Ann Inst Stat Math
DOI 10.1007/s10463-014-0457-x

Quantile regression and variable selection of partial linear single-index model

Yazhao Lv · Riquan Zhang · Weihua Zhao · Jicai Liu

Received: 30 May 2012 / Revised: 9 December 2013

Abstract Partial linear single-index model (PLSIM) is a flexible and applicable model when investigating the underlying relationship between the response and the multivariate covariates. Most previous studies on PLSIM concentrated on mean regression, based on least square or likelihood approach. In contrast to this method, in this paper, we propose minimizing average check loss estimation (MACLE) procedure to conduct quantile regression of PLSIM. We construct an initial consistent quantile regression estimator of the parametric part base multi-dimensional kernels, and further promote the estimation efficiency to the optimal rate. We discuss the optimal bandwidth selection method and establish the asymptotic normality of the proposed MACLE estimators. Furthermore, we consider an adaptive lasso penalized variable selection method and establish its oracle property. Simulation studies with various distributed error and a real data analysis are conducted to show the promise of our proposed methods.

The research was supported in part by Scientific Research Fund of Zhejiang Provincial Education Department (Y201121276), National Natural Science Foundation of China (11171112, 11101114, 11201190), Doctoral Fund of Ministry of Education of China (20130076110004), the 111 Project of China (B14019) and The Natural Science Project of Jiangsu Province Education Department (13KJB110024).

Y. Lv · R. Zhang (✉) · W. Zhao · J. Liu
Department of Statistics, East China Normal University, Shanghai 200241,
People's Republic of China
e-mail: zhangriquan@163.com

Y. Lv
Institute of Operational Research and Cybernetics, Hangzhou Dianzi University,
Hangzhou 310018, People's Republic of China

R. Zhang
Department of Mathematics, Shanxi Datong University, Datong 037009,
Shanxi, People's Republic of China

Published online: 30 March 2014

Keywords Single index · Partial linear · Quantile regression · Asymptotic normality · Minimizing average check loss estimation · Variable selection · Adaptive Lasso

1 Introduction

Semi-parametric model has been popular in the literature recently due to the explanatory power and the flexibility of modeling with multivariate covariates. See Ruppert et al. (2003) and the reference therein for more comprehensive review. Among the semi-parametric modeling literature, the partial linear single-index model (PLSIM) plays an important role, which has form

$$Y = g(\mathbf{X}^T\theta) + \mathbf{Z}^T\beta + \varepsilon, \tag{1}$$

where $\mathbf{X} \in \mathrm{R}^p$ and $\mathbf{Z} \in \mathrm{R}^q$ are covariates of the response variable Y, ε is the model error. $g(\cdot)$ is an unknown differentiable function, θ and β are unknown parameters. For identifiability, we assume $||\theta|| = 1$ and the first nonzero element of θ is positive. For convenience, we call θ index parameter, $g(\cdot)$ index function and β linear parameter.

Partial linear single-index model, as a combination of the single-index model (SIM), the widely used dimension reduction approach to avoid the "curse of dimensionality", and the partial linear model, the most popular model in semi-parametric regression (see Härdle et al. 2007), has attracted many researcher's attention and various methods have been proposed to estimate its unknown parameters and nonparametric function. Carroll et al. (1997) proposed full iteration algorithm using local linear method. As observed by Yu and Ruppert (2002), the full iteration approaches may unstable in computation. Yu and Ruppert (2002) proposed penalized spline approach to alleviate the computational difficulties, which is essentially a flexible parametric model. Xia and Härdle (2006) proposed the well-known minimizing average variance estimation (MAVE) method based on local linear approach, which gives the $\sqrt{n}$ consistent estimators of the parameters by constructive approach. Liang et al. (2010) investigate PLSIM by profile likelihood, get efficient estimation of the parameters and also consider the variable selection problem of PLSIM.

Most existing estimation procedures for PLSIM concentrate on mean regression, based on either least squares or likelihood approach. In contrast to mean linear regression, quantile regression (QR) proposed by Koenker and Basset (1978) has been widely used as a robust alternative to explore the underlying relationship between the covariates and the response, see Koenker (2005) for comprehensive review. For nonparametric regressions, Chaudhuri (1991) introduced local polynomial QR in a general multivariate setting, which is flexible but usually inapplicable in practice due to the "curse of dimensionality". In order to avoid the "curse of dimensionality" in multivariate nonparametric QR, Chaudhuri et al. (1997) considered dimension reduction by single-index modeling approach and developed the average derivative approach (Härdle and Stoker 1989) to estimate the index parameters directly. Recently, Wu et al. (2010) proposed a practical estimation procedure for SIM based on the initial estimator provided by Chaudhuri et al. (1997). Jiang et al. (2012) consider the composite quantile regression [CQR, proposed by Zou and Yuan (2008)] of SIM. Kong and Xia (2012) proposed an adaptive estimation procedure, and the estimator obtained is consistent with probability 1. However, there has been little research on the QR of PLSIM.

In real data analysis, the covariates of model (1) may include many irrelevant covariates, especially for high-dimensional $\mathbf{X}$ and $\mathbf{Z}$. In this case, sparse model is often considered superior, due to the enhancements of model predictability and interpretability. Since semi-parametric models, like (1), involve both nonparametric and parametric parts need to be estimated precisely, it is challenging to perform variable selection. Liang et al. (2010) proposed an variable selection method for mean regression of (1) by combining the profile likelihood method and the SCAD (Fan and Li 2001) penalized approach. In quantile regression of semi-parametric model, Kai et al. (2011) considered the variable selection of partial linear varying coefficient model. About the composite quantile regression of SIM, Fan et al. (2013) proposed an variable selection method by SCAD penalized method. Although, there has also been little research on the variable selection in QR PLSIM.

Motivated by the above observations, we extend the quantile regression methodology of semi-parametric models to PLSIM and propose an variable selection method, respectively. There are three major contributions of the present study. (1) We construct an initial consistent quantile regression estimation of the parameters θ and β based on multi-dimensional kernels, and then promote the estimation efficiency using one-dimensional index kernels. The final estimators derived can reach the optimal convergence rate. (2) The proposed estimation method is not sensitive to the bandwidth selection and the common "under smoothing" problem is not necessary in out estimation procedure. (3) We propose an variable selection method for quantile regression of PLSIM by combining the adaptive lasso penalized method with our proposed estimation method, which enjoys the oracle properties defined by Zou (2006).

Monte Carlo simulations with various non-normal errors and parameters changing with quantiles are conducted to show the performance of our estimation and variable selection method. In real data analysis, we apply our method to Boston housing data. Simulation and real data analysis validate the fine property of the minimizing average check loss estimation (MACLE) procedure and the adaptive lasso penalized MACLE variable selection method.

The paper is organized as follows. In Sect. 2, we introduce the estimation methodology and the calculation procedure and present the asymptotic properties of the estimators. In Sect. 3, adaptive lasso penalized quantile regression method is proposed and its oracle property is presented. Monte Carlo simulations with various error distributions are presented in Sect. 4. In Sect. 5, we apply the proposed estimation and variable selection methods to Boston housing data. Regularity conditions and technical proofs are given in the Appendix A and B.

2 Estimation methodology

In this section, we develop the semi-parametric quantile regression theory to PLSIM. Let $\rho_\tau(u) = u[\tau - I(u < 0)]$ be the check loss function at $\tau \in (0, 1)$. Quantile regression is often used to estimate the conditional quantile functions of Y, which is defined as

$$q_\tau(\mathbf{x},\mathbf{z}) = \underset{a}{\operatorname{argmin}}\ \mathrm{E}\left\{\rho_\tau(Y-a)|(\mathbf{X},\mathbf{Z}) = (\mathbf{x},\mathbf{z})\right\}.$$

The partial linear single-index model assumes that the τ-th conditional quantile function of Y can be expressed as $q_\tau(\mathbf{x},\mathbf{z}) = g_\tau(\mathbf{x}^T\theta_\tau) + \mathbf{z}^T\beta_\tau$.

Suppose $\{\mathbf{X}_i,\ \mathbf{Z}_i,\ Y_i\}_{i=1}^n$ is an independent and identically distributed sample from the model

$$Y = g_\tau\left(\mathbf{X}^T\theta_\tau\right) + \mathbf{Z}^T\beta_\tau + \varepsilon_\tau, \tag{2}$$

where the τth conditional quantile of ε_τ is zero when given $(\mathbf{X},\mathbf{Z})$. For notational convenience, we abbreviate $g_\tau(\cdot)$ as $g(\cdot)$. Theoretically, the true parameter vector (θ_τ,β_τ) satisfies that

$$(\theta_\tau,\beta_\tau) = \underset{\substack{\theta,\beta\\ ||\theta||=1,\theta_1>0}}{\operatorname{argmin}}\ \mathrm{E}\left[\rho_\tau\left(Y - g\left(\mathbf{X}^T\theta\right) - \mathbf{Z}^T\beta\right)\right]. \tag{3}$$

Note that

$$\mathrm{E}\left[\rho_\tau\left(Y - g\left(\mathbf{X}^T\theta\right) - \mathbf{Z}^T\beta\right)\right] = \mathrm{E}\left\{\mathrm{E}\left[\rho_\tau(Y - g(\mathbf{X}^T\theta) - \mathbf{Z}^T\beta)|\mathbf{X}^T\theta\right]\right\}, \tag{4}$$

where $\mathrm{E}[\rho_\tau(Y-g(\mathbf{X}^T\theta)-\mathbf{Z}^T\beta)|\mathbf{X}^T\theta]$ is the conditional expected check loss on $\mathbf{X}^T\theta$. In the following context, we will construct an empirical form of (4), by minimizing which we can derive our estimations of the unknown parameters and the unknown index function.

For given (θ,β), when $\mathbf{X}_i^T\theta$ closed to u, $g(\mathbf{X}_i^T\theta)$ can be approximated linearly by

$$g\left(\mathbf{X}_i^T\theta\right) \simeq g(u) + g'(u)\left(\mathbf{X}_i^T\theta - u\right).$$

Then the local linear approximate of $\mathrm{E}[\rho_\tau(Y - g(\mathbf{X}^T\theta) - \mathbf{Z}^T\beta)|\mathbf{X}^T\theta = u]$ will be

$$\sum_{i=1}^n \rho_\tau\left(Y_i - g(u) - g'(u)\left(\mathbf{X}_i^T\theta - u\right) - \mathbf{Z}_i^T\beta\right)\omega_{i0},$$

where ω_{i0} are non-negative weights with $\sum_{i=1}^n\omega_{i0} = 1$, typically centering at $\mathbf{x}$ and $\mathbf{x}^T\theta$. By averaging on $u_j = \mathbf{X}_j^T\theta,\ j = 1,\cdots,n$, we can get an empirical approximation of (4) by

$$\frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n \rho_\tau(Y_i - g(u_j) - g'(u_j)\mathbf{X}_{ij}^T\theta - \mathbf{Z}_i^T\beta)\omega_{ij}, \tag{5}$$

where $\mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j$, and ω_{ij} satisfies $\sum_{i=1}^n\omega_{ij} = 1$ for $\forall j = 1,\cdots,n$.

We define the quantile regression estimates of $(\theta_\tau,\ \beta_\tau)$ as

$$(\hat{\theta},\ \hat{\beta}) = \underset{\substack{\theta,\beta \\ ||\theta||=1,\theta_1>0}}{\operatorname{argmin}} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau \left(Y_i - g(u_j) - g'(u_j)\mathbf{X}_{ij}^T\theta - \mathbf{Z}_i^T\beta\right)\omega_{ij}. \tag{6}$$

The above estimation procedure can be called the MACLE method, which is parallel to the MAVE method in mean regression of PLSIM. The weights ω_{ij} can be firstly chosen as $\omega_{ij} = \frac{\mathbf{H}_b(\mathbf{X}_{ij})}{\sum_{l=1}^{n}\mathbf{H}_b(\mathbf{X}_{lj})}$, where $\mathbf{H}(\cdot)$ is a p-dimensional kernel density function, and $H_b(\cdot) = \frac{1}{b^p}\mathbf{H}(\cdot/b)$, b is the bandwidth. The initial estimations of $(\theta_\tau, \beta_\tau)$ can be derived by following steps:

Step 1. For any given θ and β, $g(\mathbf{X}_j^T\theta)$, $g'(\mathbf{X}_j^T\theta)$ can be estimated by

$$(\bar{a}_j, \bar{b}_j) = \underset{a_j,b_j}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(Y_i - a_j - b_j\mathbf{X}_{ij}^T\theta - \mathbf{Z}_i^T\beta\right)\omega_{ij}, \tag{7}$$

for $j = 1, \ldots, n$.
Step 2. The values of θ and β can be updated by

$$(\bar{\theta}, \bar{\beta}) = \underset{\substack{\theta,\beta \\ ||\theta||=1,\theta_1>0}}{\operatorname{argmin}} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau \left(Y_i - \bar{a}_j - b_jX_{ij}^T\theta - \mathbf{Z}_i^T\beta\right)\omega_{ij}. \tag{8}$$

Step 3. Iterate **Step 1** and **Step 2** until convergence. Define the final estimation of θ_τ and β_τ by $\tilde{\theta}$ and $\tilde{\beta}$.

Though the estimators $\tilde{\theta}$ and $\tilde{\beta}$ based on multi-dimensional kernels may loose efficiency, we can show that under some regular conditions they are consistent to the true values.

Lemma 1 *Let $\tilde{\theta}$ and $\tilde{\beta}$ be the estimators derived above. Suppose that condition* A.1–A.6 *in Appendix hold, $b \to 0$, and $nb^{p+2}/\log n \to \infty$. If we start the estimation procedure with θ satisfies that $\theta^T\theta_\tau \neq 0$, then we have*

$$\tilde{\theta} - \theta_\tau = o_p(1), \quad \tilde{\beta} - \beta_\tau = o_p(1).$$

Proof The proof of Lemma 1 will be presented in Appendix B. □

After we get the initial estimates of θ and β, we can improve the estimation efficiency by choosing $\omega_{ij} = K_h(\mathbf{X}_{ij}^T\tilde{\theta})/\sum_{l=1}^{n}K_h(\mathbf{X}_{lj}^T\tilde{\theta})$, where $K(\cdot)$ is an one-dimensional kernel function, $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$, h is the bandwidth. The estimation algorithm can be divided as follows:

Step 4. Given $(\tilde{\theta}, \tilde{\beta})$. Standardize $\tilde{\theta}$ s.t. $||\tilde{\theta}|| = 1$ and $\tilde{\theta}_1 > 0$. Let $a_j := g(\mathbf{X}_j^T\tilde{\theta}), b_j := g'(\mathbf{X}_j^T\tilde{\theta}), j = 1, \ldots, n$. Given $\tilde{\theta}$, $\tilde{\beta}$, get the estimates of

$a_j,\ b_j,\ j=1,\ldots,n$ by

$$(\tilde{a}_j,\ \tilde{b}_j)=\underset{a_j,b_j}{\operatorname{argmin}}\sum_{i=1}^{n}\rho_\tau\left[y_i-a_j-b_j\mathbf{X}_{ij}^T\tilde{\theta}-\mathbf{Z}_i^T\tilde{\beta}\right]\omega_{ij}\ \text{ for }\ j=1,\ldots,n, \tag{9}$$

with the bandwidth h chosen by (13) in Sect. 3.

Step 5. Given $\tilde{a}_j,\ \tilde{b}_j,\ j=1,\ldots,n$, update the estimates $(\tilde{\theta},\ \tilde{\beta})$ by

$$(\tilde{\theta},\ \tilde{\beta})=\underset{\theta,\beta}{\operatorname{argmin}}\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left[Y_i-\tilde{a}_j-\tilde{b}_j(\mathbf{X}_{ij}^T\theta)-\mathbf{Z}_i^T\beta\right]\omega_{ij}, \tag{10}$$

with ω_{ij} evaluated at $\tilde{\theta}$ and h from step 1;

Step 6. Repeat **Steps 4** and **Step 5** until convergence. Let the final estimation of (θ_τ,β_τ) by $(\hat{\theta},\hat{\beta})$.

Remark 1 In the above algorithm, θ is standardized as follows: $\theta=\operatorname{sign}(\theta_1)\theta/||\theta||$, where $\operatorname{sign}(\theta_1)$ is the sign of the first component of θ.

Here, we can call the above estimation of the parameter (θ_τ,β_τ) as the refined minimizing average check loss estimation (MACLE). This refined MACLE procedure was initially proposed by Wu et al. (2010), which is similar as the refined MAVE method proposed by Xia and Härdle (2006). After obtaining $\hat{\theta},\ \hat{\beta}$, the $g(u)$ can be estimated by the solution of a_j in (9) with $\mathbf{X}_j^T\hat{\theta}$ replaced by u, denoted by $\hat{g}(u;h,\hat{\theta},\hat{\beta})$.

In the following, we present the asymptotic property of the proposed MACLE. Let $f_Y(\cdot|\mathbf{X}^T\theta)$ and $F_Y(\cdot|\mathbf{X}^T\theta)$ be the density function and cumulative distribution function of Y condition on $\mathbf{X}^T\theta$, respectively. Let $f_{\mathcal{U}_\tau}(\cdot)$ be the marginal density function of the index $\mathbf{X}^T\theta_\tau$. We choose the kernel $K(\cdot)$ as a symmetric density function, and let

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

Theorem 1 *Suppose the conditions* A.1–A.7 *given in the Appendix hold, then we have*

$$\sqrt{n}\begin{pmatrix}\hat{\theta}-\theta_\tau\\ \hat{\beta}-\beta_\tau\end{pmatrix}\xrightarrow{L}N\left(0,\tau(1-\tau)\mathcal{D}_1^{-1}\mathcal{D}_0\mathcal{D}_1^{-1}\right), \tag{11}$$

where $\xrightarrow{L}$ *stands for convergence in distribution,* $\mathcal{D}_0=E(\mathcal{D}),\ \mathcal{D}_1=E[f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}^T\theta_\tau)\mathcal{D}],\ \mathcal{D}=\begin{pmatrix}g'(\mathbf{X}^T\theta_\tau)\tilde{\mathbf{X}}\\ \tilde{\mathbf{Z}}\end{pmatrix}^{\otimes 2},\ \tilde{\mathbf{X}}=\mathbf{X}-E(\mathbf{X}|\mathbf{X}^T\theta_\tau)$ *and* $\tilde{\mathbf{Z}}=\mathbf{Z}-E(\mathbf{Z}|\mathbf{X}^T\theta_\tau)$.

From Theorem 1, we can get $\sqrt{n}$-consistent estimators of (θ_τ,β_τ), then by (9), we can get the estimation of the nonparametric function. We present the asymptotic properties of the nonparametric part in the following:

Theorem 2 *Suppose u is an interior point of $f_{\mathcal{U}_\tau}(\cdot)$ and the regular conditions* A.1–A.7 *in the Appendix hold, we have*

$$\sqrt{nh}\left\{\hat{g}(u;h,\hat{\theta},\hat{\beta})-g(u)-\frac{1}{2}g''(u)\mu_2h^2\right\}\xrightarrow{L}N(0,\Gamma_\tau(u)),\tag{12}$$

where $\Gamma_\tau(u)=\tau(1-\tau)\nu_0 f_{\mathcal{U}_\tau}(u)^{-1}f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)^{-2}$, $f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)$ is the density of Y at $q_\tau(\mathbf{X},\mathbf{Z})$ condition on $\mathbf{X}^T\theta_\tau=u$.

According to Theorem 2, when the sample size is large, the optimal bandwidth could be derived by minimizing the asymptotic mean squared error (AMSE) from Theorem 2,

$$h_{\text{opt}}=\left\{\frac{\tau(1-\tau)}{f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)^2}\right\}^{1/5}\times\left\{\frac{\int K^2(t)\mathrm{d}t}{n[\int t^2K(t)\mathrm{d}t]^2[g''(u)]^2f_{\mathcal{U}_\tau}(u)}\right\}^{1/5}.$$

This calculation indicates that the MACLE estimator of $g(\cdot)$ enjoys the optimal rate of convergence $n^{-2/5}$. While the optimal bandwidth h_{opt} depends on some unknown values such as $f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)$, $g''(u)$ and $f_{\mathcal{U}_\tau}(u)$, whose estimations may be computational intensive. Following the similar argument as Yu and Jones (1998), we take the following rule-of-thumb bandwidth h_τ in this algorithm,

$$h_\tau=h_m\left\{\tau(1-\tau)\big/\phi(\Phi^{-1}(\tau))^2\right\}^{1/5},\tag{13}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and the cumulative distribution function of the standard normal distribution, respectively. h_m is the optimal bandwidth used in mean regression, which can be easily obtained by the plug-in method (see Ruppert et al. 1995). The approximation by (13) provides an easy approach to get the optimal bandwidth for quantile regression. We recommend Yu and Jones (1998) for detail discussion of this bandwidth approximation method.

3 Variable selection

In practice, the true model is often unknown, which may include many variables in the covariates. An under-fitted model will yield biased estimates and large residuals, while an over-fitted model may reduce the estimation efficiency. This motivates us to consider variable selection in QR of PLSIM.

We adopt the adaptive lasso idea from Zou (2006). Suppose we first fit the model by including all the predictors. Theorem 1 says that MACLE estimator, denoted by $(\hat{\theta}^{QR},\hat{\beta}^{QR})$, is $\sqrt{n}$ consistent. Then, we use $(\hat{\theta}^{QR},\hat{\beta}^{QR})$ to construct the adaptively weighted lasso penalized target function as

$$\begin{aligned}G_n(\theta,\beta)=&\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i-g\left(\mathbf{X}_i^T\theta\right)-\mathbf{Z}_i^T\beta\right)\omega_{ij}\\&+\lambda_1\sum_{j=1}^{p}\frac{|\theta_j|}{|\hat{\theta}_j^{QR}|^2}+\lambda_2\sum_{m=1}^{q}\frac{|\beta_m|}{|\hat{\beta}_m^{QR}|^2}.\end{aligned}\tag{14}$$

For the given tuning parameters (λ_1, λ_2), we obtain the penalized estimators by minimizing $G(\theta, \beta)$ with respect to θ and β with constrains $||\theta|| = 1$ and the first nonzero element of θ is positive. For the sake of simplicity, we denote the resulting estimators by $\hat{\theta}^{\lambda_1}$ and $\hat{\beta}^{\lambda_2}$.

Remark 2 Throughout this paper, we can choose different regularization parameters λ_1 and λ_2 to select the important variables of the index covariates and partial linear covariates separately. For the purpose of selecting index covariates only, we can simply set $\lambda_2 = 0$. Similarly, if we have only interest in selecting partial linear covariates, then we can set $\lambda_1 = 0$.

Remark 3 Other variable selection methods such as SCAD proposed by Fan and Li (2001) can be also used here, and the oracle property can be derived similarly. For the sake of easy computation, we choose adaptive lasso method here, which can be solved conveniently by linear programming.

Note that we need two tuning parameters in (14), λ_1 and λ_2, imposed on the linear part and the single-index part, respectively. Following the approach of Fan and Li (2004), we set $\lambda_1 = \lambda\ \mathrm{SE}(\mathbf{X}^T\hat{\theta})$ and $\lambda_2 = \lambda\ \mathrm{SE}(\hat{\beta}^T\mathbf{Z})$, where λ is the tuning parameter, and $\mathrm{SE}(\mathbf{X}^T\hat{\theta})$ and $\mathrm{SE}(\mathbf{Z}^T\hat{\beta})$ are the standard errors of the unpenalized MACLE of θ and β, respectively. The tuning parameters λ can be chosen optimally by BIC criteria. Following Wang and Leng (2007), denote

$$\mathrm{BIC}(\lambda) = \log P_\tau(\lambda) + \log(n)/n\mathrm{DF}_\lambda, \tag{15}$$

where $P_\tau(\lambda) = \sum_{j=1}^n \sum_{i=1}^n \rho_\tau(Y_i - \hat{g}(\mathbf{X}_i^T\hat{\theta}^{\lambda_1}) - \mathbf{Z}_i^T\hat{\beta}^{\lambda_2})\omega_{ij}$, DF_λ is the number of nonzero coefficients of both $\hat{\theta}^{\lambda_1}$ and $\hat{\beta}^{\lambda_2}$. We let $\hat{\lambda}(\mathrm{BIC}) = \mathrm{argmin}_\lambda \mathrm{BIC}(\lambda)$. The performance of $\hat{\lambda}(\mathrm{BIC})$ will be examined in our simulation studies in the next section.

Let $\mathcal{A}_\theta = \{j : \theta_j \neq 0\}$ and $\mathcal{A}_\beta = \{m : \beta_m \neq 0\}$. Without loss of generality, it is assumed that the correct model has regression coefficients $\theta_\tau = (\theta_{1\tau}, \theta_{2\tau})$ and $\beta_\tau = (\beta_{1\tau}, \beta_{2\tau})$, where $\theta_{1\tau}$ and $\beta_{1\tau}$ are p_0 and q_0 nonzero components of θ_τ and β_τ, respectively, and $\theta_{2\tau}$ and $\beta_{2\tau}$ are $p - p_0$ and $q - q_0$ vectors with zeros. Thus, $\mathcal{A}_\theta = \{1, \cdots, p_0\}$ and $\mathcal{A}_\beta = \{1, \cdots, q_0\}$. In addition, we define $\mathbf{X}_1$ and $\mathbf{Z}_1$ in such a way that they consist of the first p_0 and q_0 elements of $\mathbf{X}$ and $\mathbf{Z}$, respectively. We define $\tilde{\mathbf{X}}_1 = \mathbf{X}_1 - \mathrm{E}(\mathbf{X}_1|\mathbf{X}_1^T\theta_{1\tau})$ and $\tilde{\mathbf{Z}}_1 = \mathbf{Z}_1 - \mathrm{E}(\mathbf{Z}_1|\mathbf{X}_1^T\theta_{1\tau})$ similarly as in Sect. 3. In what follows, we show the adaptive lasso penalized MACLE estimators enjoy the oracle properties.

Theorem 3 *(Oracle property). Under the regular condition* A.1–A.8 *in Appendix, if* $\lambda_i/\sqrt{n} \to 0$ *and* $\lambda_i \to \infty$ *for* $i = 1, 2$, *then the adaptive lasso penalized estimators* $\hat{\theta}^{\lambda_1}$ *and* $\hat{\beta}^{\lambda_2}$ *must satisfy*

1. *Consistency in selection:* $Pr(\{j : \hat{\theta}_j^{\lambda_1} \neq 0\} = \mathcal{A}_\theta) \to 1$ *and* $Pr(\{m : \hat{\beta}_m^{\lambda_2} \neq 0\} = \mathcal{A}_\beta) \to 1$.

2. *Asymptotic normality:*

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_1^{\lambda_1} - \theta_{1\tau} \\ \hat{\beta}_1^{\lambda_2} - \beta_{1\tau} \end{pmatrix} \xrightarrow{L} N\left(0, \tau(1-\tau)\mathcal{D}_{1*}^{-1}\mathcal{D}_{0*}\mathcal{D}_{1*}^{-1}\right), \tag{16}$$

where $\hat{\theta}_1^{\lambda_1}$ *and* $\hat{\beta}_1^{\lambda_2}$ *denote the first* p_0 *and* q_0 *elements of* $\hat{\theta}^{\lambda_1}$ *and* $\hat{\beta}^{\lambda_2}$*, respectively,* $\mathcal{D}_{1*} = E[f(0|\mathbf{X}_1^T\theta_{1\tau})\mathcal{D}_*]$, $\mathcal{D}_* = \begin{Bmatrix} g'(\mathbf{X}_1^T\theta_{1\tau})\tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{Z}}_1 \end{Bmatrix}^{\otimes 2}$ *and* $\mathcal{D}_{0*} = E(\mathcal{D}_*)$.

4 Monte Carlo simulations

In this section, we consider three examples parameters changing with quantiles or with various distributed errors to assess the finite sample performance of the proposed estimation and variable selection methods, respectively.

Example 4.1 In this example, we generate 200 random samples, each consisting of $n = 100, 200$ observations, from the model

$$Y = \sin\left(\pi(\mathbf{X}^T\theta - a)/(b-a)\right) + \mathbf{Z}^T\beta + \sigma\varepsilon, \tag{17}$$

where $\mathbf{X}$ are all uniformly distributed on $[0, 1]^3$ with correlation as $\text{cor}(\mathbf{X}_i, \mathbf{X}_j) = 0.5^{|i-j|}$, $\mathbf{Z}_1, \mathbf{Z}_2$ are standard normal distributed with $\text{cor}(\mathbf{Z}_1, \mathbf{Z}_2) = 0.5$, $\mathbf{Z}_3$ is discrete distributed on $[-1, 1]$ with probability $(0.4, 0.6)$, $\mathbf{X}$. $(\mathbf{Z}_1, \mathbf{Z}_2)$ and $\mathbf{Z}_3$ are independent. $\theta = (1, 1, 1)/\sqrt{3}$, $a = 0.3912$, $b = 1.3409$, $\beta = (2, 0.5, 1)$, $\sigma = 0.1$. This model is similar as that in Carroll et al. (1997), while with more complicate linear covariates. In this simulation, we consider four error distributions for ε: $N(0, 1)$, $t(3)$, standard Cauchy and mixture of normals $0.9N(0, 1) + 0.1N(0, 10^2)$. Assume $\mathbf{X}$, $\mathbf{Z}$ and ε are mutually independent. For each error distribution, the parameters $(\theta_\tau, \beta_\tau)$, and the index function $g(\cdot)$ are estimated via series of quantile regression with $\tau = 0.3$, 0.5 and 0.7, respectively. The bias and standard deviation (Std) of the estimates of $(\theta_\tau, \beta_\tau)$ are summarized in Tables 1, 2, 3 and 4. From the tables, we can see that the quantile regression estimation of PLSIM is robust to different distributed error. Particularly, when the error follows standard Cauchy and $\tau = 0.5$, the box plots of the 200 estimates of $(\theta_\tau, \beta_\tau)$ are presented in Fig. 1. The median of the index function estimates is presented in Fig. 2. We can see that the estimates of the parameters are centered around the true values, and the median of the 200 estimated curves (black dashed line) is close to the true curve (green line).

Example 4.2 Consider the following model where the parametric part changes with the quantile,

$$Y = 5 * \exp(-(\mathbf{X}^T\theta_\tau)^2) + \mathbf{X}^T\theta_\tau + z_1e^\tau + z_2e^{1-\tau} + z_3e^{\tau(1-\tau)} + \varepsilon_\tau,$$

where $\mathbf{X} = (x_1, \ldots, x_5)^T = \Sigma^{1/2}(\mathbf{u}_1, \ldots, \mathbf{u}_5)^T$ with $\mathbf{u}_1, \ldots, \mathbf{u}_5$ mutually independent withe each $\mathbf{u}_i \sim Uniform(0, 1)$ and $\Sigma = (0.5^{|i-j|})_{1\le i,j\le 5}, \theta_\tau =$

Table 1 Monte Carlo study for Example 5.1 with normal distributed error

Estimator	Bias and Std	θ_1	θ_2	θ_3	β_1	β_2	β_3
$n = 100$							
MAVE	Bias	−0.0010	−0.0012	0.0014	−0.0019	0.0004	−0.0002
	Std	0.0177	0.0200	0.0156	0.0129	0.0126	0.0107
QR (0.5)	Bias	−0.0022	−0.0004	0.0015	−0.0019	0.0001	−0.0007
	Std	0.0198	0.0230	0.0187	0.0146	0.0157	0.0134
QR (0.3)	Bias	−0.0021	−0.0004	0.0013	−0.0021	0.0007	−0.0004
	Std	0.0201	0.0240	0.0187	0.0160	0.0163	0.0141
QR (0.7)	Bias	−0.0011	−0.0013	0.0013	−0.0022	−0.0001	−0.0013
	Std	0.0201	0.0227	0.0184	0.0159	0.0169	0.0137
$n = 200$							
QR (0.3)	Bias	−0.0051	−0.0027	0.0068	−0.0002	0.0018	0.0136
	Std	0.0174	0.0201	0.0180	0.0144	0.0144	0.0130
QR (0.5)	Bias	−0.0051	−0.0050	0.0094	0.0002	−0.0001	0.0084
	Std	0.0147	0.0171	0.0167	0.0128	0.0130	0.0124
QR (0.7)	Bias	−0.0048	−0.0045	0.0085	0.0009	0.0010	0.0072
	Std	0.0139	0.0187	0.0176	0.0135	0.0131	0.0114

Table 2 Monte Carlo study for Example 5.1 with $t(3)$ distributed error

Estimator	Bias and Std	θ_1	θ_2	θ_3	β_1	β_2	β_3
$n = 100$							
MAVE	Bias	0.0004	−0.0057	0.0023	−0.0002	−0.0002	0.0011
	Std	0.0324	0.0375	0.0320	0.0216	0.0216	0.0187
QR (0.5)	Bias	−0.0003	−0.0043	0.0027	−0.0004	−0.0006	−0.0003
	Std	0.0245	0.0313	0.0255	0.0184	0.0170	0.0156
QR (0.3)	Bias	−0.0021	−0.0040	0.0037	−0.0009	−0.0011	−0.0002
	Std	0.0263	0.0355	0.0290	0.0206	0.0202	0.0176
QR (0.7)	Bias	0.0003	−0.0047	0.0025	−0.0011	0.0008	−0.0007
	Std	0.0249	0.0307	0.0258	0.0200	0.0180	0.0169
$n = 200$							
QR (0.3)	Bias	−0.0021	−0.0067	0.0074	−0.0009	−0.0013	0.0164
	Std	0.0202	0.0256	0.0229	0.0163	0.0184	0.0152
QR (0.5)	Bias	−0.0043	−0.0069	0.0101	−0.0027	0.0014	0.0082
	Std	0.0170	0.0230	0.0180	0.0140	0.0141	0.0146
QR (0.7)	Bias	−0.0060	−0.0046	0.0095	−0.0007	−0.0013	0.0057
	Std	0.0185	0.0233	0.0188	0.0156	0.0156	0.0140

$(\tau, 2\tau, 0, 2\tau - 1, \tau - 1)^T / \sqrt{10\tau^2 - 6\tau + 2}$, $\mathbf{Z} = (z_1, z_2, z_3)^T \sim N(0, 1)^3$, $\beta_\tau = (e^\tau, e^{1-\tau}, e^{\tau(1-\tau)})^T$, $\varepsilon_\tau \sim Uniform(-\tau, 1 + \tau)$.

Table 3 Monte Carlo study for Example 5.1 with standard Cauchy distributed error

Estimator	Bias and Std	θ_1	θ_2	θ_3	β_1	β_2	β_3
$n = 100$							
MAVE	Bias	−0.0320	−0.1024	−0.0981	0.0896	−0.2025	−0.0798
	Std	0.1974	0.3507	0.2922	1.2035	1.9508	0.8807
QR (0.5)	Bias	0.0024	−0.0070	0.0010	−0.0012	−0.0016	−0.0029
	Std	0.0324	0.0430	0.0351	0.0250	0.0283	0.0215
QR (0.3)	Bias	−0.0004	−0.0076	0.0028	−0.0024	−0.0021	−0.0025
	Std	0.0373	0.0522	0.0431	0.0334	0.0370	0.0275
QR (0.7)	Bias	0.0048	−0.0102	0.0010	0.0012	−0.0039	−0.0044
	Std	0.0360	0.0480	0.0366	0.0311	0.0311	0.0265
$n = 200$							
QR (0.3)	Bias	−0.0008	−0.0127	0.0003	−0.0009	0.0016	0.0230
	Std	0.0401	0.0491	0.1054	0.0281	0.0285	0.0263
QR (0.5)	Bias	−0.0064	−0.0043	0.0011	0.0000	−0.0001	0.0074
	Std	0.0263	0.0290	0.0981	0.0181	0.0201	0.0172
QR (0.7)	Bias	−0.0039	−0.0130	0.0053	−0.0009	0.0004	0.0050
	Std	0.0262	0.0984	0.0533	0.0236	0.0213	0.0195

Table 4 Monte Carlo study for Example 5.1 with mixture of normals distributed error

Estimator	Bias and Std	θ_1	θ_2	θ_3	β_1	β_2	β_3
$n = 100$							
MAVE	Bias	−0.0063	0.0002	−0.0040	−0.0008	0.0043	−0.0002
	Std	0.0595	0.0713	0.0550	0.0410	0.0439	0.0342
QR (0.5)	Bias	−0.0028	0.0010	0.0003	−0.0004	0.0012	0.0006
	Std	0.0222	0.0268	0.0229	0.0176	0.0162	0.0136
QR (0.3)	Bias	−0.0020	−0.0003	0.0005	−0.0004	0.0011	0.0006
	Std	0.0245	0.0297	0.0237	0.0197	0.0181	0.0157
QR (0.7)	Bias	−0.0038	0.0026	−0.0004	0.0013	0.0008	0.0007
	Std	0.0230	0.0270	0.0234	0.0174	0.0187	0.0160
$n = 200$							
QR (0.3)	Bias	−0.0038	−0.0036	0.0062	−0.0001	−0.0014	0.0146
	Std	0.0180	0.0235	0.0214	0.0157	0.0161	0.0140
QR (0.5)	Bias	−0.0045	−0.0047	0.0082	−0.0000	−0.0005	0.0087
	Std	0.0168	0.0193	0.0163	0.0128	0.0133	0.0120
QR (0.7)	Bias	−0.0052	−0.0033	0.0076	−0.0006	0.0012	0.0068
	Std	0.0141	0.0185	0.0176	0.0137	0.0149	0.0128

For the combinations of different sample size n and quantile level τ, the estimation efforts of 200 times MACLE are summarized in Table 5. We can see that the MALCE method performs similarly when the parameters varying with τ.

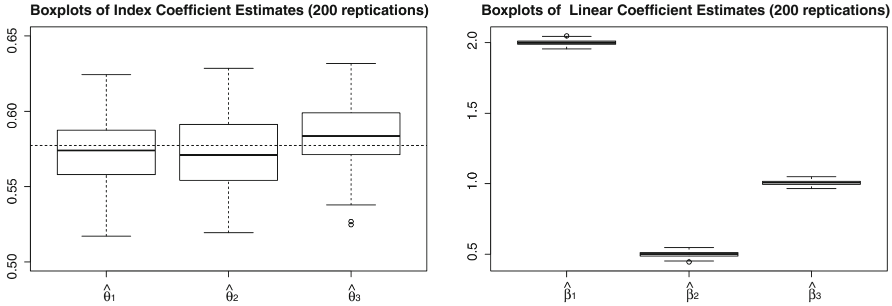

Fig. 1 Box plot of parameter estimations for Example 5.1 with Cauchy distributed error when $\tau = 0.5$

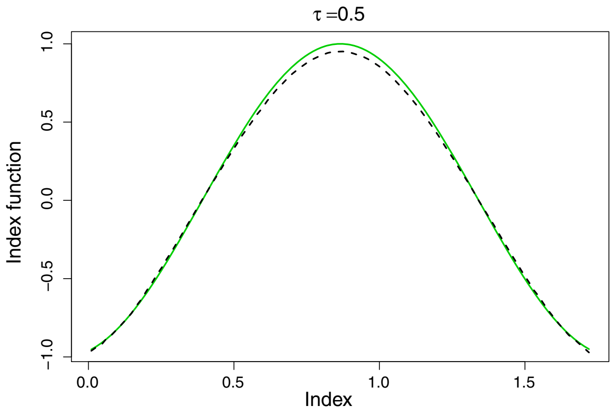

Fig. 2 The QR (0.5) estimation of the index function in Example 5.1 with Cauchy distributed error

Example 4.3 We generate 200 random samples from model (17), with $\theta = \frac{(3,1.5,0,0,2,0)^T}{\sqrt{15.25}}$ and $\beta = (3, 1.5, 0, 0, 0, 0, 0, 2)^T, \sigma = 0.1$. $\mathbf{X} = (X_1, \ldots, X_6) \sim U[0, 1]^6$, with $\text{cor}(X_i, X_j) = 0.5^{|i-j|}, \mathbf{Z} = (Z_1, \ldots, Z_8), Z_i \sim N(0, 1), j = 1, \ldots, 7$ with $\text{cor}(Z_i, Z_j) = 0.5^{|i-j|}$, Z_8 follows $b(1, 0.4)$. ε follows four types of error distribution described above. We assume $\mathbf{X}$ and $(Z_1 \ldots, Z_7)$, Z_8 and ε are independent. We conducted 250 times simulation and variable selection by adaptive lasso method for each $\tau = 0.25, 0.5$ and 0.75.

The results over 250 simulations are summarized in Table 6, where C(θ) denotes the average number of the true zero coefficients of θ that are correctly set to zero and IC(θ) be the average number of the true nonzero coefficients that are incorrectly set to zero, C(β) and IC(β) denote the corresponding average numbers for linear parameter β. In the column labeled "U-fit", we show the proportion of trials excluding any nonzero coefficient in 250 replications. We represent the probability of trials selecting the exact subset model and the probability of trials including all six significant variables and some noise variables in the columns "C-fit" ("correct-fit") and "O-fit" ("over-fit"), respectively. From Table 6, we can see that variable selection procedure can efficiently choose the true subset, which clearly shows the virtue of the adaptive lasso penalized MACLE proposed in Sect. 3.

Table 5 Summary of 200 times estimation for Example 5.2

τ	Bias and Std	θ_1	θ_2	θ_3	θ_4	θ_5	β_1	β_2	β_3
$n = 200$									
0.5	Bias	−0.0002	−0.0030	0.0079	−0.0009	0.0085	0.0022	−0.0018	−0.0005
	Std	0.0491	0.0323	0.0539	0.0575	0.0477	0.0339	0.0340	0.0357
0.8	Bias	0.0030	−0.0060	0.0049	−0.0035	0.0039	−0.0002	0.0009	0.0013
	Std	0.0494	0.0324	0.0514	0.0487	0.0450	0.0264	0.0265	0.0266
0.9	Bias	0.0036	−0.0117	0.0083	−0.0064	0.0050	0.0036	−0.0006	−0.0001
	Std	0.0616	0.0920	0.0595	0.0534	0.0535	0.0330	0.0263	0.0307
$n = 400$									
0.5	Bias	−0.0017	−0.0017	0.0050	0.0004	0.0028	−0.0009	−0.0003	0.0031
	Std	0.0343	0.0217	0.0445	0.0440	0.0308	0.0239	0.0249	0.0239
0.8	Bias	−0.0017	0.0005	0.0019	−0.0052	0.0029	0.0011	−0.0014	−0.0030
	Std	0.0298	0.0201	0.0321	0.0343	0.0307	0.0192	0.0172	0.0170
0.9	Bias	0.0010	−0.0021	0.0035	−0.0064	0.0041	0.0015	0.0000	−0.0013
	Std	0.0366	0.0239	0.0472	0.0452	0.0383	0.0206	0.0206	0.0224

Table 6 Variable selection results by adaptive lasso penalized MACLE of PLSIM

Methods	MRME(θ)	C(θ)	IC(θ)	MRME(β)	C(β)	IC(β)	U-fit	O-fit	C-fit
Standard normal									
QR (0.5)	0.328	4.940	0	1.738	4.990	0	0	0.060	0.940
QR (0.25)	0.454	4.940	0	1.679	4.990	0	0	0.045	0.955
QR (0.75)	0.526	4.915	0	1.739	4.985	0	0	0.060	0.940
t-distribution with $df = 3$									
QR (0.5)	0.364	4.965	0	1.838	4.990	0	0	0.045	0.955
QR (0.25)	0.378	4.940	0	1.859	4.995	0	0	0.045	0.955
QR (0.75)	0.421	4.915	0.005	1.816	4.985	0	0.005	0.090	0.905
Standard Cauchy									
QR (0.5)	0.328	4.920	0.005	1.364	4.990	0	0.005	0.090	0.905
QR (0.25)	0.371	4.680	0.055	1.819	4.950	0	0.055	0.255	0.690
QR (0.75)	0.387	4.765	0.015	1.691	4.970	0	0.015	0.205	0.780
$0.9N(0.1) + 0.1N(0, 10^2)$									
QR (0.5)	0.319	4.980	0	1.736	5	0	0	0.020	0.980
QR (0.25)	0.441	4.965	0	1.634	4.995	0	0	0.030	0.970
QR (0.75)	0.487	4.940	0.020	1.602	4.990	0	0.020	0.060	0.920

5 Real data analysis

In this section, we apply our estimation method to Boston housing data, which is available online at http://lib.stat.cmu.edu/datasets/bostoncorrected.txt, with some corrections and augmentation by the latitude and longitude of each observation, called the

Table 7 Parameter's estimation for Boston Housing data by penalized MACLE

Predictors τ	0.1	0.3	0.5	0.7	0.9
RM	0.3668	0.4287	0.3393	0.2442	0
log(TAX)	−0.3283	−0.2335	−0.1812	−0.1706	0
PTRATIO	−0.1776	−0.1255	−0.1832	−0.1959	0.3590
log(LSTAT)	−0.8521	−0.8637	−0.9047	−0.9343	0.9333
CRIM	−1.4152	−0.9831	−0.8312	0	0
NOX	−0.5768	0	0	0	0
DIS	−0.9576	−0.7251	−0.8414	−0.9497	−1.0471
LON	−0.6786	−0.8907	−0.8429	−0.8911	0
LAT	0	0	0	0	0
B	0.6118	0.7789	0.8497	1.0188	0
ZN	0	0	0	0	0
AGE	0	0	0	0	0
INDUS	0	0	0	0	0
CHAS	0	0	0.9193	0.7047	0

Corrected Boston House Price Data. There are 506 observations, 15 non-constant predictor variables and one response variable, corrected median value of owner-occupied homes (CMEDV). Predictors include longitude (LON), latitude (LAT), crime rate (CRIM), proportion of area zoned with large lots (ZN), proportion of non-retail business acres per town (INDUS), Charles River as a dummy variable (= 1 if tract bounds river; 0 otherwise) (CHAS), nitric oxides concentration (NOX), average number of rooms per dwelling (RM), proportion of owner-occupied units built prior to 1940 (AGE), weighted distances to five Boston employment centers (DIS), index of accessibility to radial highways (RAD), property tax rate (TAX), pupil-teacher ratio by town (PTRATIO), black population proportion town (B), and lower status population proportion (LSTAT). Following previous studies, we take logarithmic transformation on TAX and LSTAT. For simplicity, we exclude the categorical variable RAD, standardize the other covariates aside from CHAS, and recode the value of CHAS to -1(when CHAS $= 0$) and 1(when CHAS $= 1$).

Following previous studies, we construct index based on the following four predictor: RM, log(TAX), PTRATIO and log(LSTAT) and compose the partial linear part by the other ten predictors. The covariates of PLSIM may change for different τ, we rely on the variable selection method in Sect. 3 to choose the preferred model. Note that the response variable is censored from above, and therefore quantile regression is more trustful than mean regression. Furthermore, we analyze the normality of the residuals obtained by modeling the mean of CMDEV by PLSIM with index covariates and the linear parts as described above. By Shapiro–Wilk test (Shapiro and Wilk 1965), we find that the p value is less than 2.2×10^{-16}. This reminds us further that the error cannot be normal, and the mean regression based on least square is unsuitable here.

The estimation of the index and linear parameters by adaptive lasso penalized MACLE is presented in Table 7, we can see that the predictors' influence is different for

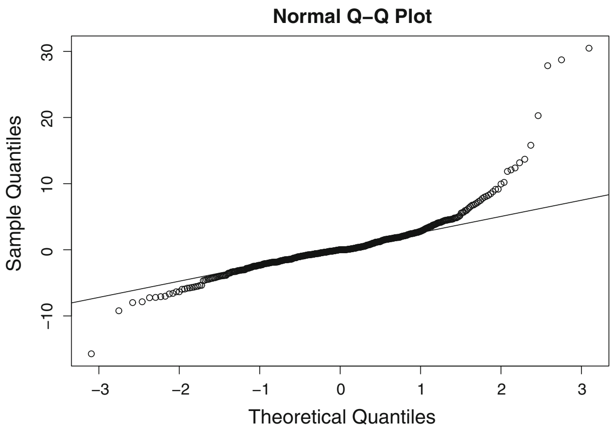

Fig. 3 QQ plot of the residual for median regression

different condition quantile of CMDEC. The norm quantile–quantile of the residual, when $\tau = 0.5$, is presented in Fig. 3, from which we can see apparently that the residuals cannot follow normal distribution. This also shows in some sense that, the MACLE of PLSIM is robust and applicable in various real data analysis in social research or scientific field with non-specific errors. In this real data analysis, we only concentrate on the variable selection in quantile regression of PLSIM. The model selection between quantile regression of single-index model and partial linear single-index model based on model complexity is still an open question to be solved in future research.

Appendix A

In order to derive the asymptotic properties, we need the following regularity conditions:

- **A.1** The kernel $K(\cdot) \geq 0$ has a compact support and its first derivative is bounded, satisfies $\int_{-\infty}^{\infty} K(z)\mathrm{d}z = 1, \int_{-\infty}^{\infty} zK(z)\mathrm{d}z = 0, \int_{-\infty}^{\infty} z^2K(z)\mathrm{d}z < \infty$ and $\int_{-\infty}^{\infty} z^j K^2(z)\mathrm{d}z < \infty, \ j = 0, 1, 2$;
- **A.2** The covariates $\mathbf{X}, \ \mathbf{Z}$ are bounded. The marginal density function of $\mathbf{X}^T\theta$, denoted by $f_{\mathcal{U}}(\cdot)$, is uniformly continuous for θ in a neighborhood of θ_τ and bounded away from 0 and ∞ on its compact support $\mathcal{U}_\tau$;
- **A.3** The condition density of Y given $\mathrm{X}^T\theta = u$, denoted by $f(\cdot|u)$ is positive and continuous in u;
- **A.4** The function $g(\cdot)$ has a continuous and bounded second derivative;
- **A.5** The conditional expectations $\mathrm{E}\left[\mathbf{Z}|\mathbf{X} = \mathbf{x}\right], \mathrm{E}\left[\mathbf{Z}\mathbf{Z}^T|\mathbf{X} = \mathbf{x}\right]$ have bounded derivatives. The conditional expectations $\mathrm{E}(\mathbf{X}|\mathbf{X}^T\theta = u), \mathrm{E}(\mathbf{Z}|\mathbf{X}^T\theta = u)$, $\mathrm{E}(\mathbf{X}^{\otimes 2} f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}^T\theta)))|\mathbf{X}^T\theta = u), \mathrm{E}(\mathbf{Z}^{\otimes 2}|\mathbf{X}^T\theta = u)$ and $\mathrm{E}(\mathbf{X}\mathbf{Z}^T|\mathbf{X}^T\theta = u)$ are twice differentiable in $u \in \mathcal{U}_\tau$, where $A^{\otimes 2} = AA^T$ for any matrix or vector A;

A.6 $\mathrm{E}\left[g'(\mathbf{X}^T\theta_\tau)^2[\mathbf{X}-\mathrm{E}(\mathbf{X})]^{\otimes 2}\right]$, $\mathrm{E}\left([\mathbf{Z}-E(\mathbf{X}|\mathbf{X})]^{\otimes 2}\right)$, $\mathrm{E}([g'(\mathbf{X}^T\theta)]^2\tilde{\mathbf{X}}^{\otimes 2})$ and $\mathrm{E}(\tilde{\mathbf{Z}}^{\otimes 2})$ are positive-definite matrix's, where $g'(\cdot)$ is the first derivative of $g(\cdot)$;
A.7 Bandwidth h satisfies that, $h \sim n^{-\delta}$ and $1/6 < \delta < 1/4$;

Remark 4 These conditions above are common in the semi-parametric literature and are satisfied in many applications. Condition A.1 simply requires that the kernel function is a proper density with finite second moment, which is required to derive the asymptotic variance of estimators; Condition A.2 guarantees the existence of any ratio terms with the density appearing as part of the denominator; Condition A.4 is a common assumption for an unknown differentiable function. Condition A.5 is common in semi-parametric regression, which is needed to derive the asymptotic distribution of $(\hat{\theta},\ \hat{\beta})$. Condition A.6 is the same as the condition (vi) in Liang et al. (2010).

Lemma 2 *Suppose $A_n(s)$ is convex and can be represented as $\frac{1}{2}s^TVs+U_n^Ts+C_n+r_n(s)$, where V is symmetric and positive definite, U_n is stochastically bounded, C_n is arbitrary, and $r_n(s)$ goes to zero in probability for each s. Then α_n, the argmin of A_n, is only $o_p(1)$ away from $\beta_n=-V^{-1}U_n$, the argmin of $\frac{1}{2}s^TVs+U_n^Ts+C_n$. If also $U_n \xrightarrow{L} U$, then $\alpha_n \xrightarrow{L} -V^{-1}U$.*

Proof This lemma comes from the basic corollary in Hjort and Pollard (1993). □

Lemma 3 *Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be independent and identically distributed random vectors, where Y_i is scalar random variable, $\mathbf{X}_i$ is p-dimensional random vector. Assume further that $E|y|^s < \infty$ and $\sup_{\mathbf{x}} \int |y|^s f(\mathbf{x}, y)dy < \infty$, $s \geq 3$, where $f(\cdot,\cdot)$ denotes the joint density of $(\mathbf{X}, Y)$. Let $K(\cdot)$ be a bounded positive function with a bounded support and satisfying a Lipschitz condition. Then*

$$\sup_{\mathbf{x}\in\Xi}\left|\frac{1}{n}\sum_{i=1}^{n}[K_h(\mathbf{X}_i-\mathbf{x})y_i - E(K_h(\mathbf{X}_i-\mathbf{x})y_i)]\right| = O_p\left[\left(\frac{\ln(1/h^p)}{nh^p}\right)^{1/2}\right],$$

provided that $n^{2\varepsilon-1}h^p \to \infty$ for some $\varepsilon < 1-s^{-1}$, where Ξ is the compact support of $\mathbf{X}$.

Proof This follows from the result by Mack and Silverman (1982). □

For simplicity, we shorthand write $\delta_n = [\ln(1/h)/nh]^{1/2}$, $\delta_\theta = |\tilde{\theta}-\theta_\tau|$, $\delta_\beta = |\tilde{\beta}-\beta_\tau|$, $\delta_\gamma = \delta_\theta+\delta_\beta$, $\mu_\theta(\mathbf{x}) = \mathrm{E}(X|\mathbf{X}^T\theta = \mathbf{x}^T\theta)$, $\nu_\theta(\mathbf{x}) = \mathrm{E}(Z|\mathbf{X}^T\theta = \mathbf{x}^T\theta)$ and $K_{ih}^\theta = K_{i,h}^\theta(\mathbf{x}) = K_h(\mathbf{X}_{i0}^T\theta)$, where $\mathbf{X}_{i0} = \mathbf{X}_i - \mathbf{x}$. For ease of expression, write $\varepsilon_i = Y_i - q_\tau(\mathbf{X}_i, \mathbf{Z}_i)$, whose τth conditional quantile is zero when given $(\mathbf{X}_i, \mathbf{Z}_i)$.

Lemma 4 *Let $\theta_d = \tilde{\theta}-\theta_\tau$, $\beta_d = \tilde{\beta}-\beta_\tau$. Suppose the regular conditions* A.1–A.7 *hold, we have*

$$\hat{g}(\mathbf{x}^T\tilde{\theta}) = g(\mathbf{x}^T\tilde{\theta}) + \frac{1}{2}g''(\mathbf{x}^T\tilde{\theta})\mu_2 h^2 - g'(\mathbf{x}^T\tilde{\theta})\mu_{\theta_\tau}(\mathbf{x})^T\theta_d - \nu_{\theta_\tau}(\mathbf{x})^T\beta_d + R_{n1}^{\tilde{\theta}}(\mathbf{x})$$
$$+ O\left(h^2(h^2+\delta_\theta+\delta_n)+\delta_\beta\delta_n+\delta_\theta\delta_\gamma\right)$$
$$\hat{g}'(\mathbf{x}^T\tilde{\theta}) = g'(\mathbf{x}^T\tilde{\theta}) + \frac{1}{h}R_{n2}^{\tilde{\theta}}(\mathbf{x}) + O(h^2+\delta_\gamma),$$

where $R_{n1}^{\theta}(\mathbf{x}) = \left[nf_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{x}^T\theta)f_{\mathcal{U}_\tau}(\mathbf{x}^T\theta)\right]^{-1}\sum_{i=1}^n K_{i,h}^{\theta}\psi_\tau(\varepsilon_i)$, $\psi_\tau(u) = u(\tau - I(u<0))$ *and* $R_{n2}^{\theta}(\mathbf{x}) = \left[nh\mu_2 f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{x}^T\theta)f_{\mathcal{U}_\tau}(u)\right]^{-1}\sum_{i=1}^n K_{i,h}^{\theta}\psi_\tau(\varepsilon_i)$ $\mathbf{X}_{i0}^T\theta$.

Proof of Lemma 4 Let $\mathbf{x}^T\theta = u$ for easy of expression, and write the estimation of $g(u)$ by $\hat{g}(u;h,\tilde{\theta},\tilde{\beta})$ to indicate the dependence on h and $(\tilde{\theta},\tilde{\beta})$. For given u, we write $\hat{a}_{(\tilde{\theta},\tilde{\beta})} := \hat{g}(u;h,\tilde{\theta},\tilde{\beta})$ and $\hat{b}_{(\tilde{\theta},\tilde{\beta})} := \hat{g}'(u;h,\tilde{\theta},\tilde{\beta})$, which are the solutions of the following minimization problem,

$$\min_{a,b}\sum_{i=1}^n \rho_\tau(y_i - a - b\mathbf{X}_{i0}^T\tilde{\theta} - \mathbf{Z}_i^T\tilde{\beta})K(\mathbf{X}_{i0}^T\tilde{\theta}/h).$$

Let

$$\hat{\eta}_n = (nh)^{1/2}\begin{pmatrix}\hat{a}_{(\tilde{\theta},\tilde{\beta})} - g(u)\\ h[\hat{b}_{(\tilde{\theta},\tilde{\beta})} - g'(u)]\end{pmatrix},\quad M_i = \begin{pmatrix}1\\ \mathbf{X}_{i0}^T\tilde{\theta}/h\end{pmatrix},$$
$$r_i(u) = -g(\mathbf{X}_i^T\tilde{\theta}) + g(u) + g'(u)\mathbf{X}_{i0}^T\tilde{\theta} + \mathbf{Z}_i^T(\tilde{\beta}-\beta_\tau),\ K_i = K(\mathbf{X}_{i0}^T\tilde{\theta}/h).$$

Thus, $\hat{\eta}_n$ minimizes

$$Q_n(\eta) = \sum_{i=1}^n\left[\rho_\tau(\varepsilon_i - r_i(u) - \eta^T M_i/\sqrt{nh}) - \rho_\tau(\varepsilon_i - r_i(u))\right]K_i.$$

Following the identity by Knight (1998),

$$\rho_\tau(u-v) - \rho(u) = -v\psi(u) + \int_0^v (I(u\le s) - I(u\le 0))\mathrm{d}s,$$

we can write

$$Q_n(\eta) = -\eta^T W_n + B_n(\eta), \tag{18}$$

where

$$W_n = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K_i M_i [\psi_\tau(\varepsilon_i) + r_i(u)],$$

$$B_n(\eta) = \sum_{i=1}^{n} K_i \int_{r_i(u)}^{r_i(u)+M_i^T \eta/\sqrt{nh}} (I(\varepsilon_i \le s) - I(\varepsilon_i) \le 0))\mathrm{d}s.$$

Let $\tilde{\mathcal{X}}$ be the σ-field generated by $\{\mathbf{X}_1^T\tilde{\theta}, \mathbf{X}_2^T\tilde{\theta}, \cdots, \mathbf{X}_n^T\tilde{\theta}\}$. Consider the conditional expectation of $B_n(\eta)$, we have

$$\begin{aligned}
\mathrm{E}\left(B_n(\eta)|\tilde{\mathcal{X}}\right) &= \sum_{i=1}^{n} K_i \int_{r_i(u)}^{r_i(u)+M_i^T \eta/\sqrt{nh}} \mathrm{E}\left(I(\varepsilon_i \le s) - I(\varepsilon_i \le 0)|u_i\right) \mathrm{d}s \\
&= \frac{1}{2} f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|u)\eta^T \left(\frac{1}{nh}\sum_{i=1}^{n} M_i {M_i}^T K_i\right)\eta \\
&\quad + \left(\frac{f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|u)}{\sqrt{nh}} \sum_{i=1}^{n} K_i r_i(u) M_i\right)^T \eta + o_p(1).
\end{aligned}$$

It can be shown that $\mathrm{Var}(B_n(\eta)|\tilde{\mathcal{X}}) = o(1)$, therefore, we have

$$\begin{aligned}
B_n(\eta) &= \frac{1}{2}\eta^T \left(\frac{f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|u)}{nh}\sum_{i=1}^{n} M_i {M_i}^T K_i\right)\eta \\
&\quad + \left(\frac{f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|u)}{\sqrt{nh}} \sum_{i=1}^{n} K_i r_i(u) M_i\right)^T \eta + o_p(1) \\
&\equiv B_{n1}(\eta) + B_{n2}(\eta) + o_p(1),
\end{aligned} \tag{19}$$

where $B_{n1}(\eta) = \frac{1}{2}\eta^T \mathbb{S}_n \eta$, $B_{n2}(\eta) = \left(\frac{f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)}{\sqrt{nh}} \sum_{i=1}^{n} K_i r_i(u) M_i\right)^T \eta$ and $\mathbb{S}_n = \frac{f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)}{nh} \sum_{i=1}^{n} M_i {M_i}^T K_i$. By using Lemma 3, we can easily get that

$$\mathbb{S}_n = \mathbb{S} + O_p(h^2 + \delta_n + \delta_\theta), \tag{20}$$

where $\mathbb{S} = f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|u) f_{\mathcal{U}_\tau}(u) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$.

Now, consider the $B_{n2}(\eta)$. Since

$$r_i(u) = \left(g'(\mathbf{X}_i^T\tilde{\theta})\mathbf{X}_i^T, \mathbf{Z}_i^T\right)\begin{pmatrix}\theta_d \\ \beta_d\end{pmatrix} - \frac{1}{2}g''(u)\left(\mathbf{X}_i^T\tilde{\theta} - u\right)^2 + O\left[\left(\mathbf{X}_i^T\tilde{\theta} - u\right)^3 + |\theta_d|^2\right],$$

we have

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)K_i r_i(u) = \sqrt{nh}\, f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)\, f_{\mathcal{U}_\tau}(u)$$
$$\times\left(-\frac{1}{2}g''(u)\mu_2 h^2 + g'(u)\mu_{\theta_\tau}(\mathbf{x})^T\theta_d + \nu_{\theta_\tau}(\mathbf{x})^T\beta_d + O(h^4+\delta_\theta^2+h^2\delta_\theta+\delta_\theta\delta_\beta)\right),$$
$$\text{and}\quad \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)K_i r_i(u)\frac{\mathbf{X}_{i0}^T\tilde{\theta}}{h} = \sqrt{nh}[O(h^3+h\delta_\gamma)]. \tag{21}$$

Combining (18), (19), (20) and (21), we have

$$Q_n(\eta) = \frac{1}{2}\eta^T\mathbb{S}\eta - W_n^T\eta + \sqrt{nh}\, f_Y(q_\tau(\mathbf{X},\mathbf{Z})|u)\, f_{\mathcal{U}_\tau}(u)$$
$$\times\begin{pmatrix} -\frac{1}{2}g''(u)\mu_2h^2 + g'(u)\mu_{\theta_\tau}(\mathbf{x})^T\theta_d + \nu_{\theta_\tau}(\mathbf{x})^T\beta_d + O(h^4+\delta_\theta^2+h^2\delta_\theta+\delta_\theta\delta_\beta) \\ O(h^3+h\delta_\gamma)\end{pmatrix}^T \eta + o_p(1).$$

By using Lemma 2, the argmin of $Q_n(\eta)$ can be expressed as

$$\hat{\eta}_n = \mathbb{S}^{-1}W_n - \sqrt{nh}$$
$$\times\begin{pmatrix} -\frac{1}{2}g''(u)\mu_2h^2 + g'(u)\mu_{\theta_\tau}(\mathbf{x})^T\theta_d + \nu_{\theta_\tau}(\mathbf{x})^T\beta_d + O(h^4+\delta_\theta^2+h^2\delta_\theta+\delta_\theta\delta_\beta) \\ O(h^3+h\delta_\gamma)\end{pmatrix} + o_p(1).$$

Note the definition of $\hat{\eta}_n$ and W_n, we complete the proof. □

Proof of Theorem 1 Given $\hat{g}\left(\mathbf{X}_j^T\tilde{\theta}\right)$, $\hat{g}'\left(\mathbf{X}_j^T\tilde{\theta}\right)$, the estimation of $g\left(\mathbf{X}_j^T\tilde{\theta}\right)$ and $g'\left(\mathbf{X}_j^T\tilde{\theta}\right)$ for $j=1,\ldots,n$, we obtain $\hat{\theta}$, $\hat{\beta}$ by

$$(\hat{\theta},\hat{\beta}) = \underset{||\theta||=1,\ \beta}{\operatorname{argmin}} \sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - \hat{g}\left(\mathbf{X}_j^T\tilde{\theta}\right) - \hat{g}'\left(\mathbf{X}_j^T\tilde{\theta}\right)\mathbf{X}_{ij}^T\theta - \mathbf{Z}_i^T\beta\right)\omega_{ij}.$$

Let

$$\hat{\gamma}^* = \sqrt{n}\begin{pmatrix}\hat{\theta}-\theta_\tau \\ \hat{\beta}-\beta_\tau\end{pmatrix},\quad M_{ij} = \begin{pmatrix}\hat{g}'(\mathbf{X}_j^T\tilde{\theta})\mathbf{X}_{ij} \\ \mathbf{Z}_i\end{pmatrix},$$
$$r_{ij} = -g\left(\mathbf{X}_i^T\theta_\tau\right) + \hat{g}\left(\mathbf{X}_j^T\tilde{\theta}\right) + \hat{g}'\left(\mathbf{X}_j^T\tilde{\theta}\right)\mathbf{X}_{ij}\theta_\tau,$$

then $\hat{\gamma}^*$ minimizes

$$\mathcal{Q}_n(\gamma^*) = \sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}\rho_\tau\left[\left(\varepsilon_i - r_{ij} - M_{ij}^T\gamma^*/\sqrt{n}\right) - \rho_\tau(\varepsilon_i - r_{ij})\right].$$

Using the Knight's identity, we can write

$$\begin{aligned}
\mathcal{Q}_n(\gamma^*) &= -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}\psi_\tau(\varepsilon_i)M_{ij}^T\gamma^* \\
&\quad +\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}\int_{r_{ij}}^{r_{ij}+M_{ij}^T\gamma^*/\sqrt{n}}[I(\varepsilon_i\le s)-I(\varepsilon_i\le 0)]\mathrm{d}s, \\
&\equiv \mathcal{Q}_{1n}(\gamma^*)+\mathcal{Q}_{2n}(\gamma^*),
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{Q}_{1n}(\gamma^*) &= -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}\psi_\tau(\varepsilon_i)M_{ij}^T\gamma^*, \\
\mathcal{Q}_{2n}(\gamma^*) &= \sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}\int_{r_{ij}}^{r_{ij}+M_{ij}^T\gamma^*/\sqrt{n}}(I(\varepsilon_i\le s)-I(\varepsilon_i\le 0))\mathrm{d}s.
\end{aligned}$$

Take $\mathcal{Q}_{2n}(\gamma^*)$ first. We shorthand write $\tilde{U}_i=\mathbf{X}_i^T\tilde{\theta},\ \tilde{U}_j=\mathbf{X}_j^T\tilde{\theta}$. Let us calculate the conditional expectation of $\mathcal{Q}_{2n}(\gamma^*)$:

$$\begin{aligned}
\mathrm{E}\left(\mathcal{Q}_{2n}(\gamma^*)|\tilde{\mathcal{X}}\right) &= \sum_{j=1}^{n}\sum_{i=1}^{n}\int_{r_{ij}}^{r_{ij}+M_{ij}^T\gamma^*/\sqrt{n}}\omega_{ij}\left[sf_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)(1+o(1))\right]\mathrm{d}s \\
&= \frac{1}{2}\gamma^{*T}\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}M_{ij}^T\omega_{ij}\right)\gamma^* \\
&\quad +\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)r_{ij}M_{ij}\right)^T\gamma^*+o_p(1).
\end{aligned}$$

Define $\mathcal{R}_n(\gamma^*)=\mathcal{Q}_{2n}(\gamma^*)-\mathrm{E}(\mathcal{Q}_{2n}(\gamma^*)|\tilde{\mathcal{X}})$. It can be shown that $\mathcal{R}_n(\gamma^*)=o_p(1)$. Therefore, we have

$$\mathcal{Q}_{2n}(\gamma^*)\equiv\mathcal{Q}_{2n1}(\gamma^*)+\mathcal{Q}_{2n2}(\gamma^*)+o_p(1),$$

where

$$\begin{aligned}
\mathcal{Q}_{2n1}(\gamma^*) &= \frac{1}{2}\gamma^{*T}\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}M_{ij}^T\right)\gamma^*, \\
\mathcal{Q}_{2n2}(\gamma^*) &= \left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}r_{ij}\right)^T\gamma^*.
\end{aligned}$$

Consider $\mathcal{Q}_{2n1}(\gamma^*)$. Denote $\mathcal{G}_n^{\tilde{\theta}} = \frac{1}{n}\sum_{j=1}^n \sum_{i=1}^n f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}M_{ij}^T\omega_{ij}$. By using Lemma 3, we can easily get that $\mathcal{G}_n^{\tilde{\theta}} = \mathcal{G} + O(h^2 + \delta_n + \delta_\gamma)$, where

$$\mathcal{G} = \begin{pmatrix} 2\tilde{W}_0 & \tilde{C}_{12} \\ \tilde{C}_{12}^T & \mathrm{E}(f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}^T\theta_\tau)\mathbf{Z}\mathbf{Z}^T) \end{pmatrix},$$
$$\tilde{W}_0 = \mathrm{E}\left(f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}^T\theta_\tau)\{g'(\mathbf{X}^T\theta_\tau)\}^2(X-\mu_{\theta_\tau}(\mathbf{X}))(X-\mu_{\theta_\tau}(\mathbf{X}))^T\right),$$
$$\tilde{C}_{12} = \mathrm{E}\left(f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}^T\theta_\tau)g'(\mathbf{X}^T\theta_\tau)(X-\mu_{\theta_\tau}(X))\mathbf{Z}^T\right).$$

Thus, we have

$$\mathcal{Q}_{2n1}(\gamma^*) = \frac{1}{2}\gamma^{*T}\mathcal{G}\gamma^* + o_p(1). \tag{22}$$

Now consider $\mathcal{Q}_{2n2}(\gamma^*)$. Note that

$$\begin{aligned} r_{ij} &= g\left(\mathbf{X}_i^T\tilde{\theta}\right) - g\left(\mathbf{X}_i^T\theta_\tau\right) - \left(g\left(\mathbf{X}_i^T\tilde{\theta}\right) - g\left(\mathbf{X}_j^T\tilde{\theta}\right) - g'\left(\mathbf{X}_j^T\tilde{\theta}\right)\mathbf{X}_{ij}^T\tilde{\theta}\right) \\ &\quad + \left[\hat{g}\left(\mathbf{X}_j^T\tilde{\theta}\right) - g\left(\mathbf{X}_j^T\tilde{\theta}\right)\right] + \left[\hat{g}'\left(\mathbf{X}_j^T\tilde{\theta}\right) - g'\left(\mathbf{X}_j^T\tilde{\theta}\right)\mathbf{X}_{ij}^T\right]\tilde{\theta} - \hat{g}'\left(\mathbf{X}_j^T\tilde{\theta}\right)\mathbf{X}_{ij}^T\left(\tilde{\theta} - \theta_\tau\right) \\ &= (1, \mathbf{X}_{ij}^T\tilde{\theta}/h)\begin{pmatrix} \hat{g}(\mathbf{X}_j^T\tilde{\theta}) - g(\mathbf{X}_j^T\tilde{\theta}) \\ h(\hat{g}'(\mathbf{X}_j^T\tilde{\theta}) - g'(\mathbf{X}_j^T\tilde{\theta})) \end{pmatrix} \\ &\quad + g'\left(\mathbf{X}_i^T\tilde{\theta}\right)\mathbf{X}_i^T\theta_d - \hat{g}'\left(\mathbf{X}_j^T\tilde{\theta}\right)\mathbf{X}_{ij}^T\theta_d - \frac{1}{2}g''\left(\mathbf{X}_j^T\tilde{\theta}\right)\left(\mathbf{X}_{ij}^T\tilde{\theta}\right)^2 + O\left[\delta_\theta^2 + \left(\mathbf{X}_{ij}^T\tilde{\theta}\right)^3\right], \end{aligned}$$

$\mathcal{Q}_{2n2}(\gamma^*)$ as

$$\mathcal{Q}_{2n2}(\gamma^*) \equiv (\mathcal{Q}_{2n21} + \mathcal{Q}_{2n22})^T\gamma^* + O(\delta_\theta^2 + h^3),$$

where

$$\begin{aligned} \mathcal{Q}_{2n21} &= \frac{1}{\sqrt{n}}\sum_{j=1}^n\sum_{i=1}^n f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)\omega_{ij}M_{ij}\left(1, \mathbf{X}_{ij}^T\tilde{\theta}/h\right)\begin{pmatrix} \hat{g}(\mathbf{X}_j^T\tilde{\theta}) - g(\mathbf{X}_j^T\tilde{\theta}) \\ h(\hat{g}'(\mathbf{X}_j^T\tilde{\theta}) - g'(\mathbf{X}_j^T\tilde{\theta})) \end{pmatrix}, \\ \mathcal{Q}_{2n22} &= \frac{1}{\sqrt{n}}\sum_{j=1}^n\sum_{i=1}^n f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)\omega_{ij}M_{ij} \\ &\quad \times\left(g'(\mathbf{X}_i^T\tilde{\theta})\mathbf{X}_i^T\theta_d - \hat{g}'(\mathbf{X}_j^T\tilde{\theta})\mathbf{X}_{ij}^T\theta_d - \frac{1}{2}g''(\mathbf{X}_j^T\tilde{\theta})(\mathbf{X}_{ij}^T\tilde{\theta})^2\right). \end{aligned}$$

Consider $\mathcal{Q}_{2n21}$ and $\mathcal{Q}_{2nn2}$ separately. Note the expression of $\hat{g}(\mathbf{x}^T\tilde{\theta})$ and $\hat{g}'(\mathbf{X}^T\tilde{\theta})$ in Lemma 4, we have

$$
\begin{aligned}
\mathcal{Q}_{2n21} &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}(1,\mathbf{X}_{ij}^T\tilde{\theta})\begin{pmatrix} R_{n1}^{\tilde{\theta}}(\mathbf{X}_j)\\ R_{n2}^{\tilde{\theta}}(\mathbf{X}_j)\end{pmatrix}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}\\
&\quad\times\left(\frac{1}{2}g''(\mathbf{x}^T\tilde{\theta})\mu_2 h^2 - g'(\mathbf{x}^T\tilde{\theta})\mu_\theta(\mathbf{x})^T\theta_d - \nu_\theta(\mathbf{x})^T\beta_d\right)\\
&\quad+O_p(h^2(h^2+\delta_\theta+\delta_n)+\delta_\beta\delta_n+\delta_\theta\delta_\gamma)\\
&\equiv T_1+T_2+o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
T_1 &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}(1,\mathbf{X}_{ij}^T\tilde{\theta})\begin{pmatrix} R_{n1}^{\tilde{\theta}}(\mathbf{X}_j)\\ R_{n2}^{\tilde{\theta}}(\mathbf{X}_j)\end{pmatrix}\\
T_2 &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)M_{ij}\\
&\quad\times\left(\frac{1}{2}g''(\mathbf{x}^T\tilde{\theta})\mu_2 h^2 - g'(\mathbf{x}^T\tilde{\theta})\mu_\theta(\mathbf{x})^T\theta_d - \nu_\theta(\mathbf{x})^T\beta_d\right).
\end{aligned}
$$

By direct calculation, we have

$$
\begin{aligned}
T_1 &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\omega_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_i)}{nf_{\mathcal{U}}(\tilde{U}_j)f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_j)}M_{ij}(1,\ \mathbf{X}_{ij}^T\tilde{\theta}/h)\sum_{k=1}^{n}\begin{pmatrix}1\\ \mathbf{X}_{kj}^T\theta/h\end{pmatrix}K_h(\mathbf{X}_{kj}^T\tilde{\theta})\psi_\tau(\varepsilon_k)\\
&= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{n}\psi_\tau(\varepsilon_k)\omega_{kj}\begin{pmatrix}\hat{g}'(\tilde{U}_j)(\mu_{\theta_\tau}(\mathbf{X}_j)-\mathbf{X}_j)\\ \nu(\mathbf{X}_j)\end{pmatrix}+o_p(1).
\end{aligned}
$$

Combining T_1 with $\mathcal{Q}_{1n}(\gamma^*)$, we have

$$
\begin{aligned}
\mathcal{Q}_{1n}(\gamma^*)+T_1^T\gamma^* &= \left[-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_\tau(\varepsilon_i)\omega_{ij}\begin{pmatrix}\hat{g}'(\tilde{U}_j)[\mathbf{X}_i-\mu_{\theta_\tau}(\mathbf{X}_j)]\\ \mathbf{Z}_i-\nu_{\theta_\tau}(\mathbf{X}_j)\end{pmatrix}\right]^T\gamma^*+o_p(1)\\
&= -\sqrt{n}\mathcal{W}_n^T\gamma^*+o_p(1),
\end{aligned}
\tag{23}
$$

where $\mathcal{W}_n=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_\tau(\varepsilon_i)\omega_{ij}\begin{pmatrix}\hat{g}'(\tilde{U}_j)(\mathbf{X}_i-\mu_{\theta_\tau}(\mathbf{X}_j))\\ \mathbf{Z}_i-\nu_{\theta_\tau}(\mathbf{X}_j)\end{pmatrix}$.

Combining T_2 with $\mathcal{Q}_{2n22}$, we have

$$\begin{aligned}
&\mathcal{Q}_{2n22} + T_2 \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=1}^{n} f_Y\left(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_j\right) \omega_{ij} M_{ij} \left(g'\left(\mathbf{X}_j^T \tilde{\theta}\right)\left(\mu_{\theta_\tau}(\mathbf{X}_j) - \mathbf{X}_j^T\right)^T \theta_d - \nu_{\theta_\tau}(\mathbf{X}_j)^T \beta_d\right) \\
&\quad + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=1}^{n} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_j)\omega_{ij} \begin{pmatrix} \hat{g}'(\mathbf{X}_{ij}^T\tilde{\theta})\mathbf{X}_{ij} \\ \mathbf{Z}_i \end{pmatrix} \begin{pmatrix} g'(\mathbf{X}_j^T\tilde{\theta})(\mu_{\theta_\tau}(\mathbf{X}_j) - \mathbf{X}_j) \\ -\nu_{\theta_\tau}(\mathbf{X}_j) \end{pmatrix}^T \begin{pmatrix} \theta_d \\ \beta_d \end{pmatrix} \\
&\quad + o_p(1) \\
&= -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\tilde{U}_j) \begin{pmatrix} g'(\mathbf{X}_j^T\tilde{\theta})(\mathbf{X}_j - \mu_{\theta_\tau}(\mathbf{X}_j)) \\ \nu_{\theta_\tau}(\mathbf{X}_j) \end{pmatrix}^{\otimes 2} \begin{pmatrix} \theta_d \\ \beta_d \end{pmatrix} \\
&\quad + o_p(1).
\end{aligned}$$

It is easy to show that

$$\mathcal{Q}_{2n22} + T_2 = -\sqrt{n} C_0 \begin{pmatrix} \theta_d \\ \beta_d \end{pmatrix} + o_p(1), \tag{24}$$

where $C_0 = \begin{pmatrix} \tilde{W}_0 & 0 \\ 0 & \mathrm{E}\left[f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}^T\theta_\tau)\nu_{\theta_\tau}(X)\nu_{\theta_\tau}(X)^T\right] \end{pmatrix}$.

Hence, combining (22), (23) and (24), we have

$$\mathcal{Q}_n(\gamma^*) = \frac{1}{2}\gamma^{*T}\mathcal{G}\gamma^* - \left[\mathcal{W}_n + \sqrt{n}C_0 \begin{pmatrix} \theta_d \\ \beta_d \end{pmatrix}\right]^T \gamma^* + o_p(1).$$

Following Lemma 2, $\hat{\gamma}^*$, the minimizer of $\mathcal{Q}_n(\gamma^*)$ can be expressed as

$$\hat{\gamma}^* = \mathcal{G}^{-1}\mathcal{W}_n + \sqrt{n}\mathcal{G}^{-1}C_0 \begin{pmatrix} \theta_d \\ \beta_d \end{pmatrix} + o_p(1).$$

Note that $\hat{\gamma}^* = \sqrt{n}\begin{pmatrix} \hat{\theta} - \theta_\tau \\ \hat{\beta} - \beta_\tau \end{pmatrix}$, we have

$$\begin{pmatrix} \hat{\theta} - \theta_\tau \\ \hat{\beta} - \beta_\tau \end{pmatrix} = \mathcal{G}^{-1}\mathcal{W}_n + \mathcal{G}^{-1}C_0 \begin{pmatrix} \tilde{\theta} - \theta_\tau \\ \tilde{\beta} - \beta_\tau \end{pmatrix} + o_p(1/\sqrt{n}). \tag{25}$$

Note the expressions of $\mathcal{G}$, C_0 and by condition A.6, we can get that $\mathcal{G}$, C_0 and $\mathcal{G} - C_0$ are all positive symmetric matrices. Therefore, $\tilde{\mathcal{G}} = \mathcal{G}^{-1/2}C_0\mathcal{G}^{-1/2}$ is also a positive matrix with all eigenvalues less than 1. Let $(\tilde{\theta}_k, \tilde{\beta}_k)$ be the estimation results of the k-th iteration in the algorithm. For each k, Eq. (25) hold with $(\hat{\theta}, \hat{\beta})$ replaced by $(\tilde{\theta}_{k+1}, \tilde{\beta}_{k+1})$ and $(\tilde{\theta}, \tilde{\beta})$ by $(\tilde{\theta}_k, \tilde{\beta}_k)$. Let $\tilde{\gamma}^k = \tilde{\mathcal{G}}^{-1/2}\begin{pmatrix} \tilde{\theta}_k \\ \tilde{\beta}_k \end{pmatrix}$, we have

$$\tilde{\gamma}^{k+1} = \mathcal{G}^{-1/2}\mathcal{W}_n + \tilde{\mathcal{G}}\tilde{\gamma}^k + o_p(1/\sqrt{n}).$$

Since all the eigenvalues of $\tilde{\mathcal{G}}$ are less than 1, by similar analysis as Xia and Härdle (2006), the convergence of the algorithm can be easily obtained. Thus, for sufficient large k, we have

$$\mathcal{G}^{-1/2}\begin{pmatrix}\hat{\theta}-\theta_\tau\\ \hat{\beta}-\beta_\tau\end{pmatrix}=\mathcal{G}^{-1/2}\mathcal{W}_n+\tilde{\mathcal{G}}\mathcal{G}^{1/2}\begin{pmatrix}\hat{\theta}-\theta_\tau\\ \hat{\beta}-\beta_\tau\end{pmatrix}+o(1/\sqrt{n}).$$

It follows that

$$(\mathcal{G}-\mathcal{G}^{1/2}\tilde{\mathcal{G}}\mathcal{G}^{1/2})\begin{pmatrix}\hat{\theta}-\theta_\tau\\ \hat{\beta}-\beta_\tau\end{pmatrix}=\mathcal{W}_n+o(1/\sqrt{n}).$$

By the Cramér–Wald device and CLT, the proof of Theorem 1 is completed. □

Proof of Theorem 2 When the parameters θ_τ and β_τ are known, given u be an inner point of $\mathcal{U}_\tau$, similar as the proof in Lemma 4, we have

$$\hat{g}(u;h,\theta_\tau,\beta_\tau)=g(u)+\frac{1}{2}g''(u)\mu_2h^2+R_{n1}^{\theta_\tau}+O(h^3),$$

and

$$\sqrt{nh}\left[\hat{g}(u;h,\theta_\tau,\beta_\tau)-g(u)-\frac{1}{2}g''(u)\mu_2h^2\right]\xrightarrow{L}N(0,\Gamma(u)).$$

From Lemma 4, we have $\hat{g}(u;h,\tilde{\theta},\tilde{\beta})-\hat{g}(u;h,\theta_\tau,\beta_\tau)=-\mathrm{E}(X|\mathbf{X}^T\theta_\tau=u)^T\theta_d-\mathrm{E}(Z|\mathbf{X}^T\theta_\tau=u)^T\beta_d+R_{n1}^{\tilde{\theta}}-R_{n1}^{\theta_\tau}+O(\delta_\theta+h\delta_n+h^3)$. Since now we have $\theta_d=O_p(1/\sqrt{n})$ and $\beta_d=O_p(1/\sqrt{n})$, we need only to show

$$\sqrt{nh}\left(R_{n1}^{\hat{\theta}}-R_{n1}^{\theta_\tau}\right)=o_p(1). \tag{26}$$

By direct calculation, we can easily get $\mathrm{Var}\left(\sqrt{nh}(R_{n1}^{\hat{\theta}}-R_{n1}^{\theta_\tau})\right)=o(1)$ when $\hat{\theta}-\theta_\tau=O_p(1/\sqrt{n})$ and $nh^4\to\infty$. Thus, (26) is hold and we complete the proof of Theorem 2. □

Proof of Theorem 3 Denote the adaptive lasso MACLE estimator of (θ_τ,β_τ) by $\hat{\gamma}^{AQR}$, let $\gamma^\tau=\begin{pmatrix}\theta_\tau\\ \beta_\tau\end{pmatrix}$, $\sqrt{n}(\hat{\gamma}^{AQR}-\gamma^\tau)=\hat{\mathbf{u}}$, $\hat{\theta}_d=\hat{\theta}^{QR}-\theta_\tau$ and $\hat{\beta}_d=\hat{\beta}^{QR}-\beta_\tau$, then $\hat{\mathbf{u}}$ is the minimizer of the following criterion:

$$G_n(\mathbf{u})=\sum_{j=1}^{n}\sum_{i=1}^{n}\omega_{ij}\left(\rho_\tau(\varepsilon_i+r_{ij}+M_{ij}^T\mathbf{u}/\sqrt{n})-\rho_\tau(\varepsilon_i+r_{ij})\right)$$

$$+\sum_{j=1}^{p}\frac{\lambda_1}{\sqrt{n}|\hat{\theta}_j^{QR}|^2}\sqrt{n}\left[|\theta_{\tau j}+\frac{u_j}{\sqrt{n}}|-|\theta_{\tau j}|\right]$$
$$+\sum_{m=1}^{q}\frac{\lambda_2}{\sqrt{n}|\hat{\beta}_m^{QR}|^2}\sqrt{n}\left[|\beta_{\tau m}+\frac{u_{p+m}}{\sqrt{n}}|-|\beta_{\tau m}|\right]. \tag{27}$$

Following the proof of Theorem 3, we write $G_n(\mathbf{u})$ as follows:

$$\begin{aligned}G_n(\mathbf{u}) &= \frac{1}{2}\mathbf{u}^T\mathcal{G}\mathbf{u}-\mathcal{W}_n^T\mathbf{u}+\sqrt{n}\left(\hat{\theta}_d^T,\ \hat{\beta}_d^T\right)C_0^T\mathbf{u}+o_p(1)\\ &\quad+\sum_{j=1}^{p}\frac{\lambda_1}{\sqrt{n}|\hat{\theta}_j^{QR}|^2}\sqrt{n}\left[|\theta_{\tau j}+\frac{u_j}{\sqrt{n}}|-|\theta_{\tau j}|\right]\\ &\quad+\sum_{m=1}^{q}\frac{\lambda_2}{\sqrt{n}|\hat{\beta}_m^{QR}|^2}\sqrt{n}\left[|\beta_{\tau m}+\frac{u_{p+m}}{\sqrt{n}}|-|\beta_{\tau m}|\right].\end{aligned} \tag{28}$$

For $1\le j\le p_0, \theta_{\tau j}\ne 0$, then $|\hat{\theta}_j^{QR}|^2\rightarrow_p|\theta_{\tau j}|^2$, and $\sqrt{n}(|\theta_{\tau j}+u_j/\sqrt{n}|-|\beta_{\tau j}|)\rightarrow u_j sgn(\theta_{\tau j})$. By Slusky's theorem, $\frac{\lambda_1}{\sqrt{n}|\hat{\theta}_j^{QR}|^2}\sqrt{n}(|\theta_{\tau j}+u_j/\sqrt{n}|-|\theta_{\tau j}|)\rightarrow_p 0$. For $p_0<j\le p, \theta_{\tau j}=0$, then $\sqrt{n}(|\theta_{\tau j}+u_j/\sqrt{n}|-|\theta_{\tau j}|)\rightarrow_p\infty$. Similar results can be derived for β. Therefore, we have

$$\begin{aligned}&\frac{\lambda_1}{\sqrt{n}|\hat{\theta}_j^{QR}|^2}\sqrt{n}\left[|\theta_{\tau j}+\frac{u_j}{\sqrt{n}}|-|\theta_{\tau j}|\right]\rightarrow_p W(\theta_j,u_j)\\ &=\begin{cases}0, & \text{if } \theta_{\tau j}\ne 0,\\ 0, & \text{if } \theta_{\tau j}=0,\ \text{and } u_j=0,\\ \infty, & \text{if } \theta_{\tau j}=0,\ \text{and } u_j\ne 0.\end{cases}\end{aligned}$$

and

$$\begin{aligned}&\frac{\lambda_2}{\sqrt{n}|\hat{\beta}_m^{QR}|^2}\sqrt{n}\left[|\beta_{\tau m}+\frac{u_m}{\sqrt{n}}|-|\beta_{\tau m}|\right]\rightarrow_p W(\beta_m,u_{p+m})\\ &=\begin{cases}0, & \text{if } \beta_{\tau m}\ne 0,\\ 0, & \text{if } \beta_{\tau m}=0,\ \text{and } u_{p+m}=0,\\ \infty, & \text{if } \beta_{\tau m}=0,\ \text{and } u_{p+m}\ne 0.\end{cases}\end{aligned}$$

Corresponding to $\gamma^0=(\theta_{10}^T,\theta_{20}^T,\beta_{10}^T,\beta_{20}^T)^T$, write $\mathbf{u}=(\mathbf{u}_{10}^T,\mathbf{u}_{11}^T,\mathbf{u}_{20}^T,\mathbf{u}_{21}^T)^T$. Then it follows that

$$G_n(\mathbf{u})\rightarrow\frac{1}{2}\mathbf{u}^T\mathcal{G}\mathbf{u}-W_n^T\mathbf{u}+\left(\hat{\theta}_d^T,\hat{\beta}_d^T\right)C_0^T\mathbf{u}$$

$$+\sum_{j=1}^{p} W(\theta_j, u_j) + \sum_{m=1}^{q} W(\beta_m, u_{p+m}) + o_p(1)$$

$$\rightarrow L(\mathbf{u}) = \begin{cases} \frac{1}{2}\mathbf{u}^T \mathcal{G}\mathbf{u} - W_n^T \mathbf{u} + \left(\hat{\theta}_d^T, \hat{\beta}_d^T\right) C_0^T \mathbf{u} & \text{when } \mathbf{u}_{11} = 0, \mathbf{u}_{21} = 0; \\ \infty & \text{otherwise.} \end{cases}$$

Note that $G_n(\mathbf{u})$ is convex in $\mathbf{u}$, and $L(\mathbf{u})$ has an unique minimizer. By the epi-convergence results of Geyer (1994) and the same arguments in the proof of Theorem 1, the asymptotic normality can be easily established. Next, consider the consistency property of model selection. Firstly, for any $j \in \mathcal{A}_\theta$, $\hat{\theta}_j^\lambda \rightarrow_p \theta_{\tau j}$ by the asymptotically normality results, and then we have $j \in \hat{\mathcal{A}}_\theta$. Thus, $Pr(\mathcal{A}_\theta \subset \hat{\mathcal{A}}_\theta) \rightarrow_p 1$. Then it is suffice to show that $\forall j' \notin \mathcal{A}_\theta$, $Pr(j' \in \hat{\mathcal{A}}_\theta) \rightarrow 0$.

Note that the minimization on (θ, β) of (14) is equal to the minimization of (27) on $\mathbf{u}$. Suppose $j' \in \hat{\mathcal{A}}_\theta$. By the KKT optimality condition of the minimization of (28), we have

$$\left| \mathcal{G}^{(j')}\mathbf{u} - \mathcal{W}_n^{j'} + \sqrt{n} \begin{pmatrix} \hat{\theta}_d \\ \hat{\beta}_d \end{pmatrix}^T C_0^{(j')} \right| = \left| \frac{\lambda_n}{\sqrt{n}|\hat{\theta}_{j'}^{QR}|^2} \right|,$$

where $\mathcal{G}^{(j')}$ is the j'th row of $\mathcal{G}$, and $C_0^{(j')}$ is the j'th row of C_0, respectively,$\mathcal{W}_n^{j'}$ is the j'-th element of $\mathcal{W}_n$. By the above asymptotical normality of $\mathcal{W}_n$, we can easily get that $\mathbf{u} = O_p(1)$, $\mathcal{W}_n^{j'} = O_p(1)$. By Theorem 1, we have $\sqrt{n}\begin{pmatrix} \hat{\theta}_d \\ \hat{\beta}_d \end{pmatrix} = O_p(1)$. Since by condition A.2, $\mathcal{G}_{(j')}$ and $C_0^{(j')}$ are all bounded by those definitions. Combining above results, we have

$$\mathcal{G}^{(j')}\mathbf{u} - \mathcal{W}_n^{j'} + \sqrt{n} \begin{pmatrix} \hat{\theta}_d \\ \hat{\beta}_d \end{pmatrix}^T C_0^{(j')} = O_p(1). \tag{29}$$

Now, consider $\frac{\lambda_n}{\sqrt{n}|\hat{\theta}_{j'}^{QR}|^2}$. Since $j' \notin \mathcal{A}_\theta, \theta_{\tau j'} = 0$, by Theorem 1, we have $\hat{\theta}_{j'}^{QR} \rightarrow_p 0$, $\sqrt{n}\hat{\theta}_{j'}^{QR} = \sqrt{n}(\hat{\theta}_{j'}^{QR} - \theta_{\tau j'}) = O_p(1)$. Combined this with $\lambda \rightarrow \infty$, we have $\frac{\lambda_n}{\sqrt{n}|\hat{\theta}_{j'}^{QR}|^2} \rightarrow_p \infty$ and that

$$\begin{aligned} &\Pr(j' \in \hat{\mathcal{A}}_\theta) \\ &= \Pr\left[\left| \mathcal{G}^{(j')}\mathbf{u} - \mathcal{W}_n^{j'} + \sqrt{n} \begin{pmatrix} \hat{\theta}_d \\ \hat{\beta}_d \end{pmatrix}^T C_0^{(j')} \right| = \left| \frac{\lambda_n}{\sqrt{n}|\hat{\theta}_{j'}^{QR}|^2} \right| \right] \rightarrow_p 0. \end{aligned}$$

We complete the proof of $\Pr(\hat{\mathcal{A}}_\theta = \mathcal{A}_\theta) \rightarrow_p 1$. Similarly, $\Pr(\hat{\mathcal{A}}_\beta = \mathcal{A}_\beta) \rightarrow_p 1$ can be also derived. □

Appendix B

In this Appendix, we outline the proof of Lemma 1 in Sect. 2. Given any θ and β, denote $\beta_d = \beta - \beta_\tau, \theta_d = \theta - \theta_\tau$, $\delta_\beta = |\beta - \beta_\tau|, \delta_\theta = |\theta - \theta_\tau|$. $\delta_{pn} = \{\log n/(nb^p)\}^{1/2}$, $\tau_{pn} = b^2 + \delta_{pn}$.

Proof of Theorem 1 For any inner point $\mathbf{x} \in \Xi$, the values of $g(\mathbf{x}^T\theta)$, $g'(\mathbf{x}^T\theta)$ can be estimated by

$$(\bar{a}, \bar{b}) = \underset{a,b}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(Y_i - a - b\mathbf{X}_{i0}^T\theta - \mathbf{Z}_i^T\beta)H_b(\mathbf{X}_{i0}), \tag{30}$$

where, $H_b(X_{i0}) = \frac{1}{b^p}H(\frac{\mathbf{X}_i - \mathbf{x}}{b})$. Then the estimation of θ_τ and β_τ can be updated by

$$(\bar{\theta}, \bar{\beta}) = \underset{\theta,\beta}{\operatorname{argmin}} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau(Y_i - \bar{a}_j - \bar{b}_j(\mathbf{X}_i - \mathbf{X}_j)^T\theta - \mathbf{Z}_i^T\beta)W_{ij}, \tag{31}$$

where $\bar{a}_j, \bar{b}_j$ is the estimated values of $g(\mathbf{X}_j^T\theta_\tau)$, $g'(\mathbf{X}_j^T\theta_\tau)$ by (30) respectively, $W_{ij} = \frac{\mathbf{H}_b(\mathbf{X}_{ij})}{\sum_{l=1}^n \mathbf{H}_b(\mathbf{X}_{lj})}$.

For convenience, denote $\varepsilon_i = Y_i - g(\mathbf{X}_i^T\theta_\tau) - \mathbf{Z}_i^T\beta$, thus the τth quantile of ε_i will equal zero when given $\mathbf{X}_i, \mathbf{Z}_i$. Denote $\begin{pmatrix} \bar{\theta} - \theta_\tau \\ \bar{\beta} - \beta_\tau \end{pmatrix} = \bar{\zeta}$, then $\bar{\zeta}$ will be the minimizer of

$$\mathcal{V}_n(\zeta) = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[\rho_\tau(\varepsilon_i - \bar{r}_{ij} - N_{ij}^T\zeta) - \rho_\tau(\varepsilon_i - \bar{r}_{ij})\right]W_{ij} \tag{32}$$

where $N_{ij} = \begin{pmatrix} \bar{b}_j\mathbf{X}_{ij} \\ \mathbf{Z}_i \end{pmatrix}$, $\mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j, \bar{r}_{ij} = \bar{a}_j + \bar{b}_j\mathbf{X}_{ij}^T\theta_\tau - g(\mathbf{X}_i^T\theta_\tau)$.

Using the Knight's identity, we can write

$$\begin{aligned}
\mathcal{V}_n(\zeta) &= -\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij}\psi_\tau(\varepsilon_i)N_{ij}^T\zeta \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij}\int_{\bar{r}_{ij}}^{\bar{r}_{ij}+N_{ij}^T\zeta}[I(\varepsilon_i \le s) - I(\varepsilon_i \le 0)]ds, \\
&\equiv \mathcal{V}_{1n}(\zeta) + \mathcal{V}_{2n}(\zeta),
\end{aligned}$$

where

$$\mathcal{V}_{1n}(\zeta) = -\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij}\psi_\tau(\varepsilon_i)N_{ij}^T\zeta,$$
$$\mathcal{V}_{2n}(\zeta) = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij}\int_{\bar{r}_{ij}}^{\bar{r}_{ij}+N_{ij}^T\zeta}(I(\varepsilon_i \le s) - I(\varepsilon_i \le 0))\mathrm{d}s.$$

Firstly, consider $\mathcal{V}_{2n}(\zeta)$. Denote $\mathcal{X}$ be the $\sigma-$field generated by $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$, then the conditional expectation of $\mathcal{V}_{2n}(\zeta)$ will be:

$$\begin{aligned}
\mathrm{E}\left(\mathcal{V}_{2n}(\zeta)|\mathcal{X}\right) &= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\int_{\bar{r}_{ij}}^{\bar{r}_{ij}+N_{ij}^T\zeta} W_{ij}\left[s f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)(1+o(1))\right]\mathrm{d}s \\
&= \frac{1}{2}\zeta^T\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)N_{ij}N_{ij}^T W_{ij}\right)\zeta \\
&\quad + \left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\bar{r}_{ij}N_{ij}\right)^T\zeta + o_p(1).
\end{aligned}$$

Define $\mathcal{R}_n(\zeta) = \mathcal{V}_{2n}(\zeta) - \mathrm{E}(\mathcal{V}_{2n}(\zeta)|\mathcal{X})$. It can be shown that $\mathcal{R}_n(\zeta) = o_p(1)$. Therefore, we have

$$\mathcal{V}_{2n}(\zeta) \equiv \mathcal{V}_{2n1}(\zeta) + \mathcal{V}_{2n2}^T\zeta + o_p(1),$$

where

$$\mathcal{V}_{2n1}(\zeta) = \frac{1}{2}\zeta^T\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)N_{ij}N_{ij}^T\right)\zeta,$$
$$\mathcal{V}_{2n2} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)N_{ij}\bar{r}_{ij}.$$

Consider $\mathcal{V}_{2n1}(\zeta)$. Denote $\mathcal{L}_n^\theta = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)N_{ij}N_{ij}^T W_{ij}$, we have

$$\mathcal{L}_n^\theta = \mathcal{L}^\theta, \tag{33}$$

where

$$\mathcal{L}^\theta = \begin{pmatrix} b^2(\theta^T\theta_\tau)^2\mathrm{E}\left[g'(\mathbf{X}^t\theta_\tau)^2 f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\right]I_{p\times p} + O(b\delta_{pn} + b^2\delta_\beta) & O(b^2 + b\delta_{pn}) \\ O(b^2 + b\delta_{pn}) & \mathrm{E}\left[f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\mathbf{Z}\mathbf{Z}^T\right] + O(b + \delta_{pn}) \end{pmatrix}.$$

Now consider $\mathcal{V}_{2n2}$. We can get

$$\mathcal{V}_{2n2} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} N_{ij} \bar{r}_{ij} f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i)$$
$$= \left\{ \begin{array}{l} \theta^T\theta_\tau(\theta^T\theta_\tau - 1)\theta_\tau \mathrm{E}\left[\left(g'(\mathbf{X}^T\theta_\tau)\right)^2 f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right] + O(b^3 + b\delta_{pn} + b^2\delta_\beta) \\ \mathrm{E}\left[\nu(\mathbf{X})\nu^T(\mathbf{X}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right](\beta - \beta_\tau) + O(b + \delta_{pn}) \end{array} \right\}. \tag{34}$$

The proof of (33) and (34) can be found in the latter part of this Appendix.

Then for the new value of β_τ, denoted by $\bar{\beta}$, we have

$$\bar{\beta} - \beta_\tau = \left\{\mathrm{E}\left[\mathbf{Z}\mathbf{Z}^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]\right\}^{-1} \mathrm{E}\left[\nu(\mathbf{X})\nu(\mathbf{X})^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right](\beta - \beta_\tau)$$
$$+ O(b + b^{-1}\delta_{pn}).$$

Note that multidimensional kernel is used in this algorithm, the above equation does not depend on the choice of θ. Replacing β as $\bar{\beta}_k$ and $\bar{\beta}$ as $\bar{\beta}_{k+1}$, we have

$$\bar{\beta}_{k+1} - \beta_\tau = \left\{\mathrm{E}\left[\mathbf{Z}\mathbf{Z}^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]\right\}^{-1} \mathrm{E}$$
$$\times \left[\nu(\mathbf{X})\nu(\mathbf{X})^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right](\bar{\beta}_k - \beta_\tau) + O(b + b^{-1}\delta_{pn}).$$

By using the regular condition A.6, $\mathrm{E}\left[\mathbf{Z}\mathbf{Z}^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right] - \mathrm{E}\left[\nu(\mathbf{X})\nu(\mathbf{X})^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right] = \mathbf{E}\left[(\mathbf{Z} - \nu(\mathbf{X}))(\mathbf{Z} - \nu(\mathbf{X}))^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]$ is a positive matrix, similar conclusion can be derived for $\mathrm{E}\left[\mathbf{Z}\mathbf{Z}^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]$ and $\mathrm{E}\left[\nu(\mathbf{X})\nu(\mathbf{X})^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]$. Thus, all the eigenvalues of $\left\{\mathrm{E}\left[\mathbf{Z}\mathbf{Z}^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]\right\}^{-1} \mathrm{E}\left[\nu(\mathbf{X})\nu(\mathbf{X})^T f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X})\right]$ are less than 1. After sufficiently many iterations, we have

$$\bar{\beta}_k - \beta_\tau = O(b + b^{-1}\delta_{pn}) \to_p 0.$$

If $\theta^T\theta_\tau \neq 0$, then following the previous results, we have

$$\bar{\theta}_k - \theta_\tau = (\theta^T\theta_\tau)^{-1}(1 - \theta^T\theta_\tau)\theta_\tau + O(\delta_\beta + b + b^{-1}\delta_{pn}).$$

Since $\delta_\beta \to_p 0$, we have

$$\bar{\theta} := \mathrm{Sgn}_1\bar{\theta}/|\bar{\theta}| = \theta_\tau + O(b + b^{-1}\delta_{pn}).$$

The proof is completed. $\square$

For the estimated $\bar{a}$ and $\bar{b}$ derived by (30), we have following results.

Lemma 5 *Suppose $b \to 0$, and $nb^{p+2}/\log n \to \infty$ and the regular conditions* A.1–A.6 *hold, we have*

$$\bar{a} = g(\mathbf{x}^T\theta_\tau) + \nu(\mathbf{x})^T\beta_d + O(b + \delta_{pn}), \tag{35}$$
$$\bar{b} = \theta^T\theta_\tau g'(\mathbf{x}^T\theta_\tau) + O\left\{\delta_\beta + b^{-1}\delta_{pn} + b\right\}, \tag{36}$$

uniformly for $\mathbf{x} \in \Xi$.

Proof of lemma 5 For convenience, let $\varepsilon_i = Y_i - g(\mathbf{X}_i^T\theta_\tau) - \mathbf{Z}_i^T\beta_\tau$, then the conditional mean of $\psi_\tau(\varepsilon_i)$ is zero when given $(\mathbf{X}_i, \mathbf{Z}_i)$. Denote $\bar{\xi} = \begin{pmatrix} \bar{a} - g(\mathbf{x}^T\theta_\tau) \\ (\bar{b} - g'(\mathbf{x}^T\theta_\tau))b \end{pmatrix}$, $N_i = \begin{pmatrix} 1 \\ b^{-1}(\mathbf{X}_i - \mathbf{x})^T\theta \end{pmatrix}$, then $\bar{\xi}$ is the minimizer of

$$V_n(\xi) = \frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_{i0})\left[\rho_\tau(\varepsilon_i - \bar{r}_i - N_i^T\xi) - \rho_\tau(\varepsilon_i - \bar{r}_i)\right], \tag{37}$$

where $\bar{r}_i = g(\mathbf{x}^T\theta_\tau) + g'(\mathbf{x}^T\theta_\tau)(\mathbf{X}_i - \mathbf{x})^T\theta - g(\mathbf{X}_i^T\theta_\tau) + \mathbf{Z}_i^T(\beta - \beta_\tau)$.

By using Knight's identity, we can rewrite $V_n(\xi)$ as

$$V_n(\xi) = V_{1n}^T\xi + V_{2n}(\xi), \tag{38}$$

where

$$V_{1n} = -\frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_{i0})N_i\psi_\tau(\varepsilon_i)$$
$$V_{2n}(\xi) = \frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_{i0})\int_{\bar{r}_i}^{\bar{r}_i + N_i^T\xi}\left[I(\varepsilon \le s) - I(\varepsilon \le 0)\right]. \tag{39}$$

By calculating the conditional expectation on $\mathcal{X}$, we can get

$$\begin{aligned}
\mathrm{E}(V_{2n}(\xi)|\mathcal{X}) &= \frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_{i0})\int_{\bar{r}_i}^{\bar{r}_i + N_i^T\xi}\mathrm{E}\left[I(\varepsilon \le s) - I(\varepsilon \le 0)\right]\mathrm{d}s. \\
&= \frac{1}{2}\xi^T\left\{\frac{1}{n}\sum_{i=1}^N \mathbf{H}_b(\mathbf{X}_{i0})f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i)N_iN_i^T\right\}\xi \\
&\quad + \left\{\frac{1}{n}\sum_{i=1}^N \mathbf{H}_b(\mathbf{X}_{i0})f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i)\bar{r}_iN_i\right\}^T\xi \\
&\equiv V_{2n1}(\xi) + V_{2n2}^T\xi,
\end{aligned}$$

where

$$V_{2n1}(\xi) = \frac{1}{2}\xi^T \left\{ \frac{1}{n} \sum_{i=1}^{N} \mathbf{H}_b(\mathbf{X}_{i0}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i) N_i N_i^T \right\} \xi,$$

$$V_{2n2} = \frac{1}{n} \sum_{i=1}^{N} \mathbf{H}_b(\mathbf{X}_{i0}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i) \bar{r}_i N_i.$$

Let $R_n = V_{2n1}(\xi) - E(V_{2n1}(\xi)|\mathcal{X})$, it is easy to show that $R_n = o_p(1)$, and then we have $V_{2n}(\xi) = V_{2n1}(\xi) + V_{2n2}^T \xi$.

Firstly, consider $V_{2n1}(\xi)$. Denote $L_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{N} \mathbf{H}_b(\mathbf{X}_{i0}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i) N_i N_i^T$, note that $\theta^T\theta = 1$, by Lemma 3, we can easily get that

$$L_n(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{x}) I_2(1 + O(\tau_{pn})),$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the marginal density function of $\mathbf{X}$ at $\mathbf{x}$. Thus,

$$V_{2n1}(\xi) = \frac{1}{2} f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{x}) \xi^T I_2 \xi (1 + O(\tau_{pn})). \tag{40}$$

Now consider V_{2n2}. Note that

$$\begin{aligned}
\bar{r}_i =& g(\mathbf{x}^T\theta_\tau) + g'(\mathbf{x}^T\theta_\tau)(\mathbf{X}_i - \mathbf{x})^T\theta - g(\mathbf{X}_i^T\theta_\tau) + \mathbf{Z}_i^T(\beta - \beta_\tau) \\
=& -\frac{1}{2} g''(\mathbf{x}^T\theta_\tau)(\mathbf{X}_{i0}^T\theta_\tau)^2 + g'(\mathbf{x}^T\theta_\tau)\mathbf{X}_{i0}^T(\theta - \theta_\tau) + o((\mathbf{X}_{i0}^T\theta_\tau)^2) + \mathbf{Z}_i^T\beta_d.
\end{aligned}$$

By using Lemma 3 and similar calculation, we can get that

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \mathbf{H}_b(\mathbf{X}_{i0}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i) \bar{r}_i = f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{x}) \nu(\mathbf{x})^T \beta_d + O(b^2 + \delta_{pn}), \\
&\frac{1}{n} \sum_{i=1}^{n} \mathbf{H}_b(\mathbf{X}_{i0}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{X}_i) \bar{r}_i b^{-1} \mathbf{X}_{i0}^T \theta \\
&= f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{x}) g'(\mathbf{x}^T\theta_\tau) \theta^T(\theta - \theta_\tau) b + O(b\delta_\beta + \tau_{pn}).
\end{aligned}$$

Thus, we have

$$V_{2n2} = f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{x}) \begin{pmatrix} \nu(\mathbf{x})^T \beta_d + O(b^2 + \delta_{pn}) \\ g'(\mathbf{x}^T\theta_\tau) b \theta^T(\theta - \theta_\tau) + O(b\delta_\beta + \tau_{pn}) \end{pmatrix}. \tag{41}$$

Combining (38), (40) and (41), we have

$$\begin{aligned}
V_n(\xi) = \frac{1}{2}\xi^T L_n(\mathbf{x})\xi - V_{1n}(\xi) + f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X}, \mathbf{Z})|\mathbf{x}) \\
\times \begin{pmatrix} \nu(\mathbf{x})^T \beta_d + O(b^2 + \delta_{pn}) \\ g'(\mathbf{x}^T\theta_\tau) b \theta^T(\theta - \theta_\tau) + O(b\delta_\beta + \tau_{pn}) \end{pmatrix}.
\end{aligned}$$

By using Lemma 2, the minimizer of (32) can be expressed as

$$\bar{\xi} = L(\mathbf{x})^{-1}V_{1n} - \begin{pmatrix} \nu(\mathbf{x})^T\beta_d + O(b^2+\delta_{pn}) \\ g'(\mathbf{x}^T\theta_\tau)b\theta^T(\theta-\theta_\tau) + O(b\delta_\beta+\tau_{pn}) \end{pmatrix}. \tag{42}$$

Note that the mean of $\psi_\tau(\varepsilon_i)$ is zero, by using Lemma 3, $V_{1n} = O(b^2+\delta_{pn})$. Note the definition of $\bar{\xi}$, and combining the result of V_{1n} with (42), we complete the proof. □

Proof of 33 To proof (33), it is sufficient to show that

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\bar{b}_j^2\mathbf{X}_{ij}\mathbf{X}_{ij} \\
&\quad = b^2(\theta^T\theta_\tau)^2\mathrm{E}\left[g'(X)^2 f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\right] I_{p\times p} + O(b\tau_{pn}+b^2\delta_\beta), \\
&\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\bar{b}_j\mathbf{X}_{ij}\mathbf{Z}_i^T = O(b^2+b\delta_{pn}), \\
&\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\mathbf{Z}_i\mathbf{Z}_i^T = \mathrm{E}\left[f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\mathbf{Z}\mathbf{Z}^T\right] + O(b+b\delta_{pn}).
\end{aligned}$$

Here, we give the details for the first equation. Let $\kappa_0(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{H}(\mathbf{X}_i-\mathbf{x})$, then by Lemma 3, we can get that

$$\kappa_0(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) + O(\tau_{pn}).$$

By Lemma 5, we have $\bar{b}_j = \theta^T\theta_\tau g'(\mathbf{X}_j^T\theta_\tau) + O(\delta_\beta + b^{-1}\delta_{pn} + b)$, thus,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\bar{b}_j^2\mathbf{X}_{ij}\mathbf{X}_{ij} \\
=&\frac{1}{n}\sum_{j=1}^{n}(\theta^T\theta_\tau)^2 g'(\mathbf{X}_j)^2 f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_j)b^2 I_{p\times p} + O(b^2\delta_\beta + b\delta_{pn}) \\
=&(\theta^T\theta_\tau)^2\mathrm{E}\left[g'(\mathbf{X})^2 f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\right] b^2 I_{p\times p} + O(b^2\delta_\beta + b\delta_{pn}).
\end{aligned}$$

□

Proof of 34 To proof (34), it is sufficient to show that

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij}\bar{b}_j\mathbf{X}_{ij}\bar{r}_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)$$

$$
\begin{aligned}
&= \theta^T\theta_\tau(\theta^T\theta_\tau - 1)\theta_\tau \mathrm{E}\left[\left(g'(\mathbf{X}^T\theta_\tau)\right)^2 f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\right] \\
&\qquad\qquad\qquad\qquad\qquad + O(b^3 + b\delta_{pn} + b^2\delta_\beta), \\
&\frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n W_{ij}\mathbf{Z}_i\bar{r}_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i) \\
&= \mathrm{E}\left[\nu(\mathbf{X})\nu^T(\mathbf{X}) f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\right](\beta - \beta_\tau) + O(b + \delta_{pn}).
\end{aligned}
$$

Here, we give the details for the second equation. By Lemma 5, we have

$$
\begin{aligned}
\bar{r}_{ij} &= \bar{a}_j + \bar{b}_j\mathbf{X}_{ij}^T\theta_\tau - g(\mathbf{X}_i^T\theta_\tau) \\
&= \bar{a}_j - g(\mathbf{X}_j^T\theta_\tau) + \left(\bar{b}_j - g'(\mathbf{X}_j^T\theta_\tau)\right) X_{ij}^T\theta_\tau + g'(\mathbf{X}_j^T\theta_\tau)\mathbf{X}_{ij}^T\theta_\tau \\
&\quad - \left(g(\mathbf{X}_i^T\theta_\tau) - g(\mathbf{X}_j^T\theta_\tau)\right) \\
&= \nu(\mathbf{X}_j)^T\beta_d + (\theta^T\theta_\tau - 1)g'(\mathbf{X}_j^T\theta_\tau)\mathbf{X}_{ij}^T\theta_\tau + \frac{1}{2}g''(\mathbf{X}_j^T\theta_\tau)(\mathbf{X}_{ij}^T\theta_\tau)^2 + O\left((\mathbf{X}_{ij}^T\theta_\tau)^3\right).
\end{aligned}
$$

Denote

$$
\begin{aligned}
C_0(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_i - \mathbf{x}) f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\mathbf{Z}_i, \\
C_1(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_i - \mathbf{x}) f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\mathbf{Z}_i(\mathbf{X}_i - \mathbf{x})^T, \\
C_2(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathbf{H}_b(\mathbf{X}_i - \mathbf{x}) f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\mathbf{Z}_i(\mathbf{X}_i - \mathbf{x})^T(\mathbf{X}_i - \mathbf{x}).
\end{aligned}
$$

By Lemma 3, it is easy to show that

$$
\begin{aligned}
&C_0(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{x})\nu(\mathbf{x}) + O(\tau_{pn}), \\
&C_1(\mathbf{x}) = O(b^2 + b\delta_{pn}), \quad C_2(\mathbf{x}) = O(b^2\tau_{pn}).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n W_{ij}\mathbf{Z}_i\bar{r}_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i) \\
&\quad = \frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n W_{ij} f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\mathbf{Z}_i \\
&\qquad \times \left[\nu(\mathbf{X}_j)^T\beta_d + (\theta^T\theta_\tau - 1)g'(\mathbf{X}_j^T\theta_\tau)\mathbf{X}_{ij}^T\theta_\tau + O\left((\mathbf{X}_{ij}^T\theta_\tau)^2\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{n}\kappa_0(\mathbf{X}_j)^{-1}\Big[C_0(\mathbf{X}_j)\nu(\mathbf{X}_j)^T\beta_d + (\theta^T\theta_\tau - 1)g'(\mathbf{X}_j^T\theta_\tau)C_1(\mathbf{X}_j)\theta \\
&\qquad\qquad + O(\theta_\tau^T C_2(\mathbf{X}_j)\theta_\tau)\Big] \\
&= \frac{1}{n}\sum_{j=1}^{n}\nu(\mathbf{X}_j)^T\beta_d\sum_{i=1}^{n}W_{ij}f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X}_i)\mathbf{Z}_i + O(b+\delta_{pn}) \\
&= \mathrm{E}\Big[\nu(\mathbf{X})\nu^T(\mathbf{X})f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\Big]\beta_d + O(b+\delta_{pn}) \\
&= \mathrm{E}\Big[\nu(\mathbf{X})\nu^T(\mathbf{X})f_Y(q_\tau(\mathbf{X},\mathbf{Z})|\mathbf{X})\Big](\beta-\beta_\tau) + O(b+\delta_{pn}).
\end{aligned}
$$

The proof of the first equation is similar. □

Acknowledgments We sincerely thank three referees and associate editor for their valuable comments that has led to a great improved presentation of our work.

References

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489.

Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, *39*, 246–269.

Chaudhuri, P., Doksum, K., Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, *25*, 715–744.

Fan, J., Li, R. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*, 1348–1360.

Fan, J., Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, *99*, 710–723.

Fan, Y., Härdle, W., Wang, W., Zhu, L. (2013). Composite quantile regression for the single-index model. In *Sonderforschungsbereich 649: Ökonomisches Risiko-(SFB 649 Papers)*. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät.

Geyer, C. J. (1994). On the asymptotics of constrained m-estimation. *The Annals of Statistics*, *22*, 1993–2010.

Härdle, W., Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of American Statistical Association*, *84*, 986–995.

Härdle, W., Gao, J., Liang, H. (2007). *Partially Linear Models*. Berlin: Springer.

Hjort, N., Pollard, D. (1993). Asymptotics for minimizers of convex processes. Preprint.

Jiang, R., Zhou, Z., Qian, W., Shao, W. (2012). Single-index composite quantile regression. *Journal of the Korean Statistical Society*, *41*, 323–332.

Kai, B., Li, R., Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of Statistics*, *39*, 305–332.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics*, *26*, 755–770.

Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.

Koenker, R., Basset, G. S. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Kong, E., Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, *28*, 730–768.

Liang, H., Li, R., Liu, X., Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, *38*, 3811–3836.

Mack, Y. P., Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields*, *61*, 405–415.

Ruppert, D., Sheather, S. J., Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, *90*, 1257–1270.

Ruppert, D., Wand, M. P., Carroll, R. J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

Wang, H., Leng, C. (2007). Unified lasso estimation via least squares approximation. *Journal of the American Statistical Association*, *102*, 1039–1048.

Wu, T., Yu, K., Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, *101*, 1607–1621.

Xia, Y., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, *97*, 1162–1184.

Yu, K., Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, *93*, 228–237.

Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, *97*, 1042–1054.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429.

Zou, H., Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, *36*, 1108–1126.