

INFERENCE FOR SINGLE-INDEX QUANTILE REGRESSION MODELS WITH PROFILE OPTIMIZATION

BY SHUIE MA AND XUMING HE

University of California, Riverside and University of Michigan

Single index models offer greater flexibility in data analysis than linear models but retain some of the desirable properties such as the interpretability of the coefficients. We consider a pseudo-profile likelihood approach to estimation and testing for single-index quantile regression models. We establish the asymptotic normality of the index coefficient estimator as well as the optimal convergence rate of the nonparametric function estimation. Moreover, we propose a score test for the index coefficient based on the gradient of the pseudo-profile likelihood, and employ a penalized procedure to perform consistent model selection and model estimation simultaneously. We also use Monte Carlo studies to support our asymptotic results, and use an empirical example to illustrate the proposed method.

1. Introduction. Quantile regression as an alternative to modeling the conditional mean function has gained attention since the seminal work of [Koenker and Bassett (1978)]. By direct modeling of the conditional quantile functions, quantile regression leads to a more comprehensive regression analysis than the least squares methods. Let Y be the response variable, and $\mathbf{X} = (X_1, \dots, X_p)^T$ be the p -dimensional covariates. For any given $0 < \tau < 1$ and $\mathbf{x} \in R^p$, a characterization of the τ th conditional quantile of Y is

$$(1) \quad G_\tau(\mathbf{x}) = \arg \min_a E\{\rho_\tau(Y - a) | \mathbf{X} = \mathbf{x}\},$$

where $\rho_\tau(s) = \tau s - sI(s < 0)$ is the quantile loss function. It is often convenient to assume that $G_\tau(\mathbf{x})$ has a specific parametric form such as $G_\tau(\mathbf{x}) = \alpha_\tau + \mathbf{x}^T \boldsymbol{\beta}_\tau$, where $(\alpha_\tau, \boldsymbol{\beta}_\tau^T)^T \in R^{p+1}$ is an unknown coefficient vector. For better flexibility, nonparametric quantile regression has also been studied in the literature, including the kernel-based methods [Bhattacharya and Gangopadhyay (1990), Fan, Hu and Truong (1994), Yu and Jones (1998)] and the spline-based methods [He and Shi (1994), Koenker, Ng and Portnoy (1994)]. When the covariates are multivariate, a fully nonparametric model suffers from the “curse of dimensionality,” but some asymptotic theory is available in Chaudhuri (1991) and He, Ng and Portnoy (1998) among others.

Received April 2015; revised October 2015.

MSC2010 subject classifications. Primary 62G08; secondary 62G20.

Key words and phrases. Quantile regression, single-index, profile principle, polynomial spline, model selection, score test.

To achieve dimension reduction, a variety of structural models have been considered in quantile regression. De Gooijer and Zerom (2003), Horowitz and Lee (2005) and Koenker (2011) proposed additive quantile regression models. Kim (2007) and Wang, Zhu and Zhou (2009) considered varying coefficient models. In this paper, we consider a semiparametric single-index model by assuming $G_\tau(\cdot)$ to be a function of a linear index $\mathbf{x}^\top \boldsymbol{\beta}_\tau$. Single-index models were first proposed by Ichimura (1993) and have been widely used in the literature. The existing methods of estimating parameters of the conditional mean in a single-index model include the backfitting algorithm [Carroll et al. (1997)], the penalized spline estimation [Yu and Ruppert (2002)], the minimum average variance estimation [MAVE, Xia, Tong and Li (1999) and Xia and Härdle (2006)], and the profile least squares estimation [Liang et al. (2010)]. For estimation of conditional quantiles under restrictive settings, asymptotic results have been obtained by Zhu, Huang and Li (2012) assuming a linearity condition on \mathbf{X} and by Zou and Zhu (2014) for i.i.d. error models. Moreover, Wu, Yu and Yu (2010) proposed a modified version of the MAVE method, and Kong and Xia (2012) refined the algorithm by introducing a penalty term. Most authors considered estimation of the index coefficient and the nonparametric function iteratively. To achieve estimation consistency, the iterative algorithm requires a consistent initial estimate of the index. Wu, Yu and Yu (2010) suggested the root- n consistent average derivative quantile estimator proposed in Chaudhuri, Doksum and Samarov (1997) as the initial estimate.

The focus of the present paper is the method of profiling for quantile estimation and inference. For a given $\boldsymbol{\beta}_\tau$, the function $G_\tau(\cdot)$ can be estimated by a spline [de Boor (2001)] by minimizing the sample analogue of (1), and the resulting estimator $\widehat{G}_\tau(\cdot, \boldsymbol{\beta}_\tau)$ of $G_\tau(\cdot)$ is a function of $\boldsymbol{\beta}_\tau$. By replacing $G_\tau(\cdot)$ with its estimate $\widehat{G}_\tau(\cdot, \boldsymbol{\beta}_\tau)$, we obtain the estimator of $\boldsymbol{\beta}_\tau$ minimizing the objective function over $\boldsymbol{\beta}_\tau$. The same idea in the profile least squares estimation via kernel smoothing has been studied in Liang et al. (2010). Empirical studies in Liang et al. (2010) show that the profile method leads to a more stable estimate than other iterative methods, which is corroborated by our numerical studies for quantile regression. We also demonstrate that the profile estimator is less sensitive to the starting value than the iterative methods.

We consider the profile approach for another important reason, that is, the approach defines naturally a single objective function over $\boldsymbol{\beta}_\tau$, making it convenient for statistical inference on the index parameter. As we know from the work of Murphy and van der Vaart (2000), among others, profile likelihood is an attractive likelihood-based approach to inference in semiparametric models. The approach we take in this paper is based on profile pseudo-likelihood, where the pseudo-likelihood corresponds to the quantile objective function. The profile pseudo-likelihood enables us to develop a score test based on the gradient of the quantile objective function for the null hypothesis of zero coefficients in $\boldsymbol{\beta}_\tau$, the performance of which is more stable than other tests based on the asymptotic variance-covariance matrix. The main difficulty in developing the score test statistic is the

lack of differentiability of $\widehat{G}_\tau(\cdot, \boldsymbol{\beta}_\tau)$ as a function of $\boldsymbol{\beta}_\tau$, and as a result, the rank score of the profile quantile loss function has to be approximated. We derive an effective approximation to overcome the lack of differentiability by considering a smooth version of the loss function. Similar to the rank score test in linear models, the proposed test is shown to have a chi-square limiting distribution under the null hypothesis.

The fact that we have a single objective function of $\boldsymbol{\beta}_\tau$ also enables us to consider, in the presence of multiple covariates, the method of penalized estimation, which shrinks some coefficients of the single-index components to zero. We apply the local linear approximation (LLA) method proposed in Zou and Li (2008), and show that the resulting estimator has desired asymptotic properties. Since the pseudo-likelihood function is not convex in $\boldsymbol{\beta}_\tau$, the convexity lemma [Pollard (1991)] and the quadratic approximation used for analyzing linear quantile regression estimators cannot be directly applied. In this paper, we verify that the profile approach leads to a consistent and asymptotically normal estimator of the index parameter, and it has the same asymptotic variance as the iterative estimator of Wu, Yu and Yu (2010).

The rest of this paper is organized as follows. We describe the profile approach to quantile regression estimation in a single-index model in Section 2, and its penalized version for variable selection in Section 3. In both cases, we provide the asymptotic properties of the proposed estimators. Some implementation details are given in Section 4. Section 5 proposes the score test based on the gradient of the profile quantile loss function. Such tests are made possible by the availability of a single objective function in the profile approach. In Sections 6 and S.1, we evaluate the finite sample properties of the proposed procedures via simulation studies. Section 7 illustrates the proposed method through an example. Some concluding remarks are given in Section 8. All technical details including detailed proofs are provided in the Appendix and on-line supplemental materials [Ma and He (2015)].

2. Profile estimation. In this paper, we consider a single-index model for the conditional quantile such that the τ th quantile of Y given $\mathbf{X} = \mathbf{x}$ is $G_\tau(\mathbf{x}^T \boldsymbol{\beta}_\tau^0)$ for some unknown index parameter $\boldsymbol{\beta}_\tau^0$ and unknown function G_τ , where \mathbf{x} lies in a compact set \mathcal{C} . For convenience, we write

$$(2) \quad Y = G_\tau(\mathbf{X}^T \boldsymbol{\beta}_\tau^0) + \varepsilon,$$

where the error term satisfies $P(\varepsilon \leq 0 | \mathbf{X}) = \tau$ for any $\mathbf{X} \in \mathcal{C}$. Note that the quantile regression is τ -specific, so is the error term ε .

Due to the nonparametric nature of G_τ , the scale of the index parameter $\boldsymbol{\beta}_\tau^0$ is not identifiable. Throughout the paper, we assume that $\boldsymbol{\beta}_\tau^0$ belongs to the parameter space

$$\Theta = \{\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T : \|\boldsymbol{\beta}\|_2 = 1, \beta_1 > 0, \boldsymbol{\beta} \in R^p\},$$

where $\|\mathbf{a}\|_2$ denotes the L_2 norm for any vector $\mathbf{a} = (a_1, \dots, a_p)^T$, and p is a fixed number, not increasing with n . The particular choice of Θ , however, is not important. To identify the function G_τ , we define the support of the function as $[a, b]$, where a and b are the infimum and the supremum of $\{\mathbf{x}^T \boldsymbol{\beta}, \mathbf{x} \in \mathcal{C}, \boldsymbol{\beta} \in \Theta\}$, respectively.

We consider statistical inference based on a random sample $(\mathbf{X}_i, Y_i : 1 \leq i \leq n)$ from the above model. Let $U(\mathbf{x}; \boldsymbol{\beta}) = \mathbf{x}^T \boldsymbol{\beta}$ be the linear index. Now for any given $\boldsymbol{\beta} \in \Theta$ and $u \in [a, b]$, we define $\tilde{G}_\tau(u, \boldsymbol{\beta})$ to be the τ th quantile function of Y given $\mathbf{X}^T \boldsymbol{\beta} = u$. It then holds that $\boldsymbol{\beta}_\tau^0 \equiv \arg \min_{\boldsymbol{\beta} \in \Theta} L_\tau^*(\boldsymbol{\beta})$, where

$$L_\tau^*(\boldsymbol{\beta}) = E[\rho_\tau\{Y - \tilde{G}_\tau(\mathbf{X}^T \boldsymbol{\beta}, \boldsymbol{\beta})\} - \rho_\tau(Y)],$$

and we assume throughout the paper that $\boldsymbol{\beta}_\tau^0$ is the unique minimizer of $L_\tau^*(\boldsymbol{\beta})$. Moreover, $G_\tau(\mathbf{X}^T \boldsymbol{\beta}_\tau^0) = \tilde{G}_\tau(\mathbf{X}^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)$. As we will make clear later, the interval $[a, b]$ does not need to be pre-specified in our estimation procedure. It is the existence of such an interval on which G_τ could be defined that is required for the theoretical results in the paper.

2.1. *Estimation of index.* First, we introduce the B-splines that will be used to approximate the unknown function \tilde{G}_τ . Let $a = t_0 < t_1 < \dots < t_N < b = t_{N+1}$ be a partition of $[a, b]$ into subintervals $I_j = [t_j, t_{j+1})$, $0 \leq j \leq N - 1$ and $I_N = [t_N, t_{N+1}]$, satisfying

$$\max_{0 \leq j \leq N} |t_{j+1} - t_j| / \min_{0 \leq j \leq N} |t_{j+1} - t_j| \leq M$$

uniformly in n for some constant $0 < M < \infty$, where $N = N_n$ increases with the sample size n . Consider the space of polynomial splines of order $m \geq 2$ on $[a, b]$. We write the normalized B spline basis of this space [de Boor (2001), page 89] as $B(u) = \{B_j(u) : 1 \leq j \leq J_n\}^T$, where $J_n = N_n + m$. In the estimation problem, the interval $[a, b]$ is unknown. In the empirical implementations, we use the minimal and maximal values of $(\mathbf{X}_i^T \boldsymbol{\beta}, 1 \leq i \leq n)$ as the two boundary points to generate B-spline basis functions $B_j(u)$ for each given $\boldsymbol{\beta}$. Outside this interval, $\tilde{G}_\tau(\cdot, \boldsymbol{\beta})$ can be defined in any natural way without affecting the results.

By the result in de Boor (2001), the quantile function can be approximated well by a spline function such that $G_\tau(\mathbf{x}^T \boldsymbol{\beta}_\tau) \approx B(\mathbf{x}^T \boldsymbol{\beta}_\tau)^T \boldsymbol{\theta}_\tau$ for some $\boldsymbol{\theta}_\tau \in R^{J_n}$. Thus, the estimators of the spline coefficients $\boldsymbol{\theta}_\tau$ and the parameter $\boldsymbol{\beta}_\tau$ are obtained by minimizing the following pseudo-likelihood function

$$(3) \quad L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \rho_\tau\{Y_i - B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\}.$$

We then adopt the profile principle described in Severini and Staniswalis (1994) and Murphy and van der Vaart (2000) for estimation of parameters in semi-parametric models to define the profile pseudo-likelihood function of $\boldsymbol{\beta}$ given

as

$$\begin{aligned}
 L_{\tau n}^*(\boldsymbol{\beta}) &= \min_{\boldsymbol{\theta} \in R^{J_n}} L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta}) = L_{\tau n}(\tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta}), \boldsymbol{\beta}) \\
 (4) \qquad &= n^{-1} \sum_{i=1}^n \rho_{\tau}\{Y_i - B(\mathbf{X}_i^T \boldsymbol{\beta})^T \tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})\},
 \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$ is the minimizer of $L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta})$ over $\boldsymbol{\theta} \in R^{J_n}$ for given $\boldsymbol{\beta}$. Thus, the proposed profile estimator of $\boldsymbol{\beta}_{\tau}^0$ is taken to be

$$\hat{\boldsymbol{\beta}}_{\tau} = \arg \min_{\boldsymbol{\beta} \in \Theta} \{L_{\tau n}^*(\boldsymbol{\beta})\}.$$

As discussed in Severini and Staniswalis (1994), in the objective function $L_{\tau n}^*(\boldsymbol{\beta})$, the solutions $\tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$ from minimizing $L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta})$ may or may not have a closed-form expression in terms of $\boldsymbol{\beta}$ depending on the model of study. Clearly, for our single-index quantile model, an explicit form for $\tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$ is not available. We then adopt the idea described in Section 6 of Severini and Staniswalis (1994) for computation. In our algorithm, the estimate of $\boldsymbol{\beta}_{\tau}$ is obtained by nonlinear optimization of the objective function $L_{\tau n}^*(\boldsymbol{\beta})$ in which $\tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$ is written as a function of $\boldsymbol{\beta}$ through optimization of $L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta})$. Moreover, it is apparent that for any given $\boldsymbol{\beta}$, the spline estimate of $\tilde{G}_{\tau}(u, \boldsymbol{\beta})$ is $\tilde{G}_{\tau n}(u, \boldsymbol{\beta}) = B(u)^T \tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$. Denote $\mathbb{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$. Moreover, for given $\boldsymbol{\beta} \in \Theta$, let $\tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$ be the minimizer of $E\{L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta})|\mathbb{X}\}$, and

$$(5) \qquad \tilde{\tilde{G}}_{\tau n}(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta}) = B(\mathbf{X}_i^T \boldsymbol{\beta})^T \tilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta}).$$

To study the asymptotic properties of the estimator, some regularity conditions are needed. Let \mathcal{H}_r be the collection of all functions on $[a, b]$ such that the m th order derivative satisfies the Hölder condition of order γ with $r \equiv m + \gamma$. That is, there exists a constant $C_0 \in (0, \infty)$ such that for each $\phi \in \mathcal{H}_r$,

$$(6) \qquad |\phi^{(m)}(u_1) - \phi^{(m)}(u_2)| \leq C_0 |u_1 - u_2|^{\gamma}$$

for any $a \leq u_1, u_2 \leq b$. We make the following assumptions:

(C1) The density function $f(\cdot)$ of $\mathbf{X}^T \boldsymbol{\beta}$ is bounded away from zero and infinity on its support, for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_{\tau}^0$.

(C2) The conditional density of $f_Y(y|\mathbf{x})$ of Y given $\mathbf{X} = \mathbf{x}$ satisfies the Lipschitz condition of order 1, and $f_Y(\tilde{G}_{\tau}(\mathbf{x}^T \boldsymbol{\beta}, \boldsymbol{\beta})|\mathbf{x}) > 0$ for $\boldsymbol{\beta} \in \Theta$ and $\mathbf{x} \in \mathcal{C}$. Also, we assume $\sup_{\mathbf{x} \in \mathcal{C}, y} f_Y(y|\mathbf{x}) < \infty$.

(C3) $\tilde{G}_{\tau}(\cdot, \boldsymbol{\beta}) \in \mathcal{H}_r$ for some $r > 3/2$, and for any $\boldsymbol{\beta} \in \Theta$.

Let a_0 and b_0 be the infimum and the supremum of $\mathbf{x}^T \boldsymbol{\beta}_{\tau}^0$ over $\mathbf{x} \in \mathcal{C}$. For $u \in [a_0, b_0]$, let $\boldsymbol{\varphi}(u) = \{\varphi_1(u), \dots, \varphi_p(u)\}^T$ be a vector of functions defined on $[a_0, b_0]$, and $E^*(\mathbf{X}|\mathbf{X}^T \boldsymbol{\beta}_{\tau}^0)$ is the minimizer of

$$E\{f_{\varepsilon}(0|\mathbf{X})\|\mathbf{X} - \boldsymbol{\varphi}(\mathbf{X}^T \boldsymbol{\beta}_{\tau}^0)\|_2^2\},$$

over $\varphi(\cdot)$, where $f_\varepsilon(\varepsilon|\mathbf{x})$ is the conditional density of ε given $\mathbf{X} = \mathbf{x}$. To be explicit, we have

$$(7) \quad E^*(\mathbf{X}|\mathbf{X}^T\boldsymbol{\beta}_\tau^0) = \frac{E\{f_\varepsilon(0|\mathbf{X})\mathbf{X}|\mathbf{X}^T\boldsymbol{\beta}_\tau^0\}}{E\{f_\varepsilon(0|\mathbf{X})|\mathbf{X}^T\boldsymbol{\beta}_\tau^0\}}.$$

Additional assumptions are now stated as follows:

(C4) Each component of $E^*(\mathbf{X}|\mathbf{X}^T\boldsymbol{\beta}_\tau^0 = u)$, as a function of $u \in [a_0, b_0] \subseteq [a, b]$, has a continuous and bounded first derivative.

(C5) There exists a constant $c_0 \in (0, \infty)$ such that

$$\sup_{\mathbb{X}} \|\partial \tilde{G}_{\tau n}(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta}) / \partial \boldsymbol{\beta} - \partial \tilde{G}_{\tau n}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0) / \partial \boldsymbol{\beta}\|_2 \leq c_0 \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau^0\|_2,$$

for $\boldsymbol{\beta}$ in any neighborhood of $\boldsymbol{\beta}_\tau^0$.

Conditions (C1)–(C4) are commonly used in the nonparametric smoothing literature; see, for example, Cui, Härdle and Zhu (2011), He and Shi (1996). Condition (C5) is a typical assumption in the regression literature, which can be easily satisfied when p is fixed. By Condition (C2) and the fact that $\varepsilon = Y - \tilde{G}_\tau(\mathbf{X}^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)$, we have that $f_\varepsilon(\varepsilon|\mathbf{X} = \mathbf{x})$ satisfies the Lipschitz condition of order 1 and $f_\varepsilon(0|\mathbf{X}) > 0$. Because $\tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta})$ is a solution to $\partial E\{L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta})|\mathbb{X}\} / \partial \boldsymbol{\theta} = \mathbf{0}$, by the implicit function theorem, Condition (C6) can follow from some smoothness condition on $f_\varepsilon(\varepsilon|\mathbf{x})$. We first state the consistency result. For any positive numbers a_n and b_n , we use $a_n \ll b_n$ to mean $a_n b_n^{-1} = o(1)$.

THEOREM 1 (Consistency). *Under Conditions (C1)–(C3), and $J_n \rightarrow \infty$ and $J_n \ll n$, we have $\|\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau^0\|_2 = o_p(1)$.*

Recall that J_n is the dimension of the spline space. To prepare for the asymptotic normality result, we use $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^T$ for any matrix \mathbf{A} , and let \mathbf{A}^+ be the Moore–Penrose inverse of \mathbf{A} . We define

$$(8) \quad \Lambda = E[f_\varepsilon(0|\mathbf{X})\{G_\tau^{(1)}(\mathbf{X}^T \boldsymbol{\beta}_\tau^0) \tilde{\mathbf{X}}\}^{\otimes 2}], \quad \Omega = E[\{G_\tau^{(1)}(\mathbf{X}^T \boldsymbol{\beta}_\tau^0) \tilde{\mathbf{X}}\}^{\otimes 2}],$$

where $G_\tau^{(1)}(\cdot)$ denotes the first-order derivative of $G_\tau(\cdot)$, and

$$\tilde{\mathbf{X}} = \mathbf{X} - E^*(\mathbf{X}|\mathbf{X}^T \boldsymbol{\beta}_\tau^0).$$

THEOREM 2 (Asymptotic normality). *Under Conditions (C1)–(C5), and if $\max\{(n \log(n))^{1/(3r-1/2)}, n^{1/(2r+2)}\} \ll J_n \ll n^{1/4}/(\log n)^{5/4}$, we have $\sqrt{n}(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau^0) \rightarrow N(0, \tau(1-\tau)\Lambda^+\Omega\Lambda^+)$, as $n \rightarrow \infty$.*

REMARK 1. Note that by the definition of Λ given in (8), we have $\Lambda \boldsymbol{\beta}_\tau^0 = \mathbf{0}$, so Λ is a singular matrix. Hence, we need to use the Moore–Penrose inverse of

Λ in Theorem 2. In model (2), if we assume $Y - G_\tau(\mathbf{X}^T \boldsymbol{\beta}_\tau^0)$ is independent of \mathbf{X} , the asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau^0)$ reduces to $\tau(1 - \tau)\{f_\varepsilon(0)\}^{-2}\Omega^+$, where $f_\varepsilon(\varepsilon)$ is the density function of ε . This special case also provides a correction of Theorem 2 of Zou and Zhu (2014).

REMARK 2. To conduct inference for $\boldsymbol{\beta}_\tau^0$ by using the asymptotic normality given in Theorem 2, we need to estimate the asymptotic covariance matrix. For estimating Ω , we can use its sample analogue. For the estimation of Λ , we refer to Section 3.4 of Koenker (2005) for possible approaches, but careful selection of smoothing parameters is needed.

2.2. *Estimation of the nonparametric function G_τ .* The spline estimator of $G_\tau(u)$ is simply $\tilde{G}_{\tau n}(u, \widehat{\boldsymbol{\beta}}_\tau) = B(u)^T \widehat{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau)$, where $\widehat{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau)$ minimizes

$$L_{\tau n}(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}}_\tau) = n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - B(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau)^T \boldsymbol{\theta}\}],$$

over $\boldsymbol{\theta}$. The following theorem presents the global convergence rate of the estimator for the nonparametric function.

THEOREM 3. *Under the conditions of Theorem 2, we have*

$$n^{-1} \sum_{i=1}^n \{\tilde{G}_{\tau n}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau, \widehat{\boldsymbol{\beta}}_\tau) - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\}^2 = O_p(J_n^{-2r} + J_n n^{-1}).$$

REMARK 3. If the number of spline basis functions J_n is of the order $n^{1/(2r+1)}$, which satisfies the order requirements given in Theorem 3 for $r > 3/2$, the optimal global convergence rate of $n^{-r/(2r+1)}$ is attained for G_τ . Also note that $G_\tau(u)$ can be consistently estimated only for $u \in [a_0, b_0]$ as specified in Condition (C4), even though the estimate of G_τ is used on a broader interval $[a, b]$.

3. Penalized estimation. One advantage of profiling is the availability of a single objective function as a function of $\boldsymbol{\beta}$, with which a common penalty toward sparsity can be introduced to regularize the coefficient when several to many covariates are present. More specifically, we consider minimization of

$$L_{\tau n}^*(\boldsymbol{\beta}) + \sum_{j=1}^p \omega_j p_{\lambda_n}(|\beta_j|)$$

with respect to $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, where $p_{\lambda_n}(\cdot)$ is a penalty function with a regularization parameter λ_n , $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^T$ is a weighting vector, and $L_{\tau n}^*(\boldsymbol{\beta})$ is defined in (4).

Many forms of the penalty functions are available, including convex penalties like LASSO [Tibshirani (1996)] and the nonconvex penalties like SCAD [Fan and

Li (2001)]. In our setting, the objective function is nonconvex even if a convex penalty is used. To have a computationally tractable and stable solution, we propose a modification of the local linear approximation (LLA) method used in Zou and Li (2008) and Kai, Li and Zou (2011) with the SCAD penalty. The details are given as follows.

Let the profile estimator of β_τ obtained from Section 2.1 be the initial estimate, denoted by $\hat{\beta}_\tau^0 = (\hat{\beta}_{\tau 1}^0, \dots, \hat{\beta}_{\tau p}^0)^T$. Also let

$$\tilde{G}_{\tau n}^{(1)}(u, \beta) = \partial \tilde{G}_{\tau n}(u, \beta) / \partial u = \{\partial B(u) / \partial u\}^T \tilde{\theta}_\tau(\beta).$$

We consider a linear approximation

$$\tilde{G}_{\tau n}^*(\mathbf{x}^T \beta, \hat{\beta}_\tau^0) = \tilde{G}_{\tau n}(\mathbf{x}^T \hat{\beta}_\tau^0, \hat{\beta}_\tau^0) + \tilde{G}_{\tau n}^{(1)}(\mathbf{x}^T \hat{\beta}_\tau^0, \hat{\beta}_\tau^0) \mathbf{x}^T (\beta - \hat{\beta}_\tau^0),$$

and define a new objective function

$$(9) \quad Q_{\tau n}(\beta) = L_{\tau n}^{**}(\beta) + \sum_{j=1}^p \omega_j p'_{\lambda_n}(|\hat{\beta}_{\tau j}^0|) |\beta_j|,$$

where

$$L_{\tau n}^{**}(\beta) = n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - \tilde{G}_{\tau n}^*(\mathbf{X}_i^T \beta, \hat{\beta}_\tau^0)\}],$$

and

$$p'_\lambda(\beta) = \lambda \left\{ I(\beta \leq \lambda) + \frac{(a\lambda - \beta)_+}{(a - 1)\lambda} I(\beta > \lambda) \right\},$$

for some $a > 2$. In our empirical work, we take $a = 3.7$, which is used in Fan and Li (2001). Given the initial estimate $\hat{\beta}_\tau^0$, the new objective function $Q_{\tau n}(\beta)$ is similar to the penalized linear quantile regression problem [Wu and Liu (2009)] with linear predictors $\tilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \hat{\beta}_\tau^0, \hat{\beta}_\tau^0) X_{ij}$, $1 \leq j \leq p$, and adjusted response

$$Y_i^* = Y_i - \tilde{G}_{\tau n}(\mathbf{X}_i^T \hat{\beta}_\tau^0, \hat{\beta}_\tau^0) + \tilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \hat{\beta}_\tau^0, \hat{\beta}_\tau^0) \mathbf{X}_i^T \hat{\beta}_\tau^0,$$

for each $i = 1, \dots, n$, and thus it can be solved via linear programming. The penalty term in (9) is a weighted ℓ_1 penalty with the weight $\omega_j p'_{\lambda_n}(|\hat{\beta}_{\tau j}^0|)$ for each $j = 1, \dots, p$. To achieve the right scaling, we choose $\omega_j = \sqrt{\tau(1 - \tau)} \hat{\sigma}_j$, where $\hat{\sigma}_j^2 = n^{-1} \sum_{i=1}^n \tilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \hat{\beta}_\tau^0, \hat{\beta}_\tau^0)^2 X_{ij}^2$. A similar scale factor is given in Belloni and Chernozhukov (2011). Based on the above discussion, we propose to use

$$\hat{\beta}_\tau^{\text{OSE}} = \arg \min_{\beta \in \Theta} Q_{\tau n}(\beta)$$

as the one-step penalized profile estimator, whose asymptotic property is given below.

Without loss of generality, we assume that the true model has the index parameter $\beta_\tau^0 = \{(\beta_{\tau 1}^0)^T, (\beta_{\tau 2}^0)^T\}^T$, where $\beta_{\tau 1}^0$ is the $p_1 \times 1$ dimensional vector with nonzero components, and $\beta_{\tau 2}^0$ is the $(p - p_1) \times 1$ vector of zeros. Accordingly, let \mathbf{X}_1 be the vector which consists of the first p_1 elements of \mathbf{X} , and $\tilde{\Lambda}_1 = E[f_\varepsilon(0|\mathbf{X}_1)\{G_\tau^{(1)}(\mathbf{X}_1^T \beta_{\tau 1}^0) \mathbf{X}_1\}^{\otimes 2}]$, and $\tilde{\Omega}_1 = E[\{G_\tau^{(1)}(\mathbf{X}_1^T \beta_{\tau 1}^0) \mathbf{X}_1\}^{\otimes 2}]$. We also partition the estimator $\hat{\beta}_\tau^{\text{OSE}}$ into two corresponding parts, $\hat{\beta}_{\tau 1}^{\text{OSE}} \in R^{p_1}$ and $\hat{\beta}_{\tau 2}^{\text{OSE}} \in R^{p-p_1}$.

THEOREM 4. *Under the conditions of Theorem 2, and $n^{-1/2} \ll \lambda_n \ll 1$, we have as $n \rightarrow \infty$, (i) (sparsity) $\hat{\beta}_{\tau 2}^{\text{OSE}} = \mathbf{0}$ with probability approaching 1; and (ii) (asymptotic normality)*

$$\sqrt{n}(\hat{\beta}_{\tau 1}^{\text{OSE}} - \beta_{\tau 1}^0) \rightarrow N(0, \tau(1 - \tau)\tilde{\Lambda}_1^+ \tilde{\Omega}_1 \tilde{\Lambda}_1^+).$$

Note that the one-step estimator $\hat{\beta}_{\tau 1}^{\text{OSE}}$ of $\beta_{\tau 1}^0$ may not achieve the same efficiency obtained under the reduced model with only p_1 covariates. However, due to the selection consistency, we can always fit the model again to estimate the parameters with full asymptotic efficiency after the inactive covariates are removed through the penalty.

4. Implementation. There are tuning parameters to be chosen in the proposed profile estimation method. We take the following steps to balance computational complexity with good accuracy.

For the unpenalized profile estimate, we use the Nelder–Mead algorithm in the R package “optim” to obtain the minimizer of $L_{\tau n}^*(\beta_\tau)$ given in (4), and then normalize the solution to obtain the estimate. We use a finite-difference approximation to calculate the gradient. In the estimation of $\hat{\beta}_\tau$, we simply use $N_n = \lfloor Cn^{1/(2m+1)} \rfloor + 1$ equally space knots for the order m B-splines to facilitate computation, where $C > 0$ is a constant and $\lfloor a \rfloor$ denotes the integer part of a number, so that the estimator of the nonparametric $G_\tau(\cdot)$ achieves the optimal convergence rate. In our empirical work, we use $C = 1$, but the estimates of β_τ are not very sensitive to this choice of C within a reasonable range. In the estimation of G_τ of Section 2.2, we allow the number of internal knots N_n to be chosen to correspond to the first local minimum of

$$\text{BIC}_1(N_n) = \log\{L_{\tau n}(\hat{\theta}_\tau)\} + \frac{\log n}{2n}(N_n + m).$$

In the penalized profile estimation, we choose the tuning parameter λ_n to correspond to the first local minimum of

$$\text{BIC}(\lambda_n) = \log[L_{\tau n}^{**}\{\hat{\beta}_\tau^{\text{OSE}}(\lambda_n)\}] + \frac{\log n}{2n}\hat{p}_1,$$

where \widehat{p}_1 is the number of estimated nonzero parameters in the selected model. A similar strategy is used in Wang, Li and Tsai (2007).

To find the minimizer of (9), we introduce some slack variables, and the problem is then equivalent to minimizing

$$\sum_{i=1}^n [\tau \xi_i + (1 - \tau) \zeta_i] + n \sum_{j=1}^p \omega_j p'_{\lambda_n}(|\widehat{\beta}_{\tau j}^0|)(\beta_j^+ + \beta_j^-)$$

subject to

$$\begin{aligned} \xi_i &\geq 0, & \zeta_i &\geq 0, \\ \xi_i - \zeta_i &= Y_i - \widetilde{G}_{\tau n}^*(\mathbf{X}_i^T(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-), \widehat{\boldsymbol{\beta}}_{\tau}^0), & i &= 1, \dots, n, \\ \beta_j^+ &\geq 0, & \beta_j^- &\geq 0, & j &= 1, \dots, p, \end{aligned}$$

where $\boldsymbol{\beta}^+ = (\beta_1^+, \dots, \beta_p^+)^T$ and $\boldsymbol{\beta}^- = (\beta_1^-, \dots, \beta_p^-)^T$ are the positive and negative parts of the vector $\boldsymbol{\beta}$ componentwise. This is a standard linear program, and our parameter estimate is the normalized value of $\boldsymbol{\beta}$ from the optimization problem.

5. Hypothesis testing. In the quantile regression literature, it is well known that the rank score (RS) test, which is a score test based on the gradient of the quantile objective function, is a powerful test on the parameter $\boldsymbol{\beta}_{\tau}$. It was demonstrated in Kocherginsky, He and Mu (2005), for example, the RS test is simple to use and has a robust performance. Thanks to the availability of a single objective function in the profile approach, we develop a profile RS test in this section. The main difficulty in this development is the lack of differentiability of $\widetilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})$ as a function of $\boldsymbol{\beta}$, and as a result, the rank score of the profile quantile loss function has to be approximated. In this section, we derive an effective approximation to overcome the lack of differentiability.

Without loss of generality, we consider the parameter vector as $\boldsymbol{\beta}_{\tau} = (\boldsymbol{\beta}_{\tau 1}^T, \boldsymbol{\beta}_{\tau 2}^T)^T$, where $\boldsymbol{\beta}_{\tau 1}$ and $\boldsymbol{\beta}_{\tau 2}$ are $p_1 \times 1$ and $(p - p_1) \times 1$ dimensional vectors, respectively. Accordingly, \mathbf{X}_i is partitioned into \mathbf{X}_{1i} and \mathbf{X}_{2i} .

We consider testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_{\tau 2} = \mathbf{0}_{(p-p_1)}$ versus the alternative hypothesis that $\boldsymbol{\beta}_{\tau 2}$ is nonzero. Let $\widehat{\boldsymbol{\beta}}_{\tau}^N$ be the estimator that minimizes $L_{\tau n}^*(\boldsymbol{\beta})$ under \mathcal{H}_0 , that is,

$$\widehat{\boldsymbol{\beta}}_{\tau}^N = \{(\widehat{\boldsymbol{\beta}}_{\tau 1}^N)^T_{p_1 \times 1}, (\widehat{\boldsymbol{\beta}}_{\tau 2}^N)^T_{(p-p_1) \times 1}\}^T,$$

where $\widehat{\boldsymbol{\beta}}_{\tau 2}^N = \mathbf{0}_{(p-p_1) \times 1}$. We now consider a smooth version of the profile quantile loss function

$$\begin{aligned} \widetilde{L}_{\tau n}^*(\boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^n [\rho_{\tau}\{Y_i - \widetilde{G}_{\tau n}(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})\}] \\ &= n^{-1} \sum_{i=1}^n [\rho_{\tau}\{Y_i - B(\mathbf{X}_i^T \boldsymbol{\beta})^T \widetilde{\boldsymbol{\theta}}_{\tau}(\boldsymbol{\beta})\}], \end{aligned}$$

where $\tilde{G}_{\tau n}(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})$ is defined in (5) and its score

$$\begin{aligned} \mathbf{s}_2(\hat{\boldsymbol{\beta}}_\tau^N) &= \{-\partial \tilde{L}_{\tau n}^*(\hat{\boldsymbol{\beta}}_\tau^N) / \partial \boldsymbol{\beta}_2\}_{(p-p_1) \times 1} \\ &= n^{-1} \sum_{i=1}^n \rho_\tau^{(1)}\{Y_i - \tilde{G}_{\tau n}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N)\} \{\partial \tilde{G}_{\tau n}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N) / \partial \boldsymbol{\beta}_2\}, \end{aligned}$$

where $\rho_\tau^{(1)}(u) = \tau I(u \geq 0) + (\tau - 1)I(u < 0)$ and

$$\begin{aligned} \partial \tilde{G}_{\tau n}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N) / \partial \boldsymbol{\beta}_2 \\ = \{\partial B(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)^T / \partial \boldsymbol{\beta}_2\} \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N) + \{\partial \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N)^T / \partial \boldsymbol{\beta}_2\} B(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N), \end{aligned}$$

which is the partial derivative of $\tilde{G}_{\tau n}(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})$ taken with respect to $\boldsymbol{\beta}_2$ with the value evaluated at $\hat{\boldsymbol{\beta}}_\tau^N$, and $\boldsymbol{\beta} = ((\boldsymbol{\beta}_1)_{p_1 \times 1}^T, (\boldsymbol{\beta}_2)_{(p-p_1) \times 1}^T)^T$. Clearly, $\tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N)$ is not a function of the sample, but it is proven in the [Appendix](#) that

$$\begin{aligned} \{\partial \tilde{G}_{\tau n}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N) / \partial \boldsymbol{\beta}_2\} &= G_\tau^{(1)}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N) \tilde{\mathbf{X}}_{2i} + o_p(1) \\ &= \tilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N) \hat{\mathbf{X}}_{2i} + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \tilde{\mathbf{X}}_{ki} &= \mathbf{X}_{ki} - E^*(\mathbf{X}_{ki} | \mathbf{X}_i^T \boldsymbol{\beta}_\tau^0) = \mathbf{X}_{ki} - \frac{E\{f_\varepsilon(0) \mathbf{X}_{ki} | \mathbf{X}_i^T \boldsymbol{\beta}_\tau^0\}}{E\{f_\varepsilon(0) | \mathbf{X}_i^T \boldsymbol{\beta}_\tau^0\}} \\ &= \mathbf{X}_{ki} - E(\mathbf{X}_{ki} | \mathbf{X}_i^T \boldsymbol{\beta}_\tau^0), \end{aligned}$$

under the assumption that $f_\varepsilon(0 | \mathbf{X}) = f_\varepsilon(0)$, and $\hat{\mathbf{X}}_{ki} = \mathbf{X}_{ki} - \hat{E}(\mathbf{X}_{ki} | \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)$ for $k = 1, 2$, where $\tilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N)$ and $\hat{E}(\mathbf{X}_{ki} | \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)$ are spline estimators of $G_\tau^{(1)}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)$ and $E(\mathbf{X}_{ki} | \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)$, respectively, given as

$$\begin{aligned} \hat{E}(\mathbf{X}_{ki} | \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N) &= B(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)^T \left\{ \sum_{i=1}^n B(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N) B(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)^T \right\}^{-1} \\ (10) \quad &\times \left\{ \sum_{i=1}^n B(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N) \mathbf{X}_{ki} \right\}. \end{aligned}$$

Recall that

$$\tilde{G}_{\tau n}^{(1)}(u, \boldsymbol{\beta}) = \partial \tilde{G}_{\tau n}(u, \boldsymbol{\beta}) / \partial u = \{\partial B(u) / \partial u\}^T \tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta}),$$

then

$$\tilde{G}_\tau^{(1)}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N, \hat{\boldsymbol{\beta}}_\tau^N) = B^{(1)}(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}_\tau^N)^T \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N).$$

To approximate the variance–covariance matrix of the score, let $\widehat{g}_{\tau k,i} = \widetilde{G}_{\tau}^{(1)}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_{\tau}^N, \widehat{\boldsymbol{\beta}}_{\tau}^N) \widehat{\mathbf{X}}_{2i}$, $\widehat{g}_{\tau,i} = \widetilde{G}_{\tau}^{(1)}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_{\tau}^N, \widehat{\boldsymbol{\beta}}_{\tau}^N) \widehat{\mathbf{X}}_i$, and $\widehat{\Omega}_{nkl} = n^{-1} \sum_{i=1}^n \widehat{g}_{\tau k,i} \widehat{g}_{\tau l,i}^T$ for $k, l = 1, 2$, and putting them together, let

$$\widehat{\Omega}_n = n^{-1} \sum_{i=1}^n \widehat{g}_{\tau,i} \widehat{g}_{\tau,i}^T = \begin{pmatrix} \widehat{\Omega}_{n11} & \widehat{\Omega}_{n12} \\ \widehat{\Omega}_{n21} & \widehat{\Omega}_{n22} \end{pmatrix}.$$

Finally, we propose an empirical score

$$\widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}_{\tau}^N) = n^{-1} \sum_{i=1}^n \rho_{\tau}^{(1)}\{Y_i - \widetilde{G}_{\tau n}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_{\tau}^N, \widehat{\boldsymbol{\beta}}_{\tau}^N)\} \{\widetilde{G}_{\tau}^{(1)}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_{\tau}^N, \widehat{\boldsymbol{\beta}}_{\tau}^N) \widehat{\mathbf{X}}_{2i}\},$$

and the test statistic

$$(11) \quad \mathcal{T}_n = n\{\tau(1 - \tau)\}^{-1} \{\widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}_{\tau}^N)\}^T \widehat{\Omega}_n^{22} \{\widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}_{\tau}^N)\},$$

where $\widehat{\Omega}_n^{22} = (\widehat{\Omega}_{n22} - \widehat{\Omega}_{n21}^T \widehat{\Omega}_{n11}^+ \widehat{\Omega}_{n12})^+$.

THEOREM 5. *Under the conditions of Theorem 2, if (Y_i, \mathbf{X}_{1i}) and \mathbf{X}_{i2} are independent given $\mathbf{X}_{1i} \boldsymbol{\beta}_{\tau 1}^0$, and $f_{\varepsilon}(0|\mathbf{x}) = f_{\varepsilon}(0)$, we have $\mathcal{T}_n \rightarrow \chi_{p-p_1}^2$.*

REMARK 4. The asymptotic result of Theorem 5 is established under a stronger condition than the null hypothesis of $\boldsymbol{\beta}_{\tau 2} = \mathbf{0}$. In more general settings, the limiting distribution is expected to be a mixture of χ_1^2 distributions. Similar to the rank score test in linear models [Kocherginsky, He and Mu (2005)], we find the rank-based test under the $\chi_{p-p_1}^2$ limiting distribution is quite robust against deviations from the assumptions we have made. Other forms of the tests, such as the Wald tests, rely on the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\tau}$, which is more difficult to estimate with a lower level of reliability in finite-sample problems; see Case 2 of Section S.1 in the supplemental materials [Ma and He (2015)] for a demonstration.

6. Simulation. In this section, we conduct simulation studies to assess the finite-sample performance of the proposed estimation and testing methods. A comparison between the proposed profile optimization method and the backfitting algorithm is also made to show that the backfitting algorithm tends to depend more seriously on the starting values. We also demonstrate that the proposed score test produces reliable results without having to estimate the conditional densities needed in the asymptotic variance–covariance of the quantile estimators. Three test cases are presented below.

6.1. Case 1.

EXAMPLE 1. We consider a sine-bump model, which is similar to a setting considered by Carroll et al. (1997) for mean regression. More specifically, the data

are generated from the model:

$$(12) \quad Y_i = G(\mathbf{X}_i^T \boldsymbol{\beta}^0) + \sigma \varepsilon_i = \frac{\sin\{\pi(\mathbf{X}_i^T \boldsymbol{\beta}^0 - A)\}}{C - A} + \sigma \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $C = \sqrt{3}/2 + 1.645/\sqrt{12}$, $\boldsymbol{\beta}^0 = \frac{1}{\sqrt{14}}(3, 2, 1)^T$, and $\sigma = 0.1$. The components of $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})^T \in R^3$ are independently generated from Uniform (0, 1), with one of the two distributions for ε_i : the Laplace distribution and the t-distribution with 3 degrees of freedoms (t-distr). Under this model, the τ th quantile of Y_i given \mathbf{X}_i is:

$$G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}^0) = G(\mathbf{X}_i^T \boldsymbol{\beta}^0) + \sigma Q(\tau),$$

where $Q(\tau)$ is the τ th quantile of ε_i .

We compare our proposed profiling algorithm with a backfitting iterative method as commonly used in the estimation of single index models. The backfitting algorithm can be described as follows. Let $\widehat{\boldsymbol{\beta}}_\tau^{(k)}$ and $\widehat{\boldsymbol{\theta}}_\tau^{(k)}$ be the estimated values of $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\theta}_\tau$ at the k th step, then the $(k + 1)$ th estimate of $\boldsymbol{\theta}_\tau$, denoted by $\widehat{\boldsymbol{\theta}}_\tau^{(k+1)}$, is obtained by minimizing

$$L_{\tau n}(\widehat{\boldsymbol{\beta}}_\tau^{(k)}, \boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - B(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau^{(k)})^T \boldsymbol{\theta}\}],$$

over $\boldsymbol{\theta}$, and the $(k + 1)$ th estimate of $\boldsymbol{\beta}_\tau$, denoted by $\widehat{\boldsymbol{\beta}}_\tau^{(k+1)}$, is obtained by minimizing

$$L_{\tau n}(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}_\tau^{(k+1)}) = n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - B(\mathbf{X}_i^T \boldsymbol{\beta})^T \widehat{\boldsymbol{\theta}}_\tau^{(k+1)}\}],$$

over $\boldsymbol{\beta}$, and then replacing $\widehat{\boldsymbol{\beta}}_\tau^{(k+1)}$ by its normalized vector $\widehat{\boldsymbol{\beta}}_\tau^{(k+1)} / \|\widehat{\boldsymbol{\beta}}_\tau^{(k+1)}\|_2$. The iteration stops at the $(k + 1)$ th step when $\|\widehat{\boldsymbol{\beta}}_\tau^{(k+1)} - \widehat{\boldsymbol{\beta}}_\tau^{(k)}\|_2 < \epsilon$ and $\|\widehat{\boldsymbol{\theta}}_\tau^{(k+1)} - \widehat{\boldsymbol{\theta}}_\tau^{(k)}\|_2 < \epsilon$ for some small value ϵ (chosen to be $\epsilon = 10^{-6}$ in our study). The number of knots N for the B-spline basis functions is selected in the same way as in the profile estimation.

The backfitting algorithm and the profile optimization are just two routes to minimizing the same objective function. The backfitting approach has been commonly used in the literature, because it involves alternating optimization problems in an easy-to-understand format. The main challenge in our setting comes from the optimization of $L_{\tau n}^*(\boldsymbol{\beta})$, which is a nonconvex function. While the profile estimator needs optimization of a single nonconvex function, the backfitting has to handle nonconvex optimization in each of its iterations. Consequently, as we confirm in our empirical studies, the backfitting algorithm is more sensitive to the choice of starting values and is more likely to reach a local minimum than the profile optimization method proposed in this paper.

We compare the performance of the proposed profile estimator (PR) and the backfitting estimator (BA) by using different initial values and by generating the error term from different distributions at $\tau = 0.75$ and $n = 200$ from 100 simulated data sets. In Figure 1, we give the boxplots of the index coefficient estimates for (a) the Laplace distribution and (b) the t-distribution with the initial value $(1, 1, 1)/\sqrt{3}$, and (c) the Laplace distribution and (d) the t-distribution with the initial value $(1, 3, 6)/\sqrt{46}$. We see that by using the initial value $(1, 1, 1)/\sqrt{3}$ which is closer to the true value, the two estimators lead to similar performance. However, for the initial value $(1, 3, 6)/\sqrt{46}$ which is not close to the true value, the backfitting algorithm generates many outlying points. The results indicate that

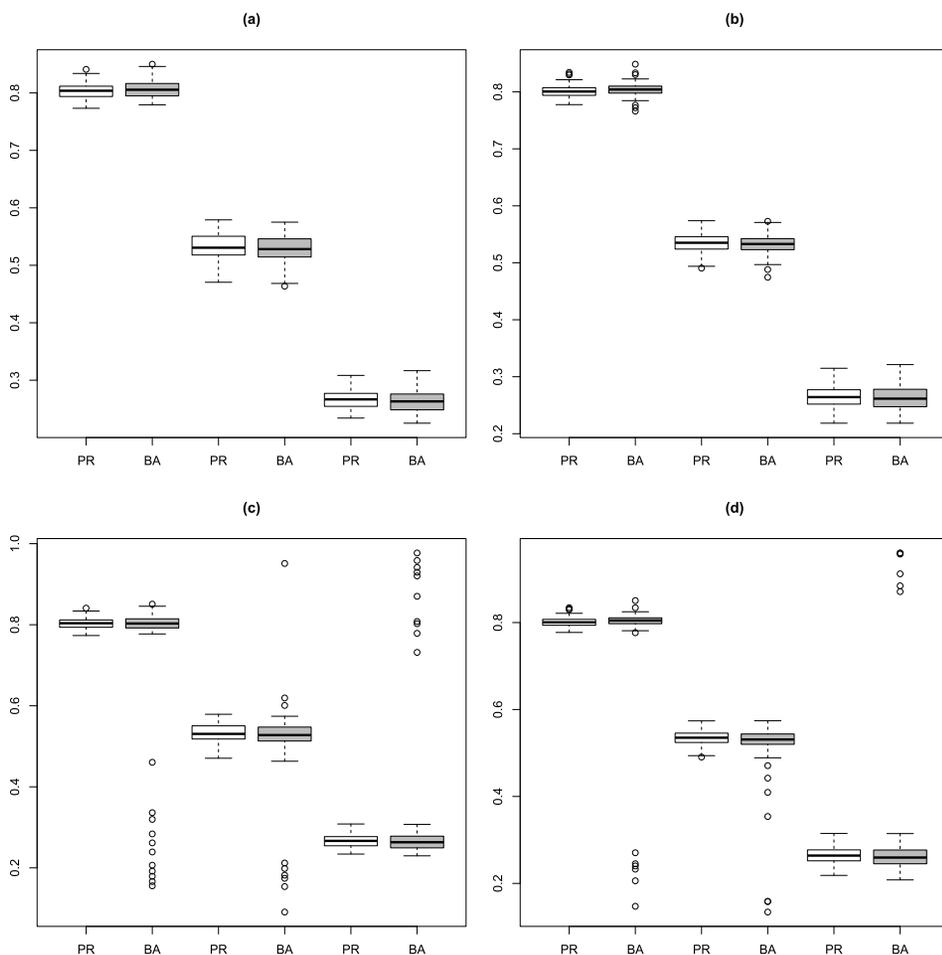


FIG. 1. Boxplot of the index coefficient estimates $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ for (a) the Laplace distribution and (b) the t-distribution with the initial value $(1, 1, 1)/\sqrt{3}$, and (c) the Laplace distribution and (d) the t-distribution with the initial value $(1, 3, 6)/\sqrt{46}$ in Example 1.

the performance of the backfitting method is more sensitive to the choice of initial values.

We conclude this example by reporting the computation time in seconds by using Dell, 3.4 GHz Intel Core with the error generated from the Laplace distribution at $\tau = 0.75$. For estimation at one data set, the computing time for the profile method is 0.342 and 0.384 seconds, respectively, by using initial values $(1, 1, 1)/\sqrt{3}$ and $(1, 3, 6)/\sqrt{46}$, and the computing time for the backfitting method is 0.555 and 0.839 seconds, respectively, by using initial values $(1, 1, 1)/\sqrt{3}$ and $(1, 3, 6)/\sqrt{46}$. We see that given a starting value the profile method needs less computation time in our empirical comparisons.

6.2. *Case 2.* In this example, we study the finite-sample performance of the proposed RS test statistic \widehat{T}_n given in (11). The data are generated from (12) just as in Case 1, except that the components of $\mathbf{X}_i = (X_{i1}, \dots, X_{i7})^T \in R^7$ are independently generated from Uniform $[0, 1]$, and the regression parameter vector takes the form $\boldsymbol{\beta}^0 = (\beta_1^0, \dots, \beta_7^0)^T = \frac{1}{\sqrt{14+4c^2}}(3, 2, c, c, 1, c, c)^T$, where c ranges from 0 to 0.2 with increment 0.02. We consider the null and alternative hypotheses:

$$(13) \quad \begin{aligned} H_0 : & \quad \beta_3 = \beta_4 = \beta_6 = \beta_7 = 0 \\ \text{versus } H_1 : & \quad \beta_j \neq 0, \quad \text{for some } j \in \{3, 4, 6, 7\}. \end{aligned}$$

In addition, 500 realizations were generated with sample size $n = 200$ and $\tau = 0.5$ to estimate the power $= \sum_{i=1}^{500} I(\widehat{T}_{n,i} > \chi_{4,\alpha}^2)/500$ at the significance level $\alpha = 0.05$, where $\widehat{T}_{n,i}$ is the value of the i th replication of \widehat{T}_n , and $\chi_{4,\alpha}^2$ is the $100(1 - \alpha)$ th quantile of χ_4^2 . Figure 2(a) displays the power function versus the c value for the three distributions of ε_i : the standard normal (thick line), Laplace distribution (dashed line) and the t-distribution (thin line). The Type I error rates for the three distributions (the power at $c = 0$) are 0.05, 0.06 and 0.05, respectively, which are close to the nominal significance level 0.05. Moreover, we can also observe that the empirical size of power increases rapidly to 1 as c increases. The results demonstrate that the proposed RS test is a useful test. Next we simulate the data from the model: $Y_i = G(\mathbf{X}_i^T \boldsymbol{\beta}^0) + \{1 + G(\mathbf{X}_i^T \boldsymbol{\beta}^0)/3\} \sigma \varepsilon_i$, where $G(\cdot)$, $\boldsymbol{\beta}^0$, σ , \mathbf{X}_i and ε_i are generated in the same way as described above. In this model, the error terms are dependent of the predictors. We use this model to evaluate the robustness of the RS test when the assumption given in Theorem 5 does not hold. Figure 2(b) displays the power function versus the c value for the three distributions of ε_i . We observe from Figure 2(b) that the RS test is quite robust against heterogeneous errors in the model.

6.3. *Case 3.* In this example, we compare the performance of the proposed one-step LLA estimation procedure with stepwise regression through backward elimination by a simulation study. In the stepwise regression, the RS test proposed

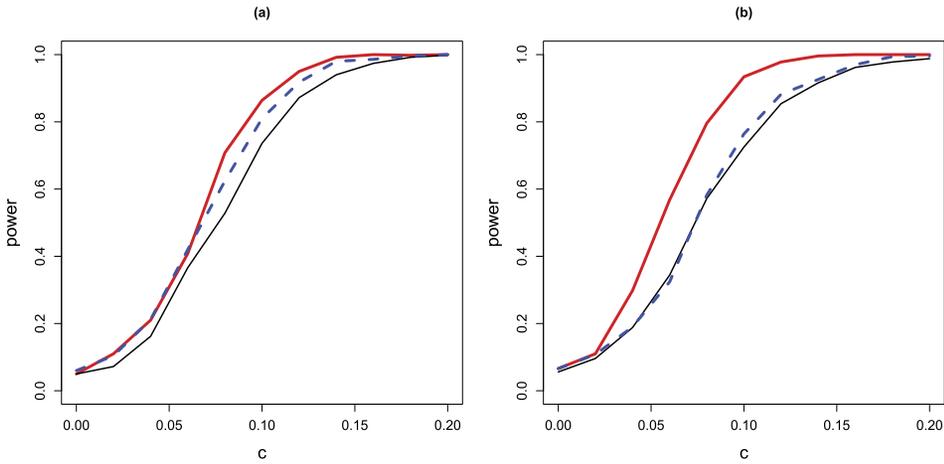


FIG. 2. Plots of power function of the RS test statistic $\widehat{\mathcal{T}}_n$ defined in (11) at significance level 0.05 for the three distributions for ε_i : standard normal (thick line), Laplace distribution (dashed line) and the t-distribution (thin line). Panel (a) is for the i.i.d. error model and Panel (b) is for the heteroscedastic error model.

in Section 5 is used as the criterion for the deletion of each variable at each step with the significance level 0.05. We generate Y_i from a model where the set of nonzero coefficients changes with τ . In this case, we consider

$$(14) \quad G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}(\tau)) = G(\mathbf{X}_i^T \boldsymbol{\beta}(\tau)) + \sigma Q(\tau),$$

where $\boldsymbol{\beta}(\tau) = \{\beta_1(\tau), \dots, \beta_7(\tau)\}^T$ with $\beta_j(\tau) = 0.5(1 + \tau)$ for $j = 1, 2, 3$ and $\beta_j(\tau) = 0$ for $j = 4, 5, 6, 7$, $G(u) = u^2$, and $\sigma = 0.1$. Note that $Q(\cdot)$ is the quantile function of the error term ε_i , which is taken to be the same as given in Case 1, but in model (14), the set of nonzero index coefficients $\beta_j(\tau)$, $j = 1, 2, 3$, changes with τ .

We use the same distribution of \mathbf{X}_i as in Case 2. To generate Y_i from this model, we simply let

$$Y_i = G(\mathbf{X}_i^T \boldsymbol{\beta}(U_i)) + \sigma Q(U_i),$$

where $i = 1, \dots, n$, and $U_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1]$. We consider the quartiles $\tau = 0.25, 0.75$, and simulate 500 data sets with $n = 200$ or 500 in the study. We use equal weights $\widehat{\boldsymbol{\beta}}_\tau^{\text{ini}} = (1, \dots, 1)/\sqrt{7}$ as the starting values. To compare the one-step LLA penalized variable selection procedure (LLA) with the backward elimination procedure (BW), Table 1 shows the proportions of the models correctly fitted (C) (exactly the relevant covariates are selected), overfitted (O) (both the relevant covariates and some irrelevant covariates are selected) and underfitted (U) (some relevant covariates are not selected). The table also reports the average true positives (TP), that is, the average number of selected covariates among the relevant

TABLE 1

Variable selection and estimation results for $\beta(\tau)$ by the one-step LLA procedure and the backward elimination (BW) procedure in Case 3. The columns of C, O and U present the percentage of correct-fitting, over-fitting and under-fitting, respectively. The columns TP and FP report the average number of true and false positives (nonzero coefficients). The columns LLA, PROFILE and ORACLE show the MSEs $\times 100$ of one-step LLA estimator, the profile estimator in the selected model and the oracle estimator

τ	n		C	O	U	TP	FP	LLA	PROFILE	ORACLE	
0.5	200	Normal	LLA	0.850	0.150	0.000	3.000	0.164	2.86	2.35	1.63
			BW	0.594	0.320	0.086	2.910	0.450	–	6.10	1.63
		Laplace	LLA	0.816	0.184	0.000	3.000	0.194	2.85	2.39	1.61
			BW	0.574	0.306	0.120	2.850	0.440	–	8.21	1.61
		t-distr	LLA	0.826	0.174	0.000	3.000	0.184	2.98	2.38	1.69
			BW	0.540	0.320	0.140	2.850	0.450	–	6.10	1.69
	500	Normal	LLA	1.000	0.000	0.000	3.000	0.000	0.69	0.58	0.58
			BW	0.800	0.200	0.000	3.000	0.245	–	0.91	0.58
		Laplace	LLA	1.000	0.000	0.000	3.000	0.000	0.68	0.55	0.55
			BW	0.774	0.226	0.000	3.000	0.265	–	0.90	0.55
		t-distr	LLA	1.000	0.000	0.000	3.000	0.000	0.72	0.59	0.59
			BW	0.740	0.260	0.000	3.000	0.305	–	0.98	0.59
0.75	200	Normal	LLA	0.944	0.056	0.000	3.000	0.066	1.18	1.06	0.89
			BW	0.700	0.210	0.090	2.900	0.290	–	4.80	0.89
		Laplace	LLA	0.920	0.080	0.000	3.000	0.100	1.38	1.26	0.98
			BW	0.684	0.206	0.110	2.876	0.280	–	6.21	0.98
		t-distr	LLA	0.916	0.084	0.000	3.000	0.104	1.38	1.26	0.98
			BW	0.614	0.226	0.160	2.820	0.334	–	8.53	0.98
	500	Normal	LLA	1.000	0.000	0.000	3.000	0.000	0.43	0.40	0.38
			BW	0.790	0.210	0.000	3.000	0.240	–	0.54	0.38
		Laplace	LLA	1.000	0.000	0.000	3.000	0.000	0.47	0.43	0.42
			BW	0.804	0.196	0.000	3.000	0.210	–	0.56	0.42
		t-distr	LLA	1.000	0.000	0.000	3.000	0.000	0.45	0.42	0.42
			BW	0.764	0.236	0.000	3.000	0.265	–	0.59	0.42

covariates, and the average false positives (FP), that is, the average number of selected covariates among the irrelevant covariates. The last three columns of Table 1 reports the estimated MSE for the LLA estimator, the profile estimator in the re-fitted model after variable selection and the oracle estimator by assuming the true model is known, denoted by LLA, PROFILE and ORACLE, respectively, by both procedures.

By the LLA penalized method, we observe that the percentages of correct-fitting get close to 1 as n increases. The same is true with the backward elimination approach but the increase is clearly slower. The MSEs of the LLA estimates are slightly higher than the PROFILE estimates under the selected model, and of

course, they are not as good as the ORACLE measures unless the correct models are identified with very high probability.

The proposed estimation procedure is computationally faster than backward elimination. We ran the above simulation experiments on Macbook Pro with 2 GHz Intel Core. At $\tau = 0.75$ and when the data are generated under Gaussian errors, we obtained that the average operation time per simulated data set in R by the one-step LLA procedure is 4.54 and 12.19 seconds for $n = 200$ and 500, respectively, but the average time by the backward elimination procedure is 30.74 and 75.15 for $n = 200$ and 500, respectively.

7. A data analysis example. In this section, we illustrate the proposed method by analyzing the CD4 cell count change in the ACTG320 study. The data come from a double-blind, placebo-controlled trial that compared the three-drug regimen (treatment) with the two-drug regimen (control) in HIV-infected patients [Hammer et al. (1997)]. In our analysis, we take the response variable $Y =$ as the CD4 count change at week 24 (CD4.24), and the covariates are $X_1 = \log(\text{CD4.0})$ (logarithm of the baseline CD4 cell counts CD4.0 at week 0); $X_2 = \log(\text{RNA.0})$ (logarithm of baseline RNA concentration at week 0); $X_3 = \text{trt}$ (binary treatment indicator, $\text{trt} = 1$ for the treatment group and $\text{trt} = 0$ for control group); $X_4 = \text{age}$; and $X_5 = \text{gender}$ (binary variable, $\text{gender} = 0$ for male and $\text{gender} = 1$ for female). We also consider the interaction effect of $\log(\text{CD4.0})$ and trt , with $X_6 = \text{trt} \times \log(\text{CD4.0})$. By removing the observations with missing values in these variables and dropping one outlier with $\text{CD4.0} = 0$, we have 855 observations in our study. We use centered and standardized values of all predictors for model fitting. We fit linear quantile regression and let the normalized estimates of the parameters be the initial estimates in our unpenalized profile estimation procedure.

Table 2 shows the estimated coefficients (Estimate) and the corresponding p -value (Pvalue) from the score test at two quantile levels $\tau = 0.5, 0.75$. We observe that the estimated coefficients of $\log(\text{CD4.0})$, $\log(\text{RNA.0})$, trt and $\text{trt} \times \log(\text{CD4.0})$ are significantly different from zero at significance level 0.05 at $\tau = 0.5$, while $\log(\text{RNA.0})$ becomes insignificant at $\tau = 0.75$. The significance of the interaction term indicates that with the treatment the baseline CD4 count has differential impact on the CD4 change over the 24-week period. Next, we

TABLE 2
The estimated coefficients at quantiles $\tau = 0.5, 0.75$ for the real data application

τ		$\log(\text{CD4.0})$	$\log(\text{RNA.0})$	trt	age	Gender	$\text{trt} \times \log(\text{CD4.0})$
0.50	Estimate	0.171	0.154	1.855	-0.001	0.049	-0.134
	Pvalue	<0.001	0.002	<0.001	0.382	0.071	<0.001
0.75	Estimate	0.434	0.016	1.678	0.011	0.161	0.066
	Pvalue	<0.001	0.357	<0.001	0.389	0.056	0.034

TABLE 3
Variable selection results and the estimated coefficients for the selected variables at quantiles $\tau = 0.5, 0.75$ by the LLA method for the real data application

τ	log(CD4.0)	log(RNA.0)	trt	Age	Gender	trt x log(CD4.0)
0.50	0.164	0.092	1.857	–	–	–0.142
0.75	0.429	–	1.726	–	–	–

perform the one-step LLA penalized variable selection procedure (LLA) with the tuning parameter selected by the BIC given in Section 4. Table 3 shows the variable selection results and the estimated coefficients for the selected variables at $\tau = 0.5, 0.75$. We see that the variable selection results are quite consistent with the score test results given in Table 2.

To better understand the impact of baseline CD4 on the response, in Figure 3 we plot the estimated quantile function $\widehat{G}_{\tau_n}(\cdot)$ and the response value (dots) versus log(CD4.0) at the sample means of log(RNA.0) and age for trt = 0 (thin line), trt = 1 (thick line) and gender = 0 (left panel), gender = 1 (right panel). We see that at $\tau = 0.5$, for both genders the estimated response value increases steadily with log(CD4.0) linearly for the treatment group, which indicates that the treatment takes similar effects on those patients with either a small or large number of CD4 cell counts at week 0. However, for the control group, the estimated response value stays around zero for small values of log(CD4.0), but it shows an increasing trend for large values of log(CD4.0). This indicates that when patient's CD4 cell counts at week 0 are below certain level, they show no change over time. When CD4 cell counts at week 0 are large, they still increase over time even without treatment, and the increasing rate is larger for patients with a larger number of CD4 cell counts at week 0. At $\tau = 0.75$, the estimated quantile function $\widehat{G}_{\tau_n}(\cdot)$ has a similar pattern as at $\tau = 0.5$ for the control group. However, for the treatment group, instead of showing a linear pattern, the response value seems to remain at a certain level for either large value of log(CD4.0) (greater than 4) or small value of log(CD4.0) (less than 1), but it shows an increasing trend when the value of log(CD4.0) is between 1 and 4. These findings on the treatment effects would be missed by a linear quantile model.

8. Discussion. In this paper, we establish the consistency and asymptotic normality of the profile estimation procedure in single-index quantile regression models. The asymptotic distributions of such estimates are the same as those obtained from other iterative methods of estimation, but the profile approach has better stability and is less sensitive to initial values of the index parameters. Moreover, the availability of a single objective function of the index parameter allows the devel-

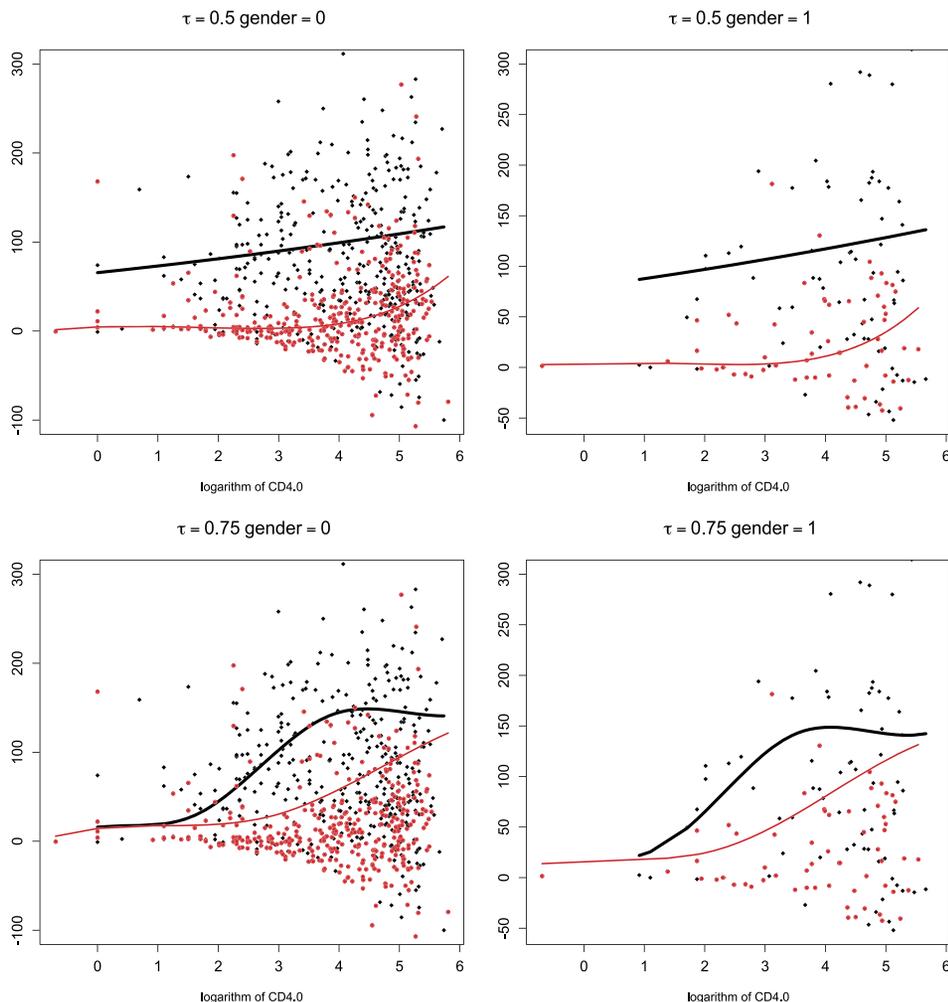


FIG. 3. Plots of the estimated quantile function $\widehat{G}_{\tau_n}(\cdot)$ and the response value (dots) versus $\log(\text{CD4.0})$ at the sample means of $\log(\text{RNA.0})$ and age for $\text{trt} = 0$ (thin line, red color), $\text{trt} = 1$ (thick line, black color) and gender = 0 (left panel), gender = 1 (right panel).

opment of a robust score test for inference on those parameters, and enables the use of penalized optimization for model selection. Both the theoretical and empirical works in the paper show that the profile pseudo-likelihood approach is valuable to estimation and inference for single-index quantile regression models.

The asymptotic theory established for the penalized estimation in this paper assumes that \mathbf{X} has a fixed dimension as n increases. The problem of handling both nonparametric estimation of G_τ and a linear index with a growing number of variables poses some challenges, and additional work is needed to investigate how the profile method works when the number of variables increases with n .

APPENDIX

For any positive numbers a_n and b_n , let $a_n \asymp b_n$ denote $\lim_{n \rightarrow \infty} a_n/b_n = c$, for a positive constant c , and $a_n \sim b_n$ denote $\lim_{n \rightarrow \infty} a_n/b_n = 1$. For any vector $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_s)^T \in R^s$, denote $|\boldsymbol{\zeta}| = \max_{1 \leq l \leq s} |\zeta_l|$ and $\|\boldsymbol{\zeta}\|_r = (\sum_{l=1}^s |\zeta_l|^r)^{1/r}$. For any matrix $\mathbf{A} = (A_{ij})_{i=1, j=1}^{s, t}$, denote $|\mathbf{A}| = \max_{1 \leq i \leq s, 1 \leq j \leq t} |A_{ij}|$ and $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq s} \sum_{j=1}^t |A_{ij}|$. For any symmetric matrix $\mathbf{A}_{s \times s}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\boldsymbol{\zeta} \in s, \boldsymbol{\zeta} \neq \mathbf{0}} \|\mathbf{A}\boldsymbol{\zeta}\|_r \|\boldsymbol{\zeta}\|_r^{-1}$.

A.1. Proof of Theorem 1. By the definition of $\widehat{\boldsymbol{\beta}}_\tau$, one has $P\{L_{\tau n}^*(\widehat{\boldsymbol{\beta}}_\tau) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)\} = 1$. For any open set $\mathcal{S}(\boldsymbol{\beta}_\tau^0)$ that includes $\boldsymbol{\beta}_\tau^0$, one has

$$\begin{aligned} &P\{L_{\tau n}^*(\widehat{\boldsymbol{\beta}}_\tau) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)\} \\ &= P\{L_{\tau n}^*(\widehat{\boldsymbol{\beta}}_\tau) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0) \text{ and } \widehat{\boldsymbol{\beta}}_\tau \in \mathcal{S}(\boldsymbol{\beta}_\tau^0)\} \\ &\quad + P\{L_{\tau n}^*(\widehat{\boldsymbol{\beta}}_\tau) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0) \text{ and } \widehat{\boldsymbol{\beta}}_\tau \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)\} \\ &\leq P\{\widehat{\boldsymbol{\beta}}_\tau \in \mathcal{S}(\boldsymbol{\beta}_\tau^0)\} + P\left\{\inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} L_{\tau n}^*(\boldsymbol{\beta}) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)\right\}. \end{aligned}$$

Next, we will show that $P\{\inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} L_{\tau n}^*(\boldsymbol{\beta}) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)\} \rightarrow 0$, which implies that $\widehat{\boldsymbol{\beta}}_\tau$ must be in any open set $\mathcal{S}(\boldsymbol{\beta}_\tau^0)$ so that $\widehat{\boldsymbol{\beta}}_\tau$ is a consistent estimator of $\boldsymbol{\beta}_\tau^0$ with probability approaching 1. Define

$$\widetilde{L}_\tau^*(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - \widetilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})\}].$$

Note that

$$\begin{aligned} &P\left\{\inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} L_{\tau n}^*(\boldsymbol{\beta}) \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)\right\} \\ &= P\left\{\inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} \{L_{\tau n}^*(\boldsymbol{\beta}) - \widetilde{L}_\tau^*(\boldsymbol{\beta})\right. \\ &\quad \left. + \widetilde{L}_\tau^*(\boldsymbol{\beta}) - L_\tau^*(\boldsymbol{\beta}) + L_\tau^*(\boldsymbol{\beta})\} \leq L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)\right\} \\ &\leq P\left[\inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} \{L_{\tau n}^*(\boldsymbol{\beta}) - \widetilde{L}_\tau^*(\boldsymbol{\beta})\} + \inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} \{\widetilde{L}_\tau^*(\boldsymbol{\beta}) - L_\tau^*(\boldsymbol{\beta})\} \right. \\ &\quad \left. + L_\tau^*(\boldsymbol{\beta}_\tau^0) - L_{\tau n}^*(\boldsymbol{\beta}_\tau^0) \leq L_\tau^*(\boldsymbol{\beta}_\tau^0) - \inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} L_\tau^*(\boldsymbol{\beta})\right] \\ &\leq P\left[\sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} |L_{\tau n}^*(\boldsymbol{\beta}) - \widetilde{L}_\tau^*(\boldsymbol{\beta})| + \sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} |\widetilde{L}_\tau^*(\boldsymbol{\beta}) - L_\tau^*(\boldsymbol{\beta})| \right. \\ &\quad \left. + |L_\tau^*(\boldsymbol{\beta}_\tau^0) - L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)| \geq \inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} L_\tau^*(\boldsymbol{\beta}) - L_\tau^*(\boldsymbol{\beta}_\tau^0)\right], \end{aligned}$$

where $L_\tau^*(\boldsymbol{\beta}) = E[\rho_\tau\{Y - \tilde{G}_\tau(\mathbf{X}^T\boldsymbol{\beta}, \boldsymbol{\beta})\}]$. Since $\boldsymbol{\beta}_\tau^0$ is the unique minimizer of $L_\tau^*(\boldsymbol{\beta})$, for every open set $\mathcal{S}(\boldsymbol{\beta}_\tau^0)$, there exists $\varepsilon > 0$, such that $\inf_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} L_\tau^*(\boldsymbol{\beta}) - L_\tau^*(\boldsymbol{\beta}_\tau^0) > \varepsilon$. Thus, it is sufficient to show that for every $\varepsilon > 0$,

$$(A.1) \quad P\left\{ \sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} |L_{\tau n}^*(\boldsymbol{\beta}) - \tilde{L}_\tau^*(\boldsymbol{\beta})| > \varepsilon \right\} \rightarrow 0,$$

$$(A.2) \quad P\left\{ \sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} |\tilde{L}_\tau^*(\boldsymbol{\beta}) - L_\tau^*(\boldsymbol{\beta})| > \varepsilon \right\} \rightarrow 0,$$

$$(A.3) \quad P\{|L_\tau^*(\boldsymbol{\beta}_\tau^0) - L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)| > \varepsilon\} \rightarrow 0.$$

To verify (A.1), note

$$\begin{aligned} & \sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} |L_{\tau n}^*(\boldsymbol{\beta}) - \tilde{L}_\tau^*(\boldsymbol{\beta})| \\ &= \sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} \left| n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - \tilde{G}_{\tau n}(\mathbf{X}_i^T\boldsymbol{\beta}, \boldsymbol{\beta})\}] \right. \\ & \quad \left. - [\rho_\tau\{Y_i - \tilde{G}_\tau(\mathbf{X}_i^T\boldsymbol{\beta}, \boldsymbol{\beta})\}] \right| \\ &\leq \sup_{\boldsymbol{\beta} \in \Theta \setminus \mathcal{S}(\boldsymbol{\beta}_\tau^0)} n^{-1} \sum_{i=1}^n |\tilde{G}_{\tau n}(\mathbf{X}_i^T\boldsymbol{\beta}, \boldsymbol{\beta}) - \tilde{G}_\tau(\mathbf{X}_i^T\boldsymbol{\beta}, \boldsymbol{\beta})| = o_p(1), \end{aligned}$$

which follows from Lemma S.3. Thus, (A.1) is proved. (A.2) can be proved by the uniform consistency theorem given in Andrews (1987). Moreover, by Lemma S.3, we have $|L_{\tau n}^*(\boldsymbol{\beta}_\tau^0) - \tilde{L}_\tau^*(\boldsymbol{\beta}_\tau^0)| = o_p(1)$ and by the weak law of large numbers, we have $|\tilde{L}_\tau^*(\boldsymbol{\beta}_\tau^0) - L_\tau^*(\boldsymbol{\beta}_\tau^0)| = o_p(1)$. Hence, (A.3) follows from the fact that

$$|L_\tau^*(\boldsymbol{\beta}_\tau^0) - L_{\tau n}^*(\boldsymbol{\beta}_\tau^0)| \leq |L_{\tau n}^*(\boldsymbol{\beta}_\tau^0) - \tilde{L}_\tau^*(\boldsymbol{\beta}_\tau^0)| + |\tilde{L}_\tau^*(\boldsymbol{\beta}_\tau^0) - L_\tau^*(\boldsymbol{\beta}_\tau^0)| = o_p(1).$$

A.2. Proof of Theorem 2. Let

$$\tilde{L}_{\tau n}^*(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n [\rho_\tau\{Y_i - B(\mathbf{X}_i^T\boldsymbol{\beta})^T \tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta})\}],$$

where $\tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta})$ is the minimizer of $E\{L_{\tau n}(\boldsymbol{\theta}, \boldsymbol{\beta})|\mathbb{X}\}$. For any $\mathbf{b} \in R^p$, define

$$\begin{aligned} \tilde{D}_{\tau n,i}(\mathbf{b}) &= \rho_\tau\{Y_i - B(\mathbf{X}_i^T(\boldsymbol{\beta}_\tau^0 + \mathbf{b}))^T \tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta}_\tau^0 + \mathbf{b})\} \\ & \quad - \rho_\tau\{Y_i - B(\mathbf{X}_i^T\boldsymbol{\beta}_\tau^0)^T \tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta}_\tau^0)\}, \\ (A.4) \quad D_{\tau n,i}(\mathbf{b}) &= \rho_\tau\{Y_i - B(\mathbf{X}_i^T(\boldsymbol{\beta}_\tau^0 + \mathbf{b}))^T \tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta}_\tau^0 + \mathbf{b})\} \\ & \quad - \rho_\tau\{Y_i - B(\mathbf{X}_i^T\boldsymbol{\beta}_\tau^0)^T \tilde{\boldsymbol{\theta}}_\tau(\boldsymbol{\beta}_\tau^0)\}, \\ \tilde{D}_{\tau n}(\mathbf{b}) &= n^{-1} \sum_{i=1}^n \tilde{D}_{\tau n,i}(\mathbf{b}), \quad D_{\tau n}(\mathbf{b}) = n^{-1} \sum_{i=1}^n D_{\tau n,i}(\mathbf{b}). \end{aligned}$$

Let $\widehat{\beta}_\tau$ be a minimizer of $L_{\tau n}^*(\beta)$ over $\beta \in R^p$ for $L_{\tau n}^*(\beta)$ given in (4). It is helpful to note that for any β in R^p , we can define the B-splines on the scaled interval $[a\|\beta\|_2, b\|\beta\|_2]$, but the value of the spline at \mathbf{x} remains invariant over the norm of β used. Hence, we have $L_{\tau n}^*(\beta) = L_{\tau n}^*(\beta/\|\beta\|_2)$. Therefore, $\widehat{\beta}_\tau$ is determined only in its direction, and by taking $\widehat{\beta}_\tau = \widehat{\beta}_\tau/\|\widehat{\beta}_\tau\|_2$, we have $\widehat{\beta}_\tau$ as the minimizer of $L_{\tau n}^*(\beta)$ over $\beta \in \Theta$. Moreover, $L_{\tau n}^*(\widehat{\beta}_\tau) = L_{\tau n}^*(\widehat{\beta}_\tau)$ which implies $L_{\tau n}^*(\widehat{\beta}_\tau) \leq L_{\tau n}^*(\beta)$ for all $\beta \in R^p$. We will show that $\sqrt{n}\|\widehat{\beta}_\tau - \beta_\tau^0\|_2 = O_p(1)$ and $\sqrt{n}(\widehat{\beta}_\tau - \beta_\tau^0)$ is asymptotically normal in the following two steps.

Step 1. We can decompose $\widehat{\beta}_\tau$ as $\widehat{\beta}_\tau = a_{\tau n}\beta_\tau^0 + s_{\tau n}^*\eta$, where η is a unit vector and orthogonal to β_τ^0 , so that $\|\widehat{\beta}_\tau\|_2 = a_{\tau n}^2 + s_{\tau n}^{*2} = 1$. Note that because $\widehat{\beta}_\tau$ is a consistent estimator of β_τ^0 , we have $|a_{\tau n} - 1| = o_p(1)$, $s_{\tau n}^* = o_p(1)$ and $a_{\tau n} = \sqrt{1 - s_{\tau n}^{*2}}$. We can further write $a_{\tau n}^{-1}\widehat{\beta}_\tau = \beta_\tau^0 + s_{\tau n}\eta$, where $s_{\tau n} = a_{\tau n}^{-1}s_{\tau n}^*$. We will show that $s_{\tau n} = O_p(n^{-1/2})$ which implies that $s_{\tau n}^* = O_p(n^{-1/2})$ and

$$\begin{aligned}
 1 - a_{\tau n} &= 1 - \sqrt{1 - s_{\tau n}^{*2}} \\
 (A.5) \quad &= \{1 - (1 - s_{\tau n}^{*2})\} / \{1 + \sqrt{1 - s_{\tau n}^{*2}}\} \\
 &= O(s_{\tau n}^{*2}) = O_p(n^{-1}),
 \end{aligned}$$

and then $\|\widehat{\beta}_\tau - \beta_\tau^0\|_2 = O_p(n^{-1/2})$. In what follows, we take $\widehat{\beta}_\tau = a_{\tau n}^{-1}\widehat{\beta}_\tau$. Clearly, $\widehat{\beta}_\tau$ is a minimizer of $L_{\tau n}^*(\beta)$ over $\beta \in R^p$. Let $\widehat{\mathbf{b}}_{\tau n} = s_{\tau n}\eta = \widehat{\beta}_\tau - \beta_\tau^0$. Then $\widehat{\mathbf{b}}_{\tau n}$ minimizes the function $D_{\tau n}(\mathbf{b}) = L_{\tau n}^*(\beta_\tau^0 + \mathbf{b}) - L_{\tau n}^*(\beta_\tau^0)$.

Denote

$$\begin{aligned}
 \Gamma_n &= -n^{-1} \sum_{i=1}^n \{\tau - I(\varepsilon_i < 0)\} G_\tau^{(1)}(\mathbf{X}_i^T \beta_\tau^0) \widetilde{\mathbf{X}}_i, \\
 \Lambda_n &= n^{-1} \sum_{i=1}^n f_\varepsilon(0|\mathbf{X}_i) \{G_\tau^{(1)}(\mathbf{X}_i^T \beta_\tau^0)\}^2 \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^T,
 \end{aligned}$$

and $c_n(\mathbf{b}) = (1 + \sqrt{n}\|\mathbf{b}\|_2)^{-1} \sqrt{n}\mathbf{b}$. Applying the weak law of large numbers to $c_n^T(\widehat{\mathbf{b}}_{\tau n})\sqrt{n}\Gamma_n$,

$$\begin{aligned}
 (A.6) \quad c_n^T(\widehat{\mathbf{b}}_{\tau n})\sqrt{n}\Gamma_n &= O_p\{\{c_n^T(\widehat{\mathbf{b}}_{\tau n})E(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T)c_n(\widehat{\mathbf{b}}_{\tau n})\}^{1/2}\} \\
 &= O_p\{\|c_n(\widehat{\mathbf{b}}_{\tau n})\|_2\},
 \end{aligned}$$

where the last equality follows from that the eigenvalues of $E(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T)$ are bounded from infinity by the fact that \mathbf{X} is distributed on a compact set. By (S.17) in

Lemma S.6 and (S.32) in Lemma S.7, we have

$$\begin{aligned}
 D_{\tau n}(\widehat{\mathbf{b}}_{\tau n}) &= -n^{-1} \sum_{i=1}^n \{\tau - I(\varepsilon_i < 0)\} G_{\tau}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_{\tau}^0) \widetilde{\mathbf{X}}_i^T \widehat{\mathbf{b}}_{\tau n} \\
 &\quad + n^{-1} \sum_{i=1}^n 2^{-1} f_{\varepsilon}(0|\mathbf{X}_i) \{G_{\tau}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_{\tau}^0)\}^2 \widehat{\mathbf{b}}_{\tau n}^T \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^T \widehat{\mathbf{b}}_{\tau n} \\
 &\quad + o_p(n^{-1/2} \|\widehat{\mathbf{b}}_{\tau n}\|_2) + O_p((\log n) J_n n^{-1} \|\widehat{\mathbf{b}}_{\tau n}\|_2^{1/2}) + o_p(n^{-1}).
 \end{aligned}$$

By the fact that $2ab \leq a^2 + b^2$, we have $2n^{-1/2} \|\widehat{\mathbf{b}}_{\tau n}\|_2 \leq n^{-1} + \|\widehat{\mathbf{b}}_{\tau n}\|_2^2$ and

$$\begin{aligned}
 &2J_n n^{-1} (\log n) \|\widehat{\mathbf{b}}_{\tau n}\|_2^{1/2} \\
 &\leq \|\widehat{\mathbf{b}}_{\tau n}\|_2 J_n^2 n^{-1} (\log n)^{2+2/10} + n^{-1} (\log n)^{-2/10} \\
 &\leq 0.5 \|\widehat{\mathbf{b}}_{\tau n}\|_2^2 J_n^4 n^{-1} (\log n)^{4+6/10} + 0.5 n^{-1} (\log n)^{-2/10} + n^{-1} (\log n)^{-2/10} \\
 &= o(1) \|\widehat{\mathbf{b}}_{\tau n}\|_2^2 + o(n^{-1}),
 \end{aligned}$$

by the assumption $J_n \ll n^{1/4}/(\log n)^{5/4}$. Then

$$\begin{aligned}
 D_{\tau n}(\widehat{\mathbf{b}}_{\tau n}) &= -n^{-1} \sum_{i=1}^n \{\tau - I(\varepsilon_i < 0)\} G_{\tau}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_{\tau}^0) \widetilde{\mathbf{X}}_i^T \widehat{\mathbf{b}}_{\tau n} \\
 \text{(A.7)} \quad &\quad + n^{-1} \sum_{i=1}^n 2^{-1} f_{\varepsilon}(0|\mathbf{X}_i) \{G_{\tau}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_{\tau}^0)\}^2 \widehat{\mathbf{b}}_{\tau n}^T \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^T \widehat{\mathbf{b}}_{\tau n} \\
 &\quad + o_p(1) \|\widehat{\mathbf{b}}_{\tau n}\|_2^2 + o_p(n^{-1}).
 \end{aligned}$$

By the definition of $\widehat{\mathbf{b}}_{\tau n}$, we have $D_{\tau n}(\widehat{\mathbf{b}}_{\tau n}) \leq 0$. Multiplying both sides of (A.7) by $n(1 + \sqrt{n} \|\widehat{\mathbf{b}}_{\tau n}\|_2)^{-2}$, we have

$$\begin{aligned}
 \text{(A.8)} \quad &c_n^T(\widehat{\mathbf{b}}_{\tau n}) \sqrt{n} \Gamma_n (1 + \sqrt{n} \|\widehat{\mathbf{b}}_{\tau n}\|_2)^{-1} + 2^{-1} c_n^T(\widehat{\mathbf{b}}_{\tau n}) \Lambda_n c_n(\widehat{\mathbf{b}}_{\tau n}) \\
 &\quad + o_p(1) + o_p(1) (1 + \sqrt{n} \|\widehat{\mathbf{b}}_{\tau n}\|_2)^{-2} \leq 0.
 \end{aligned}$$

Let n_k be a subsequence such that $\sqrt{n_k} s_{\tau, n_k} \rightarrow \infty$, then we have $\|c_{n_k}(\widehat{\mathbf{b}}_{\tau, n_k})\|_2 \asymp 1$ so that

$$c_{n_k}^T(\widehat{\mathbf{b}}_{\tau, n_k}) \sqrt{n_k} \Gamma_{n_k} (1 + \sqrt{n_k} \|\widehat{\mathbf{b}}_{\tau, n_k}\|_2)^{-1} = o_p(1)$$

by (A.6). This result and (A.8) imply that $c_{n_k}^T(\widehat{\mathbf{b}}_{\tau, n_k}) \Lambda_{n_k} c_{n_k}(\widehat{\mathbf{b}}_{\tau, n_k}) = o_p(1)$, and thus

$$c_{n_k}^T(\widehat{\mathbf{b}}_{\tau, n_k}) \Lambda_{n_k} c_{n_k}(\widehat{\mathbf{b}}_{\tau, n_k}) = o_p(1)$$

by Conditions (C2) and (C3). Recall that $\widehat{\mathbf{b}}_{\tau n} = s_{\tau n} \boldsymbol{\eta}$. On the other hand,

$$\begin{aligned} c_{n_k}^T(\widehat{\mathbf{b}}_{\tau, n_k}) \Lambda c_{n_k}(\widehat{\mathbf{b}}_{\tau, n_k}) &= (1 + \sqrt{n_k} |s_{\tau, n_k}|)^{-2} (\sqrt{n_k} s_{\tau, n_k})^2 \boldsymbol{\eta}^T \Lambda \boldsymbol{\eta} > 0 \end{aligned}$$

for any unit vector $\boldsymbol{\eta} \in R^p$ orthogonal to $\boldsymbol{\beta}_\tau^0$. Then we have a contradiction. Thus, $\sqrt{n} |s_{\tau n}| = O_p(1)$.

Step 2. By (A.7) and $\|\widehat{\mathbf{b}}_{\tau n}\|_2 = O_p(n^{-1/2})$, we can write $D_{\tau n}(\widehat{\mathbf{b}}_{\tau n})$ as

$$\begin{aligned} D_{\tau n}(\widehat{\mathbf{b}}_{\tau n}) &= \widehat{\mathbf{b}}_{\tau n}^T \Gamma_n + 2^{-1} \widehat{\mathbf{b}}_{\tau n}^T (\Lambda_n + \varrho_n \mathbf{I}) \widehat{\mathbf{b}}_{\tau n} - 2^{-1} \varrho_n \|\widehat{\mathbf{b}}_{\tau n}\|_2^2 \\ &\quad + o_p(1) \|\widehat{\mathbf{b}}_{\tau n}\|_2^2 + o_p(n^{-1}) \\ &= \widehat{\mathbf{b}}_{\tau n}^T \Gamma_n + 2^{-1} \widehat{\mathbf{b}}_{\tau n}^T (\Lambda_n + \varrho_n \mathbf{I}) \widehat{\mathbf{b}}_{\tau n} + o_p(n^{-1}) \end{aligned}$$

for any $\varrho_n = o(1)$ and $\varrho_n \notin \sigma(\Lambda_n)$, where $\sigma(\Lambda_n)$ is the spectrum of a square matrix Λ_n , so that $(\Lambda_n + \varrho_n \mathbf{I})^{-1}$ exists. Define

$$D_{\tau n}^*(\mathbf{b}) = \mathbf{b}^T \Gamma_n + 2^{-1} \mathbf{b}^T (\Lambda_n + \varrho_n \mathbf{I}) \mathbf{b}.$$

Then $\widetilde{\mathbf{b}}_{\tau n} = -(\Lambda_n + \varrho_n \mathbf{I})^{-1} \Gamma_n$ minimizes $D_{\tau n}^*(\mathbf{b})$. Moreover, $\|\widetilde{\mathbf{b}}_{\tau n}\|_2 = O_p(n^{-1/2})$. Let $\alpha_{1n}, \dots, \alpha_{rn}$ be distinct eigenvalues of Λ_n , which are self-adjoint. Then we can write $\Lambda_n = \sum_{i=1}^r \alpha_{in} S_{in}$, where S_{in} are the spectral projectors of Λ_n . Therefore, we have $\Lambda_n + \varrho_n \mathbf{I} = \sum_{i=1}^r (\alpha_{in} + \varrho_n) S_{in}$, and thus $(\Lambda_n + \varrho_n \mathbf{I})^{-1} = \sum_{i=1}^r (\alpha_{in} + \varrho_n)^{-1} S_{in}$. Hence,

$$\lim_{n \rightarrow \infty} (\Lambda_n + \varrho_n \mathbf{I})^{-1} = \lim_{n \rightarrow \infty} \sum_{i=1}^r \alpha_{in}^{-1} S_{in} = \lim_{n \rightarrow \infty} \Lambda_n^+ = \Lambda^+.$$

Therefore, by the central limit theorem, as $n \rightarrow \infty$,

$$(A.9) \quad \sqrt{n} \widetilde{\mathbf{b}}_{\tau n} \rightarrow N(0, \tau(1 - \tau) \Lambda^+ \Omega \Lambda^+).$$

Recall that $a_{\tau n}^{-1} \widehat{\boldsymbol{\beta}}_\tau = \boldsymbol{\beta}_\tau^0 + \widehat{\mathbf{b}}_{\tau n}$ and $1 - a_{\tau n} = O_p(n^{-1})$ given in (A.5). Next, we will show that

$$(A.10) \quad \|\widehat{\mathbf{b}}_{\tau n} - \widetilde{\mathbf{b}}_{\tau n}\|_2 = o_p(n^{-1/2}).$$

Hence, by (A.10), (A.9) and Slutsky's theorem, we have $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau^0) \rightarrow N(0, \tau(1 - \tau) \Lambda^+ \Omega \Lambda^+)$.

If there exists a subsequence n_k such that $\widehat{\mathbf{b}}_{\tau n_k} - \widetilde{\mathbf{b}}_{\tau n_k} = \vartheta_k n_k^{-1/2} \mathbf{u}_{n_k}$, where $\vartheta_k \asymp 1$ and \mathbf{u}_{n_k} is a unit vector. Convexity of $D_{\tau n_k}^*(\mathbf{b})$ implies that for $0 < l < \vartheta_k$ and $l \asymp \vartheta_k$,

$$(1 - l/\vartheta_k) D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k}) + l/\vartheta_k D_{\tau n_k}^*(\widehat{\mathbf{b}}_{\tau, n_k}) \geq D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k} + l n_k^{-1/2} \mathbf{u}_{n_k}),$$

so that

$$(l/\vartheta_k)\{D_{\tau n_k}^*(\widehat{\mathbf{b}}_{\tau, n_k}) - D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k})\} \geq D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k} + l n_k^{-1/2} \mathbf{u}_{n_k}) - D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k}).$$

Since $D_{\tau n_k}^*(\widehat{\mathbf{b}}_{\tau, n_k}) - D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k}) = D_{\tau n_k}(\widehat{\mathbf{b}}_{\tau, n_k}) - D_{\tau n_k}(\widetilde{\mathbf{b}}_{\tau, n_k}) + o_p(n_k^{-1})$, then

$$\begin{aligned} n_k\{D_{\tau n_k}(\widehat{\mathbf{b}}_{\tau, n_k}) - D_{\tau n_k}(\widetilde{\mathbf{b}}_{\tau, n_k})\} &\geq (\vartheta_k/l)n_k\{D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k} + l n_k^{-1/2} \mathbf{u}_{n_k}) - D_{\tau n_k}^*(\widetilde{\mathbf{b}}_{\tau, n_k})\} + o(1) \\ &= (\vartheta_k/l)n_k(l n_k^{-1/2} \mathbf{u}_{n_k})^T (\Lambda_{n_k} + \varrho_{n_k} \mathbf{I})(l n_k^{-1/2} \mathbf{u}_{n_k}) + o(1) \\ &\geq C \vartheta_k l \mathbf{u}_{n_k}^T \Lambda_{n_k} \mathbf{u}_{n_k} \end{aligned}$$

with probability approaching 1, for some constant $0 < C < \infty$. We decompose $\mathbf{u}_{n_k} = h_{1, n_k} \boldsymbol{\beta}_\tau^0 + h_{2, n_k} \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a unit vector and orthogonal to $\boldsymbol{\beta}_\tau^0$. Then $\boldsymbol{\beta}_\tau^{0T} \mathbf{u}_{n_k} = \boldsymbol{\beta}_\tau^{0T} (h_{1, n_k} \boldsymbol{\beta}_\tau^0 + h_{2, n_k} \boldsymbol{\eta}) = h_{1, n_k}$, since $\boldsymbol{\beta}_\tau^{0T} \boldsymbol{\beta}_\tau^0 = 1$ and $\boldsymbol{\beta}_\tau^{0T} \boldsymbol{\eta} = 0$. Recall that $\widehat{\mathbf{b}}_{\tau n_k} = s_{\tau n} \boldsymbol{\eta}$. Therefore,

$$\boldsymbol{\beta}_\tau^{0T} \mathbf{u}_{n_k} = \vartheta_k^{-1} n_k^{1/2} \boldsymbol{\beta}_\tau^{0T} (\widehat{\mathbf{b}}_{\tau n_k} - \widetilde{\mathbf{b}}_{\tau n_k}) = -\vartheta_k^{-1} n_k^{1/2} \boldsymbol{\beta}_\tau^{0T} \widetilde{\mathbf{b}}_{\tau n_k} = h_{1, n_k}.$$

By (A.9), we have with probability approaching 1, $\text{var}(n_k^{1/2} \boldsymbol{\beta}_\tau^{0T} \widetilde{\mathbf{b}}_{\tau n_k}) \rightarrow \tau(1 - \tau) \boldsymbol{\beta}_\tau^{0T} \Lambda^+ \Omega \Lambda^+ \boldsymbol{\beta}_\tau^0$ as $n \rightarrow \infty$. Since $\Lambda \boldsymbol{\beta}_\tau^0 = \mathbf{0}$, $\boldsymbol{\beta}_\tau^0$ is an eigenvector of Λ and orthogonal to other eigenvectors of Λ . Then we have $\Lambda^+ \boldsymbol{\beta}_\tau^0 = \mathbf{0}$. Therefore, $n_k^{1/2} \boldsymbol{\beta}_\tau^{0T} \widetilde{\mathbf{b}}_{\tau n_k} = o_p(1)$ which implies $h_{1, n_k} = o_p(1)$, and thus $h_{2, n_k}^2 = 1 - h_{1, n_k}^2 = 1 - o_p(1)$. Hence, we have with probability approaching 1,

$$n_k\{D_{\tau n_k}(\widehat{\mathbf{b}}_{\tau, n_k}) - D_{\tau n_k}(\widetilde{\mathbf{b}}_{\tau, n_k})\} \geq C \vartheta_k l \mathbf{u}_{n_k}^T \Lambda_{n_k} \mathbf{u}_{n_k} = C \vartheta_k l h_{2, n_k}^2 \boldsymbol{\eta}^T \Lambda_{n_k} \boldsymbol{\eta} > 0.$$

This contradicts with the fact that $\widehat{\mathbf{b}}_{\tau, n_k}$ minimizes $D_{\tau n_k}(\mathbf{b})$. Therefore, $\|\widehat{\mathbf{b}}_{\tau n} - \widetilde{\mathbf{b}}_{\tau n}\|_2 = o_p(n^{-1/2})$.

A.3. Proof of Theorem 3. Theorem 3 follows from Lemma S.3 and root- n consistency of $\widehat{\boldsymbol{\beta}}_\tau$ directly.

A.4. Proof of Theorem 4. Since minimizing the objective function $n Q_{\tau n}(\boldsymbol{\beta})$ is equivalent to minimizing

$$n L_{\tau n}^{**}(\boldsymbol{\beta}) - n L_{\tau n}^{**}(\boldsymbol{\beta}_\tau^0) + n \sum_{j=1}^p \omega_j p'_{\lambda_n}(|\widehat{\boldsymbol{\beta}}_{\tau j}^0|) (|\beta_{\tau j}| - |\boldsymbol{\beta}_{\tau j}^0|).$$

For $\boldsymbol{\beta}_\tau^0 + \mathbf{v}/\sqrt{n}$ in the neighborhood of $\boldsymbol{\beta}_\tau^0$, define

$$\begin{aligned} (A.11) \quad G_n(\mathbf{v}) &= n\{L_{\tau n}^{**}(\boldsymbol{\beta}_\tau^0 + \mathbf{v}/\sqrt{n}) - L_{\tau n}^{**}(\boldsymbol{\beta}_\tau^0)\} \\ &\quad + n \sum_{j=1}^p \omega_j p'_{\lambda_n}(|\widehat{\boldsymbol{\beta}}_{\tau j}^0|) (|\boldsymbol{\beta}_{\tau j}^0 + v_j/\sqrt{n}| - |\boldsymbol{\beta}_{\tau j}^0|). \end{aligned}$$

It is derived in Theorem 5 of Zou and Li (2008) that the second term of (A.11) can be expressed as

$$n \sum_{j=1}^p \omega_j p'_{\lambda_n}(|\hat{\beta}_{\tau j}^0|)(|\beta_{\tau j}^0 + v_j/\sqrt{n}| - |\beta_{\tau j}^0|) \rightarrow \begin{cases} 0, & \text{if } \beta_2 = \beta_{\tau 2}^0, \\ \infty, & \text{otherwise,} \end{cases}$$

with probability approaching 1. Therefore, by the epiconvergence results [Geyer (1994), Knight and Fu (2000)], we have $\hat{\beta}_{\tau 2}^{\text{OSE}} \rightarrow \mathbf{0}$ with probability approaching 1. Let $D_{\tau n}^{**}(\mathbf{v}) = L_{\tau n}^{**}(\beta_{\tau}^0 + \mathbf{v}/\sqrt{n}) - L_{\tau n}^{**}(\beta_{\tau}^0)$. Following similar reasoning as the proof for (A.7), it can be proved that

$$\begin{aligned} D_{\tau n}^{**}(\mathbf{v}) &= -n^{-1} \sum_{i=1}^n \{\tau - I(\varepsilon_i < 0)\} G_{\tau}^{(1)}(\mathbf{X}_i^T \beta_{\tau}^0) \mathbf{X}_i^T \mathbf{v} / \sqrt{n} \\ &\quad + n^{-1} \sum_{i=1}^n 2^{-1} f_{\varepsilon}(0 | \mathbf{X}_i) \{G_{\tau}^{(1)}(\mathbf{X}_i^T \beta_{\tau}^0)\}^2 \mathbf{v}^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{v} / n \\ &\quad + o_p(\|\mathbf{v}\|_2^2/n) + o_p(n^{-1}). \end{aligned}$$

Moreover, the consistency result that $\|\hat{\beta}_{\tau}^{\text{OSE}} - \beta_{\tau}^0\|_2 = o_p(1)$ can be proved by the same procedure in the proofs of Lemma 1. Then the asymptotic normality for $\hat{\beta}_{\tau 1}^{\text{OSE}}$ holds by the same arguments in the proofs of Theorem 2. The sparsity such that $P(\hat{\beta}_{\tau 2}^{\text{OSE}} = \mathbf{0}) \rightarrow 1$ can be proved by the same reasoning as given in Kai, Li and Zou (2011), and thus omitted. \square

A.5. Proof of Theorem 5. Let $g_{\tau k,i} = G_{\tau}^{(1)}(\mathbf{X}_i^T \beta_{\tau}^0) \tilde{\mathbf{X}}_{ki}$, $\Omega_{kl} = E(g_{\tau k,i} g_{\tau l,i}^T)$ for $k, l = 1, 2$, $g_{\tau,i} = G_{\tau}^{(1)}(\mathbf{X}_i^T \beta_{\tau}^0) \tilde{\mathbf{X}}_i$, and $\Omega = E(g_{\tau,i} g_{\tau,i}^T) = \begin{pmatrix} \Omega_{21} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$. Let $\Omega_{nkl} = n^{-1} \sum_{i=1}^n (g_{\tau k,i} g_{\tau l,i}^T)$ for $k, l = 1, 2$. First consider

$$\tilde{\mathbf{s}}_2^*(\beta_{\tau}^0) = n^{-1} \sum_{i=1}^n (g_{\tau 2,i} - \Omega_{21} \Omega_{11}^+ g_{\tau 1,i}) \rho_{\tau}^{(1)} \{Y_i - G_{\tau}(\mathbf{X}_i^T \beta_{\tau}^0)\}.$$

Because $E\{\sqrt{n} \tilde{\mathbf{s}}_2^*(\beta_{\tau}^0)\} = \mathbf{0}$ and

$$\begin{aligned} \text{Var}\{\sqrt{n} \tilde{\mathbf{s}}_2^*(\beta_{\tau}^0)\} &= \tau(1 - \tau) E\{(g_{\tau 2,i} - \Omega_{21} \Omega_{11}^+ g_{\tau 1,i})(g_{\tau 2,i} - \Omega_{21} \Omega_{11}^+ g_{\tau 1,i})^T\} \\ &= \tau(1 - \tau) E\{g_{\tau 2,i} g_{\tau 2,i}^T - g_{\tau 2,i} g_{\tau 1,i}^T \Omega_{11}^+ \Omega_{12} - \Omega_{21} \Omega_{11}^+ g_{\tau 1,i} g_{\tau 2,i}^T \\ &\quad + \Omega_{21} \Omega_{11}^+ g_{\tau 1,i} g_{\tau 1,i}^T \Omega_{11}^+ \Omega_{12}\} \\ &= \tau(1 - \tau) (\Omega_{22} - \Omega_{21} \Omega_{11}^+ \Omega_{12}). \end{aligned}$$

By the central limit theorem, we have

$$\sqrt{n} \tilde{\mathbf{s}}_2^*(\beta_{\tau}^0) \rightarrow N(\mathbf{0}_{(p-p_1)}, \tau(1 - \tau) (\Omega_{22} - \Omega_{21} \Omega_{11}^+ \Omega_{12})).$$

Next, define

$$\tilde{s}_2(\boldsymbol{\beta}_\tau^0) = n^{-1} \sum_{i=1}^n (g_{\tau 2,i} - \Omega_{n21} \Omega_{n11}^+ g_{\tau 1,i}) \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\}.$$

Since with probability approaching 1,

$$\begin{aligned} & E\{[\tilde{s}_2(\boldsymbol{\beta}_\tau^0) - \tilde{s}_2^*(\boldsymbol{\beta}_\tau^0)]^T [\tilde{s}_2(\boldsymbol{\beta}_\tau^0) - \tilde{s}_2^*(\boldsymbol{\beta}_\tau^0)]\} \\ &= n^{-1} \tau(1 - \tau) E[g_{\tau 1,i}^T \{\Omega_{21} \Omega_{11}^+ - \Omega_{n21} \Omega_{n11}^+\}^T \{\Omega_{21} \Omega_{11}^+ - \Omega_{n21} \Omega_{n11}^+\} g_{\tau 1,i}] \\ &= o(n^{-1}), \end{aligned}$$

where the last equation holds due to the fact that $\Omega_{n21} \Omega_{n11}^+ = \Omega_{21} \Omega_{11}^+ + o_p(1)$, we have $\|\tilde{s}_2(\boldsymbol{\beta}_\tau^0) - \tilde{s}_2^*(\boldsymbol{\beta}_\tau^0)\|_2 = o_p(n^{-1/2})$. In the following, it suffices to show

(A.12) $\quad \|\widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}_\tau^N) - \tilde{s}_2(\boldsymbol{\beta}_\tau^0)\|_2 = o_p(n^{-1/2}),$

(A.13) $\quad \widehat{\Omega}_n^{22} = \Omega^{22} + o_p(1),$

where $\widehat{\Omega}_n^{22}$ is given in (11).

By the smoothness conditions of G_τ given in Condition (C3), we have the non-parametric uniform convergence rates of the quantile spline estimators [Portnoy (1997)] given as

$$\sup_{1 \leq i \leq n} |\widetilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0) - G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)| = O_p\{(J_n^3 n^{-1} \log n)^{1/2} + J_n^{-r+1}\},$$

and by the smoothness condition of $E(\mathbf{X}|\mathbf{X}^T \boldsymbol{\beta}_\tau^0 = u)$ given in Condition (C4), we have the nonparametric uniform convergence rates of the least squares spline estimator [Wang and Yang (2009)] given as $\sup_{1 \leq i \leq n} |\widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta}_\tau^0) - \widetilde{\mathbf{X}}_{2i}| = O_p\{(J_n n^{-1} \log n)^{1/2} + J_n^{-1}\}$. Hence,

(A.14) $\quad \sup_{1 \leq i \leq n} |\widetilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0) \widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta}_\tau^0) - g_{\tau 2,i}| \leq \omega_n,$

and $\omega_n = O_p\{(J_n^3 n^{-1} \log n)^{1/2} + J_n^{-r+1} + J_n^{-1}\}$. By (A.14) and Theorem 2, we have $|\widetilde{G}_\tau^{(1)}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau^N, \widehat{\boldsymbol{\beta}}_\tau^N) \widehat{\mathbf{X}}_i - G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0) \widetilde{\mathbf{X}}_i| = o_p(1)$, so that result (A.13) follows. To show (A.12), we just need to verify

(A.15)
$$\begin{aligned} \Pi_{n1}(\widehat{\boldsymbol{\beta}}_\tau^N) &= n^{-1} \sum_{i=1}^n [\rho_\tau^{(1)} \{Y_i - \widetilde{G}_{\tau n}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau^N, \widehat{\boldsymbol{\beta}}_\tau^N)\} \\ &\quad - \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\}] g_{\tau 2,i} \\ &= -n^{-1} \sum_{i=1}^n \Omega_{n21} \Omega_{n11}^+ g_{\tau 1,i} \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\} + o_p(n^{-1/2}), \end{aligned}$$

$$\begin{aligned}
 \Pi_{n2}(\widehat{\boldsymbol{\beta}}_\tau^N) &= n^{-1} \sum_{i=1}^n \rho_\tau^{(1)} \{Y_i - \widetilde{G}_{\tau n}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau^N, \widehat{\boldsymbol{\beta}}_\tau^N)\} \{g_{\tau 2,i} - \widetilde{G}_{\tau n}^{(1)}(\mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_\tau^N, \widehat{\boldsymbol{\beta}}_\tau^N) \widehat{\mathbf{X}}_{2i}\} \\
 \text{(A.16)} \quad &= o_p(n^{-1/2}).
 \end{aligned}$$

By the definition of Π_{n1} , it can be further written as

$$\text{(A.17)} \quad \Pi_{n1}(\widehat{\boldsymbol{\beta}}_\tau^N) = \Pi_{n11}(\widehat{\boldsymbol{\beta}}_\tau^N, \widetilde{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau^N)) + \Pi_{n12}(\widehat{\boldsymbol{\beta}}_\tau^N),$$

where

$$\begin{aligned}
 \Pi_{n11}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= n^{-1} \sum_{i=1}^n \Pi_{n11,i}(\boldsymbol{\beta}, \boldsymbol{\theta}), \\
 \Pi_{n12}(\boldsymbol{\beta}) &= n^{-1} \sum_{i=1}^n \Pi_{n12,i}(\boldsymbol{\beta}),
 \end{aligned}$$

and

$$\begin{aligned}
 \Pi_{n11,i}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= [\rho_\tau^{(1)} \{Y_i - B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} - \rho_\tau^{(1)} \{Y_i - \widetilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})\}] g_{\tau 2,i}, \\
 \Pi_{n12,i}(\boldsymbol{\beta}) &= [\rho_\tau^{(1)} \{Y_i - \widetilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})\} - \rho_\tau^{(1)} \{Y_i - \widetilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)\}] g_{\tau 2,i}.
 \end{aligned}$$

Moreover, let $\Pi_{n11,ik}(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $\Pi_{n12,ik}(\boldsymbol{\beta})$ be the k th component in $\Pi_{n11,i}(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $\Pi_{n12,i}(\boldsymbol{\beta})$, respectively, for $k = 1, \dots, (p - p_1)$.

Since $\boldsymbol{\beta}_{\tau 2}^0 = \mathbf{0}$ under \mathcal{H}_0 , we have $\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0 = \mathbf{X}_{i1}^T \boldsymbol{\beta}_{\tau 1}^0$. For any $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ with $\boldsymbol{\beta}_2 = \mathbf{0}_{(p-p_1) \times 1}$ and $\boldsymbol{\beta}_1$ in a neighborhood of $\boldsymbol{\beta}_{\tau 1}^0$, we have $\mathbf{X}_i^T \boldsymbol{\beta} = \mathbf{X}_{i1}^T \boldsymbol{\beta}_1$. Since (Y_i, \mathbf{X}_{i1}) and \mathbf{X}_{i2} are independent given $\mathbf{X}_{i1} \boldsymbol{\beta}_{\tau 1}^0$, $\rho_\tau^{(1)} \{Y_i - B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} - \rho_\tau^{(1)} \{Y_i - \widetilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta})\}$ which is a function of Y_i and $\mathbf{X}_{i1}^T \boldsymbol{\beta}_1$ is independent of \mathbf{X}_{i2} given $\mathbf{X}_{i1} \boldsymbol{\beta}_{\tau 1}^0$. Moreover, $E(g_{\tau 2,i} | \mathbf{X}_{i1}^T \boldsymbol{\beta}_{\tau 1}^0) = \mathbf{0}$. Then we have $E\{\Pi_{n11,i}(\boldsymbol{\beta}, \boldsymbol{\theta})\} = E[E\{\Pi_{n11,i}(\boldsymbol{\beta}, \boldsymbol{\theta}) | \mathbf{X}_{i1}^T \boldsymbol{\beta}_{\tau 1}^0\}] = \mathbf{0}$. Similarly, we have $E\{\Pi_{n12,i}(\boldsymbol{\beta}) | \mathbf{X}_{i1}^T \boldsymbol{\beta}_{\tau 1}^0\} = \mathbf{0}$. In page S.26–S.27 of the supplemental materials [Ma and He (2015)], we have shown that

$$\text{(A.18)} \quad \Pi_{n11}(\widehat{\boldsymbol{\beta}}_\tau^N, \widetilde{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau^N)) = o_p(n^{-1/2}).$$

Since $E\{\Pi_{n12,i}(\boldsymbol{\beta}) | \mathbf{X}_{i1}^T \boldsymbol{\beta}_{\tau 1}^0\} = \mathbf{0}$, then $E\{\Pi_{n12,i}(\boldsymbol{\beta}) | \mathbf{X}_i\} = \mathbf{0}$. By following the same procedure as the proofs for (S.54), it can be proved by Bernstein’s inequality given in Bosq (1998) for $\delta_n \asymp n^{-1/2}$,

$$\text{(A.19)} \quad \sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau^0\| \leq \delta_n, \boldsymbol{\beta}_2 = \mathbf{0}} \left\| \Pi_{n12}(\boldsymbol{\beta}) - n^{-1} \sum_{i=1}^n E\{\Pi_{n12,i}(\boldsymbol{\beta}) | \mathbf{X}_i\} \right\|_2 = o_p(n^{-1/2}).$$

By the fact that for sufficiently small $|t|$,

$$E\{\rho_\tau^{(1)}(\varepsilon + t) - \rho_\tau^{(1)}(\varepsilon) | \mathbf{X}\} = f_\varepsilon(0)t + o(|t|),$$

we have $\|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau^0\|_2 \leq \delta_n$,

$$\begin{aligned}
 & n^{-1} \sum_{i=1}^n E\{\Pi_{n12,i}(\boldsymbol{\beta})|\mathbf{X}_i\} \\
 \text{(A.20)} \quad & = n^{-1} \sum_{i=1}^n f_\varepsilon(0)\{\tilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}, \boldsymbol{\beta}) - \tilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)\} g_{\tau 2,i} \\
 & \quad + o(n^{-1/2}).
 \end{aligned}$$

Following the same reasoning as the proof for Theorem 2, by the assumption that $f_\varepsilon(0|\mathbf{X}) = f_\varepsilon(0)$ we have

$$\begin{aligned}
 \hat{\boldsymbol{\beta}}_{\tau 1}^N - \boldsymbol{\beta}_{\tau 1}^0 & = f_\varepsilon(0)^{-1} \Omega_{n11}^+ n^{-1} \sum_{i=1}^n g_{\tau 1,i} \rho_\tau^{(1)}\{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\} \\
 \text{(A.21)} \quad & \quad + o_p(n^{-1/2}),
 \end{aligned}$$

and $\|\hat{\boldsymbol{\beta}}_\tau^N - \boldsymbol{\beta}_\tau^0\|_2 = O_p(n^{-1/2})$. Hence, by (A.20), (A.21) and Taylor’s expansion, we have

$$\begin{aligned}
 & n^{-1} \sum_{i=1}^n E\{\Pi_{n12,i}(\hat{\boldsymbol{\beta}}_\tau^N)|\mathbf{X}_i\} \\
 & = -n^{-1} \sum_{i=1}^n g_{\tau 2,i} f_\varepsilon(0) G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0) \mathbf{X}_{1i}^T (\hat{\boldsymbol{\beta}}_{\tau 1}^N - \boldsymbol{\beta}_{\tau 1}^0) + o_p(n^{-1/2}).
 \end{aligned}$$

For any $\boldsymbol{\beta}_1$ satisfying $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{\tau 1}^0\|_2 \leq \delta_n$ with $\delta_n \asymp n^{-1/2}$, define

$$\text{(A.22)} \quad \Gamma_n(\boldsymbol{\beta}_1) = n^{-1} \sum_{i=1}^n g_{\tau 2,i} f_\varepsilon(0) G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0) E(\mathbf{X}_{1i}|\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)^T (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{\tau 1}^0).$$

Then $E\{\Gamma_n(\boldsymbol{\beta}_1)\} = \mathbf{0}$ and

$$\begin{aligned}
 & E\left\{ \sup_{\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{\tau 1}^0\|_2 \leq \delta_n} \Gamma_n(\boldsymbol{\beta}_1)^T \Gamma_n(\boldsymbol{\beta}_1) \right\} \\
 & \leq \delta_n^2 n^{-1} E[g_{\tau 2,i}^T g_{\tau 2,i} \{G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\}^2 E(\mathbf{X}_{1i}|\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)^T (\mathbf{X}_{1i}|\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)] \\
 & = O(\delta_n^2 n^{-1}).
 \end{aligned}$$

Thus, we have for $\delta_n \asymp n^{-1/2}$

$$\text{(A.23)} \quad \sup_{\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{\tau 1}^0\|_2 \leq \delta_n} \|\Gamma_n(\boldsymbol{\beta}_1)\|_2 = o_p(n^{-1/2}).$$

By (S.50), (A.22), and (A.23), we have

$$\begin{aligned}
 & n^{-1} \sum_{i=1}^n E\{\Pi_{n12,i}(\widehat{\boldsymbol{\beta}}_\tau^N) | \mathbf{X}_i\} \\
 &= -n^{-1} \sum_{i=1}^n g_{\tau 2,i} f_\varepsilon(0) G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0) \widetilde{\mathbf{X}}_{1i}^T (\widehat{\boldsymbol{\beta}}_\tau^N - \boldsymbol{\beta}_\tau^0) + o_p(n^{-1/2}) \\
 &= -n^{-1} \sum_{i=1}^n g_{\tau 2,i} G_\tau^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0) \widetilde{\mathbf{X}}_{1i}^T \\
 &\quad \times \left[\Omega_{n11}^+ n^{-1} \sum_{i=1}^n g_{\tau 1,i} \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\} \right] \\
 (A.24) \quad &+ o_p(n^{-1/2}) \\
 &= -n^{-1} \sum_{i=1}^n g_{\tau 2,i} g_{\tau 1,i}^T \left[\Omega_{n11}^+ n^{-1} \sum_{i=1}^n g_{\tau 1,i} \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\} \right] \\
 &+ o_p(n^{-1/2}) \\
 &= -n^{-1} \Omega_{n21} \Omega_{n11}^+ \sum_{i=1}^n g_{\tau 1,i} \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\} + o_p(n^{-1/2}).
 \end{aligned}$$

Therefore, the result (A.15) follows from (A.17), (A.18), (A.19) and (A.24).

Denote $\widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta}) = \mathbf{X}_{2i} - \widehat{E}(\mathbf{X}_{2i} | \mathbf{X}_i^T \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_\tau^T, \boldsymbol{\theta}^T)^T$ and $\widehat{E}(\mathbf{X}_{2i} | \mathbf{X}_i^T \boldsymbol{\beta})$ is defined in (10). $\Pi_{n2}(\widehat{\boldsymbol{\beta}}_\tau^N)$ given in (A.16) can be written as

$$\Pi_{n2}(\widehat{\boldsymbol{\beta}}_\tau^N) = \Pi_{n2}(\widehat{\boldsymbol{\beta}}_\tau^N, \widetilde{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau^N)) = \Pi_{n21}(\widehat{\boldsymbol{\beta}}_\tau^N, \widetilde{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau^N)) + \Pi_{n22}(\widehat{\boldsymbol{\beta}}_\tau^N, \widetilde{\boldsymbol{\theta}}_\tau(\widehat{\boldsymbol{\beta}}_\tau^N)),$$

where

$$\Pi_{n21}(\boldsymbol{\beta}, \boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \Pi_{n21,i}(\boldsymbol{\beta}, \boldsymbol{\theta}), \quad \Pi_{n22}(\boldsymbol{\beta}, \boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \Pi_{n22,i}(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

$$\begin{aligned}
 \Pi_{n21,i}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= [\rho_\tau^{(1)} \{Y_i - \{B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\}\} - \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\}] \\
 &\quad \times \{g_{\tau 2,i} - \{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta})\},
 \end{aligned}$$

$$\Pi_{n22,i}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \rho_\tau^{(1)} \{Y_i - G_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)\} \{g_{\tau 2,i} - \{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta})\}.$$

Denote $F_n = \{(\boldsymbol{\beta}, \boldsymbol{\theta}) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau^0\|_2 \leq \delta_n, \|\boldsymbol{\theta} - \boldsymbol{\theta}_\tau^0(\boldsymbol{\beta}_\tau^0)\| \leq L_n\}$ for $\delta_n \asymp n^{-1/2}$ and $L_n \asymp J_n n^{-1/2}$. Then, by (A.14) and

$$\sup_{1 \leq i \leq n, \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau^0\|_2 \leq \delta_n} |\{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta}) - \{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0)^T \boldsymbol{\theta}\} \widehat{\mathbf{X}}_{2i}(\boldsymbol{\beta}_\tau^0)| = O(\delta_n),$$

there is a constant $0 < C < \infty$,

$$\begin{aligned}
 & E \left\{ \sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} \Pi_{n21}(\boldsymbol{\beta}, \boldsymbol{\theta})^T \Pi_{n21}(\boldsymbol{\beta}, \boldsymbol{\theta}) \right\} \\
 & \leq Cn^{-2} \sum_{i, i'} E \sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} [|\rho_\tau^{(1)}\{Y_i - \{B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \\
 & \quad - \rho_\tau^{(1)}\{Y_i - \tilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)\}| \times |\rho_\tau^{(1)}\{Y_{i'} - \{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \\
 & \quad - \rho_\tau^{(1)}\{Y_{i'} - \tilde{G}_\tau(\mathbf{X}_{i'}^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)\}|] (\omega_n^2 + \delta_n^2) \\
 & \leq C(\omega_n^2 + \delta_n^2)n^{-2} \left[\sum_{i=1}^n E \sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} |B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta} - \tilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)| \right. \\
 & \quad \left. + \sum_{i \neq i'} E \left[\sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} |B(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta} - \tilde{G}_\tau(\mathbf{X}_i^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)| \right. \right. \\
 & \quad \left. \left. \times |B(\mathbf{X}_{i'}^T \boldsymbol{\beta})^T \boldsymbol{\theta} - \tilde{G}_\tau(\mathbf{X}_{i'}^T \boldsymbol{\beta}_\tau^0, \boldsymbol{\beta}_\tau^0)| \right] \right] \\
 & = (\omega_n^2 + \delta_n^2) [n^{-1} O_p\{J_n^{-r} + (J_n n^{-1})^{1/2}\} + O_p\{J_n^{-2r} + J_n n^{-1}\}] \\
 & = o_p(n^{-1}),
 \end{aligned}$$

due to the fact that $\omega_n = O_p\{(J_n^3 n^{-1} \log n)^{1/2} + J_n^{-r+1} + J_n^{-1}\}$. Hence, $\sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} \|\Pi_{n21}(\boldsymbol{\beta}, \boldsymbol{\theta})\|_2 = o_p(n^{-1/2})$ which implies

$$\|\Pi_{n21}(\hat{\boldsymbol{\beta}}_\tau^N, \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N))\|_2 = o_p(n^{-1/2}).$$

Since $E\{\Pi_{n22,i}(\boldsymbol{\beta}, \boldsymbol{\theta}) | \mathbb{X}\} = \mathbf{0}$, then $E\{\Pi_{n2,i}(\boldsymbol{\beta}, \boldsymbol{\theta})\} = \mathbf{0}$. Moreover,

$$\begin{aligned}
 & E \left\{ \sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} \Pi_{n22,i}(\boldsymbol{\beta}, \boldsymbol{\theta})^T \Pi_{n21,i}(\boldsymbol{\beta}, \boldsymbol{\theta}) \right\} \\
 & = \tau(1 - \tau) E \sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} [\{g_{\tau2,i} - \{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \hat{\mathbf{X}}_{2i}(\boldsymbol{\beta})\}^T \\
 & \quad \times \{g_{\tau2,i} - \{B^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta})^T \boldsymbol{\theta}\} \hat{\mathbf{X}}_{2i}(\boldsymbol{\beta})\}] \\
 & = o_p(1).
 \end{aligned}$$

Then by following the same procedure as the proofs for (S.54), we have $\sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in F_n} \|\Pi_{n22}(\boldsymbol{\beta}, \boldsymbol{\theta})\|_2 = o_p(n^{-1/2})$, so that $\|\Pi_{n22}(\hat{\boldsymbol{\beta}}_\tau^N, \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N))\|_2 = o_p(n^{-1/2})$. Therefore,

$$\|\Pi_{n2}(\hat{\boldsymbol{\beta}}_\tau^N)\|_2 \leq \|\Pi_{n21}(\hat{\boldsymbol{\beta}}_\tau^N, \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N))\|_2 + \|\Pi_{n22}(\hat{\boldsymbol{\beta}}_\tau^N, \tilde{\boldsymbol{\theta}}_\tau(\hat{\boldsymbol{\beta}}_\tau^N))\|_2 = o_p(n^{-1/2}).$$

Thus, the result in (A.16) is proved.

SUPPLEMENTARY MATERIAL

Supplement to “Inference for single-index quantile regression models with profile optimization” (DOI: [10.1214/15-AOS1404SUPP](https://doi.org/10.1214/15-AOS1404SUPP); .pdf). We present several lemmas that will be used in the proof of the main theorems, and the proof of equation (A.18). Then we present Example 2 for Case 1 and additional simulation results for Case 2 in the simulation studies.

REFERENCES

- ANDREWS, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* **55** 1465–1471. [MR0923471](#)
- BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1 -penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. [MR2797841](#)
- BHATTACHARYA, P. K. and GANGOPADHYAY, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415. [MR1062716](#)
- BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd ed. Springer, New York. [MR1640691](#)
- CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. [MR1467842](#)
- CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777. [MR1105843](#)
- CHAUDHURI, P., DOKSUM, K. and SAMAROV, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25** 715–744. [MR1439320](#)
- CUI, X., HÄRDLE, W. K. and ZHU, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39** 1658–1688. [MR2850216](#)
- DE BOOR, C. (2001). *A Practical Guide to Splines*, Revised ed. *Applied Mathematical Sciences* **27**. Springer, New York. [MR1900298](#)
- DE GOOIJER, J. G. and ZEROM, D. (2003). On additive conditional quantiles with high-dimensional covariates. *J. Amer. Statist. Assoc.* **98** 135–146. [MR1965680](#)
- FAN, J., HU, T. C. and TRUONG, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Stat.* **21** 433–446. [MR1310087](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- GEYER, C. J. (1994). On the asymptotics of constrained M -estimation. *Ann. Statist.* **22** 1993–2010. [MR1329179](#)
- HAMMER, S. M., SQUIRES, K. E., HUGHES, M. D., GRIMES, J. M., DEMETER, L. M., CARRIER, J. S., ERON, J. J., FEINBERG, J. E., BALFOUR, H. H., DEYTON, L. R., CHODAKEWITZ, J. A. and FISCHL, M. A. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS clinical trials group 320 study team. *N. Engl. J. Med.* **337** 725–733.
- HE, X., NG, P. and PORTNOY, S. (1998). Bivariate quantile smoothing splines. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 537–550. [MR1625950](#)
- HE, X. and SHI, P. D. (1994). Convergence rate of B -spline estimators of nonparametric conditional quantile functions. *J. Nonparametr. Stat.* **3** 299–308. [MR1291551](#)
- HE, X. and SHI, P. (1996). Bivariate tensor-product B -splines in a partly linear model. *J. Multivariate Anal.* **58** 162–181. [MR1405586](#)
- HOROWITZ, J. L. and LEE, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100** 1238–1249. [MR2236438](#)

- ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. [MR1230981](#)
- KAI, B., LI, R. and ZOU, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Statist.* **39** 305–332. [MR2797848](#)
- KIM, M.-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35** 92–108. [MR2332270](#)
- KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. [MR1805787](#)
- KOCHERGINSKY, M., HE, X. and MU, Y. (2005). Practical confidence intervals for regression quantiles. *J. Comput. Graph. Statist.* **14** 41–55. [MR2137889](#)
- KOENKER, R. (2005). *Quantile Regression*. Cambridge Univ. Press, Cambridge. [MR2268657](#)
- KOENKER, R. (2011). Additive models for quantile regression: Model selection and confidence bands. *Braz. J. Probab. Stat.* **25** 239–262. [MR2832886](#)
- KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. [MR0474644](#)
- KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. [MR1326417](#)
- KONG, E. and XIA, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory* **28** 730–768. [MR2959124](#)
- LIANG, H., LIU, X., LI, R. and TSAI, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38** 3811–3836. [MR2766869](#)
- MA, S. and HE, X. (2015). Supplement to “Inference for single-index quantile regression models with profile optimization.” DOI:[10.1214/15-AOS1404SUPP](#).
- MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. [MR1803168](#)
- POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199. [MR1128411](#)
- PORTNOY, S. (1997). Local asymptotics for quantile smoothing splines. *Ann. Statist.* **25** 414–434. [MR1429932](#)
- SEVERINI, T. A. and STANISWALIS, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89** 501–511. [MR1294076](#)
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- WANG, H., LI, R. and TSAI, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94** 553–568. [MR2410008](#)
- WANG, J. and YANG, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19** 325–342. [MR2487893](#)
- WANG, H. J., ZHU, Z. and ZHOU, J. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37** 3841–3866. [MR2572445](#)
- WU, Y. and LIU, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19** 801–817. [MR2514189](#)
- WU, T. Z., YU, K. and YU, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.* **101** 1607–1621. [MR2610735](#)
- XIA, Y. and HÄRDLE, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.* **97** 1162–1184. [MR2276153](#)
- XIA, Y., TONG, H. and LI, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831–842. [MR1741980](#)
- YU, K. and JONES, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. [MR1614628](#)
- YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. [MR1951258](#)

- ZHU, L., HUANG, M. and LI, R. (2012). Semiparametric quantile regression with high-dimensional covariates. *Statist. Sinica* **22** 1379–1401. [MR3027092](#)
- ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. [MR2435443](#)
- ZOU, Q. and ZHU, Z. (2014). M-estimators for single-index model using B-spline. *Metrika* **77** 225–246. [MR3157984](#)

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, RIVERSIDE
RIVERSIDE, CALIFORNIA 92521
USA
E-MAIL: shujie.ma@ucr.edu
URL: <http://faculty.ucr.edu/~shujiema/>

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-2029
USA
E-MAIL: xmhe@umich.edu
URL: <http://www.xuminghe.com/>