i^c - \bar{\mathbf{x}}_0^c) = \prod_{s=1}^{d_1} \frac{1}{\bar{h}_s^0} k\left(\frac{\bar{X}_{is}^c - \bar{x}_{0s}^c}{\bar{h}_s^0}\right), \quad \Lambda_{\bar{\lambda}^0}(\bar{\mathbf{X}}_i^d, \bar{\mathbf{x}}_0^d) = \prod_{s=1}^{d_3} (\bar{\lambda}_s^0)^{I\left(\bar{X}_{is}^d \neq \bar{x}_{0s}^d\right)}.$$

Letting $\widehat{Q}_\tau(\bar{\mathbf{x}}_0; \bar{h}^0, \bar{\lambda}^0)$ be the minimizer to $\mathcal{L}_n(\alpha; \bar{\mathbf{x}}_0^c, \bar{\mathbf{x}}_0^d)$ with respect to α, we next prove that

$$\sqrt{n\bar{H}^0}\left[\widehat{Q}_\tau(\bar{\mathbf{x}}_0; \bar{h}^0, \bar{\lambda}^0) - Q_\tau(\bar{\mathbf{x}}_0) - b(\bar{\mathbf{x}}_0; \bar{h}^0, \bar{\lambda}^0)\right] \xrightarrow{d} N\left[0, \sigma_*^2(\bar{\mathbf{x}}_0)\right], \tag{A.12}$$

where $\bar{H}^0 = \bar{h}_1^0 \cdots \bar{h}_{d_1}^0$. Let

$$V(\mathbf{x}_0) = f_{\bar{X}}(\bar{\mathbf{x}}_0) f_e(0|\bar{\mathbf{x}}_0) = \xi(\bar{\mathbf{x}}_0)$$

and

$$U_n(\bar{\mathbf{x}}_0) = \frac{1}{n}\sum_{i=1}^n \eta_i(\bar{\mathbf{x}}_0) K_{\bar{h}^0}(\bar{\mathbf{X}}_i^c - \bar{\mathbf{x}}_0^c) \Lambda_{\bar{\lambda}^0}(\bar{\mathbf{X}}_i^d, \bar{\mathbf{x}}_0^d).$$

By Lemma A.1, we have

$$\sqrt{n\bar{H}^0}\left[\widehat{Q}_\tau(\bar{\mathbf{x}}_0;\bar{h}^0,\bar{\lambda}^0)-Q_\tau(\bar{\mathbf{x}}_0)\right]=\xi^{-1}(\bar{\mathbf{x}}_0)\left[\sqrt{n\bar{H}^0}\cdot U_n(\bar{\mathbf{x}}_0)\right]+o_P(1). \tag{A.13}$$

From (A.13), in order to prove (A.12), we only need to derive the limiting distribution of $U_n(\bar{\mathbf{x}}_0)$.

Let $U_n^\star(\bar{\mathbf{x}}_0)$ be defined as $U_n(\bar{\mathbf{x}}_0)$ but with $\eta_i=\eta_i(\bar{\mathbf{x}}_0)$ replaced by $\eta_i^\star=\tau-I(e_i<0)$. Then, we have

$$U_n(\bar{\mathbf{x}}_0)-E\left[U_n(\bar{\mathbf{x}}_0)\right]=U_n^\star(\bar{\mathbf{x}}_0)-E\left[U_n^\star(\bar{\mathbf{x}}_0)\right]+U_n(\bar{\mathbf{x}}_0)-U_n^\star(\bar{\mathbf{x}}_0)-E\left[U_n(\bar{\mathbf{x}}_0)-U_n^\star(\bar{\mathbf{x}}_0)\right]. \tag{A.14}$$

Similarly to the proof of Theorem 4.1 in Li, Li and Li (2018), we may show that

$$Var\left[U_n(\bar{\mathbf{x}}_0)-U_n^\star(\bar{\mathbf{x}}_0)\right]\leqslant E\left\{\left[U_n(\bar{\mathbf{x}}_0)-U_n^\star(\bar{\mathbf{x}}_0)\right]^2\right\}=o\left((n\bar{H}^0)^{-1}\right). \tag{A.15}$$

From the classical central limit theorem for the *i.i.d.* random variables, we have

$$\sqrt{n\bar{H}^0}\{U_n^\star(\bar{\mathbf{x}}_0)-E\left[U_n^\star(\bar{\mathbf{x}}_0)\right]\}\xrightarrow{d}N\left(0,\tau(1-\tau)f_{\bar{X}}(\bar{\mathbf{x}}_0)\nu_0\right). \tag{A.16}$$

In view of (A.14)–(A.16), we have

$$\sqrt{n\bar{H}^0}\{U_n(\bar{\mathbf{x}}_0)-E\left[U_n(\bar{\mathbf{x}}_0)\right]\}\xrightarrow{d}N\left(0,\tau(1-\tau)f_{\bar{X}}(\bar{\mathbf{x}}_0)\nu_0\right). \tag{A.17}$$

It remains to derive the asymptotic bias term of the local kernel quantile estimation. By the smoothness condition in Assumptions 2(ii) and 3(i), we have

$$\begin{aligned}
E\left[U_n(\bar{\mathbf{x}}_0)\right] &= E\left\{\left[\tau-I\left(e_i<-\delta_i(\bar{\mathbf{x}}_0)\right)\right]K_{\bar{h}^0}(\bar{\mathbf{X}}_i^c-\bar{\mathbf{x}}_0^c)\Lambda_{\bar{\lambda}^0}(\bar{\mathbf{X}}_i^d,\bar{\mathbf{x}}_0^d)\right\}\\
&= E\left\{\left[F_e(0|\bar{\mathbf{X}}_i)-F_e(-\delta_i(\bar{\mathbf{x}}_0)|\bar{\mathbf{X}}_i)\right]K_{\bar{h}^0}(\bar{\mathbf{X}}_i^c-\bar{\mathbf{x}}_0^c)\Lambda_{\bar{\lambda}^0}(\bar{\mathbf{X}}_i^d,\bar{\mathbf{x}}_0^d)\right\}\\
&\approx E\left[f_e(0|\bar{\mathbf{X}}_i)\delta_i(\bar{\mathbf{x}}_0)K_{\bar{h}^0}(\bar{\mathbf{X}}_i^c-\bar{\mathbf{x}}_0^c)\Lambda_{\bar{\lambda}^0}(\bar{\mathbf{X}}_i^d,\bar{\mathbf{x}}_0^d)\right]\\
&\approx \xi(\bar{\mathbf{x}}_0)b(\bar{\mathbf{x}}_0;\bar{h}^0,\bar{\lambda}^0),
\end{aligned} \tag{A.18}$$

where $\delta_i(\bar{\mathbf{x}}_0)=Q_\tau(\bar{\mathbf{X}}_i)-Q_\tau(\bar{\mathbf{x}}_0)$, and $b(\bar{\mathbf{x}}_0;\bar{h}^0,\bar{\lambda}^0)$ is defined as in (3.2). Therefore, we have

$$\sqrt{n\bar{H}}\left[U_n(\bar{\mathbf{x}}_0)+\xi(\bar{\mathbf{x}}_0)b(\bar{\mathbf{x}}_0;\bar{h}^0,\bar{\lambda}^0)\right]\xrightarrow{d}N\left(0,\tau(1-\tau)f_{\bar{X}}(\bar{\mathbf{x}}_0)\nu_0\right). \tag{A.19}$$

By (A.13) and (A.19), we prove the asymptotic normal distribution in (A.12), completing the proof of Theorem 3.2. □

The next lemma is on the uniform consistency for the marginal quantile estimation, which is crucial to prove the sure screening property in Theorem 4.1.

LEMMA A.3. *Suppose that Assumptions 1(i), 4(i), 5 and 6(i) are satisfied. Then we have*

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\left|\widehat{Q}^c_{\tau,s}(x)-Q^c_{\tau,s}(x)\right| = O_P\left(n^{-2/5}\sqrt{\log n}\right) \tag{A.20}$$

and

$$\max_{1\leqslant s\leqslant d_3+d_4}\sup_{x\in\mathcal{D}_s}\left|\widehat{Q}^d_{\tau,s}(x)-Q^d_{\tau,s}(x)\right| = O_P\left(n^{-2/5}\sqrt{\log n}\right), \tag{A.21}$$

where $\mathcal{S}_s^\star\subset\mathcal{S}_s$ such that $w_s(x)\neq 0$ for $x\in\mathcal{S}_s^\star$, $\mathcal{S}_s$ and $\mathcal{D}_s$ denote the supports for the s-th continuous and discrete covariates, respectively.

PROOF OF THEOREM 4.1. Note that the conventional τ-th sample quantile function $\widehat{Q}_\tau$ is root-n consistent, i.e.,

$$\widehat{Q}_\tau - Q_\tau = O_P\left(n^{-1/2}\right). \tag{A.22}$$

By (A.20)–(A.22), we readily have that

$$\max_{1\leqslant s\leqslant d_1+d_2}\left|\widehat{D}^c_{\tau,s}-D^c_{\tau,s}\right| = O_P\left(n^{-2/5}\sqrt{\log n}\right), \tag{A.23}$$

and

$$\max_{1\leqslant s\leqslant d_3+d_4}\left|\widehat{D}^d_{\tau,s}-D^d_{\tau,s}\right| = O_P\left(n^{-2/5}\sqrt{\log n}\right). \tag{A.24}$$

By (A.23) and (A.24) as well as the assumption that $n^{-2/5}\sqrt{\log n}=o(\varpi_n)$ and $\gamma_n^c=\gamma_n^d=\varpi_n/2$, we have

$$\begin{aligned}
& P\left(\mathcal{M}^c_\tau\subset\widehat{\mathcal{M}}^c_\tau,\ \mathcal{M}^d_\tau\subset\widehat{\mathcal{M}}^d_\tau\right)\\
= \ & P\left(\min_{s\in\mathcal{M}^c_\tau}\widehat{D}^c_{\tau,s}\geqslant\gamma_n^c,\ \min_{s\in\mathcal{M}^d_\tau}\widehat{D}^d_{\tau,s}\geqslant\gamma_n^d\right)\\
\geqslant \ & P\left(\min_{s\in\mathcal{M}^c_\tau}D^c_{\tau,s}-\max_{1\leqslant s\leqslant d_1+d_2}\left|\widehat{D}^c_{\tau,s}-D^c_{\tau,s}\right|\geqslant\gamma_n^c,\ \min_{s\in\mathcal{M}^d_\tau}D^d_{\tau,s}-\max_{1\leqslant s\leqslant d_3+d_4}\left|\widehat{D}^d_{\tau,s}-D^d_{\tau,s}\right|\geqslant\gamma_n^d\right)\\
\geqslant \ & 1-\left[P\left(\min_{s\in\mathcal{M}^c_\tau}D^c_{\tau,s}-\max_{1\leqslant s\leqslant d_1+d_2}\left|\widehat{D}^c_{\tau,s}-D^c_{\tau,s}\right|\leqslant\gamma_n^c\right)\right.\\
& \left.+P\left(\min_{s\in\mathcal{M}^d_\tau}D^d_{\tau,s}-\max_{1\leqslant s\leqslant d_3+d_4}\left|\widehat{D}^d_{\tau,s}-D^d_{\tau,s}\right|\leqslant\gamma_n^d\right)\right]\\
\geqslant \ & 1-\left[P\left(\max_{1\leqslant s\leqslant d_1+d_2}\left|\widehat{D}^c_{\tau,s}-D^c_{\tau,s}\right|\geqslant\varpi_n/2\right)+P\left(\max_{1\leqslant s\leqslant d_3+d_4}\left|\widehat{D}^d_{\tau,s}-D^d_{\tau,s}\right|\geqslant\varpi_n/2\right)\right]\\
= \ & 1-[o(1)+o(1)]=1+o(1),
\end{aligned} \tag{A.25}$$

completing the proof of Theorem 4.1. □

References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.

Angrist J., Chernozhukov, V. and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, 74, 539–563.

Bagues, M. and Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies*, 77, 1301–1328.

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. and Fernández-Val, I. (2011). Conditional quantile processes based on series or many regressors. *Manuscript* available at https://arxiv.org/pdf/1105.6154v1.pdf.

Blau, F. and Kahn, L. (1996). Wage structure and gender earnings differentials: an international comparison. *Economica*, S29-S62.

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(2), 413–425.

Chernozhukov, V. and Hong, H. (2002). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, 97, 872–882.

de la Peña, V. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability*, 27, 537–564.

Fan, J., Fan, Y. and Barut, E. (2014). Adaptive robust variable selection. *Annals of Statistics*, 42, 324–351.

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

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.

Fan, Y. and Liu, R. (2016). A direct approach to inference in nonparametric and semiparametric quantile models. *Journal of Econometrics*, 191, 196–216.

Fan, Y. and Park, S. (2006). Confidence intervals for the quantile of treatment effects in randomized experiments. *Journal of Econometrics*, 167, 330–344.

Gao, J. and Tong, H. (2004). Semiparametric non-linear time series model selection. *Journal of the Royal Statistical Society, Series B*, 66, 321–336.

Gao, J., Peng, B., Ren, Z. and Zhang, X. (2017). Variable selection for a categorical varying-coefficient model with identifications for determinants of body mass index. *Annals of Applied Statistics*, 11, 1117–1145.

Ghysels, E. and Ng, S. (1998). A semi-parametric factor model for interest rates and spreads. *Review of Economics and Statistics*, 80, 489–502

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

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.

Hall, P. and Marron, J. S. (1987). On the amount of noise inherent in bandwidth selection for a kernel density estimator. *Annals of Statistics*, 15, 163–181.

Hall, P., Racine, J. and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99, 1015–1026.

Jones, M. C. and Hall, P. (1990). Mean squared error properties of kernel estimates of regression quantiles. *Statistics & Probability Letters*, 10, 283–289.

He, X., Wang, L. and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Annals of Statistics*, 41, 342–369.

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

Koenker, R. and Bassett, G. (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.

Kong, E. and Xia, Y. (2017). Uniform Bahadur representation for nonparametric censored quantile regression: a redistribution-of-mass approach. *Econometric Theory*, 33, 242–261.

Li, D., Li, Q. and Li, Z. (2018). Nonparametric quantile regression estimation with mixed discrete and continuous data. *Manuscript*.

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

Ma, S., Li, R. and Tsai, S. L. (2017). Variable screening via quantile partial correlation. *Journal of the American Statistical Association*, 112, 650–663.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency for kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61, 405–415.

Marron, J.S., Jones, M.C. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91, 401–407.

O'Neill, J. and Polachek, S. (1993). Why the gender gap in wages narrowed in the 1980s. *Journal of Labor Economics*, 11, 205–228.

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

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

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119, 99–130.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88, 486–494.

Su, L. and White, H. (2012). Conditional independence specification testing for dependent processes with local polynomial quantile regression. *Advances in Econometrics*, 29, 355–434.

Wang, L., Wu, Y. and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107, 214–222.

Xia, X., Li, J. and Fu, B. (2018). Conditional quantile correlation learning for ultrahigh dimensional varying coefficient models and its application in survival analysis. Forthcoming in *Statistica Sinica*.

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

Zhang, P. (1991). Variable selection in nonparametric regression with continuous covariates. *Annals of Statistics*, 19, 1869–1882.

Zheng, Q., Peng, L. and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Annals of Statistics*, 43, 2225–2258.

Supplementary Document to "Nonparametric Estimation of Conditional Quantile Functions in the Presence of Irrelevant Covariates"

Appendix B: Proofs of the Technical Lemmas

In this appendix, we provide the detailed proofs of the technical lemmas which have been used in the main proofs in Appendix A. Letting

$$U_n(\mathbf{x}, h, \lambda, \delta) = \frac{1}{n} \sum_{i=1}^{n} \eta_i(\bar{\mathbf{x}}, \delta) \mathcal{K}_{h,\lambda}(\mathbf{X}_i, \mathbf{x})$$

with $\eta_i(\bar{\mathbf{x}}, \delta) = \tau - I\left(Y_i - Q_\tau(\bar{\mathbf{x}}) - (nH)^{-1/2}\delta < 0\right)$, it is easy to find that $U_n(\mathbf{x}, h, \lambda, 0) = U_n(\mathbf{x}, h, \lambda)$ defined in Lemma A.1. Let

$$\delta_i(\bar{\mathbf{x}}) = Q_\tau(\bar{\mathbf{X}}_i) - Q_\tau(\bar{\mathbf{x}}) \ \text{ and } \ \widehat{\delta}(\mathbf{x}) \stackrel{def}{=} \widehat{\delta}(\mathbf{x}; h, \lambda) = (nH)^{1/2} \left[\widehat{Q}_\tau(\mathbf{x}; h, \lambda) - Q_\tau(\bar{\mathbf{x}})\right].$$

The proof of Lemma A.1 is similar to the proof of Theorem 3.1 in Su and White (2012). Lemmas B.1–B.3 below are crucial to derive the uniform Bahadur representation in Lemma A.1.

LEMMA B.1. *Suppose that Assumptions 1, 2, 3(i) and 4 are satisfied. Then, we have*

$$(nH)^{1/2} |U_n(\mathbf{x}, h, \lambda)| = O_P\left(m_n \tilde{H}^{-1}\right) \tag{B.1}$$

uniformly over $\mathbf{x} \in \mathcal{S}^\star \times \mathcal{D}$ *and* $(h, \lambda) \in \mathcal{H}$, *where* H *and* m_n *are defined as in Assumption 4(iii),* $\tilde{H} = \prod_{s=1}^{d_2}(\tilde{h}_s \vee 1)$, *and* $\mathcal{H}$ *denotes a set of* (h, λ) *satisfying Assumption 4(ii)–(iv).*

PROOF OF LEMMA B.1. As the dimensions d_3 and d_4 are fixed, the set $\mathcal{D}$ only contains a finite number of distinct points. Hence, in order to prove (B.1), we only have to show that (B.1) holds uniformly over $\mathbf{x}^c \in \mathcal{S}^\star$ and $(h, \lambda) \in \mathcal{H}$ for each $\mathbf{x}^d \in \mathcal{D}$. Similarly to the arguments in the proof of (A.18) in Appendix A, we have

$$\begin{aligned} E\left[U_n(\mathbf{x}, h, \lambda)\right] &= E\left\{\mathcal{K}_{h,\lambda}(\mathbf{X}_i, \mathbf{x}) E\left[\eta_i(\bar{\mathbf{x}}) | \mathbf{X}_i\right]\right\} \\ &= E\left\{\mathcal{K}_{h,\lambda}(\mathbf{X}_i, \mathbf{x}) \left[\tau - F_e\left(-\delta_i(\bar{\mathbf{x}}) | \bar{\mathbf{X}}_i\right)\right]\right\} \\ &= O\left(\kappa_n \tilde{H}^{-1}\right) \end{aligned}$$

with $\kappa_n = \sum_{s=1}^{d_1} \bar{h}_s^2 + \sum_{s=1}^{d_3} \bar{\lambda}_s$, and consequently

$$(nH)^{1/2} |E\left[U_n(\mathbf{x}, h, \lambda)\right]| = O\left((nH)^{1/2}\kappa_n \tilde{H}^{-1}\right) = O\left(m_n \tilde{H}^{-1}\right) \tag{B.2}$$

by Assumption 4(iii). With (B.2), we only need to prove that

$$(nH)^{1/2}\tilde{H}\left|U_n(\mathbf{x},h,\lambda)-E\left[U_n(\mathbf{x},h,\lambda)\right]\right| = O_P\left(m_n\right) \tag{B.3}$$

uniformly over $\mathbf{x}^c \in \mathcal{S}^\star$ and $(h,\lambda)\in\mathcal{H}$.

Consider covering the compact set $\mathcal{S}^\star$ by some disjoint sets $\mathcal{S}^\star(k)$, $k=1,\cdots,K_1$, and covering the set $\mathcal{H}$ by some disjoint sets $\mathcal{H}(k)$, $k=1,\cdots,K_2$. Denote the center points of $\mathcal{S}^\star(k)$ and $\mathcal{H}(k)$ by $\mathbf{x}^c(k)$ and $[h(k),\lambda(k)]$, respectively. Let the radius of $\mathcal{S}^\star(k)$ be of order $m_n n^{-(1-\epsilon)/2-2\iota}$, and

$$\|h-h(k)\|\leqslant m_n n^{-(1-\epsilon)/2-2\iota},\ \ \|\lambda-\lambda(k)\|\leqslant m_n n^{-(1-\epsilon)/2-\iota},\ \ (h,\lambda)\in\mathcal{H}_k$$

where $\iota \geqslant c\vee(1-\epsilon)$ is a bounded constant such that $\bar{h}_s \geqslant n^{-\iota}$ for all $s=1,\cdots,d_1$, c and ϵ are defined in Assumption 4 (ii) and (iii), respectively. Note that

$$\begin{aligned}
&\sup_{\mathbf{x}^c\in\mathcal{S}}\sup_{(h,\lambda)\in\mathcal{H}}(nH)^{1/2}\tilde{H}\left|U_n(\mathbf{x},h,\lambda)-E\left[U_n(\mathbf{x},h,\lambda)\right]\right| \qquad\qquad (B.4)\\
\leqslant\ & \max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}(nH(k_2))^{1/2}\,\tilde{H}(k_2)\left|U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))-E\left[U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right]\right|+\\
&\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}\sup_{\mathbf{x}^c\in\mathcal{S}^\star(k_1)}\sup_{(h,\lambda)\in\mathcal{H}(k_2)}\left|(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda)-(nH(k_2))^{1/2}\,\tilde{H}(k_2)\cdot U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right|+\\
&\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}\sup_{\mathbf{x}^c\in\mathcal{S}^\star(k_1)}\sup_{(h,\lambda)\in\mathcal{H}(k_2)}\left|(nH)^{1/2}\tilde{H}\cdot E\left[U_n(\mathbf{x},h,\lambda)\right]-(nH(k_2))^{1/2}\,\tilde{H}(k_2)\cdot E\left[U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right]\right|,
\end{aligned}$$

where $H(k)$ and $\tilde{H}(k)$ are defined similarly to H and $\tilde{H}$ but with the components in h replaced by those in $h(k)$, and

$$\mathbf{x}(k)=\begin{pmatrix}\mathbf{x}^c(k)\\ \mathbf{x}^d\end{pmatrix},\ \ \mathbf{x}^c(k)\in\mathcal{S}^\star(k),\ \ \mathbf{x}^d\in\mathcal{D}.$$

By the smoothness condition on $k(\cdot)$ in Assumption 4(i) and following standard calculation, we may show that

$$\begin{aligned}
&\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}\sup_{\mathbf{x}^c\in\mathcal{S}^\star(k_1)}\sup_{(h,\lambda)\in\mathcal{H}(k_2)}\left|(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda)-(nH)^{1/2}\,\tilde{H}\cdot U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right|\\
&=O_P(m_n)
\end{aligned}$$

and

$$\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}\sup_{(h,\lambda)\in\mathcal{H}(k_2)}\left|\left[(nH)^{1/2}\tilde{H}-(nH(k_2))^{1/2}\,\tilde{H}(k_2)\right]U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right|$$

$$= O_P(m_n),$$

leading to

$$\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}\sup_{\mathbf{x}^c\in\mathcal{S}^\star(k_1)}\sup_{(h,\lambda)\in\mathcal{H}(k_2)}\left|(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda)-(nH(k_2))^{1/2}\tilde{H}(k_2)\cdot U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right|$$
$$= O_P(m_n). \tag{B.5}$$

Similarly, we also have that

$$\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}\sup_{\mathbf{x}^c\in\mathcal{S}^\star(k_1)}\sup_{(h,\lambda)\in\mathcal{H}(k_2)}\left|(nH)^{1/2}\tilde{H}\cdot \mathsf{E}\left[U_n(\mathbf{x},h,\lambda)\right]-(nH(k_2))^{1/2}\tilde{H}(k_2)\cdot \mathsf{E}\left[U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right]\right|$$
$$= O(m_n). \tag{B.6}$$

On the other hand, by the Bernstein inequality for independent sequence (e.g., van der Vaart and Wellner, 1996) and noting that both K_1 and K_2 are divergent to infinity at a polynomial rate of n, we can prove that

$$\begin{aligned}
& \mathsf{P}\left(\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}(nH(k_2))^{1/2}\tilde{H}(k_2)\left|U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))-\mathsf{E}\left[U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right]\right|>c_1m_n\right)\\
\leqslant\ & \sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2}\mathsf{P}\left((nH(k_2))^{1/2}\tilde{H}(k_2)\left|U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))-\mathsf{E}\left[U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right]\right|>c_1m_n\right)\\
\leqslant\ & O\left(K_1\cdot K_2\cdot\exp\{-c_1^\star\log n\}\right)=o(1),
\end{aligned}$$

where $c_1^\star$ would be a sufficiently large positive constant if c_1 is large enough. Hence, we have

$$\max_{1\leqslant k_1\leqslant K_1}\max_{1\leqslant k_2\leqslant K_2}(nH(k_2))^{1/2}\tilde{H}(k_2)\left|U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))-\mathsf{E}\left[U_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2))\right]\right|=O_P(m_n). \tag{B.7}$$

By (B.4)–(B.7), we can prove (B.3), completing the proof of Lemma B.1. □

LEMMA B.2. *Suppose that Assumptions 1, 2(i) and 4 are satisfied. Then, we have*

$$(nH)^{1/2}\tilde{H}\left|\bar{U}_n(\mathbf{x},h,\lambda,\delta)-\mathsf{E}\left[\bar{U}_n(\mathbf{x},h,\lambda,\delta)\right]\right|=O_P\left(m_n^{3/2}(nH)^{-1/4}\right) \tag{B.8}$$

uniformly over $\mathbf{x}\in\mathcal{S}^\star\times\mathcal{D}$, $(h,\lambda)\in\mathcal{H}$ *and* $|\delta|\leqslant c_2m_n$, *where*

$$\bar{U}_n(\mathbf{x},h,\lambda,\delta)=U_n(\mathbf{x},h,\lambda,\delta)-U_n(\mathbf{x},h,\lambda,0)=U_n(\mathbf{x},h,\lambda,\delta)-U_n(\mathbf{x},h,\lambda)$$

and c_2 is a sufficiently large positive constant.

PROOF OF LEMMA B.2. As in the proof of Lemma B.1, we only need to prove

$$(nH)^{3/4}\tilde{H}\left|\bar{U}_n(\mathbf{x},h,\lambda,\delta)-E\left[\bar{U}_n(\mathbf{x},h,\lambda,\delta)\right]\right| = O_P\left(m_n^{3/2}\right) \tag{B.9}$$

uniformly over $\mathbf{x}^c \in \mathcal{S}^\star$, $(h,\lambda)\in\mathcal{H}$ and $|\delta|\leqslant c_2 m_n$ for each $\mathbf{x}^d\in\mathcal{D}$. The main techniques are similar to those used in the proof of Lemma B.1. Consider covering the compact set $\mathcal{S}^\star$ by some disjoint sets $\bar{\mathcal{S}}^\star(k)$, $k=1,\cdots,\bar{K}_1$, and covering the set $\mathcal{H}$ by some disjoint sets $\bar{\mathcal{H}}(k)$, $k=1,\cdots,\bar{K}_2$. Let $\bar{\mathcal{I}}(k)$, $k=1,\cdots,\bar{K}_3$, be the disjoint intervals covering the closed interval $[-c_2m_n, c_2m_n]$. Denote the center points of $\bar{\mathcal{S}}^\star(k)$, $\bar{\mathcal{H}}(k)$ and $\bar{\mathcal{I}}(k)$ by $\mathbf{x}^c(k)$, $[h(k),\lambda(k)]$ and $\delta(k)$, respectively. In addition, we let the radius of $\bar{\mathcal{S}}^\star(k)$ and $\bar{\mathcal{I}}(k)$ be of orders $m_n^{1/2}n^{-(1-\epsilon)/4-2\iota}$ and $m_n^{3/2}n^{-(1-\epsilon)/4-\iota}$, respectively, and

$$\|h-h(k)\|\leqslant m_n^{1/2}n^{-(1-\epsilon)/4-2\iota},\ \ \|\lambda-\lambda(k)\|\leqslant m_n^{1/2}n^{-(1-\epsilon)/4-\iota},\ \ (h,\lambda)\in\bar{\mathcal{H}}(k),$$

where ι is defined in the proof of Lemma B.1. Following the proofs of (B.5) and (B.6), we may show that

$$\begin{aligned}
&\sup_{\mathbf{x}^c\in\mathcal{S}^\star}\sup_{(h,\lambda)\in\mathcal{H}}\sup_{|\delta|\leqslant c_2m_n}(nH)^{3/4}\tilde{H}\left|\bar{U}_n(\mathbf{x},h,\lambda,\delta)-E\left[\bar{U}_n(\mathbf{x},h,\lambda,\delta)\right]\right| \\
\leqslant\ & \max_{1\leqslant k_1\leqslant\bar{K}_1}\max_{1\leqslant k_2\leqslant\bar{K}_2}\max_{1\leqslant k_3\leqslant\bar{K}_3}\bar{U}_n(k_1,k_2,k_3)+O_P(m_n^{3/2}),
\end{aligned} \tag{B.10}$$

where

$$\bar{U}_n(k_1,k_2,k_3)=(nH(k_2))^{3/4}\tilde{H}(k_2)\left|\bar{U}_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2),\delta(k_3))-E\left[\bar{U}_n(\mathbf{x}(k_1),h(k_2),\lambda(k_2),\delta(k_3))\right]\right|.$$

Finally, using the Bonferroni and Bernstein inequalities, we can prove that

$$\begin{aligned}
& P\left(\max_{1\leqslant k_1\leqslant\bar{K}_1}\max_{1\leqslant k_2\leqslant\bar{K}_2}\max_{1\leqslant k_3\leqslant\bar{K}_3}\bar{U}_n(k_1,k_2,k_3)\geqslant c_2^\diamond m_n^{3/2}\right) \\
\leqslant\ & \sum_{k_1=1}^{\bar{K}_1}\sum_{k_2=1}^{\bar{K}_2}\sum_{k_3=1}^{\bar{K}_3}P\left(\bar{U}_n(k_1,k_2,k_3)>c_2^\diamond m_n^{3/2}\right) \\
\leqslant\ & O\left(\bar{K}_1\cdot\bar{K}_2\cdot\bar{K}_3\cdot\exp\{-c_2^\star\log n\}\right)=o(1),
\end{aligned}$$

where $c_2^\star>0$ would be sufficiently large if $c_2^\diamond>0$ is large enough. Hence, we have

$$\max_{1\leqslant k_1\leqslant\bar{K}_1}\max_{1\leqslant k_2\leqslant\bar{K}_2}\max_{1\leqslant k_3\leqslant\bar{K}_3}\bar{U}_n(k_1,k_2,k_3)=O_P(m_n^{3/2}), \tag{B.11}$$

which together with (B.10), leads to (B.9), completing the proof of Lemma B.2. $\square$

LEMMA B.3. *Suppose that Assumptions 1(ii), 2 and 4 are satisfied. Then, we have*

$$(nH)^{1/2}\tilde{H}\cdot \mathsf{E}\left[\bar{U}_n(\mathbf{x},h,\lambda,\delta)\right] = -V(\mathbf{x},\tilde{h},\tilde{\lambda})\delta + o\left(m_n^{3/2}(nH)^{-1/4}\right) \tag{B.12}$$

uniformly over $\mathbf{x}\in\mathcal{S}^\star\times\mathcal{D}$, $(h,\lambda)\in\mathcal{H}$ *and* $|\delta|\leqslant c_2m_n$, *where* $V(\mathbf{x},\tilde{h},\tilde{\lambda})$ *is defined in Lemma A.1.*

PROOF OF LEMMA B.3. Let $\delta_i(\bar{\mathbf{x}}) = Q_\tau(\bar{\mathbf{X}}_i) - Q_\tau(\bar{\mathbf{x}})$ as above. By Assumptions 2 and 4, we readily have that

$$\begin{aligned}
\mathsf{E}\left[\bar{U}_n(\mathbf{x},h,\lambda,\delta)\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left\{\left[\eta_i(\bar{\mathbf{x}},\delta)-\eta_i(\bar{\mathbf{x}},0)\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_i,\mathbf{x})\right\}\\
&= -(nH)^{-1/2}\delta\cdot\mathsf{E}\left[\mathcal{K}_{h,\lambda}(\mathbf{X}_i,\mathbf{x})f_e(-\delta_i(\bar{\mathbf{x}})|\bar{\mathbf{X}}_i)\right] + O\left(\delta^2(nH)^{-1}\right)\\
&= -(nH)^{-1/2}\delta\cdot f_{\tilde{\mathbf{X}}}(\bar{\mathbf{x}})f_e(0|\bar{\mathbf{x}})\cdot\mathsf{E}\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_i,\tilde{\mathbf{x}})\right] + O\left(|\delta|\kappa_n(nH)^{-1/2}+\delta^2(nH)^{-1}\right)\\
&= -(nH)^{-1/2}\delta\cdot f_{\tilde{\mathbf{X}}}(\bar{\mathbf{x}})f_e(0|\bar{\mathbf{x}})\cdot\mathsf{E}\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_i,\tilde{\mathbf{x}})\right] + o\left(m_n^{3/2}(nH)^{-3/4}\right)
\end{aligned} \tag{B.13}$$

uniformly over $\mathbf{x}\in\mathcal{S}^\star\times\mathcal{D}$, $(h,\lambda)\in\mathcal{H}$ and $|\delta|\leqslant c_2m_n$, where we have used the facts of

$$|\delta|\kappa_n = O(m_n\kappa_n) = O\left(m_n^2(nH)^{-1/2}\right) = o\left(m_n^{3/2}(nH)^{-1/4}\right)$$

and

$$\delta^2(nH)^{-1} = O\left(m_n^2(nH)^{-1}\right) = o\left(m_n^{3/2}(nH)^{-3/4}\right)$$

due to Assumption 4(iii). □

PROOF OF LEMMA A.1. Following the proofs of Lemma A2 in Ruppert and Carroll (1980) and Lemma A.5 in Su and White (2012), we may show that

$$(nH)^{1/2}\tilde{H}\cdot U_n\left(\mathbf{x},h,\lambda,\widehat{\delta}(\mathbf{x})\right) = O_P\left((nH)^{-1/2}\right) = o_P\left(m_n^{3/2}(nH)^{-1/4}\right) \tag{B.14}$$

uniformly over $\mathbf{x}\in\mathcal{S}^\star\times\mathcal{D}$ and $(h,\lambda)\in\mathcal{H}$, where $\widehat{\delta}(\mathbf{x}) = \widehat{Q}_\tau(\mathbf{x};h,\lambda) - Q_\tau(\bar{\mathbf{x}})$. By Lemmas B.2 and B.3, we readily have that

$$\left|(nH)^{1/2}\tilde{H}\left[U_n(\mathbf{x},h,\lambda,\delta)-U_n(\mathbf{x},h,\lambda)\right] + V(\mathbf{x},\tilde{h},\tilde{\lambda})\delta\right| = O_P\left(m_n^{3/2}(nH)^{-1/4}\right) \tag{B.15}$$

uniformly over $\mathbf{x}\in\mathcal{S}^\star\times\mathcal{D}$, $(h,\lambda)\in\mathcal{H}$ and $|\delta|\leqslant c_2m_n$.

For notational simplicity, we next let "$\sup_{\mathbf{x},(h,\lambda)}$" denote "$\sup_{\mathbf{x}\in\mathcal{S}^\star\times\mathcal{D}}\sup_{(h,\lambda)\in\mathcal{H}}$". Note that

$$\mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2m_n} -\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right] < c_2c_3m_n^2\right)$$

$$\leqslant \quad P\left(\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2 m_n} -\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right] < c_2c_3m_n^2,\ \Omega_{n1}\right) + P\left(\Omega_{n1}^c\right) \tag{B.16}$$

where Ω_{n1} denotes the event that

$$\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2 m_n}\left\{-\delta\left[-V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right)\delta + (nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda)\right]\right\} \geqslant 2c_2c_3m_n^2,$$

and Ω_{n1}^c is the complement of Ω_{n1}. Note that there exists a constant $\underline{c} > 0$ such that $V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right) \geqslant \underline{c}$, and thus

$$\begin{aligned}
&\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2 m_n}\left\{-\delta\left[-V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right)\delta + (nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda)\right]\right\} \\
\geqslant\ & -c_2m_n\cdot\sup_{\mathbf{x},(h,\lambda)}(nH)^{1/2}\tilde{H}\cdot|U_n(\mathbf{x},h,\lambda)| + \underline{c}c_2^2m_n^2.
\end{aligned}$$

Consequently, Ω_{n1}^c indicates that

$$-c_2m_n\cdot\sup_{\mathbf{x},(h,\lambda)}(nH)^{1/2}\tilde{H}\cdot|U_n(\mathbf{x},h,\lambda)| + \underline{c}c_2^2m_n^2 < 2c_2c_3m_n^2,$$

and

$$P(\Omega_{n1}^c) \leqslant P\left(\sup_{\mathbf{x},(h,\lambda)}(nH)^{1/2}\tilde{H}\cdot|U_n(\mathbf{x},h,\lambda)| > (\underline{c}c_2 - 2c_3)m_n\right) \to 0 \tag{B.17}$$

by letting c_2 be large enough. On the other hand, when

$$\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2 m_n} -\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right] < c_2c_3m_n^2$$

and the event Ω_{n1} jointly hold, we must have

$$\sup_{\mathbf{x},(h,\lambda)}\sup_{|\delta|=c_2 m_n}\left\{\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta) - (nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda) + V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right)\delta\right]\right\} \geqslant c_2c_3m_n^2.$$

This, together with (B.15) and the condition of $m_n^2 = o(nH)$ in Assumption 4(iii), implies that

$$\begin{aligned}
& P\left(\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2 m_n} -\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right] < c_2c_3m_n^2,\ \Omega_{n1}\right) \\
\leqslant\ & P\left(\sup_{\mathbf{x},(h,\lambda)}\sup_{|\delta|=c_2 m_n}\left\{\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta) - (nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda) + V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right)\delta\right]\right\} \geqslant c_2c_3m_n^2\right) \\
\leqslant\ & P\left(\sup_{\mathbf{x},(h,\lambda)}\sup_{|\delta|=c_2 m_n}\left|(nH)^{1/2}\tilde{H}\cdot[U_n(\mathbf{x},h,\lambda,\delta) - U_n(\mathbf{x},h,\lambda)] + V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right)\delta\right| \geqslant c_3m_n\right)
\end{aligned}$$

$$\leqslant \quad \mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}\sup_{|\delta|=c_2m_n}\left|(nH)^{1/2}\tilde{H}\cdot[U_n(\mathbf{x},h,\lambda,\delta)-U_n(\mathbf{x},h,\lambda)]+V\left(\mathbf{x},\tilde{h},\tilde{\lambda}\right)\delta\right|\geqslant c_3^{\star}m_n^{3/2}(nH)^{-1/4}\right)$$
$$\to \quad 0, \tag{B.18}$$

where $c_3^{\star}$ is a sufficiently large positive constant. With (B.16)–(B.18), we prove that

$$\mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|=c_2m_n}-\delta\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right]<c_2c_3m_n^2\right)\to 0, \tag{B.19}$$

by choosing $c_2>0$ sufficiently large.

We next consider a general case of $|\delta|\geqslant c_2m_n$. Note that $-\delta U_n(\mathbf{x},h,\lambda,\varpi\delta)$, when treated as a function of ϖ, is non-decreasing for $\varpi\geqslant 1$. For $|\delta|\geqslant c_2m_n$, we let $\delta^{\star}=\delta/\varpi^{\star}$ with $\varpi^{\star}=|\delta|/(c_2m_n)$. It is easy to find that $|\delta^{\star}|=c_2m_n$,

$$-\delta^{\star}U_n(\mathbf{x},h,\lambda,\delta)=-\delta^{\star}U_n(\mathbf{x},h,\lambda,\varpi^{\star}\delta^{\star})\geqslant-\delta^{\star}U_n(\mathbf{x},h,\lambda,\delta^{\star})$$

and consequently

$$|U_n(\mathbf{x},h,\lambda,\delta)|\geqslant\frac{-\delta^{\star}U_n(\mathbf{x},h,\lambda,\delta^{\star})}{c_2m_n}.$$

Then, by (B.19), we may show that

$$\mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta|\geqslant c_2m_n}\left|(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right|<c_3m_n\right)$$
$$\leqslant \quad \mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}\inf_{|\delta^{\star}|=c_2m_n}-\delta^{\star}\left[(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta^{\star})\right]<c_2c_3m_n^2\right)\to 0 \tag{B.20}$$

by letting $c_2>0$ be large enough.

Note that

$$\mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}(nH)^{1/2}\left|\widehat{Q}_\tau(\mathbf{x};h,\lambda)-Q_\tau(\bar{\mathbf{x}})\right|>c_2m_n\right)$$
$$\leqslant \quad \mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}(nH)^{1/2}\left|\widehat{Q}_\tau(\mathbf{x};h,\lambda)-Q_\tau(\bar{\mathbf{x}})\right|>c_2m_n,\Omega_{n2}\right)+\mathsf{P}\left(\Omega_{n2}^c\right), \tag{B.21}$$

where Ω_{n2} denotes the event that

$$\sup_{\mathbf{x},(h,\lambda)}\left|(nH)^{1/2}\tilde{H}\cdot U_n\left(\mathbf{x},h,\lambda,\widehat{\delta}(\mathbf{x})\right)\right|<c_3m_n.$$

By (B.14), we readily have that

$$\mathsf{P}\left(\Omega_{n2}^c\right)=\mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)}\left|(nH)^{1/2}\tilde{H}\cdot U_n\left(\mathbf{x},h,\lambda,\widehat{\delta}(\mathbf{x})\right)\right|\geqslant c_3m_n\right)\to 0. \tag{B.22}$$

On the other hand, by (B.20), we may prove that

$$
\begin{aligned}
& \mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)} (nH)^{1/2}\left|\widehat{Q}_\tau(\mathbf{x};h,\lambda)-Q_\tau(\bar{\mathbf{x}})\right| > c_2 m_n, \Omega_{n2}\right) \\
= \ & \mathsf{P}\left(\sup_{\mathbf{x},(h,\lambda)} \inf_{|\delta|\geqslant c_2 m_n} \left|(nH)^{1/2}\tilde{H}\cdot U_n(\mathbf{x},h,\lambda,\delta)\right| < c_3 m_n\right) \to 0. \qquad \text{(B.23)}
\end{aligned}
$$

By (B.21)–(B.23), we show that $\widehat{\delta}(\mathbf{x}) = \widehat{Q}_\tau(\mathbf{x};h,\lambda) - Q_\tau(\mathbf{x}) = O_P(m_n)$ uniformly over $\mathbf{x} \in \mathcal{S}^\star \times \mathcal{D}$ and $(h,\lambda) \in \mathcal{H}$, which together with (B.14) and (B.15), proves (A.1), completing the proof of Lemma A.1. $\square$

Proof of Lemma A.2. The main idea to be used in the proof is similar to that in the proof of Proposition 3.1 in Li, Li and Li (2018). In order to simplify the notation, throughout this proof, we let $W_i = W(\mathbf{X}_i)$ and $\zeta_i(\mathbf{X}_i) = \widehat{Q}_{(-i)}(\mathbf{X}_i) - Q_\tau(\bar{\mathbf{X}}_i)$ with $\widehat{Q}_{(-i)}(\mathbf{X}_i) = \widehat{Q}_{(-i)}(\mathbf{X}_i;h,\lambda)$. Note that the CV function can be rewritten as

$$
\begin{aligned}
\mathrm{CV}(h,\lambda) &= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - \widehat{Q}_{(-i)}(\mathbf{X}_i)\right)W_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(e_i + Q_\tau(\bar{\mathbf{X}}_i) - \widehat{Q}_{(-i)}(\mathbf{X}_i)\right)W_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(e_i)W_i + \frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\left(e_i - \zeta_i(\mathbf{X}_i)\right) - \rho_\tau(e_i)\right]W_i \\
&\overset{\mathrm{def}}{=} \mathrm{CV}_1 + \mathrm{CV}_2(h,\lambda). \qquad \text{(B.24)}
\end{aligned}
$$

Using (B.24) and noting that CV_1 does not rely on h and λ, to complete the proof of Lemma A.2, we only need to derive the asymptotic leading term for $\mathrm{CV}_2(h,\lambda)$.

Using the following identity equality from Knight (1998):

$$
\rho_\tau(x-y) - \rho_\tau(x) = y\left[I(x\leqslant 0) - \tau\right] + \int_0^y \left[I(x\leqslant z) - I(x \leqslant 0)\right]dz, \qquad \text{(B.25)}
$$

we have

$$
\rho_\tau\left(e_i - \zeta_i(\mathbf{X}_i)\right) - \rho_\tau(e_i) = \zeta_i(\mathbf{X}_i)\left[I(e_i\leqslant 0) - \tau\right] + \int_0^{\zeta_i(\mathbf{X}_i)} \left[I(e_i\leqslant z) - I(e_i\leqslant 0)\right]dz \quad \text{(B.26)}
$$

by choosing $x = e_i$ and $y = \zeta_i(\mathbf{X}_i)$ in (B.25). By (B.26), $\mathrm{CV}_2(h,\lambda)$ in (B.24) can be decomposed as

$$
\mathrm{CV}_2(h,\lambda) = \mathrm{CV}_{21}(h,\lambda) + \mathrm{CV}_{22}(h,\lambda), \qquad \text{(B.27)}
$$

where

$$\begin{aligned}
CV_{21}(h,\lambda) &= \frac{1}{n}\sum_{i=1}^{n} W_i \int_0^{\zeta_i(\mathbf{X}_i)} \left[I(e_i \leqslant z) - I(e_i \leqslant 0)\right] dz, \\
CV_{22}(h,\lambda) &= \frac{1}{n}\sum_{i=1}^{n} \zeta_i(\mathbf{X}_i)\left[I(e_i \leqslant 0) - \tau\right] W_i.
\end{aligned}$$

We next show that $CV_{21}(h,\lambda)$ is the asymptotic leading term of $CV_2(h,\lambda)$ uniformly over $(h,\lambda) \in \mathcal{H}$, while $CV_{22}(h,\lambda)$ is asymptotically negligible. By Lemma B.4 below, we have

$$CV_{21}(h,\lambda) = \frac{1}{2n}\sum_{i=1}^{n} \left[\zeta_i^*(\mathbf{X}_i;h,\lambda)\right]^2 W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left(m_n^{5/2}/(nH)^{5/4} + m_n^2/(nH^{1/2})\right) \quad \text{(B.28)}$$

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

$$\zeta_i^*(\mathbf{X}_i;h,\lambda) = \left[V(\mathbf{X}_i,\tilde{h},\tilde{\lambda})\right]^{-1}\left[\tilde{H}\cdot U_{(-i)}(\mathbf{X}_i;h,\lambda)\right], \quad \text{(B.29)}$$

$V(\mathbf{X}_i,\tilde{h},\tilde{\lambda})$ is defined in Lemma A.1,

$$U_{(-i)}(\mathbf{X}_i;h,\lambda) = \frac{1}{n}\sum_{j=1,\neq i}^{n} \eta_j(\bar{\mathbf{X}}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)$$

with $\eta_j(\bar{\mathbf{X}}_i) = \tau - I\left(e_j \leqslant -\delta_j(\bar{\mathbf{X}}_i)\right)$. Letting $\eta_j^\star = \tau - I(e_j \leqslant 0)$, we can rewrite $U_{(-i)}(\mathbf{X}_i;h,\lambda)$ as

$$\begin{aligned}
U_{(-i)}(\mathbf{X}_i;h,\lambda) &= \frac{1}{n}\sum_{j\neq i}\left[\eta_j(\bar{\mathbf{X}}_i) - \eta_j^\star\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i) + \frac{1}{n}\sum_{j\neq i}\eta_j^\star\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i) \\
&\stackrel{\text{def}}{=} U_{(-i),1}(\mathbf{X}_i;h,\lambda) + U_{(-i),2}(\mathbf{X}_i;h,\lambda). \quad \text{(B.30)}
\end{aligned}$$

Defining

$$B(\mathbf{X}_i;h,\lambda) = \left[V(\mathbf{X}_i,\tilde{h},\tilde{\lambda})\right]^{-1}\left[\tilde{H}U_{(-i),1}(\mathbf{X}_i;h,\lambda)\right]$$

and

$$T(\mathbf{X}_i;h,\lambda) = \left[V(\mathbf{X}_i,\tilde{h},\tilde{\lambda})\right]^{-1}\left[\tilde{H}U_{(-i),2}(\mathbf{X}_i;h,\lambda)\right],$$

by (B.29) and (B.30), we have

$$\frac{1}{n}\sum_{i=1}^{n}\left[\zeta_i^*(\mathbf{X}_i;h,\lambda)\right]^2 W_i f_e(0|\bar{\mathbf{X}}_i) = \frac{1}{n}\sum_{i=1}^{n} B^2(\mathbf{X}_i;h,\lambda) W_i f_e(0|\bar{\mathbf{X}}_i) + \frac{1}{n}\sum_{i=1}^{n} T^2(\mathbf{X}_i;h,\lambda) W_i f_e(0|\bar{\mathbf{X}}_i) +$$

$$\frac{2}{n}\sum_{i=1}^{n} B(\mathbf{X}_i;h,\lambda)T(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i). \tag{B.31}$$

We next consider the three terms on the right hand side of (B.31) separately. Observe that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} B^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i)\\
= \;& \frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2}\tilde{H}^2\sum_{j\neq i}\sum_{k\neq i}\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^{\star}\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i)\left[\eta_k(\bar{\mathbf{X}}_i)-\eta_k^{\star}\right]W_i f_e(0|\bar{\mathbf{X}}_i)\\
= \;& \frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2}\tilde{H}^2\sum_{j\neq i}\sum_{k\neq i,j}\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^{\star}\right]\left[\eta_k(\bar{\mathbf{X}}_i)-\eta_k^{\star}\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i)W_i f_e(0|\bar{\mathbf{X}}_i)\\
&+\frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2}\tilde{H}^2\sum_{j\neq i}\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^{\star}\right]^2\mathcal{K}^2_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)W_i f_e(0|\bar{\mathbf{X}}_i)\\
\stackrel{\text{def}}{=} \;& \Pi_{n1}(h,\lambda)+\Pi_{n2}(h,\lambda).
\end{aligned}$$

For $\Pi_{n1}(h,\lambda)$, letting

$$\eta_j(\mathbf{X}_i;h,\lambda)=\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^{\star}\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)=\left[I(e_j\leqslant 0)-I(e_j\leqslant -\delta_j(\bar{\mathbf{X}}_i))\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i),$$

by Lemma B.5, we have

$$\begin{aligned}
&\tilde{H}^2\sum_{j\neq i}\sum_{k\neq i,j}\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^{\star}\right]\left[\eta_k(\bar{\mathbf{X}}_i)-\eta_k^{\star}\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i)\\
= \;& \tilde{H}^2\sum_{j\neq i}\sum_{k\neq i,j}\mathrm{E}\left[\eta_j(\mathbf{X}_i;h,\lambda)\middle|\mathbf{X}_i\right]\cdot\mathrm{E}\left[\eta_k(\mathbf{X}_i;h,\lambda)\middle|\mathbf{X}_i\right]+O_P\left(n^{3/2}\kappa_n^2 m_n H^{-1/2}\right)\\
= \;& n\tilde{H}^2\sum_{j\neq i}\mathrm{E}^2\left[\eta_j(\mathbf{X}_i;h,\lambda)\middle|\mathbf{X}_i\right]+O_P\left(n^{3/2}\kappa_n^2 m_n H^{-1/2}\right)\\
= \;& n\tilde{H}^2\sum_{j\neq i}\mathrm{E}^2\left\{\left[F_e(0|\bar{\mathbf{X}}_j)-F_e(-\delta_j(\bar{\mathbf{X}}_j)|\bar{\mathbf{X}}_j)\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)|\mathbf{X}_i\right\}+O_P\left(n^{3/2}\kappa_n^2 m_n H^{-1/2}\right)\\
= \;& n\tilde{H}^2\sum_{j\neq i}\mathrm{E}^2\left[\delta_j(\bar{\mathbf{X}}_i)f_e(0|\bar{\mathbf{X}}_j)\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\middle|\mathbf{X}_i\right]+O_P\left(n^{3/2}\kappa_n^2 m_n H^{-1/2}\right)
\end{aligned}$$

uniformly over $i=1,\cdots,n$ and $(h,\lambda)\in\mathcal{H}$, which indicates that

$$\Pi_{n1}(h,\lambda) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\left[\xi(\mathbf{X}_i)\right]^{-2}W_i f_e(0|\bar{\mathbf{X}}_i)\sum_{j\neq i}\mathrm{E}^2\left[\delta_j(\bar{\mathbf{X}}_i)f_e(0|\bar{\mathbf{X}}_j)\bar{\mathcal{K}}_{\bar{h},\bar{\lambda}}(\bar{\mathbf{X}}_j,\bar{\mathbf{X}}_i)\middle|\bar{\mathbf{X}}_i\right]$$

$$+O_P\left((nH)^{-1/2}\kappa_n^2 m_n\right). \tag{B.32}$$

When dealing with $\delta_j(\bar{\mathbf{X}}_i)$, we need to consider two possible cases. In case (i), $\bar{\mathbf{X}}_j^c \neq \bar{\mathbf{X}}_i^c$ but $\bar{\mathbf{X}}_j^d = \bar{\mathbf{X}}_i^d$. For this case, using Assumption 3(i) and the Taylor's expansion for the quantile regression function $Q_\tau(\cdot)$ (with respect to the continuous components), we may show that

$$\begin{aligned}
& [\xi(\bar{\mathbf{X}}_i)]^{-1}\mathrm{E}\left[\delta_j(\bar{\mathbf{X}}_i)f_e(0|\bar{\mathbf{X}}_j)\bar{\mathcal{K}}_{\bar{h},\bar{\lambda}}(\bar{\mathbf{X}}_j,\bar{\mathbf{X}}_i)|\bar{\mathbf{X}}_i\right] \\
= \ & \xi(\bar{\mathbf{X}}_i)^{-1}\int_{\bar{\mathcal{S}}\times\bar{\mathcal{D}}}\xi(\bar{\mathbf{x}})\delta_j(\bar{\mathbf{x}})\bar{\mathcal{K}}_{\bar{h},\bar{\lambda}}(\bar{\mathbf{x}},\bar{\mathbf{X}}_i)d\bar{\mathbf{x}} \\
= \ & \frac{1}{2}\mu_2\sum_{s=1}^{d_1}\bar{h}_s^2\left[Q_\tau^{(ss)}(\bar{\mathbf{X}}_i)+2Q_\tau^{(s)}(\bar{\mathbf{X}}_i)\xi^{(s)}(\bar{\mathbf{X}}_i)/\xi(\bar{\mathbf{X}}_i)\right]+O_P(\kappa_n^2),
\end{aligned} \tag{B.33}$$

where $\xi(\bar{\mathbf{x}}) = f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}})f_e(0|\bar{\mathbf{x}})$ as in Section 3 and

$$\bar{\mathcal{K}}_{\bar{h},\bar{\lambda}}(\bar{\mathbf{X}}_i,\bar{\mathbf{x}}) = \prod_{s=1}^{d_1}\bar{h}_s^{-1}k\left(\frac{\bar{X}_{is}^c-\bar{x}_s^c}{\bar{h}_s}\right)\prod_{s=1}^{d_3}\bar{\lambda}_s^{I\left(\bar{X}_{is}^d\neq\bar{x}_s^d\right)}.$$

In case (ii), $\bar{X}_{js}^d \neq \bar{X}_{is}^d$. Then, we have

$$\begin{aligned}
& \xi^{-1}(\bar{\mathbf{X}}_i)\mathrm{E}\left[f_e(0|\bar{\mathbf{X}}_j)\delta_j(\bar{\mathbf{X}}_i)\bar{\mathcal{K}}_{\bar{h},\bar{\lambda}}(\bar{\mathbf{X}}_j,\bar{\mathbf{X}}_i)|\bar{\mathbf{X}}_i\right] \\
= \ & \sum_{\bar{\mathbf{x}}^d\in\bar{\mathcal{D}}}\frac{\xi(\bar{\mathbf{X}}_i^c,\bar{\mathbf{x}}^d)}{\xi(\bar{\mathbf{X}}_i)}\sum_{s=1}^{d_3}\bar{\lambda}_s I_s(\bar{\mathbf{x}}^d,\bar{\mathbf{X}}_i^d)\left[Q_\tau(\bar{\mathbf{X}}_i^c,\bar{\mathbf{x}}^d)-Q_\tau(\bar{\mathbf{X}}_i)\right]+O_P(\kappa_n^2),
\end{aligned} \tag{B.34}$$

where $I_s(\bar{\mathbf{x}}^d,\bar{\mathbf{X}}_i^d)$ is defined as in (3.2). Combing (B.32)–(B.34), we readily have

$$\Pi_{n1}(h,\lambda) = \frac{1}{n}\sum_{i=1}^{n}b^2(\bar{\mathbf{X}}_i;\bar{h},\bar{\lambda})W_i f_e(0|\bar{\mathbf{X}}_i)+O_P\left(\kappa_n^3+(nH)^{-1/2}\kappa_n^2 m_n\right) \tag{B.35}$$

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

For $\Pi_{n2}(h,\lambda)$, using the argument in the proof of Lemma B.5 below, we have

$$\begin{aligned}
\Pi_{n2}(h,\lambda) &= \frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2}W_i f_e(0|\bar{\mathbf{X}}_i)\tilde{H}^2\sum_{j\neq i}\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^\star\right]^2\mathcal{K}_{h,\lambda}^2(\mathbf{X}_j,\mathbf{X}_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}[\xi(\mathbf{X}_i)]^{-2}W_i f_e(0|\bar{\mathbf{X}}_i)\mathrm{E}\left[\left(\eta_j(\bar{\mathbf{X}}_i)-\eta_j^\star\right)^2\bar{\mathcal{K}}_{\bar{h},\bar{\lambda}}^2(\bar{\mathbf{X}}_j,\bar{\mathbf{X}}_i)|\bar{\mathbf{X}}_i\right] \\
&\quad +O_P\left((nH)^{-3/2}\kappa_n^2 m_n\right) \\
&= O_P\left(\kappa_n^2/(nH)+(nH)^{-3/2}\kappa_n^2 m_n\right) = O_P\left(\kappa_n^2/(nH)\right)
\end{aligned} \tag{B.36}$$

uniformly over $(h,\lambda)\in\mathcal{H}$. By (B.35), (B.36) and noting that $\kappa_n^3 = O\left((nH)^{-1/2}\kappa_n^2 m_n\right)$ by Assumption 4(iii), we readily have that

$$\frac{1}{n}\sum_{i=1}^{n} B^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) = \frac{1}{n}\sum_{i=1}^{n} b^2(\bar{\mathbf{X}}_i;\bar{h},\bar{\lambda})W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left((nH)^{-1/2}\kappa_n^2 m_n\right) \quad \text{(B.37)}$$

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

For the second term on the right hand side of (B.31), we have

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} T^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) \\
= & \frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2} W_i f_e(0|\bar{\mathbf{X}}_i)\tilde{H}^2 \sum_{j\neq i}\sum_{k\neq i}\eta_j^\star\eta_k^\star \mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i) \\
= & \frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2} W_i f_e(0|\bar{\mathbf{X}}_i)\tilde{H}^2 \sum_{j\neq i}\sum_{k\neq i,j}\eta_j^\star\eta_k^\star \mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i) + \\
& \frac{1}{n^3}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2} W_i f_e(0|\bar{\mathbf{X}}_i)\tilde{H}^2 \sum_{j\neq i}(\eta_j^\star)^2 \mathcal{K}^2_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i) \\
\stackrel{\text{def}}{=} & \Pi_{n3}(h,\lambda)+\Pi_{n4}(h,\lambda).
\end{aligned}$$

For $\Pi_{n4}(h,\lambda)$, using the argument in the proof of Lemma B.5 and similar to the proof of (B.36), we have

$$\begin{aligned}
\Pi_{n4}(h,\lambda) &= \frac{1}{n^2}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-2} W_i f_e(0|\bar{\mathbf{X}}_i)\tilde{H}^2\cdot \mathsf{E}\left[(\eta_j^\star)^2\mathcal{K}^2_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)|\mathbf{X}_i\right] + O_P\left((nH)^{-3/2}m_n\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left[f_{\bar{X}}(\bar{\mathbf{X}}_i)f_e(0|\bar{\mathbf{X}}_i)\right]^{-2} W_i f_e(0|\bar{\mathbf{X}}_i)\mathsf{E}\left[(\eta_j^\star)^2\tilde{\mathcal{K}}^2_{\tilde{h},\tilde{\lambda}}(\bar{\mathbf{X}}_j,\bar{\mathbf{X}}_i)|\bar{\mathbf{X}}_i\right] R(\tilde{\mathbf{X}}_i;\tilde{h},\tilde{\lambda}) \\
&\quad + O_P\left((nH)^{-3/2}m_n\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n\bar{H}}\frac{\tau(1-\tau)\nu_0}{f_{\bar{X}}(\bar{\mathbf{X}}_i)f_e^2(0|\bar{\mathbf{X}}_i)} R(\tilde{\mathbf{X}}_i;\tilde{h},\tilde{\lambda})W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left((n\bar{H})^{-1}\kappa_n + (nH)^{-3/2}m_n\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma_\diamond^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left((n\bar{H})^{-1}\kappa_n + (nH)^{-3/2}m_n\right) \qquad \text{(B.38)}
\end{aligned}$$

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

$$\sigma_\diamond^2(\mathbf{X}_i;h,\lambda) = \frac{1}{n\bar{H}}\cdot\frac{\tau(1-\tau)\nu_0}{f_{\bar{X}}(\bar{\mathbf{X}}_i)f_e^2(0|\bar{\mathbf{X}}_i)}\cdot R(\tilde{\mathbf{X}}_i;\tilde{h},\tilde{\lambda}) \ \text{ with } \ R(\tilde{\mathbf{X}}_i;\tilde{h},\tilde{\lambda}) = \frac{\mathsf{E}\left[\tilde{\mathcal{K}}^2_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j,\tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right]}{\mathsf{E}^2\left[\tilde{\mathcal{K}}_{\tilde{h},\tilde{\lambda}}(\tilde{\mathbf{X}}_j,\tilde{\mathbf{X}}_i)|\tilde{\mathbf{X}}_i\right]}.$$

For $\Pi_{n3}(h,\lambda)$, by Lemma B.6, we have

$$\Pi_{n3}(h,\lambda) = O_P\left(m_n/(nH^{1/2})\right)$$

uniformly over $(h,\lambda) \in \mathcal{H}$, which together with (B.38) and the fact of $(n\bar{H})^{-1}\kappa_n = O\left((nH)^{-3/2}m_n\right)$, leads to

$$\frac{1}{n}\sum_{i=1}^{n} T^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) = \frac{1}{n}\sum_{i=1}^{n} \sigma_\diamond^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left((nH)^{-3/2}m_n + m_n/(nH^{1/2})\right) \tag{B.39}$$

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

In addition, following the argument in the proofs of Lemmas B.5 and B.6, we may show that

$$\frac{1}{n}\sum_{i=1}^{n} B(\mathbf{X}_i;h,\lambda)T(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) = O_P\left(n^{-1/2}\kappa_n m_n\right) \tag{B.40}$$

uniformly over $(h,\lambda) \in \mathcal{H}$. Using (B.28), (B.31), (B.37), (B.39) and (B.40), and noting that

$$n^{-1/2}\kappa_n m_n + m_n/(nH^{1/2}) = o\left(m_n^2/(nH^{1/2})\right)$$

and

$$(nH)^{-1/2}\kappa_n^2 m_n + (nH)^{-3/2}m_n = o\left(m_n^{5/2}/(nH)^{5/4}\right),$$

we can show that $CV_{21}(h,\lambda)$ have the following uniform asymptotic representation:

$$CV_{21}(h,\lambda) = \frac{1}{2n}\sum_{i=1}^{n}\left[b^2(\bar{\mathbf{X}}_i;\bar{h},\bar{\lambda}) + \sigma_\diamond^2(\mathbf{X}_i;h,\lambda)\right]W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left(m_n^{5/2}/(nH)^{5/4} + m_n^2/(nH^{1/2})\right). \tag{B.41}$$

It remains to derive the asymptotic order of $CV_{22}(h,\lambda)$. Using Lemma A.1 and following the argument in the proof of (B.41), $CV_{22}(h,\lambda)$ has the following asymptotic leading term:

$$\begin{aligned}
CV_{22}^{\star}(h,\lambda) &= -\frac{1}{n}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-1}\left[\tilde{H}U_{(-i)}(\mathbf{X}_i;h,\lambda)\right]\eta_i^{\star}W_i \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-1}\left[\tilde{H}U_{(-i),1}(\mathbf{X}_i;h,\lambda)\right]\eta_i^{\star}W_i \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\left[V(\mathbf{X}_i;\tilde{h},\tilde{\lambda})\right]^{-1}\left[\tilde{H}U_{(-i),2}(\mathbf{X}_i;h,\lambda)\right]\eta_i^{\star}W_i
\end{aligned}$$

$$\stackrel{\text{def}}{=} \quad CV^{\star}_{22,1}(h,\lambda) + CV^{\star}_{22,2}(h,\lambda),$$

where $U_{(-i),1}(\mathbf{X}_i; h, \lambda)$ and $U_{(-i),2}(\mathbf{X}_i; h, \lambda)$ are defined in (B.30). By the argument similar to the analysis of $\Pi_{n1}(h, \lambda)$, we have

$$CV^{\star}_{22,1}(h,\lambda) = O_P\left(n^{-1/2}\kappa_n m_n\right) = o_P\left(m_n^2/(nH^{1/2})\right)$$

uniformly over $(h, \lambda) \in \mathcal{H}$. On the other hand, using the argument in the proofs of Lemmas B.5 and B.6, we may show that

$$CV^{\star}_{22,2}(h,\lambda) = O_P\left(m_n/(nH^{1/2})\right) = o_P\left(m_n^2/(nH^{1/2})\right)$$

uniformly over $(h, \lambda) \in \mathcal{H}$. Combining the above results, we can prove that

$$CV_{22}(h,\lambda) = CV^{\star}_{22}(h,\lambda)(1 + o_P(1)) = O_P\left(n^{-1/2}\kappa_n m_n + m_n/(nH^{1/2})\right) = o_P\left(m_n^2/(nH^{1/2})\right) \tag{B.42}$$

uniformly over $(h, \lambda) \in \mathcal{H}$. By (B.24), (B.27), (B.41) and (B.42), we prove (A.2), completing the proof of Lemma A.2. □

LEMMA B.4. *Suppose that Assumptions 1(i) and 2–4 are satisfied. Then (B.28) holds uniformly over* $(h, \lambda) \in \mathcal{H}$.

PROOF OF LEMMA B.4. Throughout this proof, we let $\mathcal{X}_n$ be a σ-field generated by $\mathbf{X}_i$, $i = 1, \cdots, n$. With standard calculation, we may show that, uniformly over $(h, \lambda) \in \mathcal{H}$,

$$\begin{aligned} E\left[CV_{21}(h,\lambda)\middle|\mathcal{X}_n\right] &= \frac{1}{n}\sum_{i=1}^{n} W_i \int_0^{\zeta_i(\mathbf{X}_i;h,\lambda)} \left[F_e(z|\bar{\mathbf{X}}_i) - F_e(0|\bar{\mathbf{X}}_i)\right] dz \\ &= \frac{1}{2n}\sum_{i=1}^{n} \zeta_i^2(\mathbf{X}_i;h,\lambda) W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left(\frac{1}{n}\sum_{i=1}^{n}\zeta_i^3(\mathbf{X}_i;h,\lambda)\right), \end{aligned} \tag{B.43}$$

where $\zeta_i(\mathbf{X}_i; h, \lambda) = \zeta_i(\mathbf{X}_i)$, making its dependence on h and λ explicitly. Following the proof of Lemma A.1, we have

$$\max_{1\leqslant i\leqslant n}\sup_{(h,\lambda)\in\mathcal{H}} (nH)^{1/2}\left|\zeta_i(\mathbf{X}_i;h,\lambda)\right| = O_P(m_n), \tag{B.44}$$

which together with (B.43), indicates that

$$E\left[CV_{21}(h,\lambda)\middle|\mathcal{X}_n\right] = \frac{1}{2n}\sum_{i=1}^{n} \zeta_i^2(\mathbf{X}_i;h,\lambda) W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left(m_n^3/(nH)^{3/2}\right)$$

$$= \frac{1}{2n}\sum_{i=1}^{n}\zeta_i^2(\mathbf{X}_i;h,\lambda)W_i f_e(0|\bar{\mathbf{X}}_i) + o_P\left(m_n^{5/2}/(nH)^{5/4}\right). \quad \text{(B.45)}$$

Furthermore, by the uniform Bahadur representation in (A.1), we have

$$(nH)^{1/2}\zeta_i(\mathbf{X}_i;h,\lambda) = (nH)^{1/2}\zeta_i^*(\mathbf{X}_i;h,\lambda) + O_P(m_n^{3/2}(nH)^{-1/4}) \quad \text{(B.46)}$$

and consequently,

$$\mathsf{E}\left[\mathrm{CV}_{21}(h,\lambda)\middle|\mathcal{X}_n\right] = \frac{1}{2n}\sum_{i=1}^{n}[\zeta_i^*(\mathbf{X}_i;h,\lambda)]^2 W_i f_e(0|\bar{\mathbf{X}}_i) + O_P\left(m_n^{5/2}/(nH)^{5/4}\right). \quad \text{(B.47)}$$

By (B.47), we only need to show

$$\mathrm{CV}_{21}(h,\lambda) = \mathsf{E}\left[\mathrm{CV}_{21}(h,\lambda)\middle|\mathcal{X}_n\right] + O_P\left(m_n^{5/2}/(nH)^{5/4} + m_n^2/(nH^{1/2})\right) \quad \text{(B.48)}$$

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

By (B.44) and (B.46), we readily have that

$$\max_{1\leqslant i\leqslant n}\sup_{(h,\lambda)\in\mathcal{H}}(nH)^{1/2}|\zeta_i^*(\mathbf{X}_i;h,\lambda)| = O_P(m_n). \quad \text{(B.49)}$$

Letting $\mathrm{CV}_{21}^*(h,\lambda)$ be defined similarly to $\mathrm{CV}_{21}(h,\lambda)$ but with $\zeta_i(\mathbf{X}_i;h,\lambda)$ replaced by $\zeta_i^*(\mathbf{X}_i;h,\lambda)$, by (B.46) and (B.49), we can prove that

$$\mathrm{CV}_{21}(h,\lambda) - \mathrm{CV}_{21}^*(h,\lambda) - \mathsf{E}\left[\mathrm{CV}_{21}(h,\lambda) - \mathrm{CV}_{21}^*(h,\lambda)\middle|\mathcal{X}_n\right] = O_P\left(m_n^{5/2}/(nH)^{5/4}\right) \quad \text{(B.50)}$$

uniformly over $(h,\lambda)\in\mathcal{H}$. We next prove that

$$\mathrm{CV}_{21}^*(h,\lambda) = \mathsf{E}\left[\mathrm{CV}_{21}^*(h,\lambda)\middle|\mathcal{X}_n\right] + O_P\left(m_n^2/(nH^{1/2})\right) \quad \text{(B.51)}$$

uniformly over $(h,\lambda)\in\mathcal{H}$. Let Ω_{n3} be the event defined as

$$\max_{1\leqslant i\leqslant n}\sup_{(h,\lambda)\in\mathcal{H}}(nH)^{1/2}|\zeta_i^*(\mathbf{X}_i;h,\lambda)| \leqslant c_4 m_n.$$

By (B.49), it is easy to prove that $\mathsf{P}\left(\Omega_{n3}^c\right)\to 0$ by choosing $c_4>0$ to be sufficiently large. Hence, to prove (B.51), we next only show that

$$\mathsf{P}\left(\sup_{(h,\lambda)\in\mathcal{H}}\left|\mathrm{CV}_{21}^*(h,\lambda) - \mathsf{E}\left[\mathrm{CV}_{21}^*(h,\lambda)\middle|\mathcal{X}_n\right]\right| > c_5 m_n^2/(nH^{1/2}),\ \Omega_{n3}\right)\to 0, \quad \text{(B.52)}$$

where c_5 is a sufficiently large positive constant. The proof of (B.52) is similar to the proof of Lemma B.1. Details are omitted here to save the space. □

LEMMA B.5. *Suppose that Assumptions 1, 2(i), 3(i) and 4 satisfied. Define*

$$\eta_j(\mathbf{X}_i;h,\lambda)=\left[\eta_j(\bar{\mathbf{X}}_i)-\eta_j^\star\right]\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i).$$

Then, we have

$$\sum_{j\neq i}\sum_{k\neq i,j}\eta_j(\mathbf{X}_i;h,\lambda)\left\{\eta_k(\mathbf{X}_i;h,\lambda)-\mathsf{E}\left[\eta_k(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\right\}=O_P\left(n^{3/2}\kappa_n^2 m_n H^{-1/2}\tilde{H}^{-2}\right)\quad\text{(B.53)}$$

uniformly over $i=1,\cdots,n$ *and* $(h,\lambda)\in\mathcal{H}$.

PROOF OF LEMMA B.5. It is sufficient for us to show that

$$\sum_{j\neq i}\left\{\eta_j(\mathbf{X}_i;h,\lambda)-\mathsf{E}\left[\eta_j(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\right\}\sum_{k<j,\neq i}\left\{\eta_k(\mathbf{X}_i;h,\lambda)-\mathsf{E}\left[\eta_k(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\right\}=O_P\left(n\kappa_n^2 m_n H^{-1}\tilde{H}^{-2}\right)\tag{B.54}$$

and

$$\sum_{j\neq i}\mathsf{E}\left[\eta_j(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\sum_{k<j,\neq i}\left\{\eta_k(\mathbf{X}_i;h,\lambda)-\mathsf{E}\left[\eta_k(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\right\}=O_P\left(n^{3/2}\kappa_n^2 m_n H^{-1/2}\tilde{H}^{-2}\right).\tag{B.55}$$

We next only prove (B.54) as the proof of (B.55) is similar.

Let

$$R_j(\mathbf{X}_i;h,\lambda)=\left\{\eta_j(\mathbf{X}_i;h,\lambda)-\mathsf{E}\left[\eta_j(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\right\}\sum_{k<j,\neq i}\left\{\eta_k(\mathbf{X}_i;h,\lambda)-\mathsf{E}\left[\eta_k(\mathbf{X}_i;h,\lambda)|\mathbf{X}_i\right]\right\}$$

$\mathcal{F}_j(i)=\sigma\{\mathbf{X}_i,(\mathbf{X}_k:k\leqslant j+1),(e_k:k\leqslant j)\}$. It is easy to verify that $\{R_j(\mathbf{X}_i;h,\lambda),\mathcal{F}_j(i)\}_{j\neq i}$ is a sequence of martingale differences with mean zero. As in the proof of Lemma B.1, we cover the set $\mathcal{H}$ by some disjoint sets $\tilde{\mathcal{H}}(k)$, $k=1,\cdots,\tilde{K}$. Let the center point of $\tilde{\mathcal{H}}(k)$ be $\left[\tilde{h}(k),\tilde{\lambda}(k)\right]$ and the size of the set $\tilde{\mathcal{H}}(k)$ guarantee that

$$\max_{1\leqslant i\leqslant n}\max_{1\leqslant k\leqslant\tilde{K}}\sup_{(h,\lambda)\in\tilde{\mathcal{H}}(k)}\left|\sum_{j\neq i}R_j(\mathbf{X}_i;h,\lambda)-\sum_{j\neq i}R_j(\mathbf{X}_i;\tilde{h}(k),\tilde{\lambda}(k))\right|=O_P\left(n\kappa_n^2 m_n H^{-1}\tilde{H}^{-2}\right).\tag{B.56}$$

On the other hand, using the expoential-type inequality for martingale differences (e.g., Theorem 1.2A in de la Peña, 1999), we may show that

$$P\left(\max_{1\leqslant i\leqslant n}\max_{1\leqslant k\leqslant\tilde{K}}\left|\sum_{j\neq i}R_j(\mathbf{X}_i;\tilde{h}(k),\tilde{\lambda}(k))\right|\geqslant c_6 n\kappa_n^2 m_n H^{-1}\tilde{H}^{-2}\right)\to 0\tag{B.57}$$

by choosing $c_6 > 0$ to be sufficiently large. With (B.56) and (B.57), we complete the proof of (B.54). $\square$

LEMMA B.6. *Suppose that Assumptions 1, 2(i), 3(ii) and 4 are satisfied. Then we have*

$$\sum_{i=1}^{n}[V(\mathbf{X}_i;h,\lambda)]^{-2}W_i f_e(0|\bar{\mathbf{X}}_i)\sum_{j\neq i}\sum_{k\neq i,j}\eta_j^\star\eta_k^\star\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i)=O_P\left(m_n n^2 H^{-1/2}\tilde{H}^{-2}\right)$$

uniformly over $i=1,\cdots,n$ *and* $(h,\lambda)\in\mathcal{H}$.

PROOF OF LEMMA B.6. The proof is similar to the proof of Lemma B.5 with some modification. It is sufficient to show that

$$\sum_{j=1}^{n}\eta_j^\star\sum_{i<j}[V(\mathbf{X}_i;h,\lambda)]^{-2}W_i f_e(0|\bar{\mathbf{X}}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\sum_{k<j,\neq i}\eta_k^\star\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i)=O_P\left(m_n n^2 H^{-1/2}\tilde{H}^{-2}\right) \tag{B.58}$$

uniformly over $i=1,\cdots,n$ and $(h,\lambda)\in\mathcal{H}$. Define

$$R_j^\star(h,\lambda)=\eta_j^\star\sum_{i<j}[V(\mathbf{X}_i;h,\lambda)]^{-2}W_i f_e(0|\bar{\mathbf{X}}_i)\mathcal{K}_{h,\lambda}(\mathbf{X}_j,\mathbf{X}_i)\sum_{k<j,\neq i}\eta_k^\star\mathcal{K}_{h,\lambda}(\mathbf{X}_k,\mathbf{X}_i)$$

and let $\mathcal{F}_j^\star=\sigma\{(\mathbf{X}_k:k\leqslant j+1),(e_k:k\leqslant j)\}$. It is easy to verify that $\left\{R_j^\star(h,\lambda),\mathcal{F}_j^\star\right\}$ is a sequence of martingale differences with mean zero. As in the proof of Lemma B.5, we cover the set $\mathcal{H}$ by some disjoint sets $\breve{\mathcal{H}}(k)$, $k=1,\cdots,\breve{K}$. Let $\left[\breve{h}(k),\breve{\lambda}(k)\right]$ be the center point of $\breve{\mathcal{H}}(k)$. In addition, the size of the set $\breve{\mathcal{H}}(k)$ can ensure that

$$\max_{1\leqslant k\leqslant\breve{K}}\sup_{(h,\lambda)\in\breve{\mathcal{H}}(k)}\left|\sum_{j=1}^{n}R_j^\star(h,\lambda)-\sum_{j=1}^{n}R_j^\star(\breve{h}(k),\breve{\lambda}(k))\right|=O_P\left(m_n n^2 H^{-1/2}\tilde{H}^{-2}\right). \tag{B.59}$$

Using the exponential-type inequality for martingale differences, we may show that

$$P\left(\max_{1\leqslant k\leqslant\breve{K}}\left|\sum_{j=1}^{n}R_j^\star(\breve{h}(k),\breve{\lambda}(k))\right|\geqslant c_7 n m_n n^2 H^{-1/2}\tilde{H}^{-2}\right)\to 0 \tag{B.60}$$

by choosing $c_7>0$ to be sufficiently large. By (B.59) and (B.60), we prove (B.58), completing the proof of Lemma B.6. $\square$

The proof of Lemma A.3 is very similar to the proof of Lemma A.1 above, and need the following three technical lemmas to obtain the uniform Bahadur representation. Define

$$U_{ns}(x,\delta)=\frac{1}{nb_1}\sum_{i=1}^{n}\eta_{is}(x,\delta)k\left(\frac{X_{is}^c-x}{b_1}\right),\ U_{ns}(x)=U_{ns}(x,0)=\frac{1}{nb_1}\sum_{i=1}^{n}\eta_{is}(x)k\left(\frac{X_{is}^c-x}{b_1}\right)$$

with $\eta_{is}(x,\delta) = \tau - I\left(Y_i - Q^c_{\tau,s}(x) - (nb_1)^{-1/2}\delta < 0\right)$ and $\eta_{is}(x) = \tau - I\left(Y_i - Q^c_{\tau,s}(x) < 0\right)$, and

$$\delta_{is}(x) = Q^c_{\tau,s}(X^c_{is}) - Q^c_{\tau,s}(x) \text{ and } \widehat{\delta}_s(x) = (nb_1)^{1/2}\left[\widehat{Q}^c_{\tau,s}(x) - Q^c_{\tau,s}(x)\right].$$

LEMMA B.7. *Suppose that Assumptions 1(i), 4(i), 5(i)(ii) and 6(i) are satisfied. Then, we have*

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}^\star_s}|U_{ns}(x)| = O_P\left(n^{-2/5}\sqrt{\log n}\right) \tag{B.61}$$

when $b_1 = \alpha_1 \cdot n^{-1/5}$ *and* $\mathcal{S}^\star_s \subset \mathcal{S}_s$ *such that* $w_s(x) \neq 0$ *for* $x \in \mathcal{S}^\star_s$.

PROOF OF LEMMA B.7. Similarly to the proof of (A.18) in Appendix A, by Assumption 5(i)(ii), we readily have that

$$E\left[U_{ns}(x)\right] = O(b_1^2) = O(n^{-2/5}) = o\left(n^{-2/5}\sqrt{\log n}\right) \tag{B.62}$$

uniformly over $1 \leqslant s \leqslant d_1 + d_2$ and $x \in \mathcal{S}^\star_s$. By (B.62), we only need to prove that

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}^\star_s}|U_{ns}(x) - E\left[U_{ns}(x)\right]| = O_P\left(n^{-2/5}\sqrt{\log n}\right). \tag{B.63}$$

As in the proof of Lemma B.1, we cover the compact set $\mathcal{S}^\star_s$ by some disjoint sets $\mathcal{S}^\star_s(k)$, $k = 1, \cdots, K^\star$. Denote the center point of $\mathcal{S}^\star_s(k)$ by $x_s(k)$ and let the radius of $\mathcal{S}^\star_s(k)$ be of order $n^{-4/5}\sqrt{\log n}$. Note that

$$\begin{aligned}
&\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}^\star_s}|U_{ns}(x) - E\left[U_{ns}(x)\right]| \\
\leqslant\ &\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}|U_{ns}(x_s(k)) - E\left[U_{ns}(x_s(k))\right]| + \\
&\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}\sup_{x\in\mathcal{S}^\star_s(k)}|U_{ns}(x) - U_{ns}(x_s(k))| + \\
&\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}\sup_{x\in\mathcal{S}^\star_s(k)}|E\left[U_{ns}(x)\right] - E\left[U_{ns}(x_s(k))\right]|.
\end{aligned} \tag{B.64}$$

By the smoothness condition on $k(\cdot)$ in Assumption 4(i), we may show that

$$\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}\sup_{x\in\mathcal{S}^\star_s(k)}|U_{ns}(x) - U_{ns}(x_s(k))| = O_P\left(n^{-2/5}\sqrt{\log n}\right) \tag{B.65}$$

and

$$\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}\sup_{x\in\mathcal{S}^\star_s(k)}|E\left[U_{ns}(x)\right] - E\left[U_{ns}(x_s(k))\right]| = O\left(n^{-2/5}\sqrt{\log n}\right). \tag{B.66}$$

By the Bernstein inequality for independent sequence (e.g., van der Vaart and Wellner, 1996) and noting that both $K^\star$ and d_1+d_2 are divergent to infinity at a polynomial rate of n, we can prove that, for $c_8 > 0$ sufficiently large

$$
\begin{aligned}
& \mathsf{P}\left(\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}\left|U_{ns}(x_s(k))-\mathsf{E}\left[U_{ns}(x_s(k))\right]\right| > c_8 n^{-2/5}\sqrt{\log n}\right) \\
\leqslant\ & \sum_{s=1}^{d_1+d_2}\sum_{k=1}^{K^\star}\mathsf{P}\left(\left|U_{ns}(x_s(k))-\mathsf{E}\left[U_{ns}(x_s(k))\right]\right| > c_8 n^{-2/5}\sqrt{\log n}\right) \\
\leqslant\ & O\left((d_1+d_2)\cdot K^\star\cdot\exp\{-c_8^\star\log n\}\right)=o(1),
\end{aligned}
$$

where $c_8^\star$ would be a sufficiently large positive constant when c_8 is large enough. Hence, we have

$$
\max_{1\leqslant s\leqslant d_1+d_2}\max_{1\leqslant k\leqslant K^\star}\left|U_{ns}(x_s(k))-\mathsf{E}\left[U_{ns}(x_s(k))\right]\right| = O\left(n^{-2/5}\sqrt{\log n}\right) \tag{B.67}
$$

By (B.64)–(B.67), we can complete the proof of Lemma B.7. □

LEMMA B.8. *Suppose that Assumptions 1(i), 4(i), 5(ii) and 6(i) are satisfied. Then, we have*

$$
\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\sup_{|\delta|\leqslant c_9\log^{1/2}n}\left|U_{ns}(x,\delta)-U_{ns}(x)-\mathsf{E}\left[U_{ns}(x,\delta)-U_{ns}(x)\right]\right| = O_P\left(n^{-3/5}\log^{3/2}n\right), \tag{B.68}
$$

where c_9 *is a sufficiently large positive constant.*

PROOF OF LEMMA B.8. Following the same line as in the proof of Lemma B.2, we can prove (B.68). Details are omitted here to save the space. □

LEMMA B.9. *Suppose that Assumptions 4(i) and 5(ii) are satisfied. Then, we have*

$$
\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\sup_{|\delta|\leqslant c_9\log^{1/2}n}\left|\mathsf{E}\left[\left[U_{ns}(x,\delta)-U_{ns}(x)\right]\right]+(nb_1)^{-1/2}V_s^c(x)\delta\right| = o\left(n^{-3/5}\log^{3/2}n\right). \tag{B.69}
$$

where $V_s^c(x)=f_s^c(x)f_{e,s}^c(0|x)$ *with* $f_s^c(\cdot)$ *and* $f_{e,s}^c(0|\cdot)$ *defined in Assumption 5(ii).*

PROOF OF LEMMA B.9. The proof of (B.69) is similar to the proof of Lemma B.3. Details are omitted here to save the space. □

PROOF OF LEMMA A.3. To save the space, we only prove (A.20) in details for the case of continuous covariates, and sketch the main idea for the proof of (A.21). Using Lemmas B.8 and B.9 as well as the argument in the proof of Lemma A.1, we may show that

$$
\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\sup_{|\delta|\leqslant c_9\log^{1/2}n}\left|U_{ns}(x,\delta)-U_{ns}(x)+(nb_1)^{-1/2}V_s^c(x)\delta\right| = O_P\left(n^{-3/5}\log^{3/2}n\right) \tag{B.70}
$$

and

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star} U_{ns}\left(x,\widehat{\delta}_s(x)\right)=o_P\left(n^{-3/5}\log^{3/2}n\right), \tag{B.71}$$

where $\widehat{\delta}_s(x)=(nb_1)^{1/2}\left[\widehat{Q}_{\tau,s}^c(x)-Q_{\tau,s}^c(x)\right]$.

Note that

$$\begin{aligned}
& P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n} -\delta\cdot U_{ns}(x,\delta)<(c_9c_{10})n^{-2/5}\log n\right)\\
\leqslant\ & P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n} -\delta\cdot U_{ns}(x,\delta)<(c_9c_{10})n^{-2/5}\log n,\ \Omega_{n4}\right)+P\left(\Omega_{n4}^c\right),
\end{aligned} \tag{B.72}$$

where c_{10} is a positive constant and Ω_{n4} denotes the event that

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n}\left\{-\delta\left[-(nb_1)^{-1/2}V_s^c(x)\delta+U_{ns}(x)\right]\right\}\geqslant 2(c_9c_{10})n^{-2/5}\log n.$$

By Assumption 5(ii), there exists a constant $c_\star>0$ such that

$$\min_{1\leqslant s\leqslant d_1+d_2}\inf_{x\in\mathcal{S}} V_s^c(x)\geqslant c_\star,$$

indicating that

$$\begin{aligned}
& \max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n}\left\{-\delta\left[-(nb_1)^{-1/2}V_s^c(x)\delta+U_{ns}(x)\right]\right\}\\
\geqslant\ & -c_9\log^{1/2}n\cdot\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}|U_{ns}(x)|+c_\star c_9^2 n^{-2/5}\log n.
\end{aligned}$$

Hence, if Ω_{n4}^c holds, we must have that

$$-c_9\log^{1/2}n\cdot\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}|U_{ns}(x)|+c_\star c_9^2 n^{-2/5}\log n<2(c_9c_{10})n^{-2/5}\log n,$$

leading to

$$P(\Omega_{n4}^c)\leqslant P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}|U_{ns}(x)|>(c_\star c_9-2c_{10})n^{-2/5}\log^{1/2}n\right)\to 0 \tag{B.73}$$

by letting c_9 be large enough. On the other hand, when the event

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n} -\delta\cdot U_{ns}(x,\delta)<(c_9c_{10})n^{-2/5}\log n$$

and Ω_{n4} jointly hold, we must have

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\sup_{|\delta|=c_9\log^{1/2}n}\left\{\delta\left[U_{ns}(x,\delta)-U_{ns}(x)+(nb_1)^{-1/2}V_s^c(x)\,\delta\right]\right\}\geqslant(c_9c_{10})n^{-2/5}\log n.$$

This, together with (B.70), implies that

$$P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n}-\delta\cdot U_{ns}(x,\delta)<(c_9c_{10})n^{-2/5}\log n,\ \Omega_{n4}\right)\to 0. \tag{B.74}$$

Combining (B.72)–(B.74), we can prove that

$$P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|=c_9\log^{1/2}n}-\delta\cdot U_{ns}(x,\delta)<(c_9c_{10})n^{-2/5}\log n\right)\to 0, \tag{B.75}$$

by choosing $c_9>0$ sufficiently large. Using the argument in the proof of (B.20), we may strengthen (B.75) to

$$P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\inf_{|\delta|\geqslant c_9\log^{1/2}n}|U_{ns}(x,\delta)|<c_{10}n^{-2/5}\log^{1/2}n\right)\to 0. \tag{B.76}$$

Observe that

$$\begin{aligned}&P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\left|\widehat{Q}_{\tau,s}^c(x)-Q_{\tau,s}^c(x)\right|>c_9n^{-2/5}\log^{1/2}n\right)\\ \leqslant\ &P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\left|\widehat{Q}_{\tau,s}^c(x)-Q_{\tau,s}^c(x)\right|>c_9n^{-2/5}\log^{1/2}n,\Omega_{n5}\right)+P\left(\Omega_{n5}^c\right),\end{aligned} \tag{B.77}$$

where Ω_{n5} denotes the event that

$$\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}U_{ns}\left(x,\widehat{\delta}_s(x)\right)<c_{10}n^{-2/5}\log^{1/2}n.$$

By (B.71), we have that for any $c_{10}>0$

$$P\left(\Omega_{n5}^c\right)=P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}U_{ns}\left(x,\widehat{\delta}_s(x)\right)\geqslant c_{10}n^{-2/5}\log^{1/2}n\right)\to 0. \tag{B.78}$$

On the other hand, by (B.76), we may prove that

$$P\left(\max_{1\leqslant s\leqslant d_1+d_2}\sup_{x\in\mathcal{S}_s^\star}\left|\widehat{Q}_{\tau,s}^c(x)-Q_{\tau,s}^c(x)\right|>c_9n^{-2/5}\log^{1/2}n,\Omega_{n5}\right)\to 0. \tag{B.79}$$

By (B.77)–(B.79), we complete the proof of (A.20).

The proof of (A.21) for the case of discrete covariate is very similar (and indeed simpler) as the involvement of the discrete kernel only affects the asymptotic bias term of the kernel quantile estimation, which has the order of $O(b_2)=O(n^{-2/5})$ by the choice of b_2. Details are omitted here to save the space. □

References

de la Peña, V. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability*, 27, 537–564.

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

Li, D., Li, Q. and Li, Z. (2018). Nonparametric quantile regression estimation with mixed discrete and continuous data. *Manuscript*.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, 75, 828–838.

Su, L. and White, H. (2012). Conditional independence specification testing for dependent processes with local polynomial quantile regression. *Advances in Econometrics*, 29, 355–434.

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