Bernoulli **23**(3), 2017, 1997–2027
DOI: 10.3150/16-BEJ802

Quantile regression for the single-index coefficient model

WEIHUA ZHAO[1], HENG LIAN[2] and HUA LIANG[3]

[1]*School of Science, Nantong University, Nantong, P. R. China, 22601*
[2]*School of Mathematics and Statistics, University of New South Wales, Sydney Australia 2052.*
E-mail: heng.lian@unsw.edu.au
[3]*Department of Statistics, George Washington University, Washington, D.C. 20052, USA*

We consider quantile regression incorporating polynomial spline approximation for single-index coefficient models. Compared to mean regression, quantile regression for this class of models is more technically challenging and has not been considered before. We use a check loss minimization approach and employed a projection/orthogonalization technique to deal with the theoretical challenges. Compared to previously used kernel estimation approach, which was developed for mean regression only, spline estimation is more computationally expedient and directly produces a smooth estimated curve. Simulations and a real data set is used to illustrate the finite sample properties of the proposed estimator.

Keywords: asymptotic normality; B-splines; check loss minimization; single-index coefficient models; quantile regression

1. Introduction

The varying coefficient model (VCM) has gained much attention in the literature since its introduction by Hastie and Tibshirani [9] and Chen and Tsay [4], for cross-sectional data and time series data, respectively. The model is given by

$$Y_i = \sum_{j=1}^{p} g_j(X_i) Z_{ij} + \varepsilon_i,$$

given observations $Y_i, X_i, \mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})^T, i = 1, \dots, n$, where X is usually called an index variable in this context. Fan and Zhang [8] proposed a two-step local linear estimator in the VCM which achieves univariate optimal convergence rate. The VCM also has wide applications in longitudinal studies, see Hoover et al. [13], Fan and Zhang [7] and Huang, Wu and Zhou [15].

In practice, varying-coefficient models are almost exclusively applied for the case X is a scalar random variable. Although it can be directly generalized to the case with multivariate index vector $\mathbf{X}$, it will cause the "curse of dimensionality." Xia and Li [30] proposed an elegant solution for multivariate $\mathbf{X}$ by introducing a single-index structure for the index vector, resulting in

$$Y_i = \sum_{j=1}^{p} g_j(\mathbf{X}_i^T \boldsymbol{\beta}) Z_{ij} + \varepsilon_i,$$

which was termed the single-index coefficient model (SICM). A more general framework was proposed recently by Ma and Song [22] in which $\boldsymbol{\beta}$ can be different for different function g_j. Another closely related model, termed the adaptive varying-coefficient linear models (Fan, Yao and Cai [6]), has the same form as SICM with $\mathbf{X} = \mathbf{Z}$. Further studies of SICM include Xue and Wang [32], Huang and Zhang [16].

As far as we know, researchers have only considered mean regression for SICM so far, which may be somewhat limiting. The parametric quantile regression introduced by Koenker and Bassett [18] has been well developed in the econometrics and statistics literature. When the distribution of the errors in the model is heavy tailed or the data contain some outliers, it is well known that median regression, a special case of quantile regression, is more robust than mean regression. More importantly, it can be used to obtain a large collection of conditional quantiles to characterize the entire conditional distribution. To construct a richer class of regression models capturing flexibly the relationships between the covariates and the response distribution, nonparametric quantile estimation has been studied in Hendricks and Koenker [12], Yu and Jones [33]. For varying coefficient models, Kim [17] studied quantile regression for independent data using splines, and Cai and Xu [2] used local polynomial estimation method for time series data. Further extensions to partially linear varying-coefficient models are considered by Wang, Zhu and Zhou [25], Cai and Xiao [1].

In this paper, we will develop theory and methodology for the quantile SICM using polynomial spline estimation. Polynomial spline estimation provides an alternative to local polynomial estimation method. The comparative advantages of spline methods were carefully documented in Li [19], among which the most notable is the computational convenience, although it is not our main intention here to promote splines. The disadvantage is that the exact bias term for the nonparametric function is harder to obtain, making demonstration of asymptotic normality for the nonparametric functions difficult.

Spline estimation for models with a single-index structure was considered in Wang and Yang [27], Ma, Liang and Tsai [21]. Theoretically, our study is complicated by two aspects: consideration of single-index structure and consideration of quantile regression which has a non-smooth objective function. In particular, appropriate definition and analysis of the projection of the parametric part on the nonparametric part, which is usually necessary in semiparametric models to demonstrate asymptotic normality of the parametric part, is more complicated than in partially linear models such as that studied in Wang, Zhu and Zhou [25].

The rest of the paper is organized as follows. In the next section, we present the estimation method using polynomial splines, and asymptotic properties of the estimators are considered. Section 3 contains our numerical results including simulation studies and an application to an environmental data set. Finally, we conclude in Section 4 with some mentioning of possible extensions. The technical proofs are relegated to the Appendix.

2. Quantile single-index coefficient models

2.1. Estimation

Consider the quantile SICM

$$Y_i = \mathbf{g}^T\left(\mathbf{X}_i^T\boldsymbol{\beta}\right)\mathbf{Z}_i + e_i,$$

where $(\mathbf{X}_i, \mathbf{Z}_i, Y_i, e_i)$ are independent and identically distributed (i.i.d.), $\mathbf{g}(\cdot) = (g_1(\cdot), \ldots, g_q(\cdot))$ are the q coefficient functions whose argument is the index $\mathbf{X}_i^T\boldsymbol{\beta}$, $P(e_i \le 0|\mathbf{X}_i, \mathbf{Z}_i) = \tau$, and $\mathbf{X}_i$ and $\mathbf{Z}_i$ are q-dimensional and p-dimensional covariates, respectively. Such a structure in mean regression was first considered in Xia and Li [30]. For identifiability, we assume $\|\boldsymbol{\beta}\| = 1$ and its first component is positive.

To take into account the unit norm constraint, we use the popular "delete-one-component" method (Yu and Ruppert [34], Cui, Härdle and Zhu [5]). We can write $\boldsymbol{\beta} = ((1 - \|\boldsymbol{\beta}^{(-1)}\|^2)^{1/2}, \beta_2, \ldots, \beta_q)^T$ where $\boldsymbol{\beta}^{(-1)} = (\beta_2, \ldots, \beta_q)^T$ is $\boldsymbol{\beta}$ without the first component. Thus, $\boldsymbol{\beta}$ is a function of $\boldsymbol{\beta}^{(-1)}$. The $q \times (q-1)$ Jacobian matrix is

$$\mathbf{J} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{(-1)}} = \begin{pmatrix} -\dfrac{\boldsymbol{\beta}^{(-1)}}{(1 - \|\boldsymbol{\beta}^{(-1)}\|^2)^{1/2}} \\ \mathbf{I}_{(q-1)\times(q-1)} \end{pmatrix},$$

where $\mathbf{I}_{(q-1)\times(q-1)}$ is the $(q-1)\times(q-1)$ identity matrix.

We use polynomial splines to approximate the components. Let $\tau_0 = a < \tau_1 < \cdots < \tau_{K'} < b = \tau_{K'+1}$ be a partition of $[a, b]$ into subintervals $[\tau_k, \tau_{k+1}), k = 0, \ldots, K'$ with K' internal knots. We only restrict our attention to equally spaced knots although data-driven choice can be considered such as putting knots at certain sample quantiles of the observed covariate values. A polynomial spline of order s is a function whose restriction to each subinterval is a polynomial of degree $s-1$ and globally $s-2$ times continuously differentiable on $[a, b]$. The collection of splines with a fixed sequence of knots has a normalized B-spline basis $\{B_1(x), \ldots, B_K(x)\}$ with $K = K' + s$. Note that it is possible to specify different K for each component but we assume they are the same for simplicity (using the same K's is reasonable when all components have the same smoothness parameter). In the empirical implementations, we use the minimal and maximal values of $\mathbf{X}_i^T\boldsymbol{\beta}$ as a and b to generate B-spline basis functions for a given $\boldsymbol{\beta}$.

We assume B-spline basis is normalized to have $\sum_{k=1}^K B_k(x) = \sqrt{K}$. Such normalization is not essential and is just imposed to simplify some expressions in theoretical derivations later. Let $\mathbf{B}(\cdot) = (B_1(\cdot), \ldots, B_K(\cdot))^T$. Using spline estimator, writing $g_j(\cdot) \approx \mathbf{B}^T(\cdot)\boldsymbol{\theta}_j$, we minimize

$$\sum_{i=1}^n \rho_\tau\left(Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T\left(\mathbf{X}_i^T\boldsymbol{\beta}\right)\boldsymbol{\theta}_j\right), \tag{1}$$

with the constraint $\|\boldsymbol{\beta}\| = 1$ and $\beta_1 > 0$, or equivalently regarding $\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{\beta}^{(-1)})$ as a function of $\boldsymbol{\beta}^{(-1)}$ and optimize over $(\boldsymbol{\beta}^{(-1)}, \boldsymbol{\theta})$.

Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ and $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{\Theta})$, where $\mathrm{vec}(\cdot)$ denotes the straightening operation of the matrix by stacking its columns. We note that $\sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j$ can also be written more succinctly as $\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\Theta}\mathbf{Z}_i = (\mathbf{Z}_i \otimes \mathbf{B}(\mathbf{X}_i^T\boldsymbol{\beta}))^T\boldsymbol{\theta}$, where $\otimes$ denotes the Kronecker product of two matrices.

2.2. Large sample properties

For proof of convergence rate and asymptotic normality, we need to "orthogonalize" the parametric part with respect to the nonparametric part using the following projection. Let $\mathcal{M} = \{m : m(\mathbf{x}, \mathbf{z}) = \mathbf{f}^T(\mathbf{x}^T\boldsymbol{\beta}_0)\mathbf{z}, Em^2(\mathbf{X}, \mathbf{Z}) < \infty\}$ be the *space of varying index functions*. In this paper, the projection of any random variable W onto $\mathcal{M}$, denoted by $E_{\mathcal{M}}[W]$, is defined as the minimizer of

$$E\big[f(0|\mathbf{X}, \mathbf{Z})\big(W - m(\mathbf{X}, \mathbf{Z})\big)^2\big],$$

with $m \in \mathcal{M}$. This definition can be extended trivially to the case where $\mathbf{W} = (W_1, \dots, W_q)^T$ is a random vector by $E_{\mathcal{M}}(\mathbf{W}) = (E_{\mathcal{M}}(W_1), \dots, E_{\mathcal{M}}(W_q))^T$.

We impose the following assumptions.

(A1) The covariates $\mathbf{Z}$, $\mathbf{X}$ are bounded.

(A2) Let $f(\cdot|\mathbf{X}_i, \mathbf{Z}_i)$ be the conditional density of e_i. We assume $f(\cdot|\mathbf{X}_i, \mathbf{Z}_i)$ is bounded and bounded away from zero in a neighborhood of zero, uniformly over the support of $\mathbf{X}_i, \mathbf{Z}_i$. The derivative of $f(\cdot|\mathbf{X}_i, \mathbf{Z}_i)$ is uniformly bounded in a neighborhood of zero over the support of $\mathbf{X}_i, \mathbf{Z}_i$.

(A3) The functions g_j are in the Hölder space of order $d \geq 2$. That is $|g_j^{(m)}(x) - g_j^{(m)}(y)| \leq C|x - y|^r$ for $d = m + r$ and m is the largest integer strictly smaller than d, where $g_j^{(m)}$ is the mth derivative of g_j.

(A4) Suppose $E_{\mathcal{M}}[X_j\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}] = \sum_{l=1}^p f_{jl}(\mathbf{X}^T\boldsymbol{\beta}_0)Z_l$, $1 \leq j \leq q$. The functions f_{jl} are in the Hölder space of order $d' \geq 1$. The order of the B-spline used satisfies $s \geq \max\{d, d'\} + 1$.

(A5) $E[f(0|\mathbf{X}, \mathbf{Z})(\mathbf{X}\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z} - E_{\mathcal{M}}[\mathbf{X}\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}])^{\otimes 2}]$ is positive definite, where for any matrix $\mathbf{A}$, $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^T$.

Boundedness of $\mathbf{Z}$ is assumed mainly for convenience of proof, which can possibly be replaced by moment conditions with lengthier arguments. Boundedness of $\mathbf{X}$ is tied to our estimation approach, typically assumed when using regression splines. Assumption (A2) on conditional density is commonly used in quantile regression (He and Shi [11], Wang, Zhu and Zhou [25]). Smoothness of g_j is required for proof of convergence rate. Smoothness of functions in the representation of $E_{\mathcal{M}}[X_j\mathbf{g}^{(1)^T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}]$ is usually used in semiparametric models to show the asymptotic normality of the parametric part. Finally, (A5) can be regarded as an identifiability assumption for semiparametric models (Li [19], Wei and He [29], Wang et al. [26]).

Theorem 1. *Under conditions* (A1)–(A5) *and that* $K \to \infty$, $K^{d+3/2}\log n/n \to 0$, *there is a local minimizer with*

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p\big(\sqrt{K/n} + K^{-d}\big).$$

In particular, the rate for $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(\sqrt{K/n} + K^{-d})$ *implies that* $\|\hat{g}_j - g_j\| = O_p(\sqrt{K/n} + K^{-d})$, *with* $\hat{g}_j(\cdot) = \mathbf{B}^T(\cdot)\hat{\boldsymbol{\theta}}_j$.

The convergence rate above takes a familiar form as in nonparametric regression with the two terms corresponding to bias and variance, respectively. The optimal choice of K is obviously $K \sim n^{1/(2d+1)}$. Under stronger assumptions on the choice of K and smoothness of nonparametric functions, we have the asymptotic normality of the index parameter $\boldsymbol{\beta}$. Note that when d' is large enough (for example $d' = d$), $K \sim n^{1/(2d+1)}$ is still contained in the permissible range. The technical condition $K^{\max\{4,d+3/2\}} \log n/n \to 0$ comes from an application of Bernstein's inequality in our proof.

Theorem 2. *Under conditions* (A1)–(A5) *and that* $K \to \infty$, $K^{\max\{4,d+3/2\}} \log n/n \to 0$, $\sqrt{n}K^{-2d+3/2} \to 0$, $\sqrt{n}K^{-d-d'} \to 0$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\big(0, \mathbf{J}\big(\mathbf{J}^T\boldsymbol{\Phi}\mathbf{J}\big)^{-1}\mathbf{J}^T\boldsymbol{\Sigma}\mathbf{J}\big(\mathbf{J}^T\boldsymbol{\Phi}\mathbf{J}\big)^{-1}\big)\mathbf{J}^T,$$

where $\boldsymbol{\Phi} = E[f(0|\mathbf{X},\mathbf{Z})(\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X} - E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}])^{\otimes 2}]$, $\boldsymbol{\Sigma} = \tau(1-\tau) \times E[(\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X} - E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}])^{\otimes 2}]$, *and the Jacobian matrix* $\mathbf{J}$ *is evaluated at the truth* $\boldsymbol{\beta}_0$.

Remark 1. We assume $\mathbf{Z}$ is bounded for convenience of proof. This assumption is only used in applying Bernstein's inequality in Lemmas 1 and 5 in the Appendix. More generally, we can assume $E\|\mathbf{Z}\|^\kappa < \infty$ for some $\kappa \geq 2$. Then, for example in the proof of Lemma 1, for some $M > 0$ that diverges with n, we can bound $|M_{ni}(\boldsymbol{\beta},\boldsymbol{\theta})|$ by MA and $E|M_{ni}(\boldsymbol{\beta},\boldsymbol{\theta})|^2$ by M^2D^2 (see the proof of Lemma 1 for the definition of $M_{ni}(\boldsymbol{\beta},\boldsymbol{\theta})$, A, D), when $\|\mathbf{Z}_i\| \leq M$ for all $i = 1,\ldots,n$. Then we will have

$$P\Big(\sup_{(\boldsymbol{\beta},\boldsymbol{\theta})\in\mathcal{N}} \big|M_n(\boldsymbol{\beta},\boldsymbol{\theta}) - EM_n(\boldsymbol{\beta},\boldsymbol{\theta})\big| > a\Big) \leq C\exp\bigg\{-\frac{a^2}{aMA + nM^2D^2} + CK\log n\bigg\} + P\big(\exists i, \|\mathbf{Z}_i\| > M\big).$$

Using Markov's inequality, we have $P(\exists i, \|\mathbf{Z}_i\| > M) \leq nE\|\mathbf{Z}_i\|^\kappa/M^\kappa$ and thus we can choose M to be larger than $n^{1/\kappa}$ so that $P(\exists i, \|\mathbf{Z}_i\| > M)$ converges to zero. Then it is easy to see that with this choice of M, to make the right-hand side of the displayed equation above converge to zero, we only need to replace the condition $K^{d+3/2}\log n/n \to 0$ in the statement of Theorem 1 by $K^{d+3/2}\log n/n^{1-1/\kappa} \to 0$. Similarly, when assuming $E\|\mathbf{Z}\|^\kappa < \infty$, the condition $K^{\max\{4,d+3/2\}}\log n/n \to 0$ in the statement of Theorem 2 need to be replaced by $K^{\max\{4,d+3/2\}}\log n/n^{1-1/\kappa} \to 0$.

To obtain the standard error estimate of $\hat{\boldsymbol{\beta}}$ based on the asymptotic normality result above, we need to estimate the conditional density $f(0|\mathbf{X},\mathbf{Z})$ and $E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}]$. For the estimate $\hat{f}(0|\mathbf{X}_i,\mathbf{Z}_i)$, we adopt the difference quotient method of [12],

$$\hat{f}(0|\mathbf{X}_i,\mathbf{Z}_i) = 2h_n\big\{\hat{\mathbf{g}}^T_{\tau+h_n}\big(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_{\tau+h_n}\big)\mathbf{Z}_i - \hat{\mathbf{g}}^T_{\tau-h_n}\big(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_{\tau-h_n}\big)\mathbf{Z}_i\big\}^{-1},$$

where the estimators $\hat{\boldsymbol{\beta}}_\tau$ and $\hat{\mathbf{g}}_\tau(\cdot)$ are obtained by (1) at quantile level τ and h_n is a bandwidth parameter tending to zero as $n \to \infty$. In our numerical studies, we choose $h_n = 1.57n^{-1/3}(1.5\phi^2\{\Phi^{-1}(\tau)\}/(2\{\Phi^{-1}(\tau)\}^2 + 1))^{2/3}$ following [12], where $\phi(\cdot)$ and $\Phi(\cdot)$ are the p.d.f. and c.d.f. of the standard normal distribution, respectively. For the estimate of $E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}_i\mathbf{X}_i]$, it can be obtained by the weighted least square method on the B-spline space by regarding $\hat{\mathbf{g}}^{(1)T}(\mathbf{X}_i^T\hat{\boldsymbol{\beta}})\mathbf{Z}_i\mathbf{X}_i$ as the response variable.

Denote the estimate of $E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}_i\mathbf{X}_i]$ as $\hat{\boldsymbol{\Delta}}_i$, $i = 1, \ldots, n$. Define

$$\hat{\boldsymbol{\Phi}} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(0|\mathbf{X}_i, \mathbf{Z}_i)\big(\hat{\mathbf{g}}^{(1)T}\big(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}\big)\mathbf{Z}_i\mathbf{X}_i - \hat{\boldsymbol{\Delta}}_i\big)^{\otimes 2}$$

and

$$\hat{\boldsymbol{\Sigma}} = \tau(1-\tau)\frac{1}{n}\sum_{i=1}^{n}\big(\hat{\mathbf{g}}^{(1)T}\big(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}\big)\mathbf{Z}_i\mathbf{X}_i - \hat{\boldsymbol{\Delta}}_i\big)^{\otimes 2}.$$

Then we have following proposition. Consistent estimate of the conditional density can in theory be obtained by kernel methods. However, this would suffer from curse of dimensionality, which is the reason the above more practical procedure is adopted. The proof is a simple corollary of Lemma 7 in the Appendix and thus its proof is omitted.

Proposition 1. *Under the conditions of Theorem 2, assuming $f(0|\mathbf{X}, \mathbf{Z})$ is consistently estimated by $\hat{f}(0|\mathbf{X}, \mathbf{Z})$, we have*

$$\hat{\boldsymbol{\Phi}} \longrightarrow \boldsymbol{\Phi} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} \longrightarrow \boldsymbol{\Sigma} \qquad \text{as } n \to \infty$$

in probability.

Based on Proposition 1, covariance of $\hat{\boldsymbol{\beta}}$ can be consistently estimated by

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = n^{-1}\mathbf{J}\big(\mathbf{J}^T\hat{\boldsymbol{\Phi}}\mathbf{J}\big)^{-1}\mathbf{J}^T\hat{\boldsymbol{\Sigma}}\mathbf{J}\big(\mathbf{J}^T\hat{\boldsymbol{\Phi}}\mathbf{J}\big)^{-1}\mathbf{J}^T.$$

3. Numerical illustrations

In this section, we first carry out some simulations to demonstrate the finite sample performance of the proposed quantile SICM estimator, and then apply SICM to an environmental data set.

In all our numerical examples, the nonparametric functions are approximated by cubic spline ($s = 4$), and the number of basis functions K is chosen by minimizing the following Schwarz Information Criterion (SIC, Schwartz [23], He and Shi [10], Horowitz and Lee [14])

$$\hat{K} = \arg\min_K \mathrm{SIC}(K),$$

where

$$\mathrm{SIC}(K) = \log\left(\sum_{i=1}^{n} \rho_\tau\left(Y_i - \sum_{j=1}^{p} Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}\big)\hat{\boldsymbol{\theta}}_j\right)\right) + \log(n) \times (pK)/(2n),$$

where $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_j$ denotes the estimated parameters for a given K.

We use a profile approach to fit the model. More specifically, for a given $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ can be easily obtained since the minimization problem is the same as that for standard quantile regression. Regarding $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\beta})$ as a function of $\boldsymbol{\beta}$, we optimize over $\boldsymbol{\beta}$ using the *constrOptim.nl()* function in R. The initial value of $\boldsymbol{\beta}$ is obtained simply by fitting a simple linear quantile regression to the data $(\mathbf{X}_i, Y_i), i = 1, \ldots, n$. A better method for initialization is probably the average derivative method similar to that proposed by Chaudhuri et al. [3]. However, we did not implement the average derivative method since the simpler method already seems to work well in our numerical studies.

3.1. Simulation studies

Example 1. Consider the following single index coefficient model

$$Y_i = g_0\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big) + g_1\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)Z_{i1} + g_2\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)Z_{i2} + \big(1 + \kappa|Z_{i1}|\big)e_i, \qquad i = 1, \ldots, n, \tag{2}$$

where the three nonparametric functions are $g_0(u) = 3\sin(\pi u)$, $g_1(u) = 2\cos(\pi u)$ and $g_2(u) = 4u^3$, and the true value of index parameters $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \beta_{03})^T = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^T$. $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})^T$ are independent random vectors uniformly distributed on $[-1, 1]^2$. $\mathbf{Z}_i = (Z_{i1}, Z_{i2})^T$ has a bivariate normal distribution with marginal variance 1 and correlation 0.5. The random error e_i follows standard normal distribution or Student-t distribution (degree of freedom 3) with location parameter $-q_\tau$, where q_τ is the τth quantile of the standard normal distribution or Student-t distribution with degree of freedom 3, which implies the corresponding τth quantile of e_i is zero. $\mathbf{X}_i$, $\mathbf{Z}_i$ and e_i are mutually independent. In addition, the quantity κ equals 0 or 1 corresponding to homoscedastic model (HM) and heteroscedastic model (HT), respectively. We focus on the quantile levels at $\tau = 0.1, 0.25$ and 0.5.

To evaluate the performance of the estimated index parameters $\hat{\beta}_j$, $j = 1, 2, 3$, we report the average of their bias (Bias), root mean square error (RMSE) and standard deviation estimate with three different methods, that is, empirical standard deviation estimate based on repeated simulations which serves as the gold standard (ESD), asymptotic standard deviation estimate based on asymptotic normality formula (ASD), and Bootstrap standard deviation estimate (BSD), over 1000 replications. We also consider the estimates of the coefficient functions $g_0(u)$, $g_1(u)$ and $g_2(u)$. The estimators $\hat{g}_j(\cdot)$ are assessed via the mean integrated squared errors (MISE), that is, $\mathrm{MISE} = \frac{1}{3}\sum_{j=0}^{2}\mathrm{ISE}_j$, where

$$\mathrm{ISE}_j = \frac{1}{n_{\mathrm{grid}}}\sum_{k=1}^{n_{\mathrm{grid}}}\big(\hat{g}_j(u_k) - g_j(u_k)\big)^2,$$

and $\{u_k : k = 1, \ldots, n_{\text{grid}}\}$ are regular grid points with $n_{\text{grid}} = 100$. The average values of these quantities over the 1000 replications are reported in Table 1 with different sample sizes. Figure 1(a) shows using boxplots the estimated values of index parameters with $\kappa = 1, n = 600, \tau = 0.5$. From Table 1, all biases of the estimated index parameters are close to 0 for different cases, which can also be seen from the boxplot of the index parameters in Figure 1(a). For the estimated standard deviation of the index parameters, the performance of three different methods are very similar for the homoscedastic model. BSD performs better than ASD, especially for heterogeneous models with small sample size for which ASD is quite different from ESD in some cases. Thus we prefer to use BSD, also for its straightforward implementation. Moreover, both the RMSEs of the index parameters and MISEs of the nonparametric functions become smaller as n increases. Figure 1(b)–(d) present the polynomial spline estimates of $g_0(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ for a "typical" sample, together with their 95% pointwise confidence intervals, which are obtained by pairs bootstrap. The typical sample is selected in such a way that its MISE is equal to the median in the 1000 replications. Such intervals actually ignore the bias in nonparametric estimation and should be interpreted with care but are nevertheless often used in exploring data due to its convenience. From the estimated curves together with their 95% pointwise confidence intervals in Figure 1(b)–(d), the estimated curves are visually close to the truth.

To further illustrate the importance of using a correctly specified model, we report the estimation result when fitting the generated data using the single-index model (SIM) and the partially linear single-index model (PLSIM), given by

$$\text{SIM}: \quad Y_i = g_0\big(\mathbf{X}_i^T \boldsymbol{\beta}\big) + \big(1 + \kappa|Z_{i1}|\big)e_i, \qquad i = 1, \ldots, n, \tag{3}$$

$$\text{PLSIM}: \quad Y_i = g_0\big(\mathbf{X}_i^T \boldsymbol{\beta}\big) + \alpha_1 Z_{i1} + \alpha_2 Z_{i2} + \big(1 + \kappa|Z_{i1}|\big)e_i, \qquad i = 1, \ldots, n. \tag{4}$$

For illustration purpose, we only consider comparison for the case of heteroscedastic model with normal error at $\tau = 0.5$. Table 2 shows the root mean square errors (RMSE) for the entire index parameter vector $\boldsymbol{\beta}$ and the ISE for the nonparametric function $g_0(\cdot)$. Figure 2 shows the estimated g_0 in the three different models (30 estimates are randomly chosen from 1000 replicates) with $n = 600$. As we can see from Table 2 and Figure 2, both misspecified models have much larger errors than SICM. We also compare the prediction performance of the SICM with SIM and PLSIM. We perform the leave-one-out cross-validation for models (2)–(4) with the prediction error given by $\text{CVPE} = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{Y}_i^{-(i)})^2$, where $\hat{Y}_i^{-(i)}$ is the predicted value for the ith response using the remaining $(n-1)$ observations. Table 3 lists the cross-validation prediction errors (CVPE) for heteroscedastic model with normal error at $\tau = 0.5$. From Table 3, it is clear that the value of CVPE for SICM is the smallest for each case, which means that using a correctly specified model would help to improve the prediction.

Example 2. In this example, the response observations are generated from the following model

$$Y_i = g_0\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) + g_1\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)Z_{i1} + g_2\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)Z_{i2} + \exp(\kappa Z_{i2}) \cdot e_i, \qquad i = 1, \ldots, n,$$

where $g_0(u) = 5\exp(u)$, $g_1(u) = 10\sin(2\pi u)$ and $g_2(u) = 2u^2$. The covariates $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})^T$ and $\mathbf{Z}_i = (Z_{i1}, Z_{i2})^T$ are generated as follows. We first generate the variables $\mathbf{W}_i =$

Table 1. Simulation results for Example 1, where ESD, ASD and BSD denote the empirical standard deviation estimate, the asymptotic standard deviation estimate and the Bootstrap standard deviation estimate with 200 resample at each simulation, respectively

			$\hat{\beta}_1$					$\hat{\beta}_2$					$\hat{\beta}_3$					
Model	τ	n	Bias	RMSE	ESD	ASD	BSD	Bias	RMSE	ESD	ASD	BSD	Bias	RMSE	ESD	ASD	BSD	MISE
HM-$N(0, 1)$	0.1	100	−0.0015	0.0272	0.0377	0.0326	0.0338	0.0011	0.0283	0.0386	0.0330	0.0408	−0.0032	0.0260	0.0337	0.0326	0.0364	1.9983
		200	0.0006	0.0172	0.0215	0.0229	0.0223	−0.0010	0.0169	0.0212	0.0230	0.0209	−0.0013	0.0175	0.0220	0.0227	0.0209	0.3606
		400	0.0002	0.0123	0.0154	0.0162	0.0152	0.0009	0.0120	0.0149	0.0162	0.0146	−0.0008	0.0118	0.0149	0.0161	0.0146	0.2390
		600	−0.0002	0.0091	0.0116	0.0133	0.0126	0.0002	0.0095	0.0119	0.0133	0.0122	0.0004	0.0096	0.0120	0.0132	0.0119	0.2008
	0.25	100	0.0002	0.0207	0.0260	0.0257	0.0265	−0.0004	0.0200	0.0255	0.0258	0.0299	−0.0016	0.0212	0.0266	0.0258	0.0291	1.1121
		200	0.0009	0.0134	0.0170	0.0180	0.0193	−0.0002	0.0133	0.0167	0.0180	0.0184	−0.0015	0.0132	0.0169	0.0180	0.0182	0.2898
		400	−0.0004	0.0093	0.0117	0.0129	0.0128	0.0009	0.0092	0.0115	0.0129	0.0126	−0.0007	0.0090	0.0113	0.0129	0.0128	0.1961
		600	−0.0003	0.0073	0.0093	0.0107	0.0102	0.0001	0.0074	0.0093	0.0106	0.0100	0.0000	0.0077	0.0096	0.0106	0.0101	0.1679
	0.5	100	−0.0015	0.0197	0.0247	0.0232	0.0297	0.0006	0.0183	0.0231	0.0233	0.0275	−0.0006	0.0189	0.0240	0.0234	0.0286	0.9146
		200	0.0013	0.0125	0.0158	0.0166	0.0178	0.0003	0.0124	0.0155	0.0167	0.0172	−0.0003	0.0125	0.0158	0.0166	0.0174	0.2650
		400	0.0009	0.0087	0.0110	0.0118	0.0117	0.0328	0.0089	0.0111	0.0117	0.0116	−0.0006	0.0083	0.0105	0.0118	0.0117	0.1848
		600	−0.0002	0.0066	0.0085	0.0099	0.0094	0.0004	0.0066	0.0083	0.0100	0.0093	−0.0004	0.0071	0.0089	0.0100	0.0094	0.1588
HT-$N(0, 1)$	0.1	100	−0.0007	0.0468	0.0603	0.0393	0.0560	−0.0009	0.0456	0.0613	0.0395	0.0517	−0.0075	0.0435	0.0548	0.0396	0.0493	3.4804
		200	0.0019	0.0288	0.0360	0.0277	0.0362	−0.0023	0.0266	0.0341	0.0280	0.0341	−0.0027	0.0264	0.0337	0.0278	0.0342	0.5301
		400	0.0010	0.0196	0.0245	0.0199	0.0245	0.0007	0.0179	0.0225	0.0200	0.0225	−0.0013	0.0178	0.0225	0.0198	0.0233	0.3380
		600	−0.0015	0.0150	0.0189	0.0165	0.0197	−0.0003	0.0143	0.0182	0.0167	0.0191	0.0010	0.0143	0.0178	0.0166	0.0191	0.2748
	0.25	100	0.0019	0.0339	0.0435	0.0307	0.0415	−0.0037	0.0337	0.0430	0.0310	0.0449	−0.0030	0.0315	0.0405	0.0308	0.0457	2.5141
		200	0.0016	0.0222	0.0282	0.0215	0.0313	−0.0006	0.0207	0.0260	0.0213	0.0265	−0.0008	0.0203	0.0259	0.0214	0.0281	0.4221
		400	−0.0006	0.0151	0.0189	0.0156	0.0204	0.0008	0.0139	0.0173	0.0154	0.0191	−0.0010	0.0135	0.0169	0.0155	0.0171	0.2714
		600	−0.0014	0.0119	0.0149	0.0127	0.0164	0.0003	0.0110	0.0139	0.0128	0.0156	0.0006	0.0113	0.0141	0.0127	0.0155	0.2242
	0.5	100	0.0019	0.0309	0.0392	0.0272	0.0383	−0.0027	0.0301	0.0382	0.0272	0.0371	−0.0030	0.0285	0.0360	0.0273	0.0376	1.5925
		200	0.0019	0.0201	0.0252	0.0194	0.0303	0.0014	0.0189	0.0237	0.0194	0.0243	−0.0009	0.0185	0.0235	0.0194	0.0265	0.3865
		400	0.0005	0.0140	0.0176	0.0141	0.0185	0.0009	0.0134	0.0168	0.0140	0.0171	−0.0008	0.0126	0.0159	0.0141	0.0170	0.2554
		600	−0.0009	0.0106	0.0136	0.0116	0.0152	0.0008	0.0098	0.0124	0.0117	0.0141	−0.0004	0.0106	0.0133	0.0117	0.0144	0.2107

Table 1. (Continued)

Model	τ	n	$\hat{\beta}_1$					$\hat{\beta}_2$					$\hat{\beta}_3$					
			Bias	RMSE	ESD	ASD	BSD	Bias	RMSE	ESD	ASD	BSD	Bias	RMSE	ESD	ASD	BSD	MISE
HM-t(3)	0.1	100	−0.0085	0.0460	0.0650	0.0448	0.0588	0.0037	0.0513	0.0839	0.0440	0.0542	−0.0088	0.0450	0.0658	0.0442	0.0534	2.1164
		200	−0.0025	0.0294	0.0383	0.0315	0.0376	0.0009	0.0292	0.0379	0.0313	0.0370	−0.0013	0.0292	0.0380	0.0314	0.0354	0.5979
		400	0.0011	0.0197	0.0249	0.0215	0.0253	0.0004	0.0187	0.0242	0.0214	0.0249	−0.0016	0.0189	0.0239	0.0218	0.0256	0.3877
		600	−0.0001	0.0160	0.0198	0.0176	0.0206	−0.0003	0.0160	0.0200	0.0175	0.0198	−0.0007	0.0157	0.0201	0.0175	0.0200	0.3107
	0.25	100	0.0033	0.0302	0.0384	0.0303	0.0349	−0.0024	0.0271	0.0355	0.0304	0.0330	−0.0047	0.0299	0.0388	0.0301	0.0350	1.5998
		200	−0.0010	0.0174	0.0219	0.0201	0.0213	0.0006	0.0169	0.0215	0.0202	0.0248	−0.0008	0.0170	0.0218	0.0201	0.0218	0.3719
		400	−0.0004	0.0115	0.0142	0.0143	0.0141	0.0002	0.0116	0.0146	0.0144	0.0169	−0.0004	0.0117	0.0147	0.0146	0.0146	0.2418
		600	−0.0005	0.0092	0.0116	0.0119	0.0116	−0.0005	0.0094	0.0117	0.0119	0.0127	0.0002	0.0096	0.0121	0.0119	0.0125	0.2000
	0.5	100	−0.0016	0.0241	0.0304	0.0252	0.0301	0.0012	0.0229	0.0293	0.0253	0.0375	−0.0019	0.0231	0.0300	0.0250	0.0317	1.1590
		200	0.0008	0.0148	0.0187	0.0176	0.0210	−0.0005	0.0139	0.0175	0.0176	0.0208	−0.0007	0.0143	0.0182	0.0175	0.0205	0.3094
		400	−0.0002	0.0093	0.0116	0.0125	0.0135	0.0003	0.0093	0.0116	0.0124	0.0132	−0.0004	0.0092	0.0116	0.0124	0.0120	0.2063
		600	−0.0004	0.0078	0.0098	0.0103	0.0105	−0.0004	0.0077	0.0097	0.0103	0.0105	0.0003	0.0078	0.0097	0.0103	0.0105	0.1731
HT-t(3)	0.1	100	−0.0126	0.0896	0.1210	0.0462	0.1146	0.0002	0.0899	0.1316	0.0469	0.1165	−0.0257	0.0769	0.1063	0.0471	0.1350	5.4440
		200	−0.0029	0.0525	0.0688	0.0380	0.0649	−0.0014	0.0496	0.0655	0.0378	0.0641	−0.0041	0.0478	0.0639	0.0380	0.0630	0.9217
		400	−0.0011	0.0328	0.0415	0.0255	0.0423	−0.0007	0.0294	0.0382	0.0256	0.0431	−0.0032	0.0298	0.0376	0.0256	0.0428	0.5687
		600	−0.0002	0.0267	0.0335	0.0214	0.0331	−0.0012	0.0248	0.0314	0.0214	0.0322	−0.0014	0.0251	0.0318	0.0213	0.0328	0.4453
	0.25	100	0.0000	0.0448	0.0593	0.0340	0.0616	−0.0019	0.0438	0.0580	0.0343	0.0706	−0.0065	0.0417	0.0526	0.0344	0.0737	3.1701
		200	0.0012	0.0283	0.0358	0.0242	0.0423	−0.0004	0.0263	0.0335	0.0242	0.0391	−0.0029	0.0265	0.0342	0.0246	0.0410	0.5538
		400	−0.0011	0.0184	0.0229	0.0173	0.0268	−0.0003	0.0172	0.0217	0.0174	0.0250	−0.0009	0.0179	0.0225	0.0173	0.0245	0.3443
		600	−0.0004	0.0150	0.0187	0.0143	0.0199	−0.0005	0.0142	0.0178	0.0142	0.0183	0.0002	0.0147	0.0185	0.0142	0.0189	0.2751
	0.5	100	−0.0016	0.0409	0.0530	0.0293	0.0621	−0.0011	0.0359	0.0479	0.0290	0.0600	−0.0035	0.0348	0.0450	0.0292	0.0510	2.7569
		200	0.0003	0.0241	0.0305	0.0205	0.0344	−0.0004	0.0218	0.0273	0.0206	0.0318	−0.0002	0.0220	0.0278	0.0207	0.0315	0.4571
		400	−0.0006	0.0150	0.0187	0.0148	0.0216	0.0003	0.0141	0.0175	0.0148	0.0197	−0.0005	0.0137	0.0175	0.0148	0.0200	0.2886
		600	−0.0004	0.0123	0.0155	0.0123	0.0169	−0.0005	0.0115	0.0145	0.0123	0.0161	0.0004	0.0119	0.0147	0.0123	0.0158	0.2326

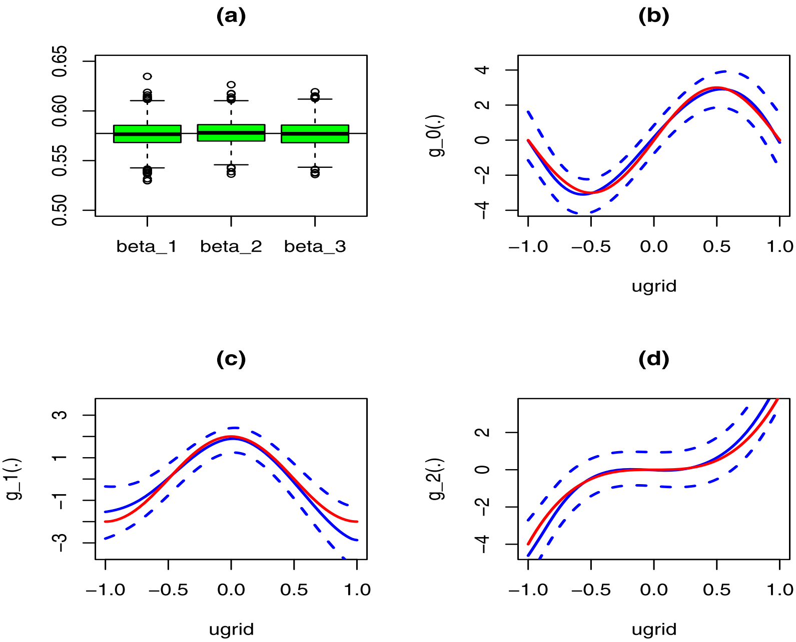

Figure 1. Boxplot of the index parameters and estimated curves of the nonparametric functions for heteroscedastic normal model with sample size $n = 600$ at $\tau = 0.5$. (a) Boxplot for index parameters; (b)–(d) Estimated curves together with their 95% pointwise confidence intervals for $g_0(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$, respectively. Red curves and blue curves represent the truth and the estimates, respectively.

$(W_{i1}, \ldots, W_{i5})^T$ from the multivariate normal distribution with mean 0 and correlation coefficients $\rho(W_{ij}, W_{ik}) = 0.8^{|j-k|}$, $1 \le j, k \le 5$, and then let $X_{ij} = \Phi(W_{ij}) - 0.5$, $j = 1, 2, 3$ and $Z_{i1} = W_{i4}$ and $Z_{i2} = W_{i5}$, where $\Phi(\cdot)$ is the accumulate distribution function of the standard normal variable. The true value of the index parameter is $\boldsymbol{\beta}_0 = (2/\sqrt{14}, 3/\sqrt{14}, 1/\sqrt{14})^T$. Settings for the random errors e_i and κ are the same as in Example 1, and e_i is independent of $\mathbf{X}_i$ and $\mathbf{Z}_i$. The main goal of this simulation is to examine whether our proposed method still performs well when the correlation between $\mathbf{X}_i$ and $\mathbf{Z}_i$ is high. In addition, we compare our procedure

Table 2. Simulation results for Example 1 for three different models with heteroscedastic normal error model at $\tau = 0.5$. The values in parentheses are the standard errors computed based on the 1000 replications

	$n = 200$		$n = 400$		$n = 600$	
Model	RMSE	MISE	RMSE	MISE	RMSE	MISE
SICM	0.0243 (0.0132)	0.3424 (0.7030)	0.0169 (0.0084)	0.1296 (0.2303)	0.0136 (0.0071)	0.0847 (0.0865)
SIM	0.0368 (0.0207)	1.5418 (1.8454)	0.0245 (0.0130)	1.2450 (1.8478)	0.0199 (0.0106)	1.0159 (1.1385)
PLSIM	0.0367 (0.0207)	1.6362 (1.9642)	0.0244 (0.0130)	1.2787 (1.8228)	0.0198 (0.0106)	1.0516 (1.1950)

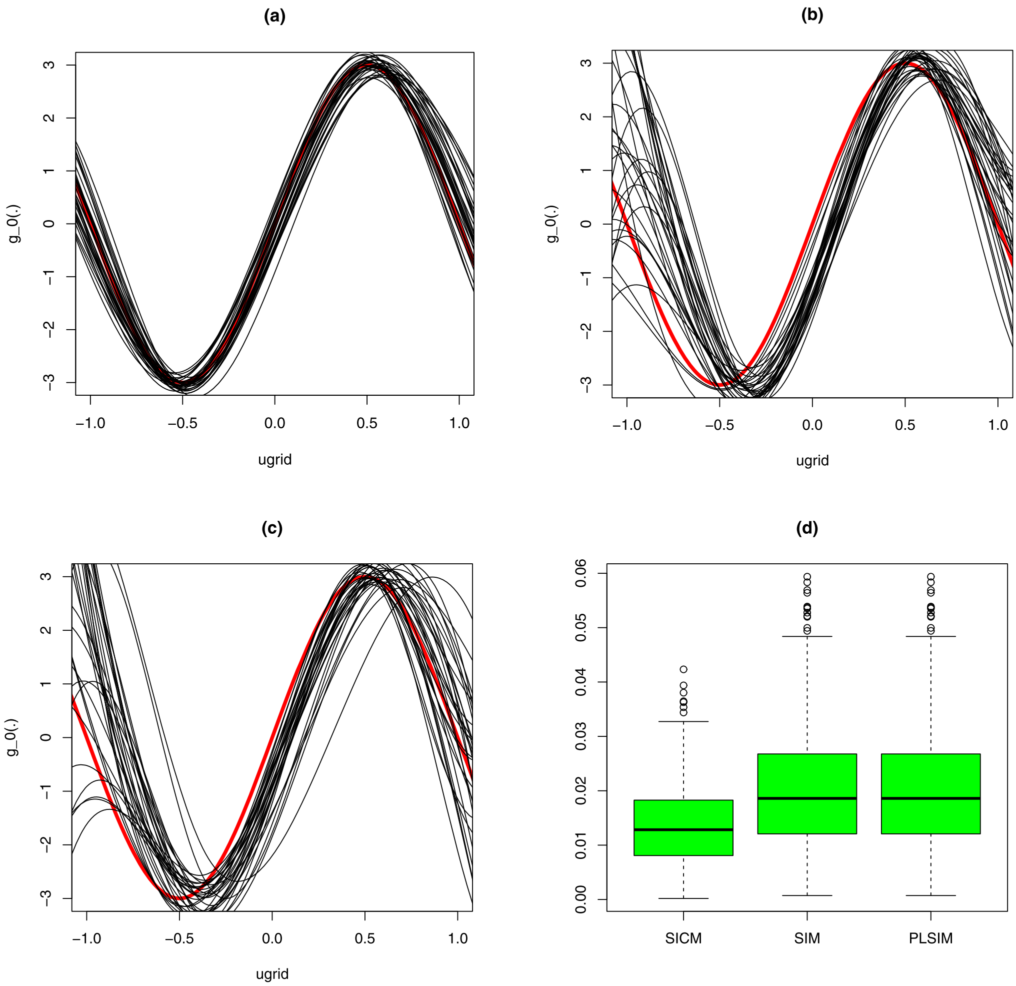

Figure 2. The median regression estimates of $g_0(\cdot)$ for three different models on generated data with heteroscedastic normal error and $n = 600$: (a) SICM; (b) SIM; (c) PLSIM. The red curve is the true function and the black curves are the estimates for 30 randomly chosen replicates. (d) Shows the boxplots of the RMSE of the index parameter.

($\tau = 0.5$) with least squares SICM (LSSICM), for which we again used polynomial splines. We also implemented a quantile version of the varying-index model of Ma and Song [22] (QRVICM) which represents a correct specification and is more flexible in allowing the index parameters for different coefficients to be different from each other. The simulation results of RMSE for $\boldsymbol{\beta}$ and MISE for nonparametric functions are shown in Table 4.

As we can see from Table 4, our proposed method still works well when the correlations between covariates are high, and both RMSE and MISE of QRSICM are smaller than least squares

Table 3. CVPE comparisons for three different models

Model	SICM	SIM	PLSIM
$n = 200$	3.9658	12.6645	12.6407
$n = 400$	2.9198	11.1792	11.1432
$n = 600$	2.7459	11.1514	11.1267

based method for the heavy-tailed $t(3)$ distribution. For the heteroscedastic model, median regression also outperforms mean regression even when the random error follows the normal distribution, which can also be seen from the boxplot of index parameters in Figure 3. In addition, since the data is generated from single-index coefficient model, the performance of QRVICM in terms of RMSE on the index parameters is inferior, although its performance in the estimation of coefficient functions is similar to QRSICM. Nevertheless, we note that QRVICM is definitely useful when the index parameters are different.

3.2. Real data analysis

In this subsection, we apply our model to an environmental data set. This data set was obtained from a respiratory study in Hong Kong, which was collected between January 1, 2000 and December 31, 2000.

In this data set, there are five daily measurements of pollutants, sulphur dioxide (in g/m^3), respirable suspended particulate (in g/m^3), nitrogen oxide (in g/m^3), nitrogen dioxide (in g/m^3) and ozone (in g/m^3), and two environmental factors, temperature (in Celsius) and relative humidity (%). The goal of the study is to examine the relationship between the levels of air pollutants and environmental factors and the daily total hospital admissions for respiratory diseases. Given measurements for multiple pollutants, it is of great interest to construct a single pollutant index to be used in predicting hospital admissions and thus a single-index structure is appropriate. For this study, we need to answer two important questions: (1) How (linear or nonlinear) the mixture of the five pollutants (pollutant index) influence of the daily total hospital admissions for respiratory diseases at different quantile levels? (2) Does the two environmental factors also have important effect on respiratory diseases at different quantile levels?

To explore whether the environmental factors have some effect on the response, in addition to the effects of the pollutants, we split the data into two subsets according to whether temperature is less than or greater than 23.8 (median temperature in the data). We then fit the single-index model with five pollutants as the predictors for the two subsets separately. The nonparametric link functions are shown in Figure 4. The difference in estimates as shown in the figure suggests that the temperature interacts with pollutants levels in a possibly complicated way.

Based on the previous discussions, we propose to use the following quantile SICM to analyze this data set:

$$\text{SICM}: \quad Y_i = g_0(\mathbf{X}_i^T\boldsymbol{\beta}) + g_1(\mathbf{X}_i^T\boldsymbol{\beta})Z_{i1} + g_2(\mathbf{X}_i^T\boldsymbol{\beta})Z_{i2} + g_3(\mathbf{X}_i^T\boldsymbol{\beta})Z_{i3} + e_i, \quad i = 1, \ldots, n, \tag{5}$$

Table 4. Simulation results for Example 2. QRSICM denotes our method at $\tau = 0.5$, LSSICM denotes the least squares method and QRVICM denotes the quantile estimate of varying-index coefficient model (Ma and Song [22]) at $\tau = 0.5$. The values in parentheses are the standard errors computed based on the 1000 replications

		$n = 200$		$n = 400$		$n = 600$	
Model	Method	RMSE	MISE	RMSE	MISE	RMSE	MISE
HM-$N(0, 1)$	QRSICM	0.0140 (0.0088)	0.2891 (0.0715)	0.0092 (0.0051)	0.2002 (0.0425)	0.0073 (0.0043)	0.1645 (0.0340)
	LSSICM	0.0118 (0.0069)	0.2857 (0.0678)	0.0078 (0.0045)	0.1989 (0.0416)	0.0063 (0.0037)	0.1629 (0.0334)
	QRVICM	0.0381 (0.0202)	0.3097 (0.0783)	0.0218 (0.0226)	0.2067 (0.0507)	0.0169 (0.0171)	0.1832 (0.0444)
HT-$N(0, 1)$	QRSICM	0.0112 (0.0074)	0.7396 (0.3185)	0.0064 (0.0041)	0.5393 (0.2095)	0.0049 (0.0030)	0.4520 (0.1601)
	LSSICM	0.0347 (0.0260)	0.7727 (0.3196)	0.0252 (0.0187)	0.5636 (0.2209)	0.0201 (0.0132)	0.4714 (0.1661)
	QRVICM	0.0588 (0.0557)	0.7637 (0.3330)	0.0510 (0.0463)	0.5602 (0.2280)	0.0423 (0.0397)	0.4670 (0.1673)
HM-$t(3)$	QRSICM	0.0154 (0.0094)	0.4635 (0.1652)	0.0106 (0.0060)	0.3146 (0.0924)	0.0082 (0.0047)	0.2593 (0.0858)
	LSSICM	0.0198 (0.0141)	0.4711 (0.1721)	0.0137 (0.0080)	0.3192 (0.0941)	0.0110 (0.0070)	0.2638 (0.0897)
	QRVICM	0.0547 (0.0500)	0.4685 (0.1740)	0.0518 (0.0479)	0.3178 (0.0889)	0.0575 (0.0476)	0.2615 (0.0833)
HT-$t(3)$	QRSICM	0.0130 (0.0087)	1.1963 (0.6769)	0.0075 (0.0046)	0.8713 (0.5222)	0.0054 (0.0033)	0.7022 (0.3088)
	LSSICM	0.0590 (0.0571)	1.2683 (0.7622)	0.0420 (0.0366)	0.9179 (0.5505)	0.0334 (0.0264)	0.7388 (0.3251)
	QRVICM	0.0640 (0.0523)	1.2183 (0.6779)	0.0545 (0.0554)	0.8783 (0.5141)	0.0468 (0.0472)	0.7272 (0.3203)

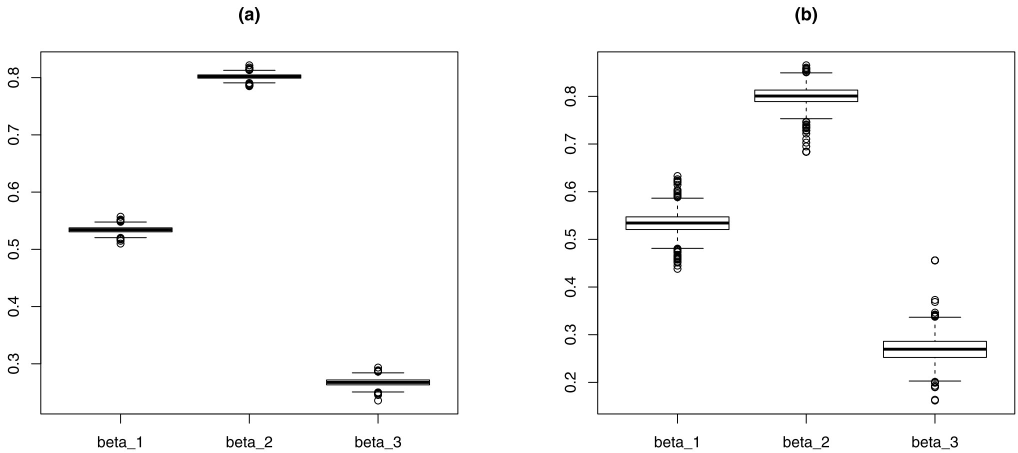

Figure 3. Boxplot of the index parameters with sample size $n = 600$ for the heteroscedastic normal error model: (a) QRSICM; (b) LSSICM.

where $\mathbf{X}_i = (X_{i1}, \ldots, X_{i5})^T$ denotes the five pollutants, and Z_{i1} and Z_{i2} denote the two environmental factors, that is, temperature and humidity. In addition, $Z_{i3} = Z_{i1} \times Z_{i2}$ denotes the interaction of the two environmental factors. The environment factors are standardized to have mean 0 and variance 1.

Table 5 lists the estimated index parameters at three different quantile levels, together with their lower bound (LB) and upper bound (UB) of 95% confidence intervals (CI), which are

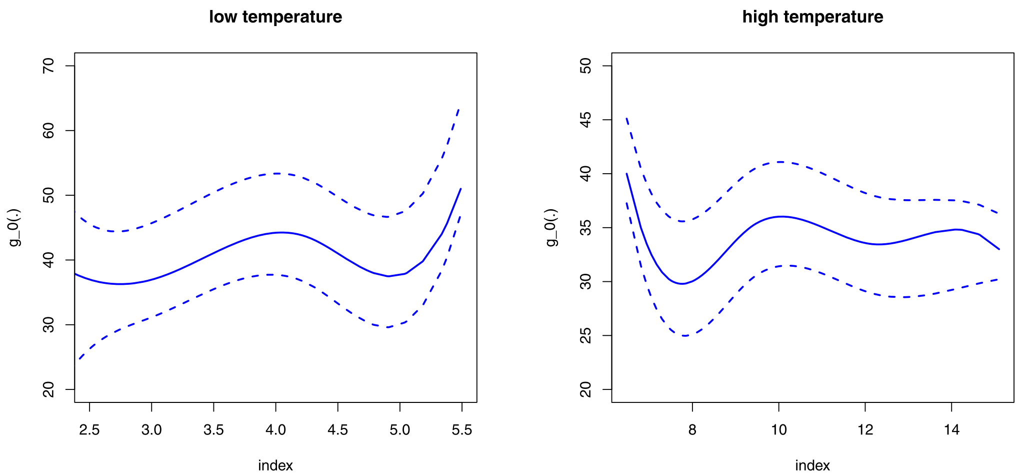

Figure 4. Estimated nonparametric link function together with their 95% pointwise confidence bands using the single-index model for the environmental data at $\tau = 0.5$: (a) Low temperature; (b) High temperature.

Table 5. The estimates (EST), lower bound (LB) and upper bound (UB) of 95% confidence interval at different quantile levels

τ		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
0.25	EST	0.7018	0.1300	0.2088	0.0958	0.6616
	LB	0.6649	0.0927	0.1612	0.0284	0.6213
	UB	0.7388	0.1673	0.2564	0.1632	0.7019
0.5	EST	0.5511	0.2850	0.1481	0.3054	0.7069
	LB	0.4781	0.2240	0.0778	0.1886	0.6454
	UB	0.6242	0.3461	0.2184	0.4223	0.7684
0.75	EST	0.8056	0.3493	0.0417	0.2088	0.4286
	LB	0.7791	0.2922	-0.0213	0.0870	0.4031
	UB	0.8320	0.4064	0.1046	0.3306	0.4541

obtained by the asymptotic result in Theorem 2. From Table 5, all the five pollutant variables are important at the significant level 0.05 except the variable X_3 at $\tau = 0.75$.

Furthermore, the estimated nonparametric curves together with their 95% pointwise confidence interval $\tau = 0.25, 0.5$ and 0.75 are given in Figures 5–7. Simultaneous confidence bands are also shown. The simultaneous bands are obtained by enlarging the pointwise confidence intervals until the band contains 95% of the bootstrap samples.

From Figures 5 and 6, the 95% confidence bands cannot completely contain the zero lines, which indicates that interactions between environmental factors and pollutants have important effects on the number of hospital admissions. However, for higher quantile $\tau = 0.75$ this effect seems to be weak.

We next compare the prediction performance of the SICM (5) with the single-index model (SIM) and the partially linear single-index model (PLSIM) at quantile $\tau = 0.5$.

$$\text{SIM}: \quad Y_i = g_0\big(\mathbf{X}_i^T \boldsymbol{\beta}\big) + e_i, \qquad i = 1, \ldots, n, \tag{6}$$

$$\text{PLSIM}: \quad Y_i = g_0\big(\mathbf{X}_i^T \boldsymbol{\beta}\big) + \alpha_1 Z_{i1} + \alpha_2 Z_{i2} + \alpha_3 Z_{i3} + e_i, \qquad i = 1, \ldots, n. \tag{7}$$

We perform the leave-one-out cross-validation for models (5)–(7) with the prediction error given by CVPE. Table 6 lists the CVPE for different models, where we also list the result of least squares based SICM method. From Table 6, it is clear that the value of CVPE for SICM is the smallest, which means that incorporating the interactions between the environmental factors and pollutants improves prediction.

4. Conclusion and discussion

In this paper, we have considered quantile regression version of single-index coefficient models. We established asymptotic properties of polynomial spline estimation method, in particular demonstrating asymptotic normality of the index parameter. Our numerical results suggest that the quantile version is useful in data modelling.

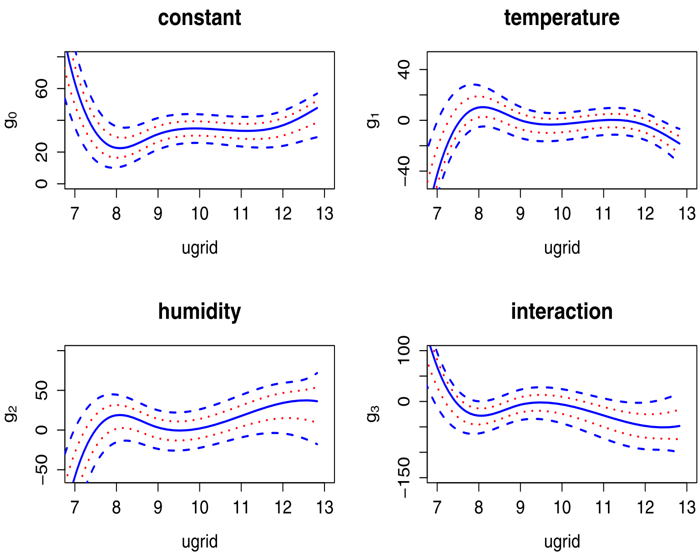

Figure 5. Estimated curves together with their 95% pointwise confidence interval (red dotted line) and 95% simultaneously confidence bands (blue dashed line) for $g_0(\cdot)$, $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ at $\tau = 0.25$ for the environmental data.

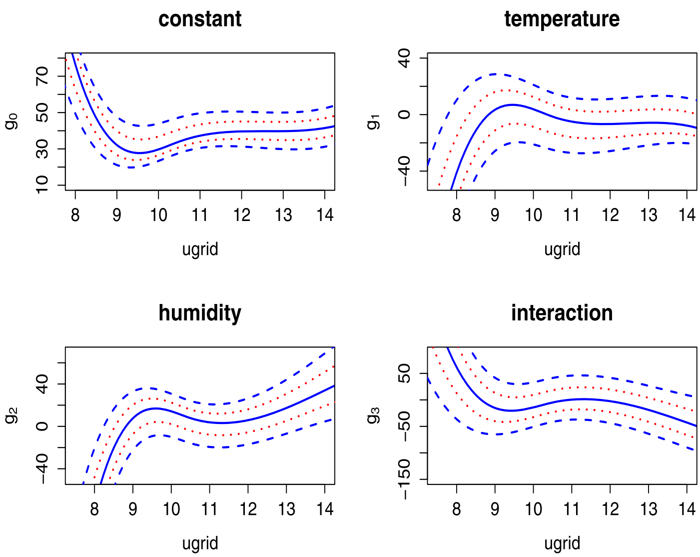

Figure 6. Estimated curves together with their 95% pointwise confidence interval (red dotted line) and 95% simultaneously confidence bands (blue dashed line) for $g_0(\cdot)$, $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ at $\tau = 0.5$ for the environmental data.

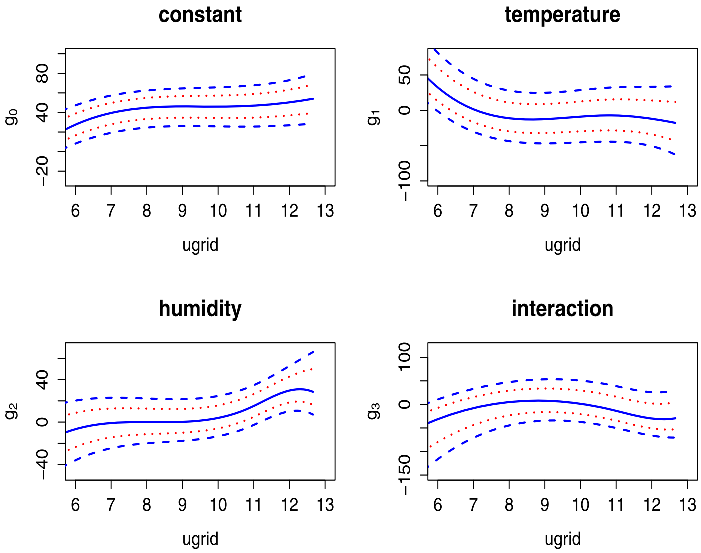

Figure 7. Estimated curves together with their 95% pointwise confidence interval (red dotted line) and 95% simultaneously confidence bands (blue dashed line) for $g_0(\cdot)$, $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ at $\tau = 0.75$ for the environmental data.

Various extensions can be considered in the future. We expect the theoretical results can be adapted to adaptive varying-coefficient linear model where $\mathbf{X}$ and $\mathbf{Z}$ are the same or the varying index coefficient model (Fan, Yao and Cai [6], Ma and Song [22]). We can consider the case of longitudinal data, for example using the approach of generalized estimating equations. Variable selection (to identify nonzero coefficient functions) can be incorporated easily by using a sparsity penalty. Furthermore, in recent years, the varying-coefficient models has been applied to high-dimensional data settings with large p (Wei, Huang and Li [28], Xue and Qu [31], Lian [20]), which is also a direction we can pursue for SICM. Finally, our environment data set is a time series data while our theory is developed under the i.i.d. setting, thus it is interesting to see if the theory can be extended to dependent observations under mixing conditions.

Appendix: Technical proofs

Let $\boldsymbol{\theta}_{0j}$ be spline coefficients in the best spline approximation of g_j with $\sup_t |g_j(t) - \mathbf{B}^T(t)\boldsymbol{\theta}_{0j}| \le CK^{-d}$, which is possible by (A3). Let $F(\cdot|\mathbf{X}, \mathbf{Z})$ be the conditional c.d.f. of e given

Table 6. CVPE of the four models for the environmental data at $\tau = 0.5$

Model	SIM	PLSIM	SICM(QR)	SICM(LS)
CVPE	50.9254	47.2058	39.8210	41.8040

the covariates. We also write the true conditional quantile $\mathbf{g}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}_i$ as m_i. In the proofs C denotes a generic positive constant which may assume different values even on the same line.

Lemma 1. *Let* $r_n = \sqrt{K/n} + K^{-d}$.

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|+\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le Cr_n} \sum_{i=1}^n \rho_\tau\left(Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j\right) - \sum_{i=1}^n \rho_\tau\left(Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\right)$$
$$+\sum_{i=1}^n\sum_{j=1}^p\big(Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j - Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\big)\big(\tau - I\{e_i\le 0\}\big)$$
$$- E\sum_{i=1}^n \rho_\tau\left(Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j\right) + E\sum_{i=1}^n \rho_\tau\left(Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\right)$$
$$= o_p\big(nr_n^2\big),$$

where the expectations are over Y_i *conditional on* $\mathbf{X}_i, \mathbf{Z}_i$ (*all expectations below are also such conditional expectations*).

Proof. As in He and Shi [11], in the proof we consider median regression with $\tau = 1/2$, $\rho_\tau(u) = |u|/2$ and the general case can be shown in the same way. Let $\mathcal{N} = \{(\boldsymbol{\beta}^{(1)}, \boldsymbol{\theta}^{(1)}), \dots, (\boldsymbol{\beta}^{(N)}, \boldsymbol{\theta}^{(N)})\}$ be a δ_n covering of $\{(\boldsymbol{\beta}, \boldsymbol{\theta}) : \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| + \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \le Cr_n\}$, with size bounded by $N \le (Cr_n/\delta_n)^{CK}$ and thus $\log N \le CK\log n$ if we choose $\delta_n \sim n^{-a}$ for some $a > 0$ (we will choose a to be large enough).

Let $M_{ni}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{2}|Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j| - |Y_i - \frac{1}{2}\sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}| + \sum_{j=1}^p(Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j) - Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j})(1/2 - I\{e_i \le 0\})$, and $M_n(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^n M_{ni}(\boldsymbol{\beta}, \boldsymbol{\theta})$. Using the Lipschitz property of $|u|$, and that for any $(\boldsymbol{\beta}, \boldsymbol{\theta})$ there exists $(\boldsymbol{\beta}^{(l)}, \boldsymbol{\theta}^{(l)})$ such that $\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(l)}\|^2 + \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(l)}\|^2 \le \delta_n^2$, we have

$$M_n(\boldsymbol{\beta}, \boldsymbol{\theta}) - EM_n(\boldsymbol{\beta}, \boldsymbol{\theta}) - M_n\big(\boldsymbol{\beta}^{(l)}, \boldsymbol{\theta}^{(l)}\big) + EM_n\big(\boldsymbol{\beta}^{(l)}, \boldsymbol{\theta}^{(l)}\big)$$
$$\le C\sum_{i=1}^n\sum_{j=1}^p\big|\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j - \mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}^{(l)})\boldsymbol{\theta}_j^{(l)}\big|,$$

which can obviously be made smaller than nr_n^2 by the Lipschitz property of the spline functions, by setting $\delta_n \sim n^{-a}$ for a large enough.

Furthermore, by simple algebra

$$\big|M_{ni}(\boldsymbol{\beta}, \boldsymbol{\theta})\big| = \left|\frac{1}{2}\left|Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j\right| - \frac{1}{2}\left|Y_i - \sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\right|\right.$$
$$\left. + \sum_{j=1}^p\big(Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j - Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\big)\big(1/2 - I\{e_i \le 0\}\big)\right|$$

$$
\begin{aligned}
&= \left| \frac{1}{2} \left| e_i + m_i - \sum_{j=1}^{p} Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big) \boldsymbol{\theta}_j \right| - \frac{1}{2} \left| e_i + m_i - \sum_{j=1}^{p} Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \right| \right. \\
&\quad \left. + \sum_{j=1}^{p} \big(Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big) \boldsymbol{\theta}_j - Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \big) \big(1/2 - I\{e_i \le 0\}\big) \right| \\
&\le \left| \sum_{j=1}^{p} Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big) \boldsymbol{\theta}_j - Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \right| \\
&\quad \times I \left\{ |e_i| \le \left| \sum_{j=1}^{p} Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big) \boldsymbol{\theta}_j - Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \right| \right. \\
&\quad \left. + \left| m_i - \sum_{j=1}^{p} Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \right| \right\}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\big| M_{ni}(\boldsymbol{\beta}, \boldsymbol{\theta}) \big| &\le \left| \sum_{j=1}^{p} Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big) \boldsymbol{\theta}_j - Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \right| \\
&\le C \sum_j \big| \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}^*\big) \boldsymbol{\theta}_j \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \big| + \big| \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) (\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j}) \big| \\
&\le C \sum_j \big| \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}^*\big) \boldsymbol{\theta}_{0j} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \big| + \big| \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) (\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j}) \big| \\
&\quad + \big| \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}^*\big) (\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j}) \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \big| \\
&\le C \big(r_n + \sqrt{K} r_n + K^{3/2} r_n^2 \big) \\
&\le C \sqrt{K} r_n \\
&=: A,
\end{aligned}
$$

where we used that $\|\mathbf{B}(x)\| \le C\sqrt{K}$ and $\|\mathbf{B}^{(1)}(x)\| \le C K^{3/2}$ at any fixed point $x \in [a, b]$.

Furthermore, we have

$$
\begin{aligned}
E \big| M_{ni}(\boldsymbol{\beta}, \boldsymbol{\theta}) \big|^2 &\le C(\sqrt{K} r_n) \sum_j E \big| \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big) \boldsymbol{\theta}_j - \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big) \boldsymbol{\theta}_{0j} \big|^2 \\
&\le C(\sqrt{K} r_n) \big(r_n^2 \big) =: D^2.
\end{aligned} \tag{8}
$$

Using Bernstein's inequality, together with union bound, we have

$$P\Big(\sup_{(\boldsymbol{\beta},\boldsymbol{\theta})\in\mathcal{N}}\big|M_n(\boldsymbol{\beta},\boldsymbol{\theta})-EM_n(\boldsymbol{\beta},\boldsymbol{\theta})\big|>a\Big)\leq C\exp\bigg\{-\frac{a^2}{aA+nD^2}-CK\log n\bigg\}.$$

The right-hand side converges to zero with $a=O(\max\{K^{3/2}r_n\log n,\sqrt{nK^{3/2}r_n^2\log n}\})=o(nr_n^2)$ (here we use the assumption that $K^{d+3/2}\log n/n\to 0$). $\square$

Now to show the convergence rate of the estimator, suppose $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|+\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=Lr_n$ for sufficiently large $L>0$. In the following, for simplicity of notation, m_i denotes the true conditional quantile $\mathbf{g}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}_i$.

Lemma 2.

$$\begin{aligned}&\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|+\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=Lr_n}\sum_i E\rho_\tau\Bigg(e_i+m_i-\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}\big)\boldsymbol{\theta}_j\Bigg)\\&-\sum_i E\rho_\tau\Bigg(e_i+m_i-\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j}\Bigg)\\&\geq L^2Cnr_n^2.\end{aligned}$$

Proof. Using the Knight's identity $\rho_\tau(x-y)-\rho_\tau(x)=-y(\tau-I\{x\leq 0\})+\int_0^y(I\{x\leq t\}-I\{x\leq 0\})\,dt$, we have that

$$\begin{aligned}&E\sum_{i=1}^n\rho_\tau\Bigg(e_i+m_i-\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}\big)\boldsymbol{\theta}_j\Bigg)-E\sum_{i=1}^n\rho_\tau\Bigg(e_i+m_i-\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j}\Bigg)\\&=\sum_i\int_{\sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}-m_i}^{\sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j-m_i}F(t|\mathbf{X}_i,\mathbf{Z}_i)-F(0|\mathbf{X}_i,\mathbf{Z}_i)\,dt\\&\geq C\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)\Bigg[\Bigg(\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}\big)\boldsymbol{\theta}_j-\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j}\Bigg)^2\\&\quad+2\Bigg(\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}\big)\boldsymbol{\theta}_j-\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j}\Bigg)\\&\quad\times\Bigg(\sum_{j=1}^p Z_{ij}\mathbf{B}^T\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j}-m_i\Bigg)\Bigg].\end{aligned}$$

We have, by Taylor's expansion,

$$\begin{aligned}
&\sum_i \left(\sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta} \right) \boldsymbol{\theta}_j - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \boldsymbol{\theta}_{0j} \right)^2 \\
&\quad \geq C \sum_i \left(\sum_{j=1}^p Z_{ij} \mathbf{B}^{(1)T} \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \boldsymbol{\theta}_{0j} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_{0j}) \right)^2 \\
&\qquad + o_p\left(n r_n^2 \right) \\
&\quad \geq C L^2 n r_n^2.
\end{aligned}$$

In the above we need the eigenvalue property as in Lemma 3 below. On the other hand, as in (8) we have similar upper bound

$$\sum_i \left(\sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta} \right) \boldsymbol{\theta}_j - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \boldsymbol{\theta}_{0j} \right)^2 \leq C L^2 n r_n^2,$$

and by the approximation property of splines,

$$\sum_i \left(\sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \boldsymbol{\theta}_{0j} - m_i \right)^2 \leq C n K^{-2d}. \tag{9}$$

Combining various bounds above, we get

$$\begin{aligned}
&E \sum_{i=1}^n \rho_\tau \left(e_i + m_i - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta} \right) \boldsymbol{\theta}_j \right) \\
&\quad - E \rho_\tau \left(e_i + m_i - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \boldsymbol{\theta}_{0j} \right) \\
&\quad \geq C L^2 n r_n^2,
\end{aligned}$$

if L is large enough. □

Lemma 3. *The eigenvalues of*

$$\frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \left(\mathbf{Z}_i \otimes \mathbf{B}^{(1)} \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \right)^T \boldsymbol{\theta}_0 \mathbf{X}_i \\ \mathbf{Z}_i \otimes \mathbf{B}\left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \end{pmatrix} \left(\left(\mathbf{Z}_i \otimes \mathbf{B}^{(1)} \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \right)^T \boldsymbol{\theta}_0 \mathbf{X}_i^T, \mathbf{Z}_i^T \otimes \mathbf{B}^T \left(\mathbf{X}_i^T \boldsymbol{\beta}_0 \right) \right)$$

are bounded and bounded away from zero with probability approaching one.

Proof. By law of large numbers, we only need to show that the matrix

$$E\left[\begin{pmatrix}\left(\mathbf{Z}\otimes\mathbf{B}^{(1)}(\mathbf{X}^T\boldsymbol{\beta}_0)\right)^T\boldsymbol{\theta}_0\mathbf{X}\\ \mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0)\end{pmatrix}\left(\left(\mathbf{Z}\otimes\mathbf{B}^{(1)}(\mathbf{X}^T\boldsymbol{\beta}_0)\right)^T\boldsymbol{\theta}_0\mathbf{X}^T,\mathbf{Z}^T\otimes\mathbf{B}^T(\mathbf{X}^T\boldsymbol{\beta}_0)\right)\right] \tag{10}$$

has eigenvalues bounded away from zero and infinity.

In turn, since $|(\mathbf{Z}\otimes\mathbf{B}^{(1)}(\mathbf{X}^T\boldsymbol{\beta}_0))^T\boldsymbol{\theta}_0-\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}|\leq CK^{-d+1}$, we only need to show that the eigenvalues of

$$E\left[\begin{pmatrix}\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}\\ \mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0)\end{pmatrix}\left(\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}^T,\mathbf{Z}^T\otimes\mathbf{B}^T(\mathbf{X}^T\boldsymbol{\beta}_0)\right)\right] \tag{11}$$

are bounded and bounded away from zero.

Under condition (A4), let $\boldsymbol{\gamma}_0$ be the $pK\times q$ matrix of spline coefficients with $\|E_{\mathcal{M}}[\mathbf{g}^{(1)T}\times(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}]-\boldsymbol{\gamma}_0^T(\mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0))\|\leq CK^{-d'}$.

Pre-multiplying (11) by

$$\begin{pmatrix}\mathbf{I} & -\boldsymbol{\gamma}_0^T\\ \mathbf{0} & \mathbf{I}\end{pmatrix} \tag{12}$$

and post-multiplying (11) by its transpose we get the matrix

$$E\left[\begin{pmatrix}\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}-\boldsymbol{\gamma}_0^T(\mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0))\\ \mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0)\end{pmatrix}^{\otimes 2}\right]. \tag{13}$$

It is easy to see that singular values of (12) are bounded and bounded away from zero, and thus we only need to show that the eigenvalues of (13) are bounded and bounded away from zero. Since $\|E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}]-\boldsymbol{\gamma}_0^T(\mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0))\|\leq CK^{-d'}$, replacing $\boldsymbol{\gamma}_0^T(\mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0))$ with $E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}]$ in (13) only makes a difference of order $o(1)$. Also noting that $f(0|\mathbf{X},\mathbf{Z})$ is bounded the bounded away from zero, we only need to consider

$$E\left[f(0|\mathbf{X},\mathbf{Z})\begin{pmatrix}\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}-E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}]\\ \mathbf{Z}\otimes\mathbf{B}(\mathbf{X}^T\boldsymbol{\beta}_0)\end{pmatrix}^{\otimes 2}\right]. \tag{14}$$

By the definition of the projection, $E[f(0|\mathbf{X},\mathbf{Z})(\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}-E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}])\times(\mathbf{Z}^T\otimes\mathbf{B}^T(\mathbf{X}^T\boldsymbol{\beta}_0))]=\mathbf{0}$ and (14) is block-diagonal and the eigenvalues of both blocks are bounded and bounded away from zero, by the property of splines and condition (A5). $\square$

The following lemma deals with one of the terms in the statement of Lemma 1.

Lemma 4.

$$\begin{aligned}\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|+\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=Lr_n}\sum_i\left(\sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_j-\sum_{j=1}^p Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\right)&\\ \times\left(\tau-I\{e_i\leq 0\}\right)=L\cdot O_p\left(nr_n^2\right).&\end{aligned}$$

Proof. For simplicity of notation, let $\varepsilon_i = \tau - I\{e_i \le 0\}$. We have

$$
\begin{aligned}
&\sum_i \Bigg(\sum_{j=1}^p Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big)\boldsymbol{\theta}_j - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j} \Bigg)\varepsilon_i \\
&\quad = \sum_i \sum_j Z_{ij} \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\varepsilon_i & (15)\\
&\qquad + \sum_i \sum_j Z_{ij} \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j})\mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\varepsilon_i & (16)\\
&\qquad + \sum_i \sum_j Z_{ij} \big(\mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}^*\big) - \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\big)\boldsymbol{\theta}_{0j} \mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\varepsilon_i & (17)\\
&\qquad + \sum_i \sum_j Z_{ij} \big(\mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}^*\big) - \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\big)(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j})\mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\varepsilon_i & (18)\\
&\qquad + \sum_i \sum_j Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j})\varepsilon_i. & (19)
\end{aligned}
$$

The term (15) obviously has order $L \cdot O_p(\sqrt{n} r_n)$. For (19), we have that $\|\mathbf{B}(\mathbf{X}_i^T \boldsymbol{\beta}_0)\varepsilon_i\|^2 = O_p(\sum_i \|\mathbf{B}(\mathbf{X}_i^T \boldsymbol{\beta}_0)\|^2) = O_p(nK)$ and thus (19) is $O_p(\sqrt{nK}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|) = L \cdot O_p(\sqrt{nK} r_n)$.

For the term (16), since $\| \sum_i \mathbf{B}^{(1)}(\mathbf{X}_i^{(1)T} \boldsymbol{\beta}_0)\varepsilon_i\|^2 = O_p(\sum_i \|\mathbf{B}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_0)\|^2) = O_p(nK^3)$ we have

$$
\sum_i \sum_j Z_{ij} \mathbf{B}^{(1)T} \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j})\mathbf{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\varepsilon_i = O_p\big(\sqrt{n} K^{3/2} r_n^2\big) = o_p\big(n r_n^2\big).
$$

With further Taylor expansion $\mathbf{B}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}^*) - \mathbf{B}^{(1)}(\mathbf{X}_i^T \boldsymbol{\beta}_0) = \mathbf{B}^{(2)}(\mathbf{X}_i^T \boldsymbol{\beta}^{**})\mathbf{X}_i^T (\boldsymbol{\beta}^* - \boldsymbol{\beta}_0)$, (17) and (18) are also of order $o_p(n r_n^2)$ and the proof is complete. □

Proof of Theorem 1. Combining Lemmas 1,2,4, we get for sufficiently large $L > 0$,

$$
\begin{aligned}
P\Bigg(&\inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = L r_n} \sum_{i=1}^n \rho_\tau \Bigg(Y_i - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}\big)\boldsymbol{\theta}_j \Bigg) \\
&> \sum_{i=1}^n \rho_\tau \Bigg(Y_i - \sum_{j=1}^p Z_{ij} \mathbf{B}^T \big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j} \Bigg)\Bigg) \to 1,
\end{aligned}
$$

and thus there is a local minimizer of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ with $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(r_n)$. □

Now we consider asymptotic normality. The challenge here is the need to perform orthogonalization appropriately. Since the parametric part is nested within the spline basis, orthogonalization is more complicated than partially linear models as studied in Wang, Zhu and Zhou [25], Wang et al. [26].

The appropriate projection $E_{\mathcal{M}}[W]$ is as defined previously. It would be clear from the proof of Lemma 6 below why it is important to use the density $f(0|\mathbf{X}, \mathbf{Z})$ when defining the projection. Let $\mathbf{\Pi}_i = \mathbf{Z}_i \otimes \mathbf{B}(\mathbf{X}_i^T \boldsymbol{\beta}_0)$ and $\mathbf{\Pi}$ be the $n \times (pK)$ matrix with rows $\mathbf{\Pi}_i^T$. The empirical version of the minimization problem corresponding to the projection is

$$\min_{\boldsymbol{\theta}} \sum_i f(0|\mathbf{X}_i, \mathbf{Z}_i)\big(W_i - \mathbf{\Pi}_i^T \boldsymbol{\theta}\big)^2,$$

with the minimizer $(\mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{\Pi})^{-1} \mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{W}$ where $\mathbf{\Gamma}$ is the diagonal matrix with diagonal elements $f(0|\mathbf{X}_i, \mathbf{Z}_i)$ and $\mathbf{W} = (W_1, \dots, W_n)^T$. Define $\mathbf{P} = \mathbf{\Pi}(\mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{\Pi})^{-1} \mathbf{\Pi}^T \mathbf{\Gamma}$.

We write

$$\begin{aligned}
&\rho_\tau\Big(e_i + m_i - \sum_j Z_{ij} \mathbf{B}\big(\mathbf{X}_i^T \boldsymbol{\beta}\big)\boldsymbol{\theta}_j\Big) \\
&= \rho_\tau\Big(e_i - \sum_j Z_{ij} \mathbf{B}^T\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)(\boldsymbol{\theta}_j - \boldsymbol{\theta}_{0j}) \\
&\quad - \sum_j Z_{ij} \mathbf{B}^{(1)T}\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j} \mathbf{X}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta}, \boldsymbol{\theta})\Big) \\
&= \rho_\tau\big(e_i - \mathbf{\Pi}_i^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \mathbf{U}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta}, \boldsymbol{\theta})\big),
\end{aligned}$$

where we defined $\mathbf{U}_i = \sum_j Z_{ij} \mathbf{B}^{(1)T}(\mathbf{X}_i^T \boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j} \mathbf{X}_i$ and

$$\begin{aligned}
&R_i(\boldsymbol{\beta}, \boldsymbol{\theta}) \\
&= \sum_j Z_{ij}\big(\mathbf{B}\big(\mathbf{X}_i^T \boldsymbol{\beta}\big) - \mathbf{B}\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\big)^T \boldsymbol{\theta}_j - \sum_j Z_{ij} \mathbf{B}^{(1)T}\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j} \mathbf{X}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\
&\quad + \Big(\sum_j Z_{ij} \mathbf{B}^T\big(\mathbf{X}_i^T \boldsymbol{\beta}_0\big)\boldsymbol{\theta}_{0j} - m_i\Big) \\
&=: R_{i1}(\boldsymbol{\beta}, \boldsymbol{\theta}) + R_{i2}(\boldsymbol{\beta}, \boldsymbol{\theta}).
\end{aligned}$$

Let $\mathbf{V} = \mathbf{U} - \mathbf{PU}$ with the ith row of $\mathbf{V}$ denoted by $\mathbf{V}_i^T = \mathbf{U}_i^T - \mathbf{P}_i^T \mathbf{U}$. To carry out orthogonalization we further write

$$\begin{aligned}
&\rho_\tau\big(e_i - \mathbf{\Pi}_i^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \mathbf{U}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta}, \boldsymbol{\theta})\big) \\
&= \rho_\tau\big(e_i - \mathbf{\Pi}_i^T\big((\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \big(\mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{\Pi}\big)^{-1} \mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{U}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\big) - \mathbf{V}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta}, \boldsymbol{\theta})\big) \\
&= \rho_\tau\big(e_i - \mathbf{\Pi}_i^T \boldsymbol{\eta} - \mathbf{V}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta}, \boldsymbol{\theta})\big),
\end{aligned}$$

with $\boldsymbol{\eta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0 + (\mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{\Pi})^{-1} \mathbf{\Pi}^T \mathbf{\Gamma} \mathbf{U}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. The proof of the following lemma is similar to Lemma 1, but requires finer analysis of $R_i(\boldsymbol{\beta}, \boldsymbol{\theta})$.

Lemma 5.

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le C/\sqrt{n},\|\boldsymbol{\eta}\|\le Cr_n}\bigg|\sum_i \rho_\tau\big(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})\big)$$
$$-\sum_i \rho_\tau\big(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big)+\sum_i \mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\varepsilon_i$$
$$-E\sum_i \rho_\tau\big(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})\big)+E\sum_i \rho_\tau\big(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big)\bigg|$$
$$=o_p(1).$$

Proof. First, we note $R_{i2}(\boldsymbol{\beta},\boldsymbol{\theta})$ is the same as $R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})$. As before, we assume $\tau=1/2$ for this proof. Note that the mathematical expressions in the statement of the lemma involves both $R_i(\boldsymbol{\beta},\boldsymbol{\theta})$ and $R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})$ which makes the proof more messy compared to varying coefficient partially linear models where $R_i(\boldsymbol{\beta},\boldsymbol{\theta})=R_{i2}(\boldsymbol{\beta},\boldsymbol{\theta})$. We have

$$\begin{aligned}
R_{i1}(\boldsymbol{\beta},\boldsymbol{\theta}) &= \sum_j Z_{ij}\big(\mathbf{B}(\mathbf{X}_i^T\boldsymbol{\beta})-\mathbf{B}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\big)^T(\boldsymbol{\theta}_j-\boldsymbol{\theta}_{0j})\\
&\quad -\sum_j Z_{ij}\big(\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta})\boldsymbol{\theta}_{0j}-\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}-\mathbf{B}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\mathbf{X}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big)\\
&= \sum_j Z_{ij}\mathbf{B}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}^*)(\boldsymbol{\theta}_j-\boldsymbol{\theta}_{0j})\big(\mathbf{X}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big)\\
&\quad -\sum_j Z_{ij}\mathbf{B}^{(2)T}(\mathbf{X}_i^T\boldsymbol{\beta}^*)\boldsymbol{\theta}_{0j}\big(\mathbf{X}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big)^2\\
&= \sum_j Z_{ij}\mathbf{B}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)(\boldsymbol{\theta}_j-\boldsymbol{\theta}_{0j})\big(\mathbf{X}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big)\\
&\quad +\sum_j Z_{ij}\mathbf{B}^{(2)T}(\mathbf{X}_i^T\boldsymbol{\beta}^{**})(\boldsymbol{\theta}_j-\boldsymbol{\theta}_{0j})\big(\mathbf{X}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big)^2\\
&\quad -\sum_j Z_{ij}\mathbf{B}^{(2)T}(\mathbf{X}_i^T\boldsymbol{\beta}^*)\boldsymbol{\theta}_{0j}\big(\mathbf{X}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big)^2.
\end{aligned}$$

It is easy to see that $|R_{i1}|\le C\sqrt{K^3/n}r_n$ and $\sum_i R_{i1}^2=O_p(r_n^2K^2+r_n^2K^4/n+1/n)=O_p(r_n^2K^2)$.

For fixed $\boldsymbol{\eta},\boldsymbol{\beta}$, let $M_{ni}(\boldsymbol{\beta},\boldsymbol{\eta})=\frac{1}{2}|e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})|-\frac{1}{2}|e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})|+(\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0+R_{i1}(\boldsymbol{\beta},\boldsymbol{\theta})))(1/2-I\{e_i\le 0\})$, we have

$$\begin{aligned}
&\big|M_{ni}(\boldsymbol{\beta},\boldsymbol{\eta})\big|\\
&\quad\le\big|\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+R_{i1}(\boldsymbol{\beta},\boldsymbol{\theta})\big|
\end{aligned}$$

$$\times I\left\{|e_i| \le \left|\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+R_{i1}(\boldsymbol{\beta},\boldsymbol{\theta})\right|+\left|\boldsymbol{\Pi}_i^T\boldsymbol{\eta}+R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\right|\right\}$$
$$\le C\left(1/\sqrt{n}+\sqrt{K^3/n}r_n\right)$$

and

$$E\left|M_{ni}(\boldsymbol{\beta},\boldsymbol{\eta})\right|^2 \le \left(1/n+K^2r_n^2/n\right)(\sqrt{K}r_n).$$

Similar to the proof of Lemma 1, by constructing a δ_n-covering and using Bernstein's inequality with union bound, we can show that

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le C/\sqrt{n},\|\boldsymbol{\eta}\|\le Cr_n}\Bigg|\sum_i \rho_\tau\left(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})\right)$$
$$-\sum_i \rho_\tau\left(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\right)+\sum_i\left(\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+R_{i1}(\boldsymbol{\beta},\boldsymbol{\theta})\right)\varepsilon_i$$
$$-E\sum_i \rho_\tau\left(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})\right)+E\sum_i \rho_\tau\left(e_i-\boldsymbol{\Pi}_i^T\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\right)\Bigg|$$

is of order $O_p(\max\{K\log n(1/\sqrt{n}+\sqrt{K^3/n}r_n)\},\sqrt{(1+K^3r_n^2)(\sqrt{K}r_nK\log n)})=o_p(1)$ (here we use that $K^4\log n/n\to 0$).

Finally, using the above bounds for R_{i1} and similar arguments, we have $\sum_i R_{i1}(\boldsymbol{\beta},\boldsymbol{\theta})(1/2-I\{e_i\le 0\})=o_p(1)$ which completes the proof. □

Lemma 6.

$$\sup_{\|\boldsymbol{\eta}\|\le Cr_n,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le C/\sqrt{n}}\Bigg|\sum_i E\rho_\tau\left(e_i-\boldsymbol{\Pi}_i\boldsymbol{\eta}-\mathbf{V}_i(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})\right)$$
$$-\sum_i E\rho_\tau\left(e_i-\boldsymbol{\Pi}_i\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\right)-\sum_i\frac{f(0|\mathbf{X}_i,\mathbf{Z}_i)}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\mathbf{V}_i\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\Bigg|$$
$$=o_p(1).$$

Proof. The appearance of both $R_i(\boldsymbol{\beta},\boldsymbol{\theta})$ and $R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})$ also causes some difficulties in proof here. By Knight's identity,

$$\sum_i E\rho_\tau\left(e_i-\boldsymbol{\Pi}_i\boldsymbol{\eta}-\mathbf{V}_i(\boldsymbol{\beta}-\boldsymbol{\beta}_0)-R_i(\boldsymbol{\beta},\boldsymbol{\theta})\right)-\sum_i E\rho_\tau\left(e_i-\boldsymbol{\Pi}_i\boldsymbol{\eta}-R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\right)$$
$$=\int_{\boldsymbol{\Pi}_i\boldsymbol{\eta}+R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})}^{\boldsymbol{\Pi}_i\boldsymbol{\eta}+R_i(\boldsymbol{\beta},\boldsymbol{\theta})+\mathbf{V}_i(\boldsymbol{\beta}-\boldsymbol{\beta}_0)} F(t|\mathbf{X}_i,\mathbf{Z}_i)-F(0|\mathbf{X}_i,\mathbf{Z}_i)\,dt$$
$$=\sum_i\frac{f(0|\mathbf{X}_i,\mathbf{Z}_i)}{2}\left\{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\mathbf{V}_i\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+R_{i1}^2+2R_{i1}\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right)$$

$$
\begin{aligned}
&+2R_{i1}\big(\mathbf{\Pi}_i\boldsymbol{\eta}+R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big)\\
&+2\big(\mathbf{\Pi}_i\boldsymbol{\eta}+R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big)\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\big\}\big(1+o_p(1)\big).
\end{aligned}
$$

We have $\sum_i R_{i1}^2 = O(r_n^2K^2) = o_p(1)$, $(\sum_i R_{i1}\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0))^2 = O_p(r_n^2K^2) = o_p(1)$ (using Cauchy–Schwarz inequality), and $(\sum_i R_{i1}(\mathbf{\Pi}_i\boldsymbol{\eta}+R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})))^2 = (r_n^2K^2)(nr_n^2) = o_p(1)$ (using Cauchy–Schwarz inequality). By the orthogonalization procedure, we used, $\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)\times \mathbf{\Pi}_i\mathbf{V}_i^T = \sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)\mathbf{\Pi}_i(\mathbf{U}_i - \mathbf{\Pi}_i^T(\mathbf{\Pi}^T\mathbf{\Gamma}\mathbf{\Pi})^{-1}\mathbf{\Pi}^T\mathbf{U}) = \mathbf{\Pi}^T\mathbf{\Gamma}\mathbf{U} - \mathbf{\Pi}^T\mathbf{\Gamma}\mathbf{U} = \mathbf{0}$. Thus, we only need to show

$$
\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0) = o_p(1),
$$

with $R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta}) = \sum_j Z_{ij}\mathbf{B}^T(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j} - m_i$.

Note that directly using $|R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})| \le CK^{-d}$ shows that the above displayed equation is of order $O_p(\sqrt{n}K^{-d}) \ne o_p(1)$ unless oversmoothing is used. Our assumptions on K allows the optimal choice $K \sim n^{1/(2d+1)}$ and thus more complicated arguments based on projection is required.

Write $\mathbf{V}_i^T = (\mathbf{U}_i^T - \mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T) + (\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T - E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T]) + (E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T] - \mathbf{P}_i^T\mathbf{U})$, a sum of three terms and we deal with each one separately.

First, by the approximation property of splines,

$$
\begin{aligned}
&\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big(\mathbf{U}_i^T - \mathbf{g}^{(1)}\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X}_i^T\big)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
&\quad = O_p\big(\sqrt{n}K^{-2d}\big)\\
&\quad = o_p(1).
\end{aligned}
$$

Then, using that $E[f(0|\mathbf{X}_i,\mathbf{Z}_i)R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})(\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T - E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T])] = 0$ which follows from our definition of projection, by direct variance calculation,

$$
\begin{aligned}
&\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big(\mathbf{g}^{(1)T}\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X}_i^T - E_{\mathcal{M}}\big[\mathbf{g}^{(1)T}\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X}_i^T\big]\big)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
&\quad = O_p\big(K^{-d}\big) = o_p(1).
\end{aligned}
$$

Finally, by condition (A4), $\|E_{\mathcal{M}}[\mathbf{g}^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\mathbf{Z}\mathbf{X}_i^T] - \mathbf{P}_i^T\mathbf{U}\| = O_p(K^{-d'} + K^{-d+3/2})$ and thus

$$
\begin{aligned}
&\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\big(E_{\mathcal{M}}\big[\mathbf{g}^{(1)T}\big(\mathbf{X}_i^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X}_i^T\big] - \mathbf{P}_i^T\mathbf{U}\big)\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
&\quad = O_p\big(\sqrt{n}K^{-d-d'} + \sqrt{n}K^{-2d+3/2}\big) = o_p(1).
\end{aligned}
$$

Thus, $\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)R_i(\boldsymbol{\beta}_0,\boldsymbol{\theta})\mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0) = o_p(1)$ and the proof is complete. □

Lemma 7.

$$\frac{1}{n}\sum_i f(0|\mathbf{X}_i,\mathbf{Z}_i)\mathbf{V}_i\mathbf{V}_i^T \to E\big[f(0|\mathbf{X},\mathbf{Z})\big(\mathbf{g}^{(1)T}\big(\mathbf{X}^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X} - E_{\mathcal{M}}\big[\mathbf{g}^{(1)T}\big(\mathbf{X}^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X}\big]\big)^{\otimes 2}\big] \quad \textit{in probability},$$

$$\frac{1}{n}\sum_i \mathbf{V}_i\mathbf{V}_i^T \to E\big[\big(\mathbf{g}^{(1)T}\big(\mathbf{X}^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X} - E_{\mathcal{M}}\big[\mathbf{g}^{(1)T}\big(\mathbf{X}^T\boldsymbol{\beta}_0\big)\mathbf{Z}\mathbf{X}\big]\big)^{\otimes 2}\big] \qquad \textit{in probability}.$$

Proof. The left-hand side is $\mathbf{V}^T\boldsymbol{\Gamma}\mathbf{V}/n = \mathbf{U}^T(\mathbf{I}-\mathbf{P}^T)\boldsymbol{\Gamma}(\mathbf{I}-\mathbf{P})\mathbf{U}/n$ where the rows of $\mathbf{U}$ are $\mathbf{U}_i^T = \sum_j Z_{ij}\mathbf{B}_k^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}\mathbf{X}_i^T$. Let $\mathbf{U}^*$ be defined similarly as $\mathbf{U}$ with $\mathbf{B}_k^{(1)T}(\mathbf{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{\theta}_{0j}$ replaced by $g_j^{(1)}(\mathbf{X}_i^T\boldsymbol{\beta}_0)$. By the approximation property of splines $\|(1/n)\mathbf{U}^{*T}(\mathbf{I}-\mathbf{P}^T)\boldsymbol{\Gamma}(\mathbf{I}-\mathbf{P})\mathbf{U}^* - (1/n)\mathbf{U}^T(\mathbf{I}-\mathbf{P}^T)\boldsymbol{\Gamma}(\mathbf{I}-\mathbf{P})\mathbf{U}\|_F = o_p(1)$ and then using the same arguments as in Lemma 1 of Wang, Zhu and Zhou [25]. The second expression is proved in the same way. □

Proof of Theorem 2. Let $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 + (\boldsymbol{\Pi}^T\boldsymbol{\Gamma}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^T\boldsymbol{\Gamma}\mathbf{U}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. By Lemmas 5, 6, and 7,

$$\begin{aligned}\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le C/\sqrt{n}}\bigg|&\sum_i \rho_\tau\big(e_i - \boldsymbol{\Pi}_i^T\hat{\boldsymbol{\eta}} - \mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})\big)\\ &- \sum_i \rho_\tau\big(e_i - \boldsymbol{\Pi}_i^T\hat{\boldsymbol{\eta}} - R_i(\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}})\big) + \sum_i \mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\varepsilon_i \\ &- \frac{n}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{\Phi}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\bigg| = o_p(1).\end{aligned} \tag{20}$$

Let $Q(\boldsymbol{\beta}) = \frac{n}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{\Phi}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) - \sum_i \mathbf{V}_i^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\varepsilon_i$ and define $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (1/n)\boldsymbol{\Phi}^{-1}\sum_i \mathbf{V}_i^T\varepsilon_i$. We have by central limit theorem

$$\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\big(0, \boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\big).$$

Note $\tilde{\boldsymbol{\beta}}$ is the minimizer of $Q(\boldsymbol{\beta})$ and $Q(\boldsymbol{\beta})$ is equal to $(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})^T\boldsymbol{\Phi}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})$ plus a term that does not involve $\boldsymbol{\beta}$. Define

$$\tilde{\tilde{\boldsymbol{\beta}}} := \mathop{\arg\min}_{\|\boldsymbol{\beta}\|=1,\beta_1>0} (\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})^T\boldsymbol{\Phi}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}).$$

By Proposition 4.1 of Shapiro [24] which works for overparametrized models (considering $\boldsymbol{\beta}$ as a function of $\boldsymbol{\beta}^{(-1)}$ and the parametrization using $\boldsymbol{\beta}$ is an overparametrization), we get that

$$\sqrt{n}(\tilde{\tilde{\boldsymbol{\beta}}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\big(0, \mathbf{J}\big(\mathbf{J}^T\boldsymbol{\Phi}\mathbf{J}\big)^{-1}\mathbf{J}^T\boldsymbol{\Sigma}\mathbf{J}\big(\mathbf{J}^T\boldsymbol{\Phi}\mathbf{J}\big)^{-1}\big)\mathbf{J}^T.$$

For any $\boldsymbol{\beta}$ with $\|\boldsymbol{\beta}\| = 1$ and $\|\boldsymbol{\beta} - \tilde{\tilde{\boldsymbol{\beta}}}\| = \delta/\sqrt{n}$ with some small $\delta > 0$. By the quadratic form of Q, we have

$$Q(\boldsymbol{\beta}) - Q(\tilde{\tilde{\boldsymbol{\beta}}}) \geq C\delta^2$$

and thus by (20),

$$\begin{aligned} P\Bigg(& \sum_i \rho_\tau\big(e_i - \boldsymbol{\Pi}_i^T \hat{\boldsymbol{\eta}} - \mathbf{V}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - R_i(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})\big) \\ & > \sum_i \rho_\tau\big(e_i - \boldsymbol{\Pi}_i^T \hat{\boldsymbol{\eta}} - \mathbf{V}_i^T(\tilde{\tilde{\boldsymbol{\beta}}} - \boldsymbol{\beta}_0) - R_i(\tilde{\tilde{\boldsymbol{\beta}}}, \hat{\boldsymbol{\theta}})\big)\Bigg) \\ & \to 1. \end{aligned}$$

Since δ is arbitrarily small, we get $\|\hat{\boldsymbol{\beta}} - \tilde{\tilde{\boldsymbol{\beta}}}\| = o_p(1/\sqrt{n})$ and the proof is complete. □

Acknowledgements

We are grateful to the insightful comments and suggestions made by Profesor Eric Moulines, an Associate Editor and the anonymous referees. The research of Weihua Zhao was partially supported by National Social Science Fund of China (15BTJ027). Liang's research was partially supported by NSF grant DMS-14-18042, and by Award Number 11529101, made by National Natural Science Foundation of China.

References

[1] Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *J. Econometrics* **167** 413–425. MR2892085

[2] Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.* **103** 1595–1608. MR2504207

[3] Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25** 715–744. MR1439320

[4] Chen, R. and Tsay, R.S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88** 298–308. MR1212492

[5] Cui, X., Härdle, W.K. and Zhu, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39** 1658–1688. MR2850216

[6] Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 57–80. MR1959093

[7] Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 303–322. MR1749541

[8] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497

[9] Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** 757–796. MR1229881

[10] He, X. and Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Anal.* **58** 162–181. MR1405586
[11] 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
[12] Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Amer. Statist. Assoc.* **87** 58–68.
[13] Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699
[14] Horowitz, J.L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100** 1238–1249. MR2236438
[15] Huang, J.Z., Wu, C.O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111–128. MR1888349
[16] Huang, Z. and Zhang, R. (2013). Profile empirical-likelihood inferences for the single-index-coefficient regression model. *Stat. Comput.* **23** 455–465. MR3070403
[17] Kim, M.-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35** 92–108. MR2332270
[18] Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644
[19] Li, Q. (2000). Efficient estimation of additive partially linear models. *Internat. Econom. Rev.* **41** 1073–1092. MR1790072
[20] Lian, H. (2012). Variable selection for high-dimensional generalized varying-coefficient models. *Statist. Sinica* **22** 1563–1588. MR3027099
[21] Ma, S., Liang, H. and Tsai, C.-L. (2014). Partially linear single index models for repeated measurements. *J. Multivariate Anal.* **130** 354–375. MR3229543
[22] Ma, S. and Song, P.X.-K. (2015). Varying index coefficient models. *J. Amer. Statist. Assoc.* **110** 341–356. MR3338507
[23] Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26. MR0184378
[24] Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *J. Amer. Statist. Assoc.* **81** 142–149. MR0830574
[25] Wang, H.J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37** 3841–3866. MR2572445
[26] Wang, L., Liu, X., Liang, H. and Carroll, R.J. (2011). Estimation and variable selection for generalized additive partial linear models. *Ann. Statist.* **39** 1827–1851. MR2893854
[27] Wang, L. and Yang, L. (2009). Spline estimation of single-index models. *Statist. Sinica* **19** 765–783. MR2514187
[28] Wei, F., Huang, J. and Li, H. (2011). Variable selection and estimation in high-dimensional varying-coefficient models. *Statist. Sinica* **21** 1515–1540. MR2895107
[29] Wei, Y. and He, X. (2006). Conditional growth charts. *Ann. Statist.* **34** 2069–2131. MR2291494
[30] Xia, Y. and Li, W.K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94** 1275–1285. MR1731489
[31] Xue, L. and Qu, A. (2012). Variable selection in high-dimensional varying-coefficient models with global optimality. *J. Mach. Learn. Res.* **13** 1973–1998. MR2956349
[32] Xue, L. and Wang, Q. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* **18** 836–856. MR2948904
[33] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628
[34] Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. MR1951258

Received April 2015 and revised October 2015